Everything You Need To Know About Undergraduate Algebra

Nathanael Chwojko-Srawley

January 6, 2025

Contents

Preliminary Chapter **i**

0.1 Sets i

0.1.1 Propositions vi

0.1.2 Induction xii

0.1.3 *Pigeon-hole Principle xii

0.1.4 Relations and Equivalence Relations xii

0.1.5 Functions xv

0.1.6 Adding Structures to Sets xxv

0.1.7 *Cardinality xxvi

0.2 Number Theory xxvii

0.2.1 GCD and Euclidean Algoritm xxvii

0.2.2 Modular Arithmetic's xxxi

0.2.3 Counting Problems xxxiv

0.3 Graphs xxxv

0.3.1 Lattices xxxvi

0.4 Categories xxxvii

0.4.1 Defining Categories xxxviii

0.5 Matrices xli

I Groups **1**

1 Introduction to Groups **5**

1.1 Basic Properties and Definitions . . . 5
1.1.1 Consequences of Group Axioms . . . 13
1.2 Examples of Groups . . . 18
1.2.1 Integers Modulo n and Cyclic Groups . . . 19
1.2.2 Symmetric Groups . . . 24
1.2.3 Dihedral Groups . . . 30
1.2.4 Quaternion Group . . . 33
1.2.5 *Matrix Groups . . . 35
1.3 Generators and Relations . . . 37
1.4 Homomorphisms and Isomorphisms . . . 40
1.4.1 Homomorphisms . . . 41
1.4.2 Isomorphisms . . . 50
1.5 Group Actions . . . 55
1.6 Cayley Graphs . . . 57

2 Sub Groups and Generating Groups 61
2.1 Basic Definition and Properties . . . 61
2.2 Examples of Subgroups . . . 63
2.2.1 Kernel and Image . . . 63
2.2.2 Center, Centralizers, and Normalizers . . . 65
2.3 Subgroups Generated by Subsets . . . 70
2.3.1 Cyclic Groups and Cyclic subgroups . . . 70
2.3.2 Generating Subgroups . . . 77
2.3.3 Free Groups . . . 80
2.3.4 Free Abelian Group . . . 88
2.3.5 *Free-Forgetful Adjoint . . . 89
2.4 Subgroup Lattices . . . 89
2.5 *Universal Property of kernel and coKernels . . . 93

3 Homomorphisms and Quotient Groups 97
3.1 Quotient Groups . . . 98
3.2 Presentations . . . 110
3.3 Application to Finite Groups . . . 112
3.4 Internal Direct Product . . . 117
3.5 Isomorphism Theorems . . . 120
3.6 Extensions . . . 127

3.7 Classifying all Finite Groups: Hölder Program . . . 130
3.8 Quotient Objects . . . 135

4 Group Actions 137
4.1 Defining Group Action . . . 137
4.2 Axiomatic representation . . . 140
4.2.1 A Word on Right-actions . . . 143
4.3 Orbit Stabalizer Theorem . . . 148
4.4 Application to Finite Groups . . . 149
4.4.1 Left Multiplication . . . 149
4.4.2 Conjugation . . . 154
4.4.3 Class Equation . . . 155
4.4.4 Sylow's Theorem . . . 159
4.4.5 Applications of Sylow's Theorem . . . 163
4.5 Actions on Groups . . . 164
4.5.1 A Word on Characteristic Groups . . . 167
4.6 Actions in general . . . 169

5 Composing Groups: Direct and Semi-Direct Products 171
5.1 Classical Definitions And Properties . . . 172
5.1.1 Fundamental theorem of Finitely Generated Abelian Groups . . . 174
5.2 Recognizing Direct Products . . . 175
5.2.1 A word on Perfect Groups . . . 181
5.3 Semi-Direct Products . . . 181
5.4 A Categorical approach of the Product . . . 188
5.4.1 The dual of the Product . . . 191
5.4.2 Universal Property of Semi-Direct Products? . . . 194

6 Further Topic in Group Theory 195
6.1 Generalization of Abelian Groups . . . 195
6.1.1 Solvable Groups . . . 197
6.1.2 Nilpotent Groups . . . 200
6.1.3 Polycyclic Groups . . . 204
6.2 Non-finitely generated abelian groups . . . 205
6.3 Linear Groups . . . 205

7 Generalization of Groups 207

7.1 Monoid 207
7.2 Groupoid 209

8 Category of Group **211**
8.1 Basic properties of Morphisms 217
8.2 Commutative diagrams 221
8.3 More abstract examples 222
8.4 Summary 222

II Rings **223**

9 Introduction To Rings **227**
9.1 Basic Properties and Definitions 228
9.1.1 Category of Ring 242
9.1.2 Summary 245
9.2 Examples of Rings 246
9.2.1 Polynomial Rings 246
9.2.2 *Formal Power Series 250
9.2.3 Matrix Rings 255
9.2.4 Group Rings 256
9.3 Ring Homomorphisms and Quotient Rings 258
9.4 Properties and Further Structures of Ideals 267
9.4.1 Operations and Generation on Ideals 267
9.4.2 Ideals and Units 273
9.4.3 Maximal Ideals 274
9.4.4 Prime Ideals 277
9.5 Ring of Fractions: Localization 280
9.5.1 Constructing the Reals? 285
9.6 Chinese Remainder Theorem 285
9.7 *Monoidal Objects 288

10 Exploring Integral Domains: Factorization and Valuation **289**
10.1 Factorization 291
10.1.1 Example of Atomic Domains: Noetherian Domains 297
10.1.2 Unique Factorization Domains 302
10.1.3 Important UFD: Principle Ideal Domains 306

10.2 Euclidean Domains . . . 310
10.3 Application: Sums of Primes and Gaussian Integers . . . 318

11 Exploring Polynomial Rings **325**
11.1 Polynomial Rings over Fields . . . 326
11.2 Polynomial Rings over UFD's . . . 329
11.3 Irreducibility of Polynomials . . . 332

III Modules **341**

12 Introduction to Modules **345**
12.1 Basic Definitions and Examples . . . 346
12.1.1 Vector-spaces . . . 351
12.1.2 Algebras . . . 354
12.1.3 Abelian Groups as Modules . . . 357
12.2 Homomorphism and Quotients of Modules . . . 359
12.2.1 Extensions and Exactness . . . 363
12.3 Generation of Modules . . . 364
12.3.1 Direct Sum of Modules . . . 366
12.3.2 Free Modules . . . 372
12.3.3 Presentations and Resolutions . . . 381
12.3.4 Finitely Generated Modules and Noetherian Modules . . . 384
12.3.5 Direct Sums and Homomorphisms . . . 387

13 Modules over Fields – Vector Spaces and Linear Algebra I **391**
13.1 Vector Spaces and their Properties . . . 393
13.1.1 Quotient Spaces and Subspaces . . . 401
13.2 Linear Transformations and Matrix Representation . . . 406
13.2.1 Change of Basis . . . 411
13.2.2 Similarity . . . 413
13.2.3 Gaussian Elimination: Finding a Nice Basis . . . 415
13.2.4 Geometric Result: Systems of Linear Equations . . . 420
13.2.5 Use Case: Reading Presentations . . . 425
13.3 Dual Spaces and Forms . . . 426
13.3.1 Dual Spaces . . . 427
13.3.2 Multilinear Forms . . . 436

13.4 Invariants of Linear Operators . . . 439
13.4.1 Trace and Determinant . . . 439
13.4.2 Characteristic Polynomial and Eigenvectors . . . 443
13.5 Geometry and Inner Product . . . 450
13.5.1 Basic Properties and Definitions . . . 451
13.5.2 Adjoint Operators in Inner Product Spaces . . . 460
13.5.3 Self-Adjoint and Normal Operators . . . 465

14 Modules Over PID's: Linear Algebra II 469
14.1 Over PID, Submodule of Free Module is Free . . . 470
14.2 Application: Linear Algebra revisited . . . 481
14.3 Rational and Jordan Canonical Form . . . 484
14.3.1 Converting a square matrix to Rational Canonical Form . . . 489
14.4 Jordan Canonical Form . . . 492
14.4.1 Changing From one Canonical Form to Another . . . 495
14.4.2 Computation: Convert to JCF . . . 495

15 Tensor Products Of Modules: Linear Algebra III 497
15.1 What's a Tensor Product? . . . 497
15.1.1 Constructing the Tensor Product . . . 499
15.2 Properties of Tensor Products . . . 506
15.3 Functors Between Mod and Scalar Extension . . . 509
15.3.1 Universal property of $S \otimes_R N$ . . . 512
15.4 Tensor Algebras and Graded Algebras . . . 515
15.4.1 Symmetric and Exterior Algebras . . . 520
15.4.2 Symmetric and Alternating Tensors . . . 528
15.5 Monoidal Categories . . . 530

16 Extension Problem – Prelude to Homological Algebra 531
16.1 Exact Sequences . . . 533
16.1.1 Relating Short Exact Sequences . . . 536
16.1.2 Splitting Sequences . . . 539
16.2 Projective Modules . . . 541
16.3 Injective Modules . . . 550
16.3.1 Example: Divisible Groups . . . 556
16.3.2 Injective Hull . . . 561
16.3.3 Duality on R-modules . . . 563

16.4 Flat Modules . . . 564
16.4.1 Flat Extensions . . . 568
16.5 Ext and Tor . . . 568
16.6 Summary . . . 570

IV Fields **571**

17 Field Theory **575**
17.1 Properties of Fields . . . 575
17.2 Field Extensions . . . 582
17.2.1 Isomorphism of Fields Extensions . . . 593
17.3 Algebraic Extensions . . . 599
17.4 Historical Detour: Straightedge and Compass Construction . . . 608
17.5 Closures of Fields . . . 609
17.5.1 Algebraic Closure . . . 609
17.5.2 Splitting Field and Normal Extensions . . . 614
17.6 Separable Extensions . . . 622
17.6.1 Separability and Embeddings . . . 628
17.6.2 Separability and Simple Extensions . . . 632
17.7 Cyclotomic Polynomials and Finite Fields . . . 635
17.7.1 Cyclotomic Polynomials and Fields . . . 635
17.7.2 Finite Fields . . . 640

18 Galois Theory **645**
18.1 Galois Extension and Galois Group . . . 647
18.1.1 Galois Correspondence . . . 655
18.1.2 Composite Fields . . . 666
18.1.3 A bit of Character Theory . . . 669
18.2 Application of Galois Theory . . . 672
18.2.1 Fundamental Theorem of Algebra . . . 672
18.2.2 Construction of Regular n-gon . . . 674
18.2.3 Symmetric Functions . . . 674
18.2.4 Solvability By Radicals . . . 675
18.3 Galois Groups of Polynomials . . . 679
18.4 *Infinite Galois Extensions . . . 681
18.4.1 Profinite Groups . . . 682

18.4.2 Infinite Galois Theory 683

19 *More Structure on Fields **689**

19.1 Ordered and Real Fields 689

19.2 Absolute Values on General Fields 693

19.2.1 Extending Valuation 698

19.2.2 Ostrowski Theorem 698

V Commutative Algebra **699**

20 More on Ring Theory, Ring Extensions, and Category Theory **703**

20.1 Categorical (co)Limits, Natural Transformations, and Adjoints 703

20.2 Noetherian Rings and Hilbert's Basis Theorem (HBT) 714

20.2.1 Gröbner Basis 719

20.3 Localization 719

20.3.1 Localization of Rings 720

20.3.2 Localization of Modules 726

20.3.3 Local Properties 728

20.3.4 Local Rings 731

20.3.5 Nakayama's Lemma 734

20.4 Completion 740

20.4.1 Constructing Completions 741

20.4.2 Filtrations 747

20.4.3 More on Graded Algebras 748

20.4.4 Artin-Rees Lemma 750

20.4.5 Associated Graded Rings: Noetherian Closed under Completion 755

20.5 Ring Extension and Integral Extension 758

20.5.1 Extension and Contraction of Ideals 763

20.5.2 Integral Extensions and Prime Ideals 766

20.6 Valuation Ring and Local Rings 768

20.6.1 Discrete Valuation Ring 771

20.7 Artinian Rings 773

21 Relating Algebra to Geometry **779**

21.1 Affine Algebraic Sets 781

21.1.1 Affine Sets 782

21.1.2 Zariski Topology . . . 792
21.2 Properties of Affine Sets and Duality . . . 797
21.2.1 Regular Functions and Duality . . . 797
21.2.2 Closed Sets and Radical Ideals . . . 801
21.2.3 Affine Variety and Prime Ideals . . . 804
21.2.4 Computation of Algebraic Sets and k-algebras . . . 809
21.2.5 Hilbert's Nullstellensatz . . . 810
21.3 Rational Maps: Classifying Varieties . . . 816
21.4 Projective Variety . . . 823
21.4.1 Definitions and Properties . . . 826
21.4.2 Properties of Projective space: Closure, Covering, Coordinates . . . 835
21.4.3 Regular Maps . . . 839
21.4.4 Products and Mappings . . . 844
21.4.5 Bézout's Theorem . . . 854
21.5 Dimension Theory . . . 855
21.5.1 Krull dimension and Transcendental Degree . . . 857
21.6 *Local Dimension: Hilbert Function . . . 860
21.6.1 Dimension of Noetherian Local Ring . . . 867
21.6.2 Smoothness: Regular Local Rings . . . 872
21.7 *Prelude to Algebraic Geometry . . . 874

22 Relating Algebra to Number Theory 883
22.1 Ring of Integers . . . 885
22.1.1 Functionals: Field Trace and Field Norms . . . 889
22.1.2 Functionals: Discriminant . . . 895
22.1.3 Abelian Group Structure . . . 902
22.1.4 Lattice Structure: Geometry of Number Rings . . . 912
22.2 Ring Structure and Dedekind Domains . . . 923
22.2.1 Fractional Ideals in Rings . . . 927
22.2.2 Prime Decomposition in Dedekind Domain . . . 931
22.2.3 Class Group of Ring of Integer . . . 937
22.2.4 Dirichlet's Unit Theorem . . . 944
22.3 Extension of Dedekind Domains and Number Fields . . . 947
22.3.1 Ideal Correspondence in Extension . . . 951
22.4 Prime Factorization in Extension . . . 958
22.4.1 Application: Quadratic Extensions and Quadratic Reciprocity . . . 968

22.4.2 Example: Eisenstein Polynomials . . . 972
22.4.3 Primes in Galois Extensions . . . 975
22.4.4 Example: Cyclotomic Fields . . . 993
22.5 Application: Kummer Theory and Kronecker Weber . . . 998
22.5.1 Kummer Theory . . . 998
22.5.2 Kronecker-Weber Theorem . . . 1003
22.6 *Modules Over Dedekind Domains . . . 1009
22.7 *Geometric Idea: 1-Dimensional Schemes . . . 1011
22.7.1 Localization of Dedekind Domains . . . 1011
22.7.2 Orders: Curves with Singularities . . . 1011

VI Representation Theory 1013

23 Semisimple Rings and Modules 1017
23.1 Decomposing Modules . . . 1018
23.2 Artin-Wederburn . . . 1027

24 Representation of Finite Groups 1039
24.1 Representations of Finite Groups . . . 1039
24.2 Character Theory . . . 1052
24.2.1 More Character Tables . . . 1070
24.2.2 Burnside's Theorem . . . 1073
24.3 Induced Representation . . . 1075
24.4 *Application to Harmonic Analysis . . . 1081

25 Representation of Lie Algebra's and Quiver Theory 1083
25.1 Lie Algebras . . . 1084
25.2 Quiver's and their Decomposition . . . 1084
25.2.1 Quiver Representation and Varieties . . . 1086
25.2.2 Decomposing Quivers . . . 1090
25.2.3 Root Systems . . . 1092
25.2.4 Gabriel's Theorem . . . 1097
25.2.5 Path Algebra of Quivers . . . 1103

VII Homological Algebra 1107

26 Abelian Categories 1113

26.1 Category of Categories . . . 1118

27 Complexes and (co)Homology 1119

27.1 Basic Definition and Properties . . . 1120

27.1.1 Homology and Cohomology . . . 1123

27.1.2 Snake Lemma . . . 1127

27.1.3 Chain Homotopies and Homotopic Categories . . . 1131

27.2 Projective and Injective Resolutions . . . 1140

27.2.1 Resolutions . . . 1146

27.2.2 Derived Category with Enough Projectives . . . 1149

27.3 Derived Functors: Ext and Tor . . . 1152

27.3.1 Derived Functor . . . 1153

27.3.2 Long Exact Sequences of Derived Functors . . . 1155

27.4 Double Complexes . . . 1165

27.4.1 Applications to Computations of Ext and Tor . . . 1172

27.5 *Formal Derived Category and Triangulated Category . . . 1174

27.5.1 Derived Categories, Formally . . . 1174

27.5.2 Triangulated Categories . . . 1175

28 Applications of (co)Homology 1179

28.1 Group Cohomology . . . 1179

28.2 K-Theory . . . 1186

28.3 Homological Dimension . . . 1186

29 Spectral Sequences 1187

29.1 Definition and Examples . . . 1188

VIII Category Theory 1191

30 Summary of Category Theory 1195

30.1 Duality . . . 1201

30.2 Functoriality . . . 1205

30.3 Natural Transformations . . . 1209

30.4 Universal Property . . . 1212

30.5 Adjoint 1214
30.6 A Second Look at Category theory 1214
30.7 Commutative Diagrams 1216

IX Appendix 1219

A Axiom of Choice 1223
A.1 Build-up 1224
A.2 Axiom of Choice Equivalences 1227
A.3 Zorn's Lemma and Transfinite Induction 1231

B Sub-Group of a Free Group is Free 1233
B.1 Using Cosets 1233
B.2 Using Tree's 1235

C Basic Topology and Topological Groups 1237
C.1 Definition and Properties 1237
C.2 Construction Topologies 1243
C.3 Compactness and Connectedness 1244
C.4 Topological Groups 1245
C.5 Completion in Topology 1247
C.5.1 Classifying Complete Archimidean Fields 1250

D Lattice Theory and Universal Algebras 1251
D.1 Lattice Theory 1251
D.2 Pointless Topology 1260
D.3 Universal Algebras 1260

E Relating Algebra to Invariant Theory 1261

F Model Theory 1263
F.1 Löwenheim-Skolem Theorem 1263

X Glossary and selected Answers 1265

G Glossary of Algebraic Structures 1267
G.1 Groups 1267

G.2 Rings 1267
G.3 Modules 1268
G.4 Universal Properties 1268
G.5 Character Tables 1268

The title of the book says it all: the intention is that you will find "everything you need to know about undergraduate algebra" in these many pages. I chose the word "everything" purposefully knowing it is a contentious word, especially in the range of material that is considered at the undergraduate level. In some universities, only the first few chapters of this book are considered part of the undergraduate curriculum, while at others the entirety of the books content and more is covered in the undergraduate program. Thus, parameters had to be set in choosing the material of this book while trying to remain true in using the word "everything" – they are the following:

1. By the end of the book, or any part of the book (ex. part on Groups, Rings, Modules, etc.), the reader should have had a deep enough understanding and intuition so as to be able to tackle graduate level mathematics. The reader should be comfortable with picking up almost any graduate level book in algebra, including topics like algebraic number theory, algebraic geometry, representation theory, lie theory, classical or geometric group theory, ring theory, homological algebra, and so forth.

2. By the end of the book, the reader should be comfortable to think in any (or at least most) major algebraic studies, to be able to more than just recite definitions and systematically follow proofs but to have a connection with the definition and the way of thinking that allowed the mathematician to come up with the proofs and ideas, which will hopefully encourage the reader to start thinking like a mathematician. The goal is that the reader has a deep pool of examples, mental structures, and intuitions that they can draw on to come up with their own ideas.

3. By the end of the book, the reader should get a sense of how interconnected mathematics is. There is over a thousand cross-references in the book interconnecting all parts of it together, and a mathematical language will be introduced early on that will allow the reader to study the connection between objects.

4. By the end of this book (or any section), the reader should be comfortable solving problems (or problems from that section). To achieve this, many exercises are added throughout the end of sections in the book. Doing as many as necessary to get comfortable with the material is very important; reading a math book and not doing the exercises is like reading a cook book without cooking – you won't get to taste the result. It is common to have a fake sense of understanding or a lack of intuition if exercises are skipped (I say this out of personal experience and that of many of the colleagues I've worked with).

Thus, in order to achieve these goals, the book is on the larger side of textbooks, and covers most of "core algebra" of undergraduate curriculums as well as most "optional" courses in algebra many undergraduate curriculum's offer. To insure the book stays within the scope of undergraduate knowledge, only material that I have personally covered or have learned about during my undergraduate years are included. Since algebra is my favourite topic, I've taken as many optional algebra courses and read as many books as possible on the subject, and so I wish to think that I've covered enough of a broad range of topics in great enough depth to use my personal experience to be a benchmark for the quantity of material[1]. There is also the question of what should be assumed to be the starting point. The solution I've done is the same as many authors solutions: have a preliminary chapter that assumes only an elementary level of knowledge of mathematics (for this book, knowledge of arithmetic's and factorization) and cover all necessary prerequisite material that will be used in the book. The preliminary chapter in this book is more extensive than most preliminary chapter to

[1]This also means that I chose to start considering some material as beyond the scope of undergraduate algebra, such as scheme theory and class field theory

insure the reader has no "knowledge gaps". In any case, I hope that the reader will find this textbook a useful accompaniment throughout their undergraduate studies in algebra.

Properties of the Book

The book covers a great multitude of topics and structures (there are at least 21 different structures in this book, including groups, rings, modules, sets, graphs, and so forth), which not only overlap in many important ways but also constantly interconnect, describe, and influence each other. Part of the goal of this book is to show how interconnected mathematics and algebra is, and so to accomplish this goal a light-version of category theory is introduced early on and gradually built upon throughout the book. It is the opinion of this author that shying away from all category theory (as some textbooks I've read have done) delays many interesting and important connections between so many parts of algebra, and gets rid of the opportunity to foreshadow some amazing results[2]. Naturally, the full generality of category theory is not needed. Similarly to how set theory is not introduced in terms of the Zermelo-Frenkel axioms but as naïve set theory, so too is category theory introduced in a simplified version, giving a sufficient amount of language and tools to study how mathematical objects interconnect while not overloading on the generalities and technicalities. Some key tools and vocabulary that will be introduced early on is that of universal properties and functors. Throughout the book, there are hundreds of instances of universal properties and functors being used, showing how pervasive of a tool they are in algebra and mathematics as a whole. Gaining an intuition by seeing these concepts naturally appear gradually but early is useful to gain deeper insights.

Each part and chapter of the book will start off with some motivation for the material. To accommodate many different learning styles, many forms of motivation are given, from historical context, to practical application, to important consequences, to how the material is connected to other areas of mathematics. The idea is to have the reader already start making connections while reading the main material. This is an approach that contrasts with Bourbaki which aims to present material minimally and coarsely, omitting any of the authors intuitions or insights and presenting an encyclopedia of mathematics. This book is not intended to be an encyclopedia for senior mathematicians to double-check their understanding of some concepts, but to be a fun and interesting yet deep and insightful pedagogical tool to encourage the reader to build up their understanding and make as many connections as possible and challenge ways of thinking by actively engaging in the interpretation of the material. Though the intuitions of this author will be present, it is my hope that the book will be written in such a way that the reader will develop their own intuition, solidifying in their own way what the material of this book has to offer.

This book is an algebra textbook, and so avoids covering material outside of algebra. However, since mathematics is a very interconnected discipline, it is sometimes impossible to avoid talking about other domains of mathematics. As an example of the interconnectedness of mathematics, many definitions will have examples that lie in different domains of mathematics. Thus, it won't be uncommon that examples of definitions will be given that are not strictly algebraic, however there will always be an algebraic example accompanying a non-algebraic example, since the book should have no prerequisites besides what is covered in the preliminary chapter. It will be necessary to cover some non-algebra material at the preliminary stage and late stage of the book. For the preliminary stage, it will be necessary to cover some basics of [naïve] set theory, number theory, graph theory, and logic. Elementary concepts not needed for algebra (such as supremum, limits, completeness

[2]For example, Cayley's theorem, the existence of a homomorphism from any ring R to $\mathrm{End}_{\mathbf{Grp}}(R)$, or representation theory in general are all naturally connected via Yonneda's lemma. Another example would be the plethora of adjoints

of $\mathbb{R}$) are not covered, and concepts that play a tangential role (such as cardinality) will be left as optional. In the later stages of the books (starting near the end of part IV on page 681), topology and geometry will start getting a more prominent role. Since these topics are so far into the book, it is perhaps not too wrong to think that the reader has a great interest in mathematics as a whole and has already covered some basics of topology, or perhaps even differential geometry. Looking through the curriculum of many universities, it seems that differential geometry is usually covered quite late, while topology is either taught as part of a 2nd or 3rd year analysis or as a separate course. Thus, the book will not rely on geometric thinking, however it will start to casually incorporate topological thinking and results to relevant theorems (which will be relatively few), starting in part V. A review of topology is given in appendix C. Finally, some intuition on the notion of the real numbers $\mathbb{R}$ and the Pythagorean theorem shall start appearing at chapter 13 for very few results.

A convention that is prominent in this book is that square brackets around words indicate that the word can be used to eliminate ambiguity. For example, we will define functions in the preliminary chapter. But in other domain of mathematics, the term function is used differently (ex. the definition of *function* in differential geometry would be defined as *real-valued function* in the preliminary chapter). Thus I will put a word in square parenthesis to show that the word can be added to the term to eliminate ambiguity. For example, though we will always say *function* throughout this book, where it might be ambiguous in general mathematical contexts and in the definition it will be written as *[set] function*.

Another convention will be that sections that have a starred in front of their title are consider as "optional". This is material that does not fit in the appendix, which is material that fills in more advanced gaps for the curious reader (ex. properties of the Axiom of Choice, or topological interpretations of some results), while at the same time not being central to the main narratives of the chapter. They were included as side-quests for the interested readers, but the reader who isn't interested can comfortably skip the section without worry of it being referenced later in the book. The exception to this referencing rule is the starred sections in the preliminary chapter; these shall only be referenced for a hand-full of proofs in the book.

Finally, as the reader goes through the book, they will gain skills and intuitions for solving many problems. Many of these skills will come from reading through all the proofs of the results, as well as the reader doing the exercises. Eventually, many of the results boxed as propositions, lemmas, and corollaries will stop being proved in detail and have part or all of it left as exercises (starting around 300 pages into the book). This pattern shall become more prominent as the reader progresses through the book and gains "mathematical maturity". This pattern reflects the changing pedagogical need of the reader as they progress through the book; it shall become more fruitful for the reader to prove results on their own rather than having the detail laid out. In fact, at that stage of the book the reader may even find that it is easier to prove the results themselves, or feel it is cumbersome to read through a proof they can do themselves[3].

What is [Abstract] Algebra?[4]

Algebra as a field of study originated in the study of numbers and equations that can be formed through the elementary arithmetic operations: addition, subtraction, multiplication, and division $(+, -, \times, \div)$. As mathematics has evolved, it was found that there exist many instances where we

[3] I, the author of this book, have had this experience the more mathematics I studied

[4] From a non-historical perspective

can find new forms of arithmetic on objects that are different from numbers with new operations. A simple example can be given by a square with corners distinctly labeled. We can rotate the square by $\pi/2$ radians (90 degrees). Doing four $\pi/2$ rotations shall result in the square with its original orientation. If we think of the symbol e as representing the original square in its original orientation, r the square rotated by $\pi/2$ radians, r^2 the square rotated by $2\pi/2$ radians, r^3 the square rotated by $3\pi/2$ radians, and r^4 the square rotate by $4\pi/2$ radians, we can imagine that $r^4 = e$. As a mathematician, we may consider this equation as a property of the "arithmetics of rotation" of a square. A similar example would be the hands of a clock: a clockhand rotates by $\pi/6$ radians every hour, and after 12 hours shall come back to its original position (as an equation, $r^{12} = e$), giving a form of "clock arithmetics". As mathematicians studied new systems, a huge quantity of new "arithmetics" where found. A list of some of the first new arithmetics systems that were found are:

1. The set of matrices can be added and multiplied
2. The set of networks (or direct graphs) can be combined to form larger graphs (in this case, the operation is the direct product)
3. Functions can be composed (In this case, the operation is composition)
4. Sets can be unioned and intersected (in this case, the operation here is set-union and set-intersection)
5. vectors can be added and multiplied by a scalar (in this case, the operations are vector addition and scalar multiplication)
6. n-gons can be rotated by multiple of $2\pi/n$ radians to get back a square that is indistinguishable from the original square (in this case, the operation is rotation by $2\pi/n$)
7. Two paths $\alpha, \beta \subseteq \mathbb{R}^2$ whose endpoints match can be concatenated (in this case, the operation is path-concatenation)

This is a short list of thousands of new "arithmetics" that were found in a huge multitude of fields (from physics, to engineering, to linguistics, to geometry, and much more), many of which share commonalities in their properties, their "algebra". A new unified way of studying these systems had to put in place in order to systematically study these systems and better understand the rules manipulating these objects, the *algebra* of these systems.

The solution was to *abstract* a list of properties these systems share and study these properties. The abstraction was called an *algebraic structure*. Any algebraic structure consists of:

- a set of *objects* and
- a set of *operations* (i.e. functions) which must follow
- certain set of *axioms* (assumed rules they must follow)

All the examples that were given above are algebraic structures, for they may all be represented as a set along with an operation or operations. Many of the operations in the above examples follow different axioms. Mathematicians have classified and given names to different types of algebraic objects depending on which set of axioms they follow. For example, the integers (with addition and multiplication) is an example of an algebraic structure called a *commutative ring*, while n-gons (with rotation by $2\pi/n$ radians) are an example of an algebraic structure called a *group*.

Abstract Algebra is the study of *family* or *types*[5] algebraic structures. As a consequence, in abstract algebra, we do not study a particular operation or "arithmetic system" such as the numbers with addition, rotations of a square, or any other given in the examples, but study operations that follow certain rules called *axioms*. This is the core idea behind abstract algebra: a new type algebraic structure is defined by either abstracting desirable properties from a collection of algebraic objects or creating axioms which are deemed representative of the studied structures. The advantage of this abstraction is we can study the property of the operations directly without any "noise" that may come from other properties of objects, giving us insights on all systems which can be represented an algebraic structure of a certain type. In this book, there shall be over 20 types algebraic structures that shall be studied, each at varying degrees of depth depending on their importance. Some of these elementary structures are, in the order presented in the book:

groups, group-actions, rings, fields, modules,
vector-spaces, [associative] algebras, chain complexes

Some other important objects that appear less often are (in no particular order):

monoids, groupoids, Lie algebras, inner product spaces, lattices, rngs, algebraic sets

Algebraic Equations

Given an algebraic structure, we may define *algebraic equations*: we have a collection of symbols called *variables* (three such common symbols in high school are x, y, z) which have restrictions placed on them. A prototypical example of such equations on numbers are *linear equations*: equations of the form $ax + b = 0$ for numbers a, b and variable x.

In abstract algebra, the operations no longer need to be addition and multiplication, and the objects no longer need to be numbers. If you know some calculus, then the following is an example of a more advanced equation[6]:

- *Differential Equations*: These are equations whose solutions are functions. The operations are addition of functions represented by $+$, multiplying by a number, and differentiating (we can treat differentiation as an operation that takes in one function and returns another). An example of a differential equation is:

  $$6f^{''} + 5cf' - f = 0$$

  where f is an at least twice differentiable function

More generally, an algebraic equation p is some composition of operations with elements of a set and variables. Often, we shall want to know if there exists an α (or multiple $\alpha_1, \alpha_2, ..., \alpha_n$) and β such that $p(\alpha) = \beta$ (or $p(\alpha_1, ..., \alpha_n) = \beta$). When looking for solutions, the set of objects of our algebraic structure matter. For example, if we restrict ourselves to working with the positive numbers, the equation $x + 5 = 0$ has no solutions, and if we allow for negative numbers then $x = -5$ is the solution. If we restrict ourselves to working with the whole numbers, then the equation $2x + 3 = 0$ has no

[5]This is similar to the term types from *type theory* which holds a more precise meaning that the current casual use of the word type. The words "family" and "type" where both used here to accommodate type theory and set theory

[6]The set of differential functions along with differentiation forms an algebraic structure known as a *lie algebra*

solutions, and if we allow for rational numbers then $x = -3/2$ is the solution. As we see with the example with differential equations, the objects we are working with need not be numbers. Indeed, as long as we define the operation $+$ correctly in more general contexts, the solution to an equation $ax + b = 0$ need not be a number (it may be a function, a matrix, a network, or a shape)!

Though looking for solutions to equations may seem like a core goal of algebra, it turns out that equations themselves present many fascinating properties that are often much more valuable to study than looking for solutions to equations. In abstract algebra, we consider equations as holding "semantic information", often holding information about a "property" of an algebraic structure. For example, the equation $r^4 = e$ from earlier is an example of an algebraic equation that tells us that rotating 4 times by $\pi/2$ radians gives us back the original square. The reader may start to imagine how this equation characterizes a square more than a pentagon or a triangle, and indeed this notion of an equation "representing" some core information become central to algebra[7]. Often, we shall be able to distinguish two particular instances of an algebraic structure by what equations the elements satisfy. More generally, algebraic equations may be hiding much more information (a famous example being the study of elliptic curves in algebraic number theory), and many mathematicians will invest a great deal of time uncovering these properties.

Goal of Abstract Algebra?

So what does an algebraist aim to achieve? In short:

1. *Solving and understanding equations*: Given an equation or system of equations, an algebraist wants to find solutions to the equation(s) (or determine that no solution exists). An algebraist will look for these solutions by "algebraic" means (for example, calculus or tools from ODE/PDE theory would be considered non-algebraic[8]). When the objects of study are not numbers, finding solutions may get quite complicated. As mentioned above, it shall also be the case that mathematicians look for intrinsic properties of algebraic equations themselves. How to think of this "information" which algebraic equations provide will be elucidated throughout the textbook.

2. *Finding properties of Algebraic Structures*: There is an incredible amount of different families of algebraic structures, each with varying properties. As can be surmised by the table of contents, a lot of energy will be dedicated to understanding the properties of algebraic structures. Some intriguing questions that shall be answered are:

   (a) When can the notion of "factorization" be defined, and when is it unique?

   (b) Given some algebraic structure, are there simple objects that can be used to construct all other objects (for example, there will be algebraic structures which will have the notion of "prime" like prime numbers)

   (c) when are objects "finite", "almost finite", or "infinite"?

   (d) Is the notion of "dimension" well-defined?

   (e) Can we define some reasonable notion of "size"?

[7] To foreshadow, the construction of free-objects shall make this rigorous, and the Nullstellensatz shall be the bridge in thinking of equations as holding semantic information and interpreting them as functions

[8] The boundary of what is an "algebraic" method to solve equations becomes more murky in advanced algebraic topics, for example in the representation of Lie Algebras or Scheme Theory. This shall not be a problem for this textbook in over 99% of cases

(f) Are some objects good at representing a problem in such a way that it elucidates some new property?

these may seem like rather abstract questions, however this is a consequence of studying the abstraction of objects, which often implies that some "core" property of the object reveals itself in the abstraction

3. *Clasifying Algebraic Structures*: In the best case scenario, all possible types of algebraic objects of a given family can be found. For example, it is one of the great accomplishments of the 20th century is finding the building blocks of an algebraic structure known as *finite groups* (requiring over 10000 pages to write down the proof)[9]. More generally, algebraist tend to look for building blocks of algebraic structures and means of creating more complicated algebraic structures given the building blocks.

4. *Developping Algebraic Techniques*: Many algebraic structures allow "shift of paradigms" that reduce a seemingly very difficult problem into a simple solved problem. For example, many shapes (circles, tori, Möbus bands, klein bottles) can be distinguished by associating an algebraic structure to them which then allow mathematicians to find properties of these shapes given properties of the algebraic structures.

The general angle of attack in solving a problem "algebraically" is:

- to find a way of inducing an algebraic structure given the object the mathematician is working on. This is very commonly done in geometry where associating different algebraic structure to a shape give different properties of the shape, depending on the properties of the algebraic structure[10].

- to find some key properties, call them axioms, then study the consequences of these axioms. This shall be done repeatedly in this textbook, and most major algebraic structures have come about through this process[11]

- to break down your problem in pieces that we *know how to put it back together* through the use of several *functions*. The more mathematics that you encounter, the more often you shall see this strategy. For example, in Measure Theory we take advantage of the vector-space structure in the process of measuring diagonal planes in $\mathbb{R}^n$, and in Fourier analysis it is key that solutions to linear PDE's are closed under function-addition to form more complicated solutions. It is important to emphasize that we know how to put it back together using *functions* or else we would have something more akin to *topology*.

Advice for Beginners

For those experiencing abstract mathematics for the first time, perhaps with a background in calculus or some computational linear algebra, the first encounter with this sort of mathematical thinking may seem a bit obtuse. This section shall give some pointers to ease into abstraction.

[9] this shall be explored further in section 3.7

[10] See a book in *algebraic topology* for more details

[11] Groups were an abstraction from bijective function, commutative rings from integers and modular arithmetic's, fields from solving polynomials, Dedekind Domain from rings of integers, and so forth

As a first-time learner, it is tempting to try and memorize the definitions, lemmas, and theorems to be able to apply them on a test, similar to how a successful strategy for at least passing most calculus tests is to memorize the rules of differentiation and integration. This only works in minimal ways for abstract algebra, and mathematics more generally. In order to achieve mastery over mathematics, the reader should really form their own intuitive understanding of the topics, an internal world that naturally holds together all the concepts which the reader should know like the back of their hand. The reader should try to internalize a reason for why the definitions exist, for why the theorems and lemmas need to be proved in the first place, and how one would come up with the questions the theorems and lemmas are asking.

At the very beginning of the readers mathematical journey, this is only partly feasible. In my personal experience, it was very much akin to being a new-born who needs help understanding the language and the basics of how to "walk". However, at your own pace, you shall start "seeing" why a certain definition was chosen, shall start asking questions to yourself that on the next page will be a theorem the author (i.e. me) already answered, and shall start feeling that the concepts you have learned are "basic concepts" that anybody would know. To start picking up these skills, there are certain steps the reader should take:

1. Do as many exercises as the reader needs until they feel they understand the concepts. Some more advanced mathematics courses even go so far as to not present any theorems/lemmas and make the students deduce these results[12].

2. Before looking at the proof of any result, try to see if you can prove it yourself. For some results, this may be very challenging, however in these cases the reader shall find that even taking a moment to think of how to tackle the question creates enormous benefit in understanding the concept, even if the reader is totally off (which I must confess I myself have been multiple times). Doing this shall also give the added benefit of giving the reader more confidence in trying to attempt problems, and being able to prove a result before a proof is presented will certainly make the reader deeply feel like they understand what is going on. It may even be that you come up with a better proof that the one presented in the book! Even if the reader was completely off, having experienced exactly what *not* to think is itself insightful (though admittedly frustrating); it is inevitable that the reader shall not have the right idea for a concept in their mind, and the sooner the reader catches their misconception by being wrong, the better.

3. Try to tie together various concepts you have learned. This increases the number of ways you can recall and use your concepts, and greatly enriches ones appreciation of the topics. This book goes to great lengths to give some intriguing connections, however it is very fruitful to any reader to start seeking out these connections.

Once the readers masters these tools and ways of thinking, by the end of the book the material will truly feel elementary and introductory.

Advice for Proofs

So far, when we tried to prove a proposition, we took the information that we had, and proceeded to find whether they imply the conclusion. However, this is not the only way to prove a proposition is

[12] I have once jokingly seen an exercise in Lang's *Algebra* which was to pick up any textbook in algebra, and prove every result from beginning to end independently

true. There are 3 logically equivalent ways of proving a proposition is correct

Direct proof, contra-positive, contradiction

(In different areas of mathematics, different proof technics will be learnt, however they boil down to doing one of these three approaches)

Preliminary Chapter

If naïve set theory is new to you, this section is going to go through a lot of terminology. The reader may want to skip to certain sections depending on their background:

1. Sets: sets, power-sets, unions and intersections. Logical quantifiers, basic propositions, vacuous truth. Induction. Relations, equivalence Relations, partial and total orders.

2. Functions: injective/surjective/bijective functions and their relation to equivalence relations. Image and restriction, composition of functions, associativity of composition, Cardinality

3. Number theory: GCD and LCM, divisibility, prime factorization, coprime numbers, the euclidean algorithm, Bézout's Identity, Modular Arithmetics, Fermat's Little Theorem

4. Graphs: basic properties

5. Category Theory: intuition, universal property, functors.

You can look over this section if you want a refresher, or skip to section 0.1.4 to start from equivalence relations, in particular if the reader is unaware of the relation between equivalence relations and functions, then this section will be important.

There are also a few starred sections that are not necessary for most of the book, and are added as they show up for a few proofs.

0.1 Sets

As in any field of study, we need to establish a common language for everyone in the field to communicate their ideas. The accepted language of mathematics today was invented by Georg Cantor in the 19th century and is called *set theory*. Before Cantor, there's wasn't one universal set of symbols and concepts mathematicians agreed upon to talk about basic concepts, with different

groups of mathematician inventing their own symbology to talk about their ideas. Nowadays, every field of mathematics grounds their work in set theory [13] – algebra is no exception.

We'll be learning a small part of set theory which concerns itself with using the tools of set theory for talking about concepts. These tools are under the label of *naïve set theory.* We will be using these tools and question little why we are using these particular tools. The reason for the term "naïve" set theory is due to history: there were a couple of important paradoxes which showed some inconsistencies in set theory [14]. From the existence of these paradoxes, a generation of mathematician's tried to free set theory from these problems [15]. Their research became known as set theory, and the previous understanding of set theory came to be known as *naïve set theory.* Naïve set theory wasn't dismissed, instead mathematicians who were interested in making a solid foundation for mathematics tried figuring out exactly how we should be using naïve set theory without contradiction, while the rest of mathematicians used naïve set theory. When we use naïve set theory, we're essentially using the results of [the new] set theory without worrying about the details. It's like taking advantage of a building without knowing the architectural details of why a support was put here, or a door was needed there. We will be taking full advantage of this pre-built building for us. Some footnotes will be added for those who want to peek at the questions that lead mathematicians of the early 20th century to investigate the details of set theory. For those with a deeper interest in the foundations of mathematics, a class on logic is recommended, or you can read one of these books: [16]

As a consequence of using naïve set theory, we'll be defining certain concepts by relying on some [very] basic intuition. For example, we will assume the idea that we can conceptualize a collection of objects (for example, imagine a bunch of trees, or a bunch of numbers), or know the basics of numbers (ex. you know how to think of the symbols $1, 2, 3, ...$ as number, and you can add/subtract/multiply/divide). Definition's after the preliminary chapter will build on the definition's presented here, and so it's not a prerequisite for chapters after the preliminary chapter to rely heavily on intuition.

Basic Definitions: Sets

In mathematics, we work with collection of things, and with "mathematical sentences" about a collection of things, which a sentence that must either be true or false. We will start by learning how to describe the collection of things, and then move on to learning about mathematical sentences. We will make a distinction between a collection of things and the things inside the collection. We call a collection of things a *set*, and the things inside a set are *elements.* Another definition of a set is provided by the founder of set theory Georg Cantor

> *A set is a Many that allows itself to be thought of as a One*

In other words, we can think of any arbitrary collection of objects put together as a new single entity. [17]. The one restriction we give to sets is that all the elements in a set must be different, in other

[13] There is a movement to replace foundation of mathematics with *type theory* – for those interested see ref:HERE

[14] Famously, Russel's paradox showed a logical problem in how we can define sets. We'll actually cover Russel's Paradox in ref:HERE (Perhaps add it to the end of set theory section?) if you're curious to check it out after reading this section

[15] one major result of this research was the axiomatization of set theory (i.e. the development of the fundamental rules) by Zermelo and Frankel, their set of axioms (i.e. fundamental rules) being called ZF in honour of them. There are a couple of axioms that can be added or removed from ZF, but ZF generally forms the core of modern mathematics. If you keep doing mathematics, you will hear statements like "This statement is true in ZF"

[16] here, some good one's in my Book folder

[17] Going deeper than this, asking questions like what are the possible things a set can contain, what are all possible sets, can there be sets that contain itself, all venture into set theory proper

words, we must have each element to be distinct.

When we write a set, we put curly braces ("{" and "}") around a collection of things (or really, symbols which could represent things [18]) which are separated by a comma.

Example 0.1.1: Example of sets

1. To write a set with the numbers $1, 2$ and 3, we write $\{1, 2, 3\}$. The numbers 1, 2, and 3 are the elements of $\{1, 2, 3\}$. Note that $\{1, 1, 2, 3\}$ would not be considered a set since the 1's repeat. The convention when a set has repeated elements is to have them "squished" into one element to create a set with only distinct elements. This would mean that we would consider the set $\{1, 1\}$ to be $\{1\}$ where we squished the two 1's into a single element. This convention also holds if the set has other elements too:

$$\{1, 1, 2, 3\} = \{1, 2, 3\}$$

   If we add an element to a set that's already there, we will also do this "squishing" [a]

2. As mentioned, sets contain can anything. The set

$$\{\text{ice-cream}, \text{cheesecake}, \text{chocolate}\}$$

   is a set that represents 3 desserts. The set represents the weekdays of the week in 5 different languages.

3. The set $\{1\}$ is a set with the element 1. We would not say that 1 and $\{1\}$ are the same thing, since $\{1\}$ is "a set containing the number one as its element" and 1 as "the number one". Using mathematical notation, $1 \neq \{1\}$

4. $\{\}$ is also a set, called the empty set. It pops up often enough that we have a special symbol for this set: $\emptyset$

5. Sets can contain sets, so $\{1, \{☺, \mathfrak{T}\}, \text{bat}\}$ is a set that contains the element 1, the element $\{☺, \mathfrak{T}\}$ (which is also a set), and the element bat.

6. Sets need not have an order, so for example

$$\{\text{ice-cream}, \text{cheesecake}, \text{chocolate}\} = \{\text{cheesecake}, \text{ice-cream}, \text{chocolate}\}$$

   This also means that a set with days of the week (written in any language) is the same if we re-arrange the days:

[a] Formally, (explain what's actually happening here in terms of ZFC?)

Sometimes, we care about the order of elements. If that is the case, we will surround them in parenthesis, "(and)", and call then n-tuples, where n represents the number of elements. For example: $(1, 2, 3)$ is a 3-tuple, and $(1, 2, 3) \neq (3, 2, 1)$ even though $\{1, 2, 3\} = \{3, 2, 1\}$ [19]. 2-tuples are

[18] Some might be dubious on whether the symbols can represent "anything". For example, is it okay to say that x represents a statement that is both true and false? Is that "a thing". This type of question of what is a "well-defined" thing is subject to deeper explorations of set theory

[19] In set theory, every mathematical object is represented as a set. This is no exception in n-tuples. A 2-tuple (a, b) would be $\{a, \{a, b\}\}$. A similar pattern follows for n-tuples

sometimes called a *pair*. If it is clear from context, it is a common convention to drop the number in front of the word tuple, i.e., if the writer consistently uses 2-tuples they might choose to just call them tuples.

Elements

As mentioned earlier, If something is inside a set, we call that an *element* or *object* of the set. We usually denote elements with lower-case letters (ex. a, b)[20]. To write "a in A" in mathematical notation, we use the symbol $\in$. For example:

$$a \in A$$

is read as "a in A" (or "a is in A" to sound more grammatically correct). If we want to say that "a not in A", we add a slash across our symbol (/):

$$a \notin A$$

For example, $1 \in \{1, 2, 3\}$, but $4 \notin \{1, 2, 3\}$. If we have the empty set $\emptyset$ (i.e. $\{\}$), then it is always the case that nothing is inside it, i.e., $a \notin \emptyset$ for any a. Sets can also contain other sets, so $\{1\} \in \{\{1\}, 2, \{\text{cheesecake}, 6\}\}$. Note that $1 \notin \{\{1\}, 2, \{\text{cheesecake}, 6\}\}$ since the set $\{\{1\}, 2, \{\text{cheesecake}, 6\}\}$ doesn't contain "the element 1" but the "the set that contains the element 1".

1: Concept Check

Is

$$\{1, 2, 3\} \in \{0, \{\{1, 2, 3\}\}, 4\}$$

Equality

The concept of equality was already used earlier based on our intuition. We will expand a bit further on how to think of equality in set theory, and we will distinguish two ways of looking at equality:

$$A = \{1, 2, 3\}$$
$$A := \{1, 2, 3\}$$

Starting with the first line, the symbol $=$ means *logically equal*. Being "logically equal" (which we will always abbreviate to saying equal[21]) means the two symbols represent the same thing. For example, the number one-half and two-quarters are equal:

$$\frac{1}{2} = \frac{2}{4}$$

[20] Here's another "beyond the scope of naive set theory" question: What are the possible elements that may exist? Question of this varieties lead to logicians like Butrard Russel to come up with *Russel's Paradox*. This paradox will actually be covered in the book near the end (see ref:HERE). It's written in such a way that you can skip to take a look at it if you're curious

[21] This is another big set-theoretic question: what does it really mean for something to be equal? Bertrand Russel and Alfred Whitehead in their book *Principia Mathematica* famously took 379 pages before they defined equality. In modern set theory, two sets are equal if they contain the same elements

even though the symbol $\frac{1}{2}$ and $\frac{2}{4}$ are different symbols. We would read $\frac{1}{2} = \frac{2}{4}$ as "one-half is equal to two-quarters". If we want to say that the two are not equal (i.e. the symbols represent different things [22]), we do the same convention as with $\in$ and put a slash across the symbol: $\frac{1}{2} \neq \frac{1}{3}$.

Example 0.1.2: Examples Of Equality

1. if we want to say two sets have the same elements, that they are each representing the same collection of objects, we write $A = B$
2. $\{1, 2, 3\} \neq \{2, 3, 4\}$. Even though both sets have the elements 2 and 3, they each contain a number the other doesn't have, and so they are not equal
3. Similarly, $\{1, 2\} \neq \{1, 2, 3\}$ since $3 \notin \{1, 2\}$
4. As seen in a elementary class in mathematics, $\frac{1}{2} = \frac{2}{4}$, but $\frac{1}{2} \neq \frac{1}{4}$ [a]

[a] Notice that in all the previous examples, we were comparing sets, while in this example we're comparing numbers. In proper set theory, every object is actually a set. Thus, even comparing $\frac{1}{2}$ and $\frac{2}{4}$ will be a comparison with sets. We will actually construct $\mathbb{Q}$ and show how it's elements are sets in section 9.5

The symbol := means the same thing as =, that is, := is also logically equal. The difference between := and = is that := is used to emphasize that the symbol on the left of := will now be used to represent the thing on the right of :=. This is most often used in definitions to make sure the reader understands that symbol will consistently be used to represent the same thing many times. Because of this, the symbol := is read as "define" – the statement $A := \{1, 2, 3\}$ can be read that "Define A to be the set $\{1, 2, 3\}$".

Set-Builder Notation

Many times, we want to have a set of elements that satisfy only certain properties. For example, from the set of all deserts, I only want the ones that contain ice-cream, or from the set of all numbers I want only the prime numbers, or from the set of all months, I want the months that end in "ber". The way we'll capture this idea with sets is by using *set builder-notation* to build a new set given an old one. If S is a set and P is some sentence describing a property (ex. being an even number, or being a prime number, or representing "this desert has ice-cream"), then we'd write

$$A = \{s \in S \mid P\}$$
$$\text{or}$$
$$A = \{s \in S \ : \ P\}$$

and is read "s in S such that [s satisfies the property P]". In the next section, we shall be more precise about what type of sentence can be used:

[22] This is another question beyond naïve set theory: Since all we're using are symbols that represent objects, where do the objects come from? One answer to this is some universal set with every possible thing as an element of it, but thinking about what exactly does it mean for a set to contain everything leads to interesting nuances (one of the problems of a set that contains everything is once again Russel's Paradox)

Example 0.1.3: Set Builder Notation

1. If

$$M = \{\text{Janurary, February, March, April, May, June,} \\ \text{July, August, September, October, November, December}\}$$

and $D = \{m \in M \mid \text{m ends in "ber"}\}$, then

$$D = \{\text{September, October, November, December}\}$$

2. If $W = \{\text{hello, (words in other languages)}\}$, and $V = \{w \in W \mid w \text{ is in Vietnamese}\}$, then

$$V = \{\text{here}\}$$

3. if $\mathbb{N} = \{0, 1, 2, 3, \cdots\}$ and $P = \{n \in \mathbb{N} \mid n \text{ is a prime number}\}$ then

$$P = \{2, 3, 5, 7, 11, \cdots\}$$

Note that $P' = \{n \in \mathbb{N} \mid \text{the only numbers that divide } n \text{ are 1 and } n\}$ is the exact same set (i.e. $P = P'$). Different properties can produce the same sets.

4. if $\mathbb{N} = \{0, 1, 2, 3, \cdots\}$ represents the set of all non-negative numbers with no decimal places, the set

$$A = \{x \in \mathbb{N} \mid x = 2y,\ y \in \mathbb{N}\}$$

would represent all even numbers [a], and can be read "x in the natural numbers such that x can be twice any natural number"

[a] This notation has it's limitations if you dig deeper in it. You can create some very interesting paradoxes (Russel's paradox being one of them). We will not construct such contradictory cases in the book

0.1.1 Propositions

We will now explore how to formulate questions and mathematical statements using mathematical language.

Intuitively, there is some understanding of the meaning of a sentence[23] . If I say "I think blue is pretty" then in your mind some information got passed on. If we make a sentence that has a truth value associated to it (the sentence can either be true or false) [24] , like "The number 2 is prime", then we call such a sentence a *proposition*. Propositions are at the heart of mathematical communication. We have already encountered propositions before when we used set-builder notations: The restriction *has to be a proposition*. For convenience, we assign a symbol to a proposition. For example, we can have $P =$ "2 is a prime number". Then P would be true. However, $Q =$ "2 is not a prime number" is a false. These two examples of propositions are examples of tautologies: propositions whose truth

[23] The nature of how we derive meaning from sentences is an interesting logical and philosophical debate that goes beyond the scope of this book. See Russel, Wittgenstein, _____, for more details

[24] Development in a field called *topoi theory* are trying to generalize our notion of proposition beyond true and false, an interesting developing field to look into

value is dependent of any variables. If instead we take:

$$R(x) = \text{“}x \text{ is a prime number”}'$$

where x may be any natural number, then this proposition depends on the value the variables x will take.

Example 0.1.4: Propositions

1. let $P(x) =$“x is a month of the year”. then

$$P(\text{September}) = \text{true} \qquad P(\text{Wednesday}) = \text{false} \qquad P(\text{Laughing}) = \text{false}$$

2. The variable x can also be other propositions. For example we can have $P(x) =$“x is a dog” and $Q(x) =$“x is an animal”.
3. Propositions can be composed, for example consider $R(x) =$“$P(x)$, then $Q(x)$” where $P(x), Q(x)$ are the propositions from the above. When is this proposition true?
4. Proposition can also take in multiple variables, for example

$$P(x, y) = (x < y)$$

Propositions of the form

$$\text{if } P(x) \text{ then } Q(x)$$

are some of the most common, allowing for relations being made between different concepts. The symbol used to symbolize an “if, then” relation is $\Rightarrow$, so

$$\text{If } P(x) \text{ then } (Q) \qquad \text{becomes} \qquad P(x) \Rightarrow Q(x)$$

Which is usually read “$P(x)$ implies $Q(x)$”. A common mistake is to assume that if $P(x) \Rightarrow Q(x)$, then $Q(x) \Rightarrow P(x)$. This need not be the case. For example, define the following two propositions:

$$A(x) = (x|2) \Rightarrow (x|4) \qquad A'(x) = (x|4) \Rightarrow (x|2)$$

When is proposition $A(x)$ true? When is $A'(x)$ true? If it is true that $P(x) \Rightarrow Q(x)$ always implies $Q(x) \Rightarrow P(x)$ we use the symbol $P(x) \Leftrightarrow Q(x)$ and read $\Leftrightarrow$ as “if and only if” (sometimes abbreviated to iff).

Vacuous Truth

Let's say we had two proposition $P(x, y) = x \Rightarrow y$, and let's say that x was false, while y is true. What is the truth value of $P(x, y)$? For example, let's say that

$$x = \text{I'm walking in the rain}$$
$$y = \text{I am wet}$$

Then if you are *not* walking in the rain, but *are* wet, then what should be the value of $P(x, y)$? Perhaps more interesting is the following example:

1. Let's say there are no cell-phones in a room. Then is the statement "all phone are turned off" true?
2. A famous such example is the following, is the statement "The king of France is not wearing pants" true or false if France does not have a king?

There are many solutions to this problem. The mathematically accepted solution to this problem is something called *vacuous truth*, which gives the following resolution:

$$(\text{False} \Rightarrow \text{True}) \text{ is True, and } (\text{False} \Rightarrow \text{False}) \text{ is } \textit{also} \text{ True}$$

The idea being that any "nonsense" statement is always "true" in the most trivial way. This interpretation of truth is not a unique one, but it is one that is very convenient as shall be seen by the reader.

Logical Quantifiers

We will now explore how to create more complicated properties by adding to our tool belt 6 "words" from which we can construct mathematical sentences:

Definition 0.1.1: Quantifiers And Connectives

We define the following 3 quantifiers to be used in propositions:

1. $\forall$: the for-all quantifier.
2. $\exists$: the *existence* quantifier
3. $\exists!$ the *unique existence quantifier*

and the following 4 connectives to be used for propositions

1. $\vee$: disjunction (or) connective. If P, Q are propositions, then $P \vee Q$ mean that if one of or both of P or Q is true, then the new poposition $P \vee Q$ is true.
2. $\wedge$: conjunction (and) connective. If P, Q are propositions, then $P \wedge Q$ mean that if both P and Q are true, then the new proposition $P \wedge Q$ is true. If one of them are false, then $P \wedge Q$ is false.
3. $\Rightarrow$: implication connective: $P \Rightarrow Q$ is false if P is true and Q is false. In other words, if P is true, then P implies Q means Q is true.
4. $\neg$ negation (not) connective. If P is a proposition, then if P is true then $\neg P$ is false, and similarly if P is false then $\neg P$ is true.

The keen reader may ask if there are any other quantifiers or connectives, and indeed there can be many more. However, it is in fact sufficient to define these connectives/quantifiers to completely describe any [first-order] mathematical sentence[25]. Using these symbols, can the reader translate some of the propositions earlier in the book using these symbols?

[25] This would be explored in a course on logic

Recall that we have defined the symbol $\in$ earlier to specify an element belongs to a set. This is a special symbol which is neither a quantifier nor a connective. It is considered a special symbol that is part of the language of set theory.

The order of quantifiers is important. For example:

$$\forall x,\ \exists y,\ x < y$$

means that for every x, there exists a y that is greater than it, which is certainly true when working with numbers. However:

$$\exists x,\ \forall y,\ x < y$$

means that there exists an x that is smaller than all y, which is certainly not possible since we can always find a smaller number.

With these, we can define certain operations we can perform on sets (don't forget to add visuals):

Definition 0.1.2: Set Operations

1. Define a subset to be:
$$A \subseteq B \Leftrightarrow (x \in A \Rightarrow x \in B)$$
Sometimes the symbol $A \subset B$ is also used. Since it is possible that $A \subset B$ $A = B$ (you can check this by the definition), the symbol $\subseteq$ will predominantly be used to emphasize this relation. If we want to make it clear that $A \neq B$, we write $A \subsetneq B$, we will write $A \subsetneq B$.

   For any set S, we always have $\emptyset \subseteq S$ and $S \subseteq S$. With this notation, we also have that $A \subseteq B$ and $B \subseteq A \Rightarrow A = B$

2. Define the intersection of two sets to be:
$$A \cap B := \{x \mid x \in A \text{ and } x \in B\}$$

3. Define the union of two sets to be:
$$A \cup B := \{x \mid x \in A \text{ or } x \in B\}$$
if $A \cap B = \emptyset$, we say that A and B are *disjoint*

4. Define difference between sets to be:
$$A \setminus B := \{x \mid x \in A \text{ and } x \notin B\}$$
sometimes also written as $A - B$. If $T \subseteq S$, then we say that $S \setminus T$ is the *compliment* of T, usually written T^c.

Another very important definition

Definition 0.1.3: Product Of Sets

Let A and B be sets. Then

$$A \times B = \{(a, b) \mid a \in A,\ b \in B\}$$

is the set of all 2-tuples (called just tuples for short). This construction can be extended indefinitely:

$$X_1 \times X_2 \times \cdots = \{(x_1, x_2, ...) \mid x_1 \in X_1,\ x_2 \in X_2,\ ...\}$$

This notation is also abbreviated as:

$$\prod_{i=1}^{n} X_i$$

For example:

Example 0.1.5: Product Of Sets

$$\{1, 2\} \times \{a, b\} = \{(1, a), (2, a), (1, b), (2, b)\}$$

and

$$\{a, b\} \times \{1, 2\} = \{(a, 1), (b, 1), (a, 2), (b, 2)\}$$

(A word on arbitrary products? i.e. $\prod_{i \in I} X_i$)

Exercise 0.1.1

1. Let $\{E_n\}_{n=1}^{\infty}$ be a collection of sets indexed by the natural numbers. Define

$$\limsup\{E_n\}_{n=1}^{\infty} = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} E_n \qquad \liminf\{E_n\}_{n=1}^{\infty} = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} E_n$$

   show that

$$\limsup\{E_n\}_{n=1}^{\infty} = \{x \mid x \in E_N \text{ for infinitely many } n\}$$
$$\liminf\{E_n\}_{n=1}^{\infty} = \{x \mid x \in E_N \text{ for all but finitely many } n\}$$

2. (Demorgans La)

Naming the Numbers and Common Sets

With the basics of set down, we can introduce some of the most common sets and their notation

Definition 0.1.4: Common Sets

1. As we've already seen, if a set has no elements, the symbol $\emptyset$ represents the empty-set
2. If a set has a single element, we call that set a *singleton*. For example, $\{1\}$ and $\{\pi\}$ are singletons.
3. If X is a set, $P(X)$ is the set of all subsets of X[a]. In set builder notation, it can be denoted $P(X) = \{S \mid S \subseteq X\}$. For example, if $X = \{a, b, c\}$, then

$$P(X) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$$

[a] The existence of this set is an axiom in ZFC

Definition 0.1.5: Sets Representing Numbers

1. $\mathbb{N} := \{0, 1, 2, 3, \cdots\}$ is the set of all *natural numbers* (some call it the *counting numbers* or *whole numbers*). Natural numbers are all positive numbers which do not have any decimal places. Sometimes, people define $\mathbb{N}$ to not have 0. We will always have $\mathbb{N}$ contain the number 0 (for reasons which will be clear much later in the book [a]). If we want only the [strictly] positive number [b], we will write $\mathbb{N}_{>0}$.
2. $\mathbb{Z} := \{\cdots - 3, -2, -1, 0, 1, 2, 3, \cdots\}$ is the set of all *integers*. Integers are the natural numbers along with their negative counter-part. Sometimes, the natural numbers are called the non-negative integers.
3. $\mathbb{Q} := \left\{ \frac{p}{q} \;\middle|\; p, q \in \mathbb{Z},\ q \neq 0 \right\}$ is the set of all *rational numbers*. For example, $\frac{1}{2} \in \mathbb{Q}$. The symbol $\mathbb{Q}$ is used since numbers of the form $\frac{p}{q}$ are called ***q**uotients*. Notice also that this is one of the instances where we use the convention that if there is a repeated element, we "squish" to a single element. If we did not take this convention, this set wouldn't be defined, since there would be many repeated numbers (for example $\frac{1}{2} = \frac{2}{4} = \frac{4}{8}$ and so forth)
4. $\mathbb{R}$ represents any number with any decimal expansion (ex. $\pi, e, 1, \frac{1}{3}$), i.e., all the numbers on the number line. If you've learned about Complex numbers or Quaternions in high school, these numbers are not real numbers, and will be defined and studied later in this book. The set $\mathbb{R}$ is called the *real numbers*. We can abuse notation (meaning use the notation in a way bends the rules a little) and write $\mathbb{R}$ as

$$\mathbb{R} := \{\cdots D_3 D_2 D_1 . d_1 d_2 d_3 \cdots \mid D_i, d_i \in \mathbb{Z}\}$$

where $\cdots D_3 D_2 D_1 . d_1 d_2 d_3 \cdots$ represents some arbitrary string of numbers put together where there can only be finitely many D_i's, but there can be infinitely many d_i's. In other word, the number

$$0.123123123123 \cdots$$

where 123 continues on forever on the right is in the set $\mathbb{R}$, but

$$\cdots 123123123$$

where 123 continues on forever on the left is not in the set $\mathbb{R}$.

[a]See section 7.1

[b]There is no complete agreement on whether 0 is positive or negative: 0 is either both positive and negative or neither positive nor negative. It usually comes down to the purposes of the context. If we allow 0 to be a positive number, we sometimes say that all numbers larger than 0 are *strictly positive*

Some find it intuitive to think about $\mathbb{R}$ has having "no gaps", which means we cannot perfectly split $\mathbb{R}$ into two using the less than and greater than operator ($<$ and $>$) without missing a number. For an example of a split, we can split $\mathbb{N}$ into $\mathbb{N}_{\frac{5}{2}^+} = \left\{x \in \mathbb{N} \mid x > \frac{5}{2}\right\}$ and $\mathbb{N}_{\frac{5}{2}^-} = \left\{x \in \mathbb{N} \mid x < \frac{5}{2}\right\}$, and we have successfully split $\mathbb{N}$ into two without missing a number (i.e. every natural number is either in $\mathbb{N}_{\frac{5}{2}^+}$ or $\mathbb{N}_{\frac{5}{2}^-}$). If we chose to split around 5, then neither the set $\mathbb{N}_{5^+}$ nor the set $\mathbb{N}_{5^-}$ would contain 5, meaning it does not split $\mathbb{N}$. With the Real numbers, there does not exist a perfect split. You might ask if $\mathbb{Q}$ has this same property, but we will prove that we can in fact always split it: We will show $\mathbb{Q}_{\sqrt{2}^+} = \left\{x \in \mathbb{Q} \mid x > \sqrt{2}\right\}$ and $\mathbb{Q}_{\sqrt{2}^-} = \left\{x \in \mathbb{Q} \mid x < \sqrt{2}\right\}$ splits $\mathbb{Q}$ (see ref:HERE).

The fact that $\mathbb{R}$ cannot be split into two using $>$ and $<$ is actually at the heart of many formal definitions of $\mathbb{R}$ which is covered in an Analysis I class. In this book, we will not need to think of $\mathbb{R}$ in this fashion, and so we'll stick with just using intuition for now. [26]

`https://math.stackexchange.com/questions/2834876/convergence-of-geometric-series-with-r1`

0.1.2 Induction

Here. The whole $(p(x) \Rightarrow p(x+1)) \Rightarrow$ here this as the formal definition? + some intuition?

0.1.3 *Pigeon-hole Principle

(Comes up in algeraic number theory, so should add it here)

0.1.4 Relations and Equivalence Relations

What does it mean two things to be equal? There are actually many ways, depending on what you want to be "equal". For example,

1. we may say that two people have the same birthday, even if they are not born in the same year.
2. more examples her

In this section, we shall explore how to capture this more general notion of equality.

Relations

We start of with the more general notion of *relation*. Many times when working with sets, some subsets can be thought of as having certain properties for example, we can split the set $\mathbb{Z}$ into even

[26]Google *Dedikend Cut construction* or *Completion of Rational Numbers* if you're curious to see a formal definition of $\mathbb{R}$

and odd numbers, or we have some numbers being bigger than others, or that some numbers are multiples of another. The following is this idea is formalized:

Definition 0.1.6: [Binary] Relation

Let X be a set. Then any collection of pairs (x, y), $x, y \in X$ is called a [binary] *relation*. If $(a, b) \in R$, we will say that x *is related to* y. If $(a, b) \in R$, we will write aRb.

More generally, we can increase the number of times we take the product (i.e. $n > 2$) (ex. $X \times X \times \cdots X$). This would be called an *n-relation*.

We shall almost always think of relations

Example 0.1.6: Relations

1. Take $\mathbb{Z}$ and define the set $<$ to be

$$(x, y) \in < \text{ if and only if } x \text{ is strictly less than } y$$

Then $(3, 4) \in <$, $(3, 3) \notin <$, and $(4, 3) \notin <$. If $(a, b) \in <$, we would usually write:

$$a < b$$

and if $(a, b) \notin <$, we would write

$$a \not< b$$

A similar relation is the $\leq$ relation, where we also allow for two numbers to be equal

2. Take $\mathbb{Z}$ and define the set R_{div} to be:

$$(x, y) \in R_{\text{div}} \text{ if and only if } x \text{ divides } y$$

Then $(2, 4) \in R_{\text{div}}$. What other elements are in this set? This set is usually denoted $|$ so that if a divides b then:

$$a|b$$

3. for a non-mathematical example, if we let X be the set of all humans, then we may define a relation "is the ancestor of". Another similar but different relation is "is the parent of" where two humans are related is one is the parent of another.

A common relation to put on objects give a sense ordering:

Definition 0.1.7: Order

Let R be a relation on X. Then R is a *partial order* if it is:

1. **reflective**: For all $x \in X$, xRx (x always relates to itself)
2. **antisymmetric**: if xRy and yRx, then $x = y$[a]
3. **Transitive**: if xRy and yRz, then xRz.

Furthermore, a partial order is a *total order* (or simply an *order*) if it is furthermore:

4. **Connected**: For each $x, y \in X$, either xRy or yRx

[a] a relation is symmetric if xRy then yRx. Hence, antisymmetry means no elements can be symmetric with another elements besides itself

The canonical total order is $\leq$ on numbers, and the canonical partial order is $\subseteq$ on sets.

Example 0.1.7: Orders

1. Let $\mathbb{N}^2 = \mathbb{N} \times \mathbb{N}$. Give some examples on $\mathbb{N}^2$ of different ordering (Ex. dictionary, and others)
2. Take any dictionary. Then they are lexicographically ordered (more words here)
3. All numbers we introduced so far come have an order defined: $(\mathbb{N}, \leq), (\mathbb{Z}, \leq), (\mathbb{Q}, \leq), (\mathbb{R}, \leq)$. This is the order define in high school (so $3 \leq 4$ or $3 \leq \pi$)
4. other cool example

Equivalence Relations

The most common type of relation that we shall see is the equivalence relation which generalizes the notion of equality:

Definition 0.1.8: Equivalence Relation

Let S be a set and $x, y, z \in S$ be elements of S. An *equivalence relation* on a set S is relation $\sim$ satisfying

1. **(reflective)** $x \sim x$ (i.e. every element must be related to itself)
2. **(symmetric)** if $x \sim y$ then $y \sim x$
3. **(Transitive)** if $x \sim y$ and $y \sim z$ then $x \sim z$

Example 0.1.8: Equivalence Relations

1. Equality ($=$) is the prototypical example of an equivalence relation. For example, If we have $a, b, c \in S$ with $aRb \Leftrightarrow a = b$, then

(a) $a = a$
(b) $a = b \Rightarrow b = a$
(c) If $a = b$ and $b = c$, then $a = c$.

In a ways, an equivalence relation is a generalization of equality.

2. Let X be any non-empty set. Then we may make all elements equivalent to each other. Similarly, we may make all elements only equivalent to themselves.
3. Let P be the set of all people. Check that "has the same birthday" is an equivalence relation
4. Two equivalence relations from high school that the reader may have encountered is with triangles. Check that being *similarto* and being *congruent to* are equivalence relations.
5. Define $\sim$ on $\mathbb{Z}$ to be $a \sim b$ if and only if $|a| = |b|$. Then $\sim$ is an equivalence relation.
6. Is $\leq$ an equivalence relation?
7. Let R be the empty relationship on any non-empty set. Is R an equivalence relation?

Given an equivalence relation $\sim$ on S, we may split S into subsets defined as:

$$(S/\sim) = \{T \subseteq S \mid x, y \in T \iff x \sim y\}$$

For example, if $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, and:

$$1 \sim 2 \qquad 3 \sim 4 \sim 5 \sim 6 \qquad 7 \sim 8 \sim 9$$

Then:

$$S/\sim = \{\{1, 2\}, \{3, 4, 5, 6\}, \{7, 8, 9\}\}$$

In fact, we may define an equivalence relation on a set S by breaking it up into disjoint sets who union of elements would be S. This is called a partition:

Definition 0.1.9: Partition

Let S be a set. Then a *partition* of S is a collection of subsets of S such that

1. for each subset S_i, S_j, $S_i \cap S_J = \emptyset$
2. $\cup_i S_i = S$

each partition is called an *equivalence class*.

0.1.5 Functions

In mathematics, it is very desirable to relate one set to another set, or said differently how two sets may "interact". For example, it may be desirable that to each positive natural number $\mathbb{N}_{>0}$, we

associate to it the number of prime factors it has, for example:

$$
\begin{aligned}
1 &\mapsto 1 \\
2 &\mapsto 1 \\
3 &\mapsto 1 \\
4 &\mapsto 2 \\
5 &\mapsto 1 \\
6 &\mapsto 2 \\
&\vdots \\
16 &\mapsto 4 \\
&\vdots
\end{aligned}
$$

Another more abstract relation we may ask for is given a set of names, say

$$\{\text{Alfred}, \text{Socrates}, \text{Buddha}, \text{Oscar Petesron}\}$$

the function returns a set containing some information about that person:

$$
\begin{aligned}
\text{Alfred} &\mapsto \{\text{Batman's Butler}, \text{Friend of Bruce}\} \\
\text{Socrates} &\mapsto \{\text{Philosopher}, \text{lived in Athens}, \text{Never wrote his work down}\} \\
\text{Buddha} &\mapsto \{\text{Real name Siddhartha Gautama}, \text{Starter of a Religion}, \text{Born in Lumbini}\} \\
\text{Oscar Peterson} &\mapsto \{\text{Greatest Jazz Pianist}, \text{Canadian}, \text{“Maharaja of the keyboard”}, \text{Lived to 82}\}
\end{aligned}
$$

As a final example, we may imagine a every location of a map of the world being represented as an element in a set. Then we may relate a point on the map to the temperature at that location. Generalizing this pattern, given some set X, we may want to assign it an element from another set Y.

Definition 0.1.10: [set] Functions

Let X and Y be sets. Then a *function* is a subset $\Gamma(f) \subseteq \{(x, y) \mid x \in X,\ y \in Y\}$ such that

1. **Total** or **Serial** (left-total): For every $x \in X$, there exists an element $(x, y) \in \Gamma(f)$
2. **Functional** (right-unique): Every $x \in X$ is associated with one and only one $y \in Y$, that is, it cannot be that $(x, y), (x, y') \in \Gamma$, for $y \neq y'$ where $y, y' \in Y$

This is usually written as:

$$f : X \to Y \qquad f(x) = y \qquad \text{or} \qquad X \xrightarrow{f} Y$$

where f is called the *function* and $\Gamma(f)$ is called the graph of the function and is defined to be:

$$\Gamma(f) = \{(x, y) \mid x \in X,\ y \in Y,\ f(x) = y\}$$

If $f : X \to Y$, X is called the *domain* and Y is called the *codomain*. The collection of all functions from two sets X, Y is denoted:

$$\mathrm{Hom}_{\mathbf{Set}}(X, Y) \qquad \text{or} \qquad Y^X \ {}^{(a)}$$

[a] Notice this is a set operation. This particular notation shall be elaborated on with more category theory in section ref:HERE

Function are also sometime called maps or mappings. Sometimes, we want to define a function, but don't need to reference to it again (in computer science, this is called an *anonymous function*). In this case, we may use the following notation:

$$x \mapsto y$$

instead of $f(x) = y$. If $X = Y$ so that $f : X \to X$, then we may think of f as a relation which has the property of being functional and total. If more than one function with the same domain and codomain is defined (say f, g with domain X and codomain Y), then the declaration is often shorthanded to $f, g : X \to Y$.

Example 0.1.9: Functions

1. The first three examples given are examples of functions, since every element in the domain get's mapped to a unique element in the codomain.
2. Some other perhaps more familiar functions to the reader are
   (a) $x \mapsto e^x$, i.e. $f : \mathbb{R} \to \mathbb{R},\ f(x) = e^x$
   (b) $\log : \mathbb{R}_{>0} \to \mathbb{R}$, $x \mapsto \log(x)$
   (c) $p(x) : \mathbb{R} \to \mathbb{R}$, $p(x) = ax^2 + bx + c$, for $a, b, c \in \mathbb{R}$
3. Let X be an arbitrary non-empty set. Then there always exists at least one function from X to X, namely:
   $$\mathrm{id}_X : X \to X \qquad \mathrm{id}_X(x) = x$$
   that is, id_X maps every element to itself.

Definition 0.1.11: Image and Restriction of Function

1. Let $S \subseteq X$ and $f : X \to Y$. Then:

$$f(S) := \{f(x) \mid x \in S\}$$

this is called the *image* of S. If $S = X$, then $f(X)$ is called the *image* or *range* of the function f, and is commonly denoted $\mathrm{im}(f)$.

2. Let $S \subseteq X$. Then $f|_S$ is a function $f|_S : S \to Y$ such that

$$f|_S(x) = f(x)$$

that is, $f|_S$ is the *restriciton* of f by S.

As a function is just some "syntactic sugar" to more easily represent the set Γ_f, two functions f, g are equal if $\Gamma_f = \Gamma_g$. In terms of function notation this means that f, g are equal if for all elements x in the domain:

$$f(x) = g(x)$$

An advantage of the left-total and right-unique property is that it is easy to compose functions together, that is it is easy to chain together the output of one function to be the input of the next:

Proposition 0.1.12: Composition of Functions

Let $f : X \to Y$, $g : Y \to Z$, $h : Z \to W$ be function. Define $(g \circ f)$ to be $(g \circ f)(x) = g(f(x))$. Then $(g \circ f)$ is a function and is called *g compose f*. The process is called *function composition*.

***Proof* :**
Make sure that $g \circ f$ satisfies the properties of functions

Example 0.1.10: Function Composition

1. Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ and $g : \mathbb{R} \to \mathbb{R}^2$ be $g(x) = (x, 0)$. Then:

$$(g \circ f)(x) = (x^2, 0)$$

A function may also be composed with itself, given the domain and codomain match:

$$(f \circ f)(x) = x^4$$

2. Let $f : X \to Y$ be any function, $\mathrm{id}_X, \mathrm{id}_Y$ the identity functions defined in example 0.1.9. Then:

$$f \circ \mathrm{id}_X = f \qquad \mathrm{id}_Y \circ f = f$$

This chaining together is in a way "state-independent": given three functions f, g, h, the function $(g \circ f)$ following by composing with h is equal to if we first just do f, then follow by taking the

combined state $(h \circ g)$[27]:

Proposition 0.1.13: Associativity of Functions

Let $f : X \to Y$, $g : Y \to Z$, $h : Z \to W$ be function. Then:

$$h \circ (g \circ f) = (h \circ g) \circ f$$

Pictorially, function composition may be represented by the following diagrams:

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[rr, bend left, "g\circ f"] & B \arrow[r, "g"] & C \end{tikzcd} \quad \text{or} \quad \begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "g\circ f"'] & B \arrow[d, "g"] \\ & C \end{tikzcd} \quad \text{or} \quad \begin{tikzcd} & C \\ A \arrow[ur, "g\circ f"] \arrow[r, "f"] & B \arrow[u, "g"'] \end{tikzcd}$$

and associativity may be drawn as:

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[rr, bend right, "g\circ f"'] & B \arrow[r, "g"] \arrow[rr, bend left, "h\circ g"] & C \arrow[r, "h"] & D \end{tikzcd}$$

***Proof* :**
This is a long and tedious proof that the reader should attempt at least once in their life.

These diagrams have the following special property: start on one of sets in the diagram (say A) and travel down the arrows to reach another set (say to D in the last diagram). Now start again at A and travel down to D via a different path in the diagram. Then notice that the compositions of functions agree with each other in the two paths. For example, take the following path:

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[rr, bend right, "g\circ f"'] & B \arrow[r, "g"] \arrow[rr, bend left, "h\circ g"] & C \arrow[r, "h"] & D \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} A \arrow[r, "f"] \arrow[rr, bend right, "g\circ f"'] & B \arrow[r, "g"] \arrow[rr, bend left, "h\circ g"] & C \arrow[r, "h"] & D \end{tikzcd}$$

Then the 1st red line is the composition $(h \circ g) \circ f$ and the 2nd red line is the composition $h \circ (g \circ f)$. Then by associativity, these two functions are equal

$$(h \circ g) \circ f = h \circ (g \circ f)$$

a diagram which satisfies this property for all paths is called a *commutative diagram*. These shall show up everywhere in this book, and mathematics more generally. They shall be further explored in section 0.4.

[27] can you think of an example where this is not the case in your everyday life? This would greatly help in understanding associativity

Injective, Surjective, and Bijective

There are two special types of functions that deserve to be pointed out: *injective* functions and *surjeciton* functions. These shall each represent some partial "reverse" of a function. Starting off with injectivity:

Definition 0.1.14: Injective

Let $f : X \to Y$ be a function. Then f is called *injective*, an *injection*, or *one-to-one* if the following condition holds: for all $a, b \in X$

$$f(a) = f(b) \;\Rightarrow\; a = b$$

To emphasize a function is injective, a hooked right arrow ($\hookrightarrow$) may be used like so:

$$f : X \hookrightarrow Y \qquad X \overset{f}{\hookrightarrow} Y \qquad X \hookrightarrow Y$$

Intuitively, being injective means we can embed the domain X directly into Y. As a consequence of this, we can always "recover" X in the following way

Proposition 0.1.15: Injective And Left Inverse

let f be an injective function. Then there exists a function $\ell f(X) \to X$ such that

$$\ell \circ f = \mathrm{id}$$

that is $(\ell \circ f)(x) = \mathrm{id}(x) = x$ is the identity function.

Proof **:**
Define $\ell : Y \to X$, $\ell(y) = \ell(f(x)) = x$. Since f is injective, this is a well-defined function (if it were not injective, then $\ell(f(x))$ would be ambiguous). Then by definition $(\ell \circ f)(x) = \ell(f(x)) = x$, as we sought to show.

Notice that the inverse function ℓ inverted the function *after* applying f. What if there is some function $r : Y \to X$ such that f reverses it? The following shows when this is the case:

Definition 0.1.16: Surjective

Let $f : X \to Y$ be a function. Then f is *surjective*, a *surjection*, or *onto* if for all $y \in Y$, there exists an x such that $f(x) = y$, i.e. $f(X) = Y$. To emphasize a function is surjective, a double-headed right arrow ($\twoheadrightarrow$) may be used like so:

$$f : X \twoheadrightarrow Y \qquad X \overset{f}{\twoheadrightarrow} Y \qquad X \twoheadrightarrow Y$$

Intuitively, being surjective means that Y is "represented" by the elements of X, i.e., every element y has one or many elements of x associated to it, which is why another word for a function to be surjective is that it is "onto". As a consequence of this, we can pullback any element $y \in Y$ to some representative $x \in X$ such that $f(x) = y$. We can capture this intuition in the following way:

Proposition 0.1.17: Surjective And Right Inverse

Let $f : X \to Y$ be a surjective function. Then there exists a (and in fact more generally many) $r : Y \to X$ such that

$$r \circ f = \mathrm{id}$$

that is $(f \circ r)(x) = \mathrm{id}(x) = x$ is the identity function.

Proof **:**
Define $r : Y \to X$ by mapping each $y \in Y$ to some element $x \in X$ such that $f(x) = y$ (such an element exists since f is surjective). Then by definition for any $y \in Y$, $(f \circ r)(y) = f(r(y)) = f(x) = y$ as we sought to show.

Another perhaps more intuitive way of seeing how surjections make every y "represented" by X is by seeing that every surjection defines an equivalence class:

Proposition 0.1.18: Surjective And Equivalence Class

Let $f : X \to Y$ be a surjective function. Then we can define an equivalence class $\sim_f$ on X by defining:

$$a \sim_f b \qquad \Leftrightarrow \qquad f(a) = f(b)$$

If $[x] \in X/\sim_f$ is an equivalence class, such that $f(x) = y$, then $[x]$ is called the *fiber* over y.

Proof **:**

1. For each $a \in X$, $f(a) = f(a)$ by right-uniqueness, hence $a \sim_f a$
2. If $a \sim_f b$ for $a, b \in X$, then $f(a) = f(b)$. By symmetry of the equivalence relation $=$, $f(b) = f(a)$, and so $b \sim_f a$
3. Can the reader show that if $a \sim_f b$ and $b \sim_f c$, then $a \sim_f c$?

If f is not surjective, this equivalence relation may still be defined on the image of f. If f is injective, then each equivalence class has size 1. Many functions that are present in mathematics are both injective and surjective; to give a simple example:

$$f : \{1, 2, 3\} \to \{a, b, c\}, \qquad f(n) = \begin{cases} a & n = 1 \\ b & n = 2 \\ c & n = 3 \end{cases} \tag{1}$$

Sch functions are given a name:

Definition 0.1.19: Bijective

Let $f : X \to Y$ be a function. Then f is said to be *bijective* or *isomorphic in* **Set** if f is injective and surjective. If X and Y are bijective, then we may write:

$$X \cong_{\mathbf{Set}} Y \qquad X \cong Y \qquad X \xrightarrow{\sim} Y$$

The intuition of bijective functions can combine the intuition behind injective and surjective functions, however they may equivalently be seen as formally defining a notion of "size". As seen in equation 1, two sets that are bijective have the same size. The issue becomes more subtle when the domain/codomain are infinite:

Example 0.1.11: Size Of Infinite Sets

Define $2\mathbb{N} = \{2n \mid n \in \mathbb{N}\}$ and $f : \mathbb{N} \to 2\mathbb{N}$ by:

$$f(n) = 2n$$

Show that f is bijective.

This gives rise to the following definition:

Definition 0.1.20: Cardinality

Let X, Y be sets. Then X is said to have the same *cardinality* as Y if there exists a bijective function $f : X \to Y$. If X, Y have the same cardinality, they are said to have the same size.

Another important property of bijective functions is that they give rise to a *unique* inverse:

Proposition 0.1.21: Bijectivity And Unique Inverse

Let $f : X \to Y$ be a bijective function. Then there exists a unique function g such that

$$g \circ f = \mathrm{id}_X \qquad f \circ g = \mathrm{id}_Y$$

where $\mathrm{id}_X : X \to X$ and $\mathrm{id}_Y : Y \to Y$ are the identity maps on the respective sets. Often, the subscript on the identity maps are dropped and the following abuse of notation is carried out:

$$g \circ f = f \circ g = \mathrm{id}$$

The function g is often denoted f^{-1}

***Proof* :**

Since f is surjective, there exists an inverse function $g : Y \to X$ such that $f \circ g = \mathrm{id}_Y$. Since f is injective, there can only be one such function (there is only element in the fiber above each $y \in Y$). Furthermore, $\mathrm{im}(f) = Y$, hence g satisfies the condition $g \circ f = \mathrm{id}_X$, as we sought to show.

2: Remark: inverse notation abuse

The symbol f^{-1} has one other of notations that the reader should be aware, in particular this symbol may represent one of two things depending on the context. Given $f : X \to Y$:

1. if f is bijective, f^{-1} represents the unique inverse function
2. $f^{-1} : P(Y) \to P(X)$ is the function

$$f^{-1}(T) = S \in P(X) \qquad S = \{x \in X \mid f(x) = y\}$$

the images of f^{-1} are called the *pre-images* of f. This notation is particularly prevalent in topology where this the pre-image function takes center stage through continuous functions, however it will also have many applications in algebra.

Given any function $f : X \to Y$, there are certain natural functions that we may define that allows us to decompose f as a composition of injective and surjective functions with an intermediary bijective function

Proposition 0.1.22: Function Decomposition

Let $f : X \to Y$ be a function. Then there exists functions such that

$$\begin{tikzcd} X \arrow[r, two heads] \arrow[rrr, bend left, "f"] & X/\sim_f \arrow[r, "\sim"] & \mathrm{im}(f) \arrow[r, hook] & Y \end{tikzcd}$$

where the 1st function is surjective, the 2nd bijective, and the 3rd injective

Proof **:**
Define $\pi_f : X \to X/\sim_f$ by

$$\pi_f(x) = [x]_f$$

that is, π_f maps x to it's equivalence class. Next, define $F : X/\sim_f \to \mathrm{im}(f)$ by

$$F([x]) = y$$

where $y \in Y$ is chosen so that $f(x) = y$. Finally, define $\iota : \mathrm{im}(f) \to Y$ by

$$\iota(y) = y$$

that is, ι maps y to itself (since $\mathrm{im}(f) \subseteq Y$). The reader should verify that these maps have the desired properties.

This proposition gives an intriguing relation between equivalence relations and functions, namely any function may be "represented" as an equivalence relation, and equivalently any equivalence relation my be "represented" by a function.

Using functions, we may define the notion of a multiset:

Definition 0.1.23: Multiset

Let X be a set. Then a *multiset* given by elements of X is a function:

$$m : X \to \mathbb{N}$$

which assigns each element $x \in X$ a multiplicity.

Example 0.1.12: Multiset

Let $X = \{a, b, c\}$ and:

$$m(a) = 3 \qquad m(b) = 1 \qquad m(c) = 0$$

Then m defines a multiset. Commonly, this set is written as:

$$\{a, a, a, b\}$$

As practice, define another multiset n and show how to define the notion of inclusion, union, intersection, and so forth.

As a closing remark, it is a common convention in the study of Analysis (one of the major studies of mathematics) that if we have a function $f : X \to Y$, we might abuse notation and use a subset of X as the actual domain. For example, $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \frac{1}{x}$ is not a function as we've defined the concept, since $f(0) = \frac{1}{0}$ is not defined [28]. If we were to re-write this to be a function, we'd write $f : \mathbb{R} - \{0\} \to \mathbb{R}$. However, it is often implied in analysis that we will restrict the domain to the points for which f is defined.

Exercise 0.1.2

1. Let $f : X \to Y$ and take $A \subseteq X$, $B \subseteq Y$. Show that $f^{-1}(A) \subseteq B \;\Leftrightarrow\; A \subseteq f(B)$
2. Let $f : X \to Y$. Then is it true that $f(A \cup B) = f(A) \cup f(B)$? What about $f(A \cap B) = f(A) \cap f(B)$? What if there were more unions or intersections?
3. Let $f : X \to Y$, and define $f^{-1} : P(Y) \to P(X), f^{-1}(Y) = \{x \in X \mid f(x) \in Y\}$. Is $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$? What about $f^{-1}(A \cap B) = f^{-1}(A \cap B)$? What if there were more unions or intersections?

   (In a way, equivalence relations capture some idea of "projecting". If you have some set S an you want to put together all equivalent objects, where equivalence here can be thought as "similar enough so that we can define the concept of equality"
4. (partitioning property. Like creating the "new objects" where we ignore the "details" that is "consumed" by the equivalence relation)
5. (natural projection?? Give $\pi : G \to \{0\}$ a natural projection that always exists?)

[28] This would be covered rigorously in a calculus or analysis class

0.1.6 Adding Structures to Sets

So far, we have been extensively working with sets and functions. Sets by themselves have no notion of size, order, arithmetic operations, geometry, among others possible properties. Many interesting subjects in mathematics deal with these additional concepts. For this reason, mathematicians often "augment", so to speak, sets to have some extra structure:

Example 0.1.13: Extra Structure

1. (Order Structure). A set $(A, \leq)$ with an order $\leq$ is called an *ordered set*. Depending on the type of order given (ex. partial oder, total order, well order etc.), the set would be called the appropraited order (ex. paritally ordere set, totally ordered set, well ordered set, etc.)

2. (Algebraic structure) The set $\mathbb{Z}$ on its own is only numbers (and so is $\mathbb{N}, \mathbb{Q}, \mathbb{R}$). If we add a function $+ : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$, where $+$ acts exactly like the familiar addition we have learnt earlier in life, then $(\mathbb{Z}, +)$ now also has addition! We can add more functions to sets, for example $(\mathbb{Z}, +, \cdot)$ where $\cdot : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ does the usual multiplication. If we augment a set with function(s), we call this an *algebraic structure*. This whole book will be dedicated to the study of algebraic structures.

3. (Topological Structures). Another common augmentation of a set is something called a topology. Given a set A, a set $\mathcal{T}_A$ is a set of subsets of A that follows three rules:

   (a) $\emptyset \in \mathcal{T}_A$ and $A \in \mathcal{T}_A$

   (b) if $\{U_i\}_{i \in I} \subseteq \mathcal{T}_A$ is some arbitrary subset of $\mathcal{T}_A$ indexed by I, then $\bigcup_{i \in I} U_i \in \mathcal{T}_A$ (i.e. an arbitrary union of sets in $\mathcal{T}_A$ is in $\mathcal{T}_A$)

   (c) if $U_1, U_2, ..., U_n \in \mathcal{T}_A$ is a finite subset of $\mathcal{T}_A$, then $\bigcap_i U_i \in \mathcal{T}_A$ (i.e. a finite intersections of sets in $\mathcal{T}_A$ is in $\mathcal{T}_A$)

   Then $(A, \mathcal{T}_A)$ is called a *topology*. If a set is augmented with a sets of sets that follow these rules, it's called a *topological structure*. Topology is the natural space in which much of Real and Complex Analysis, as well as [non-classical] geometry takes place.

4. (Measure Structures). Similar to a topological structure, a measure structure is one where

   (a) $\emptyset \in \mathcal{T}_A$ and $A \in \mathcal{T}_A$

   (b) if $U_1, U_2, ..., U_n, ... \in \mathcal{T}_A$, then $\bigcup_i U_i \in \mathcal{T}_A$ (i.e. an countable union of sets in $\mathcal{T}_A$ is in $\mathcal{T}_A$)

   (c) if $U_1, U_2, ..., U_n, ... \in \mathcal{T}_A$, then $\bigcap_i U_i \in \mathcal{T}_A$ (i.e. a countable intersections of sets in $\mathcal{T}_A$ is in $\mathcal{T}_A$)

   This structure is central to generalizing the notion of "size" for general sets and is the backbone defining the integral[a].

5. (Differential structure) If you have done calculus, then you are familiar with the notion of being differentiable. There is a field of study in which we can call "sets" differentiable. Given a set S, there is a collection of functions called an *atlas* (say, $\mathcal{A}$ represents this collection) which gives a set S a "differentiable structure", and is usually denoted $(S, \mathcal{A})$. The details of what the atlas $\mathcal{A}$ contains is left for a course in differential geometry or diffeology.

6. (Model) In section 0.1.1, we have shown how to make logical statements given sets. The full study how sets (usually augmented with other structures) may have statements which are true or provable is the domain of *model theory*. Usually, a set M along with some collection of functions and relations is augmented in such a way to be able to study the properties of logical statements. This would be covered in a class on first order logic or model theory more generally.

[a] In particular, the Lebesgue integral

This textbook is dedicated to the study of algebraic structures. It will sometimes be the case that we encounter ordered structures [29], and only near the end of the book do we encounter topological structures (p. 790) where it is assumed the reader has taken a course covering point-set topology[30]. The reason the eventual mixing of structure is because different domains of mathematics start motivating each other in further exploration. For example, many geometers and algebraists found parallels between their fields, and hence algebraic geometry was born. Similarly, it was found that algebra can be used to solve topological problems, and hence algebraic topology was found. (A word on geometric group theory and something like LCH groups or Amenable groups?)

0.1.7 *Cardinality

(Here in there in he book, cardinality will come up. For ex, in proof that an injective map from $X \to X$ where X is finite is surjective, or when we figure out the algebraic numbers are countable, and other results later in the book)

(classical diagonalization proof?)

(as exercise, do $\{0,1\}^{\mathbb{N}}$ is bijective to $[0,1]$ by thinking of binary representation?)

[29] see section 2.4

[30] Differentiable structure's will be mentioned to motivate the concept of schemes which are based off of manifolds

0.2 Number Theory

Numbers are at the heart of mathematics and algebra. In this section, the basic properties of numbers will be explored. Some history of these concepts will be given as well to give a sense of why we call these special sets numbers.

Numbers started with our intuitive ability to understand discrete quantities (ex. 5 rocks, 2 people). In the past, it was contentious whether 0 was a number (the idea being that numbers are supposed to represent a real quantity), however it is now accepted that 0 represents the notion of "nothingness", "emptyness", or "inaction". Ratios of numbers (e.x 3/4) became important when it was important to be precise on what part of a whole object we want (ex. half the money). Negative numbers became important when considering other peoples debts (you owe 2 sheep, you "have" -2 sheep). In the times of at least the Greeks, it was discovered that there are certain numbers that are irrational (ex. $\sqrt{2}$ cannot be written as $\frac{a}{b}$, where a, b are natural numbers); the square root of any prime number shown to be irrational. The square-root of negative numbers remained undefined for awhile until NAME discovered that they are useful in finding roots of polynomials of degree 3, and it wans't till Euler that square-roots of negative numbers got the symbol $i := \sqrt{-1}$. It wasn't till later that it was understood there were many more irrational numbers than rational numbers; the collection of all rational and irrational numbers form the real numbers. The collection of all numbers $a + bi$, $a, b \in \mathbb{R}$ is known as the *complex numbers*. For reasons that shall be explored later in the book[31], any further addition to the number system beyond the complex numbers start to loose properties of numbers, and so we usually consider the following sets as "numbers" (ex. $ab \neq ba$ or $(ab)c \neq a(bc)$):

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$$

We shall heavily explore the properties of all of these number systems, with the notable exception of $\mathbb{R}$. This is because the real numbers are in many ways the beginning of the exploration of "taming infinities" and start moving beyond algebra. In fact, we shall even show in section F.1 that it is impossible to construct $\mathbb{R}$ in an algebraic way.

0.2.1 GCD and Euclidean Algoritm

Definition 0.2.1: Divisibility

Let a, b be numbers. Then we say that a *divides* b, or b is *divisible by* a if there exists a number k such that

$$ak = b$$

Note that divisibility depends on the set of numbers you are working, In $\mathbb{N}$, 3 is not divisible by 5, however in $\mathbb{Q}$, 3 is divisible by 5, namely let $k = 3/5$.

Example 0.2.1: Divisibility Practice

1. Show that if $x \big| n_1$ and $x \big| n_2$, then $x \big| an_1 + bn_2$ for any $a, b \in \mathbb{Z}$.
2. Show that if $a \big| b$ and $b \big| c$ then $a \big| c$

[31] ref:HERE for details

From now on in this section, we shall limit ourselves to divisibility in $\mathbb{Z}$[32].

One special result about integers is that every number has a prime factorizaiton:

Proposition 0.2.2: Prime Factorization Of Natural Numbers

Let $n \in \mathbb{N}$. Then there exists distinct prime numbers $p_1, p_2, ..., p_k$ and natural numbers $n_1, n_2, ..., n_k$ such that n can be uniquely broken down into:

$$n = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

Proof **:**
exercise (hint: use induction)

This definition can be extended to $\mathbb{Z}$, except we loose uniqueness, the decomposition is unique up to a multiple of -1. This isn't too much of a concern, since -1 has the special property of having a multiplicative inverse, that is we may multiply -1 by a number (in this case, itself) to get to 1:

$$-1 \cdot -1 = 1$$

no other integer (besides trivially 1) can be multiplied by another integer to get to 1. Since all numbers can be broken don into multiples of primes, we can define a ay to compare ho similar to numbers are:

Definition 0.2.3: Greatest Common Divisor

Let $a, b \in \mathbb{Z}$ be to non-zero integers. Then the *greatest common divisor* between these to numbers is the greatest number n such that n divides a and n divides b.
The common notation for this fact is that:

$$n \big| a \qquad n \big| b$$

Similarly, we have the dual concept:

Definition 0.2.4: Least Common Multiple

Let $a, b \in \mathbb{Z}$ be non-zero integers. Then the *least common multiple* is the smallest integer m such that

$$a \big| m \qquad b \big| m$$

Example 0.2.2: Greatest Common Divisor

1. $\gcd(10, 5) = 5$
2. $\gcd(100, 90) = 10$
3. $\gcd(45, 49) = 1$
4. If p, q are distinct prime numbers, then $\gcd(p, q) = 0$

[32] In chapter 10, we shall re-define divisibility in a more general context

It is natural that two prime numbers have 1 as their greatest common divisor for they have no factor in common. However, it is not necessary for numbers to be prime for their gcd to be 1. We give a name to numbers with no common factors:

Definition 0.2.5: Coprime Numbers

Let $a, b \in \mathbb{N}_{>0}$. Then if $\gcd(a, b) = 1$, a and b are said to be *coprime*.

We may find the greatest common multiple by finding the prime factorization, however this is in general very inefficient on large numbers. These values can be found in an algorithmic manner relying on the fact that numbers have a notion of size[33]

Theorem 0.2.6: Euclidean Algorithm

Given two natural numbers a, b we can find their greatest common divisor by the following technic:

1. Let $a > b$. Put them into the equation $a = bk_1 + r_1$, where k is the amount of times b fits into a, and r_1 is the remainder.
2. Once you find these values, all but k shift over to the left, and you repeat the process to find r_2:

$$\begin{aligned} b &= r_1 k_2 + r_2 \\ r_1 &= r_2 k_3 + r_3 \\ r_2 &= r_3 k_4 + r_4 \end{aligned}$$

3. The algorithm continues until $r_n = 0$ (which it will reach since a, b where finite). The final non-zero remainder r_{n-1} is the greatest common divisor.

Proof **:**
exercise

As an example:

Example 0.2.3: Euclidean Algorithm
Let $a = 20$, $b = 13$. Then:

$$\begin{aligned} 20 &= 1 \cdot 13 + 7 \\ 13 &= 1 \cdot 7 + 6 \\ 7 &= 1 \cdot 6 + 1 \\ 6 &= 6 \cdot 6 + 0 \end{aligned}$$

[33] We shall see in chapter 10 that not all systems in which we can define a GCD will allow for a euclidean algorithm

so $\gcd(20, 13) = 1$, (the number above 0 in the last step). To find the $ax + by = 1$ formula, we reverse the process by converting the equation above in the following form:

$$7 - 6 = 1$$

so, put the remainder on the right hand side. Then, isolate the next equation in in the appropriate ways to substitute into this one: $7 = 13 - 6$, $6 = 13 - 7$. Since $6 = 13 - 7$, $7 = 13 - (13 - 7) = 7$, so

$$7 - (13 - 7) = 1$$

then, we repeat the same process with the step above: $7 = 20 - 13$, so:

$$20 - 13 - (13 - (20 - 13)) = 1$$

implying, we get

$$2 \cdot 20 - 3 \cdot 13 = 1 \qquad (40 - 39 = 1)$$

giving us our desired equation.

Using the euclidean algorithms, we may always find the solution to the following interesting linear equation:

Theorem 0.2.7: Extended Euclidean Algorithm (Bézout's Identity)

Let $a, b \in \mathbb{Z}$. Then there exists integers $x, y \in \mathbb{Z}$ such that

$$ax + by = \gcd(a, b)$$

Proof **:**
exercise (reverse the process)

Example 0.2.4: Bézout's Identity
Recall that $\gcd(20, 13) = 1$. To find the $ax + by = 1$ formula, we reverse the process by converting the equation above in the following form:

$$7 - 6 = 1$$

so, put the remainder on the right hand side. Then, isolate the next equation in in the appropriate ways to substitute into this one: $7 = 13 - 6$, $6 = 13 - 7$. Since $6 = 13 - 7$, $7 = 13 - (13 - 7) = 7$, so

$$7 - (13 - 7) = 1$$

then, we repeat the same process with the step above: $7 = 20 - 13$, so:

$$20 - 13 - (13 - (20 - 13)) = 1$$

implying, we get

$$2 \cdot 20 - 3 \cdot 13 = 1 \qquad (40 - 39 = 1)$$

giving us our desired equation.

Try doing the same with 220 and 38.

Bézout's Identity has the following neat interpretation: primes are fundamentally a multiplicative decomposition of numbers: any n is equal to $p_1^{n_1} \cdots p_k^{n_k}$ for appropriate primes p_i and powers n_i. However, they do have an additive relation between them, the Bézout identity. If we have two primes, we can (up to appropriate constants) add them to get to the multiplicative identity 1. In this way, we have a property of primes given via addition.

(should I add other number theory identities? Are there any other useful ones for this book? Perhaps just as exercises? They might be useful way later in algebraic number theory)

Exercise 0.2.1

1. Let $p, q, r \in \mathbb{N}$ be prime numbers greater than 3. Show that 3 divides $p^2 + q^2 + r^2$

0.2.2 Modular Arithmetic's

When counting the hours on a clock-face, we start with 12 o-clock, then add an hour to make it 1 o-clock, then another hour to make it 2 o-clock, until we circle back to 12 o-clock again, at which point we start counting again to 1 o-clock. For the rest of this section, we will say that 12 o-clock is 0 o-clock for reasons that would be explained soon. Notice how the arithmetic's of the clock different from the usual number system where for each $x \in \mathbb{N}$, there exists an n such that $n > x$. Hours on a clock could add to be zero, for example:

$$7 + 5 = 0$$

In this section, we rigorously define this notion

Definition 0.2.8: Modular Arithmetic's

Let $n \in \mathbb{N}_{>0}$. Define the equivalent relation on $\mathbb{Z}$:

$$a \sim_n b \quad \Leftrightarrow \quad a - b = kn \text{ for some } k \in \mathbb{Z}$$

The collection of these equivalences classes is denoted $\mathbb{Z}/n\mathbb{Z}$[a] and is called the *integers modulo n*. The number n is called the *modulus* and if $a \sim_n b$, we say that *a is congruent to b* modulo n. If $a \sim_n b$, then it is denoted

$$a \equiv b \bmod n \qquad \text{or} \qquad a \equiv_n b$$

The equivalence classes may be denoted as:

$$[a],\ [a]_n,\ \overline{a},\ \overline{a}_n$$

where the n is dropped if the context permits it.

[a] This notation shall be elucidated on in chapter 3

Intuitively, this equivalence relation says that the difference of two numbers is a multiple of n. When $n = 12$, then this is the usual clock arithmetic's. When n is any other number, the reader may imagine a clock-face with n digits.

Naturally, it has to be proven that $\sim_n$ is indeed an equivalence relation:

Proposition 0.2.9: Modular Arithmetic's Well-defined

Let $\sim_n$ be defined as in definition 0.2.8. Then $\sim_n$ is indeed an equivalence relation

***Proof* :**
Let $x, y, z \in \mathbb{Z}$ and let $n \in \mathbb{N}_{>0}$ be any positive natural number.

1. $x \equiv x \bmod n$ since $x - x = 0 = 0 \cdot n$
2. If $x \equiv y \bmod n$ then there exists a k st $x - y = kn$. Then $y - x = -kn = (-k)n$. Then by definition $y \equiv x \bmod n$
3. If $x \equiv y \bmod n$ and $y \equiv z \bmod n$, then $x - y = kn$ and $y - z = \ell n$ for some $k, \ell \in \mathbb{Z}$. Adding these two equations, we get:

$$x - z = x - y + y - z = (k + \ell)n$$

and so $x \equiv z \bmod n$

Since all the conditions of an equivalence relation is satisfied, $\sim_n$ is indeed an equivalence relation, as we sought to show.

Example 0.2.5: Modular Arithmetic's
Verify the following:

1. $35 \equiv 11 \bmod 12$
2. $2 \equiv -3 \bmod 5$
3. If $a \equiv b \bmod n$, then $a + t \equiv b + t \bmod n$ for any $t \in \mathbb{Z}$
4. If $a \equiv b \bmod n$, then $a^t \equiv b^t \bmod n$ for any $t \in \mathbb{N}$

Proposition 0.2.10: Properties Of Modular Arithmetic's

Let $n \in \mathbb{N}_{>0}$ and $a_1 \equiv a_2 \bmod n$, $b_1 \equiv b_2 \bmod n$. Then:

1. $a_1 + b_1 \equiv a_2 + b_2 \bmod n$
2. $a_1 - b_1 \equiv a_2 - b_2 \bmod n$
3. $a_1 b_1 \equiv a_2 b_2 \bmod n$

Furthermore:

1. $a_1 + b_1 \bmod n = a_1 \bmod n + b_1 \bmod n$ (in different notation, $\overline{a_1 + b_1} = \overline{a_1} + \overline{b_1}$)
2. $a_1 b_1 \bmod n = (a_1 \bmod n)(b_1 \bmod n)$ (in different notation, $\overline{a_1 b_1} = \overline{a_1}\,\overline{b_2}$)

Proof **:**
This is a good exercise

Proposition 0.2.11: Cancellation Properties Of Modular Arithmetic's

Let $n \in \mathbb{N}_{>0}$. Then:

1. if $a + k \equiv b + k \bmod n$ for any integer $k \in \mathbb{Z}$, then $a \equiv b \bmod n$
2. if $ka \equiv kb \bmod n$ for k coprime with n ($\gcd(k, n) = 1$) then $a \equiv b \bmod n$
3. if $ka \equiv kb \bmod kn$ for some $\mathbb{N}_{>0}$, then $a \equiv b \bmod n$

Proof **:**
good exercises

Important Results

The following are important properties of the integers modulo n:

Theorem 0.2.12: Fermat's Little Theorem

Let p be prime and $a \in \mathbb{N}_{>0}$ such that $p \nmid a$. Then

$$a^p \equiv a \bmod p$$

Proof **:**
First, we can assume that $0 \leq a \leq p - 1$, for we may always reduce a due to the properties of modular arithmetic's (hint: use the multiplicative properties of modular arithmetic's). Next, we may reduce this assertion to

$$a^{p-1} \equiv 1 \bmod p$$

For, if the above hold, then multiplying both sides by a gives the desired result.

Start by list out the first $p - 1$ positive multiples of a:

$$a, 2a, 3a, ..., (p-1)a \tag{2}$$

I first claim that no two of these are equal $\bmod p$. For the sake of contradiction, let's say two were so that $ra \equiv sa \bmod p$ for some $1 \leq r, s \leq p - 1$ where $r \neq s$. Then since p is prime and $a > 0$, $\gcd(a, p) = 1$ an hence by proposition 0.2.10:

$$r \equiv s \bmod p$$

Since $0 \leq r, s \leq p - 1$, this implies $r = s$; contradiction our earlier claim. Hence, each term in list (2) is congruent (modulo p) to a unique number $1, ..., p - 1$.

Now, multiply the terms in list (2). Then by proposition 0.2.10:

$$(1a)(2a) \cdots ((p-1)a) \equiv (1)(2) \cdots (p-1) \bmod p$$

which, when simmplifying by combining terms, is:

$$a^{p-1}(p-1)! \equiv (p-1)! \bmod p$$

Since p is prime, each number $1 \leq k \leq p-1$ is coprime to p, and hence by repeated application of proposition 0.2.11:

$$a^{p-1} \equiv 1 \bmod p$$

but now, we may simply multiply both sides by a to get $a^p \equiv a \bmod p$, as we sought to show.

The reader may ask if this can be generalized in some some way to modulo n. The key would be to generlize the step in which each number $1 \leq k \leq p-1$ is coprime with n, and indeed if $\varphi(n)$ represents the number of numbers that are coprime with n, then the identity can be generalized to:

$$a^{\varphi(n)} \equiv 1 \bmod n$$

This proof is best in the context of group theory, and so shall be left alone till then, however the persistent reader has all they need to attempt to directly prove this result.

(There are more cool proofs, but many rely on group theory or field theory. I'd like to incldue them in certain sections of the book: here)

Exercise 0.2.2

1. Let $a \equiv b \bmod n$. If $k \big| a, b$, is it the case that $(a/k) \equiv (b/k) \bmod n$?

0.2.3 Counting Problems

(Also introduce the binomial theorem?) (somehow introduce the binomial theorem. Something else too?)

0.3 Graphs

mention how this is section is not necessary to read for groups, but it will come back up when working with particular perspective on groups. It can be revisited once you make it there (perhaps make the graph intuition of groups a series of exercises?)

(See this link: (commented for compiling purposes))

We first introduce a more concrete definition of a graph in order to be able to work with it unambiguously

Definition 0.3.1: Graph

Let $V = \{v_1, v_2, ..., v_n, ...\}$ be a set of elements representing the *vertices* (also called *nodes* or *points*). Let

$$E = \{\{v_i, v_j\} \mid v_i, v_j \in V\}$$

be the set representing the *edges* of the graph (where there need not be a set $\{v_i, v_j\}$ for every pair of vertices in V). Then $\mathfrak{G} = (V, E)$ is called a *graph*.
If the edges are *pairs* instead of sets, then we call $\mathfrak{G}$ a *direct graph*

Example 0.3.1: Graphs
put different types of graphs as example? infinite, connected, cyclic, tree, completed

1. here + visual
2. here + visual
3. here + visual
4. here + visual

(define endpoints of a vertex, adjacent vertices, and if $e = \{u, v\}$ e is also called the incident of u and v)

(define simple graph: E has no repeated elements. If E can have repeated elements, it's a multigraph)

Definition 0.3.2: Graph Isomorphism

Let $\mathfrak{G} = (V_{\mathfrak{G}}, E_{\mathfrak{G}})$ and $\mathfrak{H} = (V_{\mathfrak{H}}, E_{\mathfrak{H}})$ be two graphs. A function $f : V_{\mathfrak{G}} \to V_{\mathfrak{H}}$ is called a *graph isomorphism* if for every vertex adjacent vertices $v_i, v_j \in V_{\mathfrak{G}}$, we have

$$(v_i, v_j) \in E_{\mathfrak{G}} \;\Leftrightarrow\; (f(v_i), f(v_j)) \in E_{\mathfrak{H}}$$

that is, f preserves adjacencies between vertices.
If $\mathfrak{G}$ and $\mathfrak{H}$ are isomorphic, we write $\mathfrak{G} \cong \mathfrak{H}$, and we say that $\mathfrak{G}$ is isomorphic to $\mathfrak{H}$

Example 0.3.2: examples of Graph Isomorphisms
here

These will come back in a exercise ref:HERE

0.3.1 Lattices

(These I think are important because they will be periodically mentioned. From the beginning with the partial ordering of subsets via inclusion, to using Zorn's lemma, to group/ring/module lattices, to the correspondence between the latice of divisibility in UFDs and multi-sets of irreducible elements, to the many galois correspondences we shall see, to Quivers)

Grillet has a good section on Lattices, so does Lang.

Exercise 0.3.1

1. tree is a-cyclic
2. probably counting vertices and edges problems?
3. other simple and fun graph theory problems?

0.4 Categories

(maybe at the end give the definition of a category just to expose the reader to it?)

In the last couple of section, we have been focusing on understanding mathematical objects, in particular, we tried constructing different types of objects(like sets, numbers, graphs, and binary relations) and tried showing some properties of these objects. In this section, we will focus less on the mathematical object, and more on the *relation* between mathematical objects that preserve some of the same "structure". The nouns "structure" and "relation" will be given more precise meaning through the use of *categories*.

Categories have a reputation in mathematics to be difficult to understand – a reputation I (the author of this book) do not see where it comes from. The hardest part about categories is a shift in perspective in how to think about mathematical objects (or objects more generally). Thus, before getting into what focusing on what "structure-preserving relations" means in mathematics, let's first give an analogy to get a more intuitive understanding in what we are trying to achieve:

> Let's take any Language (ex. English). We can study any given word to get some properties of it (for example, its grammatical properties, syntactic correctness, its phonetic properties in a dialect, and so forth). However, what we usually care about is not the properties of a word, but how words *relate* to other words. The meaning behind a sentence doesn't come just from the individual words, but from how the words "come together": This is called semantics [34]. As an example of semantics, we can even see that meaning words is defined in relation to other words in a sentence, i.e., how the word relates to other words. For example the word *cold* usually means the absence of heat. However, If I say "Those people are giving him a cold stare", I am not referencing to the temperature of their gaze, but some lack of friendliness in their feelings towards him – the meaning of the word is dependent on how it relates to other words. [35]
>
> If we look at language through the lens of the relation between words, we can get some insights on how language is used. An excellent example of this is translation. Let's say we're trying to translate from English to Sanskrit (or really any two languages). Both English and Sanskrit are languages, and so they have words (i.e. the objects) and rules (i.e. the relation between objects). Don't worry for a moment about exceptions to rule in particular languages (pretend they don't exist for the sake of the analogy), the main idea is that English and Sanskrit (and all languages) have words and rules that govern their use. When we compare English to Sanskrit, what we will do is see how the relations of the words in English (ex. conjugations, articles, etc.) link to those in Sanskrit (ex. declension, compounds, etc.). In a sense, we don't care about what the particular words of the respective language are. What we care is to find the word form one language that will create the same relations in the other language.
>
> For example, If I say "dog" in English, then the sentences "The dog barked", "I took the dog for a walk", or "Many dogs like to chase cats" all make sense, while "I dog the took for a walk" is not a sentence that makes sense, since the position of the word "dog" and "take" are wrong given the grammatical rules in the english languages. In Sanskrit, the grammatical rules are different: If we directly translate the sentence "I dog the took for a walk" to Sanskrit, the sentence will actually make sense! That is because the rules of

[34] Semantics is the study of the meaning that sentences (or any discourse) carries

[35] This is naturally a simplification

> Sanskrit are different than English. In fact, it is a little more subtle than this: There is no "a" (like "a walk") or "the" in Sanskrit – if you're translating from English to Sanskrit, we can't replace "a" with a Sanskrit word (since non exist). Instead, we will make sure that the words we choose in Sanskrit when translating the English sentences will have the appropriate relations in order to capture the same meaning

The entire point of this analogy is to show that it is useful to understand objects not how they "intrinsically" are, but how they "in relation" to other objects [36]. Thinking about mathematical objects by ho they relate to one another will let us be able to compare object by seeing if one mathematical object has the same type of relations with other objects.

e have already done this before: when we worked with ordered structures (ex. $(\mathbb{N}, \leq)$), we studied objects by how they related to other objects. We can even make definitions based on how objects relate. For example, in $(\mathbb{N}, \leq)$, define the *smallest* to be the object x such that for every other element $y \in \mathbb{N}$, $x \leq y$. Then it is clear that 0 is the smallest object. Hover, in $(\mathbb{Z}, \leq)$ no object is the smallest object, because for any $x \in \mathbb{Z}$, there exists $x - 1 \in \mathbb{Z}$ such that $x - 1 \leq x$. Another example of a definition based on ho an object relate, let X be the set for which any function from X is *always* injective. Then X is any singleton.

In the following section, we will try to more rigorously define a ay to talk about the relation between objects

0.4.1 Defining Categories

Let's start by giving a high-level idea of that a category is:

(Maybe say something else like: originally, we had the simple object of set that contained only information about objects. Now we will define an structure which will hold both information about relations and objects. These objects will be called categories).

Definition 0.4.1: Pre-Definition Of Category

A Category C is a collection [a] of objects and relations between objects.
We denote the collection of objects of C by $\mathrm{Obj}(C)$. For any to objects $A, B \in \mathrm{Obj}(C)$, $\mathrm{Hom}_C(A, B)$ denotes the collection of relation between objects, which we call *morphisms*. In particular, if $f \in \mathrm{Hom}_C(A, B)$, then f is called a morphism.

[a] the word *collection* is used instead of *set* for a technical reason. More on that after this definition

The fact that we said *collection* instead of *set* is because of a quirk of reason. In particular, once we properly define what a category is, we will show that sets and functions between sets will form a category. Then $\mathrm{Obj}(\mathbf{Set})$ will have to be *a set of all sets*, but it can be proven that no such set exists! This is called *Russel's Paradox*, and is addressed in chapter 30.

(Give some pre-definition of category? A collection of objects and relationships? So C is a category, $\mathrm{Obj}(C)$ are the objects, and for $A, B \in \mathrm{Obj}(C)$, $\mathrm{Hom}_C(A, B)$ is the collection of all relation between objects?

(meaning isomorphism)

[36] The philosophy behind this idea is called *structuralism*

Throughout the book, we'll be introducing more and more algebraic structures of varying levels of similarity and differences. It would be very convenient to be able to have a way to compare these algebraic structures. What we elaborate in this section will be a sort of structure for structures: a *Category*. When we introduce new algebraic

1. Bring over some stuff from the group-section to here
2. quotient objects.
3. products (fibered product as an advanced example)
4. An attempt at intuition behind universal properties? Perhaps the best first example is when we see free objects?????
5. when talking about the category of set, you can say that since we care about the relation between objects, we loose some details about sets. We won't really know where elements from one set map to another, just how they map. So, in this sense, the cardinality of sets becomes really important, since that determines what types of maps we can make (ex. injective, surjective).
   (a) Looking at sets in terms of their cardinality will actually permeate a lot of what we'll do. (Can you figure out how to explain this here??) Instead of saying F_S for a free-group, we'll write F_n where $|S| = n$, since every set S with the same cardinality are isomorphic in **Set** and free groups are defined on a set *up to isomorphism* so we will write F_S. Similarly for S_n, $\mathbb{R}^n$, and so on. Basically, wherever we see a number, we can interpret that as a secret use of this "up to isomorphism in **Set**" fact.

(lot's of words and build-up) HEREHERE

(Essentially, Given an equivalence relation $\sim$ on a set A if we want to find the object that "represents" the equivalences classes, then the object $A/\sim$ will be the on in which captures the right amount of information such that any other object that tries to captures it is always linked to $A/\sim$. In other words, $A/\sim$ is in a sense always in an object that tries to represent the equivalence classes of A)

Theorem 0.4.2: Universal Property Of Quotient Sets

Let A be a set and $\sim$ be an equivalence relation on A. Let $f : A \to Z$ be any set function such that $a \sim b \Rightarrow f(a) = f(b)$. Then there exists a unique map φ from $A/\sim$ to Z such that the following diagram commutes

$$\begin{tikzcd} A/\sim \arrow[r, "F"] & Z \\ A \arrow[u, "\pi"] \arrow[ur, "f"] & \end{tikzcd}$$

Proof **:**
Let's say $f : A \to Z$ respected where $a \sim b \Rightarrow f(a) = f(b)$. we want to ask if there exists a unique map such that

$$f = F \circ \pi$$

that is

$$f(x) = F(\pi(x))$$

Consider the map $F : A/\sim \to Z$, $F([x]) = f(x)$. This map is well-defined since if $[x] = [y]$, then $x \sim y$ so $f(x) = f(y)$.

Furthermore, this map is unique. If $F, F' : A/\sim \twoheadrightarrow Z$ were two well-defined maps, then it must be that $F(\pi(x)) = f(x) = F'(\pi(x))$ making $F = F'$, showing uniqueness, completing the proof.

0.5 Matrices

This section can be comfortably skipped until chapter 13.2. It will also be referenced in section 1.2.5 for examples of groups, but it is skippable without any loss (since it's just one out of over a dozen example of groups that will be presented).

In this book, we will work with many structures. Some of them will be very different from one another, while others share striking similarities. Many of these more similar mathematical structures will be able to be "represented" by matrices. We will define more rigorously for something to be "representable" by matrices later – for now we want to gain some basic appreciation and understanding of how to use matrices.

(History going back to China, this is the early motivator for matrices compared to the "new" one of being a useful classification theorem. Define a matrix. Show it's used to solve systems of linear equations. Explain Gaussian elimination)

(Later, Gauss developed a way to multiply matrices. The intuition shall come in chapter 13)

Part I

Groups

Groups have entered the scene of mathematics in quite a bombastic way.

(Develop on Galois' duel and how it went)

(what Galois jotted down)

(if was not made in a vacuum though, Lagrange __ years earlier can be said to be studying them, but it largely left unnoticed till Galois and soon after Cauchy's work)

(What did Galois show? Develop impossibility of the quintinc and he used this new thing he invented which will eventually be called group)

(Around the same time, two other mathematicians, Sophus Lie and Felix Klein, where studying __ in geometry, and came up with groups themselves (they called it __))

(at the same time, Gauss and Kronecker were using what would essentially be called groups to solve number theory problems)

(Each of these disciplines were using similar objects and tricks without noticing it. In __, Camille Jordan and Ernst Kummer these parallels and wrote a book defining a group on a more abstract level)

(Further development in the 19th and 20th century? Use in physics, computer science, chemistry, cryptography?)

(So is abstraction good or bad? Well, both the views above are correct. Abstraction does indeed play a very important clarifying role, by showing us that many apparently different phenomena are basically the same, and isolating the aspects of those phenomena that really matter. However, if a group is defined for us in an abstract way (I'll say more precisely what I mean by this later), then showing that it is isomorphic to a group of symmetries can make it much easier to answer questions about that group.)

1

Introduction to Groups

In this chapter, you will cover the basic properties of groups. In the process, we will prove statements that you perhaps thought were axiomatic. For example, if $ab = ac$, does that mean $b = c$? We will classify two very important sub-categories: groups with the axiom of *commutativity* ($ab = ba$), and those without it. These two categories of groups will set the narrative for how we study groups, meaning this one axiom plays an incredibly central role in the study of groups! We will cover many examples of groups to get a taste on what groups are, and define the groups which will be referenced indefinitely throughout the rest of the book. We will then cover how to compare two groups, and define the notion of equality between two groups. We'll then see how we can use groups to "represent" a system of objects. For example, we'll use a group to represent all possible ways of rotating an object. In this case, the system will consist of all possible rotations of the object. We'll conclude with a second perspective on groups by looking at the collection of all groups and introduce the mathematical language used to describe such a collection using Category Theory.

1.1 Basic Properties and Definitions

As mentioned in the preface, algebraic structures are sets along with a function (or functions) which must follow certain rules. The most common type of function used for an algebraic structure is called a *binary operation*:

Definition 1.1.1: Binary Operation

1. A *binary operation* $*$ on a set G is a function $* : G \times G \to G$. For any $a, b \in G$, we shall write $a * b$ or $*(a, b)$. Often the notation is abbreviated to ab.
2. A binary operation $*$ on a set G is *associative* if for every $a, b, c \in G$ we have $a * (b * c) = (a * b) * c$
3. If $*$ is a binary operation on a set G we say elements a and b *commute* if $a * b = b * a$. We say $*$ (or G) is *commutative* if for every $a, b \in G$, $a * b = b * a$

The star of the show in group theory is the binary operation. This function is the key mathematical structure added to a set which we study when talking about groups. When talking about different group structures, we're often talking about different binary operations.

Example 1.1.1: Binary Operations

1. Arithmetic operations are a good source of binary operations. As we've discussed in the preliminary chapter, $+$ and $\times$ are functions that take in two numbers as input, and output one number. If you want to practice proving something is a function, you can try showing that $+ : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ is a function (and since the domain is the cartesian product of two copies of the codomain, it's a binary operation). Similarly, you can do the same with $- : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$.
2. Is $-$ a binary operation with $\mathbb{N}$ (i.e. is $- : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ a binary operation)? What about $\times$ and $\div$? Which of these binary operations are associative? Are they commutative? What if you switch $\mathbb{N}$ with $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$?
3. When working with sets, it is common to take the union or intersection of sets. These are binary operations. If A is a set and $P(A)$ is its powerset, then $\cup : P(A) \times P(A) \to P(A)$ is a binary operation. Similarly for $\cap$
4. Take $\mathbb{Z}$, and recall the *gcd* and *lcm* from the preliminary chapter (ref:HERE). Then $gcd : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ is a binary operation. Similarly for *lcm*. Can you replace $\mathbb{Z}$ with $\mathbb{N}$?
5. If you have taken some multi-variable calculus, then the cross product $\times : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ is an example of a binary operation.
6. Take $E : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ to be $E(x, y) = x^y$. Check that $E(x, E(y, z))$ is not necessarily equal to $E(E(x, y), z)$ (that is $x^{(y^z)}$ is not necessarily equal to $(x^y)^z$). Hence, exponentiation is a non-assotiative operation.
7. Given two functions f and g over the same domain and codomain ($f : X \to X$ and $g : X \to X$), their composition. Thus composition of functions with the same domain and codomain forms a binary operation.

With these intuitions, we can move on to defining a group:

Definition 1.1.2: Group

A *group* is an ordered pair $(G, *)$, where G is a set and $*$ is a binary operation on G satisfying the following axioms:

1. associativity: $(a * b) * c = a * (b * c)$
2. there exists an element e in G, called the *identity* of G, such that for every $a \in G$, we have $e * a = a$
3. For each $a \in G$ there is an element $b \in G$ called an *inverse* of a, such that $ab = ba = e$. We denote this element b by a^{-1}, so that we can equivalently write $a^{-1} * a = e$

The group $(G, *)$ is called *abelian* (or commutative) if $a * b = b * a$ for every $a, b \in G$.

Commonly, if $(G, *)$ is a group, we say that G is a group under $*$.

Note: Some authors like to make it explicit that it's closed under the operation, meaning if $a, b \in G$, then $a * b \in G$. However, that is implied in the definition of binary operation (i.e. a function), since the codomain of $*$ is G.

Example 1.1.2: First Encounter With Groups

In the following examples, when we're working with numbers, think of them only as a set, and define the operations $+$, $\times$ or others in the way they were intuitively defined when learning arithmetic's.

1. $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$ are groups. We would say, for example, that $\mathbb{Z}$ is a group under $+$. For each of these examples, we would have the identity to be $e = 0$, and for each a, $a^{-1} = -a$. Note that technically, the $+$ in each group is not the same $+$, because the domains and co-domains are different. Note that $\mathbb{N}$ is *not* a group under $+$, since not every element has an inverse: There does not exist an $a \in \mathbb{N}$ such that $2 + a = 0$ (i.e. $-2 \notin \mathbb{N}$). [a]
2. $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ under $\times$ are *not* groups. $\mathbb{Q} - \{0\}$, $\mathbb{R} - \{0\}$ $\mathbb{C} - \{0\}$, $\mathbb{Q}_{>0}$, $\mathbb{R}_{>0}$ are groups under $\times$, with $e = 1$, and $a^{-1} = \frac{1}{a}$. Why is that? Take a moment to think about why removing the zero will make those sets groups. Is $\mathbb{Z} - \{0\}$ a group under $\times$? Is $(\mathbb{Z} - \{0\})$ a group under $+$?
3. For any $n \in \mathbb{N}$, $\mathbb{Z}/n\mathbb{Z}$ is a group under modulo $+$ $(\mathbb{Z}/n\mathbb{Z}, +)$. Can you show that 0 is the identity? Can you show that every number has an inverse? Note that $((\mathbb{Z}/n\mathbb{Z}), \times)$ under $\times$ *not* a group in general! It *is* a group when p is prime. You can actually pick some non-prime n and try it (ex. $n = 4$).The general proof that $(\mathbb{Z}/n\mathbb{Z}, \times)$ is a group when n is prime will come later (after we introduce theorem 2.3.8)
4. If $(A, \bullet_a)$ and $(B, \bullet_b)$ are groups, we can form a group on their cartesian product $A \times B$ called the *direct product group*, whose elements are those in the Cartesian product
$$A \times B = \{(a, b) | a \in A, b \in B\}$$
and the binary operations is defined point wise:
$$(a, b) * (c, d) = (a \bullet_a c, b \bullet_b d)$$

We can denote this group and its binary operation $(A \times B, (\bullet_a, \bullet_b))$. A common example of a direct product of group is $(\mathbb{R} \times \mathbb{R}, (+, +))$, with the usual notation of $\mathbb{R}^2$. We will dedicate a lot of energy understanding the details of the cartesian product of groups in chapter 5.

5. Define $+_2 : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ $x +_2 y = 2 + x + y$. The operation $+_2$ is associative, since

   $$(x +_2 y) +_2 z = (2 + x + y) +_2 z = (2 + x + y) + 2 + z = 2 + x + (2 + y + z) = x +_2 (y +_2 z)$$

   Furthermore, $e = -2$, since
   $$x +_2 -2 = x + 2 - 2 = x$$
   and $x^{-1} = -x - 4$ since
   $$x^{-1} +_2 x = -x - 4 + 2 + x = -2$$
   showing that $(\mathbb{Z}, +_2)$ is indeed a group. Notice that $(\mathbb{Z}, +_2)$ and $(\mathbb{Z}, +)$ are different (for example, their identity is different)[b].

6. $G = (\mathrm{Mat}_n(\mathbb{R}), \cdot)$ is almost forms a group. A binary operation exists, identity exists, and associativity holds. However, there are non-invertible matrices, meaning that $\forall M \in G$, it's not necessarily the case that $\exists N \in G$ such that $NM = \mathrm{I}$. We will explore this in section 1.2.5

7. Take $\{e\}$ to be the set with a single element, an a binary operation $\cdot$ such that $e \cdot e = e$ (that is, the binary operation that treats e like an identity). Then $(\{e\}, \cdot)$ is a group! This group is called the *trivial group*.

8. Take a set A and its powerset $P(A)$. Is $(P(A), \cup)$ a group? What about $(P(A), \cap)$? answer: [c]

9. Take $(\mathbb{Z}, -)$, where $-$ is the operation of subtraction. Then $(\mathbb{Z}, -)$ is not a group! Can you see which axiom fails?

10. Here is a more complex example which fails for the same reason as the previous one. Take $(\mathbb{R}^3, \times)$ where $\times$ represents the cross product:

    $$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_1 b_3 - a_3 b_1 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}$$

    or, if you've taken a physics or linear algebra class, you might have seen it defined as

    $$a \times b = \det \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

    where we represent the three components of the vector as $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, and we use the determinant to shorten notation. Then $(\mathbb{R}^3, \times)$ is *not* a group! The cross product is not associative! Take

    $$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$$

Then

$$\begin{pmatrix}1\\2\\3\end{pmatrix} \times \begin{pmatrix}2\\3\\4\end{pmatrix} = \begin{pmatrix}-1\\2\\-1\end{pmatrix} \qquad \begin{pmatrix}-1\\2\\-1\end{pmatrix} \times \begin{pmatrix}3\\4\\5\end{pmatrix} = \begin{pmatrix}14\\2\\-10\end{pmatrix}$$

and

$$\begin{pmatrix}2\\3\\4\end{pmatrix} \times \begin{pmatrix}3\\4\\5\end{pmatrix} = \begin{pmatrix}-1\\2\\-1\end{pmatrix} \qquad \begin{pmatrix}1\\2\\3\end{pmatrix} \times \begin{pmatrix}-1\\2\\-1\end{pmatrix} = \begin{pmatrix}-8\\-2\\4\end{pmatrix}$$

meaning $(\mathbb{R}^3, \times)$ is *not* associative, meaning it's not a group. The cross product is a different binary operation called the *lie bracket*[d].

11. If you know some complex analysis, then $G = \{z \in \mathbb{C} | z^n = 1 \text{ for some } n \in \mathbb{Z}^+\}$ is a group! Visually, this group looks like points on the unit circle in the complex plane. The main idea here is since $z^n = 1$, then elements in G will be of the form $e^{\frac{2\pi i}{n} \cdot k}$ for $1 \leq k \leq n$. If you're comfortable with multiplying complex numbers and trust me that all the elements of G are of this form, then you can check that this is a group.

    More generally, $G = \{z \in \mathbb{C} \mid \|z\| = 1\}$ is a group with multiplication (i.e. the unit circle in the complex plane). You can verify that $1 \in G$ and for any $z, w \in G$, $\|zw\| = \|z\| \, \|w\| = 1 \cdot 1 = 1$. Using these two fact, showing there exists an inverse and that G is associative follows from properties of $\mathbb{C}$ (exercise ref:HERE)

12. If you know some complex arithmetic's, show that $G = \{z \in \mathbb{C} \mid 0.1 < |z| < 10\}$ with the usual multiplication structure of $\mathbb{C}$ satisfies associativity, having an identity, and having an inverse, but is *not* a group.

13. Let A be any set. Consider the collection of every bijective function from A to itself:

    $$G = \{f : A \to A \mid f \text{ is bijective}\}$$

    Then with the binary operation of composition, $(G, \circ)$ is a group. What is the identity, and why do inverses always exist? The collection of bijective functions is a very common group that shows up in many different ways in different areas of mathematics and so has lots of notations:

    $$\mathrm{Bij}(A) \qquad \mathrm{Sym}(A) \qquad \mathrm{Aut}_{\mathbf{Set}}(A)$$

Part of the motivation behind learning group theory is in the abundance of areas groups show up in mathematics. The more domains of mathematics you learn, the more types of groups you will encounter. For example, field theory (which we explore in part IV) have *Galois groups* at its heart, Differential Geometry have *lie groups* that allow for a large classification of and construction of "manifolds" (a generalization of the notion of Euclidean space), Algebraic Geometry (which we explore in part V) use *algebraic groups* to represent more complicated objects in more familiar terms, Analysis uses *topological groups* (for example, Amenable and Locally Compact groups) in the study of measure theory, Fourier Analysis, and Functional Analysis, Algebraic topology has the *fundamental group* and *homology group* at its heart, and combinatorics uses *geometric group theory* (for example, the *Kleinian group* or *hyperbolic groups* shows up frequently).

Group can also be thought of as one of the "basic" structures that many other structures will build on. Of the algebraic structures that will be explored in this book, rings, fields, vector-spaces, and (lie and associative) algebras, are all also groups.

Overall, it is worth remembering that groups are a very common mathematical object that show up a little bit everywhere.

[a] $\mathbb{N}$ is a weaker structure called a monoid which doesn't require inverses. More on them in section 7.1

[b] I'm careful in saying they are different instead of saying they have *different group structures*. There is a subtly here on how we define two group structures being the same or different. This is addressed in section 1.4

[c] the answer is no! Like $\mathbb{N}$, these are example of monoids

[d] Explored in section 25.1

In general, when we write a group, the group operation will be clear from context, so instead of writing $(G, *)$, we will write G. We will take on this convention form here on out unless it's ambiguous. We will also assume the standard binary operation of addition for

$$\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/n\mathbb{Z}$$

unless stated otherwise (ex, if you see $\mathbb{Z}$, assume it is $(\mathbb{Z}, +)$). If we want the multiplicative structure of these numbers, we'll add a $\times$ in the superscript and implicitly omit the zero:

$$\mathbb{Z}^\times, \mathbb{Q}^\times, \mathbb{R}^\times, \mathbb{C}^\times (\mathbb{Z}/n\mathbb{Z})^\times$$

These sets are $(\mathbb{Z} - \{0\}, \times)$ [1], $(\mathbb{Q} - \{0\}, \times)$, $(\mathbb{R} - \{0\}, \times)$, and so forth. The sets $\mathbb{R}^+$ and $\mathbb{Q}^+$ shall represent the collection of positive real numbers and positive rational number with the usual multiplication. Furthermore, since we are always working with binary operations, the word "binary" is often omitted for convenience.

> **Note**: In many of the previous examples, there is some more structure usually associated with those symbols. For example, in $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$, there is a notion of some numbers being bigger than others, that is, $1 < 2$, or $n < n + 1$. These are **not** part of the definition of a group, and so cannot be used when proving general statements about groups! It will actually be very common that you work with groups with *more* structure than their group structure, so be quite careful that the examples you hold when doing exercises *only* represent the structure you want to work with at that moment

You might have noticed that some of the underlying sets of the above groups are subsets of each other: $\mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$. Does this imply that there is a relation between their group structure? For example does the usual addition in $\mathbb{Z}$ works as in $\mathbb{Q}$ with usual addition? If you think about this particular example, you'll see that the answer is yes. But being a subset doesn't always means you'll have the same (if any) group structure: $\mathbb{N} \subseteq \mathbb{Z}$, but $\mathbb{N}$ is not even a group! On the other hand, if we take $2\mathbb{Z} \subseteq \mathbb{Z}$, and give $2\mathbb{Z}$ the same operation as $\mathbb{Z}$ (that is $(2\mathbb{Z}, +)$ is our group), then $(2\mathbb{Z}, +)$ acts exactly the same way as $\mathbb{Z}$, with just fewer elements. So, we'll give a name to subsets which inherit group properties of their parent group:

Definition 1.1.3: Subgroup

Let G be a group and $H \subseteq G$. Then H is a subgroup of G if it is still a group under the same operation as G. We denote the subgroup by $H \leqslant G$ [a]

[a] In some textbooks, the symbol for subgroup is $H \preceq G$

[1] Note that $\mathbb{Z}^\times$ is not a group under $\times$ as we've pointed out earlier. However, the binary operations $\times$ is still defined on this set

You might have flinched to it saying "the same operation", since it was explicitly mentioned in the definition that for a group G, the operation is $G \times G \to G$, so H cannot have the same operation since the domain and co-domain are different. This is a common convention in algebra for structures which are similar: if we restrict the operation to only the elements of H, then $(H, *|_H)$ should be a group. In particular, if $* : G \times G \to G$ is the binary operation on G, then we say we *restrict* the binary operation to H if we can define $*|_H : H \times H \to H$, where for any two elements $h_1, h_2 \in H$ $*|_H (h_1, h_2) = *(h_1, h_2)$. If $(H, *|_H)$ is a group, we say it's a subgroup of $(G, *)$. In this way, we say that H has "the same" binary operation as G.

Example 1.1.3: Subgroups

1. Take $2\mathbb{Z} \subseteq \mathbb{Z}$ – the set of all even integers. The identity is there: $0 \in 2\mathbb{Z}$. If you add two even integers, you get an even integer $(2x + 2y = 2(x + y))$ [a]. Furthermore, if $2x \in 2\mathbb{Z}$, then there exists $-2x \in 2\mathbb{Z}$ such that $2x - 2x = 0$, meaning every element has an inverse. Finally, associativity is implied by associativity in $\mathbb{Z}$: for any $2x, 2y, 2z \in 2\mathbb{Z}$, we have that $2x, 2y, 2z \in \mathbb{Z}$, so $2x + (2y + 2z) = (2x + 2y) + 2z$. Thus $2\mathbb{Z} \leqslant \mathbb{Z}$
2. Right before the previous definition, it was mentioned that $\mathbb{Z} \subseteq \mathbb{Q}$ is an example of a subgroup (i.e. $\mathbb{Z} \leqslant \mathbb{Q}$). This is an exercise worth doing to get comfortable with the concept. Mix an match with other types of numbers to get more practice.
3. Let $\mathbb{Z}^\times \subseteq \mathbb{Q}^\times$ (i.e. $Z - \{0\} \subseteq \mathbb{Q} - \{0\}$). Is $\mathbb{Z}^\times$ a subgroup?
4. Let $H \leqslant G$ and $K \leqslant H$. Is $K \leqslant G$?
5. Here is an exercise in understanding the nuance of binary operations: Is $\mathbb{Z}/5\mathbb{Z} \leqslant \mathbb{Z}$? answer: [b]

[a] Note here how I *did* use extra properties that are not necessarily part of a group (I factored the two out). This is fine, since we're working with a *particular* example, this would not be a proof we can do on general groups

[b] Maybe I can write the in backward text. It is no, since the binary operation does not "act" the same way $(3 + 3 = 1)$ in $\mathbb{Z}/5\mathbb{Z}$ but $3 + 3 = 6$ in $\mathbb{Z}$

Groups are missing one key property that numbers usually have: commutativity. Though we already talked about them in the definition of a group, I will add it here to emphasize their importance:

Definition 1.1.4: Abelian Group

Given a group G, if $ab = ba$ for all $a, b \in G$, we call G a ***commutative*** or ***abelian*** group

Example 1.1.4: Abelian Groups

$\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/n\mathbb{Z}$ are all Abelian Groups [a]. So are $\mathbb{Q}^\times, \mathbb{R}^\times, \mathbb{C}^\times$. In fact, all examples in the example 1.1.2 were abelian groups. We haven't yet introduced groups which are not abelian. The first such group will be the Dihedral group (definition 1.2.16)

[a] This is an example where the binary operation is implied. The binary operations shown in example 1.1.2 are the standard binary operations for these sets

As mentioned earlier, being abelian/commutative is insanely powerful. Of all the properties you'll cover in basic group theory, a lot of them will be properties a group being "almost" or "kind of"

abelian. In this text, we will come across many definitions and sections that are centered around how close they are to being abelian [2]

The reason we relate a lot of conditions to commutativity is because Abelian Groups behave in much more predictable ways as compared to general groups: I'll point out where commutativity simplifies things everywhere it's applicable. The power of commutativity actually permeates a lot of algebraic structures. Because of this, we often split up domains in mathematics in "commutative" and "not-commutative". As an example of this split, some mathematicians like to reference the set of Abelian groups as **Ab** and the set of all groups as **Group** to show how different a general group versus a general abelian group is [3]. There are entire areas of mathematics where the problem has been solved in the "abelian case", while the "nonabelian case" is an active area of research[4].

If you're visual, you might want to view how every element acts on every other element:

Definition 1.1.5: Cayley Table (Multiplication Table)

Let G be a finite group (that is, a group on a finite set). Label the elements of G as $\{g_1, g_2, ..., g_n\}$, where g_1 represents the identity.
Then *Cayley table* or *multiplication table* of G is the $n \times n$ matrix whose i, j entry is the group element $g_i g_j$.

I personally prefer the term Cayley table since not all binary operations are called "multiplication", and so it can be confusing to do a multiplication table for $(G, +)$

Example 1.1.5: Cayley Table
Take $G = \mathbb{Z}/4\mathbb{Z}$. Then the Cayley tables for addition and multiplication are:

+	0	1	2	3
0	0	1	2	3
1	1	2	3	4
2	2	3	4	0
3	3	4	0	1

×	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

This can be done for any finite group, and sometimes gives valuable insights instantly. For example, notice that in the Cayley table for multiplication, for row 0, 1, and 3, they contain every element of $\mathbb{Z}/4\mathbb{Z}$, yet the row for 2 doesn't. Notice further that both tables are symmetrical, that is, position ij is equal to position ji.

In general, it can be quite cumbersome to write out an entire Cayley table, especially that the order of most groups we work with will be higher than 4, if not infinite. So throughout the document, we develop algebraic tools to avoid writing out the table.

[2] for reference, see definition 3.1.8, definition 2.2.3, and section 5.3, 6.1, and 6.1

[3] There are some technicalities here in saying it's the *set* of all groups, due to Russel's paradox. We should instead say the *class* of all sets. This is addressed in Chapter 30

[4] There are a huge number of such examples, but to point out one relevant to algebra class field theory completely classifies abelian extension of fields, while the nonabelian counterpart is a very deeply researched problem and is called *Langland's Program*. Another example that creates a split between the commutative and non-commutative case is within the study of algebraic geometry where the structures are built out of commutative rings, while the non commutative case would require more machinery from analysis and links to Fourier Transforms

Exercise 1.1.1

1. Let $G = \mathbb{Q}$, and the binary operation be average : $\mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$, where average$(x, y) = \frac{x+y}{2}$. Is $(G, \text{average})$ a group? What if $G = \mathbb{Z}/4\mathbb{Z}$?
2. Can a group structure be put on $\emptyset$ (i.e., can there exist a function $* : \emptyset \times \emptyset \to \emptyset$ such that $(\emptyset, *)$ is a group?) (In prelim chapter, make an exercise saying that $f : \emptyset \to \emptyset$ is a function, while $f : A \to \emptyset$, $|A| > 0$ is *not*, and also that $\emptyset \times \emptyset$ is vacuously well-defined)
3. (If you know some complex analysis) In example 1.1.2[10], it was stated that
$$G = \{z \in \mathbb{C} \mid |z| = 1\}$$
is a group. Prove that G is indeed a group.
4. Let $\text{Bij}(X)$ be the set of all bijections from X to X. Prove that $\text{Bij}(X)$ is a group. In section 1.2.2, we will relabel this as $\text{Sym}(X)$.
5. look at this link to get some ideas
6. (Perhaps this a superficial question, but the identity and inverse condition's can be replaced by functions, and so G would be a set with 1 binary operation, 1 "1-ary" operation, and a 0-ary operation. Introduce this to make groupoids and universal algebras more intuitive way later?
7. perhaps a question on how the identity can be replaced by a monary operation? (i.e. a perspective that aims to be a prelude to Universal Algebras?)

1.1.1 Consequences of Group Axioms

With the definition of groups down, we can deduce a lot of direct consequences form the axioms. One thing common with mathematical constructs is an abundance of propositions. These are vital to be able to prove, and once comfortable should be treated as knowledge as important as the definition of a group – I would even go so far as to say that not knowing the properties of groups means you don't really know what a group is. For example, do you really know what a frying pan is/can do if you haven't cooked with it before? Do you know the in's an out of what utensils can you use to not damage it, or what type of food will stick to it or not? You will only know once you've practiced using it.

I would strongly suggest that you attempt proving *all* of these propositions before looking at the proof, especially if this is the first time you've seen algebraic structures! It is crucial moving forward that these properties are second nature to you, since everything we build form here on out will constantly be relying on these properties.

The first thing to point out is that if we have $a * e = a$, we can deduce $e * a = e$.

Proposition 1.1.6: one-sided identity implies two-sided identity

Let G be a group and e an identity. If $a * e = a$ then $e * a = a$ (or vice-versa)

Proof **:**
Let's say $e * a = a$. We want to show that $a * e = a$ (1). If we multiply equation (1) on the right by a^{-1}, we get

$$\begin{aligned}
&= (a * e) * a^{-1} && \\
&= a * (e * a^{-1}) && \text{associativity} \\
&= a * (a^{-1}) && \text{Identity definition} \\
&= a * a^{-1} && \\
&= e && \text{inverse definition}
\end{aligned}$$

Thus, $(a * e) * a^{-1} = e$. Multiplying both sides by a^{-1} on the right, we get

$$\begin{aligned}
(a * e) * a^{-1} * a &= e * a \\
\Leftrightarrow \quad a * e &= e * a
\end{aligned}$$

as we sought to show.

Some author's put $a * e = e * a = a$ as a definition. This is a matter of preference: since it's a direct consequence of the axioms, you can put it as one of the fundamental rules a group has to follow without worry you're constraining the definition. Some authors disagree with this, and would rather keep the definition of group in fuller generality. This preference for generality is a balancing act: the more general you go, the more "complicated" the concept gets. However, if done right, you in general gain a more "nuanced" perspective with generality. In this book, the more general idea was taken, since some early proofs fail for more subtle reasons. For example, $(\mathbb{Z}, -)$ fails to be a group because of associativity, however if we put $a * e = e * a = a$ in the definition, we can also say it fails because $-a = 0 - a \neq a - 0 = a$ (for $a \neq 0$). However, as we've seen in the above proof, this fact *relies on associativity*, and therefore is hiding the fact that $(\mathbb{Z}, -)$ is still failing to be a group because of associativity.

Corollary 1.1.7: one-sided Inverse Implies two-sided Inverse

Let a, a^{-1} be elements of the group G and let e be the identity of G. Then $a * a^{-1} = e$ if and only if $a^{-1} * a = e$

Proof **:**
Let's say $b = a^{-1}$ so that $a * b = e$. Then $b = b * e = b * a * b$. Multiplying both sies by b^{-1}, we get

$$\begin{aligned}
e &= b * b^{-1} \\
&= (b * a * b) * b^{-1} \\
&= b * a * (b * b^{-1}) \\
&= b * a \\
&= a^{-1} * a && b = a^{-1}
\end{aligned}$$

as we sought to show.

Proposition 1.1.8: Basic properties of Group elements

1. The identity of G is unique. Furthermore, if for some $x \in G$, $f * x = x$, then is equal to the identity
2. For each $a, \in G$, a^{-1} is uniquely determined
3. $(a^{-1})^{-1} = a$, for all $a \in G$
4. $(a * b)^{-1} = (b^{-1}) * (a^{-1})$.
5. generalized associative law (induction on associativity).

Proof **:**
Again, I highly encourage you to attempt all of these before looking at these answers!!

1. Let $e \in G$ be the identity, that is, for all $x \in g$ $e * x = x$. Let's say there is another element $f \in G$ that is also the identity. Then

$$\begin{aligned} e &= e * f && \text{definition of identity} \\ &= f && \text{definition of identity} \end{aligned}$$

   therefore, $e = f$, showing that e is unique.

   Furthermore, if there was an f such that for only *some* $x \in G$, $f * x = x$, then

$$\begin{aligned} (f * x) * x^{-1} &= (x * x^{-1}) = e \\ \Leftrightarrow f * (x * x^{-1}) &= e \\ \Leftrightarrow f &= e \end{aligned}$$

   thus, there cannot be a "partial" identity.

2. etc.

With these in mind, here's some standard notation to simplify writing: One usually write the operation $*$ as $\cdot$ to represent an abstract binary operation, unless otherwise specified; we write $a \cdot b$ as ab. We usually also denote the identity e as 1. We'll also implement the following short hand: [5]

$$xxxxxxx \cdots x = x^n, \qquad x^{-1}x^{-1} \cdots x^{-1} = x^{-n}, \qquad x^0 = 1$$

Don't forget that last one is the identity of G. If we're using a group where specific notation is already defined (like with the $+$ operation), we can use it's specific convention. So $a + a + \cdots + a = na$, $-a + -a + \cdots + -a = -na$, and $0a = 0$ [6]. Furthermore, $a + (-a)$ is sometimes shorthanded to $a - a$.

The next result we'll cover let's us generalize our intuition of pre-abstract algebra. When learning

[5] $x^0 = 1$ since $x^0 = x^{1-1} = x^1 x^{-1} = xx^{-1} = 1$

[6] If you have covered rings, this might look familiar to the first property in proposition 9.1.5. It is quite similar in that this shorthand secretly is a ring-action on the group (but not a $\mathbb{Z}$ module since G is not necessarily abelian). This means that the intuition behind the notation is similar to the proof in 9.1.5: $0a = (1-1)a = a - a = 0$

algebra for the first time, you've probably encountered equations of the form:

$$3x = 5$$

and where asked to solve for x. The way you would do this is by dividing both sides by 3 to get the equation:

$$x = \frac{5}{3}$$

This might seem like the obvious thing to do in any situation. However, you can't always cancel out a number on one side and get a unique result! Take a moment to review your proof of uniqueness of identity and uniqueness of the inverse, and consider this thought experiment:

3: Thought Experiment

Let's take a set for which every element does *not* have an inverse. Recall (or prove) that $\mathbb{Z}/6\mathbb{Z}$ is not a group under $\cdot$. In particular, $[2]$ has no multiplicative inverse. Since $[2][2] = [0]$, then:

$$[4] = [2][2] = [2][5] = [10] = [4]$$

However, $[2] \neq [5]$. In other words, I couldn't "cancel" out the $[2]$ on the left side, i.e. for the equation $[2]x = [4]$ we get $x = [2]$ or $x = [5]$.

This cancelling out of a number on the left [or right] side is something you probably took for granted, but it's actually the consequence of the fact that every number in a group must have a unique inverse! This fact is proven in the next proposition:

Proposition 1.1.9: Cancellation Property: Unique solution to monomials

Let G be a group and let $a, b \in G$. The equation $ax = b$ and $ya = b$ have unique solutions for $x, y \in G$. In particular, the left and right cancellation properties hold in G:

1. If $au = av$ then $u = v$
2. if $ub = vb$ and $u = v$

Sometimes, this proposition is called the *Cancellation Law*

***Proof* :**
Simply use uniqueness of inverse. The point of this theorem is that if $ab = e$, for any element $a \in G$ and some element $b \in G$, then $b = a^{-1}$.

This proposition is very powerful, it means we can bring in our elementary algebra knowledge. In this sense, groups are the structure on which we generalize the idea of doing elementary algebra on. They are limited in that they only allow one operation: you probably have worked with equation with both addition and multiplication (ex. $3x^2 - 5 = 0$). We will work with such equations with 2 binary operations once we study *rings* (see part II)

With these properties, I'll take a moment to explain the power of associativity and commutativity. In all type of numbers systems we saw $(\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R})$, for any two numbers a and b:

$$aaabbbbaaabaaaabbbbbaa$$

is unambiguous. and so, we can write

$$a^3b^3a^3ba^4b^5a^2$$

Which is really powerful. If you didn't have associativity, then $aaabbbbaaabaaaabbbbbaa$ would be ambiguous, since it's possible that

$$(aaabbbbaaabaaaabbbbba)a \neq a(aabbbbaaabaaaabbbbbaa)$$

So every pair of elements would have to be bracketed since

$$abc$$

would be ambiguous [7] since it's possible that

$$(ab)c \neq a(bc)$$

and so you would be forced to have an expression like

$$((((aa)b)((bb)(ba)))(((((aa)(ba))a))((a(a(bb)))((b(bb))a))))a$$

with no simpler way of representing the expression – quite a monstrosity. There is more than $2^{22} = 4'194'304$ ways of bracketing this!! Since functions are associative, most structures in mathematics will be associative, however there are entire domain of study which concentrates on a non-associative structure [8] or drop the condition of associativity[9]. We will briefly see an alternative to associativity for an algebraic structure we will cover in part VI, chapter 25, but structures without associativity are outside the scope of this book[10]. As a consequence of associativity, we can unambiguously write:

$$x^{a+b} = x^{b+a}$$

since the order of multiplication doesn't effect the final result. We can also visualize associativity through the use of directed graphs. If we take abc, then we can visualize which tuple of elements we choose to multiply first:

$$\begin{tikzcd}
 & ((ab)c) = (a(bc)) & \\
(ab)c \arrow[ur] & & a(bc) \arrow[ul] \\
 & abc \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

this is more clear if we have 4 elements, any way we choose to bracket our elements leads to the same

[7] like $\mathbb{R}^3$ with the cross product

[8] the most famous of is called Lie Algebra, with something called the Jacobi Identity replacing associativity. Another is the logical implications: $(a \to b) \to c$ does not always give the same truth value as $a \to (b \to c)$

[9] Three such examples are *loops*, *Quasigroups*, and *magmas*

[10] See ref:HERE for books covering such topics

result:

$$\begin{tikzcd}
 & & (abcd) & & \\
((ab)c)d \arrow[urr] & (a(bc))d \arrow[ur] & (ab)(cd) \arrow[u] & a((bc)d) \arrow[ul] & a(b(cd)) \arrow[ull] \\
 & (ab)cd \arrow[ul] \arrow[ur] & a(bc)d \arrow[ul] \arrow[ur] & ab(cd) \arrow[ull] \arrow[ur] & \\
 & & abcd \arrow[ul] \arrow[u] \arrow[ur] & &
\end{tikzcd}$$

Note that adding commutativity makes this even more powerful, since

$$aba = aab = a^2b$$

and so we can group all like terms together, and write

$$aaabbbbaaabaaaabbbbbaa \qquad \text{as} \qquad aaaaaaaaaaaabbbbbbbbb$$

and collapse it to:

$$a^{12}b^{10}$$

That is, the only information we need is the number of times we multiply by a and b, with no regard to the order we multiply them! This added axiom is so powerful that it will turn out that it will constrain the number of possible groups which are commutative to relatively few: we will in fact list all possible [finitely generated] groups[11] which are commutative in chapter 5.1.1

:exercises

(Assuming some basic analysis knowledge) Because of general associativity we have that for any finite number of terms, we can bracket them however we want and will get the same result. Some curious among you might wonder if we can expand this to be able to take an infinite number of terms and have a similar result. However, this immediately becomes tricky. For example, if we take $\mathbb{Z}$, then what is $1+1+1+1+\cdots$? It should be verified that no number in $\mathbb{Z}$ will satisfy this conditions (you can always pick a larger number). Even worse, what if we take $1+1-1+1-1+\cdots$? Due to the infinite nature of this series, it does not have a resolution. We can try to limit ourselves to all infinite series that *converge*. However, there is even a limitation in this case. (move the terms around using commutativity of a non-absolutely convergent sequence). Because of this, the study of figuring out how to "add" infinitely many elements is relegated to another field of mathematics called analysis.

As an exercise to show the power of commutativity, show that if G is abelian, then :$(ab)^{-1} = a^{-1}b^{-1}$. (Once homomorphisms are introduced, show $x \mapsto x^{-1}$ and $x \to x^a$ are also homomorphisms

1.2 Examples of Groups

Now that we covered the basics, it is important to build up a family of common groups to get familiar with the how to recognize groups, and to feel very comfortable manipulating groups. We already

[11]The notion of finitely generated will be covered in section 2.3. I will address what this means once we reach theorem 5.1.5. As of writing this book, the project of classifying all abelian groups is still a very unsolved problem

showed how many previous mathematical structures were already groups (the canonical examples being $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$). We will now define new groups to show further what is possible with the definition of the group, and to build-up a repertoire of more elaborate examples.

1.2.1 Integers Modulo n and Cyclic Groups

Recall in section 0.2.2 that $\mathbb{Z}/n\mathbb{Z}$ is a collection of equivalences classes satisfying $a \sim_n b$ if and only if there exists a $k \in \mathbb{Z}$ st $a - b = kn$. Given that $\mathbb{Z}$ is a group and the results proposition 0.2.10, it is reasonable to think that it may form a group:

Proposition 1.2.1: Modular Arithmetic's Groups

Let $\mathbb{Z}/n\mathbb{Z}$ be the integesr modulo n. Then $\mathbb{Z}/n\mathbb{Z}$ forms a group under modular addition:

$$(a \bmod n) +_n (b \bmod n) = (a + b) \bmod n$$

with the binary operation denoted as $+_n$, or if the context permits simply as $+$

In almost all cases, the notion $+$ shall be used, for it will be clear from context that the elements will be part of a certain set of modular integers.

Proof **:**

(Well-defined) By proposition 0.2.10(1), the operation is well-defined

(Idenitty) For any n, there is always $\overline{0}_n \in \mathbb{Z}/n\mathbb{Z}$ for $n - n = 0$. Then for any $\overline{a}_n \in \mathbb{Z}/n\mathbb{Z}$:

$$\overline{0}_n + \overline{a})n = \overline{0 + a}_n$$

Then notice that

$$a - (0 + a) = 0 = kn \qquad (\text{with } k = 0)$$

Thus, $\overline{0 + a}_n = \overline{a}_n$, and so $\overline{0}_n$ is the identity for any $n \in \mathbb{N}_{>0}$

(Inverse) For any $\overline{a}_n \in \mathbb{Z}/n\mathbb{Z}$, pick $\overline{-a}_n$ (i.e., the equivalence class that can be represented by $-a$). Then:

$$(a + -a) - 0 = 0$$

implying that $\overline{a}_n = \overline{-a}_n = \overline{a + (-a)}_n = \overline{0}_n$, showing that $\overline{a}_n$ is indeed the inverse. Usually, a representative between $0 \leq t \leq n - 1$ so that $\overline{-a}_n = \overline{t}_n$.

(Associaivity) let $\overline{a}_n, \overline{b}_n, \overline{b}_n \in \mathbb{Z}/n\mathbb{Z}$. We want to show that

$$(\overline{a}_n + \overline{b}_n) + \overline{c}_n = \overline{a}_n + (\overline{b}_n + \overline{c}_n)$$

expanding the definition of the binary operation, we get:

$$\overline{(a + b) + c}_n = \overline{a + (b + c)}_n$$

In particular, we want to verify that:

$$[(a + b) + c] - [a + (b + c)] = kn$$

which by the associativity of the integers, we get that their difference is 0, which implies both sides are equal, completing the proof.

Definition 1.2.2: Integers Modulo n

The set $\mathbb{Z}/n\mathbb{Z}$ forms a group under modulo addition called *integers modulo n*.

Hence, for each $n \in \mathbb{N}_{>0}$, there is a group with with n elements. It has yet to be explored whether the modulo multiplication is a viable binary operation, namely if:

$$(a \bmod n) \cdot (b \bmod n) = (ab) \bmod n \tag{1.1}$$

can form a binary operation. It turns out that some more care must be put in:

Example 1.2.1: Not Always Group

Take $n = 4$. The reader should verify that $\overline{1}_4$ is the identity given the binary operation defined in equation (1.1), and by proposition 1.1.8(1) it is unique. However, notice that $\overline{2}_4$ has *no* multiplicative inverse:

$$\overline{0 \cdot 2}_4 = \overline{0}_4 \qquad \overline{1 \cdot 2}_4 = \overline{2}_4 \qquad \overline{2 \cdot 2}_4 = \overline{0}_4 \qquad \overline{3 \cdot 2}_4 = \overline{2}_4$$

Hence, $(\mathbb{Z}/n\mathbb{Z}, \cdot_4)$ is *not* a group. However, if we pick $n = 3$, notice that this forms a group!

The reader may have observed that 3 is a prime and may wonder if is the key difference. This is in fact close: the important part is to choose all elements that are *coprime* with n:

Definition 1.2.3: Euler Totient

Define $\varphi : \mathbb{N} \to \mathbb{N}$ to be

$$\varphi(n) := |\{k \in \mathbb{N} \mid \gcd(n, k) = 1\}| = \#\{k \in \mathbb{N} \mid \gcd(n, k) = 1\}$$

that is, $\varphi(n)$ counts the number of coprime

The notation φ is often re-used and so shouldn't be considered a "universal" symbol for the Euler Totient.

Definition 1.2.4: Integer Modulo n With Multiplication

Let $\mathbb{Z}/n\mathbb{Z}$ be the integers modulo n and choose representatives for each element so that the n equivalence classes may be represented with $0, ..., n-1$. Take $(\mathbb{Z}/n\mathbb{Z})^\times \subseteq \mathbb{Z}/n\mathbb{Z}$ defined to be

$$(\mathbb{Z}/n\mathbb{Z})^\times = \left\{\overline{t}_n \in \mathbb{Z}/n\mathbb{Z} \;\middle|\; \varphi(t) = 1\right\}$$

Define the binary operation to be:

$$(a \bmod n) \cdot (b \bmod n) = (ab) \bmod n \tag{1.2}$$

Then $(\mathbb{Z}/n\mathbb{Z})^\times$ is a group modulo n with multiplication.

The reader should verify as an exercise that this is indeed a group.

Cyclic Subgroups

Each group G will have something similar to the integers modulo n, which is worth exploring:

Proposition 1.2.5: Creating a Cyclic Subgroup

Let G be a group and take any $x \in G$. Let $\langle x \rangle = \left\{x^a \middle| a \in \mathbb{Z}\right\}$. Then $\langle x \rangle$ is a subgroup of G

Proof **:**
I would recommend you do this as an exercise!

Let $\langle x \rangle = \{x^a | a \in \mathbb{Z}\}$. We'll run through the group-axioms:

1. (associativity) Associativity come's from associativity in G. Let $r, s, t \in \langle x \rangle$, and consider

$$(r + s) + t$$

then since $r, s, t \in G$, and G is associative, we get that

$$(r + s) + t = r + (s + t)$$

and since $r, s, t \in \langle x \rangle$ were arbitrary 3 elements, this holds for any elements of $\langle x \rangle$, completing the proof

2. (identity) We need to show that in $\langle x \rangle$, there exists an element e such that $ex^a = x^a$. Since $0 \in \mathbb{Z}$, $x^0 = e \in \langle x \rangle$. And since for any element $y \in x$, $ey = y$, and every x^a is an element of G, $ex^a = x^a e = x^a$

3. (inverse). Let $y \in \langle x \rangle$ be an element of $\langle x \rangle$. Then by definition, for some $a \in \mathbb{Z}$, $y = x^a$. Then since $-a \in \mathbb{Z}$, $x^{-a} \in \langle x \rangle$, and as we've seen earlier:

$$x^a x^{-a} = x^{a-a} = x^0 = e$$

showing that x^a has an inverse for every a.

Thus, $\langle x \rangle$ is a group. Notice that the binary operation is identical to that of G restricted to $\langle x \rangle$, making $\langle x \rangle$ a subgroup of G, completing the proof.

Given any element $x \in G$, we can form a subgroup $\langle x \rangle \leqslant G$. The reason for it being called a cyclic subgroup comes when we consider a finite group G [12]

Example 1.2.2: Why is $\langle x \rangle$ called Cyclic?
Let G be a finite group (i.e. $|G| = n$ for some $n \in \mathbb{N}$), take $x \in G$ and consider $\langle x \rangle \leqslant G$.

1. Show that $\langle x \rangle$ is finite, that is, there is some $k \in \mathbb{N}$ such that $|\langle x \rangle| = k$

2. Show that for any x^{-a}, where $a \in \mathbb{N}$, there exists some number $b \in \mathbb{N}$ such that

$$x^{-a} = x^b$$

[12] More generally, subgroup $\langle x \rangle \leqslant G$ where $|\langle x \rangle|$ is finite

Conclude that $\langle x \rangle$ can be generated by raising number *only* to the power of positive number, i.e.

$$\{x^a \mid a \in \mathbb{Z}\} = \{x^a \mid a \in \mathbb{N}\}$$

3. show that if $|\langle x \rangle| = k$, then $x^k = e$. Use this fact to show that for any $x^a \in \langle x \rangle$, $a \leq k$.

With these facts, we can visualize a finite cyclic subgroup $\langle x \rangle$ using a graph. Let's say 1 is the identity. If we multiply 1 by x $(1 \cdot x)$, we get x. We can represent this visually with

$$1 \xrightarrow{\cdot x} x$$

if we multiply x by x, we get

$$1 \xrightarrow{\cdot x} x \xrightarrow{\cdot x} x^2$$

continuing in this fashion, we will make long "chain":

$$1 \xrightarrow{\cdot x} x \xrightarrow{\cdot x} x^2 \xrightarrow{\cdot x} x^3 \xrightarrow{\cdot x} \cdots$$

until we reach 1, at which point we can circle back to it in the graph:

$$\begin{tikzcd}
1 \arrow[r, "\cdot x"] & x \arrow[r, "\cdot x"] & x^2 \arrow[r, "\cdot x"] & x^3 \arrow[r, "\cdot x"] & \cdots \arrow[r] & x^{k-1} \arrow[llll, bend right, "\cdot x"']
\end{tikzcd}$$

which more conveniently can be written as

$$\begin{tikzcd}
 & \cdots \arrow[dl, "\cdot x"] & x^3 \arrow[l, "\cdot x"'] & \\
x^{k-1} \arrow[dr, "\cdot x"'] & & & x^2 \arrow[ul, "\cdot x"'] \\
 & 1 \arrow[r, "\cdot x"] & x \arrow[ur, "\cdot x"'] &
\end{tikzcd}$$

producing us a cycle!

Since every element can form a cyclic group, and their structure is characterized by the cardinality of $\langle x \rangle$, we will give a special name to this cardinality:

Definition 1.2.6: Order of an element

For a group G and $x \in G$, define the *order of* x to be the smallest positive integer n such that $x^n = 1$, and denote this integer by $|x|$. In this use, x is said to be of order n. If no positive power of x is the identity, the order of x is defined to be infinity and x is said to be of infinite order
Equivalently, we can also say $|x| = |\langle x \rangle|$, that is, the order of x is the order of the group generated by x

Another word for element with finite order is *torsion elements*. Elements that are not torsion elements

(i.e. with infinite order) are called *torsion-free elements*.

Example 1.2.3: Order of Elements

1. Every element of $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ under $+$ has infinite except 0. Is the same true for $\mathbb{R}^\times$ (i.e. $\mathbb{R} - \{0\}$ (answer: [a]) Note the identity always has order 1.
2. Can you find the order of the elements in $\mathbb{Z}/p\mathbb{Z}$ for prime p? What if p is not prime? What are the order of the elements of $\mathbb{Z}/60\mathbb{Z}$?

[a] no, take $-1 \in \mathbb{R}^\times$

Note: The order of the element depends on *both* the set and the binary operator. See exercise ref:HERE

Also, the word *order* is sometimes used to mean the cardinality of the group.

I highly suggest you now do a couple of exercises, since these groups will be at the foundation of finding all possible abelian group [13]

Exercise 1.2.1

1. for $n > 1$, $\mathbb{Z}/n\mathbb{Z}$ is not a group under multiplication (hint: think of identity)
2. prove that every element in a finite group has finite order
3. if $x \in G$, G a group, and $|x| = n$, then $x^{-1} = x^{n-1}$.
4. $x \in G$, G a group. Then $|x| = |x^{-1}|$. (hint: $x \cdot x \cdots x = 1$, so keep multiplying by x^{-1} till you get $1 = x^{-1} \cdot x^{-1} \cdots x^{-1}$ If x has infinite order). Though in certain infinite order groups, like $\mathbb{R} - \{0\}$ under multiplication, -1 has order 2, but 1 has order 1. Is this only true when the operation is $+$?
5. 21. is there a catch? why should n be odd?
6. $x, g \in G$, G a group. Then $|x| = |g^{-1}xg|$. Deduce that $|ab| = |ba|$ for all $a, b \in G$. (hint: simply assign an order to x, and expand $g^{-1}xg$. things will cancel out nicely. For the second part, start with ab, and multiply like so $babb^{-1}$. By what you've shown, this has the same order).
7. $x \in G$, $|x| = n < \infty$. If $n = st$ for some positive integers s and t, prove that $|x^s| = t$. (hint: how many times do you need to multiply x^s to get to 1)
8. if $x^2 = 1$ for all $x \in G$ then G is abelian.
9. if x is an element of the group G, then $H = \{x^n | n \in \mathbb{Z}\}$ is a subgroup (cf. show for all $h, k \in H$, $hk \in H$, $h^{-1} \in H$) of G (called the *cyclic subgroup* of G generated by x)
10. $A \times B$ is abelian if and only if A and B is abelian
11. Let $A \times B$ be the *direct product* group formed by the cartesian product of the group A, B as

[13] In particular, finitely generated abelian group, see section ref:HERE

shown in example 1.1.2. Show that $(a, 1)$ and $(1, b)$ commute. Deduce that the order of (a, b) is the least common multiple of $|a|$ and $|b|$.

12. any finite group G of even order contains an element of order 2. (hint: let $t(G)$ be the set $\{g \in G | g \neq g^{-1}\}$. Show that $t(G)$ has even number of elements and every non-identity element of $G - t(G)$ has order 2).

13. Other exercises you can do I'd recommend: 32, do 33, 34, 35, 36

14. Challenging exercise. Let A be a set containing numbers. Define the following

$$\mathcal{A} = \bigoplus_{i=1}^{\infty} A$$

to be the set of infinitely long tuples were *all but finitely many* elements are non zero. For example, if $A = \{1, 2, 3\}$, then

$$(0, 0, 1, 0, 2, 0, 0, 0, 0, 1, 3, 3, 2, 0, ...)$$

where the ... means there are zeros going on forever. Then this tuples is in $\mathcal{A}$. On the other hand

$$(1, 1, 1, 1, 1, ...)$$

where the ... means the ones are going on forever is not in $\mathcal{A}$.

If G is a group, let

$$\mathcal{G} = \bigoplus_{i=1}^{\infty} G$$

be the set of tuples where all but finitely many elements are non-zero, and define the operation on $\mathcal{G}$ to be pointwise:

$$(a_1, a_2, a_3, ...) \cdot (b_1, b_2, b_3, ...) = (a_1 b_1, a_2 b_2, a_3 b_3, ...)$$

Show that $\mathcal{G}$ along with this operation is a group.

15. Let G be a finite group. Show that G can be generated by elements to the power of $\mathbb{N}$ (move into the generators and relations section maybe?)

16. In some textbooks, groups are introduced more through the use of *graphs*. We will present these ideas here. (Google Frucht's Theorem)

17. Let p be a prime number. Show that $(\mathbb{Z}/p\mathbb{Z})$ is a cyclic group.

18. Let $n = 8$. What is $(\mathbb{Z}/8\mathbb{Z})^\times$? What if $n = 9$. Conclude that in general, $(\mathbb{Z}/n\mathbb{Z})^\times$ is not necessarily cyclic, but it can be.

19. (perhaps a question on the structure of $(\mathbb{Z}/n\mathbb{Z})^\times$ in general?)

1.2.2 Symmetric Groups

Let X be any set. As covered in the preliminary chapter, a bijective function $f : X \to X$ always has a (two-sided) inverse function[14], say f^{-1}. Furthermore, function composition is associative, and since

[14] This in fact characterizes bijective functions

the domain and codomain match, the composition of two bijective functions $f, g : X \to X$ give's another bijective function $g \circ f : X \to X$ (as the reader ought to check if they're not convinced). The final property to check for the collection of bijective functions on a set X to form a group is whether there exists a bijective function $\mathrm{id}_X : X \to X$ such that for all bijective functions $f : X \to X$, $\mathrm{id}_X \circ f = f \circ \mathrm{id}_X = f$. The reader should take a moment to think what this function

4: Moment

Taking a moment

The reader may have noticed that this function arises naturally by taking $f \circ f^{-1}$, namely it's the function defined to be $s \mapsto s$. Thus, given any set X, there is a natural group structure on the collection of bijections on X. This set is given a name:

Definition 1.2.7: Symmetric Group

Let X be any set. Then then:

$$S_X := \{f : X \to X \mid f \text{ is a bijection}\}$$

is called the *symmetric group* or *premutation group* of X [with respect to composition $\circ$]. If the cardinality of X is finite, say $|X| = n$, then we denote S_X by S_n.

Other common notations for the symmetric group are:

$$\mathrm{Bij}(A) \qquad \mathrm{Sym}(A) \qquad\qquad \mathrm{Aut}_{\mathbf{Set}}(A)$$

If X is finite, then the bijections S_n are determined only by the place to which elements map to and not by any property of the elements, making the bijections only dependent on the number of elements and hence the S_n notation being unambiguous. For this reason, it is often assumed that the set over which the bijections are being taken is given by[15]:

$$\{1, 2, ..., n\}$$

A mention on notation is in order: if we take $f, g \in S_X$, then the current short-hand notation for groups would suggest that the product of f and g ought to be fg and represents the short-hand $f \circ g$. However, the prevailing notation for compositions reads the functions *right-to-left*:

$$f * g = g \circ f \qquad \Rightarrow \qquad (g \circ f)(x) = g(f(x))$$

This notational quirk comes from the fact that the conventions for function composition comes much earlier than the convention for group operations. Mathematically, this shall not effect the underlying theory[16].

Symmetric groups are rather large; the number of bijections between finite sets grows:

$$|S_n| = n!$$

And hence even for $|X| = 100$, $|S_{100}| \gg 10^{100}$. For $n \in \{1, 2\}$, the bijections are rather simple, namely:

$$f_1 : \{1\} \to \{1\},\ f_1(1) = 1 \qquad g_1 : \{1, 2\} \to \{1, 2\},\ g_1 = \mathrm{id}_{\{1,2\}} \quad g_2(1) = 2\ g_2(2) = 1$$

[15] This set is also sometimes called the *indexing set* and represents labeling each element of some set of size n

[16] It shall complicate notation a little bit when defining left-group actions, which shall be addressed in due time

For larger sets, it may get difficult to track the information a bijection holds. For example, if σ is a bijection on a set with 9 elements:

$$\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

then we after the σ permutes, we might get

$$(\sigma(1), \sigma(2), \sigma(3)\sigma(4), \sigma(5), \sigma(6), \sigma(7), \sigma(8), \sigma(9)) = (2, 4, 1, 3, 7, 9, 8, 5, 6)$$

It is cumbersome to write-out the permutation by saying what each value of $\sigma \in S_n$ maps to. Instead, there is the following *matrix notation* as a shorthand:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 4 & 1 & 3 & 7 & 9 & 8 & 5 & 6 \end{pmatrix} \tag{1.3}$$

Representing f_1, g_1, g_2 in matrix notation, we would get:

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

This is better notation, however we can do even better taking advantage of the fact that a bijection is one-to-one and onto. This implies that every element is only associated with one element and each element must be represented. This allows for the following "cycle" notation. Take equation 1.3, and notice that 1 maps to 2, which maps to 4, which maps to 3, which maps to 1:

$$1 \mapsto 2 \mapsto 4 \mapsto 3 \mapsto 1$$

Furthermore, 5 maps to 7, which maps to 8, which maps to 5:

$$5 \mapsto 8 \mapsto 7 \mapsto 5$$

Finally, 6 maps to 9, and 9 maps to 6:

$$6 \mapsto 9 \mapsto 6$$

This behavior can be represented via the following *cycle notation*:

$$(1\ 2\ 4\ 3)(5\ 8\ 7)(6\ 9)$$

Each "chain" in the above is called a *cycle*, and the combination of each is called the *cycle representaiton* If a cycle contains n elements, it is called an n-cycle. In the above example, there were not any elements that map to itself. If the element 10 is added to the set and we extend σ to $\overline{\sigma}$so that $\overline{\sigma}(10) = 10$, then we would write $(1\ 2\ 4\ 3)(5\ 8\ 7)(6\ 9)(10)$. By convention, any element that maps to itself is dropped in cycle notation, and hence $\overline{\sigma}$ would in cycle notation also be represented as:

$$(1\ 2\ 4\ 3)(5\ 8\ 7)(6\ 9)$$

Hence, there is an inherent ambiguity in cycle notation on the size of the domain/codomain, since any elements that map to themselves is not taken into accost, however it is usually the element that map to a non-trivial element that are of interest and hence this ambiguity does not effect the theory. Note that:

$$(1\,2\,3) = (3\,1\,2) = (2\,3\,1)$$

Since they all actually do the same permutation, you're just starting at different positions when converting to the cycle notation.

Example 1.2.4: Cycle Representation

1. Let τ be a bijective function on a set $\{1, 2, 3, 4, 5\}$ such that:

$$(\sigma(1), \sigma(2), \sigma(3), \sigma(4), \sigma(5)) = (5, 4, 3, 2, 1)$$

   What is the cycle notation for τ

2. Cycles can be used to define permutations. If $X = \{1, 2, 3, 4\}$. Find the bijection represented by the cycle representation:

$$(1\ 2\ 3)(2\ 3\ 4)$$

   Notice that the two cycles overlap. To find τ, first find where 1 maps to in the right-cycle, and the where the resulting element maps to the left cycle; in function notation if f is the left cycle and g is the right cycle then the reader is looking for $f(g(1))$. Proceed with this for all 4 element. Remember that if an element is not in a cycle, then it is considered to map to itself.

3. Is it true that the functions represented by $(1\,2)(3\,4)$ and $(3\,4)(1\,2)$ are equal, that is:

$$(1\ 2)(3\ 4) = (3\ 4)(1\ 2)$$

   More generally, do disjoint cycles commute?

To insure the reader has the right calculation, here is an example of the product of two cycles as reference:

$$(1\ 4\ 5\ 2)(4\ 1\ 3) = (1\ 3\ 5\ 2)$$

With those extra properties of cycles, we will name one more concept for ease of labeling cycles:

Definition 1.2.8: Cycle Type

Given a cycle-representation of a permutation, you can make the cycles disjoint. Then, order them from smallest to largest. Then the cycle representation is the "length" of each cycle in that order.

Example 1.2.5: Cycle-Type

$(1\,2\,4\,3)(5\,7\,8)(6\,9)$ is equivalent to $(6\,9)(5\,7\,8)(1\,2\,4\,3)$ which we can label as a $(2, 3, 4)$-cycle type. If you have a $(1\,2)(3\,4)$ premutation, then it's a $(2, 2)$-cycle type.

A cycle of particular importance are 2-cycles.

Definition 1.2.9: Transposition

2-cycles are usually referred to as ***transpositions***

Using transpositions, any n-cycle can be decomposed into the product of transpositions. For example, take:

$$(1\,2\,4\,3)$$

Then the reader should verify that:

$$(1\,3)(1\,4)(1\,2) = (1\,2\,4\,3)$$

So, the cycle broke up into 3 transpositions. In fact, any n-cycles will break up into $n-1$ transpositions. It will break up to exactly (at most and at least) $n-1$ transpositions.

Theorem 1.2.10: n-Cycle Breakdown To Exactly $n-1$ Transpositions

Let $\sigma \in S_n$ be an n-cycle. Then σ can be broken down to the composition of exactly $n-1$ transpositions

***Proof* :**
exercise

Because of this, given any σ, which can then be written as disjoint cycles, and then into transpositions will always have a fixed number of transpositions in the product, in other words this is an invariant of a cycle σ, which deserves then a name:

Definition 1.2.11: Parity Of Cycle

Given $\sigma \in S_n$, then the parity of the cycle is the parity of the number of transpositions so if we have an n-cycle, where n is even, then it's parity is odd, since it will break down to $n-1$ transpositions. A cycle with even parity is called an *even permutation*, similarly called *odd-permutation* when the parity is odd.

Parity might seem like a weird thing to care about, but it will in fact come in very useful. So useful in fact, that there's a special name for groups based around parity:

Definition 1.2.12: Alternating Group

Let S_n be the permutation group permuting n Elements. Then $A_n \subseteq S_n$ is all cycles with *even* permutations. Note that the compositions of two cycles from A_n is even, and that the inverse of an even permutation is even, thus

$$A_n \leqslant S_n$$

Why this group is interesting requires more algebraic tools, so we'll leave at defining it for now and get back to it to show it's use.

Before moving on to properties of symmetric groups, some reader may find it interesting that at the origins of group theory, groups were defined to be subgroups of symmetric groups! Later, _____and _____have abstracted these properties and formed the modern axiomatic representation. Later in this book, it shall be shown that all groups are in fact subgroups of the appropriate symmetric group (see theorem 4.4.2).

Example 1.2.6: Interesting Exercises on Commutativity
We earlier mentioned how being commutative is a nice property, and that groups which are commutative are quite nice. It was also mentioned that S_n is in a sense the most general group that possible. So, in a sense, if we can establish properties of S_n for particular orders, that can be

quite insightful.

Here's one such property worth proving. If we take S_n, $n \geq 3$, then S_n is *not* commutative To prove this, try composing a n-cycles and a 2-cycle which are not disjoint form the left and right and see what you get

Properties of Symmetric Groups

Proposition 1.2.13: Order of element is p iff composition of disjoint p-cycles

Let p be a prime. Show that an element of S_n has order p if and only if its cycle decomposition is a product of (commuting) p-cycles.

Proof **:**
exercise

Note that if p is not prime, the "if and only if" is not necessarily true.

Example 1.2.7: Counter-Examples

1. Take
$$s = (1\,2)(3\,4\,5\,6) = s_1 s_2$$
Then $(s_1 s_2)^4 = 1$, but $|s_1| = 2$.

2. Take $(1\,2)(3\,4\,5) \in S_5$. Then
$$\begin{aligned}(1\,2)(3\,4\,5)^2 &= (3\,5\,4)\\ (1\,2)(3\,4\,5)^3 &= (1\,2)\\ (1\,2)(3\,4\,5)^4 &= (3\,4\,5)\\ (1\,2)(3\,4\,5)^5 &= (1\,2)(3\,5\,4)\\ (1\,2)(3\,4\,5)^6 &= id\end{aligned}$$
but $6 > 5$. Thus, we found a cycle of order 6 in S_5. The trick was to find two disjoint cycles such that the $\operatorname{lcm}(\sigma, \tau) > n$.

Proposition 1.2.14: σ is p-cycle for prime p then σ^k is p-cycle

If $\sigma \in S_n$ has prime order, then σ^k for every $1 \leq k \leq p$ is a p-cycle

Proof **:**
This is a simple application of proposition 1.2.13, and since we're taking $\langle \sigma \rangle$, we'll use theorem 2.3.8 (the theorem on cyclic groups). Since $|\langle \sigma \rangle| = p$, then $\langle \sigma^k \rangle = p$ for every $k < p$ ($\sigma^p = e$). This means that σ^k has prime order. Thus, by proposition 1.2.13, it is composed of p-cycles. Since $\langle \sigma \rangle = p$, and σ is a bijection (meaning we never go to another element while composing σ), then

every σ^k is a p-cycle

Note that if σ is not a p-cycle, this is not necessarily true. For example:

$$(1\,2\,3\,4)\,,(1\,2\,3\,4)^2 = (1\,3)(2\,4)\,,(1\,2\,3\,4) = (1\,4\,3\,2)$$

Notice how σ^2 broke down to two 2-cycles. The way I remember this to recall that if σ is a 4-cycle, then $|\sigma^2| = 2$, so $\sigma^2\sigma^2 = id$, which by what we've covered in the previous proposition (1.2.13), σ must be disjoint 2-cycles!! This holds true for any n-cycle, giving you an easy way of thinking about σ^k for any n-cycle with $1 \leq k \leq n$.

Proposition 1.2.15: Conjugation preserves permutation-type

If $\sigma, \tau \in S_n$, then

$$\sigma\tau\sigma^{-1} = (\sigma(1)\,\sigma(2)\,\cdots\sigma(t_k))$$

i.e., it's like applying the function σ to every element of τ

Proof **:**
This proof relies on this key observation:

$$(a\,b)(a\,c)(a\,b)^{-1} = (a\,b)(a\,c)(a\,b) = (b\,c)$$

that is, we shifted a to b (so in a sense if $\sigma = (a\,b)$ and $\tau = (a\,c)$, $\sigma\tau\sigma^{-1} = (\sigma(a), \sigma(b))$).

The rest of this proof is now a matter of keeping track of notation [a]

[a] This proof is easier prove once your introduced to group action and conjugacy classes (see definition 4.4.4)

1.2.3 Dihedral Groups

In classical geometry, an n-gon is a 2D shape with all sides of the same length. For example, a triangle is a 3-gon, a square is a 4-gon, a pentagon is a 5-gon, and so forth, while a rectangle and a circle are not n-gons. Each of these shapes can be rotated or reflected in such a way that we get back the same shape: these are called the *rigid-motion symmetries* or just *symmetries* of the n-gon. The term *rigid motion* comes from analysis/linear-algebra where a function $f : X \to X$ is called a *rigid motion* has the property that the distance between each point remains the same after the application of the function. Notationally, this would be represented as:

$$\|f(x) - f(y)\| = \|x - y\|$$

where the $\| - \| : X \to \mathbb{R}_{>0}$ is a special kind of function called the *norm* which defines size and distance information on a space[17]. We shall not use rigid motions explicitly, however we shall insure that when transforming n-gons, their they shall never stretch (ex. a square shall not become a rectangle). We shall further place the restriction that the rigid motions of the n-gons have to map to themselves, this is another way of saying that it is a *symmetry*. For example, if we choose a square (a 4-gon):

[17] Norms shall be introduced formally in section 13.5.1

and rotated it a quarter turn to the right (90 degrees to the right or $\pi/2$ radians to the right), then we'd get back a square

if we were to rotate it an eighth of a turn, we would not get the exact shape back:

Let's label quarter turns to the right r (for **r**otate). Compounding rotations, the resulting square get's rotated by π, $3\pi/2$, and 2π radians; each of these can be labeled r^2, r^3, and r^4 respectively. Another symmetry of the square would be reflections, namely there is a vertical, horizontal, and two diagonal reflections: $\sigma_h, \sigma_v, \sigma_{d1}, \sigma_{d_2}$. These are not all the symmetries of a square (for example, we transpose the left-edge with the right edge), however it is the only symmetries that are also *rigid motions* (the reader who is already comfortable with norms form linear algebra is free to prove this fact).

More formally, we may represent any n-gon with n-points, with the restrictions that any symmetry (i.e. bijection) f applied to these n-points must further follow the restrictions that the n-gon these points represent must transform as a rigid-motion (all points must remain equi-distant). This special subset of S_n with this added restriciton is called the *dihedral group*:

Definition 1.2.16: Dihedral Groups

Let S_n be the symmetric group on n elements and take the subset $D_n \subseteq S_n$ for which the premutations further follow the restriction of rigid-motions (so that these points represent an n-gon).Then D_n is called the *dihedral group*.

In order for the set D_n to be a group, one ought to prove this is indeed the csae

Proof **:**

1. Identity comes from the identity function $\mathrm{id}_n \in D_n$ since it is certainly a rigid motion.
2. The inverse come from the fact that if σ is a rigid motion, it's inverse is as well

3. The set is associative since its a collection of functions
4. The composition of two rigid motions is a rigid motion, and hence the binary operation is well-defined

Proposition 1.2.17: Properties of D_n

Let $G = D_n$. Let f represent rotation by by $2\pi/n$, and s represent a reflection about some fixed axis. Then:

1. $1, r, \ldots, r^{n-1}$ are all distinct and $r^n = 1$, so $|r| = n$
2. $|s| = 2$
3. $s \neq r^i$ for any i
4. $sr^i \neq sr^j$ for all $0 \leq i, j \leq n-1$ with $i \neq j$, so
$$D_n = \left\{1, r, r^2, \cdots, r^{n-1}, , sr, sr^2, \cdots, sr^{n-1}\right\}$$
i.e. each element can be written uniquely in the form $s^k r^i$ for some $s = 0$ or 1, and $r = 0, \cdots, n-1$
5. $rs = sr^{-1}$. In particular this shows D_n is not abelian!
6. $r^i s = sr^{-i}$ (use induction)

Proof **:**
exercise

In particular, we know now any element in D_n can be written as $s^k r^i$, with $k = 0 \vee 1$ and $0 \leq i \leq n-1$. For example, if $n = 12$:

$$(sr^9)(sr^6) = sr^9 sr^6 = sr^9 r^{-6} s = sr^3 s = s^2 r^{-3} = r^{-3} = r^9$$

This group will come up very often since it's a particularly simple.

Proposition 1.2.18: Order Of Dihedral Groupq

Let D_n be a dihedral group. Then:
$$|D_n| = 2n$$

Proof **:**
exercise.

Some authors like to put the order of the Dihedral group in the subscript, that is they write D_{2n}. This convention is rather confusing given the original intuition behind Dihedral groups, and so the D_n notation is used instead for the clarity it brings.

1.2.4 Quaternion Group

The Quaternion (or Hamiltonian) group is a group with 2 fundamental interpretation:

1. The Quaternions are a generalization of the complex numbers! We can think of the fact that $i^2 = -1$ to be the key property which distinguishes the complex numbers from other number systems. If we take $C = \{-1, 1, i, -i\} \subseteq \mathbb{C}$, then

$$\langle C | i^2 = -1 \rangle$$

will form the group which represents the rotation of $\pi/2$ in the cartesian plane. William Hamilton around the 1840s started a project to try an make 3D numbers $(1, i, j)$, the same way $\mathbb{C}$ can be thought of as 2D $(1, i)$. He couldn't manage to make them 3D, but realized the key generalization is making them "4D" $(1, i, j, k)$. He had this flash of inspiration near a bridge in 1843 and wrote down the idea under that very bridge. Today, there is a plaque commemorating the moment of inspiration.

2. It got discovered later that Quaternions efficiently capture the idea of rotating a 3D object in 3D space. At the beginning of vector calculus, some mathematician considered using quaternion instead of vectors as the pedagogical tool used to introduce the concept of 3D rotation, but history went with vector notation. With leaps in quantum mechanics in the late 20th century, quaternions were also used to represent *spin*, a fundamental property of sub-atomic particles.

Let:

$$Q = \{1, -1, i, -i, j, -j, k, -k\}$$

Such that the following relations hold:

$$i^2 = j^2 = k^2 = -1 \tag{1.4}$$
$$1 \cdot a = 1 \cdot a \qquad \forall a \in Q_8 \tag{1.5}$$
$$ij = k \qquad ji = -k \tag{1.6}$$
$$ik = j \qquad ki = -j \tag{1.7}$$
$$jk = i \qquad kj = i \tag{1.8}$$

Or, in terms of group presentation:

$$Q_8 = \langle Q | i^2 = j^2 = k^2 = ijk = -1 \rangle$$

This group is interesting, for it's the 2nd non-abelian group you've seen with order 8, D_8 being the other group. Furthermore, both of them share a similar intuitive background, being the based on rotation. However, Q_8 rotates 3D objects, while D_8 rotates 2d objects, which actually produce different group structures, a these Cayley tables show:

Example 1.2.8: Comparing D_8 and Q_8

Element	1	-1	i	$-i$	j	$-j$	k	$-k$
1	1	-1	i	$-i$	j	$-j$	k	$-k$
-1	-1	1	$-i$	i	$-j$	j	$-k$	k
i	i	$-i$	-1	1	k	$-k$	$-j$	j
$-i$	$-i$	i	1	-1	$-k$	k	j	$-j$
j	j	$-j$	$-k$	k	-1	1	i	$-i$
$-j$	$-j$	j	k	$-k$	1	-1	$-i$	i
k	k	$-k$	j	$-j$	$-i$	i	-1	1
$-k$	$-k$	k	$-j$	j	i	$-i$	1	-1

Figure 1.1: Q_8 Multiplication Table

While D_8 has

Element	e	a	a^2	a^3	x	ax	a^2x	a^3x
e	e	a	a^2	a^3	x	ax	a^2x	a^3x
a	a	a^2	a^3	e	ax	a^2x	a^3x	x
a^2	a^2	a^3	e	a	a^2x	a^3x	x	ax
a^3	a^3	e	a	a^2	a^3x	x	ax	a^2x
x	x	a^3x	a^2x	ax	e	a^3	a^2	a
ax	ax	x	a^3x	a^2x	a	e	a^3	a^2
a^2x	a^2x	ax	x	a^3x	a^2	a	e	a^3
a^3x	a^3x	a^2x	ax	x	a^3	a^2	a	e

Figure 1.2: Q_8 Multiplication Table

which the reader will notice is different

We will actually be able to show that two groups are not the same once we introduce the concept of isomorphisms (section 1.4).

Exercise 1.2.2

1. Show that there exists 4 homomorphisms $\varphi : Q_8 \to \mathbb{R}^\times$

1.2.5 *Matrix Groups

This section assumes either some linear algebra knowledge or that you have internalized some of the section 0.5 (however, some of these collection of objects goes beyond that section). All prerequisite knowledge assumed in this section shall be covered in great detail in chapter 13, section 13.2 and 13.5. Let's first recall the notation we use to describe the collection of matrices:

Definition 1.2.19: Matrix Group Under Addition

Let G be an abelian group. Then The set $M_{n\times m}(\mathbb{F})$ forms a group under element-wise operation of G, namely:

$$A * B \equiv (a_{ij} * b_{ij})_{ij}$$

If the reader is familiar with linear algebra, they may notice this group is not actually that interesting: it is equivalent to the group G^{nm}[18]. Sticking to $G = k$ is a field (notice that all fields are Abelian groups under addition), the more interesting binary operation that can be performed on matrices is matrix multiplication, since it combines elements in much more interesting ways. As you might recall, matrix multiplication requires dimensions to line up properly, so for associative to hold, we'll have the dimensions be the same. Furthermore, the identity must be I, the matrix with 1's on the diagonal, and zero's everywhere else. If there are matrices that multiply and *don't* get to I, then they cannot be in the group. Hence, the subset of invertible matrices from $M_n(k)$ must be selected in order to make a group, which the reader may recall from linear algebra implies we are taking the set for which all matrices have a non-zero determinant

Definition 1.2.20: General Linear Group

$$GL_n(F) = (\{A|A \text{ is an } n \times n \text{ matrix with entries from } F \text{ and } \det(A) \neq 0)\}$$

We will also name 3 special subgroups of GL_n since it comes up often enough. Check that these are indeed groups:

Definition 1.2.21: Special Linear Group

$$SL_n(k) = (\{A|A \text{ is an } n \times n \text{ matrix with entries from } F \text{ and } \det(A) = 1)\}$$

If $k \in \{\mathbb{R}, \mathbb{C}\}$, then geometrically these can be thought of as the volume and orientation preserving functions.

Definition 1.2.22: Orthogonal Group

$$O_n(\mathbb{F}) = \big(\{A|A^tA = AA^t = I\big)\}$$

In this group, the transformations are *distance-preserving*, which as an exercise the reader can check that for $O_2(\mathbb{R})$, all transformations are rotations and reflections across lines passing through the

[18] We can actually show this once we get to isomorphisms (definition 1.4.7)

origin[19]. Limiting to orthogonal matrices with determinant 1 gives us the set of all rotations:

Definition 1.2.23: Special Orthogonal Group

$$SO_n(\mathbb{F}) = \{A | A^t A = AA^t = I\,, \det(A) = 1\}$$

If the underlying field is clear, we sometimes write $SO(n)$.

Exercise 1.2.3

1. 6, 7. 11 If you feel like getting comfortable with Heisenberg groups
2. (If you know the dot product) Recall that if v, w are vectors then the dot product can be represented as $v \cdot w = v^T w$ where v^T is the transpose. Show that matrices in $O(\mathbb{R})$ preserve the dot product (namely $Av \cdot Aw = v \cdot w$) (perhaps move this to when dot product is introduced?? In section 13.5)

Proposition 1.2.24: Order Of $GL_n(\mathbb{F}_p)$

Let $G = GL_n(\mathbb{F}_p)$ where $\mathbb{F}_p$ is a finite field (and thus, has prime order). Then

$$|GL_n(\mathbb{F}_p)| = \prod_{i=0}^{n-1}(p^n - p^i) = (p^n - 1)(p^n - p)(p^n - p^2)\cdots(p^n - p^{n-1})$$

Proof **:**
The proof for this comes down to remembering that $GL_n(\mathbb{F}_p)$ has linearly independent columns since for any $A \in GL_n(\mathbb{F}_p)$ $\det(A) \neq 0$. We will first show it for $n = 2$. We can separate it in cases where matrices are definitely invertible, and matrices which might not might be invertible. Matrices which only have 1 entry are always non-invertible. These type of matrices are always invertible (since their determinant is always non-zero):

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \quad \begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix} \quad \begin{pmatrix} a & b \\ c & 0 \end{pmatrix} \quad \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \quad \begin{pmatrix} 0 & a \\ b & c \end{pmatrix} \quad \begin{pmatrix} a & 0 \\ b & c \end{pmatrix}$$

for $a, b, c > 0$ However

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Is not invertible if the first column is a scalar multiple of the second.

There are a total of $(p-1)^2$ types of unique columns (corresponding to the $p-1$ possibilities per element in the first column). Since the scalars are in $\mathbb{Z}/p\mathbb{Z}$, then there is $p-1$ unique scalar multiples of any row. Since there are $(p-1)^4$ total non-zero matrices, then we have

$$(p-1)^4 - (p-1)^3$$

invertible matrices with all non-zero entries.

[19] Higher dimensional analogues of "rotation" and "reflection" come from taking this interpretation in lower dimensions and claiming it generalizes to higher dimensions

What's left is to calculate the other types of matrices which are

$$2(p-1)^2$$
$$4(p-1)^3$$

So we have a total of

$$(p-1)^4 - (p-1)^3 + 2(p-1)^2 + 4(p-1)^3$$

possible invertible matrices.

TBD

1.3 Generators and Relations

(When dealing with numbers, they can represent anything, for example 5 apples, 5 cars, 5 textbooks, and so forth. However, when we defined the dihedral group. it was specifically linked to the rotation and reflections of n-gons. In this section, we will show a way to abstract away this dependency on a concrete object to define a group).

Notice that D_n can be written with r and s (reflections and rotations). Similarly, $\langle x \rangle \leqslant G$ is an element from which we can find every other element of $\langle x \rangle$. In a sense, they *generated* the group. This notion will be treated rigorously with free groups later (section 2.3.3), but intuitively this idea makes sense. For now, here how we'll define them

Definition 1.3.1: Generators

A subset S of elements of a group G with the property that every element of G can be written as a (finite) product of elements of S and their inverses is called the set of *generators* of G. If $x \subseteq G$ generates x, then we write $\langle x \rangle$. Similarly, if $x, y \in G$ are needed to generate G, we write $\langle x, y \rangle$. More generally, if $A \subseteq G$ generates G we write $\langle A \rangle$.

Example 1.3.1: Generators

1. An easy example would be 1 for $\mathbb{Z}$, since every $x \in \mathbb{Z}$ could be written as adding 1 or it's inverse. So $\mathbb{Z} = \langle 1 \rangle$.
2. By property (4) for D_n, $S = \{r, s\}$ is a set of generators of D_n, so $D_n = \langle r, s \rangle$.

Later, we'll see for finite groups that we don't need the inverse operation to generate finite groups.

For any set of generators, we have a set of relationships that they satisfy[20]. For D_n, they are

$$r^n = 1, \qquad s^2 = 1, \qquad rs = sr^{-1}$$

Why are these the relations? Because any additional relation between elements of groups may be derived from these three (p.40). Note that there can be more than one set of relations for a group (we'll see later that D_n has another set of relations that can be used).

[20] With only one exception. When a set of generators has no relationships, we call the group generated by that set a *free group*

Definition 1.3.2: Presentation of a group

if some group G is generated by a subset S and there is some collection of relations, say $R_1, \cdots, R_m$, which are all equations with elements from $S \cup \{1\}$, such that any relation among the elements of S an be deduced from these, we shall call these generators and relations a *presentation* of G, and write

$$G = \langle S | R_1, \cdots, R_m \rangle$$

Example 1.3.2: Presentation

1. Show that the presentation:

$$\langle r, s | r^n = 1, s^2 = 1, rs = sr^{-1} \rangle$$

represents the Dihedral group D_n

2. Let A be any set. Show that

$$\langle A | ab = 1, \forall a, b \in A \rangle$$

is group. In particular, show that from any set, we can define the trivial group. In section 1.2.2 and 2.3.3, we'll show two other (those times non-trivial) examples of creating groups from arbitrary sets.

Note: the right-hand side is, in the general case, a collection of arbitrary equations. It may be hard to determine when two elements in the group are equal! it may also be hard to determine the order of the group. For example

$$\langle x, y | x^2 = y^2 = (xy)^2 = 1 \rangle$$

is a presentation of a group of order 4, whereas

$$\langle x, y | x^3 = y^3 = (xy)^3 = 1 \rangle$$

is a presentation of an infinite group (exercise)

It might be tempting to only work with presentations, or to try and come up with every possible group working with generators. A considerable amount of time has been poured into this problem, and absolutely incredibly, it turns out that you can come up with presentations which are *insoluble*:

Example 1.3.3: Insoluble Presentation

The group representation by this monstrous presentation:

$$
\begin{array}{llll}
\langle & a,b,c,d,e,p,q,r,t,k & \mid & \\
& p^{10}a = ap, & pacqr = rpcaq, & ra = ar, \\
& p^{10}b = bp, & p^2adq^2r = rp^2daq^2, & rb = br, \\
& p^{10}c = cp, & p^3bcq^3r = rp^3cbq^3, & rc = cr, \\
& p^{10}d = dp, & p^4bdq^4r = rp^4dbq^4, & rd = dr, \\
& p^{10}e = ep, & p^5ceq^5r = rp^5ecaq^5, & re = er, \\
& aq^{10} = qa, & p^6deq^6r = rp^6edbq^6, & pt = tp, \\
& bq^{10} = qb, & p^7cdcq^7r = rp^7cdceq^7, & qt = tq, \\
& cq^{10} = qc, & p^8ca^3q^8r = rp^8a^3q^8, & \\
& dq^{10} = qd, & p^9da^3q^9r = rp^9a^3q^9, & \\
& eq^{10} = qe, & a^{-3}ta^3k = ka^{-3}ta^3 & \rangle
\end{array}
$$

Figure 1.3: Group for which we cannot write out what the elements are

has elements which we *cannot* write out!

The details of this is are well beyond this course, but it's existence is still interesting to know about

`https://en.wikipedia.org/wiki/Word_problem_for_groups`

You can get familiar with presentations by doing exercises in the book (p.41,42).

Generators of Symmetric Group

There are many ways of looking at the generators of S_n. The first one is perhaps the easiest one:

1. Take all adjacent transposition $S = \langle (i, i+1) \rangle$.

***Proof* :**
This proof relies on this key observation:

$$(a\,b)(a\,c)(a\,b)^{-1} = (a\,b)(a\,c)(a\,b) = (b\,c)$$

that is, we shifted a to b (so in a sense if $\sigma = (a\,b)$ and $\tau = (a\,c)$, $\sigma\tau\sigma^{-1} = (\sigma(a), \sigma(b))$).

With that in mind, we'll proceed by *induction*. Clearly if $n = 2$, then S_2 is generated by this set, so assume it generate S_{n-1} and consider S_n. Take some cycle with n in it:

$$(a_1\,a_2\,...\,a_k\,n)$$

which we've shown earlier can be decomposed as

$$(a_1\,a_2)(a_1\,a_3)\cdots(a_1\,a_k)$$

All the $(a_1\,a_j)$ are in G by the induction hypothesis (they are not in because of the generating set because we don't know if $(a_1\,a_j) = (i\,i+1)$). What's left to show is whether $(a_1\,n)$ is in G. To do this, we use the conjugation fact we brought up earlier:

$$(a_1\,n) = (a_1\,n-1)(n-1\,n)(a_1\,n-1)$$

which since $(n-1\,n) \in S$, and the other two terms are in G by induction hypothesis, it means that $(a_1\,n) \in G$. And Since any cycle can be broken down into the aforementioned

> form (where if we have a cycle, we can move k to the back, since its the same cycle that represents it), any cycle can be broken down in this fashion, and thus, we completed the proof

As a bonus, if you want to represent such a group:

$$\langle \sigma_i | \sigma_i^2 = 1 \,, \sigma_i \sigma j = \sigma_j \sigma i \text{ for } |i-j| > 1 \,, (\sigma_i, \sigma_{i+1})^3 = 1 \rangle$$

Though this is not necessary to know, but it can be good to know if you want to show something to be isomorphic to the symmetric group (since if something is isomorphic, it has the same relations).

2. Take $(1, i)$ for $2 \leq i \leq n$, that is, the set where you fix one element, then the 2nd position is all other elements. This is actually easy to show by example:

> ***Proof*** **:**
> If you can come up with a set like the last proof, then you know it generates it. Think about conjugating again.

3. This one is probably the most useful theoretically. You can actually generate a symmetric group given only 2 cycles!! Take any n-cycles from S_n, and a 2-cycle of adjacent elements in the n-cycle!

> ***Proof*** **:**
> Once I have time, I'll attempt it.

`https://en.wikipedia.org/wiki/Symmetric_group#Generators_and_relations`.

There is an elegant way to prove these once we introduce either quotient groups or conjugacy classes (in the group-action section). Thus, we won't prove these here for the nicest way to prove them is once we have those extra tools.

Exercise 1.3.1

1. exercise 7, 16, simply to know how to solve a concrete problem.
2. exercise ,17,18.
3. Do 10 to see that you can find all symmetries

1.4 Homomorphisms and Isomorphisms

When comparing D_8 and Q_8 in example 1.2.8, we resolved at looking at the Cayley tables to distinguish the two. However, this becomes very tedious to do for larger groups, and downright impossible for infinite groups. In this section, we'll make precise the notion of comparing groups.

What does it mean to compare groups? If we were comparing tuples, it is clear that $(a, b) \neq (b, a)$, since the order matters. On the other hand, if we're comparing sets, then $\{a, b\} = \{b, a\}$ since the

order doesn't matter, as long as the elements are the same. For groups, the star of the show is the binary operation. If two groups G and H shared *exactly* the same binary operation, they'd be identical in terms of the binary operation. We often encounter groups with different elements, but the binary operation acts identically in both groups.

Example 1.4.1: "Identical" groups

Let $G = \{0, 1, 2\}$ and $H = \{a, b, c\}$, and define $(G, +_G)$ and $(H, +_H)$ with binary operation's which have Cayley tables:

$+_G$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

and

$+_H$	a	b	c
a	a	b	c
b	b	c	a
c	c	a	b

Then notice we can replace 0 with a, 1 with b, and 2 with c, and the two tables would be identical. In other words, the operation $+_H$ acts exactly the same as $+_G$.

It is sometimes restrictive that the two binary operations act *exactly* the same. sometimes that are similar, but the two groups share a different structure. The take-away is that when working with homomorphisms and isomorphissm, what we care about is the binary operation between two groups behave similarly.

1.4.1 Homomorphisms

Definition 1.4.1: Homomorphism

Let $(G, \cdot)$, and $(H, \bullet)$ be groups. A map $\psi : G \to H$ such that for all $x, y \in G$,

$$\psi(x \cdot y) = \psi(x) \bullet \psi(y)$$

is called a *homomorphism*

Remember that the left side is the domains operation, and the right side is the co-domains operation. A good way of remembering the rule of a homomorphism is that the resulting image $\varphi(G) = \{\varphi(g) \mid g \in G\}$ is also a group (this is proved in proposition 1.4.2). Another way remembering the rule of a homomorphism is that both if there is a homomorphism, then in some way the binary operation must be *preserved*. It is not necessarily perfectly preserved – it's possible that only some of it is preserved. The following examples of homomorhpisms may help in gaining an intuition

Example 1.4.2: Homomorphisms

1. Take the integers with the usual addition. Let e represent even integers; o represent odd integers. Then we know that adding even with even is even, even with odd is odd, odd with even is odd, and odd with odd is even:

$$e + e = e,\ o + o = e,\ o + e = e + o = o$$

On the other hand, if we take 1 and -1 and find all possible outcomes of multiplication, we get

$$1 \cdot 1 = 1, \quad (-1) \cdot (-1) = 1, \quad 1 \cdot (-1) = (-1) \cdot 1 = -1$$

Notice how 1 can represent even integers, and -1 represent odd. In this way, multiplication on -1 and 1 "captures" the idea of adding even and odd numbers.

Let's formalize this idea by making an explicit homomorphism. First, $\{-1, 1\}$ is a group under usual multiplication $\cdot$. The element 1 is the identity since $1 \cdot 1 = 1$, $-1 \cdot 1 = -1$. Furthermore -1 it's own inverse ($-1 \cdot -1 = 1$). For associativity, most cases are essentially immediate to check, with the following case being the non-trivial case:

$$(-1 \cdot -1) \cdot -1 = 1 \cdot -1 = -1 \cdot 1 = -1 \cdot (-1 \cdot -1)$$

Any case with 1 in the question be more easily checked since $a \cdot 1 = 1 \cdot a = 1$, making the cases of $1 \cdot -1 \cdot -1$, $-1 \cdot 1 \cdot -1$, $1 \cdot 1 \cdot -1$, and so forth easy to check. Thus, $\{-1, 1\}$ is a group.

Next, let $\psi : \mathbb{Z} \to \{-1, 1\}$ be a function such that

$$\psi(x) = \begin{cases} 1 & x \text{ is even} \\ -1 & x \text{ is odd} \end{cases}$$

Then if e represents an arbitrary even number and o an arbitrary odd number:

$$\begin{gathered} \psi(e + e) = \psi(e) = 1 = 1 \cdot 1 = \psi(e) \cdot \psi(e) \\ \psi(o + o) = \psi(o) = 1 = (-1) \cdot (-1) = \psi(o) \cdot \psi(o) \\ \psi(o + e) = \psi(e + o) = \psi(o) = -1 = -1 \cdot 1 = 1 \cdot -1 = \psi(e) \cdot \psi(o) = \psi(o) \cdot \psi(e) \end{gathered}$$

Showing that ψ is indeed a homomorphism.

Generalizing the previous result, notice that our group $\{-1, 1\}$ is homomorphic to $\mathbb{Z}/2\mathbb{Z}$ (remember that by default, if we say $\mathbb{Z}/2\mathbb{Z}$ is a group, we mean $(\mathbb{Z}/2\mathbb{Z}, +)$ with modulo 2 addition). In particular

$$f : \{-1, 1\} \to \mathbb{Z}/2\mathbb{Z} \qquad f(1) = 0,\ f(-1) = 1$$

Then one should verify that that f is indeed a homomorphism. It is further a bijection, meaning that every element of the group $\mathbb{Z}/2\mathbb{Z}$ represents exactly one element from $\{-1, 1\}$, showing we can inter-change the two.

Now, take $\varphi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$, $\varphi(x) = [x]$. Then φ is a homomorphism. Indeed:

$$\begin{aligned} \varphi(x+y) &= [x+y] \\ &= [x] + [y] && \text{property of } \mathbb{Z}/n\mathbb{Z} \\ &= \varphi(x) + \varphi(y) \end{aligned}$$

If $n = 2$, then this will be exactly the homomorphism we defined earlier.

2. If you have done any calculus or analysis, you might have seen before that:

$$\frac{d}{dx}(f+g) = \frac{d}{dx}(f) + \frac{d}{dx}(g) \qquad \int f + g\,dx = \int f, dx + \int g\,dx$$

so $\frac{d}{dx}$ and $\int$ both satisfy this condition! Are they homomorphisms? In fact, they are, with the group being all differentiable functions for the 1st one, and all integrable functions for the second one. These groups require more build-up within Calculus or Analysis, and so we will not cover these operation in more detail. (More words here! Feels incomplete)[a].

3. For those with knowledge of the exponential function e^z (which we'll write as exp), then there is a natural way of relating the addition and multiplication operation of $\mathbb{C}$ and $\mathbb{C}^\times$ (as well as $\mathbb{R}$ and $\mathbb{R}^\times$). The map $\exp : \mathbb{C} \to \mathbb{C}^\times$ is surjective group homomorhpism. Restricting the domain and co-domain, we also get an injective (but not surjective) map $\exp : \mathbb{R} \to \mathbb{R}^\times$.

4. Verify that there is no bijective homomorphism from D_8 to Q_8. At the moment, this may be a tedious task, it will become easier to show later with some more properties of bijective homomorphisms.

5. Let $G_1 \times G_2$ be the direct product of two groups G_1 and G_2 as defined in example 1.1.2(4). The map $\pi_1 : G_1 \times G_2 \to G_1$ and $\pi_2 : G_1 \times G_2 \to G_2$ defined by

$$\pi_1((g_1, g_2)) = g_1 \qquad \pi_2((g_1, g_2)) = g_2$$

are homomorphisms.

6. Let G, H be any group. Then $\varphi : G \to H$ defined by $\varphi(g) = e_H$ where e_H is the identity of H is a homomorphism. This is called the *trivial homomorphism.*

7. (If you know about $\mathbb{C}$) Show that the subset $\{1, i, -1, -i\} \subseteq \mathbb{C}$ is a subgroup of $\mathbb{C}$ and that there is a bijective group homomorphism $\varphi : \mathbb{Z}/4\mathbb{Z} \to \{1, i, -1, -i\}$.

[a] However, $\frac{d}{dx}(fg)$ is not necessarily equal to $\left(\frac{d}{dx}(f)\right)\left(\frac{d}{dx}(g)\right)$ and $\int fg\,dx$ is not necessarily equal to $\left(\int f\,dx\right)\left(\int g\,dx\right)$. The first is studied in a course on Lie Groups and Lie Algebras, while the other is studied in a course in Measure Theory, in particular L^p spaces and Hölder's inequality

We will often simplify notation since lots of information is usually deduced from context. We will drop the binary operation and simply write $\varphi(ab) = \varphi(a)\varphi(b)$, given that it's clear that when we write ab the binary operation happens in the domain, and that when we werite $\varphi(a)\varphi(b)$ the binary operation happens in the codomain.

One more interpretation of a homomorphism that is a little more abstract but is enlightening in it's own way is that a homomorphism f is a function that "commutes" with the binary operation: If $*$ represent any binary operation (of any group, not just one group), then f is a homomorphism if and

only if

$$f \circ * = * \circ f$$

Concretely, if we let $*$ for a group G be represented as $*_G$ and $*$ for a group H be represented as $*_H$, then

$$f \circ *_G = *_H \circ f$$

or, if written in terms of elements and with the standard notation for binary operation:

$$f(x *_G y) = f(x) *_H f(y)$$

This again show's how we can think of commutativity as a powerful property. This idea can be captured in a diagram like so:

$$\begin{tikzcd} G \times G \arrow[r, "\varphi\times\varphi"] \arrow[d, "*_G(a,b)"'] & H \times H \arrow[d, "*_H(\varphi(a),\varphi(b))"] \\ G \arrow[r, "\varphi(ab)"'] & H \end{tikzcd}$$

which can be read as "start at $G \times G$, then either go down then right, or go right then down, and both paths should give the same result", i.e. $\varphi(ab) = \varphi(a)\varphi(b)$.

Now that homomorphisms are defined, you might ask yourself "given two groups G, and H, how do I know there is a homomorphism between them?" There is always a trivial answer to this, the homomorphism sending everything to the identity always exist [21]. This is called the *trivial homomorphism* (you should check that it's indeed a homomorphism):

$$\varphi : G \to \{e\} \qquad \varphi(g) = e$$

So the better question would be "How can I find the non-trivial homeomorphisms between these two groups?" This in some cases can be a very difficult question. If we just look at them as set-functions from G to H, then if there are n objects in one category and m in the other, there are minimum n^m possible functions! That's a ton to check. We can deduce properties of homomorphisms which will let us diminish that number:

Proposition 1.4.2: Properties of Homomorphisms

Let G, H be groups and φ be a homomorphism. Then:

1. $\varphi(1_G) = 1_H$
2. $\varphi(g^{-1}) = \varphi(g)^{-1}$
3. $\varphi(g^n) = \varphi(g)^n$
4. $\ker \varphi$ is a subgroup of G
5. $\mathrm{Im}(\varphi)$, the image of G under φ is a subgroup of H.
6. if $\psi : H \to K$ is also a homomorphism, then $\psi \circ \varphi : G \to K$ is a homomorphism, i.e., the composition of homomorphisms is a homomorphism.

[21] Elaborated in proposition 1.4.3

Proof **:**

1. $\varphi(1_G) = \varphi(1_G 1_G) = \varphi(1_G)\varphi(1_G)$, so $\varphi(1_G) = \varphi(1_G)\varphi(1_G)$. Then by proposition 1.1.9
$$\varphi(1_G)\varphi(1_G)^{-1} = \varphi(1_G)\varphi(1_G)\varphi(1_G)^{-1}$$
the left hand side becomes 1_H, and the right hand side becomes $\varphi(1_G)1_H = \varphi(1_G)$, giving us
$$\varphi(1_G) = 1_H$$
2. Hint: take $\varphi(1_G) = \varphi(gg^{-1})$, and apply linearity and proposition 1.1.9
3. Hint: $\varphi(g^n) = \varphi(g^{n-1}g)$ and use the fact that φ is a homomorphism
4. Hint: Take the set $\ker(\varphi) := \{g \in G \mid \varphi(g) = 1_H\}$, and check the 3 axioms of a group
5. Hint: Take $\varphi(G) := \{h \in H \mid \exists g \in G \text{ such that } \varphi(g) = h\}$ and check the 3 axioms of a group
6. Hint: Apply the homomorphism condition twice.

These propositions can even help us deduce whether [non-trivial] homomorphisms exist. For example, is there a non-trivial homomorhpism from $\mathbb{Z}/2\mathbb{Z}$ to $\mathbb{Z}$? Let's say $\varphi : \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}$ were such a map. By the above proposition we have $\varphi(0) = 0$. Let's say $\varphi(1) = n$. Then
$$0 = \varphi(0) = \varphi(1+1) = \varphi(1) + \varphi(1) = n + n$$
giving us $0 = n + n$, which re-arranging is $-n = n$. The only integer equal to its negative is 0, hence $n = 0$. This shows that there is no non-trivial map from $\mathbb{Z}/2\mathbb{Z}$ to $\mathbb{Z}$.

The particular structure of the group that is being preserved depends on the operation and the cardinality of the domain and co-domain. In general, to get an idea of what part of the group is being preserved, you'll have to check by doing a couple of example.

Exercise 1.4.1

1. Find all non-trivial homomorphisms between D_8 and Q_8
2. (If you know some complex analysis) Let S^1 be the circle in the complex plane. show that $f : \mathbb{R} \to S^1$, $f(x) = e^{2\pi i x}$ is a homomorphism.
3. Perhaps another exercise like the one above? Perhaps a $G \to G$ example? I think getting some ideas between which groups are homomorphic to each other to be useful for long-term mathematical career.
4. is there a non-trivial homomorphism between $\mathbb{Z}/n\mathbb{Z}$ and $\mathbb{Z}$? answer:[22]
5. Let $\varphi : G \to H$ be a homomorphism. Show that $\varphi(a+b+c) = \varphi(a) + \varphi(b) + \varphi(c)$. More generally, if $g_1, g_2, ..., g_n \in G$, then $\varphi(g_1 + g_2 + \cdots + g_n) = \varphi(g_1) + \varphi(g_1) + \cdots + \varphi(g_1)$. [hint: recall assocaitivity and induction]
6. Does there exist a surjective homomorphism $\varphi : G \to H$ where G is abelian and H is not?

[22]There does not exist a non-trivial homomorphism $\varphi : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}$, since every non-identity element of $\mathbb{Z}/n\mathbb{Z}$ ha finite order, but all elements of $\mathbb{Z}$ has infinite order.

7. Let S where $|S| = 3$ be a set and let $\mathrm{Sym}(S)$ all bijections on S (i.e. the symmetric group). Show that there exists a homomorphism $\sigma : D_6 \to \mathrm{Sym}(S)$. (it is actually an isomorphism, move this exercise?) (This is a particular example of a *group action*, which we will explore in detail in chapter 4)
8. Define the map $\varphi_g : G \to G$, $\varphi(x) = gx$. Is this map in general a homomorphism? If not in general, in a particular case?
9. Define the map $\psi_g : G \to G$, $\psi(x) = gxg^{-1}$. Is this map in general a homomorphism? If not in general, in a particular case?
10. Let $\varphi : G \to H$ be a homomorphism and $\ker(\varphi) = \{e_G\}$. Prove that φ is injective.
11. Show that there exist an injective homomorphism $\varphi_n : D_n \hookrightarrow O_2(\mathbb{R})$.

Important examples

I'll go over some homomorphisms to build a deeper understanding of the concept

The first one is deceptively boring: take $\varphi : G \to \{e\}$ such that $\varphi(g) = e$. The function φ is a homomorphism, since for any $g, h \in G$

$$\varphi(g \cdot h) = e = e \cdot e = \varphi(g) \cdot \varphi(h)$$

This seemingly boring homomorphism has actually a very important property: for *any* group G, there *always* exist a *unique* isomorphism from G to $\{e\}$. This is also true for homomorphisms with $\{e\}$ as the domain! This fact is important enough to make a proposition:

Proposition 1.4.3: Homomorphism To/From Trivial Group

Let G be any group and I represent the trivial group. Then there always exists a unique homomorphism

$$\varphi : G \to I \qquad \psi : I \to G$$

to and from I

Proof **:**
Let e_G represent the identity of G and e_I represent the element of I. We've already shown φ is a homomorphism, so let's now show uniqueness. Let's say φ' is another homomorphim with the same domain/codomain. Note that

$$\varphi'(e_G) = \varphi'(e_G \cdot e_G) = \varphi'(e_G)\varphi'(e_G) \;\Rightarrow\; \varphi'(e_G) = \varphi'(e_G)\varphi'(e_G)$$

and by proposition 1.1.9, we can multiply each side by$(\varphi')^{-1}$ and get

$$e_I = \varphi'(e_G)$$

meaning the identity *must* map to the identity. But that would then make φ' identical to φ, thus

$$\varphi = \varphi'$$

showing that φ must be unique.

A similar proof works for ψ.

So why is this property important? What can we do with it? In this case with groups, it is not used often [23]. what's more important is the concept that there *always* exists a homomorphism from any group, and that it is a *unique* homomorphism. These properties for the right homomorphisms will come back big-time when we talk about *quotient groups* (chapter 3): For *every* homomorphism there *always* exists a *unique* group (called the quotient group) which will "represent" the homomorphism.

Thus, we'll give a name to groups with this property:

Definition 1.4.4: Initial And Final Groups

Let G be a group. If for any other group H, then

1. if there exists a unique group $\varphi : G \to H$, then we say that G is an *initial group* (or *initial object*)
2. if there exists a unique homomorphism $\varphi : H \to G$, then we say that G is an *final group* (or *final object*)
3. if G is either initial or final, we say that G is a *terminal group* (or *terminal object*)

The reason for this vocabulary is clearer is you think of a graph where the edges are homomorphisms and the nodes are groups:

$$\begin{tikzcd}
 & U \arrow[d, shift left] \arrow[d, shift right] \arrow[dl, bend right] \arrow[dr, bend left] & \\
H \arrow[r, bend left] & L \arrow[l] \arrow[r] \arrow[d] & M \arrow[dl] \\
G \arrow[u, bend left] \arrow[ur] \arrow[r] \arrow[urr] & K \arrow[ur] & \arrow[ul, bend left] \arrow[ul, bend right]
\end{tikzcd} \tag{1.9}$$

You can see how there is only one homomorphism from G to any other group, and M has only one homomorphism from every other group. Thus, G is initial, H is final, and both G and H are terminal.

> **Note**: An object *can* be both initial and final. The trivial group is one such example

For now, we do not have the tools to properly analyze this result; we will gain those in chapter 2, and revisit this idea in chapter 3.

Redefining Subgroup

Moving back towards things we can prove, I'll present an important way of using homomorphisms to *define* subgroups, using the fact that subgroups as a similar structure to the original groups, and homomorphisms preserve some structure of a group. It might already be somewhat intuitive that if $H \leqslant G$, then H is pretty similar to G, especially that it must have the same group operation. With homomorphisms, we can formalize this idea:

[23] It is used in the context of Category theory, where $\{e\}$ is both the initial and final group in the category of **Group**, making it an intuitive first step towards introducing universal properties. More on this in section 8

Proposition 1.4.5: Defining Sub Group With Homomorphism

Let $(G, \cdot)$ and $(H, \bullet)$ be two groups, and $H \subseteq G$. Then H is a subgroup if and only if there exists a inclusion map $i : H \to G$ which is also a homomorphism

Proof **:**
We'll show that this definition is equivalent to definition 1.1.3, that is, that $(H, \cdot)$ is a group. Let $i : H \to G$ be an homomorphism which is also an inclusion map. All of these will come from proposition 1.4.2.

1. $(e_G = e_H)$ We need to how that H has the same identity. This comes for free with proposition 1.4.2
2. (inverse) We need to show that for any $a \in (H, \cdot)$, then there is a $b \in (H, \cdot)$ such that $ab = ba = e_G$. Since $i(a^{-1}) = a^{-1}$, then a^{-1} is a candidate. Indeed, since
$$i(e_H) = i(aa^{-1}) = i(a)i(a^{-1}) = i(a)i(a)^{-1} = aa^{-1} = e_G$$
then a^{-1} will indeed be our desired inverse
3. (associativity) Similarly to how we've down the other two, but the proof is a little more tedious to carry out. One can do this to try proving associativity at least once in their career.

This is probably one of my favourite proposition using homomorphisms: really showing how homomorphism captures the idea that an algebraic structure is preserved by defining subgroups with an inclusion map that is also a homomorphism.

Collection of Homomorphisms

Next, we will address a little more complicated object. We might ask ourselves in what other way can we use homomorphisms to study groups? If a homomorphism preserves group structure, what possible homomorphisms can we have between a group and itself? What would that give us? Let's first label the set of all such homomorphisms:

Definition 1.4.6: Set of Endomorphisms

1. Given a group G, an endomorphism is a homomorphism from an object to itself: $\varphi : G \to G$.
2. Let $\text{End}(G) := \{\varphi : G \to G \mid \forall a, b \in G,\ \varphi(ab) = \varphi(a)\varphi(b)\}$ represent the set of all endomorphisms on G

What $\text{End}(G)$ represent is the set of all possible group structure's that can be derived form G. In the example with $\mathbb{Z}$ being mapped to $\{-1, 1\}$ (example 1.4.2), notice that we can make the codomain all of $\mathbb{Z}$, and the function remains identical. In this way, we would make that function an element of $\text{End}(\mathbb{Z})$, and so this shows that the function in this example captures one of the possible sub structures of $\mathbb{Z}$ (in particular, a special kind of subgroup, we ill explore this in more detail in chapter

ref:HERE)

If we can figure out some structure on $\text{End}(G)$, we can start getting a good idea on what are all sub-structures in G. More excitingly, if $\text{End}(G)$ can be proven to be a group (with composition, $\circ$, as the operation), then we can use all the technics of group theory to study it. However, this turns out not to be the case. The problem comes when trying to prove the existence of an inverse, which doesn't necessarily exist (try proving it's a group!).

So you might ask if there is still a way to make $\text{End}(G)$ into a group; perhaps with a nice operation, we can define a [non-trivial] group structure? The one that is usually first attempted is defining addition *pointwise*, that is

$$f, g \in \text{End}(G) \qquad f + g \equiv (f + g)(x) = f(x) + g(x)$$

Where $\equiv$ means we are showing how the ne function "$f + g$" is defined by how it processes every element. With the operation defined, to check that $f + g$ is in $\text{End}(G)$, we need to check if it's a homomorphism, that is

$$(f + g)(x + y) = (f + g)(x) + (f + g)(y)$$

and going through the computation:

$$\begin{aligned}(f + g)(x + y) &= f(x + y) + g(x + y) \\ &= f(x) + f(y) + g(x) + g(y)\end{aligned}$$

and here is were we get stuck, since G is not necessarily commutative.

However, what if G was commutative? Well, the result will actually follow much more nicely!

$$\begin{aligned}&= f(x) + f(y) + g(x) + g(y) \\ &= f(x) + g(x) + f(y) + g(y) \\ &= (f + g)(x) + (f + g)(y)\end{aligned}$$

showing that $(f + g)$ is indeed a homomorphism!! Thus, $f + g$ is a well-defined operation. From here, you can check the axioms of a group to check that $\text{End}(G)$ is indeed a group. It's even an abelian group! This is an example of how being commutativity changes how a group effects.

5: Remark

In fact, more is true. If the domain of the functions are arbitrary sets, then $\text{Hom}(A, G)$ is *still* an abliean group! This fact will be very useful to us much further don the line, in part II, III, and VI.

Generalization of 2D and 3D rotations; intuition on 4D

Point Groups? (`https://en.wikipedia.org/wiki/Point_groups_in_two_dimensions`) (and the 3D one too?)

Finally, we will take a moment to detour into trying to use groups to build an intuition for 4+ dimensional objects! As we've seen, D_n represents rotations of 2D polygons, and Q_8 represents rotations of 3D objects. What about rotations and symmetries of higher dimensional objects? To find such a group, we will show that there is a homomorphism $\varphi_D : D_n \Rightarrow SO_3(\mathbb{R})$ and $\varphi_Q : Q_8 \to SO_3(\mathbb{R})$

First, we'll define φ_D:

$$r \to \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad d \to \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

TBD (do same with quaternion. Perhaps take a look at representation theory to figure out Q_8, Generalize to 4D with $SO(4)$)

Example 1.4.3: exercicss
gaining intuition on matrix multiplication

1. matrix multiplication could've been defined component-wise, so why wasn't it? Build up to (using the appropriate vocabulary, not the following one) that given $\dim M = n$, $\dim N = n$, $\mathrm{Hom}_{k-\mathbf{Vect}}(M, N) \cong_{k-\mathbf{Vect}} M_n(k)$, and we'll get $\Psi(f \circ g) = A_f \cdot A_g$ (i.e. composition becomes matrix multiplication)
2. elaborate if dimensions change?

Exercise 1.4.2

1. Let G be an arbitrary group.
    1. Prove that there always exists a homomorphism from $\mathbb{Z}$ to G
    2. Why does this not contradict the fact that I is the only initial group?

1.4.2 Isomorphisms

While homomorphisms preserve some algebraic structure, isomorphisms preserve the *entire* algebraic structure. An isomorphism can be thought of as the generalization of the concept of equality: If two groups are isomorphic, they are identical (as groups). They will only differ in the symbols which are used in representing them. The idea of an isomorphism extends far beyond group theory. Here's a fun linguistic example: if I say

one, two, three, four, five

or

un, deux, trois, quatre, cinq

Then both are capturing the idea of $1, 2, 3, 4, 5$, except in one I'm writing it in English, and in the other in French. Those two sentences ostensibly capture the same [numerical] ideas, but with different symbols

When trying to capture this idea in groups, we have to consider what it means for two group to be identical. First, the two groups would need to be of the same size, if not we cannot say they are the same. Therefore, the homomorphism must also be bijective. Next, all the relation between the two groups need to be the same. In example ref:HERE, we only preserved part of the relations.

Definition 1.4.7: Isomorphism

The map $\psi : G \to H$ is called an *isomorphism* and G and H are said to be *isomorphic* or of the same *isomorphism* type, written $G \cong H$ if

1. ψ is a homomorphism
2. ψ is a bijection

Note: Dummit showed this as the definition of isomorphism. But in a more categorical framework, this is not in general how you would introduce this. You would instead say that there exists an inverse function: $g : H \to G$ which is also homomorphic. It can be shown that this (for groups) implies a bijection. [a]

[a] If you've already studied topology, there's an example where a homomorphic bijection is not an isomorphism

When two groups are isomorphic, it means they only differ by the symbols we choose to use, they are in practice identical. The trivial isomorphism is the identity map $\mathrm{id} : G \to G$. We can also form an equivalence class of isomorphic groups: take $\mathcal{G}$ to the set of groups and let the equivalence relation be isomorphisms.

Example 1.4.4: Isomorphism Examples

Let V_4 (*Vierergruppe*, meaning 4-group, sometimes called Klein 4-group), is the group that represents the symmetries of a [non-square] rectangle. You can define it as

$$V_4 = \langle a, b | a^2 = b^2 = (ab)^2 = e \rangle$$

Show that:

$$V_4 \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$$

Where $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ has the same operation defined in the cartesian product example in ref:HERE.

A less obvious example of a group isomorphism is $\psi : \mathbb{R} \to \mathbb{R}^+$, $\exp(x) = e^x$. e^x is invertible, and thus bijective, and $e^{x+y} = e^x e^y$, preserving group structure.

Proposition 1.4.8: Properties Of Isomorphisms

(Motivated by properties of diffeomorphism in John Lee. In particular, I want to say that if $g \circ f$ and f is an isomorphism, [do we need to force g to be either injective or surjective?]so must g be)

Proof **:**

1. The composition of isomorphism is an isomorphism
2. If $g \circ f$ and g are isomorphisms, then since g is an isomorphism, so if $g^{-1} \circ g \circ f = f$, and similarly if f is an isomorphism: $g \circ f \circ f^{-1} = g$

An interesting exercise to show is if A and B are sets of the same cardinality, and S_A and S_B are two symmetric groups, then $S_A \cong S_B$. The book explains it on p.37, but personally I feel it's more intuitive if you think of it as relabeling since A and B are arbitrary sets.

Going back for a moment to hen e tried to define a group $(\text{End}(G), \circ)$ the fatal flaw as that not all homomorphisms have inverses. However, e have seen that all isomorphisms can be reversed, that is, i $\varphi : G \to H$ is an isomorphism, then there exists a ψ, usually denoted φ^{-1} such that $\varphi^{-1} : H \to G$ is an isomorphism and $\varphi^{-1} \circ \varphi = \varphi \circ \varphi^{-1} = \text{id}$ (here id is either id_G or id_H depending on whether φ is on the right or left in the composition respectively). If e take $\varphi : G \to G$ to e an isomorphism, then φ^{-1} ill *exactly* have an inverse. Excitingly, this in fact define as group!

Definition 1.4.9: Automorphism

1. if $\varphi : G \to G$ is an isomorphism from a group to itself, then φ is called an ***automorphism***
2. Let $\text{Aut}(G) := \{\varphi : G \to G \mid \varphi \text{ is an automorphism}\}$ be the collection of all automorphisms of G

Proposition 1.4.10: $\text{Aut}(G)$ is a Group

Let G be a group and $\text{Aut}(G)$ be the collection of automorphisms. Then $(\text{Aut}(G), \circ)$ forms a group

Proof **:**
Proving it's a group should be doable at this point

Example 1.4.5: Automorphisms
Simplest example are identity maps. Other examples come form Linear Algebra with bijective linear functions (since $\varphi(a)\varphi(b) = \varphi(ab)$ for a linear function, it *is* an example)

6: Foreshadowing

The concept of automorphisms will not be limited to groups. Many algebraic and non-algebraic structures will have the concept of automorphism defined. Most excitingly, the set of automorphisms for *any* of these objects *always* forms a group! Because of this, we will take a hole chapter to study (TBD chap 4? better transition?)

:exercises

(make sub questions) example 1.1.2[5], we showed that $(\mathbb{Z}, +)$ and $(\mathbb{Z}, +_2)$ had a "different structure" by showing the identity and inverses are different. Show that they are in fact isomorphic. This means that they are not a good example of two different groups on the same set since algebraically, they are identical.

Show that $D_8 \ncong \mathbb{Z}/8\mathbb{Z}$. With this knowledge, conclude that we can define two different groups on the set of 8 elements, giving a proper example of two different groups structure on the same underlying set.

(Challenging) Try to come up with two different structure on the underlying set of $\mathbb{Z}$ (Hint: recall the structure of $\bigoplus_i^\infty(\mathbb{Z}/2\mathbb{Z})$ and show there is a bijection between $\mathbb{Z}$ and this set, so that you can define a group structure on $\mathbb{Z}$ relative to this bijection. Is this group isomorphic to $\mathbb{Z}$?)

In the following section, we will explore classifying all possible groups on n elements where n has small order (larger orders of n will come when we have more tools)

First Use of Isomorphisms

In mathematics, there's a whole set of theorems that we use to show that classes of mathematical objects, like groups, are isomorphic to each other or not. Such theorems are called *classification theorems*. As a simple example, we can classify all non-abelian groups of order 6, in particular:

7

any non-abelian group of order 6 is isomorphic to S_3

Holding the theorem to be true for now, we can conclude that $D_6 \cong S_3$, and $GL_2(\mathbb{F}_2) \cong S_3$ without needing to find an explicit map. Not all groups of order 6 are isomorphic to S_3 because they're not all abelian. However, all abelian groups of order 6 are isomorphic to $\mathbb{Z}/6\mathbb{Z}$! Thus, any group of order 6 is isomorphic to either S_3 or $\mathbb{Z}/6\mathbb{Z}$. This theorem will be proved in exercise ref:HERE.

In general, it is difficult, sometime unfeasible, to find an explicit isomorphism mapping. Here are some necessary, but not sufficient, conditions for isomorphisms:

Proposition 1.4.11: Isomorphism Conditions

Let $\psi : G \to H$ be an isomorphism. Then:

1. $|G| = |H|$
2. G is abelian if and only if H is abelian
3. for all $x \in G$, $|x| = |\psi(x)|$

Proof **:**
Fairly simple - left as an exercise.

Example 1.4.6
This proposition shows that S_3 and $\mathbb{Z}/6\mathbb{Z}$ are not isomorphic, because one is abelian and the other is not. Additionally, $(\mathbb{R} - \{0\}, \times)$ and $(\mathbb{R}, +)$ are not isomorphic, because $-1 \in \mathbb{R}^\times$ has order 2, and no element in $(\mathbb{R}, +)$ has order 2

There's a cool introduction to how two groups could be homomorphic given the presentation of a group, which will require free groups to prove formally. I'll omit this for now (p.52, bottom).

Example 1.4.7

1. Take D_n and it's presentation $\langle r, s | r^n = s^2 = 1, sr = r^{-1}s \rangle$. Now take $n = mk$, $k \geq 3$ and consider d_{2k}. Then

$$\psi : d_n \to d_{2k} \qquad \text{by} \qquad \psi(r) = r_1, \text{ and } \psi(s) = s_1$$

   is a homomorphism. In other words, the operation between Dihedral groups which are "multiples" are the same (think of a triangle and hexagon. two hexagon rotations is 2 triangle rotation, angle wise).

2. an example showing D_6 and S_3 are isomorphic. p.53

Proposition 1.4.12: Isomorphism properties

Let $\psi : G \to H$ be an isomorphism, and $\Psi : G \to H$ an isomorphism (if the context permits it)

1. Let $\psi : G \to H$ is a homomorphism. Prove that the identity maps to the identity
2. $\psi(x^n) = \psi(x)^n$, for all $n \in \mathbb{Z}$.
3. $|\Psi(x)| = |x|$ for all $x \in G$. Deduce that any two isomorphic groups have the same number of elements of order n for each $n \in \mathbb{Z}^+$. Is the result true for ψ? (hint: think of the "trivial homomorphism" which sends every element from G to $\{1\}$.)
4. If Ψ exists, then G is isomorphic if and only if H is isomorphic (easy)
5. If you know where the generators map, you know where all the entries map

Proof **:**
I'll do one of them:

1. $$1_H \psi(x) = \psi(x) = \psi(1_G \cdot x) = \psi(1_G) \cdot \psi(x) \Rightarrow 1_H = \psi(1_G)$$

The rest is good to do yourself to really understand the proposition.

The last point in proposition 1.4.12 is one to keep in mind. I found it particularly useful when trying to construct isomorphisms between groups and S_n (which you will extensively once you cover group-actions).

Example 1.4.8

Prove that $SL_2(\mathbb{F}_3)$ generated by $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ is isomorphic to the quaternion group of order 8

Proof. Label the first matrix A and the second matrix B. To start, bring up the presentation of

Q_8 to get an idea of it's structure. See if you can bring it down to two generators. What are the properties of these generators (i.e. representation). Does A and B match it? If so, what's then the isomorphism. □

Prove that a homomorphism on S_n means it will preserve the parity of the cycle

Proof. Check cases □

1.5 Group Actions

(I was thinking on whether this section should stay or not, since it seemed placed here unnecessarily. However, I think this should stay for motivational purposes. In this section, introduce the intuition of a group action on regular n-gons, or on a clock, or on symmetric groups, and perhaps even matrices when talking about equivalence much later. Then, in the group-action section, we can talk more about the theoretical sides, like the universal prop, the Orbit-Stabilizer theorem, Sylow's theorem, groups acting on groups, etc.)

(Maybe use Quaternions because of their application to quantum physics? So can say 3d rotation or spin of particle?) Recall how D_8 can represents all the rotations on a square. This is not the unique thing that D_8 can represent. What if we present the following graph:

here

Then $\mathrm{Aut}(\mathcal{G})$ will be D_8. The point here is that D_8 capture some more abstract idea than just the symmetries of a square. The group D_8 represents "how" a system changes overtime. Since the action must be invertible, it must be that however it changes our system, we can go backwards. This naturally leads to the idea of *symmetry*. In other words, group actions can be thought of as finding the *symmetries* of an object.

What if there are other systems D_8 can act on? We need a general way of describing how D_8 would act on a system. We'll represent any system using a set. For example, for the rotations of a square, we'll represent each corner as an element in our set

here

Or, for a graph, we'll represent each vertex as an element of our set:

here

This system is *static*, as in we only captured one state of the object. We want to be able to capture the state of an object as it changes. To do that, we want to know where every element will "move to". For example, Given the corners of a square, if we apply r to it, then we want

$$1 \mapsto 2 \quad 2 \mapsto 3 \quad 3 \mapsto 4 \quad 4 \mapsto 1$$

This is actually a bijection relative to r: $\sigma_r : \{1, 2, 3, 4\} \to \{1, 2, 3, 4\}$. So, if we take $\mathrm{Sym}(\{1, 2, 3, 4\})$, we'll capture all possible ways of a system changing. Notice that this is exactly $S_{\{1,2,3,4\}} \cong S_4$. However, S_4 might be capture too many ways a system can change: $|D_8| = 8$ but $|S_4| = 4! = 24$.

As we've seen, S_4 would represent all possible ways of moving the vertices of a square, which is more than just rotating and reflecting a square. We might not want all possible ways of a system changing, just all possible way D_8 might change a system. To get to this limitation, we would want to make sure that we choose a subgroup of S_4 that only represents the rotations of a square. The most natural way of doing this is by having a homomorphism between D_8 and S_4:

$$\sigma : D_8 \to S_4$$

Let's see why this is the natural way of thinking about it: homomorphisms preserve group structure. If each D_8 maps to a single element of S_4, then we have injectively mapped D_8 into S_4. We say in ref:HERE that the image of a homomorphism is a subgroup of the codomain, so $\sigma(D_8) \leqslant S_4$. Furthermore, since σ is a homomorphism, it will make sure that the binary operation on D_8 is compatible with doing multiple "change of states" in our system:

$$\sigma(a \cdot b) = \sigma(a) \circ \sigma(b)$$

Thus, we can use homomorphisms to capture exactly the limitation we are looking for in our system.

Definition 1.5.1: Group Action

Let G be a group and let S be a set. Let $\mathrm{Sym}(S)$ represent the symmetric group given the set S (i.e. all possible bijections from S into itself). Then was call ρ a *group-action* from G into $\mathrm{Sym}(S)$

$$\rho : G \to \mathrm{Sym}(S)$$

if ρ is a homomorphism. Sometimes, the shorthand $G \curvearrowright S$ is used to represent a group G acting on a set S ($\curvearrowright$ is read as *acts on*).

This definition is useful for a high-level overview. However, it's sometimes useful to exactly the properties of σ has. (Words here in finding the axioms for he following definition)

Since ρ must be a homomorphism, then $\rho(a \cdot b) = \rho(a) \circ \rho(b)$. Thus, for any $x \in S$

$$\rho(a \cdot b)(x) = \rho(a)(\rho(b)(x))$$

As a corollary, we get that

$$\rho(e) = \rho(e \cdot e) = \rho(e) \circ \rho(e) \Rightarrow \rho(e) = \mathrm{id}_{\mathrm{Sym}(X)}$$

that is, the identity of G maps to the identity map on $\mathrm{Sym}(X)$, which leads us to saying that for any $x \in X$

$$\rho(e)(x) = x$$

With these, we can redefine a group-actions using only axioms

Definition 1.5.2: Group Actions

A group action of a group G on a set A is a map from $G \times A$ to A (written $\psi(g, a)$ or $g \cdot a$, for all $g \in G$ and $a \in A$). Satisfying the following properties:

1. $g_1 \cdot (g_2 \cdot a) = (g_1 g_2) \cdot a$ for all $g_1, g_2 \in G, a \in A$
2. $1 \cdot a = a, \qquad$ for all $a \in A$.

When unambiguous, we shall write $g \cdot a$ as ga.

We will show in details how these two axioms are the equivalent to the first definition in theorem ref:HERE. In the square example, A represents the set containing the 4 corners of the square, and G is D_8. Note that $g \cdot a$ is not a binary operation, and that the "typical" group action associated with our intuition of what D_8 represents is not the unique one, but the most common one.

Example 1.5.1

1. Let $G = D_8$ and V be the underlying set such that

   establish the limitations on the vertices

   Then $G \cdot V =$ __ defines a group action that "rotates" and "reflects" the square

2. Let $G = GL_n(\mathbb{R})$ act on $S = \mathbb{R}^n$ by $A \cdot \vec{x} = A\vec{x}$. This will satisfy the two axioms by the properties of matrix multiplication.

3. Let $G = S_n$ act on $S = \{1, 2, ..., n\}$ by $\sigma \cdot i = \sigma(i)$.

4. Take $G = \mathbb{R}$ and $S = \mathbb{R}^3$. Then
$$r \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} ra \\ rb \\ rc \end{pmatrix}$$
   defines a group action. This particular group action might be very familiar: it's like your typical vector space! All vector spaces are group-action, but they represent an action with a more advanced algebraic structure [a]. We will cover these in chapter ref:HERE.

5. (HERE) A Rubik's cube. Twisting the cube in different ways can be thought of as a group

6. If $\mathrm{Hom}_{\mathbf{Poly}}(\mathbb{R}, \mathbb{R})$ represents all polynomial equation, then we can define an action $\mathbb{R} \curvearrowright \mathrm{Hom}_{\mathbf{Poly}}(\mathbb{R}, \mathbb{R})$ by:
$$t \cdot f = f \circ x_t$$
   where $x_t(a) = a + t$. so $f(a)$ becomes $f(a + t)$. Check that this is a group-action TBD.

[a] They are an example of a ring-action on abelian groups, and a vector space is with a commutative division ring

Though this original intuition on group-actions is very motivating, in this book we'll mainly be applying group-actions on other groups or itself ($G \times G \to G$). For example, we can define:
$$g \cdot a = ga$$
where G is a group and the action is the group operation. We will do this action and
$$g \cdot a = gag^{-1}$$
to try and find insights on the structure of groups. These actions will turn out to be invaluable in producing many important results in group theory, as we'll see in chapter 4.

1.6 Cayley Graphs

Another way we try to find group-structure is by creating a graph using the generators of a group. This graph gives a visual on how the group elements interact with each other:

Definition 1.6.1: Cayley Graph

Let G be a group and let S be a generating set of G. The *Cayley Graph* $\Gamma = \Gamma(G, S)$ is a colored directed graph constructed as follows:

1. Each element $g \in G$ is assigned a *vertex*: the vertex set $V(\Gamma)$ of Γ bijectivity corresponds with G
2. Each generator $s \in S$ is assigned a colour c_s
3. For any $g \in G$ and $s \in S$, the vertices corresponding to the elements g and gs are joined by a direct edges of colour c_s. Thus, the edge set $E(\Gamma)$ consists of pairs of the form (g, gs), with $s \in S$ providing the colour

some popular, but not necessary, conventions are: [24]

1. The set S is finite
2. $S = S^{-1} = \left\{s^{-1} \middle| s \in S\right\}$
3. $e \notin S$ (i.e., the identity is not in S)

Example 1.6.1: Cayley Graphs

1. Let $G = \mathbb{Z} = \langle 1 \rangle$. Then we can say that $S \cup S = \{1, -1\}$. This gives us the graph:

$$\cdots \quad -3 \longleftarrow -2 \longleftarrow -1 \longleftarrow 0 \longrightarrow 1 \longrightarrow 2 \longrightarrow 3 \quad \cdots$$

   Take a moment to try and see what the Cayley graph of $\mathbb{Z} \times \mathbb{Z}$ would be (the answer is given in a later example)

2. Let $G = S_3 = \langle a, b | a^2 = b^3 = 1, ab = b^2 a \rangle$. Then if b is represented by red edges and a is represented by green vertices, we get:

$$\begin{tikzcd}
1 \arrow[r] \arrow[d, bend right] & b \arrow[r] \arrow[d, bend right] & b^2 \arrow[d, bend right] \arrow[dl, bend left] \arrow[dll, bend left] \\
a \arrow[r] \arrow[u, bend right] & ab \arrow[r] \arrow[u, bend right] & ab^2 \arrow[u, bend right]
\end{tikzcd}$$

3. Let $G = \mathbb{Z}/4\mathbb{Z}$. Then the Cayley graph is:

$$\begin{tikzcd}
0 \arrow[r] & 1 \arrow[r] & 2 \arrow[r] & 3 \arrow[lll, bend right]
\end{tikzcd}$$

[24] These conventions are usually followed in *geometric group theory*

4. Let $G = \mathbb{Z} \times \mathbb{Z}$. Then the Cayley graph is:

$$
\begin{tikzcd}
 & & & \vdots & & & \\
 & (-2,2) \arrow[r] & (-1,2) \arrow[l] & (0,2) \arrow[r] & (1,2) \arrow[r] & (2,2) & \\
 & (-2,1) \arrow[r] \arrow[u] & (-1,1) \arrow[l] \arrow[u] & (0,1) \arrow[r] \arrow[u] & (1,1) \arrow[r] \arrow[u] & (2,1) \arrow[u] & \\
\cdots & (-2,0) \arrow[r] \arrow[u] \arrow[d] & (-1,0) \arrow[l] \arrow[u] \arrow[d] & (0,0) \arrow[r] \arrow[u] \arrow[d] & (1,0) \arrow[r] \arrow[u] \arrow[d] & (2,0) \arrow[u] \arrow[d] & \cdots \\
 & (-2,-1) \arrow[r] \arrow[d] & (-1,-1) \arrow[l] \arrow[d] & (0,-1) \arrow[r] \arrow[d] & (1,-1) \arrow[r] \arrow[d] & (2,-1) \arrow[d] & \\
 & (-2,-2) \arrow[r] & (-1,-2) \arrow[l] & (0,-2) \arrow[r] & (1,-2) \arrow[r] & (2,-2) & \\
 & & & \vdots & & &
\end{tikzcd}
$$

if we let $S = \{\text{here}\}$, then the Cayley graph is

show that there is no twist

5. Let $G = D_\infty = \langle r, s \mid rs = sr^{-1} \rangle$ (i.e this is D_n but with no $r^n = 1$ relation) and let $S = \{r, s, r^{-1}, s^{-1}\}$. Then the Cayley graph is:

show the twist

6. $GL_2(\mathbb{Z}_2)$

Just by looking at the graphs, you might be able to see some interesting patterns. For example, by taking the direct product, we found it produces parallel graphs. We see in S_3 and $\mathbb{Z}/n\mathbb{Z}$ that arrow point *to* the identity represent the *order* of the element. Different Cayley graphs can also give different insights: notice that the Cayley graph for $\mathbb{Z} \times \mathbb{Z}$ with $S = \{\text{here}\}$ was simply two parallel lines with every node being connected with the one above it. On the other hand, the Caley graph for D_∞ looks like the one for $\mathbb{Z} \times \mathbb{Z}$, except it has a "twist" introduced to it[25].

In most of this textbook, Cayley graphs will not play a prominent role. They come back here and there to give a visual way of looking at new concepts introduced[26]. However, they are the foundation for an area of group theory called *geometric group theory*. For the study of geometric group theory, some knowledge of analysis is needed (ex. metrics). For those interested in such explorations, cite:HERE

[25] In chapter 5, we will learn about a way to construct D_∞ from $\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z}$, and will denote it as $\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$. Some mathematicians will call "$\rtimes$" the *crossed product* because it introduced a "cross" or "twist" in the Cayley graph

[26] For example, they will come back when working with free groups and semi-direct products

2

Sub Groups and Generating Groups

Now that we've studied some fundamental aspects of group theory and important examples, let's dive deeper in the mechanics used in analyzing and finding groups. The first way we'll be studying group is by trying to take smaller groups from the larger group and seeing if we can find information through the pieces of the group.

We will start by trying to find subgroups that exist for *any* group G. These subgroups are sometimes called *natural subgroups*, since they always exist for a group, as though the axioms of the group "naturally" let these subgroups exist. We will then move on to constructing groups given some subset of a group, and show how the basic operations of $\cup$ and $\cap$ interact with groups. Finishing off our exploration of generating groups, we will define the concept of "most general group"; it will turn out that the free groups can be thought of as the groups that "generate" all other possible groups. We will end this chapter by exploring a new way of visualizing the subgroups of finite groups through the use of a graph (in particular a lattice) that will become useful when trying to find how to construct larger groups.

2.1 Basic Definition and Properties

First, a reminder of what it means to take a smaller group from the same group:

Definition 2.1.1: Subgroup

Let G be a group with operation $*$. The subset H of G is a *subgroup* of G if H is a nonempty, closed under the restricted operation $*|_H$, that is, it is closed under products and inverse. We shall write $H \leqslant G$.

Recall that the binary operation $*|_H$ on H is a restriction of the binary operation $*$ on G. In particular, if $* : G \times G \to G$ is the binary operation on G, then we say we *restrict* the binary operation to H if we can define $*|_H : H \times H \to H$, where for any two elements $h_1, h_2 \in H$ $*|_H (h_1, h_2) = *(h_1, h_2)$. If $(H, *|_H)$ is a group, we say it's a subgroup of $(G, *)$. In this way, we say that H has "the same" binary operation as H. We will abuse notation and denote the binary operation of the subgroup $H \leqslant G$ as the same operation as the group G, that is, we will write $(G, \cdot)$ and $(H, \cdot)$ instead of $(G, \cdot)$ and $(H, \cdot|_H)$. If we have $H \leqslant G$ and $H \subsetneq G$, then we will denote this as $H < G$

Example 2.1.1

1. If $H = \{1\}$, then for any G, $H \leqslant G$. H is called the trivial subgroup. Similarly, by definition $G \leqslant G$ always.
2. $\mathbb{Z} \leqslant \mathbb{Q} \leqslant \mathbb{R}$ with the operation of addition.
3. $\mathbb{Z}^+ \nleqslant \mathbb{Z}$ under addition. Can you prove it?
4. Take the set of all symmetries ($\{1, \sigma\} \subseteq D_8$) and all rotations ($\{1, r, r^2, r^3\} \subseteq D_8$). These are both subgroups.

The subgroup H inherits a lot of properties we have proven in section 1.1.1. For example, for any equation of elements in H, it is also an equation of elements in G, and so will have the same solution(s) or lack thereof. Furthermore, the identity will remain the same, since every group must have a unique identity. Similarly for inverses.

For computational purposes, here is a useful proposition that warps up the conditions of a subgroup to be able to more quickly identify whether a subset is or is not a subgroup:

Proposition 2.1.2: The Subgroup Criterion

A subset H of G is a subgroup if and only if

1. $H \neq \emptyset$
2. for all $x, y \in H$, $xy^{-1} \in H$

Furthermore, if H is finite, then it suffices to check that H is non-empty and closed under its operation.

Proof **:**
Let $H \subseteq G$

($\Rightarrow$) Let $H \leqslant G$. Then $1 \in H$, so $H \neq \emptyset$. Take $x, y \in H$. If $y \in H$, then $y^{-1} \in H$. Since H is closed under its operation, then $xy^{-1} \in H$

($\Leftarrow$) Let $H \subseteq G$ and consider $(H, \cdot|_H)$.

(a) **(associativity)** Associativity comes for free since the binary operation on H is the same, so if $x, y, z \in H$, then $x(yz) = (xy)z$

(b) **(identity)** Since $H \neq \emptyset$, there must exist some $x \in H$. Then by the 2nd condition, $xx^{-1} = 1 \in H$, so $1 \in H$.

(c) **(inverse)** Let $x \in H$. Since $1 \in H$ then $1x^{-1} = x^{-1} \in H$.

(d) **(well-defined)** Let's make sure that if $x, y \in H$ then $xy \in H$ (making sure that the binary operation is still a function). By what we've just shown, $y^{-1} \in H$. Thus $x(y^{-1})^{-1} = xy \in H$ (where $(x^{-1})^{-1} = x$ was proved in proposition 1.1.8). Thus, $\cdot|_H$ is indeed a function

Thus, $H \leqslant G$

Now, let H be a finite group. Then for any $x \in H$, there can only be finitely many distinct elements $x_1, x_2, ..., x_n, ...$, meaning for some $a, b \in \mathbb{N}$, $x^a = x^b$ with $a < b$. This means that $x^{b-a} = 1$, and so for any $x \in H$, x has finite order. Thus, we get that $x^{-1} = x^{n-1}$, where $x^n = 1$. Thus, we automatically get H is a subgroup if H is closed under its binary operation

There are a lot of good exercises if you don't feel comfortable showing a particular subset is a subgroup or not on p.48

Exercise 2.1.1

1. Prove or disprove that the following are groups: (make sub question thing later): $\mathbb{Q}^\times \leqslant \mathbb{R}$, $D_3 \leqslant D_4$
2. Let $H, K \leqslant G$. Is $H \cup K$ always a group? Prove or give a counter-example. If not, then identify what condition(s) is needed for $H \cup K$ to be a group.
3. Prove that G cannot have a subgroup H with $|H| = n - 1$, where $n = |G| > 2$.
4. Let $K \leqslant H$ and $H \leqslant G$. Prove that $K \leqslant G$

2.2 Examples of Subgroups

In this section, we will explore "natural" subgroups as well as subgroups that arise given any homomorphism

2.2.1 Kernel and Image

Given any homomorphism $\varphi : G \to H$, are there subsets that can be created using φ which will be subgroups of either G or H? Naturally, the image of a homomorphism is a subgroup of H ($\varphi(G) \subseteq H$), essentially by definition. We might ask ourselves if there exists some subgroup of G that we can construct using φ?

commented old text here:

For the first subgroup, consider the following set

$$\ker(\varphi) \subseteq G \qquad \ker(\varphi) = \{g \in G \mid \varphi(g) = 1\}$$

then $\ker(\varphi) \leqslant G$

Proof. We will use the subgroup criterion. Take $a, b \in \ker(\varphi)$. Then using properties of homomorphisms:

$$1 = \varphi(ab^{-1}) = \varphi(a)\varphi(b^{-1}) = \varphi(a)\varphi(b)^{-1}$$

and since $\varphi(a) = \varphi(b) = 1$, we get

$$\varphi(a)\varphi(b)^{-1} = 11^{-1} = 11 = 1$$

so $ab^{-1} \in \ker \varphi$

□

the set $\ker(\varphi)$ will thus be given a name:

Definition 2.2.1: Kernel

Let $\varphi : G \to H$ be a homomorphism. Then the set

$$\ker(\varphi) = \{g \in G \mid \varphi(g) = 1\}$$

is the *kernel* of φ

8: Question to Ponder

Let's say $\varphi : G \to H$ and $\psi : G \to H$ are two homomorphisms with the same domain and codomain. If $\ker(\varphi) = \ker(\psi)$, does that imply $\varphi = \psi$?

As we have discussed earlier, $\varphi(G)$ is a group:

$$\varphi(G) \subseteq H, \quad \varphi(G) = \{h \in H \mid \varphi(g) = h \text{ for some } g \in G\}$$

For completeness of this section, we put the proof here:

Definition 2.2.2: Image

Let $\varphi : G \to H$ be a homomorphism. Then the set

$$\varphi(G) = \{h \in H \mid \varphi(g) = h \text{ for some } g \in G\}$$

or

$$\operatorname{im}(\varphi) = \{h \in H \mid \varphi(g) = h \text{ for some } g \in G\}$$

is the *image* of G with φ.

Proof. We will use the subgroup criterion as well as proposition 1.4.2. Since $\varphi(e_G) = e_H$, we have that $\varphi(G)$ is non-empty.

Next, take $a, b \in \varphi(G)$, meaning there exists an $x, y \in G$ such that $\varphi(x) = a$ and $\varphi(y) = b$. Since $\varphi(y^{-1}) = \varphi(y)^{-1}$, we have

$$ab^{-1} = \varphi(x)\varphi(y)^{-1} = \varphi(xy^{-1}) \in \varphi(G)$$

thus, $ab^{-1} \in \varphi(G)$, meaning $\varphi(G)$ passes the subgroup criterion, as we sought to show □

Given that $\varphi(G) \subseteq H$, $\varphi(G)$ has structure of H. However, since $\varphi(G)$ also has some of it's structural information based on G, it would not be surprising if we can somehow characterize $\varphi(G)$ through G. In fact, this will the first link done in the (probably now illusive) chapter 3.

Exercise 2.2.1

1. Do exercises with group actions showing that the kernel of the conjugation action is the centralizer

2.2.2 Center, Centralizers, and Normalizers

Given any group G, does there exist [non-trivial][1] subsets $H \subseteq G$ which will *always* be subgroups (i.e. $H \leqslant G$), regardless of the operation? It is clear that not all subsets will form subgroups. For example, take the set

$$T(G) = \{x \in G \mid x^n = 1, \text{ for some } n \in \mathbb{N}_{>0}\}$$

this set is called the *torsion subset* (i.e. the set of all element of finite order). If $T(G) \leqslant G$, then it must pass the subgroup criterion for all possible groups G. However, take

$$G = \langle x, y \mid x^2 = y^2 = 1 \rangle$$

both $x, y \in T(G)$. However, $xy^{-1} = xy \notin T(G)$ (note that $y^2 = 1 \Rightarrow y = y^{-1}$) – the element $(xy)^n = xyxyxyxy \cdots$ will never equal to 1, there are no relation that will permit the appropriate cancellation to occur. In particular, note that there is no relation for G which makes $xy = yx$ which would allow the values to commute and cancel. This is rather difficult to prove at this stage in the book; we will come back to this in section 2.3.3. Overall, $T(G) \nleqslant G$ for all groups G. Notice that if G was abelian then $T(G)$ *will* always be a subgroup (see exercise ref:HERE).

The question arises then if there are properties for which a subset will *always* be a subgroup for any group G regardless of the operation on G. These types of natural subgroups are in general quite hard to find, however they do exist – the rest of this section will be exploring examples of such subsets.

The first one we will explore is a great example of the power of commutativity: Let

$$Z(G) \subseteq G, \; Z(G) = \{g \in G \mid gx = xg \text{ for all } x \in G\}$$

then $Z(G) \leqslant G$

[1] the trivial subset being just the identity $\{e_G\}$, which is always the simplest subgroup which has essentially no "interesting" structure, and hence trivial

Proof. We will use the subgroup criterion. For this whole proof, let x be an arbitrary element of G. First note that $1x = x1 = x$, and so $1 \in Z(G)$. Next, take any $a, b \in Z(G)$. Since $bx = xb$, multiplying on the left and right by b^{-1} on both sides of the equation:

$$b^{-1}bxb^{-1} = b^{-1}xbb^{-1} \quad \Rightarrow \quad xb^{-1} = b^{-1}x$$

and since equality is symmetric, $b^{-1}x = xb^{-1}$, meaning $b^{-1} \in Z(G)$. Finally, we have

$$ab^{-1}x = axb^{-1} = xab^{-1}$$

meaning $ab^{-1} \in Z(G)$, proving it's a subgroup by the subgroup criterion □

This subgroup is special enough to be given a name:

Definition 2.2.3: Center

Let G be an arbitrary group. Define

$$Z(G) := \{g \in G | gx = xg \text{ for all } x \in G\}$$

to be the set of elements commuting with all the elements of G. This subset is called the *center* of G.

9: trivia

The letter Z is because the German word for center is *Zentrum*.

Example 2.2.1: Center of Groups

The provided examples should be verified by the reader as practice

1. If G is abelian, then $Z(G) = G$. Thus $Z(\mathbb{Z}) = \mathbb{Z}$, $Z(\mathbb{R}) = \mathbb{R}$, and so forth
2. $Z(D_4) = \{1, r^2\}$
3. $Z(Q_8) = \{1, -1\}$

The Center of other subgroups we've encountered in section 1.2 will be left as exercises.

This should start giving you an idea of the power of commutativity. In fact, it is even more powerful than we need it to be. The notion of the center can be weakened a bit and still remain a group. Take $A \subseteq G$ (notice that A is just a subset). Then define

$$C_G(A) \subseteq G, \; C_G(A) = \{g \in G \mid gag^{-1} = a \text{ for all } a \in A\}$$

Note that $gag^{-1} = a$ is equivalent to $ga = ag$. The change in equation is due to a further generalization we will shortly make. Before we get to that, we must first show that $C_G(A)$ is indeed a group

Proof. We will use the subgroup criterion. First, since for any $a \in A$, $1a = a1 = a$, $1 \in C_G(A)$

Next, take $x, y \in C_G(A)$. Since $y \in C_G(A)$, $yay^{-1} = a$ for all $a \in A$. Multiplying on the left by y^{-1} and right by y, we get $a = y^{-1}ay = (y^{-1})a(y^{-1})^{-1}$, showing that $y^{-1} \in C_G(A)$. Finally, to show $xy^{-1}a(xy^{-1})^{-1} = xy^{-1}ayx^{-1} = a$, we use associativity and the fact that $x, y, y^{-1} \in C_G(A)$:

$$\begin{aligned} xy^{-1}axy^{-1} &= xay^{-1}yyx^{-1} \\ &= xax^{-1} \\ &= a \end{aligned}$$

and so $xy^{-1} \in C_G(A)$, showing that $C_G(A) \leqslant G$ by the subgroup criterion □

Since it's a subgroup that will always exist, it is special enough to be given a name:

Definition 2.2.4: Centralizer

Let G be an arbitrary group, and $A \subseteq G$[a]. Define

$$C_G(A) := \{g \in G | gag^{-1} = a \text{ for all } a \in A\}$$

This subset of G is called the *centralizer* of A in G. Since $gag^{-1} = a$ if and only if $ga = ag$, $C_G(A)$ is the set of elements of G which commute with every element of A.
If A is a subgroup of G and $C_G(A) = G$, then A is called a *central subgroup*

[a] Notice A is a *subset*, not necessarily a subgroup

If we take, $A = G$ then $C_G(G) = Z(G)$. Equivalently, notice that:

$$\bigcap_{g \in G} C_G(g) = Z(G)$$

that is, it is the elements which are commutative with *all* elements (see exercise ref:HERE)

If $A \leqslant G$ such that $C_G(A) = G$, that means that every element of G commutes with A. It is left as an exercise (ref:HERE) to show that all central subgroups are contained in the center (hence the name central)

Example 2.2.2: Centralizer of Groups

Let $A = \{1, r, r^2, r^3\} \subseteq D_8$ and $B = \{1, i\} \subseteq Q_8$

1. $C_A(D_8) = A$. Notice how this set is larger than $Z(D_8)$, since $r^3 \in C_A(D_8)$ but $r^3 \notin Z(D_8)$
2. $C_B(Q_8) = \{1, -1, i, -i\}$. Similar to the previous example, notice that $C_B(Q_8)$ is larger than $Z(Q_8)$.

We can generalize this result even further: so far, we have written $ga = ag$ for all $g \in G$. we can re-write this as $g\{a\} = \{a\}g$, where we define $g\{a\} = \{gx \mid x \in \{a\}\}$ and $\{a\}g = \{xg \mid x \in \{a\}\}$ This might seem to be making the situation overly-complex, however the great advantage of this approach is we can now general to taking a general set A

$$gA = Ag \qquad gA = \{ga \mid a \in A\} \;\; Ag = \{ag \mid a \in G\}$$

What this difference does is that it weakens the condition for an element to "directly" commute ($ga = ag$) to "indirectly" commuting, or more precisely commuting with respect to the set A: for

every $a \in A$, there exists an $a' \in A$ such that $ga = a'g$. It will not always be the case that if $A \subseteq G$, then $gA = Ag$

Example 2.2.3: Counter-Example and Example
Let $A = \{r, \sigma\} \subseteq D_4$. Take $g = r$. Then:

$$rA = \{r^2, r\sigma\} \qquad Ar = \{r^2, \sigma r\} = \{r^2, r^3\sigma\}$$

notice that $rA \neq Ar$
However, if $A = \{1, r, r^2, r^3\}$ then for any $g \in D_4$, $gA = Ag$. As an example, let $g = \sigma$. Then

$$\sigma A = \{\sigma, \sigma r, \sigma r^2, \sigma r^3\} \qquad A\sigma = \{\sigma, r\sigma, r^2\sigma, r^3\sigma\} = \{\sigma, \sigma r, \sigma r^2, \sigma r^3\}$$

verifying other values of D_4 by hand shows that $gA = Ag$ for all $g \in D_4$

Therefore, we can define the following set of elements that satisfy $gA = Ag$:

$$N_G(A) \subseteq G,\ N_G(A) = \left\{g \in G \;\middle|\; gAg^{-1} = A \text{ for all } g \in G\right\}$$

Again, $N_G(A)$ is a subgroup of G. The proof is almost identical to the proof for $C_G(A)$ (change a for A in the proof). We therefore give a name to this special subgroup:

Definition 2.2.5: Normalizer

Take $gAg^{-1} = \left\{gag^{-1} | a \in A\right\}$. Define the *normalizer* of A in G to be the set

$$N_G(A) := \left\{g \in G | gAg^{-1} = A \text{ for all } g \in G\right\}$$

If A is a group and $N_G(A) = G$ (that is, $gAg^{-1} = A$ for all $g \in G$) we call A a *normal subgroup* of G

Notice that if we have $g \in C_G(A)$, then $gag^{-1} = a$ with $a \in A$, which means that $gAg^{-1} = A$, and so $g \in N_G(A)$. It follows quickly that $C_G(A) \leqslant N_G(A)$ (see exercise ref:HERE)

Example 2.2.4: Normalizer of Groups
Let $A = \{1, r, r^2, r^3\} \subseteq D_8$ and $B = \{1, i\} \subseteq Q_8$

1. $N_A(D_8) = D_8$. Notice how $Z(G) \subseteq C_A(D_8) \subseteq N_A(D_8)$, that is

   $$\{1, r^2\} \subseteq \{1, r, r^2, r^3\} \subseteq D_8$$

   Since A is also a subgroup of G, A is also an example of a normal subgroup of D_8

2. $N_B(Q_8) = \{1, -1, i, -i\}$ First, it is clear that $1 \in N_B(Q_8)$. To show that $i \in N_B(Q_8)$, notice that

   $$iB = \{i, -1\} = Bi$$

   so $i \in N_B(Q_8)$. Similarly, for -1 and $-i$. For $j, k \in Q_8$, we know that $ij = k$, $ji = -k$ and $ik = j$, $ki = -j$ (see section 1.2.4).

   $$jB = \{j, k\} \neq \{j, -k\} = Bj \qquad kB = \{k, j\} \neq \{k, -j\} = Bk$$

meaning $j, k \notin N_B(Q_8)$. Similarly for $-j, -k$. Thus, we are left with $N_B(Q_8) = \{1, -1, i, -i\}$. Notice that $C_B(Q_8) = N_B(Q_8)$, that is, the normalizer doesn't necessarily have to be bigger than the centralizer

3. If $D = \{1, (1\ 2)\} \subseteq S_3$, what is $N_D(S_3)$? What about $C_D(S_3)$?

The notion of the normalizer is not simply generalization for the sake of generalization, it in fact comes up as a very important property: the kernel of a homomorphism has is always normal!

Proposition 2.2.6: Kernels Are Normal Subgroups

Let $\varphi : G \to H$ be a homomorphism, and take $\ker(\varphi) \leqslant G$. Then $\ker(\varphi)$ is a normal subgroup, that is

$$N_G(\ker(\varphi)) = G$$

In other words, for any $g \in G$, $g(\ker(\varphi)) = (\ker(\varphi))g$

Proof **:**
Let $\varphi : G \to H$, and for notational convenience, let $K = \ker(\varphi)$.

Take any $g \in G$, and $k \in K$, that is, $k \in G$ such that $\varphi(k) = e_H$. We want to show that $gK = Kg$, or equivalently $gKg^{-1} = K$, for all $g \in G$. Take $gkg^{-1} \in gKg^{-1}$. Then

$$\varphi(gkg^{-1}) = \varphi(g)\varphi(k)\varphi(g^{-1}) = \varphi(g)1\varphi(g^{-1}) = \varphi(g)\varphi(g^{-1}) = \varphi(gg^{-1}) = e_H$$

thus, $gkg^{-1} \in K$, so for some k', $k' = gkg^{-1}$. Since this is true for arbitrary $k \in K$, we get $gKg^{-1} = K$, as we sought to show

The fact that kernels are normal subgroups will become very important in chapter 3 This brings us to the end of our generalizations. This section will be closed with a quick remark about normal subgroups. We did not talk about them much here. However, they are at the heart of chapter 3. For now, you can keep them in the back of your mind.

Exercise 2.2.2

1. $C_G(Z(G)) = G$ and deduce that $N_G(Z(G)) = G$
2. If $A, B \subseteq G$, and $A \subseteq B$, then $C_G(B) \leqslant C_G(A)$
3. For each S_3, D_8, Q_8, compute the centralizers of each element and find the center of each group. Does Lagrange's Theorem (ex 19 section 1.7) simplify the work?

   These next 3 subgroups from the next 3 questions will come back when we talk about group actions (and the conjugacy action, section 4.4.2)

4. Find the centre of $GL_3(\mathbb{C})$
5. Find the centralizer of $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$ in $GL(3, \mathbb{C})$

6. Find the centralizer of $\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}$ in $SL(5, \mathbb{C})$, the 5×5 complex matrices having determinant equal to 1
7. there are a lot of good exercises in the book
8. . (re-work exercise since $T(G)$ given as an example of why $T(G)$ not considered *natural* torsion group) Let G be an abelian group. Take $T(G) = \{x \in G \mid \exists n \in \mathbb{N} \text{ such that } x^n = 1\}$. This group is called the *torsion subgroup* of G. Show that $T(G)$ is indeed a group. is $T(G)$ a group is G is not abelian?
9. prove that $\bigcap_{g \in G} C_G(g) = Z(G)$
10. Let $A \leqslant G$ such that $C_G(A) = G$. Prove that $A \leqslant Z(G)$
11. (advanced, maybe move to another section?) Show that $C_G(S) \leqslant N_G(S)$ by showing $C_G(S)$ is the kernel of the homomorphsims $N_G(S) \to \mathrm{Sym}(S)$ by $g \mapsto (x \mapsto gxg^{-1})$ (i.e. $T(g)(x) = T_g(x) = gxg^{-1}$)
12. Maybe do a Galois correspondence question? this wiki-entry made me think of it: here
13. For any group G, and any subset $A \subseteq G$, is it true that $Z(G) \leqslant C_G(A)$?

2.3 Subgroups Generated by Subsets

All groups have an underlying set. We can thus ask ourselves many questions about groups related to it's underlying set. For example, If we have two subgroups $H, K \leqslant G$, is $H \cap K$ or $H \cup K$ also a subgroup? What if we extend to finite or infinite intersection or union? On another hand, we can ask ourselves if there exists subsets $A \subseteq G$ that can in some way generate G using the binary operation. If generating all of G is too strong of a condition, can A always generate a subgroup of G? These are the types of questions that will be asked in this section.

Note that the technics in this section are very common throughout algebra (and many other areas of mathematics). Asking whether a subset of the underlying of an algebraic object "generates" the entire object (or part of the object) will consume our time in many of the following chapters of this book (especially chapter 13).

2.3.1 Cyclic Groups and Cyclic subgroups

We will start our exploration subsets $A \subseteq G$ which are singletons: $|A| = 1$. The first question we will ask ourselves is if there is a subgroup $H \leqslant G$ which contains A. The answer to this is always yes: $G \leqslant G$ is a subgroup and by definition $A \subseteq G$. Though this is a valid answer, this is a very boring answer (in mathematical lingo, we'd say it is the trivial case). Thus, we look for a more interesting question: what is the *smallest* subgroup of G that contains A[2]. If $H \leqslant G$ is the smallest subgroup

[2] the notion of "smallest" is well-defined if $|G|$ is finite, but get's a bit tricker if $|G|$ is infinite (ex. $2\mathbb{Z} \trianglelefteq \mathbb{Z}$ but $|2\mathbb{Z}| = |\mathbb{Z}|$. We will show in the section 2.3.2 how to be more precise with the notion of smallest

containing A, H is called a *cyclic subgroup*. If the entire group G is the smallest subgroup containing A, then G is called a *cyclic group*. To find the cyclic subgroup or groups given a subset $\{a\} = A$, a similar strategy will be used from when the order of an element (definition 1.2.6) was discussed. Cyclic groups they pop-up quite often due to their simplicity.

Definition 2.3.1: Cyclic groups

Let G be a group. The group G is *cyclic* if G can be generated by a single element, i.e., there is some element $x \in G$ such that $G = \{x^n | n \in \mathbb{Z}\}$ (where the operation is usually denoted as multiplication)

Sometimes, the notation C_n used to represent arbitrary cyclic groups of order n (with multiplication notations)

We say G is generated by x (or x generated G), and write $G = \langle x \rangle$. It was proven in proposition 1.2.5 that $\langle x \rangle$ is indeed a group. The existence of a C_n for every n is because there exists the cyclic group $\mathbb{Z}/n\mathbb{Z}$ for every n (or $\langle r_n \rangle$ for $r \in D_n$). Later in this section, we will show that $\mathbb{Z}/n\mathbb{Z}$ is the *only* cyclic group of order n up to isomorphism, meaning we will be able to conclude that $C_n \cong \mathbb{Z}/n\mathbb{Z} \cong \langle r_n \rangle$. Since C_n usually has multiplication notation, for this section we will use 1 to represent the identity.

Note: The fact that $\langle x \rangle$ is the smallest subgroup containing x forced by construction: if there was a subgroup $H \leqslant G$ such that $H \subsetneq \langle x \rangle$, then there would be number n such that $x^n \in \langle x \rangle$ but $x^n \notin H$. Do you see how this contradicts that H is a subgroup?

Example 2.3.1

1. As mentioned earlier, the usual candidate for this example if $\mathbb{Z}/n\mathbb{Z}$, that is, addition modular n, which can be generated by $\overline{1}$. So $\langle \overline{1} \rangle = \mathbb{Z}/n\mathbb{Z}$. Notice there is some ambiguity here: How can you know if $\overline{1}$ represents the of order $5, 10, 324$? In general, you can't. It is up using the context to determine the order. If the context requires clarity, a subscript will be added: $\langle \overline{1}_n \rangle = \mathbb{Z}/n\mathbb{Z}$ or $\langle [1]_n \rangle = \mathbb{Z}/n\mathbb{Z}$
2. D_n is *not* cyclic. Try finding a single element in D_n which generates all of D_n
3. Is Q_8 cyclic? What about $\mathbb{Z}$? What about $\mathbb{Z} \times \mathbb{Z}$?

Proposition 2.3.2: $|C_n| = |x|$

If $C_n = \langle x \rangle$, then $|C_n| = |x|$ (where if one side of this equality is infinite so is the other).

***Proof* :**
The idea is to show that that if $|x| < \infty$, then if two elements differ in exponent and are smaller than the group, $a \neq b < n$, then $x^a \neq x^b$, which can be shown by contradiction. For the case where the exponent is greater than or equal to n, $t \geq n$, x^t, such an elements can always equal to an element of order less than n. In particular, if $t \geq n$, then $x^t = x^{t-kn}$ for appropriate k for which

$0 \leq t - kn < n$. Thus, in the $t \geq n$ case, the proof comes down to the first case that was proved. For the $|x| = \infty$ case, assume $x^a = x^b$, $a \neq b$, and proceed by contradiction (i.e. $x^{b-a} = 1$).

Note: This does not imply that for *any* $x \in C_n$, $|x| = |C_n|$. For example, $\mathbb{Z}/6\mathbb{Z}$ is cyclic since $\mathbb{Z}/6\mathbb{Z} = \langle \overline{1} \rangle$. However, $|\overline{2}| = 3$, meaning $\overline{2}$ cannot generate $\mathbb{Z}/6\mathbb{Z}$. On the other hand, $\overline{1}$ is not the only element of order 6 in $\mathbb{Z}/6\mathbb{Z}$, can you think of a second? Is there more than two?

Corollary 2.3.3: Cyclic $\Rightarrow$ Abelian

If G is cyclic, then G is abelian

Proof **:**
Let G be a cyclic group. Take some $g \in G$ that generates it. Take some $a, b \in G$ so $g^n = a$ and $g^m = b$ then

$$ab = g^n g^m = g^{n+m} = g^{m+n} = g^m g^n = ba$$

thus, G is also abelian.

You might not be satisfied with this proof: why in the exponent can we commute n and m? The answer comes the associativity of groups. Since

$$a(aa) = (aa)a$$

it makes no difference to write

$$a^{1+2} \qquad a^{2+1}$$

or more generally $a^{n+m} = a^{m+n}$ due to proposition 1.1.8[5].

Since every element in a cyclic is determined by the exponent, having some information about what power of the generator leads to the identity can give us some information on smaller powers of the generator that also lead to the identity.

Proposition 2.3.4: Powers of an Element

Let G be a group, $x \in G$ and let $m, n \in \mathbb{Z}$. If $x^n = 1$ and $x^m = 1$, then $x^d = 1$, where $\gcd(m, n) = d$. In particular, if $x^m = 1$ for some $m \in \mathbb{Z}$, then $|x|$ divides m, $|x| \big| m$

Proof **:**
This proof relies on Euclid Algorithm (see section 0.2. Let $d = \gcd(a, b)$, so that $m = dm'$ and $n = dn'$. Since $x^m = x^n = 1$, then we have $x^{dm'} = x^{dn'} = 1$. Thus, we have

$$1 = x^{dm'} = \underbrace{x^d x^d \cdots x^d}_{m' \text{ times}}$$

$x^d = 1$ or $x^d = k$ $(k \neq 1)$ so $k^{m'} = 1$. If $x^d = 1$, then we're done, so let's say $x^d = k$. Then,

similarly for $n = dn'$:

$$1 = x^{dn'} = \underbrace{x^d x^d \cdots x^d}_{n' \text{ times}}$$

where $x^d = 1$ or $x^d = k$ (same k) so $k^{n'} = 1$. Thus, we get $k^{n'} = k^{m'} = 1$. However, here lies the contradiction. Without loss of generality, let's say $n' < m'$ (if $n' > m'$, simply switch the order of the elements in the proceeding proof) so that $m' = n' + \ell$. Then since $\gcd(n', m') = 1$, $k^{m'} = k^{n'+\ell} = k^{n'} k^{\ell} = k^{\ell} \neq 1$. If it did, then m' would be a multiple of n', but then $\gcd(m', n') \neq 1$

Thus, $x^{\gcd(n,m)} = 1$, as we sought to show

Corollary 2.3.5

Let G be a group (not necessarily cyclic), let $x \in G$, and let $a \in \mathbb{Z} - \{0\}$.

1. if $|x| = \infty$, then $|x^a| = \infty$
2. if $|x| = n < \infty$, then $|x^a| = \frac{n}{\gcd(n,a)}$. If a divides n, then $|x^a| = \frac{n}{a}$.

Proof **:**

1. If $|x^a| < \infty$, then there would exist an n such that $(x^a)^n = x^{an} = 1$. (if an is negative, take $-an$). But then x would have finite order, a contradiction
2. Since $x^n = 1$, $1 = 1^a = (x^n)^a = x^{na}$. By proposition 2.3.4, $x^{gcd(a,an)} = 1$

With this idea in mind, let's do what we did before and try to see if we can classify cyclic groups.

For cyclic groups, there's in fact an incredibly simple isomorphism which will be introduced showing that if two cyclic groups have the cardinality (same order), they are in fact isomorphic, and thus can be treated as interchangeable! This means that when dealing with future theorems about cyclic groups, once we found it applies to a cyclic group of a certain order, we know it applies to any cyclic group of the same order! This also means there are only countably many finite cyclic groups, each with a predictable structure.

Theorem 2.3.6: Isomorphism between same-order Cyclic Groups

Any two cyclic groups of the same order are isomorphic. More specifically,

1. if $n \in \mathbb{Z}^+$ and $\langle x \rangle$ and $\langle y \rangle$ are both cyclic groups of order n, then the map

$$\psi : \langle x \rangle \to \langle y \rangle \qquad x^k \mapsto y^k$$

   is well defined and is an isomorphism

2. if $\langle x \rangle$ is an infinite cyclic group, the map

$$\psi : \mathbb{Z} \to \langle x \rangle \qquad k \mapsto x^k$$

   is well defined and is an isomorphism

This proof is not too difficult and so a guide has being laid out for the reader to deduce the proof

***Proof* :**
First, show the function is well defined, that is, if $x^a = x^b$ (for possibly different a and b), then $y^a = y^b$. This uses proposition 2.3.4. Then show the map is a homomorphism. To prove it's an isomorphism, you can either

1. Show there exists a homomorphism $\varphi : \langle y \rangle \to \langle x \rangle$ such that $\varphi \circ \psi = \psi \circ \varphi = \mathrm{id}$ (technically, $\mathrm{id}_{\langle x \rangle}$ or $\mathrm{id}_{\langle y \rangle}$ respectively)
2. Show ψ is injective. At which point subjectivity comes from the fact that both groups have the same cardinality and are finite.
3. Show ψ is surjective. At which point injectivity comes from the fact that both groups have the same cardinality and are finite.

For the infinite case, the map is "clearly" well-defined: if $a = b$, then $x^a = x^b [= x^a]$. By the exponent laws, this map is homomorphic. Then to prove this map is an isomorphism, you can either:

1. Show there exists a homomorphism $\varphi : \langle x \rangle \to \mathbb{Z}$ such that $\varphi \circ \psi = \psi \circ \varphi = \mathrm{id}$ (technically, $\mathrm{id}_{\mathbb{Z}}$ or $\mathrm{id}_{\langle x \rangle}$ respectively)
2. Show the map is injective and surjective (not that injectivity (resp. surjectivity) no longer implies surjectivity (resp injectivity) in the infinite case, even if the cardinalities are the same. Can you think of a surjective but not injective map from $\mathbb{Z}$ to $\mathbb{Z}$? An injective but not surjective map from $\mathbb{Z}$ to $\mathbb{Z}$? see exercise ref:HERE (in Cardinality section))

This let's us properly conclude what was stated at the beginning of this section: if G is s cyclic group of order n (ex. $\mathbb{Z}/n\mathbb{Z}$ or $\langle r \rangle$, $r \in D_n$) then

$$\mathbb{Z}/n\mathbb{Z} \cong \langle r \rangle$$

Hence, instead of referencing a specific cyclic group, it is sometimes useful to just write C_n (some authors like Z_n) to represent some arbitrary cyclic group of order n, and use $\mathbb{Z}$ for an infinite cyclic group (instead of C_∞ or Z_∞)

Now that we classified all cyclic groups, we can ask ourselves the question of which elements generate one of these groups. this next proposition leads us to the answer

Now that we know how many cyclic sub-groups there are, and how these groups can be generated, we can explore what kind of sub-groups we get from cyclic groups. Quite naturally, every subgroup of a cyclic sub group is cyclic. It will turn out that if you have a cyclic group of infinite order, it has an infinite number of cyclic subgroups, and if its of finite order, lets say n, then every sub groups is of an order that divides n. That is , if $H \leqslant G$, G is cyclic, $|H| = a$, then $a|n$.

Proposition 2.3.7: x relation to cyclic group H

Let $H = \langle x \rangle$.

1. Assume $|x| = \infty$. Then $H = \langle x^a \rangle$ if and only if $a = \pm 1$.
2. Assume $|x| = n < \infty$. Then $H = \langle x^a \rangle$ if and only if $(a, n) = 1$. In particular, the number of generators of H is $\psi(n)$, where ψ is Euler's totient function. (which gives the number of relatively prime numbers to it).

Proof. Dummit p.57-58 □

Finally, we'll put together all our knowledge into one theorem for easy reference:

Theorem 2.3.8: elements that generate cyclic subgroup

Let $H = \langle x \rangle$ be a cyclic group

1. Every subgroup of H is cyclic. More precisely, if $K \leqslant H$, then either $K = \{1\}$ or $K = \langle x^d \rangle$, where d is the smallest positive integer such that $x^d \in K$
2. If $|H| = \infty$, then for any distinct non negative integers a, b, $\langle x^a \rangle \neq \langle x^b \rangle$. Furthermore, for every integer m, $\langle x^m \rangle = \langle x^{|m|} \rangle$, where $|m|$ denotes the absolute value of m, so that the nontrivial subgroups of h correspond bijectivity with the integers $1, 2, 3, ...$
3. If $|H| = n < \infty$, then for each positive integer a dividing n, there is a unique subgroup of H of order a. This subgroup is the cyclic group $\langle x^d \rangle$, where $d = \frac{n}{a}$. Furthermore, for every integer m, $\langle x^m \rangle = \langle x^{gcd(n,m)} \rangle$, so that the subgroups of H correspond bijectivity with the positive divisors of n.

intuition

Proof **:**
This proof involves a lot of counting.

1. exercise
2. exercise

Also, once you have Lagrange's Theorem (3.3.1), this proof will become much easier.

Note that the finite group examples of theorem 2.3.8 will eventually easily generalize to perhaps the most use theorem of the course: Lagrange Theorem (theorem 3.3.1), which sates that the order of any subgroup divides the order of a group ($H \leqslant G \Rightarrow |H| \big| |G|$).

The next theorem is in a sense a generalization of part (3) of 2.3.8, since it uses this fact. Once I proof Lagrange, I'll reference this theorem to make sure this makes sense. I do want to introduce this fact here because it feels quite appropriate as a generalization of the previous theorem:

Corollary 2.3.9: if G has order p, then it is cyclic

If G is a group of prime order p, then G is cyclic, hence $G \cong Z_p$.

***Proof* :**
First show that G is cyclic. Let $x \in G$. since $|G| = p$, then $|G| > 1$ so let $x \neq e$. We can always take $\langle x \rangle \subseteq G$. Then by Lagrange theorem (generalization of part 3 of theorem 2.3.8), we know that $\langle x \rangle \big| |G|$, However, $|G|$ is prime, so only 1 and G divide it. Since $x \neq e$, then $|\langle x \rangle| \neq 1$, so $|\langle x \rangle| = p$, meaning that G was generated by a single element, and so is cyclic. By corollary 2.3.3, it is abelian.

To show its isomorphic to $\mathbb{Z}_p$ should now be easy.

Note that this doesn't mean that if $H \leqslant G$ where $|H| = p$ for some prime that $H \in Z(G)$!

Example 2.3.2: : $H \notin Z(G)$
Take D_8. Note that $\langle \sigma \rangle \leqslant D_8$ and $|\langle \sigma \rangle| = 2$, and so it's cyclic and abelian. However. $r\sigma = \sigma r^3$ meaning that $\sigma \notin Z(D_8)$. However, it is in the centralizer $C_{D_8}(\sigma) = \{g \in G | g\sigma = \sigma g\}$ for $\sigma\sigma = e = \sigma\sigma$, which can also be "obvious" since $\langle \sigma \rangle$ is abelian, and so it would be in the centralizer.

The intuition behind this is that for something to be in the center, it must commute with *all* of G. It is possible to take a subgroup that only commutes (as we've seen with the idea of the centralizer). This is why the intersection under definition 2.2.3 is true: it must commute with everything.

Note: You cannot take $\langle x \rangle \leqslant G$ and do $G - \langle x \rangle$. G is not necessarily a group. Take D_8, and take $K = \langle \sigma \rangle$. Then consider $H = D_8 - K$. Then $r^3 \cdot r\sigma = \sigma$, showing that H is not closed under multiplication. Thus, subtraction of groups is not well-defined. However, it will turn out that under the right circumstances, *dividing* and *multiplying* groups will be well-defined! This will be the subject of the next chapter (3)

interesting exercise

1. number 9
2. number 13
3. number 16
4. Knowing Cyclic groups, we can look at another great example of a homomorphism: $\psi : \mathbb{Z} \to Z_n$, where $Z_n = \langle x \rangle$,
$$\psi(a) = x^a$$
Prove that it's a homomorphism, then figure out what elements of $\mathbb{Z}$ map to an element of Z_n.
5. the product of two cyclic subgroups is not necessarily cyclic!

2.3.2 Generating Subgroups

We now turn our attention to generating groups or subgroups using more than 1 element. When studying cyclic groups, we asked whether given an element $x \in G$, does there exists a smallest group $H \leqslant G$ such that $x \in H$. It was shown that taking the closure of x by taking all powers of x resulted in such a group. Phrased differently, the subgroup $\langle x \rangle$ is the smallest subgroup of G that contains the set $\{x\}$, i.e., if $H \leqslant G$ and $x \in H$, then $\langle x \rangle \subseteq H$ (in fact $\langle x \rangle \leqslant H$). We now replace the singleton $\{x\}$ with some arbitrary subset $A \subseteq G$.

As mentioned in the beginning of this chapter, generating subgroups (or minimal subgroups) from a group is not unique to Group Theory. In the following chapters studying other algebraic structures (ex. ideals, modules, vector-spaces, algebras, fields, etc.), this theme of trying to find subobjects or minimal subobjects that contain a subset A from the original object will repeat itself often, and so will the general method for finding these subobject or minimal subobject. To find a minimal subobject there are two possible general approach: If A is a subset of the object, take all possible subobjects containing A and intersect them. In the case of groups, if $A \subseteq G$, then take all possible subgroup H_A containing A and intersect them. Another way of finding the minimal subobject is instead by approaching it from the "other direction": If A is a subset of the object, then let $\overline{A}$ to be the set of all possible combinations of elements of A using the binary operation of G. In the case of groups, of $A \subseteq G$, then $\overline{A}$ will be all the possible combination of elements of A. The main result of this section is that these two methods will actually coincide and form the same subgroup. In proceeding chapters, some details will be omitted due to them being covered in this section, so it is advised that this section be well understood.

We start with the first method. We need show that the intersection of still forms a group as just stated

Proposition 2.3.10: intersection of Groups

Let G be a group and let $\mathcal{H}$ be any collection of subgroups of G. Then

$$\bigcap_{H \in \mathcal{H}} H$$

Is a subgroup of G.

Proof **:**
Let $K = \bigcap_{H \in \mathcal{H}} H$, and take $a, b \in K$. Since $b \in K$, b is in every subgroup H in $\mathcal{H}$. Thus, b^{-1} is in every subgroup H in $\mathcal{H}$, and so $b^{-1} \in K$. Similarly, ab^{-1} would be in every subgroup H in $\mathcal{H}$ (since each subgroup passes the subgroup test), so $ab^{-1} \in K$. Thus, by the subgroup test, K is a subgroup of G

Therefore, the following concept can be defined:

Definition 2.3.11: Closure

let G be a group and $A \subseteq G$. Then

$$\langle A \rangle = \bigcap_{\substack{H \subseteq G \\ A \subseteq H}} H$$

is called the *closure of* A, or the subgroup *generated by* A. If A is a finite set $\{a_1, a_2, ..., a_n\}$, we usually write $\langle a_1, a_2, ..., a_n \rangle$, and write $\langle A, B \rangle := \langle A \cup B \rangle$ for the union of two sets.

We have already been using this notation when $A = \{x\}$. Notice that $\langle A \rangle$ must be the smallest subgroup containing A. If $H' \leqslant G$ and $A \subseteq H'$, then by definition H' is one of the subgroup that were intersected in the aforementioned definition, since $\langle A \rangle$ is equal to the intersection, $\langle A \rangle \subseteq H'$. This then indeed capture the idea of minimal subgroup. Furthermore, $\langle A \rangle$ always exists, since at minimum $A \subseteq G$ (since $G \leqslant G$), so G is part of the sets that will be intersection. If it was not specified that there always exists a subgroup that contains A, then it is possible that $\langle A \rangle = \emptyset$ (i.e. $\langle A \rangle$ might not exist). Finally, $\langle A \rangle$ is the unique smallest subgroup containing A (similarly to how $\langle x \rangle$ is unique in the sense that all cyclic subgroups of the same order are isomorphic to it). If there exists another group K' that was also the smallest subgroup containing A, then K' would be in the intersection above, so $\langle A \rangle \subseteq K'$. But since K' is supposed to be the smallest subgroup containing A, it is also the case that $K' \subseteq \langle A \rangle$. Therefore $\langle A \rangle = K'$, showing uniqueness as well. Thus, $\langle A \rangle$ always exists, and is the smallest unique subgroup of G containing A.

The limit with this approach is that there isn't much of an idea of what an element in $\langle A \rangle$ would look like, we simply know of it's existence. The second approach to finding the minimal object will give us a better insight into this problem by changing perspective from this high-level and starting with the elements of A and trying to "generate" a subgroup containing A. To that end, define the set

$$\overline{A} := \{a_1^{\sigma_1} a_2^{\sigma_2} \cdots a_n^{\sigma_n} \mid n \in \mathbb{Z} \text{ and } a_i \in A,\ \sigma_i = \pm 1, 1 \leq i \leq n\}$$

where if $A = \emptyset$ then $\overline{A} = \{1\}$. The set $\overline{A}$ contains all possible finite products of the elements from A. Note that the a_i's need not be distinct, as the element a^2 for $a \in A$ would be written $a^1 a^1$. It is also easy to verify that this set is indeed a subgroup of G using the subgroup criterion (consider that n can vary in any possible length). The interesting question is whether this set is also the smallest minimal subgroup containing A, and indeed it is:

Proposition 2.3.12

Let G be a group and $A \subseteq G$ be a subset of G. Then

$$\langle A \rangle = \overline{A}$$

Proof **:**
Will do at a later time.. p. 63 in Dummits book.

From now on, the notation $\langle A \rangle$ will be used to represent the subgroup generated by A.

If G is abelian, then it is much easier to find a subset that generates G. (show that if each $a_i \in A$ of a generating set has finite order, G must have finite order). (Point out that it is possible that

$a\alpha b^\beta = a^\gamma b\delta$ is possible even if $a\alpha \neq a^\gamma$ and $b^\beta \neq b^\delta$). (We'll show in section 5.1.5 that if G is finitely generated (i.e. $\langle A\rangle = G$ and $|A|$ is finite), we can find all possible ways we can generate G).

If G is not abelian, then as mentioned in remark (ref:HERE word problem), it is exceedingly difficult to determine the structure of G. As an illustrative example, consider:

$$G = \langle a, b \mid a^2 = b^2 = 1\rangle \tag{2.1}$$

Notably, $ab \neq ba$, $|a| = 2$ and $|b| = 2$. Then we see that:

$$abababa....$$

is a new element every time we alternate between a and b! Thus, even though G in equation 2.1 defined in a relatively simple way, with two elements that each have order 2, it still contains an infinite number of elements.

Generally, there can be many sets that generate a group: for example $\langle r, \sigma\rangle = \langle r\sigma, \sigma\rangle = D_n$, or $\langle 2, 3\rangle = \langle 7, 5\rangle = \mathbb{Z}/10\mathbb{Z}$. Notice that no proper subset of $\{2, 3\}$ generates $\mathbb{Z}/10\mathbb{Z}$, however there do exist smaller subsets of $\mathbb{Z}/10\mathbb{Z}$ that generate $\mathbb{Z}/10\mathbb{Z}$ (namely, $\{1\}$). Thus, the cardinality of the minimal generating subset cannot be determined by simply finding a minimal generating subset: it would have to be $\min\{|A| \mid A \subseteq G,\ \langle A\rangle = G\}$[3]. Notice too that if $n = 4$ then $a = \sigma$ and $b = r\sigma$ have order 2, while D_4 has order 8, showing how it is not always the case that all elements of D_4 will be written of the form $a^\alpha b^\beta$ $(\alpha, \beta \in \mathbb{Z})$ like in the abelian case.

(there are actually some cool observations D&F make here I want to ponder over later)

(If we just start with a set A that is not a subset of a group, then we can further expand this notation if we put relations on A. When explaining/defining a relation (for some subset $A \subseteq G$, $B \subseteq G$, $a_1a_2a_3\cdots a_n = b_1b_2b_3\cdots b_m$), we can always re-write it as $b_m^{-1}b_{m-1}^{-1}\cdots b_2^{-1}b_1^{-1}a_1a_2\cdots a_{n-1}a_n = 1$. In this way, any set of generators of the form $w_1 = w_1', w_2 = w_2', ...w_k = w_k'$ can be transformed to $v_1 = (w_1')^{-1}w_1$ so that we then say $v_1 = v_2 = \cdots = v_k = 1$. Mention what are the limits of what type of relation we can put on are explored in the next section)

Cover the concept of finite generation. Ask The question

10: Food for Thought

If G is finitely generated, is every subgroup of G finitely generated? In other words, if

$$G = \langle S|R\rangle$$

where S is a set of elements and R is the relation on those elements, then can you find a subgroup $H \leqslant G$ such that you need an infinite number of elements to generate H? There is an answer to this question which will be given as an exercise ref:HERE, but it can be useful to contemplate this question now as an exercise in the types of questions a mathematicians might ask themselves. You can further ask if $\varphi : G \to H$ is a homomorphsims, and G is finitely generated, is $\varphi(G)$ finitely generated?

The answer to this question is in ref:HERE.

Exercise 2.3.1

[3] If G is infinite, it might be that we require the stronger notion of an infimum: $\inf\{|A| \mid A \subseteq G,\ \langle A\rangle = G\}$, see ref:HERE (some analysis textbook)

1. Show that every finitely generated subgroup of $\mathbb{Q}$ is cyclic
2. Let (Presentation for V_4. This group is called the Klein 4-group, or *Viergruppe*. Prove that $V_4 \cong \mathbb{Z}_2 \times \mathbb{Z}_2$, showing that a group has a non-unique representation.
3. If $H, K \leqslant G$, is $H \cup K$ a group? If $H \cup K$ is a subgroup in some case, what relation would there be between H and K (does on include the other?)

2.3.3 Free Groups

(somewhere define "formal combination". Will need it for defining polynomial rings and tensors, and other free objects more generally)

So far, when we were generating groups from sets, we either took a subset of G (and so took the relation between elements of G), or we defined the set given a presentation. For example, We can make a group from $\{a, b\}$ by defining [4]

$$\langle a, b | a^n = 1, b^2 = 1, ab = b^{-1}a \rangle$$

what happens if we get rid of one or two of these relations? What about if we get rid of all three – we make a group "free" of relations besides necessary the ones needed to form a group? We explore this question in this section.

Since this group has no restrictions, it can be thought of as the "most general" group on that number of elements. In fact, something more subtle is going on here: since $\{a, b\}$ (or any set A to that matter) is essentially an arbitrary set, we're asking if from *any* arbitrary set, is there a way of defining a group using as little "extra group structure" as possible, or put another way, as free from relations as possible. If we can define such a group given $\{a, b\}$, is there a way of "defining" another group generated by two elements using this more "free group" by adding relations to the "free group"?

The answers to this question lies in exploring how to make a *function* (or to be pedantic a *set-function*) from any set A to a group G

$$f : A \to G$$

be "lifted" to a *group-homomorphism*

$$\varphi : F(A) \to G$$

where $F(A)$ will be our "free group" on A (the notation $F(A)$ is suggestive that this group will indeed be called a free group). The word "lifted" here means that if we restrict $F(A)$ to A, then $\varphi|_A = f$. This can be visualized with the following diagram (called a commutative diagram):

$$\begin{tikzcd} F(A) \arrow[r] & G \\ A \arrow[u] \arrow[ur] & \end{tikzcd}$$

Thus, we turn a set function into a group-homomorphism while still keeping the structure of the set-function. More generally, if we have $f : A \to G_2$, then if there exists a function $f_2 : A \to G_1$ and a homomorphism $\varphi : G_1 \to G_2$ such that $f_2 = \varphi \circ f_1$, then φ is called a lift of f_1. Equivalently, φ is

[4] This is D_n

a lift of f_1 is the following diagram commutes:

$$\begin{tikzcd} G_1 \arrow[r, "\varphi"] & G_2 \\ A \arrow[u, "f_1"] \arrow[ur, "f_2"'] & \end{tikzcd}$$

Why are lifts the key concept? It is because not all set-function can be lifted to group homomorphism:

Example 2.3.3: Lifts and non-Lifts

1. Let $A = \{a\}$, $G_1 = \mathbb{Z}/5\mathbb{Z}$, and $G_2 = \mathbb{R}$. Define

   (a) $f_1 : A \to G_1,\ f_1(a) = 0$

   (b) $f_2 : A \to G_2,\ f_2(a) = 4$

   Does A lift to G_1, that is:

   $$\begin{tikzcd} \mathbb{Z}/5\mathbb{Z} \arrow[r] & \mathbb{R} \\ \{a\} \arrow[u, "f_1"] \arrow[ur, "f_2"'] & \end{tikzcd}$$

   that is, does there exist a homorphism $\varphi : \mathbb{Z}/5\mathbb{Z} \to \mathbb{R}$ such that $f_2 = \varphi \circ f_1$ If we can find a homomorphism φ such that

   $$\begin{tikzcd} \mathbb{Z}/5\mathbb{Z} \arrow[r, "\varphi"] & \mathbb{R} \\ \{a\} \arrow[u, "f_1"] \arrow[ur, "f_2"'] & \end{tikzcd}$$

   then a lift exists. Let's say there was such a φ. Then we have that

   $$4 = f_2(a) = (\varphi \circ f_1)(a) = \varphi(0) = 0$$

   that is $4 = 0$ in $\mathbb{R}$. Note that since φ is a homomorphism, we are forced to have $\varphi(0) = 0$ as we showed in proposition 1.4.2. However, we already know that $4 \neq 0$. Thus, no such homomorphism can exist. Therefore, these is lift from $\{a\} \xrightarrow{f_1} \mathbb{R}$ to $\mathbb{Z}/5\mathbb{Z} \to \mathbb{R}$

2. What if we swap 0 and 4? Then:

   $$0 = f_2(a) = (\varphi \circ f_1(a)) = \varphi(4)$$

   Before, we were forced to say $\varphi(0) = 0$. However, this time we are only restricted to saying $\varphi(4) = 0$. There does indeed exist a homomorphism from $\mathbb{Z}/5\mathbb{Z}$ to $\mathbb{R}$ which satisfies this property: $\varphi(x) = 0$ (i.e., the trivial homomorphism). No other homomorphism can exist, since if $\varphi(4) \neq 0$, then $0 = \varphi(0) = \varphi(4 + 4 + 4 + 4 + 4) = 5 \cdot \varphi(4) \neq 0$ – a contradiction.

3. Let's take the same example as before, but let $f_1'(a) = 1$. Like last time, let's assume a homomorphism φ existed such that $f_2 = \varphi \circ f_1'$. Thus

   $$1 = f_2(a) = (\varphi \circ f_1')(a) = \varphi(1)$$

Thus, $\varphi(1) = 1$. As a consequence

$$\varphi(1+1) = \varphi(1) + \varphi(1) = 1 + 1 = 2 \cdots$$

meaning $\varphi(n) = n$. However, this implies that

$$0 = \varphi([0]) = \varphi([5]) = 5$$

but $0 \neq 5$. Thus, no such homomorphism exists, and so there is no morphism from (f_1', G_1) to (f_2, G_2).

4. Let $G_1 = G_2 = \mathbb{Z}/5\mathbb{Z}$ and let $A = \{a\}$. Let
   (a) $f_1 : \mathbb{Z}/5\mathbb{Z} \to \mathbb{Z}/5\mathbb{Z},\ f_1(a) = 1$
   (b) $f_2 : \mathbb{Z}/5\mathbb{Z} \to \mathbb{Z}/5\mathbb{Z},\ f_2(a) = 2$

   (TBD: want to show how there can be more than 1 (non-trivial) morphism between these two objects of $\mathcal{F}^A$)

So what group can take the place of $F(A)$ (if any) that will always satisfy a lift? Perhaps it is a group we have already encountered. We have shown in example 1.3.2 and section 1.2.2 how to define the trivial and symmetric group from arbitrary sets. Let's label the trivial group on A as $T(A)$, and the symmetric group on A as S_A. Though this shows that there is more than one way of defining a group operation on arbitrary sets, this does not help us in finding the most "general" group that can be defined on a set. Let's take the earlier presentation of D_n. Clearly, the trivial group has no information about this group. The group S_2 has 2 element: the trivial element and (1 2). However, $|D_n| = 2n$, meaning it has many more elements than S_2. Thus, such a group cannot be used as the "most general group on the set $\{a, b\}$"

In terms of intuition in relation to lifts, we see that:

$$A \to G \implies T(A) \to G$$

will not in general be a lift (choose some set $|A| > 1$ and a nontrivial group G). Similarly,

$$A \to G \implies S_A \to G$$

will not be a lift in general.

The problem with both these examples is that they in fact define a lot of structure on their underlying set. Instead, we will try to make a structure which puts the minimal amount of relations between objects to make it a group so that hopefully $F(A) \to G$ can always be a lift! Before I show the answer, I encourage you to stop reading for a moment try to think of another way you can define a group on some arbitrary set A, and perhaps think about how this group can be thought of as the "most general".

The High-level Idea

The first time you'll see this construction, you might think we added a lot of fluff in the way of the answer. However, I assure the reader that the longer build-up will have larger rewards! The construction we're doing here will repeat itself multiple times in the book in subtly different ways.

This longer construction allows us to link all these constructions as examples of *free objects* [5], and more generally *universal properties* [6] later down the line. This will be the first example in the book of a universal property, it's the *universal property of free groups.*

Let's say A is an arbitrary set. Our goal is to construct a group, which we will label $F(A)$, from which every other group which contains A can be derived. First, we will be "relating" A to any group G. Let's take all set-functions:

$$f : A \to G$$

Where A is a set that maps into the group G [7]. Let's label this function along with its co-domain G as (f, G). Choosing all possible f and G will make all possible functions from A to every possible group, essentially telling us all ways A relates to any group G. The fact that there is no restriction on the functions from A to G is a way of saying there is *no* group structure on A which limits the types of maps we can form. Remember that when we lift the map $f : A \to G$ to $\varphi : F(A) \to G$, we need that $\varphi|_A = f$, so the fact that the map can be arbitrary means that $F(A)$ must accommodate for any function from A to G. As we have shown in example 2.3.3, not all groups can accommodate this. Let's label the set of all possible (f, G) as $\mathcal{F}^A$:

$$\mathcal{F}^A = \{(f, G) | f : A \to G\}$$

Next, we want to be able to relate the groups in the co-domain. We want to define $F(A)$ to be able to lift from a set-function to a group-homomorphsims. What we will do is define the following relation between any two set's (f_1, G_1), (f_2, G_2): we will say there is a *morphism* between (f_1, G_1), (f_2, G_2)

$$(f_1, G_1) \xrightarrow{\varphi} (f_2, G_2)$$

if the following diagram commutes:

$$\begin{tikzcd} G_1 \arrow[r, "\varphi"] & G_2 \\ A \arrow[u, "f_1"] \arrow[ur, "f_2"'] & \end{tikzcd}$$

that is, where $f_2 = f_1 \circ \varphi$. You can think of the morphism φ as a homomorphism that is restricted by f_1 and f_2. In fact, every example in examples 2.3.3 can be re-written to be morphisms from $\mathcal{F}^A$. Take a moment to go back and translate the previous notation to the new one, and think about what $\mathcal{F}^A$ with its objects and morphisms represent. We've essentially taken any group homomorphism between any two groups $G_1 \xrightarrow{\varphi} G_2$ and also made a relation between A and these groups. In fact, as reader's who have read section 0.4 and 8 might have guessed, we have defined a category

[5] If you're curious to know where else we will encounter free objects, we will see *free monoids* in section 7.1 (which will turn out to be isomorphic to the set of all polynomials!), and *free-modules* in section 12.3.2

[6] You'll see these everywhere, to name a few: theorem 3.1.12, theorem ref:HERE (product), theorem ref:HERE (coproduct), proposition 9.2.3, theorem 9.5.3, theorem 15.3.1, theorem 15.1.2, and more!

[7] It might be more intuition for every f to be *injective*, since this way we'd get that every G contains the set A. We can indeed do this, and we will actually get the same result. For reasons that will be elaborated later in this section, the more general approach is desirable to be able to repeat this trick later in the textbook

Proposition 2.3.13: $\mathcal{F}^A$ is a category

The set $\mathcal{F}^A$ along with the morphisms form a category (we can write $\mathfrak{F}^A S$ as a shorthand to represent $\mathcal{F}^A$ and it's morphisms). In particular:

1. For every $(f, G) \in \mathfrak{F}^A$, there exists an identity morphism $1_{(f,G)} : (f, G) \to (f, G)$
2. If $\varphi, \psi \in \mathfrak{F}^A$ are morphisms

$$\varphi : (f_1, G_1) \to (f_2, G_2) \qquad \psi : (f_2, G_2) \to (f_3, G_3)$$

then $\psi \circ \varphi$ is a morphism.

***Proof* :**
First, it should be clear that composition of morphisms will be a morphism in $\mathcal{F}$. If $\varphi, \psi \in \mathcal{F}$ are composable morphisms (i.e. the codomain of φ matches the domain of ψ) then

$$\begin{tikzcd} G \arrow[r, "\varphi"] \arrow[rr, bend left, "\psi\circ\varphi"] & H \arrow[r, "\psi"] & K \\ & A \arrow[ul, "f"] \arrow[u, "f'"] \arrow[ur, "f''"'] & \end{tikzcd} \quad \Rightarrow \quad \begin{tikzcd} G \arrow[rr, "\psi\circ\varphi"] & & K \\ & A \arrow[ul, "f"] \arrow[ur, "f''"'] & \end{tikzcd}$$

that is, the morphism $\psi \circ \varphi \in \mathcal{F}^A$ with domain (f, G) and codomain (f'', K) exists

Next, we will check each object $(f, G) \in \mathcal{F}^A$ has an identity morphism. Take an arbitrary $(f, G) \in \mathcal{F}^A$, so that $A \xrightarrow{f} G$. Then the homomorphism $\mathrm{id}_G : G \to G$, $\mathrm{id}_G(x) = x$ satisfies

$$\begin{tikzcd} G \arrow[r, "\mathrm{id}_G"] & G \\ A \arrow[u, "f"] \arrow[ur, "f"'] & \end{tikzcd}$$

If we take any $\varphi \in \mathcal{F}^A$ where $\varphi : (f, G) \to (f', H)$, then $\mathrm{id}_H\, \varphi = \varphi\, \mathrm{id}_G = \varphi$, showing that identities exist.

The fact that any compossible triple is associative comes from function composition.

With the category $\mathfrak{F}^A$ finally defined, we can ask the following question

Does $\mathfrak{F}^A$ have an initial object[8]?

That is, does there exists (f^*, G^*) such that for any object $(f, G) \in \mathfrak{F}^A$, there exists a unique morphism φ^* such that

$$\begin{tikzcd} G^* \arrow[r, "\varphi^*"] & G \\ A \arrow[u, "f^*"] \arrow[ur, "f"'] & \end{tikzcd}$$

[8]see definition 1.4.4

Notice that this is the same lifting question we have asked about a theoretical "free group" $F(A)$ near the beginning of this section, but re-phrased! You might ask if such a group even exists? If such a group does exist, then have found a group which let's us relate it to any other group. We will call such a group (if it exists) a *free group*. Essentially, $F(A)$ must be able to relate to any other group, making it the "most general" group. But what is $F(A)$? In the following section, we will construct such a group that will indeed be the initial object of $\mathfrak{F}^A$.

Constructions of $F(A)$

Let's first give the answer to this question when $A = \{a\}$. Then the group $F(\{a\})$ is $\mathbb{Z}$. and set map $j : \{a\} \to \mathbb{Z}$ is $j(a) = 1$. Let (f, G) be any element of $\mathcal{F}^A$. Let's say $f(a) = g$. Then $g = f(a) = (\varphi \circ j)(a) = \varphi(j(a)) = \varphi(1)$. That is, the generator of $\mathbb{Z}$ maps to g. By proposition ref:HERE,

$$\varphi(n) = g^n$$

and since this was forced by proposition ref:HERE, φ is unique.

In the case of $\emptyset$, it would be $\{0\}$, which checks out since we saw that $\{0\}$ is the initial objects of groups (proposition 1.4.3). In the case were $A = \{a\}$ (that is, an arbitrary singleton). Then HERE $(\mathbb{Z})$.

For the more general case, we will define a new group.

Let A be an arbitrary set. For every element $a \in A$, let it correspond to some element $b \in B$. For the sake of what's to come, we will instead of label these element a^{-1}, and the set they belong to A'. Then, define a *word* to be any ordered tuple

$$(a_1, a_2, ..., a_n) \qquad a_i \in A \cup A'$$

We usually write a word as the literal putting together of letters side by side:

$$a_1 a_2 \cdots a_n$$

Definition 2.3.14: Word

Let A be some set. Then let $W(A)$ denote all possible combinations of *words*. TBD

Example 2.3.4: Example

1. here
2. here

Define reduction, define reduced word,

Lemma 2.3.15: reduced word lemma

reduced word lemma

Proof **:**
here

Proposition 2.3.16: Word Group

$(W(A), R)$ is a group

Proof **:**

1. (associativity)
2. ($\exists$ identity)
3. ($\exists$ inverse)

Theorem 2.3.17: Universal Property Of Free Groups

The following are each the same statement, just different ways of stating it. Let $f : A \to G$ be a function from a set to a group, $i : A \to F(A)$ to be $i(a) = a$ (i.e. $i : A \hookrightarrow F(A)$), and $F(A) = (W(A), R)$.

1. $(i, F(A))$ is the initial object of $\mathfrak{F}$.
2. The set $(i, F(A))$ satisfies the universal property for free groups on A.
3. Then there exist a unique homomorphism $\varphi : F(A) \to G$ such that f splits with i, that is, $f = i \circ \varphi$, i.e., the following diagram commutes

$$\begin{tikzcd} F(A) \arrow[r, "\varphi"] & G \\ A \arrow[u, "i"] \arrow[ur, "f"'] & \end{tikzcd}$$

we call the group component of $(i, F(A))$ the *free group* on A.

Proof **:**
We're essentially showing that given map $f : A \to G$, we can extend it to a map $\varphi : F(A) \to G$. So, let f be any map from A to some group G, and i be defined as in the theorem. TO define φ, we can use the information provided by f essentially completely. For any a and $a^{-1} \in A \cup A'$,

$$\varphi(a) = f(a) \qquad \varphi(a^{-1}) = f(a)^{-1}$$

notice that since A doesn't contain A, so the function f doesn't map a^{-1} anywhere. However, by construction, $F(A)$ does have a^{-1}, and so we define a place for it to map. To define φ on words with multiple letters by concatenating the result of the image, that is

$$\varphi(a_1 a_2 \cdots a_n) := \varphi(a_1)\varphi(a_2)\cdots\varphi(a_n)$$

This function is well-defined, since if $w = w'$, then

$$\varphi(w) = \varphi(a_1)\varphi(a_2)\cdots\varphi(a_n) = \varphi(w')$$

furthermore, this functions splits f since for any $a \in A$

$$f(a) = \varphi(i(a)) = \varphi(a)$$

which is true by definition

Now, notice that for any $w \in W(A)$, $\varphi(w) = \varphi(R(w))$. For example, if $w = abb^{-1}a^{-1}a$, then

$$\begin{aligned}\varphi(w) &= \varphi(abb^{-1}a^{-1}a)\\ &= \varphi(a)\varphi(b)\varphi(b^{-1})\varphi(a^{-1})\varphi(a)\\ &= f(a)f(b)f(b)^{-1}f(a)^{-1}f(a)\\ &= f(a)\\ &= \varphi(a)\\ &= \varphi(R(w))\end{aligned}$$

With this fact, we have that φ will be a homomorphism If $w, w' \in W(A)$, then

$$\varphi(w \cdot w') = \varphi(R(ww')) = \varphi(ww') = \varphi(w)\varphi(w')$$

establishing that $\varphi : F(A) \to G$ is a homomorphism!

Finally, we need to show that it's unique. This comes down to seeing that by the properties of homomorphisms, the way we've defined is the only way we can define a homomorphism from $F(A)$ to G. Let's say φ' was another homomorphsims from $F(A)$ to G. Then by the commutative diagram, it must be that

$$f(a) = \varphi'(i(a)) = \varphi'(a)$$

but that is exactly the definition of φ. The fact that φ' must split with f forces $\varphi = \varphi'$.

Since f an G was arbitrary, we showed that there always exists a unique homomorphism φ from $F(A)$ to G such that $f = \varphi \circ f$ – as we sought to show

Corollary 2.3.18: Free Group Is Unique

let A and B be sets such that $|A| = |B|$. Then

$$F(A) \cong F(B)$$

Proof **:**
This comes down to properties of terminal objects in categories (initial objects are *a fortiori* terminal objects).

Since free groups on sets of the same cardinality are isomorphic, we usually write F_n to denote the free group with n generators.

Definition 2.3.19: Freely Generated

Let S be a set. Then we say that S *freely generates* a group G if $G = F(S)$.

Example 2.3.5: Free Groups

1. Earlier we said that $F(A) \cong \mathbb{Z}$. If we take $f : \{a\} \to \mathbb{Z}$, $f(a) = 1$, then by the universal property of free groups, there exists a group-homomorphism: $\varphi : F(A) \to \mathbb{Z}$, where $\varphi(1) = 1$. Then it's evidently true that φ defines an isomorphism between $F(A)$ and $\mathbb{Z}$ (which you should check if it's unclear), and so
$$F(A) \cong \mathbb{Z}$$

2. The free set on 2 elements, say $F(\{x, y\})$ is a group we have encountered. The easiest way to show this group is by showing part of its Cayley graph:

Figure 2.1: Cayley Graph of F_2

this graph in fact goes on forever. The length of the edges where halved every time to make the diagram nicer. [a]

[a] If you have done any algebraic topology, this is the fundamental group of the figure 8!

Some other properties from Clara Löh's book? Geometric group theory p. 17

2.3.4 Free Abelian Group

We can repeat this same construction, except limit our attention to abelian groups. In particular, given a set A an any map $f : A \to G$ to an abelian group, we want to lift it to an abelian group $F^{ab}(A)$ in the domain of the homomorphism $\varphi : F^{ab}(A) \to G$.

The construction is actually basically identical if we impose a commutativity condition on $F(A)$, that is, for any $a, b \in F^{ab}(A)$, $ab = ba$. If you were uncomfortable with the proof of the universal property of free group, it is worth going over it again, then go over it a 2nd time but with only abliean groups and $F^{ab}(A)$ in mind.

What we will be focusing on this section is finding a nicer group that is isomorphic to $F^{ab}(A)$.

As a closing remark on the construction of free and free abelian groups, (How it is plausible that $F(A)$ is empty or a singleton or something smaller than A so this idea of "lift" is not true. However, our construction at the end will show that it is indeed a lift. In many similar constructions in the book (ref:HERE), this construction will still be lifts, however we will encounter some (the first one in chapter 15 in which there is an initial object which is not a lift)

2.3.5 *Free-Forgetful Adjoint

Is this the right place to ask this question? It's asking whether we have really defined the "right" idea of general group. The way we make this idea rigorous is by asking if we have a functor $F : \mathbf{Group} \to \mathbf{Set}$, that forgets the structure, is there a functor $G : \mathbf{Set} \to \mathbf{Group}$ that "regains" some structure? It cannot regain all of it because sets don't have enough information for that, but is there a way in which we can make sure that our free construction is the "mirror-image" of forgetting the structure?

Build up to this: If $L(_) : \mathbf{Group} \to \mathbf{Set}$ takes a group G and returns the underlying set: $\mathcal{L}((G, \cdot)) = G$, and $\mathcal{F}(_) : \mathbf{Set} \to \mathbf{Group}$ takes a set and returns a group: $\mathcal{F}(S) = ((F(S), \cdot_{F(S)})$, then

$$\mathrm{Hom}_{\mathbf{Set}}(S, \mathcal{L}(G)) \cong_{\mathbf{Set}} \mathrm{Hom}_{\mathbf{Group}}(\mathcal{F}(S), G)$$

where we there is $\cong_{\mathbf{Set}}$ in the middle because Hom_{Set} doesn't have a group structure. If it did, we would ask it to preserve that too.

2.4 Subgroup Lattices

In this section, we will show how to use graphs in order to visually see the subgroup structure of a group.

Before introducing a rigorous definition, let's first do an example: Let's do the subgroup lattice of Q_8. We must first know all the subgroups of Q_8. With some computation, the subgroups that are found are

$$1,\ \langle -1 \rangle,\ \langle i \rangle,\ \langle j \rangle,\ \langle k \rangle,\ Q_8$$

where $1 = \{e\}$. By the definition of Q_8, any other set of generators will fall into these groups. For example, notice that $\langle i, j \rangle = \langle k \rangle$ (similarly for i, k and j, k), and $\langle i, j, k \rangle = Q_8$. To construct the lattice, start by imagining Q_8 on the "top" and the identity group on the "bottom"‘

$$Q_8$$

$$1$$

then, apply the first rule of doing the lattice:

1. draw a line if one group is a subgroup of another group.

Since $1 \leqslant Q_8$, then

$$\begin{tikzcd} Q_8 \arrow[d, dash] \\ 1 \end{tikzcd}$$

Next, pick one of the groups, let's say it's $\langle -1 \rangle$. Now apply the second rule of drawing a lattice:

2. if $H \leqslant K \leqslant G$, then remove the line from H to G (if one is already drawn) and add a line from H to K, and K to G. In other words, only draw a line between subgroups A and B if and only if there is no group H such that $A \leqslant H \leqslant B$.

Since $1 \leqslant \langle -1 \rangle \leqslant Q_8$, our diagram will now look like this:

$$\begin{tikzcd} Q_8 \arrow[d, dash] \\ \langle -1 \rangle \arrow[d, dash] \\ 1 \end{tikzcd}$$

similarly, we can add $\langle i \rangle$ like so:

$$\begin{tikzcd} Q_8 \arrow[d, dash] \\ \langle i \rangle \arrow[d, dash] \\ \langle -1 \rangle \arrow[d, dash] \\ 1 \end{tikzcd}$$

Next, notice that $\langle i \rangle \nleqslant \langle j \rangle$, and $\langle j \rangle \nleqslant \langle i \rangle$, however both $\langle i \rangle$ and $\langle j \rangle$ have $\langle -1 \rangle$ as a subgroup and are both subgroups of Q_8. The same situation occurs for $\langle k \rangle$. Thus, the updated lattice will be:

$$\begin{tikzcd} & Q_8 \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\ \langle i \rangle \arrow[dr, dash] & \langle j \rangle \arrow[d, dash] & \langle k \rangle \arrow[dl, dash] \\ & \langle -1 \rangle \arrow[d, dash] & \\ & 1 & \end{tikzcd}$$

since we have exhausted our list of possible subgroups, we have finished the subgroup lattice of Q_8.

More generally, the orientation doesn't matter (we could've had 1 on top and Q_8 on bottom, or going from left to right). What matter's is we know how to translate a line into an appropriate subgroup (i.e. If 1 is on top and Q_8 is on bottom, then a line now becomes a supergroup inclusion $H \geqslant G$). Note also that the particular shape the lattice took doesn't really matter, a pretty diamond-like shape was chosen to make the lattice more readable.

If G is an infinite group, then a lattice doesn't always exist. For example, if $G = \mathbb{Q}$, then for any group $1 \leqslant H \leqslant G$, there exists a group K such that $1 \leqslant K \leqslant H$, meaning we cannot follows condition (2). Thus, we usually work with finite groups. Certain infinite groups permit subgroup lattices (see exercise ref:HERE)

With this build-up, we can formally define a subgroup lattice:

Definition 2.4.1: Group Lattice

Let G be a finite group and $\mathcal{G}$ be the collection of subgroups of G. Then $\Gamma = (\mathcal{G}, E_{\leqslant})$ is the graph where there is an edge between a group $A, B \in \mathcal{G}$ if and only if there does not exist a group K such that $A \leqslant K \leqslant B$. When Γ is visualized, we put G on top and the identity group 1 on bottom.

Example 2.4.1: Group Lattices

Many examples, and the problem's that might arise with infinite lattices, and how some infinite groups permit lattices

1. Let $G = \mathbb{Z}/2\mathbb{Z} = C_n$. Then by theorem ref:HERE, the subgroup lattice is easily determined by the divisors of n. Thus ,we get:

$$\begin{tikzcd}
 & & \mathbb{Z}/8\mathbb{Z} \arrow[d, dash] \\
 & \mathbb{Z}/4\mathbb{Z} \arrow[d, dash] & \langle 2 \rangle \arrow[d, dash] \\
\mathbb{Z}/2\mathbb{Z} \arrow[d] & \langle 2 \rangle \arrow[d, dash] & \langle 4 \rangle \arrow[d, dash] \\
0 & 0 & 0
\end{tikzcd}$$

Note that $\mathbb{Z}/2\mathbb{Z} = \langle 1 \rangle$, or $\mathbb{Z}/4\mathbb{Z} = \langle 1 \rangle = \langle 3 \rangle$, or that $\mathbb{Z}/8\mathbb{Z} = \langle 1 \rangle = \langle 3 \rangle = \langle 5 \rangle = \langle 7 \rangle$

2. More generally, if $G = \mathbb{Z}/p^n\mathbb{Z}$, then

$$\begin{tikzcd}
\mathbb{Z}/p^n\mathbb{Z} = \langle 1 \rangle \arrow[d, dash] \\
\langle p \rangle \arrow[d, dash] \\
\langle p^2 \rangle \arrow[d, dash] \\
\vdots \arrow[d, dash] \\
\langle p^n \rangle = 0
\end{tikzcd}$$

3. If $G = \mathbb{Z}/n\mathbb{Z}$ where n is a non-prime power, we can get some interesting diagram. For example:

$$\begin{tikzcd}
 & \mathbb{Z}/20\mathbb{Z} \arrow[dl, dash] \arrow[d, dash] \\
2 \arrow[d, dash] \arrow[dr, dash] & 5 \arrow[d, dash] \\
4 \arrow[dr, dash] & 10 \arrow[d, dash] \\
 & 0
\end{tikzcd}$$

4. Let $G = V_4$ (the Klein 4 group from exercise ref:HERE). Then

$$\begin{tikzcd}
 & V_4 \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
\langle a \rangle \arrow[dr, dash] & \langle b \rangle \arrow[d] & \langle c \rangle \arrow[dl, dash] \\
 & 0 &
\end{tikzcd}$$

Notice that this is the same lattice as the Quaternions (i.e. Q_8) that we have shown at the beginning. However, $V_4 \ncong Q_8$ (worth doing as an exercise). This shows that having the same group lattice is not a sufficient condition for two groups to be isomorphic (though it is a necessary condition)

Theorem 2.4.2: Subgroup And Sublattices

Let $H \leqslant G$. Then the lattice of H is the sublattice of G of groups contained in H

***Proof* :**
This fact comes straight from the definition of a lattice. Since $H \leqslant G$, then if the subgroup lattice of H where to differ from the subgroup lattice of G whose subgroups are contained in H, then the lattice of G would be wrong – a contradiction.

Example 2.4.2: Example
Give D_4, Q_8, $\mathbb{Z}/n\mathbb{Z}$.

(TBD on phrasing) We can also ask ourselves if homomorphisms preserve some group lattice. In particular, if $\varphi : G \to H$ is a homomorphism, and $\varphi(G)$ is the image of φ, can we say something about the group lattice of $\varphi(G)$ given the group lattice of G? We know that $\varphi(G)$ will be sublattice of H by theorem 2.4.2. However, finding the lattice of $\varphi(G)$ is a little more tricky. We will recieve our answer next chapter – the name of the theorem giving the answer will be the 4th isomorphism theorem.

11: Foreshadowing

$\varphi(G)$ will be have "dual" to the subgroup lattice condition. That is, while subgroup's will have a sublattice of groups contained in H, $\varphi(G)$ will have the subgroup lattice of groups *contained* within a special group (which we will discover in chapter 3

Exercise 2.4.1

1. Let $G = \mathbb{Q}$. If $1 \leqslant H \leqslant G$, show that there always exists a group H' such that $1 \leqslant 1 \leqslant H' \leqslant H$.

2.5 *Universal Property of kernel and coKernels

This section can be comfortably skipped, as its contents will not come back until chapter ref:HERE. This section covers how to think about the kernel as a universal object. As you can probably guess by the fact that there is a universal property, the concept of the kernel of a homomorphism is not limited to groups. They will appear often in other sections. Most importantly, in chapter ref:HERE, (why do we need kernels when working with derived categories)

In the following seciton, $0 : G \to H$ will represent the homomorphism $\varphi(G) = e_H [= 0]$. This is indeed a homomorphism (as we've seen in proposition 1.4.3).

Theorem 2.5.1: Universal Property Of The Kernel

Let $\varphi : G \to H$ be a homomorphism. Then the inclusion map $i : \ker(\varphi) \hookrightarrow H$ is the final object in the category of group-homomorphsism $\alpha : G \to H$ such that $\varphi \circ \alpha$ is the trivial map. In other words, every group homomorphism $\alpha : K \to G$ such that $\varphi \circ \alpha$ is trivial factor through

$$\begin{tikzcd} K \arrow[r, "\alpha"'] \arrow[rr, bend left, "0"] \arrow[dr, "\exists!\alpha'"'] & G \arrow[r, "\varphi"'] & H \\ & \ker(\varphi) \arrow[u, hook, "i"'] & \end{tikzcd}$$

(put this somewhere as a reference to double-check if they got the category right?) Let $C_{\varphi,K}$ be the category who's objects be homomorphsism $\alpha : K \to G$ such that

$$\begin{tikzcd} K \arrow[r, "\alpha"'] \arrow[rr, bend left, "0"] & G \arrow[r, "\varphi"'] & H \end{tikzcd}$$

If (K, α) and (J, β) are two objects in this category, define the morphism $\gamma : (K, \alpha) \to (J, \beta)$ to be a homomorphism from K to J such that

$$\begin{tikzcd} K \arrow[r, "\alpha"'] \arrow[rr, bend left, "0"] \arrow[dr, "\gamma"'] & G \arrow[r, "\varphi"'] & H \\ & J \arrow[u, "\beta"'] & \end{tikzcd}$$

commutes.

Proof **:**
Let $\alpha : K \to G$ such that $\varphi \circ \alpha = 0$, that is

$$\varphi(\alpha(K)) = 0$$

then $\alpha(K) \subseteq \ker(\varphi)$. If there was a homomorphism $\overline{\alpha} K \to \ker(\varphi)$, it would have to factor through:

$$\alpha = i \circ \overline{\alpha}$$

but this forces $\overline{\alpha}$ to be α restricted to K, forcing uniqueness and existnece.

(Also explain the idea of the dual question, or that it exists), namely

Theorem 2.5.2: Universal Property Of The coKernel

Let $\varphi : G \to H$ be a homomorphism. Then the inclusion map $i : \ker(\varphi) \hookrightarrow H$ is the final object in the category of group-homomorphsism $\alpha : G \to H$ such that $\varphi \circ \alpha$ is the trivial map. In other words, every group homomorphism $\alpha : K \to G$ such that $\varphi \circ \alpha$ is trivial factor through

$$\begin{tikzcd} G \arrow[r, "\varphi"'] \arrow[rr, bend left, "0"] & H \arrow[r, "\beta"'] \arrow[d, "\pi"] & Q \\ & G/H \arrow[ur, "\exists!\beta'"'] & \end{tikzcd}$$

which we will anwser in chapter 3?

Also, for cokernel, this useful exact sequence might be helpful:

$$0 \to \ker(f) \to V \xrightarrow{f} W \to \operatorname{coker}(f) \to 0$$

3

Homomorphisms and Quotient Groups

In this chapter, we take a closer look at homomorphisms and how they relate different groups. We start our exploration by getting a better understanding of the image of a homomorphsim. For example, if $\varphi : G \to H$ is a homomorphism, then we saw that $\varphi(G)$ is a group (re-prove this if it is not clear). Since φ is a function, then as we saw in φ making $(G/\sim_\varphi)$ bijective to $\varphi(G)$ (in particular, $x \sim_\varphi y \;\Leftrightarrow\; \varphi(x) = \varphi(y)$), which we can see by the φ making $(G/\sim_\varphi)$ bijective to H (in particular, $x \sim_\varphi y \;\Leftrightarrow\; \varphi(x) = \varphi(y)$), which we can see by the function:

$$[g] \mapsto \varphi(g)$$

The collection $(G/\sim_\varphi)$ is called the *fibers* of φ[1]. Since there is a bijection between $\varphi(G)$ and $(G/\sim_\varphi)$, and $\varphi(G)$ is a group, can we define a group structure on $(G/\sim_\varphi)$? Furthermore, is there some "minimal" and "unique" group structure that we can put on $(G/\sim_\varphi)$ so that there is a cannonical identification between $\varphi(G)$ and $(G/\sim_\varphi)$; those comfortable with the category theory presented so far will recognize this as asking for the universal properties of quotient objects in the category **Grp**. Due to this universal property, the name of the group structure on $(G/\sim_\varphi)$ satisfying the universal property will be the *quotient group*. The exploration of this question is at the center of the first section of this chapter. Due to the axioms of groups, it will turn out that there is a lot of "regularity" in the collection of fibers which allow for an easy representation of $(G/\sim_\varphi)$.

After finishing the first part, we will have developed the appropriate tools to get back to what we started in section 1.3 and define these notions more formally, i.e., how do we impose relations on groups. about algebraic structures more generally. about algebraic structures mor egenerally.

For the rest of the chapter, we will use our new theory to give us some very powerful results. We will see that when working with finite groups, the order of the group will give us a lot of information

[1] Given $\varphi : G \to H$ and an element $h \in H$, the fiber of h is all elements that map to h by φ

about it's subgroup structure, we will use our understanding of quotient groups to learn how to build some new types of groups, and we will learn 4 isomorphism theorems which give us lots of information on group structure.

We will end this chapter by introducing one of the most fundamental classification theorem in finite group theory: *Jordan Hölder's Program.* In the last few section, we will show how there are groups that can be considered the "building blocks" of any other finite groups, and show how we can put these groups together to form other groups.

3.1 Quotient Groups

Let $\varphi : G \to H$ be a homomorphism. We start by exploring the structure of $\varphi(G)$. Since $\varphi(G)$ is a set, we know by proposition 0.1.17 in the preliminary chapter that there is a bijection between $G/\sim_\varphi$ and $\varphi(G)$ such that for any function $\varphi : G \to H$ where $a \sim b \Leftrightarrow \varphi(a) = \varphi(b)$, we have

$$\begin{tikzcd} G \arrow[r, "\varphi"] \arrow[d, "\pi"'] & H \\ G/\sim \arrow[ur, "\bar{\varphi}"'] & \end{tikzcd}$$

We now try to see if there is an equivalent universal property for quotients in **Grp**. To that end, we would then want $G \xrightarrow{\pi} G/\sim$ to be a homomorphism, that is

$$\pi(a \cdot b) = \pi(a) \bullet \pi(b)$$

or, written in terms of equivalence relations,

$$[a \cdot b] = [a] \bullet [b]$$

We would need to have that $\bullet$ is independent of representation. That is, if we have $x, y \in [a]$, then we want

$$x \sim y \Leftrightarrow xb \sim xb \tag{3.1}$$

since we would also want that

$$[b \cdot a] = [b] \bullet [a]$$

we must have that

$$x \sim y \Leftrightarrow bx \sim by \tag{3.2}$$

These are necessary condition in order for $(G/\sim, \bullet)$ to be a group, since if once of these condition's fail, then we won't have a group. For examle, if $x \sim y \not\Rightarrow xb \sim yb$, then $[x][b] \neq [xb]$.

These are necessary conditions, since if $G/\sim$ was a group, then these must hold. We can ask if they are sufficient conditions, that is, condition's such that if they are followed, then $(G/\sim, \bullet)$ must be a group. It turns out that these two condition's are also sufficient, meaning defining a group structure on $G/\sim$ is equivalent to the equivalence relation respecting these two properties:

Theorem 3.1.1: $(G/\sim, \cdot)$ is the quotient object of Grp

Let G be a group and $G/\sim$ be the set of equivalence relations given $\sim$. Then $G/\sim$ is a group under $\cdot$ if and only if

$$\forall g \in G,\ (x \sim y \Leftrightarrow gx \sim gy \text{ and } xg \sim yg)$$

Furthermore, If $\varphi : G \to H$ is another homomorphism such that $g \sim g' \Rightarrow \varphi(g) = \varphi(g')$, then there exists a homomorphism φ such that the following diagram commutes:

$$\begin{tikzcd} G/\sim \arrow[r, "\varphi"] & H \\ G \arrow[u, "\pi"] \arrow[ur, "\varphi"'] & \end{tikzcd}$$

We say that $\sim$ is *compatible* with the group-structure on G

***Proof* :**
The proof of the $\Rightarrow$ direction we have just showed above the theorem, so we'll show that these two equivalence relations are sufficient conditions for to define $(G/\sim, \cdot)$.

First, note that the operation $\cdot : G/\sim \times G/\sim \to G/\sim$ $[a] \cdot [b] = [ab]$ is well-defined, since if we take any other two representative $a' \in [a]$ and $b' \in [b]$, then since

$$b \sim b' \Leftrightarrow ab \sim ab' \text{ and } a'b \sim a'b'$$

and

$$a \sim a' \Leftrightarrow ab \sim a'b \text{ and } ab' \sim a'b'$$

Thus, putting pieces together using transitivity, we get

$$ab \sim ab' \sim a'b' \text{ or } ab \sim a'b \sim a'b'$$

both showing that $ab \sim a'b'$, meaning $\cdot$ is well-defined. Next, we need to show that $(G/\sim, \cdot)$ is indeed a group

1. Let $[a], [b], [c] \in G/\sim$. Then

$$\begin{aligned} ([a][b])[c] &= ([ab])[c] \\ &= [(ab)][c] \\ &= [(ab)c] \\ &= [a(bc)] \\ &= \vdots \\ &= [a]([b][c]) \end{aligned}$$

2. I claim that $[e_G] \in G/\sim$ is the identity. For any $[g] \in G/\sim$,

$$[g][e_G] = [ge_G] = [g] = [e_Gg] = [e_G][g]$$

as we sought to show

3. Finally, for any $[g] \in G/\sim$, I claim that $[g]^{-1} = [g^{-1}]$. Indeed:

$$[g][g^{-1}] = [gg^{-1}] = [e_G]$$

showing that $[g^{-1}]$ is indeed the inverse of G

Thus, $(G/\sim, \bullet)$ is indeed a group.

Finally, we must show that G is universal with respect to homomorphisms $\varphi : G \to G'$ such that $g \sim g' \Rightarrow \varphi(g) = \varphi(g')$. Note that we already have that such a function exists, since φ is also a set function, so we already have that there exists a unique set-funciton $\varphi : G/\sim \to G'$ such that $\varphi([g]) = \varphi(g)$. Thus, we only need to check that φ is a homomorphism. We get this pretty much be definition:

$$\varphi([a][b]) = \varphi([ab]) = \varphi(ab) = \varphi(a)\varphi(b) = \varphi([a])\varphi([b])$$

showing that φ is indeed a homomorphism, and must be unique since there is only one possible set function (and hence homomorphism) that satisfies this property (there are no more functions to check).

Since $\bullet$ is the unique structure on $G/\sim$ that respects the universal property, we will drop it without worry Since $\bullet$ is the unique structure on $G/\sim$ that respects the universal propety, we will drop it without worry that there be any confusion, that is well write

$$[a][b] = [ab]$$

$\therefore$

Definition 3.1.2: Quotient Group Via Congruence Relation

Let G be a group and $\varphi : G \to H$ a group homomorphism giving the induced equivalence relation $\sim_\varphi$. Then the group $G/\sim_\varphi$ from theorem 3.1.1 is called a *quotient group* and the equivalence relation $\sim_\varphi$ is called a *congrunce relation*.

The fact that equivalence relations that respect 3.1 and 3.2 are the only ones that are The fact that equivalence relations that respect 3.2 and 3.2 are the only ones that are congruence relations is very powerful, since this greatly limits what type of equivalence relations we have. For the 3.1, 3.2, or both, that is we shall find all possible congruence relations, as well as all possible one-sided congruence relations. We start by characterizing the equivalences classes of $G/\sim$ satisfying 3.1. We start by characterizing the equivalences classes of $G/\sim$ satisfying 3.1. there is any information we can figure out about the equivalence classes:

Proposition 3.1.3: Equivalence Classes Using property 3.1

Let $G/\sim$ be a quotient set given $\sim$ with proprety 3.1. Then

1. The equivalence class $K \in G/\sim$ where $e_G \in K$ is a subgroup of G ($K \leqslant G$)
2. Given any other equivalence calss $[g] \in G/\sim$, for any $a, b \in [g]$

$$a \sim b \;\Leftrightarrow\; a^{-1}b \in K \;\Leftrightarrow\; aK = bK$$

where the $aK = \{ak \mid k \in K\}$ and $bK = \{bk \mid k \in K\}$

***Proof* :**

1. Since $K \subseteq G$, we need to show that for any $a, b \in K$, $ab^{-1} \in K$. Since $e_G \in K$, then

$$e_G \sim a$$

since $\sim$ is a congruence relation:

$$a^{-1} \sim e_G$$

using property 3.1. Using property 3.1 again, we get

$$a^{-1}b \sim b \sim e_G$$

where the last part comes from the transitivity of equivalence relations, and so

$$a^{-1}b \sim e_G \Rightarrow a^{-1}b \in K$$

meaning $K \leqslant G$. Note that if we worked with property 3.2, we would instead be showing that $ab^{-1} \in K$.

2. Let $[g]$ be an equivalence class and $a, b \in [g]$ so that $a \sim b$.

Since $\sim$ respects 3.1 $a^{-1}b \sim e_G$, implying $a^{-1}b \in K$. Next, take any element $x \in a^{-1}bK$. Then $x = a^{-1}bk$. Next, since $a^{-1}b \in K$ and $k \in K$, $x \in K$. Thus, $a^{-1}bK \subseteq K$. This implies that $bK \subseteq aK$, since if $x \in bK$, then $x = bk$, and so $x = a(a^{-1}bk)$, and since $a^{-1}b \in K$ and $k \in k$, we get $x = ak'$ for some $k' = a^{-1}bk \in K$, an so $x \in aK$. Using symmetry of $\sim$, we'll get that $aK \subseteq bK$, implying $aK = bK$. Thus, we showed that

$$a \sim b \Rightarrow a^{-1}b \in K \Rightarrow aK = bK$$

Note that if we were using property 3.2, then we'd get $Ka = Kb$ since we'd be multiplying on the right.

To make this chain of implication go full circle, assume $aK = bK$. By setting $k = e_G$, we get $ae_G = a \in bK$, and so $ab^{-1} \in K$. Since K is the equivalence class containing e_G, we have

$$ab^{-1} \sim e_G \Leftrightarrow a \sim b$$

by 3.1, thus showing that $aK = bK \Rightarrow a \sim b$, completing the loop of implication, thus showing that

$$a \sim b \Leftrightarrow a^{-1}b \in K \Leftrightarrow aK = bK$$

as we sought to show

This shows that for any equivalence relation satisfying 3.1 (and similarly 3.2), it's equivalence classes will be of the form aK resp. Ka aK Ka where a is the representative of an equivalence class and K is the equivalence class containing e_G. Thus, an equivalence class will determine a subgroup K which then determines a representation for the equivalence classes of $G/\sim$. equivalence class will determine a subgroup K which then determines the equivalence classes of $G/\sim$.

Corollary 3.1.4: Quotient Set Size

let G be a a finite group and $G/\sim$ be the quotient sets with respect to an equivalence relation $\sim$ satisfying equation 3.1 or 3.2 (or both). Then every quotient set is of the same size, that is

$$|aK| = |bK| \qquad \forall a, b \in G$$

Proof **:**
exercise (see ref:HERE)

Since these cosets represent our equivalence relations, we will give a special name to these quotient sets:

Definition 3.1.5: Cosets

Let $H \leqslant G$. Then for every $g \in G$, we can define

$$gH = \left\{gh \middle| h \in H\right\}$$

we call this a [left]-***coset***. We can also define the ***right***-coset:

$$Hg = \left\{hg \middle| h \in H\right\}$$

If additive notation is used for G ($(G, +)$), it is common to denote the coset as

$$g + H \qquad H + g$$

The set of left-cosets is represented as G/H, an the set of right-cosets is represented as $G \setminus H$.

What equivalence relations that satisfies poposition 3.1.3? In other words, we're asking for a converse to proposition 3.1.3[1], which stated that if $\sim$ respected property 3.1 (or 3.2), then the equivalence class containing e_G is a subgroup. The answer is yes:

Proposition 3.1.6: Subgroups And Equivalence Classes

Let $H \leqslant G$ be a subgroup of G. Define the equivalence relation

$$a \sim_L b \Leftrightarrow a^{-1}b \in H$$

then $\sim_L$ respects property 3.1

***Proof* :**
Let's first show that $\sim_L$ is indeed an equivalence relation

1. If $a \sim a$, then $a^{-1}a = e \in H$, which is the case since H is a group
2. Let $a \sim b$. Then $a^{-1}b \in H$, implying $(a^{-1}b)^{-1} = b^{-1}a \in H$, implying $b \sim a$
3. Let $a \sim b$ and $b \sim c$. Then $a^{-1}b, b^{-1}c \in H$. Then $a^{-1}bb^{-1}c = a^{-1}c \in H$, implying $a \sim c$.

Thus, $\sim_L$ is indeed an equivalence relation.

To show it satisfies property 3.1, notice that for any $g \in G$

$$a \sim b \to a^{-1}b \in Ha^{-1}(g^{-1}g)b \in H \Rightarrow (ga)^{-1}gb \in H \Rightarrow ga \sim_L gb$$

showing that $\sim_L$ satisfies property 3.1

This proposition shows that any subgroup of H will define an equivalence relation $\sim_L$ respecting property 3.1, and thus by proposition 3.1.3[2] partition G (even better, equally partition G due to corollary 3.1.4). Similarly, we can show that the relatoin $a \sim_R b \Leftrightarrow ab^{-1} \in H$ gives an equivalence relation which satisfies property 3.2. Note that since H partitioned G, only $eH = H$ is a subgroup of G since only it contains the identity.

Combining proposition 3.1.3 and 3.1.6 we get the following one to one correspondence:

$$\{H|H \leqslant G\} \leftrightarrow \left\{ \begin{array}{c} \text{equivalence relations } \sim_L \\ \text{on } G \text{ satisfying} \\ \text{property 3.1} \\ \\ \end{array} \right\}$$

Similarly:

$$\{H|H \leqslant G\} \leftrightarrow \left\{ \begin{array}{c} \text{equivalence relations } \sim_R \\ \text{on } G \text{ satisfying} \\ \text{property 3.2} \\ \\ \end{array} \right\}$$

3.2, showing that the subgroups of G give us all possible such equivalence relations, that is all one-sided congruence relations. Because of this, we can without ambiguity write G/H to represent $G/\sim$ with an equivalence relation $\sim$ satisfying property 3.1, and similarly for $G \setminus H$. 3.2, showing that the subgroups of G give us all possible such equivalence relations. Because of this, we can without ambiguity write G/H to represent $G/\sim$ with an equivalence relation $\sim$ satisfying property 3.1, and similarly for $G \setminus H$.

$G \setminus H$ satisfying $\sim_R$? The answer is actually no: G/H satisfying $\sim_R$? The answer is actually no:

12: counter-example

Take $G = D_6$. Then $H = \langle \sigma \rangle$ is a subgroup, and has left cosets $eH(= H), rH, r^2H$. giving us

$$\{e, \sigma\}, \{r, r\sigma\}, \{r^2, r^2\sigma\}$$

and the right-cosets are

$$\{e, \sigma\}, \{r, \sigma r\}, \{r^2, \sigma r^2\}$$

using the fact that $r\sigma = \sigma r^2$ and $r^2\sigma = \sigma r^2$, we can re-write the right cosets as

$$\{e, \sigma\}, \{r, r^2\sigma\}, \{r^2, r\sigma\}$$

showing that they partition G differently than the left cosets!
This means that $\sim_L$ defined by H does not respect property 3.2. If it did, then

$$r \sim_L r\sigma \Rightarrow r(r\sigma)^{-1} = r\sigma r^2 = rr\sigma \in H$$

however, $r^2\sigma \notin H$, showing that it does not satify the propety

Remember that we showed in proposition 3.1.3 that *both* the properties need to be satisfied in order for G/H to be a group. Since every $\sim$ with these properties corresponds to a subgroup H, it will come down to a special subgroup H which has both properties

(build-up normal subgroups somehow!) We have in fact already encountered the property which will always make H induce both $\sim_L$ and $\sim_R$

(I don't like this "we've already encountered". Would be better to build up the same way we did before, with the "neccessary and sufficient" narrative from before)

Proposition 3.1.7: When H Induces Both $\sim_L$ And $\sim_R$

The relatoins $\sim_L$ and $\sim_R$ induce the same cosets corresponding to H if and only if H is normal

Proof **:**
If H is normal, then $gH = Hg$, meaning the left and right cosets are the same and so H induces both $\sim_L$ and $\sim_R$. Conversly, if $\sim_L$ and $\sim_R$ are both respected, and take $x \in aH$, so $xa^{-1} \in H$, meaning $xa^{-1} = h$. Manipulating this, we get that $ax^{-1} = h^{-1}$ so $x^{-1} = a^{-1}h^{-1}$, and so $x^{-1} \in a^{-1}H$. Then since $\sim_R$ is respected, $x \sim a$. Since $\sim_L$ is repsected then $x^{-1}a \in H$, meaining $x^{-1}a = h$ or $x^{-1} \in Ha^{-1}$. Then by similar reasoning we get $x \in Ha$, showing that $aH \subseteq Ha$. doing the same argument backwards we get that $Ha \subseteq aH$ giving us $aH = Ha$, which is exactly the condition of normality

(TBD, the converse of my proof is very messy)

Definition 3.1.8: Normal Subgroup

Let $H \leqslant G$ such that $aH = Ha$. Then H is a normal subgroup, and we denoted it $H \trianglelefteq G$

We've already encountered normal subgroups before

Example 3.1.1: Normal Subgroups

1. for any group G, $Z(G) \leqslant G$. Take any $y \in G$ and consider $yx \in yZ(G)$. Then by definition of $Z(G)$, $yx = xy$, and so $[xy] = yx \in Z(G)y$, showing that $Z(G)$ is normal
2. (another such example?)

(after seeing how important normal subgroups are, here are a couple of properties about them:)

Proposition 3.1.9: properties of normal sub-groups

Let N be a normal sub-group of a group G. Then the following are equivalent

1. $N \trianglelefteq G$
2. $N_G(N) = G$
3. $gN = Ng$
4. the operation on left cosets of N in G as described in the previous proposition turns the set of cosets into a group.
5. $gNg^{-1} \subseteq N$ for all $g \in G$
6. if G is abelian and $H \leqslant G$ then $H \trianglelefteq G$

Proposition 3.1.7 is very important. The equivalence relatoin $\sim$ corresponding to H has both property 3.1 and 3.2 by proposition 3.1.6 and so by proposition 3.1.3 G/H is a group! Since normal subgroups are so important, we'll use the notation $\trianglelefteq$ instead of $\leqslant$ to say that H is a normal subgroup of G ($H \trianglelefteq G$) Given our new found knowledge, we will give a new definition of quotient group

Definition 3.1.10: Quotient Group

Let $H \trianglelefteq G$ be a normal subgroup of G. The *quotient group* G/H, sometimes written G mod H (and read G modulo H or G mod H for short) is the group $(G/\sim, \bullet)$ where $\sim$ is defined in proposition 3.1.6, and $\bullet$ is defined by

$$aHbH := (ab)H$$

An easy way to remember that this operation is defined when is H is normal is by remembering $a(Hb)H = abHH = abH$. For notation, we can write $aH = \bar{a} = [a]$ for some representative u and $G/H = \bar{G}$. Thus, going all the way back to $\pi : G \to G/\sim$, and use our knowledge to re-write this as Thus, going all the way back to $\pi : G \to G/\sim$, and use our knowledge to re-write this as

$$\pi : G \to H \qquad \pi(g) = gH$$

where π is a homomorphism if $H \trianglelefteq G$. If $H \ntrianglelefteq G$, then π is *not* a homomorphism.

Since π is a homomorphism, it's kernel is a subgroup:

$$\ker(\pi) = \{g \in G \mid \pi(g) = eH\}$$

and since $eH = H$, this means that $\ker(\pi) = H$. Since H can be any normal subgroup, this means that H is always the kernel of some map

$$\text{Normal} \Rightarrow \text{Kernel}$$

Is the converse true, that is, the kernel is always normal? If so, that would be really powerful, since the kernel would then always be a normal subgroup of G, and so any homomorphism $\varphi : G \to H$ would induce a quotient group G/K, where $K = \ker(\varphi)$. It turns out that kernels are always normal subgroups!

Proposition 3.1.11: Kernel Then Normal

Let $\varphi : G \to H$ be a homomorphism and let $K = \ker(\varphi)$. Then $K \trianglelefteq G$

***Proof* :**
Let K be the kernel of $\varphi : G \to H$, and take any $g \in g$. then

$$\varphi(gkg^{-1}) = \varphi(g)\varphi(k)\varphi(g^{-1}) = \varphi(g)\varphi(k)\varphi(g)^{-1} = \varphi(g)\varphi(g)^{-1} = e_H$$

and so $gkg^{-1} \in K$, showing that $gK = Kg$, meaining $K \trianglelefteq G$

This means that given any homomorphism $\varphi : G \to H$, we can *always* define an equivalence relation on G given the kernel of φ since $\ker(\varphi)$ is normal!

Since any normal subgroup will be the kernel of some homomorphism, and the kernel of any homomorphism will be normal, we get that the two are essentially synonymous in group theory:

$$\text{Normal} \Leftrightarrow \text{Kernel}$$

(A word how with this fact, we can find the structure of G/H by relating it to the image of $\varphi(G)$?)

With our new knowledge on quotient groups, we re-visit the universal property of quotients (theorem 3.1.1) with the intend on being able

(words here on bringing back intuition from uni. prop. of set, but now in Grp it's more prevalent, since we're linking group structures. Notice that since $H \subseteq \ker(\varphi)$, that means H will make a "finer" as compared to $\ker(\varphi)$, and since both will respect the corresponding $\sim_H$ induced by H, we know by theorem 3.1.1 that it must factor through G/H. THis result is important enough to be given a name:)

Theorem 3.1.12: The First Isomorphism Theorem

Let $\Phi : G \to H$ be a homomorphism and let $K = \ker(\varphi)$. If $N \trianglelefteq G$ where $N \subseteq \ker(\varphi)$, then there exists a unique function Φ such that the following diagram commute:

$$\begin{tikzcd} G/N \arrow[r, "\Phi"] & H \\ G \arrow[u, "\pi"] \arrow[ur, "\varphi"'] & \end{tikzcd}$$

where $\Phi = \varphi \circ \pi$, and π is the natural projection

Proof **:**
We already know this universal property in terms of congruence relations. We simply must translate the information of this theorem to the one required for theorem 3.1.1

So far, the question is saying that if

$$N \subseteq \ker(\varphi) \Leftrightarrow (\forall n \in N,\ \varphi(n) = e_H) \Rightarrow \varphi(ab^{-1}) = e_H$$

for $a, b \in N$ moving things around we get

$$\varphi(ab^{-1}) = \varphi(a)\varphi(b)^{-1} = e_G \ \Leftrightarrow\ \varphi(a) = \varphi(b)$$

thus giving us

$$(\forall a, b \in N),\ ab^{-1} \in N \Rightarrow \varphi(a) = \varphi(b)$$

and, since $ab^{-1} \in N$ is the condition defining the induced equivalence relation by N (proposition 3.1.6), we get

$$a \sim b \Rightarrow \varphi(a) = \varphi(b)$$

which put's us exactly at the inition condition for theorem 3.1.1, which gives us that the diagram commutes, as we sought to show.

What theorem 3.1.12 (the universal property of quotient groups) tells us is that all the information a homomorphism φ gives us can be captured via the quotient group G/K (TBD?)

Corollary 3.1.13: Canonical Decomposition For Groups

Let $\varphi : G \to H$ be any group homomorphism. Then it may be decomposed into

$$\begin{tikzcd} G \arrow[r, two heads, "\pi"'] \arrow[rrr, bend left, "\varphi"] & G/\ker(\varphi) \arrow[r, "\cong", "\Phi"'] & \mathrm{im}(G) \arrow[r, hook, "\iota"'] & H \end{tikzcd}$$

where Φ is an isomorphism

Proof **:**
This proof is simply piecing together all the results with figured out so far. Since $\operatorname{im}(G) \leqslant H$, then there exists an inclusion map ι which is a homomorphism. We've established that $\ker(\varphi)$ is normal, and that any normal subgroup will induce a quotient group $G/\ker(\varphi)$, meaning π (the natural projection) is a well-defined homomorphism.

To show $\Phi : G/\ker(\varphi) \to \operatorname{im}(G)$, is an isomorphism, note that we already have that Φ is given by the canonical decomposition of quotient sets. By the universal property of quotient groups, we have that it's a homomorphism. Since bijective homomorphisms are isomorphisms, we have that it's an isomorphism, completing the proof.

Corollary 3.1.14: Surjective Homomorphisms

Let $\varphi : G \to H$ be a surjective homomorphism, and let $K = \ker(\varphi)$. Then

$$G/K \cong H$$

Proof **:**
For $G/K \cong H$, notice that $H = \operatorname{im}(G)$ in corollary 3.1.13.

This last corollary is very useful and one you will use constantly in both theoretical and practical context. It is saying that if we want to know the structure of G/K, if we can find a surjective homomorphism from G to H with kernel K, then $G/K \cong H$, i.e., H will represent the structure!

(Integarte this somehow?? Maybe later)

$$1 \xrightarrow{\iota_1} K \xhookrightarrow{\iota_K} G \overset{\varphi}{\twoheadrightarrow} H \xrightarrow{\pi} 1$$

where 1 is the trivial group, ι_1 and ι_K are the inclusion homomorphisms, and π is the projection of H onto the trivial group.

The 2nd form presented in corollary 3.1.14 lets us nicely split up the pieces of any given homomorphism. This representation is called a *short exact sequence*. We will return to short exact sequences soon to flesh out many intersting results when precieving which groups can be put in the 3 non-trivial positions, that is

$$1 \xrightarrow{\iota_1} A \hookrightarrow B \twoheadrightarrow C \xrightarrow{\pi} 1$$

(bunch of examples: perhaps a reminder that if $H \leqslant G$ the $i : H \to G$ shows that G contains H??)

Exercise 3.1.1

Throughout these exercises, let $\varphi : G \to H$ be a homomorphism between the two groups G and H, and $K = \ker(\varphi)$

1. Show that $\varphi^{-1}(Q)$ (the pre-image of Q) is a group. If Q was a normal subgroup, $Q \trianglelefteq H$, show that $\varphi^{-1}(Q)$ is also normal. Is it true that if $H \trianglelefteq G$ then $\varphi(H) \trianglelefteq \varphi(G)$?
2. Let $h \in H$. We call $\varphi^{-1}(h)$ a *fiber* of φ over h. Let $F_h = \varphi^{-1}(H)$. Show that
    1. For any $u \in F_h$, $F_h = uK$

2. For any $u \in F_h$, $F_h = Ku$

3. Let $aKbK = \{akbk' \mid a, b \in G, k, k' \in K\}$ be a set, and $K \trianglelefteq G$. Show that

$$aKbK = abK$$

where $abK = \{abk \mid a, b \in G, k \in K\}$ [hint: for $aKbK \subseteq abK$, recall that $aK = Ka$ since K is normal. For $aKbK \supseteq abK$, note that $1 \in K$ so $1 \cdot a \in Ka$. Take advantage of normality to complete the proof]

4. Let $R \leqslant G$. Show that the cosets of G/R are all of equal size.

5. Show that if $aB = 1B$, then $a \in B$

6. Let $H, K \trianglelefteq G$. Show that $H \cap K \trianglelefteq G$

7. Generalizing the previous result, If $\{H_i\}_{i \in I}$ is some collection of normal subroups of G show that $\cap_{i \in I} H_i \trianglelefteq G$

8. Let G be finitely generated and let $N \trianglelefteq G$ be finitely generated. Prove that G/N is finitely generated.

9. For any group G, show that $Z(G) \trianglelefteq G$

10. Show that $n\mathbb{Q} = \mathbb{Q}$, where we're abusin' notation in writing $n\mathbb{Q}$ to mean $\{nq \mid q \in \mathbb{Q}\}$ even though multiplication is not defined on $\mathbb{Q}$ (Groups that have this property are called *divisible* or *injective* groups. More on them in section 16.3) [maybe any divisible group has a natural 2nd group structure? cause in a sense, every element can be "divided", and so it respects the property of inverse, which is also unique by the fact that the element is a unique group element, and the other axioms follow naturally]

11. Let $G \cong H$, $N \trianglelefteq G$ and $K \trianglelefteq H$. If $N \cong K$ is it true that $G/N \cong H/K$? Prove or show a counter example. hint: [2]

12. (Useful in Lie Groups) Let $GL_n(\mathbb{R})$ be the set of invertible $n \times n$ matrices over $\mathbb{R}$ (more generally, if you know linear algebra, over a field $\mathbb{F}$ with characteristic $\neq 2$ or 3 if $n = 2$). What are the normal subgroups of $GL_n(\mathbb{R})$?

13. We've shown that if we have a map $f : G \to H$ and $N \trianglelefteq H$, we can define a map $\overline{f} : G \to H/N$. Is the converse true? That is, if we have a map $f : G \to H/N$, do we have a map $F : G \to H$ such that $f = \pi \circ F$? [Hint: Consider maps with domains and codomains of $\mathbb{Z}$ and $\mathbb{Z}/n\mathbb{Z}$, $n > 1$]

14. Show that every subgroup of Q_8 is normal. Hence, every subgroup being normal does not imply the group is abelian[3]

[2] reverse text? This might be giving the answer? Consider $G = H = N = \mathbb{Z}$, $K = 2\mathbb{Z}$

[3] On the other hand, Q_8 is special: Every non-abelian group for which all subgroups are normal is of the form $G \cong Q_8 \times B \times D$ where B is something called an elementary abelian 2-group and D is called a torsion abelian group whose elements are of odd order. A group where all subgroups are normal is called a Dedekind group, and a non-abelian Dedekind group is called a Hamiltonian group.

3.2 Presentations

We have seen in section 2.3.3 that if we have any set map from a set S to a group G, $f : S \to G$, then it can be extended it to a homomorphism from the group $F(S)$ to G: $\varphi : F(S) \to G$, $\varphi|_S = f$. We've also seen that if we have a surjective homomorphism $G \to H$, then $G/N \cong H$ (where N is the kenrel of the homomorphism). We can combine these results: Converting to $(G, *)$ notation for groups for a moment, given an identity map $\mathrm{id} : G \to (G, *)$, which is a prefectly valid surjective set function from the set G to the underlying $\varphi : (F(G), *_F) \to (G, *)$. By the first isomorphism theorem, if N is the kernel, then $(F(G)/N, (*_F)_\varphi) \cong (G, *)$. Thus, $\varphi : F(G) \to (G, *)$. By the first isomorphism theorem, if N is the kernel, then $F(G)/N \cong (G, *)$. Thus, every group is the quotient of a free group! We now go back to our usual convention of groups being written as G where we drop the binary operation, where we will consider it clear from context when a map is a set map and when it is a homomorphism.

Let's generalize this result a bit: Let's say $S \subseteq G$ such that $\langle S \rangle = G$, and consider $f : S \to G$. Then by the universal property of free groups, we have that $\varphi : F(S) \to G$ is a group homomorphism. Since $\langle S \rangle = G$, and all possible words are in $F(S)$, then by construction of $F(S)$ and φ, the map φ is a surjective group homomorphism. Thus, for the appropriate kernel N, $F(S)/N \cong G$.

Since $F(S)$ is free from relations, but G has relations, φ somehow imposed relations onto $F(S)$. This means that N captured the idea of the relations imposed on $F(S)$ to make $F(S)/N \cong G$. As observed in section 2.3, any relations of a group can be written as

$$r_1 = r_2 = \cdots = r_k = 1$$

If we let $r_i = \prod_{i \in I_{r_i}} s_i$ to be some arbitrary finite[4] product of elements of S and it's inverses. Then if If we let $r_i = \prod_{i \in I_{r_i}} s_i$ to be some arbitrary product of elements of S and it's inverses. Then if we let $R = \{r_1, r_2, ..., r_k\}$, we can write $G = \langle S | r_1 = r_2 = \cdots = r_k = 1 \rangle$ as

$$G = \langle S | R \rangle$$

for short hand. Since $F(S)/N \cong G$, it must be that $R \subseteq N$. We might be tempted to say $\langle R \rangle \subseteq N$, however the closure *need not be normal*, while N must be a normal subset. Instead, we'll define the noraml closure:

Definition 3.2.1: Normal Closure

Let $S \subseteq G$ be a subset of a group G. Then define the intersection of all normal subgroups of G containing S

$$\langle S \rangle^{\trianglelefteq} = \bigcap_{\substack{N_i \trianglelefteq G \\ S \subseteq N_i}} N_i$$

to be *normal closure of S*.

$\langle S \rangle^{\trianglelefteq}$ is indeed a normal subgroup by proposition ref:HER. Since $r_1 = r_1 = \cdots = r_k = 1$ in G, then $[r_1] = [r_2] = \cdots = [r_k] = [1] \in F(S)/N$, and so $r_1, r_2 ... r_k \in N$. We thus have that $\langle R \rangle^{\trianglelefteq} \subseteq N$.

Next, let's say we have some $x \in N$. Then that means then $[x] = [1]$, so for some $n' \in N$, $xn' = 1$. But then, xn' must be some combinations of elements $r_1, r_2, \cdots, r_k$, since by the presentation of G, those are the elements that can equal to 1. Thus $N \subseteq \langle R \rangle^{\trianglelefteq}$. Therefore $N = \langle R \rangle^{\trianglelefteq}$

[4] Infinite products would require more care to be well-defined, see EYNTKA real analysis I

From this, we can formally define a presentation

Definition 3.2.2: Presentation

Let G be any group. Then a *presentation* of G is a pair of sets (S, R) where $\langle R \rangle^{\trianglelefteq}$ is the kernel of the homomorphism $F(S) \to G$. The elements of S are called the *generators* of G, and the elements of R are called the *relations* of G:

$$\langle S|R \rangle := F(S)/\langle R \rangle^{\trianglelefteq}$$

We say G is finitely generated if given a presentation (S, R), $|S|$ is finite. We say G is *finitely presented* if both $|S|$ and $|R|$ are finite.

Notice how this statement is like the dual equivalent of Cayley's theorem (4.4.2): In Cayley's theorem, every group is the subgroup of a symmetric group. Here, every group is the quotient of a free group.

Example 3.2.1

1. $\langle A|\emptyset \rangle$ is the presentation of the free group $F(A)$
2. dihedral group

Note: Let $N \trianglelefteq G$. Clearly, $\langle N \rangle^{\trianglelefteq} = N$. However, it is rather difficult to determine whether $S \subseteq N$ is a smallest set that satisfies $\langle S \rangle^{\trianglelefteq} = N$. (More words about the word problem?)

Corollary 3.2.3: Finitely Generated Groups

Let G be a group. Then G is finitely generated if and only if G is isomorphic to the quotient group of a finitely generated free group.

Proof **:**
Let $F(S)/N \cong G$. If $F(S)$ is finitely generated, then $F(S)/N$ is finitely generated (see exercise ref:HERE) (Though in that exercise, $N \trianglelefteq G$ is finitely generated given the regular closure, not the normal closure... must think about that)

On the other hand, let G is finitely generated, say $S \subseteq G$ such that $\langle S \rangle = G$ and $|S| = n$. Then there exists an inclusion map $S \hookrightarrow G$. By the universal property of free groups, this extends to a map $\varphi : F(S) \twoheadrightarrow G$. This map is surjective since S is a generating set for G. Then by the 1st isomorphism theorem, $F(S)/\ker(\varphi) \cong G$, fulfilling our requirement.

(ref:HERE if you do geometric group theory or if this is referenced in other places)

In this book, this will be the end our theoretical journey at looking at groups as $F(S)/N \cong G$. A whole field of study has emerged because of this universal property called *geometric group theory*, which we do some exposition for in the appendix (see ref:HERE). For those interested in geometric group theory, I recommend (ref:HERE, perhaps Clara Löh's book and that othe guy whose book is

harder).

3.3 Application to Finite Groups

As mentioned earlier, thinking about cosets in themselves is also very useful. In fact, a theorem which you will you to death in this course relies on cosets. With the fact that cosets partition your group, you can deduce this strikingly important result:

Theorem 3.3.1: Lagrange Theorem

If G is a finite group and H is a subgroup of G, then the order of H divides the order of G (i.e., $|H|\big||G|$), and the number of left cosets of H in G equals $\frac{|G|}{|H|}$

Proof **:**
Let $H \leqslant G$. Then you know that H partitions G into equal-sized cosets by proposition ref:HERE (it is an exercise), and so if $|G| = n$, and $|H| = k$, then you know $k|n$, implying $|H|\big|G$. Furthermore, since if $H \trianglelefteq G$, then

$$|G/H| = |G|/|H|$$

Example 3.3.1

1. As a quick reminder, we used this theorem in (*corollary* 2.3.9)to get a useful result a bit early. You can now go back and perhaps check your intuition to make sure it makes sense.
2. Try going back to (*theorem* 2.3.8) and proving it using Lagrange's Theorem. You will find it much easier, almost trivial even.

As the examples show, Lagrange is incredibly powerful as it let's us relate the order of a group and it's subgroup, which is a fact used extensively.

Definition 3.3.2: Index

If G is a group (possibly infinite) and $H \leq G$, the number of left cosets of H in G is called the index of H in G and is denoted by $|G : H|$.

With this notation, we can re-write Lagrange's Theorem as

$$|G| = |H||G : H|$$

Corollary 3.3.3: x to the order of G

If G is a finite group and $x \in G$, then the order of x divides the order of G. In particular $x^{|G|} = 1$ for all $x \in G$.

Corollary 3.3.4: index of H is 2 $\Rightarrow$ H is normal

If The $H \leqslant G$ and the index of H is 2, then H is normal

***Proof* :**
Since $[G : H] = 2$, we know that there are only 2 cosets: eH and xH for the correct $x \in G$. We want to show that H is normal, that is, for every element $g \in G$, $gH = Hg$ (which in our case there is only e and x which form distinct cosets) . Since $eH = He$, by commutativity of identity element from last homework, it sufices to show that

$$xH = Hx$$

for this, we'll take advantage of the index of H. Since the index of H is 2, we know that

$$H \cup xH = G \qquad \text{and} \qquad H \cup Hx = G$$

since the cosets of H partition G, as we showed in part 1 of this homework. Since $xH \subset G$, we know that

$$xH \subset H \cup Hx$$

therefore

$$\forall h \in H, \text{ either } xh = h'x \qquad \text{or} \qquad xh = h' \qquad \text{for some appropriate } h' \in H$$

let's say that for the sake of contradiction $xH \neq Hx$, so we can rule out the first option for a moment, leaving us with $xh = h'$. However, that would imply that $xh \in H$, which is means xh is in the first coset. However, we've shown in the first question of this homework that cosets are distinct, making this a contradiction. Therefore, we're forced to say $xh = h'x$ for appropriate h', however that implies that $xh \in Hx$. Since this holds for any $h \in H$, and we can repeat the same process for Hx, we have that

$$xH = Hx$$

meaning that all cosets "commute", and so G is normal, as we sought to show.

13: Why this doen't necessarily work when index greater than 2

Counter-example: Take $G = D_3$, i.e. the triangle group, and $H = e, \sigma$. Then

$$\{e, \sigma\} \cup \{r, r\sigma\} \cup \left\{r^2, r^2\sigma\right\} = G$$

and

$$\{e, \sigma\} \cup \{r, \sigma r\} \cup \left\{r^2, \sigma r^2\right\} = G$$

i.e. the union of the left and right cosets both must equal G. This idea is very important as it forces a relation between left and right cosets. now consider rH. Since the unions of both sets must match, we know that

$$rH \subseteq G \Rightarrow rH \subseteq H \cup Hr \cup Hr^2$$

we knwo that $rH \nsubseteq H$ since that would make the two cosets *not* distinct, and so we have

$$rH \subseteq Hr \cup Hr^2$$

which is true! test it out (while recalling that $r\sigma = \sigma r^2$). In the case we since the index being 2 situation, that would make $xH \subseteq Hx$, and vice versa for Hx too, and so it forces normality.

Because of the power of Lagrange, one might ask themselves if the converse is true: if a number divides the order of a group $n \big| |G|$, does that mean there's a subgroup $H \leqslant G$ such that $|H| = n$? That turns out to be not true:

Example 3.3.2: : Counter-Example

We'll show that $6 \big| |A_4| \not\Rightarrow H \leqslant A_4, |H| = 6$. First, let's figure out the elements of A_4:

1. the identity
2. all even permutation with max order of 4: $(\cdot\,\cdot)(\cdot\,\cdot)$
3. Either the 2 transpositions are disjoint, or they overlap with 1 element. They cannot overlap with 2 elements because then $(\cdot\,\cdot)(\cdot\,\cdot) = (\cdot\,\cdot)$, making it an odd-permutation.
4. If they overlap with 1 element, then $(\cdot\,\cdot)(\cdot\,\cdot) = (\cdot\;\cdot\;\cdot)$, thus we have all 3-cycles, and all permutation represented by 2 disjoint 2-cycles
5. By counting, we can find 8 3-cycles, and 3 $(2, 2)$-cycle types, for a total of 12 cycles (when including the identity), meaning $6 \big| |A_4|$

Let's now suppose there exists a subgroup $H \leqslant A_4$ of order 6. To analyze it we'll take advantage of the fact that the intersection of two subgroups is a subgroup (2.3.10). If you take all $(2, 2)$ cycle type elements, plus the identity, you will get a group $(K \leqslant A_4)$ [a]. Take

$$H \cap K = H' \qquad \Rightarrow \qquad H' \leqslant H\,, H' \leqslant K$$

By Lagranges Theorem, $|H'| \big| |H|$ and $|H'| \big| |K|$ which implies the order is either 1 or 2. If $|H'| = 1$, then it only has the identity. That would mean that H and K are practically disjiont, meaning

that if we take 2 non-identity elements $h \in H, k \in K$, then $h \cdot k = g$ we cannot produce an element $g \in H$ or $g \in K$ since then by closure, we'd get $k \in H$ and $h \in K$. Doing this for the case of H:

$$\begin{aligned} hk &\in H \\ \Rightarrow h^{-1}hk &\in H \\ \Rightarrow k &\in H \\ \Rightarrow H \cap K &\supseteq \{e\} \end{aligned}$$

Thus, we'd get $|H| \cdot |K|$ total elements. But $|H| \cdot |K| = 6 \cdot 4 = 24$, which is greater than the number of elements we have: $|A_4| = 12$. Thus, $|H'| \neq 1$, so we must have that $|H'| = 2$. This must also imply that H has an element $h \in H$ that is both a product of two disjoint cycles, while the other 4 elements are 3-cycles.

If $|H'| = 2$, then we'll show that it will also lead to a contradiction. Note that since $|H| = 6$, then $|A_4 : H| = 2$, so by corollary (3.3.4), $H \trianglelefteq A_4$, meaning $\forall \sigma \in A_4$, $\sigma H = H\sigma$, or equivalently $\sigma H \sigma^{-1} = H$

Now, take take the cycle of type $(2, 2)$ in H from earlier: $h = (i\,j)(k\,l)$ with $\{i, j, k, l\} = \{1, 2, 3, 4\}$ and the 3-cycle $\sigma = (i\,j\,k) \in A_4$. Then we have that

$$\sigma h \sigma^{-1} = (j\,k)(i\,l) \neq (i\,j)(k\,l)$$

And since H is normal, $\sigma k \sigma^{-1} \in H$. However, that would imply that there is a *second* cycle of type $(2, 2)$, meaning that $|H \cap V| \geq 3$, again a contradiction.

Therefore, there cannot be a subgroup of order 6 in A_4, as we sought to show

[a] Trivia: This group is isomorphic to the Klein 4-group V_4

We showed how the converse is not true, but are there partial converses that are true? There are actually two famous theorems which talk about partial converses: *Cauchy's Theorem* and *Sylow's Theorem*. Both these theorems rely on prime-number properties to give us a partial converse. Sylow's theorem goes a step further in telling us how many groups of a particular order we can find (so more than there exists one). However, by the way we'll present Sylow's theorem, it will require group-actions, the topic of the next section. However, Cauchy's theorem technically doesn't. There is an exercises in the book for doing this theorem, so I'll link it here for now for when I have time to do it:

Theorem 3.3.5: Cauchy's Theorem

If G is a finite group and p is prime diving $|G|$, then G has an element of order p

Proof **:**
proof outlined in exercise 9.

There is a simpler proof which is related to another theorem called Sylow's Theorem (4.4.13) in a similar way that the Inverse Function Theorem and Implicit function Theorem are related: proving one of them is relatively hard, but then proving the other from the first is a super easy consequence. Since Sylow's theorem *is* proved and covered extensively, I proved this theorem using Sylow's theorem once it is introduced (see 4.4.16)

Cauchy's theorem is super important. It pop's a lot when trying to find groups of smaller order:

Example 3.3.3
If G is a finite abelian group with no normal subroups (besides 0 and G)[a], then $G \cong Z_p$.

Proof. Since G is abelian, if $H \leqslant G$, then $H \trianglelefteq G$ (since $aH = Ha$ by use of commutative property). However, G is simple, so either $H = \{e\}$ or $H = G$. By Cauchy's theorem, we also know that if $p \big| |G|$ (which once must as $|G|$ is composed of prime-factors), then there exists an element of order p (meaning a subgroup of order p). However, that's a contradiction as we've just shown. Therefore, either $G = \{e\}$ or $G = \langle x \rangle$ where $|G| = p$. Either way, G, is cyclic and abelian, which by 2.3.6 means that

$$G \cong Z_p$$

□

[a] maybe introduce simple groups earlier?

Note that G was a general finite group. If we have that G is abelian, then the converse is actually *completely* true! This turns out to simply be by the fact that if $\langle x \rangle = p$ and $\langle y \rangle = q$ for prime p, q then $\langle xy \rangle = pq$

Corollary 3.3.6: Cauchy's Theorem For Abelian Groups

If G is a finite abelian group, and n divides G, then there is a group of order n

First, we prove it for if $n = p$ is prime

Proof **:**
p.102 in book

Now let's prove it in general

Proof **:**
We'll for the first time proceed by induction! If $|G| = 1$, we're done, so let's say $|G| = n$, and if there is a an element $m < n$ such that $m \big| |G|$ then there exists an $H \leqslant G$ such that $|H| = m$. The edge cases are the same as Cauchy's theorem, so they'll be omitted.

By Cauchy's theorem, we know that if $p \big| |G|$ then $P \leqslant G$ such that $|P| = p$. Since $|G| > 1$, we can assume such a p exists. Then since G is abelian, $P \trianglelefteq G$, so G/P is defined. In particular

$$|G/P| = n/p$$

Which means we found our smaller group G/P. Now for any $m \big| |G/P|$, there exists a $\overline{H}$ such that $\overline{H} = m$. By the 4th Isomorphism theorem, we know there exists a $H \leqslant G$ such that

$$H = H/N$$

meaning that $|H| = \frac{n}{p} \cdot p = n$, showing that there is group of order H, as we sought to show.

Here's also another proof I found

Let G be a finite abelian group. By Cauchy's theorem, it has a subgroup of order p for any prime divisor of $|G|$. We use mathematical induction on the factors of $|G| = n$. *Suppose* that there was a subgroup of order k, for some factor k of n.

We will show that if $kp \mid n$, then there is a subgroup of order kp, for a prime number p. By the induction hypothesis, there is a subgroup H of order k. Now, since $|G/H| = |G|/k$ is still divisible by p, G must have another element of order p not contained in H. Otherwise the quotient

$$G/H$$

would have order divisible by p and yet no element of order p! (Notice that the order of an element $\overline{g} \in G/H$ must divide the order $g \in G$ so no other element can give us an element of order p.)
This shows that there is an element x of order p not in H. But then

$$\langle x \rangle H$$

has order $|\langle x \rangle H| = |x||H| = kp$ (since the group is abelian) and so there is a subgroup of order kp. This completes the inductive step.

Figure 3.1: Alternative proof

This theorem should show you the power of commutativity! If a group is commutative, you have an an incredibly simple group structure! If you know the order of the group, you know the entire substrucutres of the group, giving you a lot of information! This very fact will be key when categorizing every abelian group in chapter 5! This means that a good chunk of the groups studied in group theory are non-commutative groups, since the structure of commutative groups is pretty simple and well-understood.

Be mindful that Lagrange's theorem gives you information about *finite* groups. In general, arithmetics on the infinite won't be defined in a meaningful way for us. In fact, we can get quite some weird results, like subgroups being isomorphic to their parent group!

Example 3.3.4: Subgroups Of $\mathbb{Z}$
Take the group $\mathbb{Z}$, and any subgroup $n\mathbb{Z} \leqslant \mathbb{Z}$ for $n \in \mathbb{N}_+$. Show that

$$n\mathbb{Z} \cong \mathbb{Z}$$

that is, every (non-trivial) subgroup of $\mathbb{Z}$ is isomorphic to $\mathbb{Z}$!

3.4 Internal Direct Product

Though it might seem inconspicuous at first, the introduction of a normal subgroup will let us be able to define the multiplication of subgroups that from subgroups in themselves!! The idea behind this idea is to decompose a group into smaller subgroups, giving an insight on its structure. In chapter 5, we will explore in greater detail the results of this decomposition. For now, we will explore a couple

of properties of internal direct product.

We'll first define some notation:

Definition 3.4.1: HK (Internal Direct Product)

Let H, K be subgroups of a group and define

$$HK = \{hk | h \in H, k \in K\}$$

Note: This is *not* necessarily a group! Take $\langle (123) \rangle = H \leqslant S_4$ and $\langle (3\,4) \rangle = K \leqslant S_4$. Then

$$HK = \{\text{id}, (123), (132), (123)(34), (132)(34), (34)\}$$

and $(132)(34) = (3421)$, but $(3421)^2 \notin HK$, meaning HK is not closed!

Proposition 3.4.2: cardinality of HK

Let H, K be finite subgroups of a group. Then

$$|HK| = \frac{|H||K|}{|H \cap K|}$$

Proof **:**

Notice that HK is the union of left-cosets of K:

$$HK = \bigcup_{h \in H} hK$$

Thus, it suffices to find the number of distinct left-cosets hK.

If $h_1K = h_2K$ then $h_2^{-1}h_1K = K \Rightarrow h_2^{-1} \in K$. Since $h_2^{-1}h_1 \in H$, this means that

$$h_2^{-1}h_1 \in H \cap K \Rightarrow h_2^{-1}h_1H \cap K = H \cap K \Rightarrow h_1H \cap K = h_2H \cap K$$

And so we reduced the problem to counting the number of distinct $H \cap K$ cosets. This is useful, since $|H \cap K| \big| H$, and so by Lagrange's theorem

$$|H : H \cap K| = \frac{|H|}{|H \cap K|}$$

Thus, there are that many left cosets of K. Since K partitioned HK, it means that every element of the cosets are unique, so to find $|HK|$ we now get

$$|HK| = \frac{|H|}{|H \cap K|}|K| = \frac{|H||K|}{|H \cap K|}$$

as we sought to show

Proposition 3.4.3: When is HK a group

If H and K are subgroups of a group, HK is a subgroup if and only if $HK = KH$

Proof **:**
Let $H, K \leqslant G$.

($\Rightarrow$) Let's say $HK \leqslant G$. We want to show that $HK = KH$, that is, for any $hk \in HK$, there exists an $h' \in H$, $k' \in K$ such that $hk = k'h'$. Since HK is a group, it is closed under its operation. Thus, for any $h_1k_1, h_2k_2 \in HK$, $h_1k_1h_2k_2 \in HK$. Since $HK = \langle hk | h \in H, k \in K \rangle$, there is a $h_3 \in H$ and $k_3 \in K$ such that

$$h_1k_1h_2k_2 = h_3k_3$$

re-arranging, we get that

$$h_3h_1k_1 = k_3k_2^{-1}h_2$$

By letting $h = h_3h_1$, $k = k_1$, $h' = h_2$, and $k' = k_3k_2^{-1}$, we get that

$$hk = k'h'$$

as we sought to show (more words on how we can choose these values and still let h, h', k, k' be "arbitrary")

($\Leftarrow$) Let's say $HK = KH$. We want to conclude that HK is a subgroup of G. We'll show that HK is closed under its binary operation – the other axioms follows straightforwardly.

Corollary 3.4.4: Condition on K when Internal Direct Product is Present

If H and K are subgroups of G and $H \leq N_G(K)$, then HK is a subgroup of G. In particular, if $K \leqslant G$ then $HK \leqslant G$ for any $H \trianglelefteq G$. Furthermore, if H and K are normal subgroups, then $HK \trianglelefteq G$.

Proof **:**
The first part of the corrollary was proven in the last proposition. What is left to prove if both H and K are normal subgroups, then $HK \trianglelefteq G$. (TBD – is this true?)

Defining when HK is a group is intersting, but it would be even more useful if we can figure out the structure of HK. For example, Can we say that $HK \cong H \times K$? The answer is not always! In fact, we will find a way of exactly figuring out the structure of HK in chapter 5 once we have learned the appropriate tools in chapter 4

14: Foreshadowing

Thinking about HK like in corollary 3.4.4 will become the defining feature when we discuss *semi-direct* products in chapter 5.

Since we have now defined how we can take the "product" of two groups, we can ask the following

intersting question: If $\mathcal{N}$ is some subset of normal subgroup G that is "closed" under multiplication, does it form a group? The reader should fiddle around with this before the answer is given. The same question will be asked in the next part of the book where we will have the equivalent of normal subgroup for rings (an algebraic structure with 2 binary operations). It with the "normal" objects of rings. The answer for groups shall be presented in chapter 7. with the "normal" objects of rings. In the case for groups, such a collection will not form a group.

Thus, we will leave internal direct product alone, with only an exception made for the next section, until we have accumulated more tools to discuss them again.

3.5 Isomorphism Theorems

We know turn back to how quotient groups are inseparably related to homomorphisms. Mathematicians who were studying homomorphisms wanted to understand the relation between groups given certain properties, for then we can gain new perspective into problems being solved. For example, if we have a group $HK \leqslant G$, and $K \trianglelefteq G$, can we always determine what HK/K is? Is there *always* an answer to this question, no matter what group HK we pick? Another question is what if we "recusively" take quotients: If $H \trianglelefteq K \trianglelefteq G$, what is $(G/H)/(K/H)$? Is there a simplification that will make this more understandable? Yet another such question is how much can we deduce of the subgroups of G/N (i.e. the lattice of G/N) given what we know about G?

In this section, we'll show these results:

1. We'll sum up the results of of theorem 3.1.12, but give it the name that is more commonly used
2. We will look at equivalent ways of looking at internal direct products, since they pop up a lot (especially chapter 5)
3. An easier way of thinking of quotient groups whose quotient is a quotient group.
4. the subgroup lattice of G and the subgroup lattice of G/N

We first recall theorem 3.1.12, but given it its classical name:

Theorem 3.5.1: First Isomorphism Theorem

If $\Phi : G \to H$ is a homomorphism of groups, then $\ker(\Phi) \trianglelefteq G$ and $G/\ker\Phi \cong \Phi(G)$.
Similarly, If $N \trianglelefteq G$, we can say there exists a unique function φ such that the following diagram commute:

$$\begin{tikzcd} G \arrow[r, "\pi"] \arrow[dr, "\Phi"'] & G/N \arrow[d, "\varphi"] \\ & H \end{tikzcd}$$

where $\Phi = \varphi \circ \pi$, and π is the natural projection

Recall that the powerful part of this theorem is that it is the universal property of quotient groups: for *any* homomorphism $\varphi : G \to H$, there exists a unique group G/N (which represents the equivalence class, and thus the homomorphism, within its group structure) along with the unique homomoprhism

π [5]which let's us link G to G/N such that $\Phi = \pi \circ \varphi$, meaning we can study the results of the homomorphism Φ by looking at group information of G/N.

Also, remember that if we take $\Phi(G) \leqslant H$, and $N = \ker(\Phi)$ so that

$$\begin{tikzcd} G \arrow[r, "\pi"] \arrow[dr, "\Phi"'] & G/\ker(\Phi) \arrow[d, "\varphi"] \\ & \Phi(G) \end{tikzcd}$$

then

$$G/\ker(\Phi) \cong \Phi(G)$$

Example 3.5.1

1. we can get some information about S_n. We can define the function $\epsilon : S_n \to \{-1, 1\}$ whether if σ is even then $\epsilon(\sigma) = 1$, and if σ is odd then $\epsilon(\sigma) = -1$. Note that we can treat $\{-1, 1\}$ as a group if $-1 \cdot -1 = 1, 1 \cdot -1 = -1, -1 \cdot 1 = 1, 1 \cdot 1 = 1$. Then ϵ would actually be a homomorphism. Then $\ker(\epsilon) = A_n$, that is, the cycles of even permutation, also known as the alternating groups. By the first isomorphism theorem, we then know that

$$S_n/A_n \cong \epsilon(S_n) = \{\pm 1\}$$

and so, using Lagrange, we get that

$$2 = |S_n/A_n| = \frac{|S_n|}{|A_n|} \qquad \Rightarrow \qquad |A_n| = \frac{1}{2}|S_n|$$

2. Let's say we have two groups G, H with $\varphi : G \to H$ being a surjective homomorphism. Then we know that

$$G/\ker(\varphi) = H$$

in other words, if we have a surjective homomorphism, we can always view the structure of H through the perspective of the quotient group $G/\ker(\varphi)$!

Corollary 3.5.2: properties as consequence

Let $\varphi : G \to H$ be a group homomorphism

1. φ is injective if and only if $\ker \varphi = 1$
2. $|G : \ker \varphi| = |\varphi(G)|$ [a]

[a]If you have taken linear algebra, this is the Rank-Nullity theorem equivalent for Group Theory

The second Isomorphism theorem let's us view the quotients of internal direct products (definition 3.4.1) form a different angle.

[5]In particular, it's a *natural homomorphism*. The word natural here has a technical meaning which we explore in section 30.3. We will explore the concept once we explore the non-natural choice for basis for vector-spaces, which will let us contrast what it means to have a "natural" versus a "non-natural" function

Theorem 3.5.3: The Second or Diamond Isomorphism Theorem

Let G be a group, let A and B be subgroups of G and assume $A \leq N_G(B)$. Then AB is a subgroup of G, $B \trianglelefteq AB$, $A \cap B \trianglelefteq A$, and $AB/B \cong A/A \cap B$.

The name diamond comes form this diagram where the arrow represent subgroups, and a red arrow represents norma subgroup:

$$\begin{tikzcd}
 & G \arrow[d] & \\
 & AB \arrow[dl] \arrow[dr, "\trianglelefteq"] & \\
A \arrow[dr] & & B \arrow[dl] \\
 & A \cap B \arrow[d] & \\
 & \{e\} &
\end{tikzcd}
\quad \Rightarrow \quad
\begin{tikzcd}
 & G \arrow[d] & \\
 & AB \arrow[dl] \arrow[dr] & \\
A \arrow[dr, "\trianglelefteq"] & & B \arrow[dl] \\
 & A \cap B \arrow[d] & \\
 & \{e\} &
\end{tikzcd}$$

and

$$\frac{AB}{B} \cong \frac{A}{A \cap B}$$

The reason this theorem is intersting is because we often enough want to take the internal direct prodcut of groups, and deduce some structural information form it with quotients, and this theorems gives us a a tool do do so (see exercise ref:HERE)

Proof **:**
Since $B \trianglelefteq G$, by proposition 3.4.3, $AB \leqslant G$. Another way of seeing this is that AB is the collection of cosets Ab. Then if $\pi : G \to G/A$ is the cannonical projection, $AB = \pi^{-1}(\pi(B))$. The image of a subgroup is a group (ref:HERE) and the pre-iamge of a subgroup is a group (see exercise ref:HERE). Thus AB is a subgroup.

To show $B \trianglelefteq AB$, we need to show that for any $ab \in AB$, $abB(ab)^{-1} = abBb^{-1}a^{-1} = B$. To do so, recall that B is normal, so that $kBk^{-1} = B$, and so

$$abBb^{-1}a^{-1} \overset{!}{=} aBa^{-1} = B$$

where $\overset{!}{=}$ comes from the fact that B is closed. Since ab was a general element of AB, it follows that $B \trianglelefteq AB$.

Finally, to show that $\frac{B}{A \cap B} \cong \frac{AB}{B}$, define the map $\varphi : A \to AB/B$, where $\varphi(a) = aB$. We'll show this is a surjective homomorphism. This function is indeed a homomorphism, since

$$\varphi(ab) = abB \overset{!}{=} aBbB = \varphi(a)\varphi(b)$$

where $\overset{!}{=}$ comes from exercise ref:HERE. Alternativly, Since $B \trianglelefteq AB$, $\pi : AB \to AB/B$ is a

well-defined group-homomorphism. Then one can show that restricting to $\pi|_A A \to AB/B$ is also a group-homomorphism.

To show that φ is surjective, notice that every element of AB/B can be written as abB. Since $b \in B$, then

$$abB = aB$$

it follows that an arbitrary coset abB has a map to it (i.e. $\varphi(a) = aB = abB$), showing that φ is surjective. Thus, by corollary 3.1.14, $A/\ker(\varphi) \cong AB/B$. To find $\ker(\varphi)$, we simply put together the information given to us:

$$\ker(\varphi) = \{a \in A \mid \varphi(a) = 1B = B\} = \{a \in A \mid aB = B\} \overset{!}{=} \{a \in A \mid a \in B\} = A \cap B$$

where $\overset{!}{=}$ comes from exercise ref:HERE. Thus, substituing $\ker(\varphi)$ by $A \cap B$, we get

$$\frac{A}{A \cap B} \cong \frac{AB}{B}$$

as we sought to show

One way to remember this theorme is through the following mnemonic:

If we have $\frac{AB}{B}$, then add a $A \cap _$ to the top and bottom, giving us

$$\frac{A \cap AB}{A \cap B} = \frac{A}{A \cap B}$$

Another way to think about the 2nd isomorphism theorem is to consider the case where $A \cap B = \{e\}$. Then this theorem tells us that

$$\frac{AB}{B} \cong A \qquad \frac{AB}{A} \cong B$$

Heuristically, we can think in this situation that we are "cancelling out" the B's. However, be mindful to take this as a heuristic – this is *not* what is actually happening!

The third Isomorphism theorem lets quickly simplify quotient groups quotiented *by* a quotient group:

Theorem 3.5.4: The Third Isomorphism Theorem

Let G be a group and let H and K be normal subgroups of G with $H \leq K$. Then K/H, G/H and

$$(G/H)/(K/H) \cong G/K$$

If we denote the quotient by H with a bar, this can be written

$$\bar{G}/\bar{K} \cong G/K$$

Proof **:**
Since $H \leqslant K$, then $K/H \trianglelefteq G/H$. To show this, first note that K/H is indeed a subgroup of G/H. For any $[x], [y] \in K/H$, $[x][y]^{-1} = [xy^{-1}] \in K/H$ since $xy^{-1} \in K$. Furthermore, for any $[g] \in G/H$, and any coset $[k] \in K/H$, $[g][k] = [k'][g]$, $k' \in K$ (i.e. $[g](K/H) = (K/H)[g]$), since

both K and H are normal. Thus $K/H \trianglelefteq G/H$.

Now, define the map

$$\varphi : G/H \to G/K \qquad gH \mapsto gK$$

This map is well-defined: If $g_1H = g_2H$, then $g_1h_1 = g_2h_2$. Since $H \leqslant K$, then $h_1, h_2 \in K$, so $g_1h_1 = g_2h_2$ still holds in G/K for any $h_i \in H$. Thus $\varphi(g_1H) = \varphi(g_2H)$, showing the function is well-defined.

Now, notice this map is a homomorphism and surjective. It is a homomorphism since

$$\varphi(g_1Hg_2H) \overset{!}{=} \varphi(g_1g_2H) = g_1g_2K \overset{!}{=} g_1Kg_2K = \varphi(g_1H)\varphi(g_2H)$$

where the $\overset{!}{=}$ equalities come from exercise ref:HERE. For surjectivity, since $H \leqslant K$, then for any gK, we know there is at least the coset gH that maps to it (there might be more, since $H \leqslant K$). Hence, by corrollary 3.1.14

$$(G/H)/\ker(\varphi) \cong G/K$$

to find $\ker(\varphi)$, we put together the information we have so far:

$$\begin{aligned}\ker(\varphi) &= \{gH \in G/H \mid \varphi(gH) = K\} \\ &= \{gH \in G/H \mid gK = K\} \\ &= \{gH \in G/H \mid g \in K\} \\ &= K/H\end{aligned}$$

Thus, replacing $\ker(\varphi)$ with K/H we get

$$(G/H)/(K/H) \cong G/K$$

as we sought to show

An easy heuristic to remember $(G/H)/(K/H) \cong G/K$ is to think about how to simply fraction's of fractions, that is, the "flip and cancel" trick

$$\frac{\frac{1}{2}}{\frac{4}{2}} = \frac{1}{2}\frac{2}{4} = \frac{1}{4}$$

Note however that this is *not* waht's happening here – it is just a useful heuristic to remember the 3rd Isomorphism Theorem. Note too that for the simplification to occur, you need that H is in the denominator of both quotients, you cannot have a different value.

Example 3.5.2: Example
Using the 3rd isomorphism theorem?

Finally, we will relate the subgroup structure (the lattice) of G/N with G. It will actually be precisely a sublattice of G that contains N! This means that going to G/N will preserve subgroup relations:

Theorem 3.5.5: The Fourth or Lattice Isomorphism Theorem (Correspondence Theorem)

Let G be a group and let N be a normal subgroup of G. Then there is a bijection from the set of subgroups H of G which contain N onto the set of subgroups $\bar{H} = H/N$ of G/N.

$$\{H | H \leqslant G,\ H \text{ contains } N\} \leftrightarrow \left\{\overline{H} \leqslant \overline{G}\right\}$$

In particular, every subgroup of $\bar{G}$ is of the form H/N for some subgroup H of G containing N (namely, its pre-image in G under the natural projection homomorphism from G to G/N). This bijection has the following properties between the two lattices

Proof **:**
Let $\overline{G} = G/N$. Let $\overline{H} \leqslant \overline{G}$. Then since the pre-image of a subgroup is a group, take $H = \pi^{-1}(\overline{H})$. Then by definition, $\pi(H) = \overline{H}$, meaning for every subgroup of $\overline{G}$, there is a subgroup of G that maps to it.

Furthermore, $N \leqslant H$. Since $\pi : H \to H/N$, $\pi^{-1}(1N) \subseteq H$ (note that it's a subset and not element since it's possible that the pre-image contains multiple elements). By definition

$$\pi^{-1}(1N) = \{h \in H \mid \pi(h) = N\} = \{h \in H \mid hN = N\} = \{h \in H \mid h \in N\}$$

and since $\pi^{-1}(1N) \subseteq H$, it must be that $N \subseteq H$, as we sought to show

This fact has many very nice consequences:

Corollary 3.5.6: Equivalent way of looking at 4th Isomorphism Theorem

1. $A \leqslant B$ if and only if $\overline{A} \leqslant \overline{B}$
2. if $A \leqslant B$ then $|B : A| = |\overline{B} : \overline{A}|$
3. $\overline{\langle A, B\rangle} = \langle \overline{A}, \overline{B}\rangle$
4. $\overline{A \cap B} = \overline{A} \cap \overline{B}$
5. $A \trianglelefteq G$ if and only if $\overline{A} \trianglelefteq \overline{G}$

Proof **:**
Each of these can be proved using a simialr technic as the proof for the 4th Isomorphism Theorem

Another way of looking at the first part of the Correspondence Theorem is in the following slight generalization: Let G be a group and N a normal subgroup so that G/N is well-defined. If $\mathcal{G}$ contains all subgroups of G, for any $H \in \mathcal{G}$, the map $\pi : H \to G/N$, $\pi(h) = hN$ is well-defined. So, H maps to a subgroup of G/N. If $\mathcal{N}$ represents all subgroups of G/N, then there is a surjective map

$$\mathcal{G} \twoheadrightarrow \mathcal{N}$$

and if we take $\mathcal{G}|_N$ to be all subgroups that contain N, then there is a bijection between $\mathcal{G}|_N$ and $\mathcal{N}$:

$$\mathcal{G}|_N \leftrightarrow \mathcal{N}$$

Example 3.5.3: Using Correspondence Theorem

1. Let $G = Q_8$, and take $\langle -1 \rangle \trianglelefteq Q_8$. Then in the following lattice, the double lines represent the sublattice of Q_8 which is an isomorphic copy of the lattice of $Q_8/\langle -1 \rangle$

$$\begin{tikzcd}
 & & Q_8 \arrow[dl, Rightarrow] \arrow[d, Rightarrow] \arrow[dr, Rightarrow] & & \\
 & \langle i \rangle \arrow[dr, Rightarrow] & \langle i \rangle \arrow[d, Rightarrow] & \langle i \rangle \arrow[dl, Rightarrow] & \\
 & & \langle i \rangle \arrow[d, dash] & & \\
 & & \{1\} & &
\end{tikzcd}$$

2. Let $G = D_4$ and let $\langle r^2 \rangle \trianglelefteq D_4$. Then with the same definitions as the last example, we get:

$$\begin{tikzcd}
 & & & D_4 \arrow[dl, Rightarrow] \arrow[d, Rightarrow] \arrow[dr, Rightarrow] & & & \\
 & & \langle s, r^2 \rangle \arrow[dl, dash] \arrow[d, dash] \arrow[dr, Rightarrow] & \langle r \rangle \arrow[d, Rightarrow] & \langle rs, r^2 \rangle \arrow[dl, Rightarrow] \arrow[d, dash] \arrow[dr, dash] & & \\
\langle s \rangle \arrow[drrr, dash] & \langle r^2 s \rangle \arrow[drr, dash] & & \langle r^2 \rangle \arrow[d, dash] & & \langle rs \rangle \arrow[dll, dash] & \langle r^3 s \rangle \arrow[dlll, dash] \\
 & & & 1 & & &
\end{tikzcd}$$

Notice that several subgroup of G (for example $\langle s \rangle$) become elements in a larger subgroup (in this case $\langle s, r^2 \rangle$). Thus, every subgroup of $\overline{G}$ might have multiple subgroups of G that map to it, but a unique subgroup which contains N.

A good example how the map of subgroup of G to G/N is surjective is by the following example: Notice that $Q_8/\langle -1 \rangle \cong D_8/\langle r^2 \rangle \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, and $\langle -1 \rangle \cong \langle r^2 \rangle$, but $Q_8 \ncong D_8$. We will explore this idea in more depth in the next section.

Exercise 3.5.1

1. In the following questions, we're going to prove what's sometimes called the 5th Isomorphism Theorem (also known as the *Butterfly Lemma*), which for our purposes will offer an alternative proof to the main result of section 3.7. HERE

3.6 Extensions

Recall (or revisit) that if $H \leqslant G$ (and G is finite), then the subgroup lattice of H is the sublattice of G of all groups *contained* in H (theorem 2.4.2), so *a fortiori* if $K \trianglelefteq G$ is a normal subgroup then the subgroup lattice of K is the sublattice of G of all subgroups contained in K. Even hen G is infinite, the set of subgroup of K is the set o subgroup of G which are contained in K. The Correspondence Theorem has shown us that the "dual" is true for quotient groups (except it orks for *all* groups, not just finite): If $K \trianglelefteq G$, then the subgroup lattice of G/K is the sublattice of G of all groups *containing* K. In this way, K and G/K together capture the subgroup structure of a group! In this section, we will explore how we can find a group structure given *only* the information from K and G/K.

More succintely, Given $K \trianglelefteq G$, the group K and G/K contain the subgroup structure of G. However, K and G/K don't capture the full structure of a group: in the process of quotienting, G/K losses some information:

Example 3.6.1: Subgroup And Quotient Group Limits

Take $G = D_8$ and $K = \{1, r, r^2, r^3\} \cong C_4$. Then $D_8/K \cong C_2$ – let's label this K'. Then given just K and K', e can't guarantee that $K' = D_8/K$, other groups can take the place of D_8. For example:

$$H = C_4 \times C_2 \qquad \varphi : C_4 \times C_2 \to C_2,\ \varphi((a, b)) = ((0, b))$$

Here, H is another group here the kernel is C_4 and the quotient is C_2.

It must be noted though that there is still information within cosets: for any $x \in G$, then $[x] = xK \in G/K$, so for any $y \in xK$, $y = xk$ for appropriate $k \in K$ (proposition 3.1.3). This means for every coset, there is some information about how the elements relate. An easy example of this can be seen if $K = Z(G)$.

Example 3.6.2: Relation Within Cosets

Take $G/Z(G)$ (as an aside, try proving that $Z(G) \trianglelefteq G$ alway, exercise ref:HERE). Take any $[x] \in G/Z(G)$, and any $a, b \in [x]$. Then $a = xz_a$ and $b = xz_b$, so $ab = xz_axz_b = xz_bxz_a = ba$ (since $z_a, z_b \in Z(G)$). Thus, within any coset of $G/Z(G)$, the elements commute! This is not nesesarily the case if the elements are *not* within the same coset (ex. $G = D_7$).

However, as we saw in example 3.6.1, there can be more than one group (say G') we can construct from K and K' (where $G'/K = K'$). We will call groups which can be built in this fashion *extension*:

Definition 3.6.1: Extension

Let A and B be groups. If $G/A \cong B$, then we say G *is an extension of* A *by* B.

The idea behind this name is that we want to be able to restrict G to A (i.e. A must be able to map injectively and homomorphically into G; $A \hookrightarrow G$), similarly to how we can restrict functions (the dual of restricting functions is also called *extending* functions, where the extended function must be exactly the same as the orginal funciton when restricting).

As we saw at the beginning of this sectino, quotient groups are the prefect candidate for extension of groups since every group can be decomposed into a [normal] subgroup and quotient group (TBD – perhaps put this earlier)

Since all normal subgroups are also kernels, we can re-write what it means to be an extension in terms of homomorphisms. In the following definition, let I represent the trivial group.

Definition 3.6.2: Short Exact Sequence

Let A, B, C be group. If there exists an injective homomorphism f and a surjective homomorphism g such that

$$I \xrightarrow{\iota} A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{\pi} I$$

where $\mathrm{im}(f) = \ker(g)$, ι is the trivial injective $(0 \mapsto 0)$ map and π is the total projective map $(c \mapsto 0)$, then we say that B *is an extension of* A *by* C.

Notice that $\mathrm{im}(\iota) = \ker(f)$ since f is injective, $\mathrm{im}(g) = \ker(\pi)$ since g is surjective, and $\mathrm{im}(f) = \ker(g)$ by definition. The fact that the image of the previous function is the kernel of the next function is no coincidence – this property is called *exactness* [6]. If we make a longer chain of homomorphisms with this property (ex. $X_1 \xrightarrow{f_1} X_2 \xrightarrow{f_2} X_3 \xrightarrow{f_3} X_4 \xrightarrow{f_4} \cdots$), then such a chain is called an *exact sequence* (we will return to these in chapter 16). The exact sequence in definition 3.6.2 is the shortest non-trivial sequence with the trivial group at the beginning and end, and so it is called a short exact sequence.

For our purposes, the fact that $\mathrm{im}(f) = \ker(g)$ means that $B = C/A$, meaning that B is an extension of A by C. (exercise: does $G \to H$ imply a short exact sequence? If no, can it)

Example 3.6.3: Short Exact Sequences

In the following examples, let 1 represent the trivial group

1. If $\varphi : G \to H$ is a homomorphism and $K = \ker(\varphi)$, we always have

$$1 \to K \hookrightarrow G \xrightarrow{\varphi} \varphi(G) \to 1$$

is a short exact sequence. If φ is surjective, then

$$1 \to K \hookrightarrow G \xrightarrow{\varphi} H \to 1$$

In this way, every surjective homomorphism represents an extension G of K by H.

2. Let A and B be two arbitrary groups. Then

$$1 \to A \hookrightarrow A \times B \xrightarrow{(a,b)\mapsto(0,b)} B \to 1$$

is a short exact sequence. In this way, any group can be extended by any other group to the extensions $A \times B$. This is sometimes called the *trivial extension* [a]

3. Take $A = B = 2$. Then

$$1 \to C_2 \xrightarrow{f} C_4 \xrightarrow{g} C_2 \to 1$$

where $f(x) = 2x$ and $g(0)$ and $g(2)$ both map to 0, while $g(1)$ and $g(3)$ map to 1. It is easy to verify that f and g are indeed homomorphism and that the above sequence is exact.

[6] You might have seen the term *exact vector-field*. This concept is in fact the same, however it would take us a long time before we can link the two (see ref:HERE)

4. Similarly to the last example, Let $A = C_2$ and $B = C_8$ and $C = C_4$. Can you find the appropriate maps such that
$$1 \to C_2 \to C_8 \to C_4 \to 1$$
is a short exact sequence? Is the following
$$1 \to C_2 \to D_4 \to C_4 \to 1$$
also a short exact sequence with the appropriate maps?

[a] In fact, once we cover homological algebra in chapter 16, the idea of this being the trivial extension will become rigorous

An extension of an X group by an X group is X. when is this statement true when X is:

1. cyclic
2. If M and A are abelian, G need not be abelian
3. finite
4. finitely generated
5. noetherian

It is in general difficult to tell when when given a short exact sequence if the sequence splits. In section 16.2, we will classify which *abelian groups* in a short exact sequence of abelian groups split[7] (abelian groups which always split in a short exact sequence of abelian groups be called *projective groups*). For general groups, The best tool we have is trying to identify when C acts on B, (mabye there are some identification of when this is possible??)

(maybe make HNN extensions an exericise? Or more it to an Algebraic topology section if one ever exists?)

(

You may also think of the short exact sequence:
$$1 \to H \to G \to G/H \to 1$$
Then since H is the kernel, we can break down G into zH for some representative $z \in G/H$. It is natural to ask if we can think of as G as "$(G/H) \cdot H$". However, this is not so simple. It sometimes works, it is called a split extension, and in general you require the fundamental theorme of extension of groups (which shows that G is a subgroup of the wearth product of H and G/H. We alos use homology to see how much G "fails" to be a direct product of H and G/H)

)

Exercise 3.6.1

1. Let $0 \to \mathbb{Z} \to B \to A \to 0$ be an extension, where B is an abelian group and A is free. Show that B splits. [Interestingly, The converse of this statement (known as Whitehead's Problem) is undecidable in ZFC, that is, it is a statement that could neither be true nor false! This idea would be explored with Gōdel's Incompleteness Theorems]

[7] It will turn out that all abelian groups are $\mathbb{Z}$-modules, and so this will be a consequence of classifying when modules split with the space case of when the module is over $\mathbb{Z}$

2. Is it save to say that if $0 \to A \to G \to B \to 0$ where $|B|$ is finite, then G is A-by-finite (and so virtually A)? That would be another cool way of thinking about the B term in a short exact sequence!

3.7 Classifying all Finite Groups: Hölder Program

With the isomorphism theorems under our belt, we are ready to tackle one of the central questions posed by group-theorists: How can we classify all groups? This sections start by building up the mental models used in order to comprehend the way classification efforts.

We can approach the problem of classification in many different ways. We briefly mentioned the idea trying to construct arbitrary presentation's of groups to try and create groups, but we've mentioned that this can create "insoluble presentation", meaning presentaion in which we cannot represent every element. In general, it is very hard to find groups in that manner. Perhaps instead we can start with the groups we knwo and see if there is a way to "decompose" them. The first thing you might try is decomposing the group into cyclic sub-groups. This is certainly possible (take an element, generate a group from it, remove those elements from your origianl group, continue doing this till you've exhauted the group [8]). But how would you then put them back together? For example, we know that $\langle r \rangle, \langle s \rangle \leqslant D_4$. With what we've learnt, is there a way to represent D_4 using these? In fact, we have just seen that we can but them back together using short exact sequences! The group D_4 is an extension of C_2 by C_4.

However, we must be more careful using extensions when we are trying to classify groups; are cyclic subgroups really the way to go? In fact, this is not the case: there exists cyclic groups which are an extension of smaller groups. We would want to be able to have some fixed subgroups we can use as our "building puzzles" which are not extensions of any other groups. In this section we will define these "smalles puzzle pieces" or "atoms of groups". Try taking a moment to see if you can come up with a more suitable definition of "smallest puzzle piece" for groups and have an idea of an example of a group following your definition.

We will end this section by being able to declare a theorem that took group theorists the better part of the 20th century to prove – the classification of all finite groups! As a quick note before proceeding, this classification effort is in fact much simpler for *abelian groups*. In chapter 6.1, section ref:HERE, we will continue this exploration in the finite abelian case.

Defining our Terms

First, we will have to define what would be our "puzzle piece". Since we are using extensions, we know there is a non-trivial homomorphism from a group G to another group H if G has a non-trivial normal subgroup. If G has *only* trivial subgroups, the structure of G is not "composed" of other possible structure but is as "simple" as possible. This motivates the folloing definition.

[8] This naturally works well in groups of *countable* size, one has to be more careful for groups of *uncountable* size – the axiom of choice is involved (see chapter A for details on the axiom of choice)

Definition 3.7.1: Simple Groups

A (finite or infinite) group G is called *simple* if $|G| > 1$ and the only normal subgroups of G are 1 and G.

These groups are called "simple" because there are no more group-structures within them [9]. To undersatnd what "no more group-structures within them" means, recall the 4th isomorphism theorem. If N is a noraml subgroup, then ther is a direct correspondence between the subgroups of G/N and the subgroups of G containing N. If there is no $N \trianglelefteq G$ besides $\{1\}$ and G, then the lattice of G can't be split into G/N and G. So in a sense, that lattice is as small as it gets. (comment on infinite case?)

Example 3.7.1: Simple Groups

We know that by Lagrange's theorem that if G is prime, then it's only subgroups must be 1 and G, and so groups of order p are examples of simple groups. If G is abelian, then it can in fact be shown that every simple subgroup is isomorphic to $\mathbb{Z}/p\mathbb{Z}$ for appropriate prime p, since those are the only subgroups which no non-trivial subgroups (and so no normal subgroups)

There are also examples of non-abelian simple Groups, both infinite and finite (the smallest finite one (of order 60) will be explored in detail in chapter 4, section ref:HERE).

Looking at simple groups a little more abstractly for a moment, they play an analogous role to primes in arithmetic's of $\mathbb{N}$ as in they cannot be "factored" into smaller [meaningful] pieces[10] (i.e. N and G/N). There is even a "unique factorization theorem" based on this idea. First, we'll define what is meant by "factorize" a group

(link to extensions)

Definition 3.7.2: Subnormal Series

Let G be a group. Then

$$G = G_1 \geqslant G_2 \geqslant G_3 \geqslant \cdots \geqslant G_n = \{1\}$$

is called a *series* (or *tower*) for the group G. If every $G_i \trianglelefteq G$, then G the series is called a *normal series*. If $G_i \trianglelefteq G_{i-1}$ for every $1 \leq i \leq n$, then we call the series a *subnormal series*:

$$G = G_1 \trianglerighteq G_2 \trianglerighteq \cdots \trianglerighteq G_n = \{1\}$$

Sometimes, [normal/subnomral] series are written in the ascending, rather than descending order:

$$\{1\} = G_0 \leqslant G_1 \leqslant \cdots \leqslant G_n = G$$

I chose to write them in descending order since down the line we will be comparing some types of series to other types, and if everything is written in descending order, then we'd have a scendario

[9] The term simple is one that comes up often enough in matheamtics. It has a sligthly more nuanced meaning than just the simplest object (ex. the trivial group is not a simple group, it's "too simple to be simple". See this nlab article on simple objects

[10] meaningful here meaning there is a notion of combining groups using short exact sequence, unlike general groups can't be

were we start writing $G_{n-i} \trianglelefteq B_i$ ineated of $G_i \trianglelefteq B_i$, which I feel will make the simple relations a little more notationaly cumbersome than it needs to be.

You might ask whether this a subnormal series would imply a noraml series, that is, that $N_i \trianglelefteq G$? However, that doesn't need to be the case:

Example 3.7.2: Normality *not* transitive
Here's an example of how normality is *not* transitive. Let $G = D_8$. Let $H \trianglelefteq G$ such that $H = \langle r\sigma, \sigma r\rangle$. Next, take $K \trianglelefteq H$ such that $K = \langle r\sigma\rangle$. Note that

$$K \trianglelefteq H \trianglelefteq G$$

but $K \ntrianglelefteq G$

thus a subnormal series need not be a noraml series. However, a normal series is a subnormal series: if $N_{i+1} \trianglelefteq G$, then $gN_{i+1}g^{-1} = N_{i+1}$ for every $g \in G$. If we restrict to every $n \in N_i$, then since $N_i \subseteq G$, it is already the cse that $nN_{i+1}n^{-1} = N_{i+1}$. Thus $N_{i+1} \trianglelefteq N_i$. Thus:

$$\text{normal series} \Rightarrow \text{subnormal series} \qquad \text{subnormal series} \not\Rightarrow \text{normal series}$$

Thus, subnormal series can be thought of as the decomposition of a group.

Example 3.7.3: Example
here. In particualr, give a two subnromal series for a group, one more refined than the other

In (reference previous example), the subnormal serise cannot be refined further, since every N_i/N_{i+1} is simple. We give a name to such a serise:

Definition 3.7.3: Composition Series

In a group G, a sequence of groups

$$1 = N_0 \leqslant N_1 \leqslant N_2 \leqslant \cdots \leqslant N_{k-1} \leqslant N_k = G$$

is called a *composition series* if $N_i \trianglelefteq N_{i+1}$ and N_{i+1}/N_i is a simple group, for all $0 \leq i \leq k-1$. If the above sequence is a composition series, the quotient group N_{i+1}/N_i are called the *composition factors* of G.

Every finite group has a composition series. Since the group is finite, there exists a maximal subgroup [11], so that $N \trianglelefteq G$ so that G/N is simple. If $N = \{1\}$, then G is already simple, since the only proper normal subgroup of G is $\{1\}$, meaning the only other normal subgroup is G. If the group G/N was not simple, then there would exist a group $N'/N \trianglelefteq G/N$. Then by the 4th isomorphism theorem, $N' \trianglelefteq G$, where N' is a group containing N. However, N is a maximal subgroup, so it must be that $N = N'$, so we'd get $N'/N = N/N \cong \{1\} \trianglelefteq G/N$, and so G/N is simple. We can then continue the same process with N, taking $N_1 \trianglelefteq N$ to be the maximal normal subgroup. Since G was finite this process will be end.

[11] note that maximal subgroups by definition are proper subgroups

Example 3.7.4: Some Composition Series
Take $\mathbb{Z}_p \times \mathbb{Z}_q$. Then

$$\{e\} \trianglelefteq \mathbb{Z}_p \times \{e\} \trianglelefteq \mathbb{Z}_p \times \mathbb{Z}_q$$
$$\{e\} \trianglelefteq \{e\} \times \mathbb{Z}_q \trianglelefteq \mathbb{Z}_p \times \mathbb{Z}_q$$

It is clear that the quotient of any two will produce simple groups

Take Q_8 Then

$$\{e\} \trianglelefteq \langle -1 \rangle \trianglelefteq \langle i \rangle \trianglelefteq Q_8$$
$$\{e\} \trianglelefteq \langle -1 \rangle \trianglelefteq \langle j \rangle \trianglelefteq Q_8$$
$$\{e\} \trianglelefteq \langle -1 \rangle \trianglelefteq \langle k \rangle \trianglelefteq Q_8$$

Where I figured out the composition series with the help of the lattice of subgroups to trace possible paths and figuring out whether they're normal or not. The quotient so any of these factors will be isomorphic to $\mathbb{Z}_2$, which is simple, and thus there are composition series.

another example is:

$$1 \trianglelefteq \langle s \rangle \trianglelefteq \langle s, r^2 \rangle \trianglelefteq D_8 \qquad \text{and} \qquad 1 \trianglelefteq \langle r^2 \rangle \trianglelefteq \langle r \rangle \trianglelefteq G$$

Note that not all infinte groups admit a compositions series

Example 3.7.5: Example
I think $M_2(\mathbb{Q})$ or $GL_2(\mathbb{Q})$ is an example.

However, some groups do have compositions series. Take the group

$$D_\infty = \langle x, y | y^2 = 1,\ xy = y^{-1}x \rangle$$

this is called the *infinite dihedral group*. Notice that $\langle x \rangle \trianglelefteq D_8$: if we take some $g \in \langle x \rangle$, then since $\langle x \rangle$ is cyclic, then $g = x^a$ for some a. Conjgating x^a, we get $yx^ay^{-1} = yyx^a = x^a$, meaning $yxy^{-1} \in \langle x \rangle$. Using this, we can show that $\langle x \rangle \trianglelefteq D_\infty$. Thus, $D_1 = D_\infty / \langle x \rangle$ is well-defined. We need to show that D_1 is Let's take some $g \notin \langle x \rangle$ so that $[g] \neq [e]$. (buidlling up to show that)

$$C_2 \cong D_\infty / \langle x \rangle$$

Which is simple. Further more, C_2 is itself a simple abelian group. Thus, our composition seris is

$$D_\infty \trianglerighteq \langle x \rangle \trianglerighteq \{1\}$$

Some of infinite groups are close to being composition series, in the sense that $G = \cup_{i=1}^{\infty} G_i$ where G_i / G_{i+1}. Such groups will be addressed further [book ref:HERE]

(A word on how composition series are just simple extensions:

We saw that a finite group G always has a composition series, and hence a subnormal series:

$$G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq \cdots \trianglerighteq G_n = \{1\}$$

Then notice that every subnormal series can be broken down into short exact sequences:

$$1 \longrightarrow G_{n-1} \longrightarrow G_{n-2} \longrightarrow G_{n-2}/G_{n-1} \longrightarrow 1$$

$$1 \longrightarrow G_i \longrightarrow G_{i-1} \longrightarrow G_{i-1}/G_i \longrightarrow 1 \quad 1 \longrightarrow G_1 \longrightarrow G_0 \longrightarrow G_0/G_1 \longrightarrow 1$$

thus, every subnormal series can be thought of as a series of series of extensions G_0/G_1, G_1/G_2, ... G_{n-1}.)

One might ask about if a composition series is supposed to be unique factorization, what about there being 3 composition series for Q_8, and what about the re-arranging we've done of $\mathbb{Z}_p \times Z_q$? It turns out that the *quotients* of each N_{i+1}/N_i will be unique *up to rearrangement*! That is the content of the next theorem:

Theorem 3.7.4: Jordan-Hölder's Theorem

Let G be a finite group with $G \neq 1$. Then

1. G has a composition series
2. The composition factors in a composition series are unique, namely, if $1 = N_0 \leqslant N_1 \leqslant \cdots \leqslant N_r = G$, and $1 = M_0 \leqslant M_1 \leqslant \cdots \leqslant M_s = G$, are two composition series of G, then $r = s$ and there is some permutation π of $\{1, 2, ..., r\}$ such that

$$M_{\pi(i)}/M_{\pi(i)-1} \cong N_{\pi(i)}/N_{\pi(i)-1} \qquad 1 \leq i \leq r$$

Proof **:**
apparently doable. This material wasn't covered in class and there's a test coming up so I'll get back to it another time

With that definition of factorization, we can now ask one of the biggest questions in group theory:

Theorem 3.7.5: The Hölder Program

1. Classify *all* finite simple groups
2. Find all ways of "putting simple groups together" to from other groups

Proof **:**
The proof for the 1st point took around 100 years of work, with the effort of around 100 mathematicians, writing between 300-500 papers, totalling 5000-10000 pages. It is left as some light reading for the reader ☺

The second point is yet to be complete.

The result of all that work means the following single incompassing theorem can confidently be write down:

Theorem 3.7.6: Types of Finite Groups

There is a list consisting of 18 (infinite) families of simple groups and 26 (or 27 [a]) simple groups not belonging to these families (the *sporadic* simple groups) such that every finite simple group is isomorphic to one of the groups in this list

[a] some include Tits group

Example 3.7.6

Here are a couple such family of groups:

1. $\{Z_p | p \text{ is prime } \}$
2. $\{SL_n(\mathbb{F})/Z(SL_n(\mathbb{F})) | n \in \mathbb{Z}^+, n \geq 2, \text{ and } \mathbb{F} \text{ is a finite field}\}$
3. A_n, $n \geq 5$.

Examples after this start to get more complicated. I found this wikiepdia article outlining more of the history of this process.

here

(somehow mention "generalization of abelian group section", where I can expan on how nilpotent and solvable groups give us some nice properties to make some groups easier to reprsent)

:exercise Let $H \trianglelefteq G$ such that there does not exist a $K \trianglerighteq H$ such that $H \trianglelefteq K \trianglelefteq G$. In other words if $K \trianglerighteq H$, then either $K = H$ or $K = G$. Such a normal subgroup is called the maximal normal subgroup. Show that H is a maximal normal subgroup if and only if G/H is a simple group

3.8 Quotient Objects

Link to Epimorphisms, link the idea of the "dual" is sub-object. Some intuition then perhaps on how the sub and quotient object together capture the idea of the entire group G. Leads to the idea of extensions.

In the previous chapter, we introduced the equivalent definition of subgroup in proposition 1.4.5. In it, we had an injective map $i : H \to G$. We will now ask what mathematicians call the natural *dual* question: What happens if we instead consider a surjective function $j : G \to H$ which is also a homomorphism?

(Link to Ker Object?)

4

Group Actions

In section 1.5, we saw

Groups are abstract object that have many applications (build-up to saying how we can use them to represent concrete things, systems. Maybe an example using $\mathbb{R}$ to represent time acting on some sytem represented by $\mathrm{Sym}(X)$)

Yet another way of seeing groups is (by seeing that they can act on other groups, or themselves! Thinking of $G \to \mathrm{Aut}(G)$ means that we are diving into the structure of G, so to speak, and using itself to analyze that structure!)

4.1 Defining Group Action

Let's return back to the world of sets and see what sets can represent

1. If $X_1 = \left\{(x, y) \in \mathbb{R}^2 \;\middle|\; x^2 + y^1 = 1\right\}$, then X_1 represents a circle:

   Img here

2. Similarly, if $X_2 = \left\{(x, y, z) \in \mathbb{R}^3 \;\middle|\; x^2 + y^2 + z\right\}$, then X_2 represents a sphere

   Img here

3. If $X_3 = \{\text{points} \mid \text{graph}\}$, then X_3 is a graph. For exmaple if $X_3' = HERE$, then visually, we get

   Img Here

4. If $X_4 = \{1, 2, 3, 4\}$, we can let each number represent the side of a square. That way, we get that

here

5. In general, if $|X_4'| = n$, we can let that represnt an n-gon:

Img here

6. A set can have the smae size but represent different things. For example, we can represent a tetrahedron with X_5 $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, which looks like

Img here

7. Any set can also represent the elements of a group, since every group has an underlying set

Each of these represnt some object. The set doesn't contain any more information than the object. For example, the set X_4, which we declared is representing the corners of a square, only has the information that there are 4 corners – in a sense, the object it represents is "static", unchanging. The set X_4 doesn't contain information about whether the object can be moved or changed (for example, rotated).

What we seek is a way to represent how an object can *change* given some action is performed. For example, If we want to say we have rotated the square a quarter turn to the right (i.e. 90 degrees, or $\pi/4$ radians), then visually it looks like

Img here

The way we can capture this "action" preformed on the square, which is represented by the set X_4, is through a bijective function! If

$$f(1) = 2, \quad f(2) = 3, \quad f(3) = 4, \quad f(4) = 1$$

then f captured the idea of rotating the square represented by X_4. For the sake of foreshadowing, we will label the function f as r.

$$r(1) = 2, \quad r(2) = 3, \quad r(3) = 4, \quad r(4) = 1$$

We can similarly define a function, which again I'll label r^2 for the sake of foreshadowing, which rotates our square two quarters

$$r^2(1) = 3, \quad r^2(2) = 4, \quad r^2(3) = 1, \quad r^2(4) = 2$$

or, we can rotate the square across the vertical axis. Let σ be the function such that

$$\sigma(1) = 2, \quad \sigma(2) = 1, \quad \sigma(3) = 4, \quad \sigma(4) = 3$$

An observant reader would notice that this is essentially identical to how we defined the dihedral group D_8 back in section 1.2.3. Since each function we defined as far are bijective function, they are all part of S_4 (or equivalently $\mathrm{Sym}(X_4)$). Thus, what we have roughly shown is that there is a subgroup $S_4' \leqslant S_4$ such that every element of D_8 maps to some bijection in S_4': $D_8 \to S_4'$. This means

there is an homomorphism from D_8 into $\mathrm{Sym}(X_4)$: $\sigma : D_8 \to \mathrm{Sym}(X_4)$. In fact, even more is true: if we apply the function $r(x)$ follow by $r(x)$ we get $r^2(x)$, that is $(r \circ r)(x) = r^2(x)$ for all $x \in X_4$. More generally, it should be verified that for any function $a, b \in S_4$, $(a \circ b)(x) = (ab)(x)$. Thus, not only is there a bijection between D_8 and S_4', but there is a *homomorphism*! This homomorphism "captures" the idea of how D_8 will "act" on the set X_4 by showing how every element changes the set S_4.

We can generalize this idea: instead of X_4 we can have some set X, and instead of D_8, we can have some group G. If we define a homomorphism $\sigma : G \to \mathrm{Sym}(X)$. The homomorphism σ will represent how the group G acts on the set X. Homomorphisms from a group to $\mathrm{Sym}(X)$ (for any X) are given a name:

Definition 4.1.1: Group Action

Let G be a group and let S be a set. Let $\mathrm{Sym}(S)$ represent the symmetric group given the set S (i.e. all possible bijections from S into itslf). Then was call ρ a *group-action* from G into $\mathrm{Sym}(S)$

$$\rho : G \to \mathrm{Sym}(S)$$

if ρ is a homomorphism. Sometimes, the shorthand $G \curvearrowright S$ is used to represent a group G acting on a set S ($\curvearrowright$ is read as *acts on*).

Re-do these examples but with the other noation. Also, be clear that $\rho(g)$ is a permutation, so explain the idea of $\rho(g)(x)$ since it is a function.

Example 4.1.1: Group Actions

These are some concrete examples to show such functions exist

1. these next ones are inspired from this post `https://gowers.wordpress.com/2011/11/06/group-actions-i/`
2. Let $G = A_4$ be the alternating group on 4 elements and let T represent a regular tetrahedron with vertices labeled:

   IMG HERE

   We can think of $T = \{1, 2, 3, 4\}$ with the understanding that the number's correspond to the vertices. Let's say we want to rotate the tetrahedron 120 degrees ($\pi/3$ radians) while fixing 1. With some thought, the permutation $(123) \in A_4$ will accomplish this:

   IMG HERE

   (More words on this)
3. Let's take the same secenario, but now rotate a squre (lead up to how this is D_8)
4. (3 line example from link, this is *not* a faithful action)
5. Take $G = \mathbb{R}^\times$ and $S = \mathbb{R}^2$. Then

   $$r \cdot s = r \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} rx \\ ry \end{pmatrix}$$

   This will be a group action.

6. Take $G = \mathbb{R}^\times$ and $S = \mathbb{R}^2$. Then

$$r \cdot s = r \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} rx \\ r^{-1}y \end{pmatrix}$$

This is a different group action with the same group and set

7. G acts on itself by $(g, h) \mapsto gh$. similarly to last one, $(g, h) \mapsto hg^{-1}$ is also an action (the right action).

8. Take a group G. We introduced the shortand notation of

$$a \cdot a \cdot a \cdot \cdots \cdot a = a^n$$

in section 1.1 using generalized associativity. From this fact, we can define an action from $\mathbb{Z}$ onto G:

$$z \cdot a = a^z$$

This group action is well-defined. For $x, y \in \mathbb{Z}, g \in G$

(a) $x \cdot (y \cdot g) = x \cdot g^y = (g^y)^x = g^{yx} = g^{xy} = (xy) \cdot g$

(b) $1 \cdot g = g$

We can shorthand $\rho(g)(x)$ to $g(x)$, or even $g \cdot x$ if we're careful. Then we get

$$\rho(g \cdot h)(x) = (\rho(g) \circ \rho(h))(x)$$

which transaltes to

$$(gh)(x) = g(h(x)) \qquad (gh) \cdot x = g \cdot (h \cdot x) \tag{4.1}$$

Since we also have that $\varphi(e)(x) = x$ for all x, since e must map to the identity, and so be the identity permutation, we usually write

$$e(x) = x \qquad e \cdot x = x \tag{4.2}$$

(Already introduce Kaley's theorem to show how G can represent itself?? Would be a cool example, would just need to show the action is faithful. Maybe a word on faithful actions already?)

4.2 Axiomatic representation

Checking that a homomorphism exists between G and $\mathrm{Sym}(X)$ can be difficult at times. For this reason, we endeavour for simpler conditions which will characterize this homomorphism.

[What we'll do is find conditinos on the functions in $\mathrm{Sym}(x)$ which will let us define group actions. Function in $\mathrm{Sym}(x)$ are of the form $S \to S$. To not forget the group information, let's carry around that too: $G \times S \to S$. This is basically the same as before (instead of $\varphi(x)(y)$ we have $\varphi(x, y)$. We transformed it form something that takes one variable twice into a function that takes in 2 variables.

15: Remark

curryingRmk This process is called *uncurrying*, and going the other way is called *currying*)

]

[build up how $a \cdot (b \cdot c) = (ab) \cdot c$ is natural. Then why we need to care about the identity because of what we found]

These two conditions are necessary conditions, that is, given we have to represent a group action, these two will be consequences. Are the sufficient conditions, that is, will they define a goup-action on their own? The answer is in fact yes, and is what we'll prove for the rest of this section:

Theorem 4.2.1: Axiomatic Representation

Let G act on A. Consider

$$\sigma_g : A \to A \qquad \sigma_g(a) = g \cdot a$$

then the following are true:

1. for each $g \in G$, σ_g is a *permutation* of A
2. the map from $\varphi : G \to S_A$ defined by $g \mapsto \sigma_g$ is a *homomorphism*

Note: I say homomorphisms to S_A and not symmetric groups to make it clear from the beginning that the relation is with a special subgroup of the symmetric group. It has *some* of the symmetric-group structure.

Proof **:**

For (i), we need to show that σ_g follows the rule of permutation, that is, it is a set-map from $A \to A$ that is bijective (which for our case, we'll use the equivalent fact that it has a 2-sided inverse). This should be easy, for by construction $\sigma_{g^{-1}}$ should be the inverse. Indeed:

$$(\sigma_g \circ \sigma_{g^{-1}})(a) = (gg^{-1})(a) = e \cdot a = a$$

showing that $\sigma_g \circ \sigma_{g^{-1}}$ is the identity permutation, which shows σ_g^{-1} is a right inverse. We can equivalently show it's a left-inverse since $g^{-1}g = gg^{-1}$, and hence σ_g defines a permutation. Doing conventional proof by injectivity and surjectivity is also insightful to see where group-action axioms are used. To prove injectivity, we need to show that if

$$g \cdot a = g \cdot b \Rightarrow a = b$$

and since it is a *group* that we're dealing with, then

$$\begin{aligned} g \cdot a &= g \cdot b \\ \Rightarrow g^{-1}(g \cdot a) &= g^{-1} \cdot (g \cdot b) \\ \Rightarrow (g^{-1}) \cdot a &= (g^{-1}g) \cdot b && \text{associativity} \\ \Rightarrow (e \cdot a) &= e \cdot b \\ \Rightarrow a &= b && \text{identity} \end{aligned}$$

and for surjectivity, for any $b \in S$, there must exist a a such that $g \cdot a = b$ and indeed

$$\begin{aligned} &g \cdot a = b \\ &\Leftrightarrow a = g^{-1} \cdot b \end{aligned}$$

and since $g^{-1} \cdot b \in S$ (by well-definedness of group action function), such an a exists. Thus, all 3 properties of group-actions were needed to prove bijectivity.

Therefore, since every g is associated with a permutation σ_g, we can ask ourselves if G and S_A are related, and indeed they are by homomorphism. It turns out that we just have enough properties with group actions to show this. Take $\varphi : G \to S_A$, $g \mapsto \sigma_g$. This map is trivially well-defined, so all we need to check is the homomorphism: $\varphi(gh) = \varphi(g) \circ \varphi(h)$ (recall that the operation on S_A is function composition). Indeed, for any a:

$$\begin{aligned}
\varphi(gh)(a) &= \sigma_{gh}(a) \\
&= (gh) \cdot a \\
&= g \cdot (h \cdot a) && \text{key group-action axiom} \\
&= \sigma_g(\sigma_h(a)) \\
&= (\sigma_g \circ \sigma_h)(a) \\
&= (\varphi(g) \circ \varphi(h))(a)
\end{aligned}$$

showing that the homomorphism is indeed well-defined, as we sought to show

(Copy from other thm) In a sense, what theorem 4.2.1 shows is showing injectivity and homomorphism. We now need to show that the function is surjective. We'll then show surjectivity, that for every homomorphism, there exists a group-action.

Let's say we had homomorphism f. If a group-action mapped to it, then we'd have surjectivity. If a group-action maps to it, then $g \cdot a = f(g)(a)$, by our function. Then

$$\begin{aligned}
(gh) \cdot a &= f(gh)(a) \\
&= (f(g) \circ f(h))(a) && \text{homomorphism condition} \\
&= G(H(a)) && \text{abbreviated } f(g)(f(h)(a)) \\
&= g \cdot (h \cdot a)
\end{aligned}$$

Similarly, we want to show that $e \cdot a = a$

$$e \cdot a = f(e)(a) = f(ee)(a) = (f(e) \circ f(e))(a)$$

again by the homomorphism condition, this implies that

$$f(e)(a) = (f(e) \circ f(e))(a) = f(e)(f(e)(a))$$

for every a. Since f is a bijection, it has an inverse, hence

$$(f(e))^{-1}(f(e)(a)) = f(e)(a) \Rightarrow \mathrm{id}(a) = f(e)(a)$$

showing that:

$$e \cdot a = a$$

And thus, for every homomorphism, there is an associated group-action, completing the proof

Since the axiomatic representation can be used to define a homomrophism, we could have equivalently defined group actions like so:

Definition 4.2.2: Group Action

Let G be a group an S be a set. A function $\varphi : G \times S \to S$ is a *group action* if

1. $\varphi(g, \varphi(h, s)) = \varphi(gh, s)$ (also written as $g \cdot (hs) = (gh) \cdot s$)
2. if e is the identity of G, then $\varphi(e, s) = s$

Notationally, $g \cdot s$ or gs is written to represent the function (g, s)

Many textbooks define group actions as these two axioms, and then build up to show that this will define a unique homomorphsim from G to S, and call this homomorphism the *premutation represetnation*:

Definition 4.2.3: Permutation Representation

If G is a group acting on a set A. Then ***premutation representation*** of action by G is the homomorphisms of G into the symmetric group S_A ($\cong S_n$) which satisfies the group action G on A. We shall say a given action of G on A *affords* or *induces* the associated permutation representation of G.

In this book, the first definition was chosen to be the leading definition, since the definition of group action presetend at the beginning of this chapter generalizes very naturally into a great multitudes of domains in matheamtics [1]. As we'll see soon, instead of letting S be a set, we can let S be any other algebraic (and even non-aglebraic) object, and still define a group action. If we swap the group G for another algebraic object, we will get a [algebraic-object]-action (part III is dedicated to the study of ring actions on abelian groups).

Personally, I also find it easier to remember there existing a homomorphism rather than remembering the axioms. This might not seem like much of a problem since there are only 2 axioms, but it might become easier once we reach introduce other actions (like modules) which will have many more properties. It might then be easier to remember the singel condition that the map $A \to \mathrm{Aut}_C(B)$ is a homomorphism.

Exercise 4.2.1

1. Dummit says we can think of this analogously to how there is a matrix representation of linear transformations. (Turn this into an exercise? Mabye that every S_n is a linear group?)

4.2.1 A Word on Right-actions

What we defined here is technically the *left* group action. There is an analogous *right* group action, which is defined as

$$(g, s) \mapsto sg^{-1}$$

[1] For ex. Module theory are rings (covered in part II)acting on abelian groups. We will study Modules extensively (in part III). Other this definition is found is dynamical systems in a class on differential equations, or represetntation theory (see part ref:HERE), or in many geometric settings like representation of affine spaces, or in topology in ref:HERE, and more generally as action in category theory (see section ref:HERE)

since if we didn't have the inverse, associativity would break:

$$(gh) \cdot s = sgh \qquad \Rightarrow \qquad g \cdot (h \cdot s) = (h \cdot s)g = (sh)g = shg \neq sgh$$

Joe also mentioned that the inverting g is also to preserve when the group-action is commutative:

$$s \cdot g = g \cdot s$$

but I haven't encountered this anywhere yet

Also how you can define a right-group action from a left one using $a \cdot g = g^{-1} \cdot a$, which is also why we tend to do what we just did up there.

— used to be section here

Let's say G is a group acting on S. Recall the definition of the group-action kernel (definition 4.3.4) that a kernel of $\psi : G \times S \to S$ ($\ker(\psi)$) that contains all $g \in G$ where $g \cdot s = s$ for every $s \in S$. With the permutation representation of a G, we can think of the kernel of $\varphi : G \to S_A$ ($\ker(\varphi)$) as all elements g who map to σ such that $\sigma_g(s) = s$. Since φ is a homomorphism, then $\ker(\varphi)$ is a normal subgroup.

Let's take two elements $a, b \in G$, and let's say that $a, b \in gK$ where K is the kernel of the associated permutation representation, and gK is a coset. What does that tell us about how a, b act on S? Since $a, b \in gK$, $a = gk$, $b = gk'$, so:

$$\varphi(a)(s) = \varphi(gk)(s) = \varphi(g)(s) = \varphi(gk')(s) = \varphi(b)(s) \qquad \forall s \in S$$

that is, a, b will act the same way on s! That means that if we let $G/\ker(\varphi)$ be the group acting on A, we know have that our action ψ' is a *faithful* group action!

If that makes sense to you, then you have a good grasp on the permutation representation of the action.

So the next question is, can we gain information about A? A is just a set, so any "information" we give it means information of the group-actions that corresponds with A. So far, the only extra thing we've labeled are the sets $X_s = \{g \cdot s | g \in G\}$ which represents where all the points s can get mapped to. With the information that the group action links to a group partly defined on G, that means that we can bring in the theory of cosets. Since G is a group, and G_a (the stability of a, definition 4.3.2) is a subgroup, then $|G_a| \big| |G|$, or equivalently G_a partitions G via it's cosets. Since we have cosets partitioning G, then we can have an equivalent notion for the orbit partitioning A. The difference will turn out to be that if we have $|G_a|$, then $|X_a|$ will be $|G : G_a|$. You can think of this as $|G| = |X_a||G_a|$, so you can think of it as a sort of Lagrange theorem on orbits.

The way we will define the orbits here might seem strange at first, but recall that if you define an equivalence relation, you've defined a partition. Next, if we take X_a. Then $a \sim b$ if and only if $a \cdot g = b$, if $a \sim b$ then we know that $g \cdot b = a \Rightarrow b = g^{-1} \cdot a$, which implies that $b \in X_a$. So, since cosets can also be thought of as partition, you can see how we can the concept of equivalence relations in the next theorem.

Furthermore, we will soon introduce the concept of a transitive action, which will be inspired by the transitivity property of an equivalence relation.

Theorem 4.2.4: Orbit-Stabilizer Theorem

Let G be a group acting on a nonempty set A. The relation on A defined by

$$a \sim b \qquad \text{if and only if} \qquad a = g \cdot b \, (a = \sigma_g(b)) \text{ for some } g \in G$$

is an equivalence relation that defines X_a, that is, if we fix some $a \in X_a$, then the above equivalence relation will list out all elements of X_a.
Furthermore, for each $a \in A$, we can define X_a, and

$$|X_a| = |G : G_a|$$

i.e. the index of the stabilizer of a, where G_a is the stabilizer of a ($\{g \in G | g \cdot a = a\}$)

Proof **:**
Proving the equivalence relation is like all other times, so it will be skipped

Note that since if $a \sim b$, then there exists a g such that $g \cdot a = b$. By transitivity, we'd then get another element g' such that $g' \cdot a = c$. We can do this for all elements, to get all possible combinations. However, all these elements can be reduced to the form $g \cdot a$. So, when thinking about the equivalence class, we can think of a as a "representative". Furthermore, we can then take our $g \cdot a$ and multiply it by ever element of g to get to all possible elements in the equivalence class, thus let $\mathcal{C}_a$ be the equivalence class of a, so

$$\mathcal{C}_a = \{g \cdot a | g \in G\} = [g \cdot a]$$

Now, we need to show that the equivalences classes indeed are of cardinality $|G : G_a|$. Note that gG_a are the left-cosets. If we can bijectivity map the $\mathcal{C}_a$ to the cosets G_a (that is gG_a for every $g \in G$), then there would be the same number of cosets as elements in the equivalence classes, which would imply the cardinality to the index of the stabilizers. More rigorously, we want to show that

$$\varphi_a : \mathcal{C}_a \to G/G_a \qquad g \cdot a \mapsto gG_a$$

is a map form $\mathcal{C}_a$ to the left-cosets of G_a in G which is bijective. Here are a couple of computations to see what's happening with this function

1. Let's fix G_a as our stabilizer.
2. if $g \in G_a$, then $[g \cdot a] = [e \cdot a] = [a]$, showing how we clearly remain in the same equivalence class
3. if $g \notin G_a$, then $[g \cdot a] \neq [a]$, meaning their not in the same equivalence class. However, since G_a is a group, we can make cosets, and we know that since $g \in gG_a$, then we know it in *that* equivalence class. This comes down to knowing that G_a is a group to once again take advantage of Lagrange's theorem to be able to split up the group G in that manner.
4. Let's say $g \in gG_a$ (implying $g \notin G_a$ by Lagrange). If $a, b \in C_A$, then there exists some $g \in G$ such that $g \cdot a = b$ As we've seen before, $[g \cdot a] = [b]$. If we take $g' \in G_a$, then

$$[b] = [g \cdot a] = [g \cdot (g' \cdot a)] = [gg' \cdot a] \qquad gg'G_a = gG_a$$

that chain of equation should give you an idea on why this is true now (since g' will just keep a, and G_a will partition since it's a group)

First, we need to show this map is well defined. If $g \cdot a = g' \cdot a$, then is $gG_a = g'G_a$? Let's say $g \cdot a = a'$. Then for every $x \in gG_a$, $x \cdot a = (gx') \cdot a = g(x' \cdot a)g \cdot a = a'$. Since $g' \cdot a = a'$, $g'(x \cdot a) = g' \cdot a = a'$ meaning the elements of the coset $g'G_a$ also bring those elements to the same place, and so $gG_a = g'G_a$.

Next is to show bijectivity. This map is clearly surjective, since for any $gG_a \in G/G_a$, if we look at the representative g, then the element $g \cdot a$ is an element of $\mathcal{C}_a$ by construction. To show injectivity, suppose $gG_a = hG_a$. Then $h^{-1}g \in G_a$, implying $h^{-1}g \cdot a = a$, implying $g \cdot a = h \cdot a$, proving injectivity [a]. Thus, our map is bijective, and so A splits up the same way the stabilizers do, showing that A is a disjoint union of sets with respect to this equivalence relation.

[a] the implications are actually if and only if's

Note that the equivalences partitioning the underlying set doesn't tell us if G_a partitions the group. Actually, it can't since $e \in G_a$ for every a.

Also, don't mix the cardinality of the set and of the group; it seems very tempting with the two linked by the Orbit-Stabilizer Theorem. Take for example

$$S_{10} \times N \to N \qquad N = \{1, ..., 10\}, \qquad \sigma \cdot n = \sigma(n)$$

that is, the action which just permutes the set N. Note that there exists $\sigma_1, ..., \sigma_n$ such that $\sigma_i(1) = i$, so this action is hits all the elements. So the orbit is all of N for every X_n

$$X_n = N \qquad \forall n \in N$$

which only tells us that:

$$|X_n| = |N| = |S_n : (S_n)_a| \Rightarrow |(S_n)| = |(S_n)_a| \cdot |N| \Rightarrow |(S_n)_a| = \frac{10!}{10} = 9!$$

Note also that this doesn't tell us if the orbits are of the same size, simply that they partition. Exercise 9 on page 117 outlines when the orbits are the smae size.

One thing it does tell us is that the orbit's size is determined by a group, meaning two things: we can use group theory! so it means we can use all the group-theoretic things we found to find it's order and other properties of it to determine the size of the orbit. Secondly, if we have some nice groups as stabilizers (say, for example $C_G(g_i)$, the centralizer of a point), then we can deduce some properties/limitations of the orbit size (say that $|G : Z(G)| = n$, then since $Z(G) \leqslant C_G(g_i)$, then $|G : C_G(g_i)| \leq n$. Stabilizers being Centralizers will come up when dealing with the conjugacy action (section 4.4.2).

Recall how earlier we said that Orbits are like Cosets. With this theorem, we see how one-to-one the intuition between these two are (actually identical by the proof). To get a hang of this, here's an example

Example 4.2.1: : Concrete Case
Let R be the set of all polynomials with integer coefficients in the independent variables x_1, x_2, x_3, x_4

$(R = \mathbb{Z}[x_1, ..., x_4])$, and let S_4 act on R by permuting it's indices like so:

$$\sigma p(x_1, x_2, x_3, x_4) = p(x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(3)}, x_{\sigma(4)},)$$

Find the polynomials in the orbit of S_4 in R containing $x_1 + x_2$

Proof. Recall the intuition of representatives. Basically, asked to find $(S_4) \cdot (x_1 + x_2)$, where $x_1 + x_2$ is just one of the representatives of this orbit. We want to use the orbit stabilizer theorem to limit our search through all 24 polynomials (find the "zeroe"'s so to speak), i.e.:

$$(S_4)_{x_1+x_2} = \{g \in S_4 | g \cdot p(x) = p(x)\}$$

that is, all permutations which keep $x_1 + x_2$ the same. Since it's commutative, we know that $(1\,2)$ will keep it the same. Since x_3, x_4 are not part of it, $(3\,4)$ will keep it the same. Therefore, $(1\,2)(3\,4)$ will keep it the same. add in the identity, and that will actually be all possibilities. Anything else will mean shifting one of the variables. Thus, we have that $|(S_4)_{x_1+x_2}| = 4$, so by the Orbit-Stabilizer Theorem

$$S_4 \cdot (x_1 + x_2) = |S_4 : (S_4)_{x_1+x_2}| = 24/4 = 6$$

Thus, we know we're looking for 6 elements. By looking at some elements left in S_4, we see that elements which will produce different results are $(3\,4)$, $(1\,3)$, $(2\,4)$, $(2\,3)$..., giving us

$$S_4 \cdot (x_1 + x_2) = \{(x_1 + x_2), (x_1 + x_3), (x_1 + x_4), (x_2 + x_3), (x_2 + x_4), (x_3 + x_4)\}$$

□

Similarly to how simple groups only have one normal subgroup, there is the equivalent notion in group-actions:

Definition 4.2.5: Transitive Action

The action of G on A is called *transitive* if there is only one orbit, i.e., given any two elements $a, b \in A$, there is some $g \in G$ such that $a = g \cdot b$

I'll explain later. I found this to be insightful:

`https://encyclopediaofmath.org/wiki/Transitive_group#References`

One more definition I found useful to define

Definition 4.2.6: Free Group Action

A group action is said to be free if all it's stabilizers are the trivial group $(\{e\})$

For a bit, I confused free and transitive group-actions.

Also, want to eventually include

Lemma 4.2.7: Burnside'S Lemma

$$|X/G| = \frac{1}{|G|}\sum_{g\in G}|X_g|$$

Proof. `https://en.wikipedia.org/wiki/Burnside%27s_lemma` □

4.3 Orbit Stabalizer Theorem

(Build up like in Aluffi)

The main motivating example to think about (which also was studied in the creation of group actions, and so motivated some vocabulary) is the group of invertible 3×3 matrices with determinant of 1 ($SO(3)$), which represent all rotation, acting on the unit sphere S^2. If we take any point, say the north pole $\{N\}$ then

$$SO(3) \cdot N = S^2$$

that is, the entire sphere. However, if we limit to the group $SO(2)$ rotating the point about the x-axis, then:

$$SO(2) \cdot N = \{N\}$$

which makes sense, since we're at the pole. If we consider a point D which is not a pole, then:

$$SO(2) \cdot D = \{\text{latitude line intersecting } D\}$$

Sometimes, we're interested in this set (with a subgroup acting on only 1 element). This set is called the *orbit* of D (i.e. our point). If you think about the visual with the sphere, the naming should make sense.

Definition 4.3.1: Orbit

Let G act on a set S. Then

$$X_a = \{g \cdot a | g \in G\} \qquad \text{or} \qquad Ga = \{g \cdot a | g \in G\}$$

That is, X_a (or Ga or $G \cdot a$) represent all the elements g will take a to, for every $g \in G$.

But what about the example with the north pole? That doesn't look like an "orbit". And it seems to be doing something a little different too: all the action of G return the same point. When this happens, we give a special name to the subgroup that doesn't change that point:

Definition 4.3.2: Stability Group or Stabilizer of s

if $s \in S$, the *stability group of* s is $\left\{g \in G \middle| g \cdot s = s\right\}$. This is often denoted G_s

Proposition 4.3.3: Stabilizer is a group

the stabilizer G_a is a group.

If this happens for every point, then we also have a name for this:

Definition 4.3.4: Kernel Of Group G

if for every element $s \in S$ $g \cdot s = s$, then g is said to be in the kernel of G.

Note: This is a touch confusing at first, but both G_a and X_a have an element of the set in the subscript. While G_a represent all elements $g \in G$ that keep a at a, X_a is part of the underlying set A that we get when we take every $g \cdot a$. Here is how I sometimes remember which is which. If $G_a = G$, then that means every element of G sends a to a, so $X_a = \{a\}$, so in this way they're kind of opposites in a very natural way. The relation between the two will be expanded shortly (namely, $|G| = |G_a||X_a|$)

Because of this, thinking about the Orbit as $G \cdot a$ will eventually lead to thinking about a is a "representative" of the orbit, just like if we have aK a is a representative of the coset! This will be made rigorous once we prove the Orbit Stabilizer Theorem (4.2.4)

Note that since the kernel of an action is elements which stabilize *all* points, then

$$\ker(G) = \bigcap_{a \in A} (G_a)$$

If this happens for *no* points, then we also have a name for this:

Definition 4.3.5: Faithful Group Action

An action is ***faithful*** if its kernel is the identity

4.4 Application to Finite Groups

(Words here)

4.4.1 Left Multiplication

So far, our group-action was acting on arbitrary sets. We'll now make it act on itself. Perhaps the first natural group action can come to mind is:

$$g \cdot a = ga \qquad \forall g \in G,\, a \in G$$

Since we saw that group-actions have a permutation representation, and G is finite, we can label the elements of G $(g_1, g_2, ..., g_n)$ so that we can think of the previous action as

$$gg_i = g_j$$

or in permutation terms:

$$\sigma_g(i) = j$$

What this lets us do is see how all the elements act when multiplying any other element. This is a good example where group actions are seen as being more general than homomrophism: the map $\varphi_g : G \to G$ sending $h \mapsto gh$ is not a homomorphism unless $g = e$.

Example 4.4.1: Left Multiplication Permutation Representation
The following show how to think about group-actions on themselves and how to think of the equivalent permutation representation.

1. Take $G = \{1, a, b, c\}$ to be the Klein 4-group. Label $1, a, b, c$ with $1, 2, 3, 4$ respectively. With this, we can find the permutation representation of a:

$$\begin{aligned} a \cdot 1 &= a \text{ and so } \sigma_a(1) = 2 \\ a \cdot a &= 1 \text{ and so } \sigma_a(2) = 1 \\ a \cdot b &= c \text{ and so } \sigma_a(3) = 4 \\ a \cdot c &= b \text{ and so } \sigma_a(4) = 3 \end{aligned}$$

   Showing that with this labeling, we get that $\sigma_a = (1\,2)(3\,4)$. One can similarly get that σ_1 is the identity, and

$$\sigma_b = (1\,3)(2\,4) \qquad \sigma_c = (1\,4)(2\,3)$$

   and if we map $x \mapsto \sigma_x$, then this map is indeed a homomorphism form G to S_4, as can be checked.

Example 4.4.2: exercise

1. Show that the group action of left multiplication is always a free group action (see definition ref:HERE).

Consequences of Left Multiplication Action

This is useful since we can get an idea of how a element of a group actions on every other element, but we won't get much information about the structure of the group. However, we can get this if we take the following group action

$$\varphi : (G \times G/H) \to G/H \qquad g \cdot aH = gaH$$

Note that H need not be normal!! We are doing this to partition G. This is in a sense a generalization of the previous action, since if $H = \{e\}$, the group actions will in fact be the previous one we defined.

We can again consider the permutation representation, so that if the index of H is finite, then we can label the cosets $a_1H \cdots a_mH$ and:

$$\sigma_g(i) = j \qquad \text{if and only if} \qquad ga_iH = a_jH$$

Example 4.4.3
Let $G = D_8$ and $H\langle s\rangle$. You can do the same process as the last question to come up with a permutation representation for the cosets.

So as we mentioned in our intuition, group actions can get us some more information about the group they act on. In this case, what extra information do we get and why? The "why" in my mind, is explained by thinking of a group action representing *all* possible combinations of elements of G with itself (or G with G/H). Because of this, the action is kind of capturing all possible scenarios of ways of thinking about how elements interact with each other at once. Another way of thinking that way is by considering the permutation representation: since every element g maps to a homomorphism, from G to S_n, then every element is like a function, which has information about where it maps every point in the domain! Thus, we might be able to figure some information out as a consequence of this. The following theorem's sum up what we can get from this group action:

Theorem 4.4.1: Result of Group Acting on it's Cosets

Let G be a group, let H be a subgroup of G, and let G act by left-multiplication on a set A of left cosets of H in G (which we'll denote G/H). Let π_H be the associated permutation representation afforded by this action. Then

1. G acts transitively on G/H
2. the stabilizer in G of the point $1H \in G/H$ is the subgroup H
3. the kernel of the action (i.e. the kernel of π_H) is $\bigcap_{x\in G} xHx^{-1}$, and $\ker(\pi_H)$ is the largest normal subgroup of G contained in H.

***Proof* :**
we'll prove each of these separately:

1. Since $A = G/H$, and $g \cdot aH = (ga)H$, then it's simply take $g = ba^{-1}$ then
$$(ba^{-1}) \cdot aH = (ba^{-1}a)H = bH$$
and since b was arbitrary, that means that aH is associated by transitivity to bH for any b, and so the action is transitive
2. Consider all points $g \in G$ such that $g \cdot 1H = H$. Then we'd have
$$H = g \cdot 1H = (g1)H = gH$$
That is, $gH = H$, which is only true when $g \in H$, completing the proof

3. The kernel of this action is $\bigcap_{a \in A} G_a$. that is

$$\begin{aligned}
\ker(\varphi) &= \bigcap_{xH \in G/H} \left\{ g \in G \middle| g \cdot xH = gxH = xH \right\} \\
&= \bigcap_{xH \in G/H} \left\{ g \in G \middle| = x^{-1}gxH = H \right\} \\
&= \bigcap_{xH \in G/H} \left\{ g \in G \middle| = x^{-1}gx \in H \right\} \\
&= \bigcap_{xH \in G/H} \left\{ g \in G \middle| = g \in xHx^{-1} \right\} \\
&= \bigcap_{x \in G} xHx^{-1}
\end{aligned}$$

which is exactly what we were going for. Note that if $y \in H$, then $xyx^{-1} \in \bigcap_{x \in G} xHx^{-1}$. And since this set was constructed by taking every possible "normalizer per element", this means that this must be the *largest* normal subgroup of H.

This leads us to one of the larger "philosophical" theorems of this course

Theorem 4.4.2: Cayley's Theorem

Every group is isomorphic to a subgroup of some symmetric group. If G is a group of order n, then G is isomorphic to a subgroup of S_n

Proof **:**
Let $H = 1$, so we have that $G \times G/\{1\} \to \overline{G}$. Now take the premutation representation: $G \to S_{\overline{G}}$. This is a homomorphism; if the kernel is trivial, then it's an isomorphism. Indeed

$$G \cdot \overline{g} = \{g \in G | g \cdot \overline{g} = \overline{g}\} = \{e\}$$

Since $G \cong \overline{G}$, so we can use the uniqueness of the identity. Thus

$$\ker \varphi = \{e\}$$

Since the kernel is trivial, G is isomorphic to its image in S_G.

This is an absolutely fascinating result. For the range of places group-actions are used to "represent" actions (like Lorenz Transformations for moving through time, or the conservation laws in physics), all of them can ultimately be thought of through the intuition of symmetry (in the case of the conservation laws, conservation in momentum can be thought of was translations in nspace, and conservation of energy can be thought of as translation in time)

16: trivia

When group theory started, it actually strated with the study of symmetries just like this! Then, someone axiomized it, making the symmetric group a particular group, and then Cayley realised that the original "intuition" to think about it in terms of symmetries can still be applicable

I said "philosophical" theorem, because though it gives us an insight on the nature of groups, using this theorem is usually not practical. $|S_n| = n!$, which can be **huge**. For example, if $n = 100$,

$$|S_{100}| = 93'326'215'443'944'180'000'000'000'000'000'000'000'000'000'000$$
$$'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000$$
$$'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000$$

Working with the subgroup form it might not actually be much nicer than working with G. There are instances where we *can* deduce extra information ,because we *do* have a homomorphism to S_n meaning we can perhaps use the 1st Isomrophism Theorem or other Isomorphism theorem and deduce important information. In a sense, Compositiones series are a better intuition for "base" or representation using simple groups than this representation. However, this is also one of my favourite theorems of the course, since we now get a very clear idea of to what level we should be thinking about groups! And it gives a great light to why symmetric groups are really important.

Corollary 4.4.3: generalizing normality when index is 2

If G is a finite group of order n and p is the smallest prime dividing $|G|$, the many subgroup of index p is normal

Note: In general, a group of order n need not have a subgroup of index p. For example, A_4 has no subgroup of index 2, as we've shown that there is no subgroup of order 6 when finding a counter-example to the converse of Lagrange (3.3.1)

Proof **:**
p.120 Dummit and Foote

Here are a couple of exercises I found insightful:

Example 4.4.4: Using Group Actions to Get information
Here is a result of a question in the book I thought was worth pointing out. We know that $|Q_8| = 8$. Since we can have Q_8 act on itself, by Cayley's theorem, we have that $Q_8 \cong S_8$ where S_8 denotes a subgroup of S_8 defined by the image of Q_8 in it' permutation representation. Note that Q_8 acted on another set of cardinality 8. What if it acted on some A, such that $|A| \leq 7$? Then it's impossible that the kernel be empty, that the group-action be faithful. However, I think it's still possible for it to be transitive (if we take S_8 acting on A).

Proof. Let Q_8 act on some set A such that $|A| \leq 7$. Then we know that $|Q_8 : (Q_8)_a| = |Q_8 \cdot a|$ where $Q_8 \cdot a \subset A$. Since $|A \leq 7$, then $|Q_8 \cdot a| \leq 7$, since it's also a subset. However, that implies

that $|Q_8| = |Q_8 \cdot a||Q_8|$ where the $Q_8 \cdot a$ will never have 8 elements, meaning $(Q_8)_a$ must always have 2 or more elements! Since a was arbitrary, this is true for every orbit. Note that for Q_8, any subgroup *must* contain -1. Thus, -1 *must* be in every stabilizer, and hence it's also in the kernel, and hence the kernel is non-trivial, and so we don't have an isomorphism. □

Example 4.4.5: : D_8 in S_8
For practice, let's represent D_8 as a subgroup of S_8. This would be one worth remembering since it gives you some idea of how imposing some structure on S_8 by taking a subgroup of it can create other groups.

Exercise 4.4.1

1. We have seen that for any group, there exists an embedding $G \to S_{|G|}$. Show that for any *finite* group, there exists an embedding $S_n \to GL_n(\mathbb{F})$ ($\mathbb{F}$ or $\mathbb{F}_n$?). Thus, every finite group is a subgroup of the matrix group.
2. (very hard) is this the case for all infinite groups? [No, but I can't recall one]

4.4.2 Conjugation

We will now define a different group-action to find different structural information on a group.

Let the group action be $\varphi : G \times G \to G$,

$$g \cdot a = gag^{-1}$$

This one is useful in that it will let us get information about the centralizers/normalizer (2.2.4, 2.2.5) of G (which should already make some sense if you think about what the group-action is doing, look back at the definition of centralizer and normalizer).

One can check the actions to see that it is indeed a group action. When we related two elements before, we said that $a \sim b$ if and only if $a = g \cdot b$ for some g. we'll be doing the same here, but this time we call it *conjugacy*.

Definition 4.4.4: Conjugacy reminder

Two element $a, b \in G$ are said to be ***conjugate*** if there is some $g \in G$ such that $b = gag^{-1}$ (i.e. if and only if they are in the same orbit of G acting on itself by conjugation). The orbits of G acting on itself by conjugation are called the ***conjugacy class*** of G

Exercise 4.4.2

1. If G is abelian, then $g \cdot a = gag^{-1} = a$, therefore, $G_a = G$, meaning also that the action is not transitive, since $X_a = \{a\}$.
2. If $|G| > 1$, then the group action will not be transitive, since $G_1 = G$, so $X_1 = \{1\}$.

We'd like to use theorem 4.2.4 to find the order of the orbits. We first find the order of the

stabilizers. Let $a \in G$. then:

$$G_a = \{g \in G | g \cdot a = gag^{-1} = a\} = C_G(a)$$

that is, we get that it's the centralizers of a in G, and so $X_a = |G : G_a| = |G : C_G(a)|$

As before, we don't get much if we simply act on elements by conjugation. We generalize this by acting on sets:

$$\varphi : G \times (G) \to (G) \qquad g \cdot S = gSg^{-1}$$

In case the idea of conjugating a set is not very clear:

Definition 4.4.5: Two Sets Being Conjugated

Two subsets $S, T \subseteq G$ are said to be *conjugate in* G if there is some $g \in G$ such that $T = gSg^{-1}$ (i.e. if and only if they are in the same orbit of G acting on its subsets by conjugation)

Now, by theorem (4.2.4), we know that $|X_S| = |G : G_S|$. And by definition of stabilizer, we know that

$$G_S = \{g \in G | g \cdot S = gSg^{-1} = S\} = N_G(S)$$

that is, the stabilizers are the normalizers of S in G, therefore, $|X_S| = |G : N_G(S)|$.

Corollary 4.4.6: Partitioning Relative To Conjugate Group-Action

The number of conjugates of a subset S in a group G is the index of the normalizer of S,

17: What Joe did in class:

Consider the left action $L : (g, h) \mapsto gh$ and right action $R : (g, h) \mapsto hg^{-1}$. Notice that L and R commute, that is

$$L_g \circ R_g = R_g \circ L_g : x \mapsto gxg^{-1}$$

or more generally

$$R_g \circ L_k = L_k \circ R_g : x \mapsto gxk^{-1}$$

We call the resulting function C_g the *conjugation*. The function C_g is another action of G on G, and a very common one. the orbit of C_g look like this: $\{gxg^{-1} | g \in G\}$, and is called the *conjugate class of* x.
Fun fact, the centralizer of C_g would be $Z(G)$ since we'd need that $gxg^{-1} = x \Rightarrow gx = xg$ for all $g \in G$.

Note: C_g can be thought as a left or right group action. For a right group-action, take

$$(g, h) \mapsto g^{-1}hg$$

4.4.3 Class Equation

Let's stick to the action by an element. What information does this group-action give us about G? We saw before that we can partition G, meaning that we can split $|G|$ into the sum of $|G : C_G(g_i)|$ for

appropriate i. However, since we're conjugating, we can ask ourselves what are the conjugacy classes if $g \in Z(G)$ or $g \notin Z(G)$ (definition 2.2.3). If $g \in Z(G)$, then $hgh^{-1} = g$ for every h, so $h \cdot g = h$, so $X_g = \{g\}$, meaning that the conjugacy class of every element in the Center has a cardinality of 1. We can thus split up $|G|$ in a new way which will give us a bit more information about it:

Theorem 4.4.7: Class Equation

Let G be a finite group and let $g_1, ..., g_n$ be the representatives of distinct conjugacy classes of G not contained in $Z(G)$ of G. Then

$$|G| = |Z(G)| + \sum_{i=1}^{n} |G : C_G(g_i)|$$

Proof **:**
You already know that the orbits partition G. What you're doing is just squishing the orbits of elements from $a \in Z(G)$ into just $Z(G)$ cause those are the exact elements whose orbits are singletons.

Though the theorem might seem obvious, splitting up a group relative to it's centralizers actually can come in quite handy. We will show a couple of examples, including a very neat result for conjugacies of permutations, and that A_5 is simple.

Example 4.4.6
here

1. The class equation doesn't really give information on abelian groups since $|G| = |Z(G)|$, which we already knew
2. Let G be nonabelian. For example, let $G = Q_8$. Since cyclic groups are abelian (corollary 2.3.9), then $\langle g \rangle \leqslant C_G(g)$. Thus, we can find information about it's conjugacies more easily, especially since we know that the conjugacy class has order $|G : C_G(a)|$. Take $\langle i \rangle \leqslant C_G(i) \leqslant Q_8$. Since $i \notin Z(G)$, and $|G : \langle i \rangle| = 2$, then $\langle i \rangle$ is normal, and thus $C_G(i) = \langle i \rangle$. Thus, we know that $|G : C_G(i)| = 2$, and so i has 2 conjugates in Q_8, namely i and $-i = kik^{-1}$.

Our first general result we can use this theorem for is deducing that groups of prime power order p^α must have non-trivial centres ($|Z(G)| > 1$)

Theorem 4.4.8: Centers Of Prime-Order Groups

If p is prime and P is a group of prime power order p^α for some $\alpha \geq 1$, then P has a nontrivial center $|Z(P)| > 1$

Proof **:**
By the class equation, we know that

$$|P| = |Z(P) + \sum_{i=1}^{n} |P : C_P(g_i)|$$

for appropriate p_i's. Since $C_P(p_i)$ is a group, $|C_P(p_i)| \big| p^\alpha$, meaning $|C_p(p_i)| = p^\beta$ for some $\beta_i \leq \alpha$. Note that $|C_P(p_i)| = p^\alpha$, then $Z(P) = C_p(P)$, so the p would divide the centre, so let's assume there isn't a $C_P(p_i)$ such that $|C_P(p_i)| = p^\alpha$. Conversely, if $|C_P(p_i)| = 1$, then $|P : C_P(p_i)| = p^\alpha$, but then the centralizer would be empty, but it must always have $1 \in Z(P)$ So we get $\beta_i < \alpha$. Continuing, we get that $|P : C_P(p_i)| = p^{\beta_i'}$ for $\beta_i' = \alpha - \beta_i$. Since $\beta_i < \alpha$, $\beta_i' < \alpha$, so $p|P : C_P(p_i)|$ for every p_i. Similarly, we know that $|Z(P)| \big| p^\alpha$, meaning that $Z(P) = p^\beta$ for it's own beta. Let's say $\beta = 0$. Thus:

$$p^\alpha = 1 + \sum_{i=1}^{n} p^{\beta_i'}$$

Since $1 < \beta_i' < \alpha$, that means that you can think of the sum on the left as decomposing to a bunch of p's that need to be added up, and similarly to the right. Thus, having a 1 will guarantee the two sides won't add up then, and hence, $\beta \geq 1$, and so $p \big| |Z(P)|$, and so $|Z(P)| > 1$.

Corollary 4.4.9: P and $|P| = P^2$ then P is abelian

If $|P| = p^2$ for some prime p, and P is abelian. More precisely, P is isomorphic to either Z_{p^2} or $Z_p \times Z_p$

Proof **:**
p.125

Application to S_n

We can also get some nifty results on symmetric groups. S_n is a really important group when studying group actions, so we'll take a moment to explore it's properties. We can take this moment to exercise our understanding of conjugate classes by finding all the conjugate classes of S_n. Let's start with S_4 for simplicity. Let's take a particular example:

$$(1234)(123)(4321) = (234)$$

Notice that the "order" of the element is the same, so we wouldn't have (1234) and (123) being in the same conjugate class. Next, we can think of what happens if we "combine" two permutation like

$$(123)(34)(34)$$

however, (34) is it's own opposite, so we get back to (123). So these two can be said to have different "cycle types", but are the same element. Overall, if we take $\sigma \in S_n$, and write it as a product of disjoint cycles, *then* the cycle type determines the conjugacy class, and vice versa. So

$$S_4 : \{e\}, \{(ab)\}, \{(abc)\}, \{(abcd)\}, \{(ab)(cd)\}$$

We can expand on that last sentence. Here is a very useful fact about symmetric groups. Any cycle can be broken down to transpositions (2-cycles)

$$(123)(246)(35) = (15)(13)(16)(14)(12)$$

$$(256)(145)(53) = (13)(15)(12)(16)(14)$$

$$(123)(456) - (46)(45)(13)(12)$$

Because of this fact, we can write

$$S_n = \langle (ab) \big| 1 \leq a < b \leq n \rangle$$

Note that because $(12)(24)(56)(56) = (12)(24)$, the number of transpositions need not be well-defined.

Note: **Move this over to closer to the beginning of group actions.** However, there is something we can define out of it.

Definition 4.4.10: Parity

The *parity* (even or odd) of the number of transposition needed to expresses $\sigma \in S_n$ *is* well defined. We say that σ is even or odd accordingly

Example 4.4.7
We'd have

1. (12) odd
2. $(123) = (13)(12)$ even
3. in general, $(123...2k)$ is odd, while $(123...2k+1)$ is even
4. we'd also have that if a cycle σ, τ are even/odd, and we multiply them, then the parity of the result would be like multiplying the parities (so odd*odd = odd, even*even = even, odd*even = even, even*odd=even)

1. when two elements are conjugate – define *cycle-type* as consequences
2. examples from book
3. Since Centralizer is stabilizer of stabilizer under conjugation, and orbit is conjugation class, and we know about conjugation class now, so conjugation class gives predictable formula
4. bonus, can conclude: Stabilizer with m-cycles in S_n is _____(p.127)
5. A_5 is simple
6. $A_n \trianglelefteq S_n$ since $\ker(\epsilon) = A_n$. Bonus: $V_4 \leqslant A_4$

4.4.4 Sylow's Theorem

We've now arrived at one of the more powerful theorems of the course. The following buildup will let us find the number of groups of a particular order! In other words, is very powerful in it's use of classifying groups of a given order.

Personally, I also find it interesting how prime-numbers come in here. It shows how related the ideas decomposing numbers into primes comes in when "decomposing" groups. To show this, we'll start by labeling groups of prime-power order:

Definition 4.4.11: p-Group

Suppose p is a prime. A group P such that $|P| = p^a$ for some $a \in \mathbb{N}$ is called a p-group

Example 4.4.8
here

1. $\mathbb{Z}/p\mathbb{Z}$
2. vector space V, since $|V| = |\mathbb{F}_p^n| = p^n$
3. cyclic group of order p^a

Note every group however might have a p-group as a subgroup. It turns out though that every group will have something closely related:

Definition 4.4.12: Sylow p-group

Suppose G is a finite group such that $|G| = p^a m$, where $(m, p) = 1$, i.e. $p \nmid m$ (meaning that α is the largest value possible, if $p^\beta n = p^\alpha m$, where $\beta < \alpha$, then $p|n$). Then a subgroup $P \leqslant G$, whose order is p^a then G has a subgroup of order p^a, then P is called a p-Sylow subgroup, or Sylow p-group.

Note: Since $p \nmid m$, α is the largest possible number. Because of this, we have some notion that G is a *maximal* p-group.

Example 4.4.9
here

1. Take S_3 and $\langle r \rangle$ as a subgroup such that $\{e, r, r^2\}$. Note that $|S_3| = 2 \cdot 3$ so $\langle r \rangle$ is a Sylow 3-group. Similarly, if $\langle \sigma \rangle$ such that $|\langle \sigma \rangle| = 2$, it is a Sylow 2-group.

I mentioned that a Sylow p-group will always exist, but I didn't prove it. One might ask if Sylow p-groups always exist for *any* finite group? One can further ask if we find Sylow p-subgroups, are they related, or how many Sylow p-subgroups an arbitrary finite group G would have. Instead of thinking any finite-group for now, we can simply and restrict our attention to groups of order $p^\alpha m$, where $p \nmid m$. This simplification is actually incredibly powerful, since we can show that *all* such

groups have Sylow p-subgroups, and further that they are *all* isomorphic to each other. This is precisely the contents of *Sylow's Theorem*:

Theorem 4.4.13: Sylow's theorem

Let G be a group such that $|G| = p^a m$, $p \nmid m$, then

1. **(Existence)** there exists a p-Sylow subgroup
2. **(Similarity)** For fixed p, the p-Sylow subgroup are conjugate (i.e. you can relate Sylow p-groups via conjugation, and by 4.5.2, this means they're isomorphic!)
3. **(number of them)** For fixed p, the number of p-Sylow subgroups, sometimes denoted n_p or $n_p(G)$, Satisfies $n_p \equiv 1 \bmod (p)$. Also $n_p(G) | \frac{|G|}{|N_G(P)|}$ if P is a Sylow p-subgroup. Wikipedia phrases it this way:
   (a) $n_p | m$, which is the index of the Sylow p-subgroups
   (b) $n_p \equiv 1 \bmod p$
   (c) $n_p = |G : N_G(P)|$ where eP is any Sylow p-subgroup of G and N_G denotes the normalizer

***Proof* :**
To simplify the proof, we'll first show a lemma

Lemma 4.4.14: Intersecting Sylow p-subgroups with p-subgroups

Suppose P is a Sylow p-subgroup of G, and let Q be any p-subgroup of G. Then $Q \cap N_G(P) = Q \cap P$.

Proof. The ($\supseteq$) direction is clear. So we'll proof ($\subseteq$). Let $H = Q \cap N_G(P) \subseteq N_G(P)$. We'll take advantage of the fact that the order of multiplicity of the prime power that divides P (the α in p^α is the largest possible value – *maximal*). Since Q is a p-group, then $Q \cap N_G(P)$ must also be a p-group, and hence H is a p-group.

HP is a subgroup, then $HP = \frac{|P| \cdot |H|}{|P \cap H|}$. Note that $P, H, P \cap H$ are all p-groups. so $|HP|$ is a power of p But since $HP \supseteq P$, and that P is a maximal p-group, then $HP = P$, which implies

$$H \subseteq P$$

And since $H \subseteq Q$, then $\Rightarrow Q \cap N_G(P) \subseteq P$, as we sought to show. □

Now we can start the proof. We will do a proof by induction. If $|G| = 1$, then it's trivially true since we would just have G as the Sylow p-subgroup and $n_p = 1$, so let $|G| = p^\alpha m$, $p \nmid m$, and that it's true for all appropriate $|H| < |G|$.

We first want to show that the Sylow p-subgroup exists. If $p || Z(G)|$, then since $Z(G)$ is abelian, we know by Cauchy's theorem (3.3.5) that there exists a subgroup $N \leqslant Z(G)$ with $|N| = p$. Since $N \leqslant Z(G)$ then $N \trianglelefteq G$ (since it's in the normalizer). So we can consider $G/N = \overline{G}$. If$|G| = p^\alpha m$, where $(m, p) = 1$ (which is what we're given for G) then $|\overline{G}| = p^{\alpha-1} m$. Thus, we found a group of

smaller order satisfying our induction hypothesis, so by assumption there exists some $\overline{P}$ such that $|\overline{P}| = p^{\alpha-1}$. From this, we can define $P \leqslant G$ such that $P/N = \overline{P}$. Then

$$|P| = |P/N| \cdot |N| = p^{\alpha}$$

showing that is indeed maximal, and hence a Sylow p-subgroup.

So let's say $p \not\big| |Z(G)|$. We know by the class equation that

$$|G| = |Z(G)| + \sum_{i=1}^{n} |G : C_G(g_i)|$$

where the g_i's are representatives of central conjugacy classes. If we had that $p\big||G : C_G(g_i)|$ for every g_i, then since $p\big||G|$, then that'd imply that $p|Z(G)$, a contradiction for we assumed $p \nmid Z(G)$, so we know there is some g_i such that

$$p \nmid |G : C_G(g_i)| \Rightarrow p \nmid k$$

for some appropriate k. Then, by Lagranges theorem, $|C_G(g_i)| = p^{\alpha}k'$ such that $k \cdot k' = m$ and $p \nmid k'$. Note that $C_G(g_i)$ is a smaller group satisfying the induction hypothesis, and so there exists a Sylow p-subgroup P for $C_G(g_i)$ such that $|P| = p^{\alpha}$. Further, by construction, this must also be a Sylow p-subgroup for G (sine $|G| = p^{\alpha}m$ and $|P| = p^{\alpha}$), hence there exists a Sylow p-subgroup, completing the induction, and hence showing every Group of order $p^{\alpha}m$ has a Sylow p-subgroup.

We know turn to proving part 2 and 3. Or at least, I will in the future (p.140), decomposing this theorem has proved to be a boring task with the way Dummit is doing it.

So, let's get back to $|G| = p^{\alpha}m$. This seems specific at first glance, why can't this be true in general? However, we know well that *every* number k can be broken down to $k = p^{\alpha}m$ where $p \nmid m$ for appropriate p. We know that $k = p_1^{\alpha_1} \cdot \ \cdots \ \cdot p_n^{\alpha_n}$ by the fundamental theorem of arithmetics, and that $p_i \nmid p_j$ for $i \neq j$, so choose some G such that $|G| = k$, and lets say that

$$k = p_1^{\alpha_1} \cdot \ \cdots \ \cdot p_n^{\alpha_n}$$

and so $p_i\big||G|$. We can thus re=write $k = p_i^{\alpha_i}m$ where $p_i \nmid m$. We know have the condition to apply Sylow's theorem.

I also called Sylow's theorem a partial converse of Lagrange's Theorem (3.3.1), mainly that if $H \leqslant G$, $|H|\big||G|$. Sylow's theorem states that for every prime factor p of the order of a finite group G (i.e. $p\big||G|$), then there exists a Sylow p-subgroup of order p^n, and furthermore, we know how many there are, and that their internal structure is all isomorphic (which makes Sylow's theorem give us more information than Cauchy Theorem (3.3.5), which would just say there exists a group of order p).

Corollary 4.4.15: Consequences of Sylow's theorem

Let P be in a Sylow p-subgroup of G. Then the following are equivalent

1. P is the unique Sylow p-subgroup of G, i.e., $n_p = 1$
2. P is normal in G
3. P is characteristic in G
4. All subgroups generated by elements of p-power order are $p$groups , i.e., if X is any subset o G such that $|x|$ is a power of p for all $x \in X$, then $\langle X \rangle$ is a p-group.

Proof **:**
p.142. Some implications are easy, other's a little more work is needed

18: How I approach Sylow Theorem Question

1. Figure out which prime you want to work with (if it's not been limited for you yet). Change $|G| = n$ to $|G| = p^\alpha m$ $(p \nmid m)$
2. Figure out if there are any properties of the group you're about to work with which might be useful to know ahead of time (ex. D_2n always has a cyclic normal group $\langle r \rangle$)
3. Apply Sylow's theorem and get all the numbers you need. Do $n_p \equiv 1 \bmod p$ to get the possibilities, $p|m$ where m is the index to limit it, then either $n_p = |G : N_G(P)|$ to get the final answer or brute for some (if the order is small enough)

Example 4.4.10

1. Show that a Sylow p-subgroup of D_n is cyclic and normal for every odd prime p (recall that $\langle r \rangle$ is cyclic)
2. Exhibit all Sylow 3-subgroups of A_4 and all Sylow 3-subgroups of S_4
3. Exhibit two distinct Sylow 2-subgroups of S_5 and an element of S_5 that conjugates one into the other

   Proof. This one involves recalling the subgroup of S_8 that represents D_8. □

4. Let $|G| = pqr$ where p, q, r are prime and $p < q < r$. Prove that G has a normal Sylow subgroup of either p, q, or r

   Proof. Apply Sylow's theorem to all of them. When you reach the part of trying to get the amount of Sylow $p/q/r$-subgroups, take the largest smallest non-1 number. Since the groups are of prime order, and are all *distinct*, we can do the trick of showing we have too many elements: Take the elements of prime order p, q, r. Since all the groups are prime, they are cyclic and so by theorem 2.3.8 you know the amount of them (say order p, then $p - 1$

elements generate it). Then if you add it all up, you'll get a number greater than pqr, a contradiction since $|G| = pqr$, meaning one of the $n_{p/q/r}$ has to be 1 in order to diminish the number of elements, and by corollary 4.4.15, we get that it must be normal. □

Also, since we now have Sylow's Theorem, we can prove Cauchy's theorem for non-abelian groups with Sylow:

Theorem 4.4.16: Cauchy's Theorem Proved

If G is a finite group and p is prime diving $|G|$, then G has an element of order p. See 3.3.5

***Proof* :**

Let G be a group such that $p\big||G|$. that means that $|G| = p^\alpha m$ where $p \nmid m$. So by Sylow's theorem, we have that $P \in \mathrm{Syl}_p(G)$, $|P| = p^\alpha$.

Now, take an element $x \in P$. Consider $\langle x \rangle \leqslant P$. Let $H = \langle x \rangle$. Since $|H|\big||P|$, we have that $|H|$ can be $1, p, p^2, ..., p^\alpha$. If $|H| = p^\alpha$, then that means that P is cyclic, and so we can create the subgroup $\langle x^{p^{\alpha-1}} \rangle$, and then by theorem 2.3.8 $|\langle x^{p^{\alpha-1}} \rangle| = \frac{p^\alpha}{(p^\alpha, p^\alpha - 1)} = p$, and we're done. If P is not cyclic, then $|H| < p^\alpha$. Since $|H|\big||P|$, then $|H| = p^\beta$ for some $\beta < \alpha$. We still have that H is cyclic, so we can once again take $K = \langle x^{\beta-1} \rangle$, and since

$$|\langle x^{\beta-1} \rangle| = \frac{p^\beta}{(p^\beta, p^{\beta-1})} = p$$

we have yet again found a subgroup of order p. Since the subgroup operator is transitive, we

$$K \leqslant H \leqslant G \Rightarrow K \leqslant G$$

meaning we found a group of order p in G, as we sought to show

4.4.5 Applications of Sylow's Theorem

Here are explicit orders Dummit went over. The goal of this is to get comfortable with some general cases of orders of groups.

Example 4.4.11: : Groups of Order pq, p, q prime $p < q$

Let $|G| = pq$. And consider $P \in \mathrm{Syl}_p(G)$ and $Q \in \mathrm{Syl}_q(G)$. we know that

$$n_p \cong 1 \bmod (p) \qquad n_q \cong 1 \bmod (q)$$

meaning and by the 2nd part of Sylow's theorem, $1 + kp|q$ and $1 + kq|p$. We also know that P and Q are cyclic by theorem 2.3.8. So let $P = \langle x \rangle$ and $Q = \langle y \rangle$.

Since $q > p$, then only $k = 1$ is satisfied for the equation $1 + kq|p$, so we know that $n_q = 1$, and so Q is the only Sylow p-subgroup, and is consequentially normal (corollary 4.4.15), thus $Q \trianglelefteq G$.

For n_p, since q is prime, the only possibilities are that $n_p = 1$ or $n_p = q$. For a moment, let's consider the case where $n_p = 1$ and $P \trianglelefteq G$. We do this because we'll show that if this scenario occurs, then $G \cong Z_{pq}$. TBD

Example 4.4.12: : Groups of Order 30
p.143

Example 4.4.13: : Groups of Order 12
p.144

Example 4.4.14: : Groups of Order p^2q, $p \neq q$
p.144

Example 4.4.15: : Groups of Order 60
p.145

Corollary 4.4.17: Simplicity of A_5

A_5 is simple

Simplicity of A_n

As you've seen, S_n has very little structure to it, being the group that could represent all groups. For example, we know that $Z(S_n) = 1$ for $n \geq 3$, and $|S_n| = n!$, showing how big and inter-related things are. However, We get to deduce one *big* thing with Sylow's theorem. If $n \geq 5$, then there will *always* exists a normal subgroup of S_n!

Exercise 4.4.3

1. Show that a group G is abelian if and only if every Sylow subgroup is normal

4.5 Actions on Groups

(Maybe move this under Sylow's Theorem?)

Perhaps introduce $\mathrm{Aut}_{\mathbf{Set}}(X) = \mathrm{Sym}(X)$ earlier? Like in the

We have been thinking of premutation representation in terms of S_n. We can perhaps add to this intuition of what S_n "represents" by recalling the definition of Automorphism (definition 1.4.9). If we think of bijections as the isomorphisms of the category of sets, then the set of permutations on a set A is then the automorphisms of A: $\mathrm{Sym}(A)$, so in terms of sets, $\mathrm{Sym}(A) \cong_{\mathbf{Set}} S_n$:

$$\varphi : G \to \mathrm{Sym}(A)$$

And as we've seen with the study of group-actions, we usually preform an group-action on itself: $\mathrm{Aut}(G)$

But you might ask yourself why would we want to take $\mathrm{Aut}(G)$? Well, if we are homomorphic/isomorphic to $\mathrm{Aut}(G)$, this will tell us a lot more about the information we get from thinking about the

group action, since it's not *just* that the group can be represented through permutations, but it is an even *stronger* case, since the group can be represented as permutations *which preserve groups structure.* For that reason, if the subgroup of S_n the premutation representation of G maps to is some automorphism, then we say that the group-action is a *structure preserving premutation.* What this will imply is that subgroups of G will be sent to subgroups of G (though not necessarily the same subgroups), elements of order n sent to elements of order n (again not necessarily the same ones), and other things isomorphisms let's us now. What this means we'll demonstrate after a concrete example.

Since $\mathrm{Aut}(G) \leqslant S_G$, then $\mathrm{Hom}(G, \mathrm{Aut}(G)) \leqslant \mathrm{Hom}(G, S_G)$, meaning that if we find a homomorphism from G to $\mathrm{Aut}(G)$, then studying such homomorphisms is equivalent to studying group actions (theorem 4.2.1), so for defining our group-action, we'll define the appropriate homomorphism.

We can actually sightly generalize $\mathrm{Aut}(G)$. We know that if $N \trianglelefteq G$, G acts on N by conjugation

$$\varphi : G \times N \to N \qquad n \mapsto gng^{-1}$$

However, if $H \leqslant G$, then it's not necessarily defined since H is not normal so ghg^{-1} might not be in H. However, this is fixed if we take $N_G(H) \leqslant G$, that is, all elements of G which are normal in G. Now, we can define the following homomorphism:

$$\psi : N_G(H) \to \mathrm{Aut}(H) \qquad g \mapsto (h \mapsto ghg^{-1}) \text{ (i.e. } \psi(g)(h) = ghg^{-1})$$

You can quickly check that

$$\begin{aligned}
&\Psi(ab) \\
\Leftrightarrow &\varphi(ab)(h) = h(ab)h^{-1} = hah^{-1}hbh^{-1} = (hah^{-1})(hbh^{-1}) = \varphi(a)(h)\varphi(b)(h) = (\varphi(a) \circ \varphi(b))(h) \\
\Leftrightarrow &\Psi(a)\Psi(b)
\end{aligned}$$

To show it's indeed a homomorphism. Note $\ker(\psi) = \{g \in G | \varphi(g) = id\} = \left\{g \in G \middle| ghg^{-1} = g\right\} = C_G(H)$, that is, the centralizers.

Since we're dealing with homomorphisms, recall that the study of homomorphisms and their kernels is equivalent to the study of normal groups, and so $N_G(H)/C_G(H)$ is a normal group, and by the first isomorphism theorem (3.1.12): $N_G(H)/C_G(H) \cong I \leqslant \mathrm{Aut}(H)$.

19: Foreshadowing

We will specify one subcase form this construction. If $H \trianglelefteq G$ and $K \leqslant G$, then we can define:

$$\varphi : K \to \mathrm{Aut}(H) \qquad k \mapsto (h \mapsto khk^{-1})$$

Since $K \leqslant N_G(H) = G$, this construction holds. The interesting thing about this construction is that we can ignore now the group G and just focus on K and H. Why is this interseting? It will be the *key* idea when trying to define *semi-direct* products. I'll point back to this once we get there (section 5).

Note that it's not isomorphic to $\mathrm{Aut}(H)$, but it's isomorphic into its image ($\Psi(N_G(H)) = I \leqslant \mathrm{Aut}(H)$) which *strictly* has the Automorphism of conjugation! We actually give a special name to this sub-group of $\mathrm{Aut}(G)$

Definition 4.5.1: Inner Automorphism

Let $\mathrm{Aut}(G)$ be the set of automorphisms. Let $\mathrm{Inn}(G) \leqslant \mathrm{Aut}(G)$ denote all automorphisms defined by conjugation, that is,

$$\varphi_g \in \mathrm{Inn}(G) \qquad \varphi_g(h) = ghg^{-1}$$

Now, the fact that this group-action permutation representation fits within $\mathrm{Aut}(H)$ tells us that if S is a subgroup, gSg^{-1} is a subgroup (actually isomorphic to S). Furthermore, if $h^n = e$, then $((ghg^{-1})^n = e)$. More rigorously:

Corollary 4.5.2: Conjugacy and subgroups

If K is any subgroup of the group G and $g \in G$, then $K \cong gKg^{-1}$

Proof **:**
For $K \cong gKg^{-1}$, Take $G = H$, from the previous theorem so that $N_H(H) = H$. Thus we have

$$H \to \mathrm{Aut}(H)$$

as a homomorphism, meaning that $\varphi(H)$ is a subgroup of $\mathrm{Aut}(G)$. TBD

We'll get a little more comfortable with $\mathrm{Aut}(G)$ for a moment

Example 4.5.1
here

1. Let $G = \mathbb{Z}/p\mathbb{Z}$, p prime. G is cyclic, then we know that if $\psi \in \mathrm{Aut}(G)$ then

   $$\psi(k) = \psi(1) + \cdots + \psi(1)$$

   so ψ can take 1 to any (other) generator. So

   $$\mathrm{Aut}(G) = \mathrm{Aut}(\mathbb{Z}/p\mathbb{Z}) = (\mathbb{Z}/p\mathbb{Z})^\times$$

   and if n is not prime, then

   $$\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z}) = (\mathbb{Z}/n\mathbb{Z})^\times$$

   where the elements of the multiplicative group is where we took all the elements relatively prime to n.

2. Let V be a vector space over $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. Then we know that since V is finite dimensional then $V \cong \mathbb{F}_p^n$, that is v can be represented as n-tuples of elements form $\mathbb{F}_p$. Then

   $\mathrm{Aut}(V)$ are the maps $V \to V$ that preserve addition, which is $GL(n, \mathbb{F}_p)$ acting by matrix multiplication.

This example was a bit hard to come up with, so is there an easier way to look at the structure of $\mathrm{Aut}(G)$?. It will turn out that looking at $\mathrm{Aut}(G)$ would be like looking at the conjugacy classes!

This is becasue Inner Automorphisms group (the group which you can think of as representing the conjugacy classes) quotients the Automorphism and produce a quotient group, that is, $\mathrm{Inn}(G)$ is a *normal* subgroup!

Definition 4.5.3: Outer Automorphism

Given G, $\mathrm{Aut}(G)$, $\mathrm{Inn}(G)$, then

$$\mathrm{Aut}(G)/\mathrm{Inn}(G)$$

is defined. We call this the ***Outer-Automorphism***, and is denoted $\mathrm{Out}(G)$.

Proof **:**
We'll prove that $\mathrm{Inn}(G)$ is indeed a normal subgroup. simple proof, just gotta get used to quotient by function spaces

Note also that this means that the structure of Out is the antithesis of Inn by construction.

To show another link to conjugacy classes:

Proposition 4.5.4: If G is Abelian, Then $\mathrm{Inn}(G)$ Is Trivial

Let G be an abelian group. Then every Inner-automorphism is the trivial map, and thus, $\mathrm{Out}(G) \cong \mathrm{Aut}(G)$

Proof **:**
If G is abelian, then $\varphi_g(x) = gxg^{-1} = x$, meaning that every conjugation map is the identity, so we only have $Id \in \mathrm{Inn}(G)$, meaning $\mathrm{Aut}(G)/\mathrm{Inn}(G) \cong \mathrm{Aut}(G)$.

Thus, the more abelian a group is, the the larger $\mathrm{Out}(G)$ becomes.

(incorporate this somehow?) Similarly, there is a surjective homomorphism $f : G \to \mathrm{Inn}(G)$, $g \mapsto f_g$, where f_g is conjugation by g. Then the kernel of this map is $Z(G)$. Thus, we have

$$1 \longrightarrow Z(G) \longrightarrow G \longrightarrow \mathrm{Inn}(G) \longrightarrow 1$$

In this way, $\mathrm{Inn}(G)$ is the *dual* of $Z(G)$.

4.5.1 A Word on Characteristic Groups

(With $\mathrm{Inn}(G)$, we can define normal subgroups differently: H is normal if and only if for any $\sigma \in \mathrm{Inn}(G)$, $\sigma(H) = H$. In other words, normal subgroups are the ones that map to themselves. If H is not normal, then $\sigma_x : H \to \sigma_x(H)$ and $\sigma_y : H \to \sigma_y(H)$ can have different domains. However, if H is normal, then for any i, $\sigma_i : H \to H$ is well-defined, letting us "restrict" $\mathrm{Inn}(G)$ to $\mathrm{Inn}(H)$. Is this true in general, if $H \trianglelefteq G$ then we can restrict $\mathrm{Aut}(G)$ to $\mathrm{Aut}(H)$? The answer is no: (example). We will thus give a name to groups that have this added property:)

Definition 4.5.5: Characteristic In G

A subgroup H of a group G is called *characteristic* in G, denoted $H \operatorname{char} G$ or $H \blacktriangleleft G$, if every automorphism of G maps H to itself, i.e., $\sigma(H) = H$ for all $\sigma \in \operatorname{Aut}(G)$.

The notation $H \blacktriangleleft G$ is not as common as $H \operatorname{char} G$. However, I preffer it as a reminder that it's a strengthening of a normal subgroup which has notation $H \trianglelefteq G$ (where $\trianglelefteq$ is now ubiquitously accepted notation).

Example 4.5.2

1. If PQ is a subgroup of G where P and Q are normal, then P, Q are characteristic.
2. Show that $Z(G) \blacktriangleleft G$
3. Let $G' = \left\{xyx^{-1}y^{-1} \mid x, y \in G\right\}$. G' is called the *commutator subgroup*, and captures some idea of how abelean our group is. It will be very important in section 5.2 and 6.1.1. Show that $G' \blacktriangleleft G$

Proposition 4.5.6: Characteristic Subgroups and transitivity

1. characteristic subgroups are normal
2. If H is the unique subgroup of G of a given order, then H is characteristic in G
3. If H char K, and $K \trianglelefteq G$, then $H \trianglelefteq G$

When we worked with compositione series (definition 3.7.3), we saw that normality is *not* transitive. This is very annoying, since transitivity is a very powerful axiom (for example, with equivalence relations). Thus, this generalization can come in handy?

Proof **:**

1. Since $\varphi(H) = H$ for every $\varphi \in \operatorname{Aut}(G)$, then it's also true for every $\varphi \in \operatorname{Inn}(G)$, so
$$\varphi(H) = gHg^{-1} = H$$
And since this was true for every g, H is normal.
2. If H is a unique subgroup, then $gHg^{-1} = H$, since we know conjugacy maps are isomorphic, which means order-preserving, which means it must be the same group, However, that implies if that for $\varphi \in \operatorname{Inn}(G)$, $\varphi(H) = H$. Similarly. Similarly for any automorphism, since H is a subgroup of itself, and automorphisms preserve subgroup structure, then $\varphi(H) = H$, thus $H \blacktriangleleft G$
3. Since H char K, then for all the $\varphi \in \operatorname{Aut}(K)$, $\varphi(K) = K$. Since $K \trianglelefteq$, we know that for $\psi \in \operatorname{Inn}(G)$, $\varphi(K) = gKg^{-1} = K$ (not this is not necessarily true for all automorphisms, meaning K is not necessarily characteristic of G). Since H char K, we can restrict the domain $\psi|_H$, but keep the conjugation *from* G, so we have
$$\varphi|_H(H) = gHg^{-1} = H$$

thus, $H \trianglelefteq G$

Lemma 4.5.7: Showing Characteristic With Generators

If $H \leqslant G$, and $\varphi : G \to \mathrm{Aut}(G)$, then we only need to check the generators of G to check that H is normal (here)

Proof. finish later □

Here is something I found by doing some tinkering with characteristic subgroups. If we have $H \leqslant G$, Then is $\mathrm{Aut}(H) \leqslant \mathrm{Aut}(G)$? Take a moment to think about it

Now that you've taken a moment to think about it, you might have seen that there the question isn't defined. If $\sigma \in \mathrm{Aut}(G)$, then then the domain and co-domain is $\sigma : G \to G$. However, in $\mathrm{Aut}(H)$. we'd need the domain and co-domain to be H. If we tried to restrict every σ to σ_H, then we'd get $\sigma|_H : H \to \sigma(H)$, with no guarantee that $\sigma(H) = H$. However, if $H \operatorname{char} G$, we *do* get this fact for every $\sigma \in \mathrm{Aut}(G)$. Therefore, we can meaningfully write $H \operatorname{char} G$ then $\mathrm{Aut}(H)$" $\leqslant''$ $\mathrm{Aut}(G)$, where $\leqslant$ is written in quotes since it's not actually a subgroup. It might be better to write $\mathrm{Aut}(G)|_H = \mathrm{Aut}(G)$. Thus, characteristic subgroups have a use in that we can define "sub-actions" (maybe?). (TBD) Really interestingly!!, this type of thinking abou defining some notion of subaction becomes important in Galois theory, in particular in the 3rd fundamenal galois theorem (theorem 18.1.16)

Exercise 4.5.1

1. G be a group, and let Let $H = \{g \in G \mid \sigma(g) = g,\ \sigma \in \mathrm{Inn}(G)\}$. Show that $H = Z(G)$. In fancer vocabulary, we say that $Z(G)$ is the fixed point group of $\mathrm{Inn}(G)$ [2]
2. Let $G_\sigma = \{g \in G \mid \sigma(g) = g\}$, that is G_σ is the set of all elements of G that are fixed by σ. Show that $G_s \trianglelefteq G$.
3. Show that if $\mathrm{Aut}(G)$ is abelian, then $G_\sigma \blacktriangleleft G$.
4. (Double check this) in exercise ref:HERE (in section 2.1 I think), we showed that if G is abelian, then $T(G) \leqslant G$. Show that if $\mathrm{Aut}(G)$ is abelian, then $T(G)$ is a subgroup.
5. (Go to that website characteristing $\mathrm{Aut}(G)$ when G is finitely generated (or just finite?) and abelian which says because the product and co-product coincide, it's actually very simple!)

4.6 Actions in general

(Talk about if $c \in C$ is an object in a cateogry, $\mathrm{Aut}_C(c)$ is a group! Thus, group appear very naturally *everywhere* because of this)

(perhaps talk about how this might be where the intuition that all groups are in some fundamental sense the representation of "symmetries", because an automorphism maps an object identically back to itself).

[2] We can strengthen the notion of $Z(G)$ by defining a group $L(G)$ called the *absolute center* of G by letting $L(G)$ be the fixed subgroup of $\mathrm{Aut}(G)$. Show that $L(G) \blacktriangleleft Z(G)$

5

Composing Groups: Direct and Semi-Direct Products

(Current intuition: We've learned that any finite group is an extension of simple groups. How now do we concretely know the structure of G given H and $G/H = K$? This can in general be a very hard question, however there are cases in which this question is much more managable: the *direct produc* and *semi-direct product*. As it turns out, if G is an extention of H by K, then the relations of G can purely be described by H and K by something called the *wreath product* which a sort of action on H given by K including the semi-direct product; this is the *Krasner–Kaloujnine universal embedding theorem*. This theorem is more difficult and of little use to the general project of introducing algebra in this book, but the spirit of undersatnding the structure of G given H and K is going to be a theme that shall crop up a lot, and deeply explored in part VII)

I got a couple of intuition on direct products

1. Adding group structure to products. There is a notion of products with sets: X, Y can give $X \times Y$. It is very easy to show that if X and Y have group structures $((X, *_X), (Y, *_Y))$, then the product of these form a very natural group. This product also doesn't break under infinitely many products too [1]

2. Sort of the dual to the previous intuition: instead of building up groups, we're going to *decompsoe* groups into direct product of groups. We've learnt how to make subgroups out of groups by looking for elements with special properties (ex. commutativity) or generating subgroups with a subset of elements. We've found smaller groups by learning about how to properly define the idea of quotienting two groups with the help of homomorphisms. Now, we'll sort of combine the two. We are going to try and decompoes groups (ex. G) to into smaller

[1] If you've seen topology, you know that this is not always true in every Category. In the **Top** Category, you will/have seen that the product topology on infinitely many spaces need some special conditions

groups whose direct product would be G (ex. $G = H \times K$), or when we *almost* have a direct product ($G = H \rtimes K$)

3. This intuition overlaps with the last point. If we have $H, K \trianglelefteq G$, so that we don't have to think about $H \cap K$, then we can form $HK = G$ (or if $H \cap K = \{e\}$, $H \times K \cong G$). We can generalize this, and have $H \trianglelefteq G$ and $K \leqslant G$, an we will from here define a new product $H \rtimes K = G$ to be able to construct some differnet groups. Remebmer how I said a lot of stuff in math defines how you're *almost* abelian? Here is another example of the idea of being commutative as being the point from which we are trying to generalize. This one will heavily rely on intuition presented about the automorphisms! TBD

5.1 Classical Definitions And Properties

You might recall example 1.1.2[4], we showed we can create a group-structure given two groups. We revisit this definition here in generality:

Definition 5.1.1: Direct Product

The ***direct product*** of $G_1 \times G_2 \times \cdots$ of groups $G_1, ..., G_n, ...$ with the operations $*_1, ..., *_n, ...$ respectively is the set of sequences $(g_1, ..., g_n, ...)$ (or n-tuples if the number of groups is finite) where $g_9 \in G_i$ with the operation defined componentwise:

$$(g_1, ..., g_n, ...) * (h_1, ..., h_n, ...) = (g_1 *_1 h_1, ..., g_n *_n h_n, ...)$$

Sometimes, the direct product is called the *external direct product* to differentiate it from the *internal direct product* (see definition 3.4.1)

Example 5.1.1

1. Think of $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ with component-wise addition: $(\mathbb{R} \times \mathbb{R}, (+, +))$ or $(\mathbb{R}^2, (+, +))$.
2. The groups can also be arbitrary. Let $G_1 = \mathbb{Z}$, $G_2 = S_3$ and $G_3 = GL_2(\mathbb{R})$. Then $G_1 \times G_2 \times G_3$ with the operation defined component wise is indeed a group

For the direct product, there are a few general properties you can prove as an exercise:

Proposition 5.1.2: properties of Direct Products

Let every G_i be a group.

1. $|\prod_i G_i| = \prod_i |G_i|$
2. For each fixed i, the set of elements of G which have the identity of G_j in the jth position for all $j \neq i$ and arbitrary elements of G_i in position i is a subgrou pof G isomorphic to G_i:
$$G_i \cong (1_{g_1}, ..., g_i, 1_{g_{i+1}}, ...)$$
3. For each fixed i, define $\pi_i : G \to G_i$ by
$$\pi_i((g_1, ..., g_n)) = g_i$$
then π_i is a surjective homomorphsim with
$$\begin{aligned}\ker(\pi_i) &= \{(g_1, ..., g_{i-1}, 1, g_{i+1}, ..., g_n) | g_j \in G_j \text{ for all } j \neq i\} \\ &\cong G_1 \times \cdots \times G_{i-1} \times G_{i+1} \times G_n\end{aligned}$$
4. Under the identificatiosn in part (1), if $x \in G_i$ and $y \in G_j$ for some $i \neq j$, then $xy = yx$

Proof **:**
exercise

(Comment on that last part of the proposition with $xy = yx$. Since $(X, 0)$ and $(0, Y)$ are "independent" from each other (i.e., they both kind of stay in their own lanes, when multiplying we just do component wise), it comes at the "cost", or a necessary consequence of it, is that we must treat those element having an *abelian* relation. This is another example of how something being abelian makes thing very powerful – if we want those two to be "independent", we impose an abeilan condition, we impose an abeilan condition)

Exercise 5.1.1

1. Does there exist a group such that $G \times G \cong G$? hint: [2]

   Let G be a finitely generated abelian group. Show that every subgroup is also a finitely generated abelian group.

In definition 5.1.1, the components of the direct product need not be the same. If they are the same, we can introduce the following notation:

$$G^4 = G \times G \times G \times G$$

which repesents the direct product of the group G 4 times.

We can equivalently write it this way

is equivalent to $\left\{f : \{0, 1, 2, 3\} \to G \middle| \ f \text{ is a function}\right\}$ where f is any function! (Poof of this here).

[2] Think about $G = \prod_{n=1}^{\infty} \mathbb{Z}$

Example 5.1.2: Example
here

Show how this is useful in generalizing to (for example):

$$(\mathbb{Z}/2\mathbb{Z})^{\mathbb{R}} = \{f : \mathbb{R} \to \mathbb{Z}/2\mathbb{Z}\}$$

more generally, we can write A^B!

Maybe as a bonus (so stared section) show the link between A^B and the universal property of the product.

Example 5.1.3: Exercies

1. If $H \leqslant G_1 \times G_2$, is it true that there always exists subgroups $H_1 \leqslant G_1$ and $H_2 \leqslant G_2$ such that
$$H \cong H_1 \times H_2$$
if yes prove it, if not give a counter-example.

5.1.1 Fundamental theorem of Finitely Generated Abelian Groups

This theorem is stated, but not proved till till much later (chapter 14, theorem 14.1.5). The main idea that is trying to be presented here is the power of of being abelian: If we have an abelian, it is *such* a nice group that we can classify them all with isormophism to $\mathbb{Z}$ and $\mathbb{Z}_{p^\alpha}$ for appropriate power-prime. It will also help in our indeavour to classify groups. With this representation, it becomes much much easier. Here are two terms we introduce now to present the theorem:

Definition 5.1.3: Finitely Generated Group

A group is finitely generated if $\exists A \subseteq G$ such that $\langle A \rangle = G$.

Definition 5.1.4: Free Abelian Group

$\mathbb{Z}^r = \mathbb{Z} \times \cdots \times \mathbb{Z}$. $\mathbb{Z}^r$ is called the *free abelian group of rank r*.

Theorem 5.1.5: Fundamental Theorem of Finitely Generated Abelean Groups

Let G be a finitely generated group G. Then:

$$G \cong \mathbb{Z}^r \times \mathbb{Z}_{n_1} \times \cdots \mathbb{Z}_{n_s}$$

for some integers $r, n_1, ..., n_s$ with the following conditions:

1. $r \geq 0$ and $n_j \geq 2$ for all j, and
2. $n_{i+1} | n_i$ for $1 \leq i \leq s-1$

Furthermore, the expression is unique (p.173 for clarification).

For those interested in a proof of this theorem for finite group see section ref:HERE [3]

> ***Proof*** **:**
> We prove this in theorem 14.1.5. We will give a purely group-theoretical proof of this fact in ref:HERE.

What we'll do for the rest of this section is give the vocabulary to talk about abelian groups.

Notice also that it's for *finitely generated abelian groups*. What about infinitely generated abelian groups? We can simplify for a moment and ask about countably generated abelian groups. We have solved this for p-groups, but we have yet to solve it in the general case:

`https://math.stackexchange.com/questions/3927834/classification-of-countably-infinite-abelian-groups`

Definition 5.1.6: free rank and invariant factors

r is called the *free rank* or *betti number*, while $n_1, ..., n_s$ are called the *invariant factors*. The whole expression is called the *invariant factor decomposition of* G.

will finish later

Exercise 5.1.2

1. Let $A \subseteq B$ two bwo finitely generated free abelian groups each of rank n. Show that B/A has rank 0.
2. Let $G = (\mathbb{Z}/n\mathbb{Z})^\times$. Decompose G into a product of cyclic groups with additive notation.

5.2 Recognizing Direct Products

Let's say we have a (not necessarily abelian) group G. Can G be decomposed into the direct product of two smaller groups? This question is what we'll be exploring in this section.

The 4th identity of proposition 5.1.2 gives us a very valuable insight on how to think of direct products. If we have two groups G_1 and G_2, we showed that creating a set of 2-tulpes with component-wise operation creates a new group. With the fact that components commute ($xy = yx$, if $x \in G_1$, $y \in G_2$), we can also define the direct product in terms of the following:

$$G_1 \times G_2 = \langle x, y | xy = yx, x \in G_1, y \in G_2 \rangle$$

where the operation is done when everything is "moved to the appropriate place"

Example 5.2.1: presentation example
Let $G_1 = S_3$ and $G_2 = \mathbb{Z}_5$. Take $x = (1\ 2)4$ and $y = 4(2\ 3)2$. Then

$$xy = (1\ 2)44(2\ 3)2 = (1\ 2)(2\ 3)442 = (1\ 2\ 3)2$$

[3] For those who want to see how the difficulties of generalizing this theorem to infinitely (particularly countably) generated abelian groups and want to see a particular generalization, google Prüfer Theorems

You might ask yourself what's the point of this 2nd perspective? The idea is that we can now ask ourselves when can a group be decomposed into smaller groups. In other words, if we had a group G, can we split it into two parts G_1, G_2 such that $G \cong G_1 \times G_2$?

Thus, we will explore exactly how far a two subgroups (or more generally subsets) of a group are from being commutative, and then use that information to find out when we can decompose groups.

We start with defining a way to capture *how commutative* a groups/subgroups/elements are:

Definition 5.2.1: Commutator and Commutator Group

Let G be a group, let $x, y \in G$, and let A, B be nonemtpy subset of G

1. Define $[x, y] = x^{-1}y^{-1}xy$. $[x, y]$ is called the *commutator* of x and y
2. Define $[A, B] = \langle [a, b] \big| a \in A,\ b \in B \rangle$, the group geenrated by the commutators of elements from A and B
3. Define $G' = \langle [x, y] | x, y \in G \rangle$ to be the subgroup of G generated by commutators of all element sfrom G. G' is called the *commutator subgroup* of G

Notice that if x and y commute, then $[x, y] = 1$. If all the element of G commute, then $G' = \{1\}$. Thus, the commutator captures an idea of how much the elements of G commute

Proposition 5.2.2: Commutator Subgroup Properties

Let G be a group, $x, y \in G$, and $H \leqslant G$. Then

1. $xy = yx[x, y]$
2. $H \trianglelefteq G$ if and only if $[H, G] \leqslant H$
3. for any automorphism σ of G, $\sigma([x, y]) = [\sigma(x), \sigma(y)]$, that is, $G' \operatorname{char} G$. As a consequence, G/G' is abelian
4. G/G' is the largest abelian quotient of G. That is, if $H \trianglelefteq G$ and G/H is abelian, then $G' \leqslant H$. Conversely, if $G' \leqslant H$, then $H \trianglelefteq G$ and G/H is abelian
5. (Universal Abelian Quotient) if $\varphi : G \to A$ is any homomorphism of G into an abelian group A, then φ factors through G', i.e., $G' \leqslant \ker(\varphi)$ and the following diagram commutes:

$$\begin{tikzcd} G \arrow[r] \arrow[dr, "\varphi"'] & G/G' \arrow[d] \\ & A \end{tikzcd}$$

(A word on how:) These proofs are a good review of many previous concepts covered in this textbook, and so I highly recommend you try to do all of these yourself before doing the proof – you'll gain a much stronger intuition behind commutator subgroups.

***Proof* :**

1. $xy = yxx^{-1}y^{-1}xy = xy$
2. Let $H \trianglelefteq G$. Then $gH = Hg$, meaning every element of H, $gh = h'g$ for some $h, h' \in H$. Thus
$$[H, G] = \langle h^{-1}g^{-1}hg | h \in H, g \in G\rangle = \langle hh' | h, h' \in H\rangle$$
thus, $[H, G]$ is generated by elements of H, meaning $[H, G] \leqslant H$. Conversly, if $[H, G] \leqslant G$, then for every $h^{-1}g^{-1}hg \in [H, G]$, there is an h' such that
$$h^{-1}g^{-1}hg = h' \Rightarrow g^{-1}hg = hh'$$
and since this is true for any $h \in H$, $g^{-1}Hg = H$, or if we replace g with g^{-1}, we get the more conventional $gHg^{-1} = H$, meaning $H \trianglelefteq G$.
3. Let $[x, y] \in G'$. If $\sigma \in \mathrm{Aut}(G)$, then
$$\sigma([x, y]) = \sigma(x^{-1}y^{-1}xy) = \sigma(x)^{-1}\sigma(y)^{-1}\sigma(x)\sigma(y) = [\sigma(x), \sigma(y)]$$
and since $\sigma(x), \sigma(y) \in G$, that means' σ maps commutators to commutators. Since σ is bijective (i.e., has a two-sided inverse):
$$\sigma(G') = G$$
for any σ, and so $G' \operatorname{char} G$, and so $G' \trianglelefteq G$ (proposition 4.5.6).
Since G/G' is well-defined, notice that for $[x], [y] \in G/G'$,
$$[xy] = [xyy^{-1}x^{-1}yx] = [yx]$$
and so, G/G' is abelian.
4. Let $H \trianglelefteq G$ such that G/H is abelian. Then for any $[x], [y] \in G/H$, $[xy] = [yx]$, implying
$$[xyx^{-1}y^{-1}] = [1] \Rightarrow xyx^{-1}y^{-1} \in H$$
showing that $G' \leqslant H$. Conversly, if $G' \leqslant H$, then for any $g \in G$ and any $h \in H$
$$ghg^{-1}h^{-1} \in H \Rightarrow ghg^{-1}h^{-1} = h'$$
for some $h' \in H$, but then $ghg^{-1} = h'h$. Since h, h', g were all arbitrary, that means that $gHg^{-1} = H$ for any g, and so H is normal, and so G/H is well-defined. Furthermore, for any $[x], [y] \in G/H$,
$$[xy] = [xyy^{-1}x^{-1}yx] = [yx]$$
since $y^{-1}x^{-1}yx \in H$. Thus, G/H is also abelian.
Another way of proving this is by taking advantage of isomorphism theorems (take a moment to think about it if you didn't see it this way as good practice!).
If G/G' is abelian, then every subgroup is normal, and since $G' \leqslant H$, $H/G' \trianglelefteq G/G'$ is well-defined. By the 4th isomorphism theorem, $H \trianglelefteq G$, and so G/H is well-defined. Then, by the 3rd isomorphism theorem:
$$G/H \cong (G/G')(H/G')$$
and since G/G' is abelian, so is it's quotient, and so G/H is abelian.

5. This is another way of writing (4), but in terms of homomorphisms.

   Let $\varphi : G \to A$ be a homomorphism from G to any abelian group G, and let $\ker(\varphi)$ be the kernel. We need to find a map from G to G/G' and a map G/G' to A such that the diagram commutes, that is, if π is the map from G ot G/G' and ψ is the map from G/G' to A, then $\varphi = \psi \circ \pi$.

   Since G is abelian, $G/\ker(\varphi)$ is abelian. Thus, by part (4), $G' \leqslant \ker(\varphi)$. Since G' is also a normal subgroup of G, G/G' is well-defined, the natrual projection map exists and is well-defined:

   $$\pi : G \to G/G' \qquad \pi(g) = gG'$$

   Now, for the map from G/G' to A, we need to make sure that it splits with φ, meaning it is dependent on where φ sends it's elements, and that it is independent of the representatives of G/G', that is, if

   $$[g] = [gx^{-1}y^{-1}xy] \Rightarrow \psi(g) = \psi(gx^{-1}y^{-1}xy)$$

   In fact, φ with the co-domain of G/G' will do just that. Let $\psi = \overline{\varphi}$ represent the function:

   $$\overline{\varphi} : G/G' \to A \qquad \overline{\varphi}([g]) = \varphi(g)$$

   this function is well-defined, since if $[g] = [gx^{-1}y^{-1}xy]$ for any $x, y \in G$

   $$\begin{aligned}
   \overline{\varphi}(gx^{-1}y^{-1}xy) &= \varphi(gx^{-1}y^{-1}xy) && \text{by definition} \\
   &= \varphi(g)\varphi(x^{-1})\varphi(y^{-1})\varphi(x)\varphi(y) && \varphi \text{ is a homomorphism} \\
   &= \varphi(g)\varphi(x)\varphi(x^{-1})\varphi(y)\varphi(y^{-1}) && A \text{ is abelian} \\
   &= \varphi(g)\varphi(xx^{-1})\varphi(yy^{-1}) && \varphi \text{ is a homomorphism} \\
   &= \varphi(g)\varphi(1)\varphi(1) \\
   &= \varphi(g) && \varphi(1) = 1 \\
   &= \overline{\varphi}([g]) && \text{by definition}
   \end{aligned}$$

   Thus, this function is indeed well-defined! It came down to φ already being defined for us, and A being abelian. One should also check that $\overline{\varphi}$ is indeed a homomorphism (exercise).

   Finally, we need to show that $\varphi = \overline{\varphi} \circ \pi$. If $g \in g$. Then

   $$(\overline{\varphi} \circ \pi)(g) = \overline{\varphi}(\pi(g)) = \overline{\varphi}([g]) = \varphi(g)$$

   and since this is true for all $g \in G$, then $\varphi = \overline{\varphi} \circ \pi$, meaning φ does indeed split! Thus, the diagram does indeed commute:

   $$\begin{tikzcd}
   G \arrow[r] \arrow[dr, "\varphi"'] & G/G' \arrow[d] \\
   & A
   \end{tikzcd}$$

(develop universal property and commutative groups)

Example 5.2.2

1. If G is abelian, then for any $x, y \in G$, $[x, y] = xyx^{-1}y^{-1} = xx^{-1}yy^{-1} = 1$, and so $G' = 1$, and $G/G' \cong G$.
2. Example with D_8
3. Example with D_n
4. Example with S_5
5. Example with F_2

With this, we can answer now give criterion to recognize direct products. First, a lemma:

Lemma 5.2.3: Representation's of HK

Let H and K be subgrousp of a group G. The number of distinct ways of writing each element of the set HK in the form hk, for some $h \in H$ and $k \in K$ is $|H \cap K|$. In particualr, if $H \cap K = 1$, then each element of HK can be written uniquely as a product of hk, for some $h \in H$ and $k \in K$.

Proof **:**
see proposition 3.4.2

Theorem 5.2.4: Recognition Theorem For Direct Products

Let G. If H, K are subgroups such that

1. H and K are normal in G, and
2. $H \cap K = \{1\}$

Then

$$HK \cong H \times K$$

Conersly, if $HK \cong H \times K$, then $H \cap K = \{1\}$ and both H and K are normal subgroups of G.

If $HK = G$, then $G \cong H \times K$. Similarly, If G is isomorphic to the diret product of M and N (where M and N need not be subgroups of G) if $H \cong M$, $K \cong N$, and $G = HK$.

The propety that $H \cap K = 1$ has a name:

Definition 5.2.5: Complement

Let H be a subgroup of G. A subgroup K of G is caleld a *complement* for H in G if $G = hK$ and $H \cap K = 1$.

Thus, we can rephrase the previous theroem as a group is a derict product if there exists a normal complement for some proper normal subgroup of G.

***Proof* :**
We'll first prove a lemma that we'll use for both direction: If HK is a group and $|H \cap K| = 1$, then $hk = kh$.

Proof. Since $H \trianglelefteq G$, $khk^{-1} \in H$, and so $(khk^{-1})h^{-1} \in H$. Similarly, $k(hk^{-1}h^{-1}) \in H$. Since $H \cap K = 1$, we have

$$khk^{-1}h^{-1} = 1 \Rightarrow kh = hk$$

showing that every element of H commutes with every element of K. Note that does does not show that HK is abelian. For example, we did not show that for any $k, k' \in HK$, $kk' = k'k$. □

If $HK \cong H \times K$, we know that $|HK| = |H \times K| = |H||K|$, implying that $|H \cap K| = 1$, fulfilling the first requirement (ONLY IN THE FINITE CASE!!). Next, since HK is indeed a group (being isomorphis to $H \times K$), then by corolalry 3.4.4, at least one of H or K are normal. We need to show that both H and K are normal. We know by proposition 5.1.2 that $(H, 1)$ and $(1, K)$ are normal. We can't just push H or K through to $(H, 1)$ or $(1, K)$ since we don't have control over what isomorphism we have, and we can't force H to map to $(H, 1)$ [a]. Since $HK \cong H \times K$ their pre-image are also normal groups. Let's say φ is the isomorphism. Then since both of these preimage are isomorphic to H and K repsectively, then we can create an isomorphism $\psi : HK \to HK$ where $\psi(hk) = \varphi^{-1}(h, k)$. Since φ^{-1} is an isomorphism, this map is well-defined and bijective. THis map is also a homomorphism since:

$$\begin{aligned}
\psi(h_1k_1h_2k_2) &= \psi(h_1h_2k_1k_2) && \text{lemma at the beginning} \\
&= \varphi^{-1}((h_1h_2, k_1k_2)) \\
&= \varphi^{-1}((h_1, k_1) \cdot (h_2, k_2)) \\
&= \varphi^{-1}(h_1, k_1)\varphi^{-1}(h_2, k_2) \\
&= \psi(h_1k_1)\psi(h_2'k_2)
\end{aligned}$$

Thus, H and K can be asociated with $(H, 1)$ and $(1, K)$, and must also be normal – completing this direciton.

For the other way, we need to find an isomorphism from HK to $H \times K$. We'll first establish some properties of HK to show such a function exists.

Since H and K are normal, then *a fortiori*, we met the condition for HK to be a subgroup by proposition 3.4.3, and so by the lemma done earlier, every element of H commutes with elements of K.

Now, by lemma 5.2.3, the elements of HK can be uniqueley written as hk, and so we can define the function

$$\varphi : HK \to H \times K \qquad hk \mapsto (h, k)$$

This function is a homomorphism, since $h_1k_1h_2k_2 = h_1h_2k_1k_2$. Since the right hand side is of the form $h'k'$ for some $h' \in H$, $k' \in K$, this is the unique way of representing it, and so:

$$\begin{aligned}
\varphi(h_1k_1h_2k_2) &= \varphi(h_1h_2k_1k_2) \\
&= (h_1h_2, k_1k_2) \\
&= (h_1, k_1)(h_2k_2) \\
&= \varphi(h_1k_1)\varphi(h_1k_2)
\end{aligned}$$

From here, it is easy to see that φ is a bijection, since every element of $H \times K$ must uniquely be mapped to by an element hk in HK, and so φ is an isomrophism.

[a]Ex. Take $(\mathbb{Z}/6\mathbb{Z})(\mathbb{Z}/6\mathbb{Z}) \to \mathbb{Z}/6\mathbb{Z} \times \mathbb{Z}/6\mathbb{Z}$ and map the element order 2 and 3 into different places in the image

Another way of thinking about this theorems is by saying that the *interal direct product* (definition 3.4.1) and *external direct product* (definition 5.1.1) of two subgroups H and K of G are equivalent if the two conditions in theorem 5.2.4 match. If $H \cap K \neq \{1\}$ or one of them is not normal, then $HK \ncong H \times K$. In the next section, we'll explore a generalization of the direct product, the *semi-direct product*, and show that HK can *always* be written as $HK \cong H \rtimes K$, where $\rtimes$ represents the semi-direct product

Note that if we have an abelian group G, finitely generated or not, (theorem 5.1.5 can't be used), and we have $H, K \leqslant G$, (really $\trianglelefteq G$ since G is abelian), then it's always the case that $HK \cong H \times K$. This is another example of the niceties of being commutative.

5.2.1 A word on Perfect Groups

We saw that if a group G is abelian, then $G' = \{1\}$. What if we had the complete opposite: if $G = G'$? Such a group has a name:

Definition 5.2.6: Perfect Group

Let G be a group and G' be the commutator subgroup. If $G = G'$, then G is called a *prefect group*

Example 5.2.3: Perfect Groups

1. A_5 is an example of a perfect group.

I'm not sure yet how yet they are important. They seemed to be linked to solvable groups.

5.3 Semi-Direct Products

As we mentioned earlier, not all groups can be broken down into [non-trivial] direct products:

Example 5.3.1: $G \ncong H \times K$
Let $G = D_6 = \langle \sigma, r | \sigma^2 = r^3 = 1,\ \sigma r = r^{-1}\sigma \rangle$. By Lagrange, we know non-trivial proper subgroups of D_6 would have order 2 or 3. We can computaitonly determine that $\langle \sigma \rangle$ and $\langle r \rangle$ are the only subgroups of order 2 and 3 respectively. Let $\langle \sigma \rangle$ represent the subgroup of order 2, and $\langle r \rangle$ represent the subgroup of order 3. Since $|D_6| = 6$, then $\langle r \rangle \trianglelefteq D_6$ (corollary 3.3.4). Since one of the subgroups are normal, by proposition 3.4.3 $\langle r \rangle \langle \sigma \rangle$ is a group, and by proposition 3.4.2, the cardinality of the group is 6. Since $|D_6| = 6$, it must be that $\langle r \rangle \langle \sigma \rangle \cong D_6$. However, we saw under the definition of 3.1.10 that $\langle 2 \rangle \ntrianglelefteq D_6$, and so by the recognition theorem for direct products:

$$D_6 \ncong \langle \sigma \rangle \times \langle r \rangle$$

If there were an isomorphism between these two, then no matter what isomorphism we chose, we will encounter a problem from the fact that $rs = sr^{-1}$. let's say φ was some isomorphism from $\langle s\rangle\langle r\rangle$ to D_6. Then $\varphi(ab) = (s, r)$ for appropriate a, b where $a \in \{e, \sigma\}$ and $b = \{e, r, r^2\}$ Then:

$$(e, r^2) = (s, r) \cdot (s, r) = \varphi(abab) = \varphi(aab^{-1}b) = \varphi(e) = (e, e)$$

but $(e, r^2) \neq (e, e)$, and therefore this isomrophism *must* fail, no matter how it's defined.

Since, $\langle\sigma\rangle \cong \mathbb{Z}/2\mathbb{Z}$ and $\langle r\rangle \cong \mathbb{Z}/3\mathbb{Z}$, we can swtich over to that notation. Then we can say that

$$D_6 \ncong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$$

The question might arise on whether there is a suitable generalization of the direct product to classify the above group. More generally, is there a way to capture the structure of HK, where H and K are *arbitrary* groups (i.e. not subgroups of a larger group G). That question is explored in this section. Note that if G was abelian then automatically, $K \trianglelefteq G$, since every subgroup of an abelian group is normal. Thus, this distinction becomes interesting only when G is non-abelian.

The way we'll do this is by trying to fix the fact that elements of the group H and K will no longer always commute with each other. Note that we always have that $HK = KH$ (proposition 3.4.3), but that didn't imply $hk = kh$. In the dihedral example, we had $rs = sr^{-1}$. If we were to try to write this as a product, we'd get

$$(r, 1)(1, s) = (1, s)(r^{-1}, 1)$$

Another example would be S_3. The group S_3 is another group of order 6 with subgroups of order 2 and 3, for example: $(1\ 2)$ and $(1\ 2\ 3)$). Since $|S_3| = 6$, we again have $|(1\ 2\ 3)| \trianglelefteq S_3$ and $\langle(1\ 2\ 3)\rangle\langle(1\ 2)\rangle \cong S_3$. However, one can easily check that $(1\ 2) \ntrianglelefteq S_3$ (try taking $g = (1\ 2\ 3)$ and see if $g\langle(1\ 2)\rangle g^{-1} = \langle(1\ 2)\rangle$). We again get that $ab \neq ba$, and instead, they commute in a different manner. let $a = (1\ 2)$ and $b = (1\ 2\ 3)$. Then you can verify that:

$$ab = b^2a$$

meaning when a and b comute, b becomes b^2. Thus, we can write

$$S_3 = \langle a, b | a^2 = b^3 = 1,\ ab = b^2a\rangle$$

If were to try to write this as a product, we'd get that

$$(a, 1)(1, b) = (1, b^2)(a, 1)$$

So the question is "Is there a more general way to capture how elements in a group commute"?

As a way to motivate the coming definition, Let's start by having a larger group G, and recall that to define HK as a group, we need $H \trianglelefteq G$ and $K \leqslant G$.

Since HK is a group, we already have that $(h_1k_1)(h_2k_2) = h_3k_3$ for some $h_3k_3 \in HK$ (similarly to how in a direct product, $(a_1, b_1)(a_2b_2) = (a_1a_2, b_1b_2) = (a_3, b_3)$). More precicely, we take $h_1k_1, h_2k_2 \in HK$, we can re-write $h_1k_1h_2k_2$ as $h_3k_3 \in HK$. We showed this in proposition 3.4.3. Doing it again here, notice exactly how we get to write $h_1k_1h_2k_2$ as h_3k_3:

$$\begin{aligned}(h_1k_1)(h_2k_2) &= h_1k_1h_2(k_1^{-1}k_1)k_2\\ &= h_1(k_1h_2k_1^{-1})k_1k_2\\ &= (h_1(k_1h_2k_1^{-1}))k_1k_2\\ &= h_3k_3\end{aligned}$$

where $h_3 = h_1(k_1 h_2 k_1^{-1})$ (since $H \trianglelefteq G$, $gh_1g^{-1} \in H$) and $k_3 = k_1 k_3$.

To get to this construction, we heavily relied on *starting* with the group G and all the properties of the subgroups H, and K (ex. $H \cap K = \{e\}$, $H \trianglelefteq G$). Relying so heavily on a bigger group G does not help in our generalization of the direct product: when defining an arbitrary direct product $G_1 \times G_2$, we did not rely on some bigger group G from which G_1 and G_2 were normal subgroups. Thus, we have to figure out a way to do this construction by only relying on information on H and K.

To do this, we must be able to describe the operation between $(h_1k_1) \cdot (h_2k_2)$ without using the normality of $H \trianglelefteq G$. It is easy for the k elements since $k_3 = k_1k_2$, so the question about about defining $k_1h_2k_1^{-1}$.

We have in fact already encountered a solution to this problem. Since H is normal in G, the group K *acts* on H by conjugation:

$$k \cdot h = khk^{-1} \qquad \text{for } h \in H, k \in K$$

Since H is normal, $khk^{-1} = h' \in H$, meaning if we know how every k acts on every h, we can drop the necessity for a larger group G. In fact, all this information *can* be captured with a group-action:

$$\varphi : K \to \mathrm{Aut}(H)$$

meaning we now only rely on the product of K, H, and the homomorphism φ which captures how H and K interact. Importantly for defining a group with this, $\mathrm{Aut}(H) = \mathrm{Aut}_{\mathbf{Grp}}(H)$, not $\mathrm{Aut}_{\mathbf{Set}}(H)$, and so

$$\varphi(x)(ab) = \varphi(x)(a)\varphi(x)(b)$$

Thus, we've eliminated the need for a larger group G!

With this idea, we can define a our new product:

Definition 5.3.1: Semi-Direct Product

Let H and K be groups and let φ be a homomorphism form K into $\mathrm{Aut}(H)$. Let $\cdot$ denote the (left) action of K on H determined by φ. Let G be the set of ordered pairs (h, k) with $h \in H$ and $k \in K$ and define the following multiplication on G:

$$(h_1, k_1)(h_2, k_2) = (h_1k_1 \cdot h_2, k_1k_2)$$

we usually denote $G = H \rtimes_\varphi K$, or $H \rtimes K$ if the context permits it

Notice how we are not using a bigger group, but a lot of intuition can be derived from thinking about H and K as subgroups of a larger group, showing how subgroups still play an important role in the analysis and conception of semi-direct products, but it is certainly not necessary as we will point out in the examples.

The $\rtimes$ notation is to remember that H is the normal subgruop of $H \rtimes K$ (as we'll prove soon). It might be evident that in most cases $H \rtimes K \ncong K \rtimes H$, since we'd be defining different actions. More on this when we show some examples.

Another way of remembering which group is acting on which is with a pop-culture reference (curtosy of one of my professors): if you are aware of what a "Kamehameha" is from dragon ball, then the symbol "$\rtimes$" can be thought of as Goku's hand preforming the Kamehameha.

If you're unaware of the cultural reference, here's a youtube video I found (or you can google it):

`https://www.youtube.com/watch?v=_ct3ccrU4AI`

As silly as it may be, this turned out to be the most efficient way for me to remember which groups acts on which.

Before we get too hasty, we should prove that this is indeed a group and that all the properties we were talking about earlier hold:

Proposition 5.3.2: Properties Of Semi-Direct Product

1. This multiplication makes G into a group of order $|G| = |H||K|$
2. The sets $\{(h,1)|h \in H\}$ and $\{(1,k)|k \in K\}$ are subgroups of G and the maps $h \mapsto (h,1)$ and $k \mapsto (1,k)$ for $k \in K$ are isomorphisms of these subgroups with the groups H and K respectively:
$$H \cong \{(h,1)|h \in H\} \qquad K \cong \{(1,k)|k \in K\}$$
3. $H \trianglelefteq G$
4. $H \cap K = \{e\}$
5. for all $h \in H$ and $k \in K$, $khk^{-1} = k \cdot h = \varphi(k)(h)$

Proof **:**

1. To show $H \rtimes K$ is a group, we need to show it obeys the axioms of a group
   (a) **associativity**. Let $(a,x),(b,y),(c,z) \in H \rtimes Z$. Then
$$\begin{aligned}((a,x)(b,y))(c,z) &= (a\ x \cdot b, xy)(c,z)\\ &= (a\ x \cdot b\ (xy) \cdot c, xyz)\\ &= (a\ x \cdot b\ x \cdot (y \cdot c), xyz)\\ &= (a\ \varphi(x)(b)\ \varphi(x)(y \cdot c), xyz)\\ &= (a\ \varphi(x)(b\ y \cdot c), xyz)\\ &= (a\ x \cdot (b\ y \cdot c), xyz)\\ &= (a,x)(by \cdot c, yz)\\ &= (a,x)((b,y),(c,z))\end{aligned}$$
   where I switched over to homomorphism notation to show that since $\varphi(x)$ is a homomorphism (note that $\varphi(x) : K \to K$, not $\varphi : K \to K$), then $\varphi(x)(st) = \varphi(x)(s)\varphi(x)(t)$.
   (b) **Identity**: I claim that $(1,1)$ is an identity. In particular, for any $(x,y) \in H \rtimes K$
$$(1,1)(x,y) = (1\ 1 \cdot x, 1y) = (x,y) = (x,y)(1,1)$$

(c) **Inverse**: I claim that for any $(h,k) \in H \rtimes K$, $(h,k)^{-1} = (k^{-1} \cdot h^{-1}, k^{-1})$ is an inverse. Indeed:

$$\begin{aligned}(h,k)(k^{-1} \cdot h^{-1}, k^{-1}) &= (hk \cdot (k^{-1} \cdot h^{-1}), kk^{-1}) \\ &= (h(kk^{-1}) \cdot h^{-1}, 1) \\ &= (h\ 1 \cdot h^{-1}, 1) = (hh^{-1}, 1) \\ &= (1,1)\end{aligned}$$

and similarly for $(h,k)^{-1}(h,k)$.

Thus, $H \rtimes K$ is a group.

2. Let $\hat{H} = \{(h,1) | h \in H\}$ and $\hat{K} = \{(1,k) | k \in K\}$. Notice that for any $a, b \in H$,

$$(a,1)(b,1) = (ab \cdot 1, 1) = (ab, 1)$$

and for any $a, b \in K$

$$(1,a)(1,b) = (1a \cdot 1, ab) = (1, ab)$$

where $a \cdot 1 = 1$ was shown in ref:HERE (in the $\operatorname{Aut}\mathbf{Grp}(G)$ section should be said we can now do)

$$\varphi(x)(a+b) = \varphi(x)(a) + \varphi(x)(b)$$

leading us to this fact.

Thus, we can see that the inculsion map is a homomorphism, and that the map from H to $\hat{H}$ and K to $\hat{K}$ is isomorophic. By this identification,

3. Since $K \cong \hat{K}$ and $H \cong \hat{H}$, it's clear then that $H \cap K = \{1\}$.

4. If we identify elements k with $(1,k)$ and h with $(h,1)$ Then notice that

$$\begin{aligned}khk^{-1} &= (1,k)(h,1)(1,k)^{-1} \\ &= ((1,k)(h,1))(1,k^{-1}) \\ &= (k \cdot h, k)(1,k^{-1}) \\ &= (k \cdot h\ k \cdot 1, kk^{-1}) \\ &= (k \cdot h, 1) \\ &= k \cdot h\end{aligned}$$

i.e., we have that $khk^{-1} = k \cdot h$. In other words $kh = \varphi(k)(h)k$, so commuting kh made k act on h.

5. Apparnelty, $\hat{K} \leqslant N_G(\hat{H})$. Since $G = HK$ (I guess from earliers intuition?) then "certainly" $H \leqslant N_G(H)$, so we have $N_G(H) = G$¡ i.e, $H \trianglelefteq G$, completing (3)

Point (4) in particular is worth pointing out since this is were we generalized our intuition. In the coming example, play close attention to exactly how the action commute elements.

Remember that the semi-direct products are a generalization of direct products. Exactly how is what this next theorem identifies:

Proposition 5.3.3: When The Semi-Direct Product is a Direct Product

Let H and K be groups and let $\varphi : K - \mathrm{Aut}(H)$ be a homomorphism. Then the following are equivalent:

1. The identity (set) map between $H \rtimes K$ and $H \times K$ is a group homomorphism hence an isomrophism
2. φ is the trivial homomorphism form K into $\mathrm{Aut}(H)$
3. $K \trianglelefteq H \rtimes K$

Proof **:**
p.178

Example 5.3.2: Semi-Direct Products

1. Let H be any abelian group (even of infinite order) and let $K = \langle x \rangle \cong Z_2$. be the group of order 2. Define $\varphi : K \to \mathrm{Aut}(H)$ by mapping x to the automorphism of inversion on H, so that the associated action is $x \cdot h = h^{-1}$, or, written in homomorphism form:
$$\varphi(x)(h) = h^{-1} \qquad \varphi(e)(h) = h$$
Note that this is an automorphism, for
$$f_x : H \to H, \qquad f_x(h) = h^{-1} \qquad f_x(xy) = (xy)^{-1} = x^{-1}y^{-1} = f_x(x)f_x(y)$$
since H is abelian, the map is it's own inverse (making it bijective).
Thus, we can define $H \rtimes_\varphi K$. To distinguish between the group-action and the group operation, I'll ues the $+$ symbol to represent the operation for both groups. To get comfortable with this, let's do a computation:
$$(a_1, x)(a_2, x) = (a_1 + x \cdot a_2, x + x) = (a_1 a_2^{-1}, 0)$$
$$\begin{aligned} xhx^{-1} &= (0, x)(h, 0)(0, x^{-1}) \\ &= (0 + x \cdot h, 0 + x)(0, x^{-1}) = (-h, x)(0, x) \\ &= (-h + x \cdot 0, x + x) \\ &= (-h - h, 0) \\ &= (-h, 0) \\ &= h^{-1} \end{aligned}$$
Thus, we have:
$$xhx^{-1} = h^{-1} \Rightarrow xh = h^{-1}x$$
in the special case where $H = Z_n$, then $\mathbb{Z}/n\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z} \cong D_n$!!! if $H = \mathbb{Z}$, then we'd say $G = D_\infty$.
2. Let $H = \mathbb{R}^2$ and $K = SO(2)$.
3. Since simple groups have no normal subgroups, simple groups cannot be decomposed into semi-direct products, showing further how they are like "prime numbers" for groups.

More examples in the book p.178

With the idea of the semi-direct product in mind and a couple of examples, we end this section by given sufficient and necessary conditions to recognize semi-direct products:

Theorem 5.3.4: Recognition Theorem For Semi-Direct Products

Suppose G is a group with subgrosup H and K such that

1. $H \trianglelefteq G$
2. $H \cap K = 1$

Let $\varphi : K \to \mathrm{Aut}(H)$ be the homomorphism define dby mapping $k \in K$ to the automrophis mof left conjugation by k on H. Then

$$HK \cong H \rtimes K$$

In particualr, if $G = HK$, with H and K satisfying the aforementioend conditions, then

$$G \cong H \rtimes K$$

Put another way, a group is a semi-derict product if there exists a complement for some proper normal subgroup of G.

***Proof* :**
p. 180

(A word on how not all groups are semi-direct products. Show example. So Semi-direct product is actually a special kind of extension (will show in exercises). This type of extensions (called a split extension) will come back again later with modules (part III, chapter 16).

There is a product that does represent any extension by the two groups on the left and right (sort of); it is called teh wreath product, and it comes along with the universal embedding theorem.

`https://en.wikipedia.org/wiki/Wreath_product`

Definition 5.3.5: Split Sequecne

Given a short exact sequence

$$1 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 1$$

we say that it *splits* if there eexists a homomorphism $\psi : H \to G$ such that $\psi \circ f = \mathrm{id}_G$ (so in a sense, also embedd C in B). We can write this as the commutative diagram:

$$\begin{tikzcd}
1 \arrow[r] & A \arrow[r] & B \arrow[r, "f"] & C \arrow[r] & 1 \\
 & & & C \arrow[ul, "\psi"] \arrow[u, "\mathrm{id}"'] &
\end{tikzcd}$$

Point out how with this we can make $B \cong C \rtimes_\alpha \ker(\varphi) = C \rtimes A$, where the action $\alpha : C \to \mathrm{Aut}(A)$, $\varphi(c)(a) = c \cdot a = \psi(c)^{-1} a \psi(c) [\in A]$.

If the action is trivial, then the extension can be represented as a direct product

$$1 \longrightarrow A \longrightarrow A \times C \longrightarrow C \longrightarrow 1$$

So is every short exact sequence split? The answer is unfortunately no

Example 5.3.3: Counter-Example
Take

$$1 \longrightarrow C_2 \longrightarrow C_4 \longrightarrow C_2 \longrightarrow 1$$

With the first component mapping $0 \mapsto 0$ and $1 \mapsto 2$. while in the 2nd map, $0, 2$ map to 0 and $1, 3$ map to 1.

5.4 A Categorical approach of the Product

The idea of being able to take the "product" of two objects is not unique to sets and groups. As we continue to explore more and more algebraic structures, we will keep asking ourselves a similar question: Does algebraic structure X have the flexibility/restriction to define products? [4] What we'll be doing then is finding a property which we can use to define what a product is in different algebraic settings.

Let's pick two arbitrary groups A and B. We defined product in the classical definition as the new group $(A \times B, (\cdot_A, \cdot_B))$. Let's denote this new group C. The key idea is that C is a group that contains groups A and B, and you can "recover" A and B from C. The way we can recover these two groups from C is simply by projcting to them:

$$\pi_A : C \to A,\ (a, b) \mapsto a \qquad \pi_B : C \to B,\ (a, b) \mapsto b$$

It should be checked that π_A and π_B are both surjective homomorphisms. Notice that $\pi_A(C) \cong A$ and $\pi_A(C) \cong B$. In this way, we have "recovered" A and B from C. Furthermore, C contains no more information than information about A and B. So essentially, if there exists a group C that most efficienly stores A and B, in particular if there are two functions (in our case π_A, π_B) which let's us recover A and B, then we can say C is the product of A and B. let's write this group and these two functions (C, π_A, π_B)

But is C the most efficient group for storing information about the group A and B? We can answ3re this by relating C to any other group, as is common in category theory. Let's say we have a group $G \in \mathrm{Obj}(Grp)$ were we have a homomorphism $G \xrightarrow{f_A} A$ and $G \xrightarrow{f_B} B$ (so some structe of G is in A and B). Then does there exist a homomorphism $f : G \to C$ which can compose with π_A and π_B and produce f_1 and f_2? In other words: (Rushed)

[4] As usual with category theory, the question of whether the product exists goes beyond algebra: Topological spaces, uniform spaces, graphs, ordered sets, and other domains of mathematics will also have products!

$$\begin{tikzcd}
 & G \arrow[d, "f"] \arrow[ddl, bend right, "f_A"'] \arrow[ddr, bend left, "f_B"] & \\
 & C \arrow[dl, "\pi_A"] \arrow[dr, "\pi_B"'] & \\
A & & B
\end{tikzcd}$$

In a sense, it's saying that whatever structure G has of A and B, C must have it too, since we can pass through C wihtout loosing that structure.

We can now foramlize this idea with categories: Let $C_{A,B}$ be a category. The objects are

$$\begin{tikzcd}
 & X \arrow[dl, "f_A"'] \arrow[dr, "f_B"] & \\
A & & B
\end{tikzcd}$$

and it can be represented as (X, f_A, f_B). If (X, f_A, f_B) and (Y, f_A, f_B) are objects, then a morphism $f : (X, f_A, f_B) \to (Y, f_A, f_B)$ exists if the following diagram commutes:

$$\begin{tikzcd}
 & X \arrow[d, "f"] \arrow[ddl, bend right, "f_A"'] \arrow[ddr, bend left, "f_B"] & \\
 & Y \arrow[dl, "\pi_A"] \arrow[dr, "\pi_B"'] & \\
A & & B
\end{tikzcd}$$

Example 5.4.1: Morphissm In $C_{A,B}$
here. Examples and counter-examples

Proposition 5.4.1: $C_{A,B}$ Froms A Category

Let $C_{A,B}$ be defined as above. Then $C_{A,B}$ forms a category

***Proof* :**
Left as an exercise that everything works out.

We can ask whether $C_{A,B}$ has *final* objects, That is, an object $(G, f_A, f_B) \in C_{A,B}$ where for any $(H, g_A, g_B) \in C_{A,b}$, $\mathrm{Hom}((H, g_A, g_B), (G, f_A, F_B))$ is a singleton. In fact, (C, π_A, π_B) is such an object. In other words, we'll show that (C, π_A, π_B) must be the only group with this property since the axioms of groups and homomorphisms forces it to be.

Theorem 5.4.2: Universal Property Of The Product For Groups

Let A and B be groups in G. Then $A \times B$ is a group, $\pi_A : A \times B \to A$ and $\pi_B : A \times B \to B$ are homomorphisms. For any group H, if $f_A : H \to A$ and $f_B : H \to B$, then there exists a unique f such that $f_A = \pi_A \circ f$ and $f_B = \pi_B \circ f$. The morphism f is denoted $f_A \times f_B$.

Proof **:**
We showed that $(A \times B, \cdot_A, \cdot_B)$ in proposition ref:HERE. Next, let $(A \times B, \pi_A, \pi_B)$ be as above and let $(H, f_A, f_B) \in C_{A,B}$. Define

$$f : H \to A \times B \qquad f(h) = (f_A(h), f_B(h))$$

Then we can see that $f_A(h) = (\pi_A \circ f)(h) = \pi_A(f_A(h), f_B(h)) = f_A(h)$, and similalry for f_B, showing that the diagram indeed comutes, and so f is indeed a morphism in $C_{A,B}$.

To show that f is unique, let's assume there is another morphism f'. Then

$$f_A(h) = (\pi_A \circ f')(h) = \pi_A(f'(h)) = \pi_A(f'(h)_A, f'(h)_B) = f'(h)_A$$

where $f'(h)_A$ and $f'(h)_B$ represent the components of $f'(h)$. Similarly, $f'(h)_B = f_B(h)$. Therefore, $f'(h) = (f_A(h), f_B(h))$. However, that is exactly what f does:

$$f(h) = (f_A(h), f_B(h)) = f'(h)$$

thus $f = f'$, showing uniqueness. Thus (C, π_A, π_B) is the final object of $C_{A,B}$. In other words, $A \times B$ respects the universal property of the product in Group.

We can further generalize this for an arbitrary direct product, that is if $(G_i)_{i \in I}$ is a family of groups, then their direct product $G = \prod_{i \in I} G_i$ along with $\pi_i : G \to G_i$ for every $i \in I$ will be a final object in $C_{(G_i)_{i \in I}}$.

Theorem 5.4.3: Universal Property For Arbitrary Direct Product Of Groups

For every group H, and every I-indexed family of morphisms $f_i : H \to G_i$, there exists a unique morphism $f : Y \to X$ such that the following diagrams commute for every $i \in I$:

$$\begin{tikzcd} H \arrow[d, "f"] \arrow[dr, "f_i"] & \\ G \arrow[r, "\pi_i"] & G_i \end{tikzcd}$$

Proof **:**
Similar to 5.4.3. All the steps generalize without a problem.

Exercise 5.4.1

1. Given that you've seen the universal property of products for groups, what do you think it

would be for sets? What would be the generalization for an aribtary category? (For a hint, see the preliminary chapter or section 5.4)

5.4.1 The dual of the Product

What happens if you flip the arrows? Instead of looking at $C \xrightarrow{\pi_A} A$ and $C \xrightarrow{\pi_B} B$, we look at $A \xrightarrow{i'_A} C$ and $B \xrightarrow{i'_A} C$, and for all groups G, if we have $A \to G$ and $B \to G$, then there exists a homomorphism $f : C \to G$ such that the following diagram commutes

$$\begin{tikzcd}
A \arrow[dr, "i'_A"] \arrow[ddr, bend right, "f'_A"'] & & B \arrow[dl, "i'_B"'] \arrow[ddl, bend left, "f'_B"] \\
& C \arrow[d, "F"] & \\
& G &
\end{tikzcd}$$

Does $(A \times B, i_A, i_B)$ where $i_A(a) = (a, 1)$ and $i_B(b) = (1, b)$ work? Let's break down the restrictions of this commutative diagram to check. By the above commutative diagram, we need that $f'_A(a) = F(i_A(a)) = F((a, 1))$ and that $f'_B(b) = F(i_B(b)) = F((1, b))$. So we have

$$F((a, 1))F((1, b)) = f'_A(a)f'_B(b)$$

Since F must be a homomorphism, then:

$$F((a, 1))F((1, b)) = F((a, 1)(1, b)) = F((a, b)) = F((1, b)(a, 1)) = F((1, b))F((a, 1))$$

that is $f'_A(a)f'_B(b) = f'_B(b)f'_A(a)$, that is, the image of A and B must always commute. However, we can see that there will exist a group G and homomorphisms f'_A f'_B where $f'_A(a)f'_B(b) \neq f'_B(b)f'_A(a)$:

Example 5.4.2: Counter-example
Let $A = \mathbb{Z}_2$ and $B = D_6$ so that we get

$$\begin{tikzcd}
\mathbb{Z}_2 \arrow[dr, "i_A"] & & D_6 \arrow[dl, "i_B"'] \\
& \mathbb{Z}_2 \times D_6 &
\end{tikzcd}$$

Let $G = D_6$, $f'_A : \mathbb{Z}_2 \to D_6$, $f'_A(x) = \sigma^x$, $f'_B : D_6 \to D_6$, $f'_B(x) = x$ (i.e. f'_B is the identity map). Then is it true that there exists a homomorphism F such that:

$$\begin{tikzcd}
\mathbb{Z}_2 \arrow[dr, "i_A"] \arrow[ddr, bend right, "x \mapsto \sigma^x"'] & & D_6 \arrow[dl, "i_B"'] \arrow[ddl, bend left, "\mathrm{id}"] \\
& \mathbb{Z}_2 \times D_6 \arrow[d, "F"] & \\
& D_6 &
\end{tikzcd}$$

By definition $F((a,1)) = F(i_A(a)) = f'_A(a) = \sigma^a$ and $F((1,b)) = F(i_B(b)) = f'_B(b) = b$, so

$$F((a,b)) = \sigma^a b$$

and by what we just saw, we must also have that $F((a,b)) = b\sigma^a$ for every $a \in A$ and $b \in B$. However

$$F((1,2)) = \sigma r^2 \neq r^2\sigma$$

showing that F is *not* a well-defined homomorphism!

It should be noted that if we *only* stuck to abelian group, then this wouldn't be a problem, since we always have that $ab = ba$, so $F((a,b))$ is well-defined. In fact, it is worth checking that if A and B are abelian groups, $((A \times B), i_A, i_B)$ *does* respect the commutative diagram, and is the initial object in the category who's objects are triples (G, f_A, f_B) (where G *has* to be abelian) and morphsism are the aformentioned commutaive diagram.

The problem here came that $A \times B$ actually added structure on it. As we saw in proposition 5.1.2[5], for any $x, y \in A \times B$, $xy = yx$. (Talk about how then the product is a way to store information by somehow adding conditions to let the storage happen??)

So does there exist a group and maps will be the initial object in this category? The answer to this question is actually yes! It is a group we have not yet encountered. The main idea here is that this group is supposted to be the most generic way of having A and B (so as little relations btween A and B) so that the homomorphism from this group to any G does not preserve these properties. We have encountered a similar situation like this before when defining the *free group* (see section 2.3.3). Something analogous is defined called the *free product*.

Definition 5.4.4: Free Product

Let H and K be groups, S_H be a generating set for H and S_K be a generating set for K, R_H and R_K be the relations given the generating sets respectively (at this point, I already showed presentations, so I should be able to introduce R_H and R_K rigorously). Then

$$H * K = \langle S_H \cup S_k | R_H \cup R_K \rangle$$

Example 5.4.3: Free Products

Take $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z}$. Mention that this is the group mentioned in section re:HERE (chap 2 near beginning) that as a group for which the torsion subset as not a group

Exercise 5.4.2

1. Show that $(H * K, i_H, i_K)$ respects the universal property of the co-product in **Grp**.

Infinite abelian case

It should be mentioend again that if we restrict to **Ab**, then the universal property of the coproduct is fulfilled by $H \times K$ (notice that $H * K = H \times K$). Is this also true in the infinite case: given some groups $G_1, G_2, \cdots$, does $G_1, G_2 \times ...$ respect the universal property of the coproduct? The answer is no!

Example 5.4.4: Example

Let's say $G = \prod_{i=1}^{\infty} \mathbb{Z}$. Let's say that every component group is represented by G_i. For every G_i, define the inclusion map $G_i \xrightarrow{\iota_i} G$. Let's say $G_i \xrightarrow{f_i} \mathbb{Z}$, $f_i(x) = x$ Then if G respects the universal property of the coproduct, then there exists a function F such that

$$f_i(g_i) = F(0, 0, ..., g_i, ..., 0, ...)$$

This is true for every component of F. Now here is the question: for some elements $1 \in G_1$, $1 \in G_2 ... 1 \in G_i, ...$, what is:

$$F(1, 1, ..., 1, ...)$$

By linearity, we can see that

$$F(1, 1, ..., 1, ...) = 1 + F(0, 1, 1, ..., 1, ...)$$

and we can keep going this way. However, the problem is that we will continue indefenetily. More precicely, there is no number $y \in \mathbb{Z}$ such that $y = F(1, 1, 1, ..., 1, ...)$. Thus, F is *not* a well-defined function!

The problem came from the fact that with the flexibility of having an infinite sum, we run into the possibility of never terminating our sum, that is, we can create a scenario where we have series which do not converge (TBD?).

A group still exists which can fix our scenario and respct the universal property for coproducts in the infinite case:

Definition 5.4.5: Direct Sum Of Groups

Let $\{A_i\}_{i \in I}$ be a family of groups. Then the *direct product*

$$A = \bigoplus_{i \in I} A_i$$

is a subset of $\prod_{i \in I} A_i$ consisting of all elements $(x_i)_{i \in I}$ with $x_i \in A_i$ such that $x_i = 0$ for all but finitely many indices i (cofinitely many indices).

The proof that this is a group was an exercise ref:HERE (in section 1.1). Note that in the finite case, $\bigoplus_{i=1}^{n} G_i = \prod_{i=1}^{n} G_i$. It only differs once we reach the infinite case

Theorem 5.4.6: Universal Property Of Coproducts In Ab

Let I be an index for the family of groups $G_{i \in I}$. Then $G = \bigoplus_{i \in I} G_i$ respects the universal property of the corpoduct in **Ab**.

Proof **:**

This proof is very similar to the infinte case one and the one for the usual universal property of the direct product.

5.4.2 Universal Property of Semi-Direct Products?

I want to bring in the intuition of a product that when elements commute it introduces a twist. Show this through Cayley graphs

here

Then generalize the how the universal property of products with coslice categories to fibered categories.

However, it also seems like there isn't a universal property for semi-direct products, since it's not unique (saw that wiki entry where many semi-direct products shared the same structure, so I'm not sure where this yet goes in the "global landscape of mathematics"

`https://en.wikipedia.org/wiki/Semidirect_product#Generalizations`

Also, to show lack of uniqueness:

$$(C_3 \rtimes D_8) \cong (Q_{12} \rtimes C_2) \cong (D_{12} \rtimes C_2) \cong (V \rtimes D_6)$$

There is also no necessary and sufficient conditions that are known for semi-direct products, although they can be characterzied by being a right-split exact sequence.

6

Further Topic in Group Theory

Miscelaneous things I don't know how to classify yet

More on J.H. program? More on solvable groups? introduce metabelian groups? nilpotent goups? (Have to mention Krasner–Kaloujnine universal embedding theorem and make exercises defining the wreath product)

6.1 Generalization of Abelian Groups

So far we saw that

$$\text{cyclic} \Rightarrow \text{Abelian}$$

We found a way to find how close agroup is to being abelian using commutator groups. If $G' = \{1\}$, then G is abelian. Otherwise, G is not abelian. We'll develop further generalization's from this concept.

Definition 6.1.1: Metabelian Group

Let G be a group. If $G^{(1)} \neq \{1\}$, but $G^{(2)} = \{1\}$, we sy that G is a *metablean* group

Proposition 6.1.2: Properties Of Metabelian Groups

Let G is metabelian, then

1. any subgroup is metabelian
2. If $\varphi : G \to H$ is a homomorphism, then $\varphi(G)$ is metabelian

Proof **:**

1. Let $H \leqslant G$ be a subgroup of G, and consider H''. If, $H'' \subseteq G'' = \{1\}$, and so H is metabelian as well. This fact can easily be shown direclty. Notice that $H' \subseteq G'$, since if $x = aba^{-1}b^{-1} \in G$, then this is also a commutator for the group G, and so $x \in G'$. By the same logic, the group H'' will be a subset of G'', meaning H'' is trivial as well.
2. Let $\varphi : G \to H$ be a homomorphism and G a metabelian group, and consider $\varphi(G) \leqslant H$. We'll show that homomorphisms preserve commutator subgroups to show that $\varphi(G)$ is metabelian. In paticualr, if $N \subseteq G$, then $\varphi(N') = \varphi(N)'$.
   (a) ($\subseteq$) Let $x \in \varphi(N')$. Then $x = \varphi(xyx^{-1}y^{-1}) = \varphi(x)\varphi(y)\varphi(x)^{-1}\varphi(y)^{-1}$. And so by definition, $x \in \varphi(N)'$.
   (b) ($\supseteq$) Let $x \in \varphi(N)'$. Then $x = \varphi(x)\varphi(y)\varphi(x)^{-1}\varphi(y)^{-1} = \varphi(xyx^{-1}y^{-1})$. And so by definition $x \in \varphi(N')$.

   Thus
   $$\varphi(G)'' = \varphi(G')' = \varphi(G'') = \varphi(\{1\}) = \{1\}$$
   and so $\varphi(G)$ is also metabelian.

Example 6.1.1: Metabelian Groups

1. Any abelian group is always metabelian
2. if A and B are abelian groups, and $A \rtimes B$ is well-defined, then $A \rtimes B$ is metabelian.

Proposition 6.1.3: Extension of Abelian Groups

Let G be a group. If $H \trianglelefteq G$ is abelian and G/H is abelian, then G is metabelian. More generally, if G extends an abelian group N by an abelian group M, then G is metabliean, i.e, the following is a short exact sequence

$$1 \longrightarrow A \longrightarrow G \longrightarrow M \longrightarrow 1$$

Furthermore, the converse is true too, namely if $H \trianglelefteq G$ and G/H are abelian, then $G'' = \{1\}$, showing this is an equivalent definition

***Proof* :**
Let $A \trianglelefteq G$ and G/A be abelian. Then by proposition ref:HERE, $G' \subseteq A$, meaning $G' = \{1\}$. Evidentally then, $G'' = \{1\}$. Conversly, if $G'' = \{1\}$, then G' the elements of G' commute. Note that this does not imply $G' \trianglelefteq G$ immediately, since we did not esatblish those elements commute with every element of G, and so we must show that $G' \trianglelefteq G$. Let $g \in G$ and $h \in G'$. By definition, $ghg^{-1}h^{-1} \in G'$. Since $h \in G'$, $ghg^{-1}h^{-1}h = ghg^{-1} \in G'$, and thus $G' \trianglelefteq G$. Thus, by proposition ref:HERE, G/G' is also abelian completing the other direction.

Thus this is an equivalent definition of metabelian.

6.1.1 Solvable Groups

Metabelian groups are a simple example of a group called a *solvable* group. Instead of looking at the 2nd commutator subgroup being trivial, we can ask if the nth commutator subgroup will be trivial:

Definition 6.1.4: nth Commutator Group

Let G be a group and $G' = G_{(1)}$.

$$G_{(n)} = [G_{(n-1)}, G_{(n-1)}] = \left\{aba^{-1}b^{-1} \middle| a, b \in G_{(n-1)}\right\} = (G_{(n-1)})'$$

represents the nth commutator group.

Definition 6.1.5: Solvable Groups

A group G is *solvable* if there is a subnormal series with commutator subgroup:

$$G = G_0 \trianglerighteq G_{(1)} \trianglerighteq G_{(2)} \trianglerighteq \cdots \trianglerighteq G_{(n)} = \{1\}$$

such a subnormal series is called the *derived series*. I $G_{(n)} = 1$, we say G has *solvable length* n.

Example 6.1.2: Solvable Groups

1. Every abelian group is solvable, since $1 \trianglelefteq G$ is a normal series.
2. Any metabelian group is solvable, with the series is caleld the *derived series of* G.

$$1 = G'' \trianglelefteq G' \trianglelefteq G$$

   Since G/G' is abelian, and $G'/G'' \cong G'$ which is abelian.

There are equivalent definitions of solvability that are worth keeping in mind to have a variety of intuitions.

Lemma 6.1.6: characteristic of nth commutator

Let G be a group. Then $G_{(i+1)} \operatorname{char} G_{(i)}$ for any $i \in \mathbb{N}$. Consequently, $G_{(i+1)} \trianglelefteq G_{(i)}$.

Proof **:**
We have already proved this in proposition 5.2.2. Then normality comes from proposition 4.5.6.

We can also prove normality directly. We want to show that for any $g \in G_{(i)}$, $gG_{(i+1)}g^{-1} = G_{(i+1)}$. Take $h \in G_{(i+1)}$. Then by definition of $G_{(i+1)}$, $ghg^{-1}h^{-1} \in G_{(i+1)}$. Since $h \in G_{(i+1)}$,

$$ghg^{-1}h^{-1}h = ghg^{-1} \in G_{(i+1)}$$

implying $gG_{(i+1)}g^{-1} = G_{(i+1)}$, and so $G_{(i+1)} \trianglelefteq G_i$. since i was arbitrary, we see that this holds true for any nth commutator subgroup, as we sought to show.

Theorem 6.1.7: Equivalent Definitions Of Solvable group

The following are equivlent

1. G is solvable
2. There exists a subnormal series

$$G = A_0 \trianglerighteq A_1 \trianglerighteq A_2 \trianglerighteq \cdots A_n = \{1\}$$

such that A_i/A_{i+1}. Such a subnormal series is caleld a *subnormal series with abelian factors.*

Proof **:**
$(1 \Rightarrow 2)$ We want to construct a derived series. However, we already have one: since $G_{(i+1)} \trianglelefteq G_{(i)}$, let $A_i = G_i$.

$(2 \Rightarrow 1)$ Since $G/A_{(1)}$ is abelian, then by proposition 5.2.2 $G_{(1)} \leqslant A_1$. Then inductively, let $G_{(i)} \leqslant A_i$. then since A_i/A_{i+1} is abelian, there exists a group $G^* \leqslant A_{i+1}$ such that A_i/G^* is abelian. Then, take $f : G_{(i)} \to A_i/G^*$, $x \mapsto [x]$. Consider $[x], [y] \in f(G_{(i)})$. Then since $f(G_{(i)})/G^*$ is a subgroup of A_i/A_{i+1}, it is also abelian, so

$$[xy] = [yx] \Rightarrow xyg = yx \qquad g \in G*$$

moving variables around we get

$$xyx^{-1}y^{-1} = [x, y] = g$$

since x^{-1}, y^{-1} was arbitrary, we see that $G_{(i+1)} \leqslant G* \leqslant A_{i+1}$, that is

$$G_{(i+1)} \trianglelefteq A_i$$

Thus, once we reach A_n, we see that

$$G_{(n)} \trianglelefteq A_n = \{1\}$$

thus $G_{(n)} = 1$, as we sought to show.

The use of the 2nd definition is to show that finding any solvable series for a group is enough to show a group is solvable. The advantage of the 1st definition is that it is the most refined solvable series of

a group, since the commutator subgroup is smallest subgroup which makes the quotients A_i/A_{i+1} abelian.

Proposition 6.1.8: Properties Of Solvable Groups

Let G be solvable. Then

1. subgroups of solvable groups are solvable
2. if $\varphi : G \to H$ is a homomorphism, then $\varphi(G)$ is solvable
3. if M is an extension of a solvable group by a solvable group, then M is solvable.

Proof **:**
For (1) and (2), the commutator subgroup definition makes these proofs much easier. To get used to both definitions, we'll prove the statement using both:

1. Let $G_{(n)} = \{1\}$, and consider $H \leqslant G$. Then as we've seen, $H_{(n)} \leqslant G_{(n)}$ meaning H is solvable.

   Using the other definition, we'll construct a solvable series for H. Let $G \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq \cdots \trianglerighteq G_n$ be a solvalbe series for G. Take the homomorphism

   $$\varphi_i : H \cap G_i \to G_i/G_{i+1} \qquad \varphi(x) = xG_{i+1}$$

   which has kernel $(H \cap G_i) \cap G_{i+1} = H \cap G_{i+1}$. Since it's a kernel, it is a normal subgruop of $H \cap G_i$, and the quotient $H \cap G_i/H \cap G_{i+1}$ maps injectivly into G_i/G_{i+1}. Since it is commutative, then so is $H \cap G_{i+1}$. Thus

   $$H \trianglerighteq H \cap G_1 \trianglerighteq H \cap G_2 \trianglerighteq \cdots \trianglerighteq H \cap G_n$$

   is a solvable series for H.

2. We've shown that $\varphi(G'') = \varphi(G')' = \varphi(G)''$ (proposition 6.1.2). Continuing inductively, we'd get $\varphi(G_{(n)}) = \varphi(G)_{(n)}$. We can see that the derived series will be preserved.

   Using the other definition, let $\overline{G}$ be the quotient gorup of G, and let $\overline{G}_i$ be the image of G_i into $\overline{G}$. Then

   $$\overline{G} \trianglerighteq \overline{G}_1 \trianglerighteq \cdots \trianglerighteq \overline{G}_n = \{1\}$$

   is a solvable series for $\overline{G}$

3. Let $N \trianglelefteq G$ and G/N be solvable. We'll show then that G has a series. Since N and $\overline{G}$ is solvable, then they both have a solvable series:

   $$\overline{GG}_1 \trianglerighteq \overline{G}_2 \trianglerighteq \cdots \trianglerighteq \overline{G}_n = \{1\}$$

   $$N \trianglerighteq N_1 \trianglerighteq N_2 \trianglerighteq \cdots \trianglerighteq N_m = \{1\}$$

   Thus, by the 3rd isomrophism theorem $\overline{G}_i$ in G we get $G_i/G_{i+1} \cong \overline{G}_i/\overline{G}_{i+1}$. Combining these series, we get

   $$G \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_n = N \trianglerighteq N_1 \trianglerighteq \cdots \trianglerighteq N_m = \{1\}$$

   which is a solvable series for G, completing the proof.

Corollary 6.1.9: Finite P-Groups

Let G be a finite p-group. Then G is solvable.

Proof **:**
Let's use induction on the order of G. The base case is trivial, so let's consider an order $|G| > 1$. Then by hw exexerice (or corollary 4.16 in Milne Group Theory book) $Z(G)$ has non-trivial center. Since $Z(G)$ is normal in G, $G/Z(G)$ is well-defined. Since $Z(G)$ is non-trivial $G/Z(G)$ has smaller order. Thus, by the induction hypothesis, $G/Z(G)$ is solvable. Since $Z(G)$ as a group is commutative, by proposition 6.1.8(2), G is solvable.

There is actually one more incredibly powerful theorem for which groups are solvable:

Theorem 6.1.10: Feit-Thompson's Theorem

Let G be a finite group of odd order. Then G is solvable.

Proof **:**
Huge proof requiring 255 pages.

From this statement, we can state that either a finite group is solvable, or it has an element of order 2 (see exercice 1.2.1)!

Theorem 6.1.11: Solvability Condition

The finite group is solvable if and only if for every divisor n of $|G|$ such that $\left(n, \frac{|G|}{n}\right) = 1$, G has a subgroup of order n .

Proof **:**
very difficult proof not in textbook.

6.1.2 Nilpotent Groups

(put somehere ho to classify finite abelian groups) We've defined solvable groups using nth commutator groups to give a notion of generalization from being commutative. What if we used the center instead? SDince $Z(G) \trianglelefteq G$, $G/Z(G)$ is well-defined. To generalize, we can define:

$$Z^2(G) = Z(G/Z(G))$$

that is, the 2nd center is the center of the quotient $G/Z(G)$. If we can continue this chain with $Z^n(G)$, and eventually get a commutative group $z^n(G) = G$, we call such a group a *nilpotent group*:

Definition 6.1.12: Nilpotent Groups

Let G be a group. If $Z^m(G) = G$. for some $m \in \mathbb{N}$, then G is said to be nilpotent. The smallest such m is caleld the *(nilpotency) class of G*.

Example 6.1.3: Nilpotent Groups

1. If G is abelian, then $Z(G) = G$, meaning all abelian groups are nilpotent
2. If G is meta-abelian, Then $G/Z(G)$ is abelian (since $G' \subseteq Z(G)$), meaning $Z(G/Z(G)) = G/Z(G)$, and so is nilpotent with nilpotency class 2.
3. Every p-groups are an example of nilpotent groups. Let G be a finite p-group, so that $|G| = p^a$ for some $a \in \mathbb{N}$. If $Z(G) = G$, we're done. If not, then since $Z(G)$ is non-trivial (proposition ref:HERE), $|Z(G)| \mid |G|$, so $|Z(G)| = p^{a_2}$ for some $a_2 < a$. But then $Z(G)$ is again a p-group, so we can continue this process. Thus, we get an ever shrinking sequence of nubmers $a > a_2 > a_3 \cdots a_n$, till eventually it must be that either $a_n = 1$ or $a_n = 0$. If $a_n = p$, then the associated group is cyclic, implying it's abelian, and we're done. If $a_n = 0$, then we have the trival subgroup, which is trivially abelian.
4. This is not so much an example as future motivation: Nilpotent groups make an appearance in the classification of Lie Groups.

Notice that for any $[g] \in Z^2(G)$, $[g, x] \in Z(G)$ for any $x \in G$. To see this, pick $[x] \in G/Z(G)$ (i.e., the coset which can have x as it's reprsentative). Then since $[g] \in Z^2(G)$, $[gx] = [xg]$. Then $gxz = xgz'$ for some $z, z' \in Z(G)$. Then$g^{-1}x^{-1}gx = z'z^{-1} \in Z(G)$. Thus, $Z^2(G) \subseteq Z(G)$. This holds true for any $Z^n(G)$ (defined inductively). Thus, can also get a subnormal series, called the *upper central series*:

$$\{1\} \subseteq Z(G) \trianglelefteq Z^2(G) \cdots Z^n(G) \cdots$$

If this series terminates with $Z^n(G) = G$, then we say that G is a *nilpotent group*.

Proposition 6.1.13: Equivalent Definition

The following are equivalent:

1. G is nilpotent (i.e. there exists an *upper central series*)
2. There exists a *lower central series*

$$G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq \cdots \trianglerighteq G_n = \{1\}$$

   where $G_i = [G, G_{i-1}] = \left\{ghg^{-1}h^{-1} \middle| g \in G, h \in G_{i-1}\right\}$.

3. There exists a *central series*

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_n = \{1\}$$

   where $G_i/G_{i+1} \leqslant Z(G/G_i)$, or equivalently, $[G, G_{i+1}] \leqslant G_i$

Proof **:**
(1) $\Rightarrow$ (2) Let G be nilpotent. Then we saw that if $g \in Z^{i+1}(G)$, then $[g, x] \in Z^i(G)$, meaning $[G, Z^{i+1}(G)] \leqslant Z^i(G)$. Then decomposing recusively, we see that $G_{i+1} = [G, G_i] \trianglelefteq Z^i(G)$. Showing that every G_i exists, and will have to terminate at G_n. Furthermore, by the same justification as $G_{(i)}$, $G_{i+1} \trianglelefteq G_i$. Thus, they form a subnormal series

(1) $\Rightarrow$ (3) easy to do

(2) $\Rightarrow$ (1)

With the lower central series, we can gain an intuition on why the group is called nilpotent. Let G be any group and $g \in g$ be any element. Define the *adjoint action*

$$\mathrm{ad}_g : G \to G \qquad \mathrm{ad}_g(x) = [g, x]$$

Then if G has nilpotency class n, then $(\mathrm{ad}_g)^n(x) = e$, for all $x \in G$. In this sense, the adjoint action is nilpotent. Note that if this is true for each element, this does not imply that G is nilpotent

Example 6.1.4: Example
here

Such groups are caleld *Engel groups*, and come up in Lie Algebra.

Proposition 6.1.14: Nilpotent Then Solvable

Let G be a nilpotent group. Then G is solvable

Proof **:**
Notice that $G_{(n)} \trianglelefteq G_n$, meaning the there must be a solvable length of G, since for some n, $G_{(n)} \subseteq G_n = \{1\}$, implying $G_{(n)} = 1$, showing that G is solvable.

Note that the converse is not true

Example 6.1.5: Solvable, but not Nilpotent
let

$$G = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \middle| a, b, c \in \mathbb{R}, ab \neq 0 \right\} \leqslant GL_2(\mathbb{R})$$

Then I encourage you to check that $Z(G) = \{aI | a \neq 0\}$ where I is the identity matrix. Furthermore $Z^2(G) = Z(G/Z(G)) \cong \{1\}$, i.e., we don't resolve to the an abelian group but a trivial group, and so G is not nilpotent.

However, G is solvable. We can form a solvable series using $H = \left\{ \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} | a \in \mathbb{R} \right\}$. It should be verified that $H \trianglelefteq G$, and that $G/H \cong \mathbb{R}^\times \times \mathbb{R}^\times$ (meaning it's abelian), and that $H \cong (\mathbb{R}, +)$, meaning H is abelian, so $G'' = H' = \{1\}$, showing it is indeed solvable.

Another example would be S_3. Since $S_3 \cong \mathbb{Z}_3 \rtimes \mathbb{Z}_2$, and both $\mathbb{Z}_3$ and $\mathbb{Z}_2$ are nilpotent (since their abelian), then the semi-direct product of nilpotent groups is not necessarily nilpotent. More generally, the extension of a nilpotent group by a nilpotent group is not necessarily nilpotent.

Thus, we have the folowing chain of implications

$$\text{cyclic} \Rightarrow \text{abelian} \Rightarrow \text{nilpotent} \Rightarrow \text{solvable} \Rightarrow \text{finitely generated (show } A_5 \text{ not solvable)}$$

So nilpotent groups don't interact well with extensions. However, the fac that their slightly stronger than extensions, but more general then abelian groups, brings us some useful properties:

Proposition 6.1.15: Properties Of Nilpotent Groups

Let G be a nipotent group. Then

1. If H is a proper subgruop of G, then H is a proper normal subgruop of $N_G(H)$. We call this property the *normalizer property*, and we sometimes say that the "normalizer grows"
2. Every Sylow subgroup of G is normal
3. G is the direct product of its Sylow subgroups
4. if d divides the oder of G, then G has a noraml subgroup of order d

In fact, (4) implies that G is nilpotent, giving us another definition of nilpotency.

Proof **:**

1. Let G be a nilpotent group and $H \leqslant G$. We want to show that $H \trianglelefteq N_G(H)$. We'll do this by induction on the order of G. This is clearly true when $|G| = 1$, so let $|G| = n$. IF G is abelian, then for any $H \leqslant$, $N_G(H) = G$, so we get our result trivially. So let's say G was not abelian.

   if $Z(G) \nleqslant H$, then

   $$here$$

   If $Z(G) \leqslant H$, then $H/Z(G)$ is contained in $G/Z(G)$. By proposition ref:HERE, $G/Z(G)$ is nilpoent. Thus, there es a subgroup of $G/Z(G)$ which normalizer $H/Z(G)$ and $H/Z(G)$ is a proper subgroup. Thus, by the correspondence theorem, $H \trianglelefteq N_G(H)$. TBD

2. Let P be a Sylow p-subgroup, and consider $N = N_G(P)$. Since $P \trianglelefteq N$, by proposition 4.4.15, $P \operatorname{char} N$. Since

   $$P \operatorname{char} N \trianglelefteq N_G(N)$$

   Then by transitivity of characteristic subgruops, we have that $P \trianglelefteq N_G(N)$. Then by Sylow's Theorem, $N_G(N) \trianglelefteq N$. By part (1), we get that $N = G$, showing that $P \trianglelefteq G$, that is, Sylow p-subgroups are normal subgroups of G

3. Let $p_1, p_2, ..., p_n$ represent distinct pimres dividing the order of G

`https://en.wikipedia.org/wiki/Nilpotent_group#Properties`

Exercise 6.1.1

1. Prove that any nilpotent group of class 3 or lower is metabelian
2. Show that if $\mathrm{Aut}(G)$ is abelian, then G is nilpotent of class 2

6.1.3 Polycyclic Groups

We've extended a abelian group using *central extensions* and *abelian extensions* (TBD). In this section, we'll once again be extending an abelian group, but this time we'll ask what happens if we extend a cyclic group in such a way that G_i/G_{i+1} is cyclic?

Definition 6.1.16: Polycyclic Group

Let G be a group. Then if

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_n = \{1\}$$

st G_i/G_{i+1} is cyclic, Then G is said to be polycyclic. The series G admits is caleld a *subnormal series with cyclic factors*. If $G_n = \{1\}$, then we say that G is polycyclic with length n.
If $n = 2$, then we say that G is a *metacyclic group*.

Example 6.1.6: polycyclic groups
By the fundamental theroem of finitely generated group, you can prove that all finitely generated abelian group are in fact polycyclic (exercice).

word on saying *the* subnormal series with cyclic factors? (up to isomorphism?)

Proposition 6.1.17: Polycyclic Then Finitely Generated

Let G be a polycyclic group. Then G is finitely generated.

Proof **:**
Let G be polycyclic, so that

$$G = G_0 \trianglerighteq G_1 \trianglerighteq \cdots \trianglerighteq G_n = \{1\}$$

is the subnormal series with cyclic factors. Since G_0/G_1 is cyclic, let' say $[g_0] \in G_0/G_1$ generates G_0/G_1. Similarly, let's say $[g_i] \in G_i/G_{i+1}$ such that $\langle g_i \rangle = G_i/G_{i+1}$. Eventually we'll get that $\langle g_{n-1} \rangle = G_{n-1}/G_n \cong G_{n-1}$, meaning G_{n-1} is cyclic. Then we get that

$$1 \to G_{n-1} \to G_{n-2} \to G_{n-2}/G_{n-1} \to 1$$

i.e., we extend a cyclic group by a cyclic group. Notice if we take any $x \in G_{n-2}$ then we get $[x] \in G_{n-2}/G_{n-1}$, so $[x] = [g_{n-2}^{\alpha}] = [g_{n-2}]^{\alpha}$ so

$$x = (g_{n-2}g_{n-1})^{\alpha}$$

and since x was arbitrary, this mean that $\langle g_{n-2}, g_{n-1} \rangle = G_{n-2}$. continuing inductively, we'll get that

$$\langle g_i | 1 \leq i \leq n \rangle = G$$

showing that G is finitely generated.

Note that the converse is false:

Example 6.1.7: Finitely Generated, But Not Polycyclic
Let $G = A_5$, which is a simple non-cyclic group. Since it simple, it's subnormal series is only

$$A_5 \trianglerighteq \{1\}$$

and so, it cannot be polycyclic.

Then it can be checked that if we restrict to finitely generated groups, we get:

cyclic $\Rightarrow$ abelian $\Rightarrow$ nilpotent $\Rightarrow$ Polycyclic $\Rightarrow$ solvable $\Rightarrow$ finitely generated (show A_5 not solvable)

6.2 Non-finitely generated abelian groups

(Prüfer's Theroems!)

6.3 Linear Groups

Perhaps add a section here on linear groups and some more complicated properties because of how important they are to Lie Algebra and Invariant Theory?

7

Generalization of Groups

In this chapter, I will cover common algebraic objects which are generalizations of groups.

7.1 Monoid

Having a unique inverse is one of the more powerful axioms you can add to an algebraic structure. Recall how the uniqueness of a linear solution is bassed off of the fact that the invere exists and is unique (proposition 1.1.9). This property is also called the *cancellation property*. There are many mathematical structures which do not have this ($\mathbb{N}$ would be a group if inverses existed). For this reason, the study of this structure is also of interest to many

Definition 7.1.1: Monoid

A *Monoid* is an ordered pair $(M, *)$ where M is a set and $*$ is a binary operation on M satisfying the following axioms

1. associativity: $(a * b) * c = a * (b * c)$
2. there exists an element e in G, called the *identity* of G such that for every $a \in G$, we have $a * e = e * a = a$ [a]

The Monoid $(M, *)$ is called *abelian* (or commutative) if $a * b = b * a$ for every $a, b \in M$

[a] it can in fact be proven that if $a * e = a$ then $e * a = a$ (or vice-versa); an indispensable exercise worth doing

Example 7.1.1
Naturally, all groups are monoids. I'll give non-group example of monoids:

1. The natural numbers $\mathbb{N}$
2. Given a set A and it's powerset $P(A)$ (set of all substs), then $(P(A, \cup))$ is a monoid with $\emptyset$ as the identity.
3. Given a set A and it's powerset $P(A)$, then $(P(A), \cap)$ is a monoid with A as the identity.
4. The set $\{0, 1\}$ with any boolean operation ($\Rightarrow$, NOR, AND, etc.) is a monoid. More commonly, the set is relabled as $\{\text{False}, \text{True}\}$
5. The set of strings, given an alphabet is a free monoid.
6. Given a set A, the set of all functions $f : A \to A$ is a monoid! This is incredibly powerful, since we can now put an algebraic structure on sets themselves!! Any set has a default algebraic structure! Further wil be true soon: When we study rings, rings will be an abelian group, along with a second operation which will be a monoid!!

   There are far reaching consequences of this when relating algebra to other fields. If you can put a abliean group structure on something, then taking the set of all functions will now produce you a ring! In algebraic topology, this is exact trick is used to be able to compare rings with topologies, which all comes down to
7. Given the collection of all normal subgroups of a group, their multiplications form a monoid structure! functions having a natural monoid structure!

Recall that $\text{Aut}(G)$ is a group, but $\text{End}(G)$ was not since not every element has an inverse (section 1.4)? This then makes $\text{End}(G)$ a monoid

Proposition 7.1.2: G Abelian Implies End(G) Monoid

here

Proof. Let G be an abliean group and consider $\text{End}(G)$.

Associativity holds by

The identity naturally exists, since $1 : G \to G$, $1(g) = g$ is definable and $\forall r \in \text{End}(G)$, $r1 = 1r = r$. □

Example 7.1.2
exercice: Show that if $G = D_6$, then $\text{End}(G)$ is not a group.

There is a lot of intereting theory for monoids which I will simply sum up for now to get to the main reason I wanted to bring them up

1. Recall that I said that groups have the cancellation property because of the unique inverse of every element. It is possible that there are monoids with this property that are not groups. Such monoids are called monoids *with the cancellation property*. The natural numbers is an

example of such a monoid. Monoids with the cancellation property can be extended to form groups (called the Grothendieck construction). If the monoid is finite, then it implies that it's a group

2. Recall that the set of all homomorphisms between a group and itself forms a group. With a monoid, we can weaken the condition of the functions being homomorphisms. Given a set S and every set-function $f : S \to S$ with the operation defined point-wise ($(f + g)(x) = f(x) + g(x)$) is also a monoid.

3. The same way we defined group-actions with groups, we can define monoid-actions with the same axioms. These pop-up in computer-science in the study of automata.

4. In functional programming, monads is a central tool in programming. Monads are monoids! The objects on which they are monoids is a more complicated subject, and will be left for the appendix on category theory [1]

5. free-monoids play a pivitol role in modeling the behavior of machines (ex. Turing machines). (DEVELOP ON THIS)

Though all these reaon are good reasons to look into monoids, I bring them up for their use in Category Theory. More particularly, we can define a monoid to be a category with a single element!

Exercise 7.1.1

1. standard monoid questions
2. In the following exercices, we'll test your understanding of the idea of defining quotient objects by building up to defining a quotient monoid
   a) Recall the difference between an equivalence relation and a congruence relation. What condition do we need in order that $[x][y] = [xy]$ in a monoid? (I saw that the regular one for groups fails, and it works if there's an inverse – investigate this further for it seems to be an interesting question to do!!)

7.2 Groupoid

Recall how at the very begining under definition 1.1.2, I remarked how some authors prefer to say groups are not closed under multiplication? Some authors introduce this convention to be able to later drop this axiom in order to define the more general notion of *groupoid*. A groupoid is a group which is not closed under the binary operation! [2]

[1] For those who are impatient or know some Category theory, a monoid on a category of functors which are derived given by a particular adjoint

[2] Technically, this means it's not a binary operation, because binary operations are functions, which need to be closed, but the spirit of the sentence is correct

Definition 7.2.1: Groupoid

A groupoid is a set G with a unary operation $^{-1} : G \to G$ and a patial function $* : G \times G \rightharpoonup G$ where the following axioms hold:

1. Where $a * b$ and $b * c$ are defined, $(a * b) * c = a * (b * c)$. Conversaly, if $(a * b) * c$ and $a * (b * c)$ is defined, then so are both $a * b$ and $b * c$ and $(a * b) * c = a * (b * c)$
2. $a^{-1} * a$ and $a * a^{-1}$ are always well-defined
3. if $a * b$ is defined, then $a * b * b^{-1} = a$ and $a^{-1} * a * b - b$

Lot's of cool stuff I'll fill in later

Semi-groups? If you introduce wreath product, can be intersting to talk about in conjuction with automata?

8

Category of Group

Throughout the book, we'll be introducing more and more algebraic structures of varying levels of similarity and differences[1]. It would be very convenient to be able to have a way to compare these algebraic structures. In this section, we elaborate more on the concept of *Categories* which were introduced in section 0.4 . When we introduce new algebraic structures, we will show how the algebraic structure also forms a *category*, and by doing so we will open up the ability to compare and contrast all sorts of properties. For example, we saw in example 1.1.2 that if G and H are groups we can define $G \times H$ to be a group under point-wise operations. Will the same type of construction be possible for algebraic structures in coming chapters? For rings? Group-actions? Fields? Vector-space? Will all of these objects have a notion similar to subgroup? Will they all have a notion of homomorphism/isomorphism, or a similar notion to group action? The answer will be yes to some and no to others! In order for these structure to have or not have a property, say for example the ability to do a "direct product", there must be a way of saying "the direct product of groups follows the same principle as the direct product of rings/group-actions/fields/vector spaces/etc." This will be accomplished with the language of category theory.

As mentioned in section 0.4, the core concept of categories is that we want to be able to have a structure that holds *objects* (ex. groups), and a particular way to *relate* these objects (ex. homomorphisms) – see the analogy in the aforementioned section for an analogous point of reference. Different categories can have the same objects but different relations, or the same relations but different objects, or different objects and relations. Commonly, a relation between two objects will be a way to change from one object to another while preserving some of the original structure of the object. These types of relations are called *structure preserving* relations. As an example of relations between objects that preserve some structure, take homomorphism, which relates two groups if there is some common group-structure. The fact that homomorphisms preserve (some or all of) the structure can also be

[1] In this book, there will be 15 algebraic objects that can be considered "key objects" to algebra, while we also work with 6 others non-algebraic objects throughout the book (ex. sets, graphs, ordered sets, etc.)

found in the etymology of the word: the word *homomorphism* come from *homo-* meaning "same" [2] and *morphism* meaning shape. Since homomorphsims preserve (some or all) group structure, they in a sense change a shape from one group to another while keeping some of the "sameness" between the two. Many different mathematical objects have a relation between them that preserve some of it's structure (topologies with topology-preserving functions, graphs and graph-preserving functions, or ordered sets with order-preserving functions to name few), which is why when categories were first introduced, the word *morphism* came to be the name used to describe relations between objects (i.e. if A and B are related in a category, there is a morphism from A to B). The word morphism is a generalization of the word homomorphism to capture the idea of relation's between objects.

I was purposefully being vague in the start by saying that a category has a "particular way to relate objects" instead of just jumping to saying that morphisms preserve some of the structure of the objects. At the beginning of category theory in the 1940s, the types of relations morphisms represented were relations that preserved (some or all of) the structure of the objects. As time went on, more mathematical objects were shown to form categories using morphisms that didn't preserve structure in a way we saw with homomorphisms, and the idea that some of the structure is preserved in some context can be a little of a stretch (see exercise ref:HERE for an example (order on sets)). This is good to know early on to not get the wrong idea of categories, however in the undergraduate level, 99% of the time morphism in categories will be about preserving some/all the structure of a particular object.

So overall, Categories will be collections of objects as well morphisms to relate those objects. With this very high-level idea, let's give some concrete examples before introducing the definition of a category. If we take all the groups and all the possible homomorphism between groups; that will in fact from the category of groups, sometimes denoted **Grp** or **Group**. Another example would be the collection of all sets and all function between sets; that will form the category of sets, sometimes denoted **Set**. We'll soon bring in the definition of a category to prove that it will follow the axioms of a category, for now we will build-up some terminology in order to talk about these collections.

Definition 8.0.1: Category Of Grp and Set

1. Let **Grp** represent the collection of all groups and homomorphism between Groups and **Set** the collection of all sets and functions between sets.
2. Let $\text{Obj}(Grp)$ represent all the objects in the category of **Grp**. Similarly for $\text{Obj}(Set)$.
3. For any $G, H \in \text{Obj}(Grp)$, Let $\text{Hom}_{\mathbf{Grp}}(G, H)$ represent all homomorphisms from G to H. Similarly, if $A, B \in \text{Obj}(Set)$, let $\text{Hom}_{\mathbf{Set}}(A, B)$ represent all function form A to B.

In many ways, properties of these collections was the starting point for the first category theorists definition of a category, so it is worth taking a closer look at the properties. To not clutter the analysis, we will stick with **Grp** to build-up to the definition of a category.

When we relate two group using homomorphsims, we always have a domain and a co-domain:

$$\varphi : G \to H$$

This means that relations are *directional*: a homomorphism from G to H is different from a homomorphism from H to G. In some structure, direction of the relation doesn't matter (for example,

[2] Other words with homo- would be homeostasis (same state), homochromatic (same colour), or homogeneous (same kind/nature, or uniform structure). It's antonym would be hetero- (heterostatis, heterochromatic, heterogenous)

in equivalence relations, $x \sim y$ if and only if $y \sim x$). In categories, we pay careful attention at the direction of the arrow, and so when generalizing, a morphism from an object A to B is not the same from a morphism from an object B to A.

When we work with homomorphism, we showed in proposition 1.4.2 that the composition of two homomorphisms is a homomorphism. Particularly, if have a homomorphism from G to H, and a homomorphism from H to K, then there will always exist a homomorphism from G to K:

$$\begin{tikzcd} G \arrow[r, "f"] \arrow[rr, bend right, "g \circ f"'] & H \arrow[r, "g"] & K \end{tikzcd}$$

So, if there is a way to get from G to K through homomorphism, there must be a way to "get there directly" too. Morphisms will have the same property.

Finally, notice that for every group there always exists a "trivial" homomorphism:

$$1_G : G \to G \qquad 1_G(g) = g$$

that is, the homomorphism takes G to itself. In a sense, what is being said here is that every object trivially relates to itself. Another way of thinking that 1_G is trivial is that when composed with other homomorphisms (with appropriate domain or co-domain), it does not change them. If:

$$\alpha : G \to H, \ \beta : F \to G$$

were homomorphsims, then

$$\alpha \circ 1_G = \alpha \qquad 1_G \circ \beta = \beta$$

We want this same property to exist for categories: That every object in a category has a *trivial morphism* which can compose with morphisms (with appropriate domain or co-domain) and not affect it

You might have noticed that all of these properties are about the morphisms of a category. This is a common philosophy in category theory: What's important isn't the objects, but the how an objects relates to other object! It is arguable that the more mathematics you will encounter, the more it will be clear that it is the way objects can relate to object objects that i the defining feature of many objects. In fact, many algebraic objects that we will see will be defined by how they relate to other objects.

With this, we will finally define a category:

Definition 8.0.2: Category

A ***category*** C consists of[a]

1. A collection of *objects*, denoted $\mathrm{Obj}(C)$
2. A collection of *morphisms*, denoted $\mathrm{Mor}(\mathbf{C})$

where

- Each morphism has a specified ***domain*** and ***codomain*** object. Sometimes, this is written as $f : X \to Y$, where f is the morphism with domain X and codomain Y [b]. For any $A, B \in \mathrm{Obj}(C)$, $\mathrm{Hom}_C(A, B)$ denotes all morphisms form A to B.
- Each object has an ***identity morphism*** $1_X : X \to X$
- For each pair of morphisms f, g with the codomain of f equal to the domain of g (ex. $f \in \mathrm{Hom}\, C(A, B)$, $g \in \mathrm{Hom}_C(B, C)$), there exists a specified ***composite morphism*** gf (or $g \cdot f$)[c] whose domain is equal to the domain of f and whose codomain is equal to the codomain of g, i.e.

$$f : X \to Y \qquad g : Y \to Z \qquad \Rightarrow \qquad gf : X \to Z$$

such that

1. For any $f : X \to Y$, the composites $1_Y f$ and $f 1_X$ are both equal to f (i.e. unital with the identity homomorphism)
2. For any composable triple of morphisms f, g, h, the composites $h(gf)$ and $(hg)f$ are equal (i.e. associative), and so we can write hgf unambiguously.

[a] The word *collection* being used instead of *set* was purposeful; we will get to this

[b] Note that f **need not be a function**! Namely, the morphism need not have a unique codomain

[c] the notation is purposefully not $\circ$ so as to not draw too much of an analogy of gf with the process of composing functions to demonstrate that in some cases gf is *not* function composition

Example 8.0.1: Examples of Categories

1. It was mentioned that groups along with homomorphisms from a category. Indeed, by proposition 1.4.2 the composition of two homomorphsims is a homomorphism, so $gf = g \circ f$. Homomorphsims are associative since function composition is associative (proposition 0.1.12), and for any $G \in \mathrm{Obj}(Grp)$, the map $\mathrm{id} :\to G$, $\mathrm{id}(g) = g$ is the identity morphisms since for any $f : F \to G$ and $h : G \to H$, $f \circ \mathrm{id} = f$ and $\mathrm{id} \circ h = h$.
2. If we take all groups, and restrict to taking only the isomorphisms, we will again have a category; check that all the axioms are still satisfied. Such a category is called a *groupoid* (the reason for this name will be made evident in exercise ref:HERE (show with 1 element is group, then define more general structure groupoid with no identity, which we expand more upon in section 7.2))
3. If we take all abelian groups and use homomorphsims, this will again be a category. This category is significant enough to deserve it's own notation: this category will denoted $\mathbf{Ab}$.

4. If we take all sets and all functions between sets, we will get the category of **Set**.

   Notice that every morphism in **Grp** can be thought of as a morphism in **Set** if we "forget" the group structure of the domain and co-domain, and the homomorphism condition of the morphism of **Grp**. In This way, the category of **Grp** and the category of **Set** are related if we "forget" some of the additional structure of **Grp**. Is it possible to have the reverse? That is, starting with **Set**, is it possible to somehow "upgrade" it to **Grp**? Is there more than one way of doing this? Is there a "natural" way of doing it, that is, is there a way of performing this upgrade in such a way that it follows from the properties of groups and sets without needing to add any more conditions? These questions will be further explored in section 2.3.3. generally.

5. If we take the collection partially ordered sets, that the collection of $(S, \leq)$ where $S \in \text{Obj}(Set)$ and $\leq$ is a possible partial order on S, then we can define a notion similar to homomorphisms that preserve (some or all of) the structure of the order, i.e. we can define a $f : (S, \leq_S) \to (R, \leq_R)$ where f is the appropriate morphism. Can you think of what this morphism should be?

6. In section 0.3, we introduced graphs and graph isomorphisms. These also form a category! There is also a notion of "graph homomorphisms" mapping vertices to adjacent vertices and a generalization of the notion of *graph colouring*. For those interested in this topic, see ref:HERE.

Before continuing with the language of category, we will add one more example of a category which is a little more abstract than the previous categories but one we will constantly use. Namely, we can take categories themselves to be objects and define a notion of morphism between categories! This category will be denoted **Cat**. Since morphisms relate objects, the morphisms in the category of **Cat** allows us to find commonalities between different categories! We have in fact already described a category morphism in example 8.0.1[4]: If we take **Grp**, we can forget the group structure of every group and the homomorphism property of every homomorphsims and we will end up with **Set**. In effect, we did $(G, *) \to G$ and mapped the homomorphism f where $f : (G, *_G) \to (H, *_H)$ to the function f $f : G \to H$ where the function f is identical to the homomorphsims f except it no longer requires the homomorphism property (in fact, it is impossible for it to have the homomorphism property since sets don't have a binary operation). We will call category morphisms *Functors* (as a generalization of the word function)

Definition 8.0.3: [Covariant] Functor

Let C, D be two categories. Then a *Functor* F will consist of:

1. a function between $\text{Obj}(C)$ and $\text{Obj}(D)$. If $a \in C$ is an object, then $F(a) \in D$ denotes the object in D.
2. a function between $\text{Mor}(\mathbf{C})$ and $\text{Mor}(\mathbf{D})$. If $f \in C$ is a morphism in C, then $F(f)$ denotes the morphism in D.

such that

1. F must preserves the domain and codomain of each morphism: $f : A \to B$ must map to $F(f) : F(A) \to F(B)$
2. F must map identities to identities: $F(\text{id}_f) = \text{id}_{F(f)}$
3. F must preserve composition: $F(g \cdot_C f) = F(g) \cdot_D F(f)$.

Example 8.0.2: Functor

1. For any category C, there is always a trivial functor sending objects to themselves and functions to themselves
2. The relation above between **Group** and **Set** is indeed a functor: $F : \textbf{Group} \to \textbf{Set}$ Such a functor is called a *forgetful functor* since it forgets some of the structure of the category. In the coming chapters, many of the objects are augmentation of sets. All of these categories will have a forgetful functor $F : \mathbf{C} \to \textbf{Set}$.

Some authors prefer to be more specific about the second condition and define it instead as the following: for any two object $a, b \in \text{Obj}(C)$ there exists a function $\text{Hom}_C(a, b) \to \text{Hom}_D\, F(a), F(b)$. We will see many more examples of functors throughout this book. For now, it is left as an optional exercise to prove that categories and functors together form a category[3].

So far, all the categories we have dealt were all in some fundamental sense based on sets: each object was either a set with an augmentation (ex. $(G, *)$ or $(S, \leq)$) or a collection of sets following certain restrictions (ex. graphs). Because of this, every object in every category so far has the notion of having "elements". However, there is nothing in the definition of an object in a category that restricts objects to needing elements:

Example 8.0.3: Non-Set Category

Let $\mathbb{N}$ be the set of all natural numbers and $\leq$ be the usual order. Then we define a category C whose objects are the numbers of $\mathbb{N}$ and whose morphism is $\leq$, that is for every $x \in \mathbb{N}$, $\leq$ is the morphism that relates x to numbers greater than or equal to number y, which we will label $x \leq y$. Notice that $\leq$ is not a function, since x is not "mapped" to a single element but to every element greater than it. Furthermore, $\leq$ is the only morphism in in this category. Composition will return $\leq$ (since it's the only morphism) and we see that if $x \leq y$ and $y \leq z$ then their composition results

[3]There is a technicality about the fact we are taking "the category of all categories" in the sense that this is self-referential creating a contradiction usually called Russel's Paradox. There is a definition of *small category* which circumnavigates this problem which was not introduced so as to no overload new terminology

in $x \leq z$.

This is indeed a category: $\leq$ has a domain and co-domain (if you like you can picture $\leq: x \to y$ as long as it is kept in mind that $\leq$ is not a function). For any $x \in \mathbb{N}$, x has an identity morphism, namely the only morphism $\leq$, (you can picture $\leq \cdot \leq = \leq$).

More generally, any set S with some partial order forms a category. For example, if X is a set, then $(P(X), \subseteq)$ is a category whose objects are all subsets of X and the morphism is the partial order induced by $\subseteq$ (i.e. $U \subseteq V$ then $U \to V$).

(Another would be to simply draw diagrams!!)

The set $\mathbb{N}$ was an arbitrary choice in the previous construction: we could have chosen $\mathbb{R}$ or $\{0\}$ or an arbitrary set with an arbitrary ordering. We can visualize the category in the example by imagining a directed graph where every edge represents the morphism between the domain and codomain:

$$\begin{tikzcd} 0 \arrow[r] \arrow[rrr, bend right] & 1 \arrow[r] \arrow[rr, bend left] & 2 \arrow[r] & 3 & \cdots \end{tikzcd}$$

Such a directed graph is called a *commutative diagram*. The full reason behind the name is a little abstract without the proper build-up[4], however we can still understand the etymology given the following illustrative example: imagine we start at the number 1, and every horizontal arrow represents multiplying on the right by x, and every vertical arrow represents multiplying on the right by y, and we form a square like so:

$$\begin{tikzcd} 1 \arrow[r, "\cdot x"] \arrow[d, "\cdot y"] & x \arrow[d, "\cdot y"] \\ y \arrow[r, "\cdot x"] & yx = xy \end{tikzcd}$$

This square shape is so common that the any directed graph represented morphisms came to be called a commutative diagram. In section 8.2, we shall further explore the use of commutative diagram to understand properties of categories.

8.1 Basic properties of Morphisms

I clearly don't have enough knowledge to write this yet:

`https://en.wikipedia.org/wiki/Monomorphism`.

Read this and come back.

When we studied function, we started by classifying injective and surjective functions due to their simplicity and universality. It would be nice if there are some nice equivalent, or at least similar, notions of injectivity and subjectivity for morphisms in general categories. Since objects need not have elements, the definition has to take this generality into account. Fortunately, there is already a property of injective and surjective functions that was characterized in terms of functions in

[4]We will often care about morphisms between categories commuting in some "natural" way. This will lead to the study of natural transformation from which the name commutative diagram will takes it's name

proposition 0.1.15 and 0.1.17: it was shown that all injective functions have a left inverse and all surjective functions have a right-inverse. These notions can be generalized in the following way:

Definition 8.1.1: Monomorphism

Let $f : A \to B$ be a morphism. Then if for any morphism $\alpha : Z \to A$, $\beta : Z \to A$, if

$$f \circ \alpha = f \circ \beta \Rightarrow \alpha = \beta$$

then we call f a monomorphism

Intuitively, monomorphisms should take you to a "unique place" similarly to how injective functions have a single element to represent them.

Proposition 8.1.2: In Set, monomorphism if and only if Injective

Let $f : A \to B$ be a set function. Then f is injective if and only if it is a monomorphism

***Proof* :**

($\Rightarrow$) Let's say f is injective, and that $f \circ \alpha = f \circ \beta$. We want to show that $\alpha = \beta$. By proposition 0.1.15, there exists a left-inverse f^{-1}. Now, notice that

$$\begin{aligned}
(f^{-1} \circ f) \circ \alpha &= f^{-1} \circ (f \circ \alpha) && \text{associativity of composition} \\
&= f^{-1} \circ (f \circ \beta) && \text{by assumption} \\
&= (f^{-1} \circ f) \circ \beta && \text{associativity of composition}
\end{aligned}$$

meaning we get $\mathrm{id}_A \circ \alpha = \mathrm{id}_A \circ \beta$, and so

$$\alpha = \beta$$

as we sought to show

($\Leftarrow$) Let's say f is a monomorphism so that $f \circ \alpha = f \circ \beta$ implies $\alpha = \beta$. Let's say f wasn't injective, that is, for some $a, b \in A$, $f(a) = f(b)$ but $a \neq b$. Choose $B' = f(B)$ to be the domain of α and β where α and β both agree where to send elements with the exception of α sending $f(a)$ to a and β sending $f(a)$ to b. Then

$$f \circ \alpha = f \circ \beta$$

but $\alpha \neq \beta$ a contradiction. Since those elements were arbitrary, it must be the case that,

$$f(a) = f(b) \Rightarrow a = b$$

showing injectivity.

One can also prove this directly by letting $Z = \{p\}$ (i.e., a singleton) and using the monomorphism property to conclude that

$$f(a) = f(b) \Rightarrow a = b$$

Similarly, we can define an epimorphism:

Definition 8.1.3: Epimorphism

Let $f : A \to B$ be a morphism.

If for any function $\alpha : Z \to B$, $\beta : Z \to B$, if

$$\alpha \circ f = \beta \circ f \Rightarrow \alpha = \beta$$

then we call f a epimorphism

Proposition 8.1.4: In Set, epimorphism if and only if Surjective

Let $f : A \to B$ be a set function. Then f is surjective if and only if it is a epimorphism

Proof **:**
Exercise

We have mentioned earlier that in category theory, we define objects based on how they relate to other objects. We have in fact done this before: When we defined a subgroup based on a group. We mentioned how this more complicated definition generalizes we

Definition 8.1.5: Subobject

Let C be a category and $X \in C$ be an object in this category. We say $Y \in C$ is a *suboject* of X is there exists a monomorphism

$$Y \to X$$

(maybe details with the "up to isomorphism" thing)

To demonstrate what I mean, Let's take an arbitrary set X, and an arbitrary set Y. When can I say that Y is a subset of X, or vice versa? You probably guessed it: if one can be part of included in another, one is a subset of the other. Written in terms of functions: if there exists an inclusion map $i : Y \to X$, then Y is a subset of X ($Y \subseteq X$). We then can define a set being a subset as follows:

If there exists an inclusion map $i : X \to Y$, then X is a subset of Y ($Y \subseteq X$).

If both inclusion maps exist, then there is two injections, and so by Schröder–Bernstein theorem (see appendix A), there exists a bijection between X and Y ($X = Y$). In this way, we can actually *define* a set to be a *subset* of another set if there exists an inclusion map from Y *to* the bigger set. If we weaken our need for i to be an inclusion map, to i being an injective map, we can start comparing sets with different elements [5]

We can apply this now to groups! First, recall (or revisit) what we did in proposition 1.4.5. Since we have proven the proposition is equivalent to the definition of subgroup, we will bring it here, but label it as a second definition of subgroup:

[5] More philosophically, it doesn't actually matter what are the elements of a set; what matters is that there *are* arbitrary distinct elements in a set. So being an inclusion or injective is actually "equivalent" in a broader sense. How so? This question will be answered in appendix 30 with more build-up to the greater philosophy of mathematics

Since the same technic is used in defining both concepts, we can say each are similar concepts or types of "objects". Thus, we can say what we have here is applied the more general idea of *sub-object* to both situations.

Notice how for sets, if there are is an injection going both way, we say that the two sets are bijective and $X = Y$? This is actually *exactly* the same with groups!

Example 8.1.1: sub-objects

1. **subobjects in Grp**: Recall in proposition 1.4.5 where we showed that if $(H, \bullet)$ is a group and $H \subseteq G$, then H is a subgroup if and only if the inclusion map $i : H \to G$ is a homomorphism. This proposition was the proof that **Grp** had subojects!
2. **subobjects in Set**:
3. sub-order?

(maybe mention the question of switching the monomorphism for an epimorphism and how that let's you define something we'll call *quotient objects*, which we'll get too soon?)

(build-up to the idea of isomorphism in category theory)

Definition 8.1.6: Isomorphism In C

Let C be a category, and f be a morphism. Then f is said to be an *isomorphism* if HERE.

Example 8.1.2: isomorphism in different Categories

1. In **Grp**, HERE
2. In **Set**, HERE
3. In the category of ordered sets, HERE

Proposition 8.1.7: Isomorphism Generalized

Let G, H be groups. Then if there exists an injective homomorphism $G \to H$ *and* and injective homomorphism $H \to G$, then we say that G and H are *isomorphic* and we write $G \cong H$.

Notice how I didn't say that $G = H$. That is because I said we have an *injective* homomorphism. I did not say that it's necessarily an *inclusion* map. If it were, we can justify saying $G = H$ completely. However, group theoretically, there is no difference between G and H, making them "equivalent". In this way $G \sim H$ in terms of equality $=$. So, combining the symbols $\sim$ and $=$ gets us $G \cong H$.

The great advantage of defining subgroup/subset in this way is that we can now define what it means to be a sub-object in many other areas of mathematics! We can even ask whether it is *possible* to define the notion of sub-object; there are some areas in mathematics where the concept of an injective map cannot be defined!![6]

[6] The category of posets $(P, \leq)$ is an example. See appendix 30

Also, perhaps define Hom, Aut, and End here? Perhaps Also give the category of **Set**? To be able to give an idea of what a category is and to bet able to define $\mathrm{Aut}_{\mathbf{Set}}(A)$ for group actions. Introduce the notation

$$\mathrm{Sym}(X) := \mathrm{Aut}_{\mathbf{Set}}(X)$$

to use it later?

Show how for any category C, and any object $X \in C$, then $\mathrm{Aut}_C(X)$ forms a group! In this way, groups appear everywhere. The fact that $\mathrm{Aut}_C(G)$ is a group will come back big-time when studying *group actions*.

Next talk about initial and final objects. This leads to universal properties. Bring up how $\{e\}$ was an initial and final object in **Group**. Bring up how initial and final objects are unique, so once we proved on exists, we proved it's the only one up to isomorphism.

8.2 Commutative diagrams

As mentioned countless times, categories try to capture the idea of relations between objects. Very often, this is done by forcing some relation between morphisms. We've already seen an example of this in the category of **Set** and **Grp** by saying the composition of two functions (or homomorphisms) is again a function (or homomorphism). We represented this by the diagram:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[rr, bend right, "h"'] & Y \arrow[r, "g"] & Z \end{tikzcd}$$

where we usually denote h as $g \circ f$. We can form all sorts of relations between objects and enforce the fact that compositions of functions must lead to other arrow (TBD). We call this a commutative diagram

(category can be projected down to a graph. I you only project (PROPER PROPERTY HERE), then you get a special graph called a *quiver*, which try to be a generalization of algebraic structures (TBD).)

Definition 8.2.1: Commutative Diagram

Give a nice definition here

Example 8.2.1: Example
Must re-phrase

1. Recall that a group homomorphsims is a function $\varphi : G \to H$ which preserves the group action. If we take G, and H represent the underlying set, and $(G, \cdot_G)$, $(H, \cdot_H)$ represent the set with the underlying operation, then we can what's the most *natural* requirement for the operation to be preserved? Note that if we drop the operation preserving part of φ, we can think of the inputs of the binary operation as ^{a}

$$(\varphi \times \varphi) : G \times G \to H \times H \qquad (\varphi \times \varphi)(a, b) = (\varphi(a), \varphi(b))$$

Then, we can combine this with the operations on the group to get the following diagram:

$$\begin{tikzcd} G \times G \arrow[r, "\varphi\times\varphi"] \arrow[d, "\cdot_G"] & H \times H \arrow[d, "\cdot_H"] \\ G \arrow[r, "\varphi"] & H \end{tikzcd}$$

Thus, from a categorical perspective, the best way to define an operation to be preserving is if this diagram *commutes*, meaning:

$$\begin{tikzcd} (a,b) \arrow[r, dotted, no head] \arrow[d, mapsto] & {} \arrow[d, dotted, no head] \\ a \cdot b \arrow[r, mapsto] & \varphi(a \cdot b) \end{tikzcd} \qquad \begin{tikzcd} (a,b) \arrow[r, mapsto] \arrow[d, dotted, no head] & (\varphi(a), \varphi(b)) \arrow[d, mapsto] \\ {} \arrow[r, dotted, no head] & \varphi(a) \cdot \varphi(b) \end{tikzcd}$$

which in it's familiar form would be $\varphi(a \cdot b) = \varphi(a)\varphi(b)$

2. maybe bring up how 1.9 is not necessarily a commutative diagram since the arrow's aren't necessarily restricted through composition.

[a]The reason for this construction, as we shall soon see, relies on the *universal property of products*. For now, since it's something all readers would know at this point, we leave it as intuitive

8.3 More abstract examples

(Give an example of a slice category or comma category, and a fibered category. Explain how the point of these categories and many like them is to capture certain relations between objects.

8.4 Summary

In this section, we defined a category, showed a couple of examples, a couple of more abstract categories, how to think relating objects using commutative diagrams, and generalized the idea of subobject and isomorphism to categories. Throughout all of group theory, many more concepts will be introduced. Many of these structure can seen in other algebraic structures we will encounter. From here on out, there will be an extra section in each chapter dedicated to generalizing concepts to see the deeper idea behind them.

Part II

Rings

Now that we've the essentials covered groups, we can move on to structures which have 2 binary operations on the set. These structures are called *rings* [7] As with commutative groups, commutative rings will be very different from rings, so much so that one can think of ring theory as really being 2 seperate domains put under one label: commutative rings and non-commutative rings. However, similary to how all groups share some common structure (a notion of homomorphism, isomorphism theorem, direct product, free group, group-action, etc.), so do rings, and so we will be building on the foundation of ring theory.

We may wonder how to interpret adding a second binary operation to an abelian group? Does this restrict our structure? Does this limit what we can prove? Does it give more tools to work with when proving? The way I like to percieve adding this second operation is that there are more properties that we have at our disposal to work with. The "type" of new property we add also determines what properties we work with (for example, as is briefly addressed in appendix C, section C.4, we can add a topological structure to groups which gives us different tools we can use to study groups). In the case of adding a new binary operations, we gain more "algebraic" tools to work with. For example, we have an understanding of how the prime numbers of $\mathbb{N}$ interact when we add them[8] or when we multiply them, but we have little knowledge of the interaction of prime numbers between these two binary operations. Furthermore, for any abelian group, we can always add at least one new binary operation that satisfies the axioms of a ring[9] (which we will present near the end of section 9.1.1). In that way, all groups can be made into rings, and we can study them as rings.

For the 1st operation, we will always use additive notation. For example, if we're taking the internal direct product of two groups in this ntoation, we will write $H + K$, since we will reserve the multiplicative notation for the 2nd binary operation.

[7] The term *ring* was probably a contraction of David Hilbert's term *Zahlring*, literally translated as *number ring*. The word ring was probably coined because of the cyclic nature of the numbers Hilbert was working with. More details on this discussion can be found here: `https://math.stackexchange.com/questions/61497/why-are-rings-called-rings`

[8] All be it to a very limited degree

[9] Similarly to how all abelian groups are topological groups with the discrete topology

9

Introduction To Rings

With Groups extensively covered, what's "basic" has now changed. We need not so extensively elaborate intuitions for homomorphisms, quotients, products, or actions. Instead, we'll show how these concept grow when considering *rings*. Fundamentally, a ring is a set with two operators (hence, is a triple $(R, +, \cdot)$ where $+$ and $\cdot$ catisfy certain axioms[1]). We will show some really interesting examples which we gain when we introduce a second operator, and take advantage of some of the categorical language we have built-up to shed some more light on the nature of rings (which will be noticably different than that of groups).

As with groups homomorphisms, "ring homomorphisms" will play a central role in the study of rings. There will be an equivalent to "normal groups" which we will extensively study: the equivalent for rings is called *ideals*. Manipulating and combining ideals is at the core of ring theory. After exploring ideals, we will take a moment to look at a way to embedd any ring R into a larger ring that permits the notion of "division" (as we'll see, unlike groups, not all elements of a ring will have an inverse with the second operation). Through this construction, we will show how to take $\mathbb{Z}$ and create $\mathbb{Q}$.

[1]These axioms can be said to be "natural" in some way. These exploration's will be done later in the chapter, see ref:HERE

9.1 Basic Properties and Definitions

Definition 9.1.1: Ring

A *ring* R is a set together with two binary operations $+$ and $\times$ (called addition and multiplication), $(R, +, \times)$ satisfying the following axioms:

1. $(R, +)$ is an *abelian* group, where the identity is denoted by 0
2. **(Mult. Associativity)** $(a \times b) \times c = a \times (b \times c)$ for all $a, b, c \in R$
3. **(Mult. Identity)** There exists an $e_\times \in R$ such that for every $r \in R$, $e_\times \times r = r \times e_\times = r$. The element $e_\times$ is usually labeled 1.
   - Sometimes, we use *rng* to denote a ring without multiplicative identity.
4. **(Distributivity)** For all $a, b, c \in R$

$$a \times (b + c) = (a \times b) + (a \times c) \qquad \text{and} \qquad (a + b) \times c = (a \times c) + (b \times c)$$

   The two equations seperately are caleld left and right distribuitivity respectively.

If $a \times b = b \times a$ for all $a, b \in R$, we call the ring a *commutative ring*.

We'll usually write $a \times b$ as $a \cdot b$ or ab for convinence. The convention that the multiplicative identity is 1 and additive identity is 0 is essentially universal, and so it is safe enough to have it as part of the definition of a ring.

20: Remark

Some author's do not include 1 in the definition of a ring (we called rings without identity *rngs*); if that is a case, a ring with a multiplicative identity is usually called a unital ring. Any *rng* R can be extended to a ring by taking $\mathbb{Z} \oplus R$ with identity $(1, 0)$ (What is the multiplication here that makes $\mathbb{Z} \oplus \mathbb{R}$ into a ring?)[a], at which point we can work with $\mathbb{Z} \oplus R$ to take advantage of the properties of a ring with idenitty (of which there are many advantages). The most common place in mathematics where mathematicians usually don't want rings to have identities is in analysis; for those inclined for examples, see exercises ref:HERE-ref:HERE. In Algebra, it is now pretty rare to find a textbook that define's a ring without a multiplicative identity.

[a] More generally, taking the direct sum with any [unital] ring will do, however different choices might lead to better result in different fields, for those interested see exercise ref:HERE

Just like associativity was a powerful axiom, so is distributivity for linking the $+$ operation and the $\times$ operation. If you have any arbitary expression:

$$(a + b)(c + d + e(f + g(i + j)kl(m + n)op))q$$

It can be simplified to the sum of monomials:

$$ac + bc + ad + bd + aefq + befq + aegiklmopq + begiklmopq + aegjklmopq$$
$$+begjklmopq + aegiklnopq + begiklnopq + aegjklnopq + begjklnopqq$$

There are non-distributive algebraic structures. However, the study of non-distributive structures seems to be much scarser than that of non-associative structures or structures without inverse. [2]. Distribuitivity seems to be how most structure link the two operations; at the minimum it usually would be distributive on only one side [3].

Note: R being an abelian group is unavoidable given the distributive property

$$a + b + a + b = (1 + 1)(a + b) = a + a + b + b \Rightarrow a + b = b + a$$

Another way to think about G being abelian is that we want to concentrate on the *ring* operation now, so if we can make the group-operation very-well behaved (as we've shown commutativity does in chapter 5), it will not obtrude our study of the ring operation. This is a common technic.

Example 9.1.1: Ring Examples

1. Taking any abelian group G, and define $a \cdot b = 0$. This is the *trivial ring*. If If $G = \{0\}$, then we call this the *zero ring*. In fact, all $ab = 0$ for all $a, b \in R$ if and only if R is the zero ring:

   Proof. If R is the zero ring, then we already have that $0 \cdot 0 = 0$. Thus, we must prove the converse

   Let's say $\forall a, b \in R$, $ab = 0$. Then $1 \cdot 1 = 0$, meaning $1 = 0$

   Then, notice that

   $$\begin{aligned} &\Leftrightarrow\ 0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a \\ &\Leftrightarrow\ 0 \cdot a = 0 \cdot a + 0 \cdot a \\ \Leftrightarrow\ 0 \cdot a - 0 \cdot a &= 0 \cdot a + 0 \cdot a - 0 \cdot a \qquad \text{additive inverse} \\ &\Leftrightarrow 0 = 0 \cdot a \end{aligned}$$

   Thus, we have $0 = 0 \cdot a = 1 \cdot a = a$ for all a, and so all element equal 0, showing it's a zero ring. □

   In fact, this proof has a very powerful corollary: if $1 = 0$, then we have the zero ring! This property usually breaks many parts of many proofs, so we usually say that $1 \neq 0$ in most theorems and propositions. In some textbooks, this is even part of the axioms of a ring

2. The integers $\mathbb{Z}$ is the canonical example of a ring; when ring theory started, a lot of study was with trying to characterize properties of integers. Note that $ab = ba$, making $\mathbb{Z}$ a *commutative ring*.

3. $\mathbb{Z}/n\mathbb{Z}$ is a [commutative] ring with identity

4. $\mathbb{R}, \mathbb{Q}, \mathbb{C}$ are all [commutative] rings with identity. Naturally, $\mathbb{N}$ is not [a]

5. If R and S are rings, then $R \times S$ is also a ring with operation component wise. [b]. What is the identity of $R \times S$?

[2] I found this paper: generalized De Morgan algebra
[3] Something called a *Quasifields* is an example

6. If you have taken Linear Algebra, then all fields are an example of [commutative] rings

7. Recall the group Q_8 (see 1.2.4). Using it, we will define the set $\mathbb{H} = \{a + bi + cj + dk \mid a, b, c, d \in \mathbb{R}\}$ where the elements bi, cj, an dk are distinct elements of $\mathbb{H}$, and are not another element in $\mathbb{R}$ or Q_8 (think of a, b, c, d as coefficients). Then we can make $\mathbb{H}$ into a ring by defining addition to be ponitwise:

$$(a + bi + cj + dk) + (a' + b'i + c'j + d'k) = (a + a') + (b + b')i + (c + c')j + (d + d')k$$

and multiplication through repeated use of the distributive property, which distributes to:

$$\begin{aligned}(a + bi + cj + dk)(a' + b'i + c'j + d'k) = &(aa' - bb' - cc' - dd') \\ &+ (ab' + ba' + cd' - cd')i \\ &+ (ac' - bd' + ca' + db')j \\ &+ (ad' + bc' - cb' + da')k\end{aligned}$$

Then $\mathbb{H}$ is called the *(real) Hamiltonians* (real since we could have defined them over $\mathbb{C}$). They form a non-commutative ring, and can be considered an extension of the complex numbers $\mathbb{C}$ (there are in fact two copies of $\mathbb{C}$ inside $\mathbb{H}$, more in that in exercise ref:HERE)

8. If X is a set and A is a ring, then the colleciton of all function $f : X \to A$ are a ring with function addition and multiplication being done point-wise: $(f + g)(x) = f(x) + g(x)$, $(fg)(x) = f(x)g(x)$. All ring structure on this function space comes from A. A concrete example of this is if $X = [0, 1]$, $A = \mathbb{R}$, and f is continuous $f : [0, 1] \to \mathbb{R}$. These types of rings show up a lot in analysis with the colleciton of function from a space X (a topological space, or measurable space, a manifold, etc.) to $\mathbb{R}$ or $\mathbb{C}$ giving many properties of the domain of the functions.

9. The set $2\mathbb{Z}$ is not a ring since it does not have an identity. It is thus a *rng*.

10. From section 1.2.5, we saw that $(M_n(G), \cdot)$ is not necessarily a group since inverses were not not always defined. However, this is not a requirements for ring multiplication! It is very tedious and a bit hard at this stage to check that associativity and distributivity hold (there will be a simple proof of this in corrollary 13.2.5), but they are verifiable and show that $(M_n(R), +, \cdot)$ is a ring. Note that $M_n(R)$ is not commutative. For example

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

Matrix rings will be explored in more detail soon, and will have a whole chapter (in particular chapter 24) dedicated to the study of when matrix rings "act" on finite groups (in a similar way groups acted on sets in group-actions).

11. Recall when I asked if $P(X)$ is a group under $\cup$? Here is another similar question with rings. Take $P(X)$ again, and define addition and multiplication as

$$A + B = (A - B) \cup (B - A) \qquad A \cdot B = A \cap B$$

prove that it's a [commutative ring]. Another thing you can show is that for any $Y \in P(X)$, $Y^2 = Y$. If a ring has this property, it is said to be a *boolean ring.* [c]

12. Let G be any abelian group. Then $(\mathrm{End}_{\mathbf{Grp}}(G), +, \circ)$ is a ring. In this way, any ring is formed from an abelian group. [d]

[a] $\mathbb{N}$ is a semi-ring with $+$ and $\cdot$

[b] Though this might feel "trivially true", the fact that taking the direct product of two rings will produce the same algebraic structure, in this case a ring, is not always true! See 9.1.8 for a counter-example

[c] This ring comes up in *measure theory*, where the set of measurable sets can be made into a boolean ring

[d] The general term for taking [abelian] objects (ex. abelian groups) from an category C and working with $\mathrm{End}(X)$ as the new object and morphisms that also preserve $\circ$ is **Ab**-enriched category given the abelian category C

When working with rings, we will assume the usual ring structure for $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{Z}/n\mathbb{Z}$, and use the same notation if it's clear in context. The notation $\mathbb{Z}^\times, \mathbb{Q}^\times, ...$ will soon be repurposed for another use later in this section.

Note that every ring is naturally an abelian group as well as a monoid (see section 7.1). When working with homomorphisms for rings, both structure need to be preserved in some way for the image of a funciton to be a ring. There is in fact one more sneaky criterion that needs to be added in order for a the image of a function be be a ring, and hence a *ring homomorphism*:

Definition 9.1.2: Ring Homomorphism

Let R and S be rings. Then:

1. *ring homomorphism* is a map $\psi : R \to S$ satisfying
   (a) $\varphi(a + b) = \varphi(a) + \varphi(b)$
   (b) $\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$
   (c) $\varphi(1_R) = 1_S$
2. If $\varphi : R \to S$ is a ring homomorphism, and there exists a ring homomorphism $\psi : S \to R$, then φ is a ring isomorphism

Note that just like for group homomorphisms, it can be proven that definition 9.1.2[2] is equivalent to a ring homomorphism being bijective.

Note that the reason we add the condition that $\varphi(1_R) = \varphi(1_S)$ is so that the image $\varphi(R)$ is also a ring. If a ring is not defined to have a multiplicative identity (as is done for *rngs*), then the condition of $\varphi(1_R) = \varphi(1_S)$, is not necessary. Furthermore, where $\psi(e_G) = \psi(e_H)$ is a provable property given that ψ is a group-homomorphism and e_G, e_H are the respective identities of the domain and codomain, the same is not true for rings, that is, the first 2 conditions of a ring homomorphism can be satisfied, however the resulting image will not be a ring (take $f : \mathbb{Z} \to \mathbb{Z}$, $x \mapsto 2x$, see exercise ref:HERE).

Many times, mathematical objects will have multiples different structures defined on them, but we sometimes want to only work with some of them[4]. Thus, a group homomorphism between rings will only preseve group structure, a while ring homomorphism preserves both. Ring homomorphisms will be further explored in section 9.3, where we will show how to relate the structure of $\varphi(R) \subseteq S$ to R similarly to how it was done with quotient groups.

From definition 9.1.2, we can define a sub-ring:

[4] An example with full of different structures is $\mathbb{R}$ which can come with a group, ring, and order and structure, but also structures that we haven't covered, some of which are non-algebraic (ex. completeness (see uniformly continuous functions and uniform spaces), or topological)

Definition 9.1.3: Sub-Ring

Let R be a ring and $S \subseteq R$ a subset. Then S is a *subring* of R if it is still a ring under the same operations as R. In particular, a subring of a ring R is a subset $S \subseteq R$ with a ring operation such that the inclusion map $i : S \to R$ is a ring homomorphism [a].

[a] See proposition 1.4.5

Corollary 9.1.4: sub-ring condition

a *subring* of the ring R is a subgroup of R that is closed under multiplication

Proof **:**
see exercise ref:HERE

It is simpler to prove a subset is a subring as compared to subgroups: you need to show it' an abelian sub-group that is still closed under multiplication. Notice that there is no special symbol for a subring like there as for subgroup. This will be a recurring pattern as we continue to explore algebra: there is no special symbol for "subobjects". We will always say "Let S be a subring of R" or "Let $S \subseteq R$ be a subring of R". Another consideration to keep in mind is that the identity of a subring need not be the same (since an inclusion homomorphism $S \hookrightarrow R$ need not map the identity to itslef): If $R \times S$ is a ring, $R \times \{0\}$ is a subring. What is it's [multiplicative] identity?

Proposition 9.1.5: Properties of Rings

Let R be a ring. Then:

1. $0a = a0 = 0$ for all $a \in R$
2. $(-a)b = a(-b) = -(ab)$ for all $a, b \in R$ (recall $-a$ is the additive inverse)
3. $(-a)(-b) = ab$ for all $a, b \in R$
4. The multiplicative identity is unique and $-a = (-1)a$.
5. Generalized Distribuitivity: Distributivity can be extended inductively.

Proof **:**
All of these proofs take advantage of the abelian structure of R and distributivity. For example:

1. Using the fact that 0 is an additive identity and distribuitivity, we get:

$$0 \cdot a = (0 + 0)a = 0 \cdot a + 0 \cdot a$$

Since $0 \cdot a \in R$, there exists an additive inverse, thus

$$0 \cdot a - 0 \cdot a = 0 \cdot a + 0 \cdot a - 0 \cdot a$$

implying $0 = 0 \cdot a$. A similar argument can be made for $0 = a \cdot 0$

the other 3 proofs continue in a similar fashion.

:exercise Show that $(-1)^2 = 1$.

Consequences of no Multiplicative Inverse

Now that we know what is needed to be a ring, let's put names to common rings with properties added on. For each of these properties, I would recommend to go back and see which rings we covered have these properties. If you want to verify your intuition, you can check glossary G.2.

The first thing is that the multipication operation is missing one of the axioms of a group: elements need not have a multiplicative inverse. If every element has a multiplicative inverse, we will give a name to such a ring:

Definition 9.1.6: Division Ring

A ring R with identity 1, where $1 \neq 0$ is called a *division ring* (or a *skew field*) if every non-zero element $a \in R$ has a multiplicative inverse, $b \in R$ such that $ab = 1$. It can be proven this implies that $ba = 1$

If furthermore, multipication is commutative, the resulting ring is a commutative division ring, whcih is called *field*:

Definition 9.1.7: Field

Let F be a divison ring. If F is also *commutative*, then F is called a *field*.

Sometimes, the symbol k (from German Körperor[5]) or $\mathbb{F}$ will be used for a field The field is the "smallest" mathematical structure in which we can perform all the standard arithmetic operations $+, -, \times, \div$ in the usual way. If R is a field, then it is as though there is two compatible abelian group structures (both are associatve, have an identity, have an inverese, both commute), the two operations being linked through distributivity. In particular, it's like $(R, +)$ and $(R - \{0\}, \cdot)$ are combined through distributivity. . If R is a field, we can call the sub-rings of R a sub-field. For example, $\mathbb{Q}$ is a subfield of $\mathbb{R}$.

(TBD revise this in light of the following fact: If F and K are field, $(F, +) \cong_{\mathbf{Grp}} (K, +)$ and $F^\times \cong_{\mathbf{Grp}} K^*$, then is it true that $F \cong_{\text{Ring}} K$? I saw that no, and in principle there can be uncountably many combinations of possible field structures)

Example 9.1.2: Exercise

Prove that any subfield of $\mathbb{R}$ must contain $\mathbb{Q}$

Proof. Let $K \subseteq \mathbb{F}$ be a subifeld. First show the identity of $\mathbb{R}$ is the identity of K. Then $\mathbb{Z}$ is there (why?). Then $\mathbb{Q}$ is there (why?) □

define a quandle (an exampleis matrices over non-commutative rings and $(a^b)^c = (a^c)^{(b^c)}$. Show that conjugation of a group defines a quandle. This is more useful in set theory

[5]some words about whatever German Mathematician pioneered fields that lead to this notation

Zero divisors and Units

If a ring is not a division ring, then the cancellation property does not necessarily hold (see proposition 1.1.9). In groups, we constantly take advantage of the fact that $ab = ac \Rightarrow b = c$. This proof relied on the [unique] inverse of elements. However, there are rings in which this does not happen

Example 9.1.3

Let $R = \mathbb{Z}/6\mathbb{Z}$ with the usual addition and multiplication. First, show that it's a ring.

Then, notice that $[2]$ times any number in $\mathbb{Z}/6\mathbb{Z}$ is not $[1]$ (i.e. there does not exist $[a]$ such that $[2][a] = [1]$):

$$[2][1] = [2]\,, [2][2] = [4]\,, [2][3] = [0]\,, [2][4] = [2]\,, [2][5] = [4]$$

Furthermore, there exists an element ($[3]$) such that $[2]$ times it gives the additive identity: $[2][3] = [0]$

So it breaks down! You might wonder if for any $a \in R$ if there exists an element b such that $ab = 0$ or $ab = 1$, but that need not be the case:

Example 9.1.4: No Multiplicative Inverses

In $\mathbb{Z}$ with the usual addition and multiplication, take any $a \in \mathbb{Z}$, $a \notin \{0, 1\}$. Then there does not exist an element b such that $ab = 1$, which is shown by $\mathbb{Z}$ not having multiplicative inverse, but also $ab = 0$ cannot happen in $\mathbb{Z}$ for any $b \in \mathbb{Z}$

Thus, one cannot imply the other in general

You might further wonder if for any $a \in R$ that only $ab = 0$ happens or only $av = 1$ happens, that is, if there exists an element $b \in R$ such that $ab = 0$, then there does *not* exist an element $v \in R$ such that $av = 1$ (or vice versa). However, this again is not the case! What if $ab = 0$, or $ab = 1$, does that means $ba = 0$ or $ba = 1$? Also no (in the non-commutative case)!

Example 9.1.5

1. Take S to be the set of all infinite sequences. This set, under component-wise addition, is a group. Then take $\mathrm{End}(S)$, making function composition multiplication. Then this is a ring with identity.

2. Prove that

$$P(a_1, a_2, ..., a_n, ...) = (a_1, 0, ..., 0, ...) \tag{9.1}$$
$$L(a_1, a_2, ..., a_n, ...) = (a_2, a_3, ..., a_{n-1}, ...) \tag{9.2}$$
$$R(a_1, a_2, ..., a_n, ...) = (0, a_1, ..., a_{n+1}, ...) \tag{9.3}$$
$$0(a_1, a_2, ..., a_n, ...) = (0, 0, ..., 0, ...) \tag{9.4}$$
$$1(a_1, a_2, ..., a_n, ...) = (a_1, a_2, ..., a_n, ...) \tag{9.5}$$
$$\tag{9.6}$$

   are all elements in $\mathrm{End}(S)$ $(L, R, P, 0, 1 \in \mathrm{End}(S))$

3. Show that $LP = 0$, but $LR = 1$. Then, show that $PL \neq 0$ (so $LP = 0$, but $PL \neq 0$)

4. You just showed that $LR = 1$. Show that $RL \neq 1$.

Since it's not commutative, it's not imediately clear that since we found an example where $ab = 0$ but $ac = 1$, then we can find the same but multiplying from the left side. However, we can construct exactly such an example from the previous example:

1. Define $\text{End}(S)^{op}$ to be the object where function composition is reversed:

$$f \cdot g = g \circ f$$

   show that $\text{End}(S)^{op}$ is still a ring. Convince yourself that all the previous elements $(1) - (5)$ are in $\text{End}(S)^{op}$.

2. Show that $PL = 0$, but $RL = 1$. Then, show that $PL \neq 0$ (so $LP = 0$, but $PL \neq 0$)

So, we cannot carry over our intuition of the cancellation property into general rings! However, thanks to associatvity, there is *one* relation we can prove: if $ab = 0$, then there does not exist a c such that $ca = 1$. That is, if there exists an element b such that $ab = 0$, then there does not exist a c such that $ca = 1$:

Proposition 9.1.8: Partial Result for ab and ca

Let $a \in R$ such that there exists an element $b \in R$ such that $ab = 0$ (or $ba = 0$). Then there does not exist an element c such that $ca = 1$ (or $ac = 1$), that is:

$$\begin{matrix} ab = 0 \\ ca \neq 1 \end{matrix} \quad \text{and} \quad \begin{matrix} ab = 1 \\ ca \neq 0 \end{matrix}$$

***Proof* :**
Take $a \in R$ as in the proposition. Let's say there exists a $b \in R$, $b \neq 0$, such that $ab = 0$. Let's say there is a c such that $ca = 1$, $c \neq 0$. Then:

$$b = 1b = (ca)b = c(ab) = c(0) = 0$$

a contradiction since $b \neq 0$

Let's say there exists a $b \in R$, $b \neq 0$, such that $ba = 1$. Let's say there is a c such that $ac = 0$, $c \neq 0$.

$$c = 1c = (ba)c = b(ac) = b \cdot 0 = 0$$

a contradiction since $c \neq 0$

Note that as a consequence of this theorem, if an element a has a left-sided inverse l and ar right-sided inverse r, then $l = r$, since:

$$l = l1 = l(ar) = (la)r = 1r = r$$

Since multiplicative inverses can either multiply to 1, to 0 or to neither, we will take a moment to name elements that can multiply into 1, 0, or neither:

Definition 9.1.9: Zero Divisors and Units

Let R be a ring.

1. Let $a \in R$ be a nonzero element. Then a is called a *left* (resp. *right*) *zero divisor* if there is a nonzero element $b \in R$ such that ab (resp. ba) equals to 0. If $ab = ba = 0$, then a is called a *two-sided zero divisor*, or simply a *zero-divisor*

2. Assume $1 \neq 0$. An element u of R is called a left (resp. right) *unit* in R if there is some v in R such that vu (resp. uv) equals to 1. If $uv = vu = 1$. , then u is called a *two-sided unit*, or simply *unit* for short

If R is commutative, then there is no distinction between left and right zero-divisors/units.

The term zero-divisors comes from the following observation: if $ab = n$ in a commutative ring for simplicity, then we may say that $a|n$ and $b|n$ (i.e. a, b divide n). Then if $ab = 0$, then "$a\big|0$" and "$b\big|0$", so a, b divide 0 (they are *zero-divisors*). This notion of divisibility shall further be explored in chapter 10. Note that if u is a unit of R so that there is some v such that $uv = vu = 1$, then v is unique (think of the same proof for groups). The same is not the case for zero-divisors (try playing around in $\mathbb{Z}/10\mathbb{Z}$). Since any unit must be a two-sided inverse, we can define the following:

Proposition 9.1.10: Group of Units

Let R be a commutative ring, and let $R^\times$ denote the set of all units of R. Then $R^\times$ is a group under $\cdot$, and is called the *group of units of* R

Proof **:**
Associativity comes from associativity of multiplication, multiplicative identity comes from the fact that 1 is a unit, and inverses come from the definition of units.

It only remains to show that $\cdot$ is indeed a well-defined operation; does it map units of R to units of R. Take $a, b \in R^\times$. Since both are units, then there exists a u and v st $au = ua = bv = vb = 1$. Then I claim that vu is the element such that $(vu)(ab) = 1$. Indeed:

$$(vu)(ab) = v(ua)b = v \cdot 1 \cdot b = v \cdot b = 1$$

Thus $ab \in R^\times$, showing it is indeed well-defined as we sought to show

Sometimes, the group of units of R is denoted $U(R)$, R^*, or $E(R)$ (from German *Einheit*, meaning "oneness") Notice how this let's us characterize division rings and fields in a new way: if every non-zero element of a ring (resp. commutative ring) is a unit, the ring is then a division ring (resp. field). Note too that it is not always obvious when a ring can be characterized as a field in this way:

Example 9.1.6: Every Element is a Unit

Let D be a not-a-perfect square rational number [a] Then take $\mathbb{Q}(\sqrt{(D)}) \subseteq \mathbb{C}$,

$$\mathbb{Q}(\sqrt{(D)}) := \left\{ a + b\sqrt{(D)} \;\middle|\; a, b \in \mathbb{Q} \right\}$$

Note that if $D > 0$, then $\mathbb{Q}(\sqrt{D}) \subseteq \mathbb{R}$. This set is indeed a ring since

$$\begin{aligned} &\forall x, y \in \mathbb{Q}(\sqrt{5}) \qquad a + b\sqrt{D} - c - d\sqrt{D} = (a - c) + (b - d)\sqrt{D} \\ &\forall x, y \in \mathbb{Q}(\sqrt{5}) \qquad (a + b\sqrt{D})(c + d\sqrt{D}) = (ac + bdD) + (ad + bc)\sqrt{D} \end{aligned}$$

showing it's indeed a [commutative] ring.Note that D being not-a-square implies every element can be written uniquely. If it were a square, say 9, then

$$2\sqrt{9} = 6$$

meaning two elements were written the same way! Because of this, if a, b are non-zero, then $a^2 - b^2 D$ is non-zero. For example, if $D = 4$, then

$$4^2 - 2^2 4 = 16 - 16 = 0$$

but Since D is not-a-square,

$$a^2 - b^2 D = 0 \Rightarrow a^2 = b^2 D \Rightarrow \frac{a^2}{b^2} = D \Rightarrow \frac{a}{b} = \sqrt{D}$$

which since D is not-a-square, it's square-root is not a rational number [b], hence a contradiction. Thus, $a^2 - b^2 D$ is non-zero if a, b are non-zero. The point of this is to notice that

$$(a + b\sqrt{D})(a - b\sqrt{D}) = a^2 - b^2 D$$

then $a + b\sqrt{D}$, $a - b\sqrt{D}$ must both be non-zero! Thus, we can define

$$\frac{a - b\sqrt{D}}{a^2 - b^2 D}$$

to be the inverse of $a + b\sqrt{D}$! This shows that every non-zero element of $\mathbb{Q}(\sqrt{D})$ is a unit, and hence it's a field!

[a] perfect square number: $x \in X$ such that $\sqrt{x} \in X$
[b] A similar proof to that of $\sqrt{2}$ is not rational!

This also means that since every element in a field has a multiplicative inverse, a field has *no* zero-divisors.

As we saw, if x is *not* a zero divisor, this does not imply that x *is* a unit, and vice-versa, as we have seen with the integers. However, in the finite case, this *does* hold

Proposition 9.1.11: In finite ring, not zero divisor $\Rightarrow$ unit

If R is a finite ring, and $x \in R$ is *not* a zero-divisor, then x is a unit

***Proof* :**
The key idea of finiteness is that we can take advantage of forcing injective maps being surjective. Withought loss of generality, when there are zero divisors, we take the left zero divisors. The proof for right zero-divisors works identically by switching the order of the elements.

Let's say $x \in R$ is not a left zero divisor so that for no element $y \in R$ there is $xy = 0$. If $x = 1$, we're done, so let's say $x \neq 1$. Define the map $y \mapsto xy$. [a]. Since for no y there is $xy = 0$, then this map is injective. Since R is finite and the map is from R to R, then this map is also surjective. Thus, there exists a y' such that $xy' = 1$. It remains to show that $y'x = 1$.

The element x can either be a right zero-divisor or it can not be. If it is not a right-zero divisor, then the map the map $y \mapsto yx$ would be injective, and hence surjective, and we can find an element y'' such that $y''x = 1$. So, let's say x was a right zero-divisor. If it was a right zero-divisor, then there would exist a $z \in R$ such that $zx = 0$, and x would be a right, but not a left zero-divisor. However, there exists a y' such that $xy' = 1$, which is a contradiction by proposition 9.1.8. Therefore, in the finite case, if x is a right zero-divisor, it would also have to be a left zero-divisor

Therefore, x is not a zero-divisor, so we can again define the map $y \mapsto yx$ which is also injective, and so there exists a y'' such that $y''x = 1$. Finally, notice that:

$$y' = 1 \cdot y' = (y''x) \cdot y' = y'' \cdot 1 = y''$$

showing that $y' = y''$ and so $xy' = y'x = 1$, fulfilling the condition for x being a unit, as we sought to show.

[a] Tangentially, note that this function is not in general a ring homomorphism – can you see why?

No Zero-Divisors: Integral Domains

Numbers as they are generally known do not have zero divisors (ex. $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$). A large part of the study of rings is the study of number theory, and so defining rings which do not have zero divisors is important for ease of communication:

Definition 9.1.12: Integral Domain

A commutative ring with identity $1 \neq 0$ is called *integral domain* if it has a no zero divisors

Example 9.1.7: Integral Domains

1. $\mathbb{Z}$ is the canonical example, even the name-sake
2. $\mathbb{Q}$ and $\mathbb{R}$ are integral domains. Note that every (nonzero) elements are also units
3. $(\mathrm{GL}_n(\mathbb{F}), +, \cdot))$ is an integral domain
4. $M_n(F), +, \cdot)$ is not an integral domain, since

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

5. Let $F = \mathrm{Hom}_{\mathbf{Set}}(X, R)$ be a collection of functions where $|X| > 2$ $R \neq \{0\}$. As noted in example 9.1.1 this is a ring. This ring is not an integral domain. Let $a, b \in X$, $a \neq b$, and $x \in R$, $x \neq 0$. Then define a function where $f(a) = 0, f(b) = x$, and the rest of the elements of f map to zero, and define $g(a) = x, g(b) = 0$, and the rest of the elements of g map to

zero. Note that f and g are not the zero function since they are not identically zero, but $fg = 0$. Importantly

$$fg(a) = f(a)g(a) = 0x = 0 \qquad fg(b) = f(b)g(b) = x0 = 0$$

The set of functions from a set X to a ring R is rarely an integral domain when $|X| > 0$ and $R \neq \{0\}$, since it would have to be the case that $fg = 0$ implies that either f is zero everywhere or g is zero everywhere. However, they do exist (an important example in complex analysis is the ring of power-series, or more generally the ring of analytic functions).

Sometimes the term *domain* used for the same type of ring without the commutative property:

Definition 9.1.13: Domain

A ring in which there are no zero-divisors is called a domain

The nicest property of integral domains is that the cancellation law holds!

Proposition 9.1.14: Cancellation Law for Integral Domains

Assume a, b, c are elements of any ring with a not a zero divisor. If $ab = ac$, then either $a = 0$ or $b = c$.

Proof **:**

$$\begin{aligned} ab &= ac \\ \Leftrightarrow a(b-c) &= 0 \\ \Leftrightarrow a = 0 \text{ or } b - c &= 0 \\ \Leftrightarrow a = 0 \text{ or } b &= c \end{aligned}$$

With this property of integral domains, proposition 9.1.11 can be expanded upon:

Corollary 9.1.15: Finite Integral Domain is a Field

If R is a finite integral domain, then R is a field

Just like in proposition 9.1.11, the trick in this proof is a neat use of function to capture the idea of how an element acts on every other element in an object, not unlike how we can think when working with group actions:

Proof **:**
Let R be a finite integral domain. and let a be a nonzero element of R. By the cancellation law the map $x \mapsto ax$ is an injective function. To see this, let's say it was not injective so $ax = ay$ but $x \neq y$. Then

$$ax - ay = 0 \quad \Rightarrow \quad a(x - y) = 0$$

Since a is non-zero and there are no zero divisors, it must be that $x - y = 0$, so $x = y$.

Now, since the map is injctive, then $|\varphi(R)| = |R|$, meaning it must be surjcetive [a]. In particular,

there is some $b \in R$ such that $ab = 1$. A similar procedure can be done to show that there exists a c such that $ca = 1$. Since units are unique (see exercise ref:HERE) $c = b$, so a is a unit in R. Since a was arbitrary nonzero element, R is a field.

[a]Note that for finite sets, it's impsosible to have a set with "less" elements, but be smaller cardinality, while it is possible for infinite sets, like $2\mathbb{Z} \subseteq \mathbb{Z}$, but $|2\mathbb{Z}| = |\mathbb{Z}$

Note: If we drop commutativity and add that every element is a unit (getting a divsion ring), then this theorem still holds (partly by proving that every finite division ring is necessarily commutative)! The proof of this is relegated to Rerpresentation theory in part VI since it requires a lot more build-up, but it's interesting to note.

Though integral domains are nice, they loose a every important property: If R is an integral domain, then $R \times R$ is *not* an ingeral domain

Example 9.1.8: product of integrals domains is not an integral domain
Let R be an arbitrary integral domain. Then take $R \times R$, with addition and multiplication done point-wise. Then

$$(1, 0) \cdot (0, 1) = (0, 0)$$

meaning there are zero-divisors, and so $R \times R$ is not a integral domain A particular consequence of this is that the direct product of fields is not a field (since fields are integral domains) More generally, integral domains do not satisfy the universal property of products.

A summary of all these properties is in section 9.1.2.

Exercise 9.1.1

1. An element $e \in R$ is called *idempotent* if $e^2 = e$. Does there exist a ring R that has exactly 3 idempotent elements? Find one or show it's not possible.
2. We'll prove that every finite integral domain is a ring using the pigeonhole principle
   a) Show... click here for this proof
3. We will define the complex numbers. Take the real numbers $\mathbb{R}$, and let i be some variable. Define

   $$\mathbb{C} := \langle \{\mathbb{R}, i\} \mid i^2 = -1 \rangle$$

   That is, we will take all possible combinations of the elements from $\mathbb{R}$ and the element i, and impose the relation $i^2 = -1$ on i. To get an idea of what's in $\mathbb{C}$:

   $$i \in \mathbb{C},\ ci \in \mathbb{C},\ a + bi \in \mathbb{C} \cdots$$

   Prove that $\mathbb{C}$ is a ring. We will later give a slightly more nuanced definition of of $\mathbb{C}$, see ref:HERE.
4. Let' say $s \neq 1$ and $s^2 = s$. This propety is called idempotence. Show that the only idempotent element in an integral domain is 0 and 1

5. Let R be a ring with $1 \neq 0$. Is $\varphi : R \to R$, $\varphi(r) = 0$ a ring-homomorphism? If so, prove it, and if not, would it be a rng-homomorphism (i.e. would the image be a rng)?

6. Show that the only idempotent elemnt in a field‘ is 0 or 1 (there exists a proof using properties of a field but not properties of integral domain)

7. A field is said to be characteristic n if there exists a number n st $\underbrace{1+1+1\cdots+1}_{n \text{ times}} = 0$. If no such number exists, we say that the field has characteristic 0. Give an

8. Let $u \in R$ be a unit and $R \to S$ a ring homomorphism. Then $\varphi(u)$ is a unit. Is the same true for zero-divisors? What about non-zero-divisors? example (or examples) of fields with characteristic 0, 2, and 5.

9. Show that if $\mathbb{F}$ is a field of characteristic 0, then $\mathbb{Q} \subseteq \mathbb{F}$.

10. (For those who know some Analysis and want examples of *rngs*) In the process of defining the Fourier series, it is necessary to define the convolution product between two continuous functions (on compact support, or more generally L^2)

11. (ibid last comment) Let X be a locally compact space (i.e. for every $x \in X$, there exists a compact set $K \subseteq X$ such that $x \in K$, that is every point has a compact neighbourhood). Let $C_0(X)$ be the set of continuous complex functions $f : X \to \mathbb{C}$ that vanish at infinity (for every $\epsilon > 0$, there exists a K such that $|f(x)| < \epsilon$ for all $x \in X - K$). Then $(C_0(X), +, \cdot)$ with function addition and multiplicatoin is a rng, but not a ring (namely, the function $1(x)$ does not vanish at infinity). This rng is used quite often in analysis: it comes up in Fourier analysis, in defining Radon Measures, and in the study of C^*-algebra to name a few examples (see Folland or ref:HERE).

12. (do the example in Commutative algebra 2 (Rings, ideals, modules) at timestamp 13:33)

    (somehow make this a comment??) Loosely speaking, rngs correspond to locally compact spaces, while rings correspond to compact spaces. For analyists, working over compact spaces is much to restrictive ($\mathbb{R}^n$ is not compact), while locally compact spaces are much more natural to work with ($\mathbb{R}^n$ is locally compact). Furthermore, recall we have mentioned we can turn a rng into a ring by doing $\mathbb{Z} \oplus R$ to some rng R. If we were interested in keeping nice analytical properties, we would instead take $\mathbb{R} \oplus R$. (ex. $\mathbb{R} \oplus C_0(X)$). This would correspond to the 1-point compatification of X: $C_0(X \cup \{\infty\}) \equiv \mathbb{R} \oplus C_0(X)$!

13. Show that $(F, +, \cdot)$ be a ring. Show that $(F, +, \cdot)$ is a field if and only if $(F, +)$ is an abelian group and the set $F^\times$ is equal to $F - \{0\}$.

14. Let R be a ring and R^{op} denote the ring where the underlying set and addition operation is the same, and the multiplication is defined as $x * y = yx$ where yx happens in R. Show that $\text{id} : R \to R^{\text{op}}$ is not necessarily a ring isomorphism (it is sometimes called an *anti-isomorphism*). When is it a ring isomorphism? Show that if R is a division ring, then the map $x \mapsto x^{-1}$ *is* a ring isomorphism.

15. show that if $x^n = 0$, then $(x - 1)$ is an invertible element. An element where $x^n = 0$ is called a *nilpotent element*.

16. recall that $(x + y)^n = \sum_{k=1}^{n} \binom{n}{k} x^n y^{n-k}$. Show that the binomial formula holds if and only if $xy = yx$.

17. I the set of all left (resp. right) invertible elements a group? How that $Z(R)$ is a subring of R.

18. let G be a cyclic group of order n. For each $k \in \mathbb{Z}$, define $f : G \to G$, $f_k(g) = kg$. The n the map $k \mapsto f_k$ defines a ring isomorphism $\mathbb{Z}/n\mathbb{Z} \cong_{\mathbf{Ring}} \mathrm{End}_{\mathbf{Ring}}(G)$ and a group isomorhpism $(\mathbb{Z}/n\mathbb{Z})^* \cong_{\mathbf{Grp}} \mathrm{Aut}_{\mathbf{Grp}}(G)$

9.1.1 Category of Ring

We now approach rings a 2nd time, but this time, we will take a more birds-eye view of the category or rings. We will see if this category has some properties different than the category of groups, and whether there is a general way of extending groups to form rings.

Proposition 9.1.16: Category of Ring

The collection of Rings along with ring homomorphisms form a category. This category will be denoted **Ring**.

Proof **:**
See exercise ref:HERE

Since **Ring** is a category, it will be verified what are the initial/final objects, and whether there are product, sub-, and quotient objects, and whether quotient objects can be classified through kernels. We can further ask about the relation between **Group** and **Ring**, and ask whether there is an analogous concept of ring-actions. In this section we will drop the subscript and write $\mathrm{Hom}(R, S)$ or $\mathrm{End}(R)$ instead of $\mathrm{Hom}_{\mathbf{Ring}}(R, S)$ or $\mathrm{End}_{\mathbf{Ring}}(R)$

The fact that sub-objects and product objects exist in rings (that is, subrings and product rings) has already been explored in definition 9.1.3 and example 9.1.1[5], though not in categorical rigour (which the reader can practice in exercise ref:HERE if they are so inclined). We will mention some more about product of rings and the consequences of integral domains not having products as an interesting example of category with no products. Quotient objects will be revisited in more detail in section 9.3.

Rings and $\mathbb{Z}$

The first thing to check is whether **Ring** has initial and final objects. In definition 1.4.4, we introduced the concept of initial and final group. We found there to not really be an intersting initial and final group: only the identity group. The same is true for the final object of rings. However, the zero ring cannot be the initial: if R is the zero ring, and we have a homomorphism $R \to S$, then S must be the zero ring (see exercise ref:HERE). There is still an initial object in the category or rings: we have our first example of a non-trivial initial object!

Proposition 9.1.17: $\mathbb{Z}$ is an Initial Ring

The ring $\mathbb{Z}$ is an intial ring

***Proof* :**
Let R be any ring (with identity) and let $\varphi : \mathbb{Z} \to R$ such that

$$\varphi(z) = 1 + 1 + 1 \text{ (z times) } = z \cdot 1$$

this is a ring homomorphism. For

$$\varphi(a+b) = (a+b) \cdot 1 = a + b = \varphi(a)\varphi(b) \qquad \varphi(ab) = (ab) \cdot 1 = a(b \cdot 1) = (a \cdot 1)(b \cdot 1) = \varphi(a)\varphi(b)$$

Furthermore, if $\varphi' : \mathbb{Z} \to R$ was another ring homomorphism, then by proposition 9.1.5

$$\varphi'(1) = 1\,, \varphi'(2) = 1 + 1\,, \cdots$$

that is, by the fact that the identity *must* be distributed in this fashion, we are forced that

$$\varphi = \varphi'$$

making it unique

Note that if rings are not defined to have a multiplicative identity, then there is no unique ring wihtout a multiplicative identity, see exercise ref:HERE (exercise 2.16 in Aluffi part III)

So how can we use this result? The way I think about it is something "integer-like" in every ring [with identity]. It might not be commutative, it might not be injective (the codomain might even be finite), or it might be uncountably big (like $\mathbb{R}$ or $M_n(\mathbb{C})$) but there is still something integer like. For this reason, asking questions like "what properties of the integers do rings have" is actually a very sensible question! In fact, that will be essentially what we do in chapter 10!

For example, if we take $x \in \mathbb{Z}$, then $x = p_1 p_2 \cdots p_n$, that is x has a *unique prime factorization.* In the next chapter, we will study rings which have this similar property (they will be called *unique factorization domains*). Similarly, to find the $gcd(a, b)$, we can use the euclidean algorithm which algorithmically tells us how much the set of divisors of the two numbers intersect. This is very handy since we can in very few steps find the common divisors of two numbers of extremely large numbers. Rings in which the notion of the euclidean algorithm is well-dfined will be called *euclidean domains.* Finally, a little more difficult but important concept, any number can be decomposed into *unique* prime numbers in $\mathbb{Z}$. We will find the equivalent of "numbers" in rings to be "ideals" (the original name was "ideal numbers" of rings). If every ideal can be generated by a single element $x \in R$, the ideal is called a *principal ideal.* Thus, a ring which is has "numbers" (ideals) which are generated by one element (just like a number just needs itself to describe, well, itself) will be called a *principal ideal domain.*

Thus, in the folloing chapters, some basic theory will extensively be used. Take a moment to review section 0.2 for a refresher on some basic tools of number theory.

Rings and $\mathrm{End}(G)$

You might ask how should we choose which of the axioms are "nice" axioms to add to our structure. One way of doing so is by thinking about the multiplication action being permutations of the elements, not unlike ho the group operation can be thought of as the permutation of elements (i.e. any group

$(G, +)$ is a subgroup of $\mathrm{Sym}(G)$) [7]

We saw in definition 1.4.6 that if G is an abelian group that $\mathrm{End}(G)$ is an agelian group under pointwise addition. Notice that there is actually already a 2nd operation, since function can compose:

$$(\mathrm{End}(G), +, \circ)$$

Then, it can be checked that composition here will make this into a ring (as shown in example 9.1.1)

You might then ask converse: given ring $(R, +, \cdot)$, can we find a ring isomorphism to $\mathrm{End}((R, +))$? The answer is yes! This will actually turn out to be the Cayleys' theroem for rings (theroem 4.4.2)! We will not cover the ring version of Cayley's theorem, for a more general theorem called Yonneda's Lemma cover's the idea of using objects to "represent" other objects (ref:HERE)

If you want to check, the natural ring embedding from $(R, +, \cdot)$ to $\mathrm{End}(R, +)$ is

$$r \mapsto (x \mapsto xr) \qquad \text{or} \qquad r \mapsto (x \mapsto rx)$$

depending on the convention on composition.

:exercse

Let $\mathrm{Aut}_{\mathbf{Ab}}(R)$ be the set of group-automorphisms on R. Show that $U(R) \cong_{\mathbf{Ab}} \mathrm{Aut}_{\mathbf{Ab}}(R)$

let R be a commutative ring. Show that $\lambda : R \to \mathrm{End}_{\mathbf{Ab}}(R)$ by $r \mapsto (y \mapsto ry)$ is indeed a ring homomorphism. If R is not commutative, show that it's almost abelian.

Products of Rings

Rings have products: Given any to rings A and B, there exists a ring $A \times B$ along ring homomorphisms $\pi_A : A \times B \to A$ and $\pi_B : A \times B \to B$ such that for any C here $f_A : C \to A$ and $f_B : C \to B$, there exists a unique $f! : C \to A \times B$ such that the folloing diagram commutes:

$$\begin{tikzcd}
 & C \arrow[d, "f!"] \arrow[ddl, bend right, "f_A"'] \arrow[ddr, bend left, "f_B"] & \\
 & A \times B \arrow[dl, "\pi_A"] \arrow[dr, "\pi_B"'] & \\
A & & B
\end{tikzcd}$$

In this way, products in **Ring** are like products in **Grp**. A similar situation occurs for coproducts.

However, as mentioned earlier, if we retirct to *integral domains* and take the category whose objects are integral domains (this is indeed still a category as can be verified), they are no longer products, as demonstrated in exmaple 9.1.8!

Since since all fields are integral domains, this means that products are not defined for fields! As a consequence of this, we cannot repeat the same trick as in chapter 5 in creating larger groups using products. Thus, extensions of fields (similar to how they were defined for groups in section 3.6) will have to be used to study the creation of larger fields. To study field extensions, we must first study ring homomorphisms more extensively; the particular way in which ring homomorphissm will be used in constructing larger fields requires more build-up, and so the study of fields is relegated to part IV.

[7] Research **Ab**-enriched categories for information about this process

9.1.2 Summary

Keeping track of these properties and their relation to everything can be a bit to take in. It might help to have a visual too:

$$
\begin{tikzcd}
 & \text{field} & \\
\text{integral domain} \arrow[ur, "\forall a \exists b \ st \ ab{=}1 \ (+1)"] & & \text{division ring} \arrow[ul, "ab{=}ba"'] \\
\text{commutative ring} \arrow[u, "\forall a \nexists b \ st \ ab{=}0"] & & \text{domain} \arrow[u, "\forall a \exists b \ st \ ab{=}1 \ (+1)"'] \arrow[ull, "ab{=}ba"] \\
 & \text{ring} \arrow[ul, "ab{=}ba"] \arrow[ur, "\forall a \nexists b \ st \ ab{=}0"'] & \\
 & \text{rng} \arrow[u, dashed, "\text{add identity } 1"] &
\end{tikzcd}
$$

The *rng* can actually have more arrow's extending from it to commutative ring, domain, and integral domain, but the diagram would be cumbersome, so I opted to make it a dashed line from *rng* to *ring*, since, it is not needed to "upgrade" those rings until we get to division ring and fields. I added a (+1) where we really need the multiplicative identity to define the next concept.

Notice that if a commutative ring only has units, it is then a commutative division ring, and hence a field. The line as not drawn so as to not clutter the diagram.

look at comment:

The following categorical results are also noted:

1. $\mathbb{Z}$ is the initial object of rings
2. All rings are subrings of $\mathrm{End}((R, +))$
3. Products exist for rings, but not for integral domains (and so not for fields).

Exercise 9.1.2

1. Since $\mathbb{Z}$ is initial, for any ring R and natural homomorphism $\mathbb{Z} \to \mathbb{R}$, $\mathrm{im}(\mathbb{Z})$ exists is a subring. Show that $\mathrm{im}(\mathbb{Z})$ is the *smallest* subring of any ring. Show further that $\mathrm{im}(\mathbb{Z})$ is either $\mathbb{Z}_n$ for any $n > 0$ or $\mathbb{Z}$. If $|\,\mathrm{im}(\mathbb{Z})| < \infty$, then we say the ring R has *characteristic* n, and if $\mathrm{im}(\mathbb{Z}) \cong \mathbb{Z}$ then R has *characteristic 0*.
2. Let R be a ring and define R^{op} to be the ring with the same underlying set and addition operation, and multiplication defined by $x * y = yx$, where yx happens in R. Show that R^{op} is a ring, and that $(_)^{\mathrm{op}} : \mathbf{Ring} \to \mathbf{Ring}$ is a functor (where do ring homomorhpisms map to?). Is it covariant or contravariant?
3. Let $a, b \in R$ such that $ab = ba$. Show that ab is invertible if and only if a and b are invertible. Generlize your result to show that if $ab \neq ba$, then if ab invertible, then a is left-invertible, and b is right-invertible.

4. Show that $\mathrm{End}_{\mathbf{Ab}}(\mathbb{Z}) \cong_{\mathbf{Ring}} \mathbb{Z}$.

9.2 Examples of Rings

The following are rings that will show up time and time again, and being aquainted with them is very useful. Many of these rings will be the starting point to a new field of study in mathematics. For example, Polynomial rings are pivotal in Commutative Algebra (see part V), and when the polynomial ring takes on coefficients from a field (i.e. a polynomial ring over a field), then they form the foundation for field theory (see part IV). Matrix rings are a common the "representation objects", the same way $\mathrm{Sym}(X)$ was an object that can be used to "represent" groups (see part VI chapter 23). Group rings can be thought of as the dual of the group of unity of rings, and also appear frequently in representation theory (see part VI chapter 24) and group cohomology (see part VII chapter 28).

There is also an optional section on formal power series covering the elementary operations that can be performed on formal power series. These results are important for a class in complex analysis, however the formal power series ring will come up as examples for certain concepts introduced later in the book as well. These examples of formal power series will not be critical in understanding any material, and so they can be comfortably skipped.

9.2.1 Polynomial Rings

Polynomial Rings are one of the first examples we'll give of a [commutative] extension of a ring. Let R be a commutative ring [8]. What we'll be doing is extending R by an arbitrary unkown element x. In a sense, what we're doing is saying x is now an element of R, an then we figure out how to define this new structure (which we label $R[x]$) so that it's also a ring. Since $x \in R[x]$, then $x \cdot x \in R[x]$. Since we know nothing of x, we can only say that $x \cdot x = x^2$ is another element. We can continue this process inductively getting us every x^k for all $k \in \mathbb{N}$. For any other element r: $r \cdot x = rx$ where rx is a new element of $R[x]$. A similar reasoning to creating new elements applies to adding: $r + x$ is a new element. However, it need not be the case that an element has an inverse, so we do not define an x^{-1} such that $xx^{-1} = 1$[9]. All of these properties together give us the following definition:

[8] In the exercises, there are questions concerning the study of non-commutative R forming a polynomial ring and why it is often preferred for R to be commutative

[9] polynomial rings where the indeterminate x has an inverse are called the *ring of fractions of R[x]*. see section 9.5

Definition 9.2.1: Polynomial Ring

Let R be a ring and x be an indeterminante. Then the sum

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

with $n \geq 0$ and each $a_i \in R$ is called a *polynomial* in x with coefficients a_i in R. The set of all such polynomials is called the ring of *polynomial in the variable x with coefficients in R* (or *polynomial ring* if context permits) and will be denoted $R[x]$. Addition is defined as

$$(a_n x^n + \cdots + a_1 x + a_0) + (b_n x^n + \cdots + b_1 x + b_0) = (a_n + b_n)x^n + \cdots (a_1 + b_1)x + (a_0 + b_0)$$

where a_n or b_n may be zero in order for addition of polynomials of different degrees to be defined. Multiplication is perfomed by defining $(a_i x^i)(b_j x^j) = a_i b_i x^{i+j}$ and then extending to all polynomials by the distributive law; this is called the *convolution product*
The following are pertinent definitions with regards to polynomials.

1. If $a_n \neq 0$, then the polynomial is said to be of *degree n*, and $a_n x^n$ is called the *leading coefficient* (where the leading coefficeint of the zero polynomial is taken to be 0).
2. The polynomial is *monic* if $a_n = 1$.

For reference, here is the general formula for polynomial multiplication (i.e. the convolution product)

$$\left(\sum_{i=0}^{\deg(p)} p_i x^i\right) \cdot \left(\sum_{i=0}^{\deg(q)} q_j x^j\right) = \sum_{k=0}^{\deg(p)+\deg(q)} \sum_{l=0}^{k} p_l q_{k-l} x^k$$

21: Remark

If you have done Fourier Analysis (or intend to in the future), then the convolution of two continuous (more generally measurable) functions follow's the exact same properties as polynomial multiplication. In fact, polynomial multiplication would be a discrete convolution.

Note that R is always a subring of $R[x]$, which we can see through the natural inclusion map [10] $R \hookrightarrow R[x]$ of $r \mapsto r$.

Usually, when we manipulate polynomials we treat them as functions. It must be stressed that we are *not* treating the elements of $R[x]$ as functions. The study of the elements of $R[x]$ as functions will be done in part V, which will establish the largest connection between algebra and geometry in the 20th century.

If $R = F$ is a field, then every coefficient is a unit, meaning we can divide each coefficient a polynomial to make it *monic*, that is,

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \overset{1/a_n}{\mapsto} x^n + a'_{n-1}x^{n-1} + \cdots + a'_1 x + a'_0$$

where $a'_i = \frac{a_i}{a_n}$. We shall see that often-times many results *up to multiplication by unit* (meaning if the result holds on p, it will hold on cp, where c is a unit). For this reason, it is worth keeping the term *monic* in mind.

[10] the map is natural since there is no other information needed to define this map besides R and $R[x]$. For a more rigorous definition of natural, appendix 30 needs to be read to reach section 30.3

We can extend the definition of a polynomial ring to a polynomial ring over several variables by simply making $R[x]$ a polynomial ring over y, that is

$$R[x, y] = (R[x])[y] = R([y])[x] = R[y, x]$$

Continuing inductively, we can define a ring over a finite number of variable this way:

$$R[x_1, x_2, ..., x_{n+1}] = (R[x_1, x_2, ..., x_n])[x_{n+1}]$$

This process can be extended "indefinitely"[11] to define $R[x_1, x_2, ..., x_n, ...]$, in which every polynomial (i.e. elements of $R[x_1, x_2, ...,]$) has finitely many terms, and each term can have multiples of any of variables of $R[x_1, x_2, ..., x_n, ...]$, for example:

$$x_{30}^2 x_7^{23523} x_2 - r x_{100} x^{39} \in R[x_1, x_2, ...]$$

Multiplication in $R[x_1, x_2, ...]$ is the same as the multiplication in polynomial rings over finitely many variables, where we know the value for any term be expanding out the multiplication far enough to group like-terms. If $x_1^{d_1} x_2^{d_2} \cdots x_k^{d_k}$ is a term, then the *total degree* of the term is

$$d = d_1 + d_2 + \cdots + d_k$$

and the tuple $(d_1, d_2, ..., d_k)$ is the multi-degree of the term.

Properties of Polynomial Rings

To motivate the next theorem, let's say we have two polynomials of degree 1, $ax + b, cx + d \in R[x]$. We can ask what is the degree of $(ax + b)(cx + d) = acx^2 = (ad + bc)x + bd$. We might expect the degree to be 2. However, it is possible that a and c are zero-divisors: take $4x + 1, 4x + 1 \in (\mathbb{Z}/8\mathbb{Z})[x]$. In fact, in the example just stated, $4x + 1$ is in fact a unit in $(\mathbb{Z}/8\mathbb{Z})[x]$! If R is an integral domain, then some of its structure will be simpler:

Proposition 9.2.2: Degree Of Polynomials

Let R be an integral domain and let $p(x), q(x)$ be nonzero elements of $R[x]$. Then

1. degree $p(x)q(x)$ = degree $p(x)$ + degree $q(x)$
2. the units of $R[x]$ are the units of R
3. $R[x]$ is an integral domain

Proof **:**

1. This comes from the definition of multiplication of polynomials and that R is an integral domain, as the highests order term will be $x^{\deg(p(x)+\deg(q(x)))}$ and there is no zero-divisors to cancelt out the term.

[11] For those uncomfortable with how the fact that the above process of defining a polynomial ring will always only produce polynomial rings with finitely many variables, the construction of $R[x_1, x_2, ...]$ is done using a *limit*, see ref:HERE

2. Let $u \in R[x]$ be a unit so that there exists a $v \in R[x]$ such that $uv = 1$. Then by the degree formula, $\deg(uv) = \deg(u) + \deg(v)$ [a]. Since $\deg(uv) = 0$, then it must be that $\deg(u) = \deg(v) = 0$, showing that $v \in R$

3. Let $p(x), q(x) \in R[x]$ such that $p(x)q(x) = 0$. Then $0 = \deg(0) = \deg(p(x)q(x)) = \deg(p(x) + \deg(q(x))$, and so both $p(x), q(x)$ must be of degree zero and so in R. Since R is an integral domain, then it must be that $p(x) = q(x) = 0$, showing that $R[x]$ is an integral domain.

[a] notice again that R must be an integral domain for this to work

The defining property of commutative rings is that they are the free objects of commutative rings.

Theorem 9.2.3: Universal Property of Polynomial Rings

Let S be a commutative ring and let

$$\varphi : R \to S$$

be any ring homomorphism and let $s \in S$ be any element of S. Then there is a unique ring homomorphism

$$\psi : R[x] \to S$$

such that $\varphi(x) = s$ and which makes the following diagram commute

$$\begin{tikzcd} R \arrow[r, "\varphi"] \arrow[d, "f"] & S \\ R[x] \arrow[ur, "\psi"'] & \end{tikzcd}$$

A nuance that must be mentioned is that polynomial rings may be over different rings. We thus may thus ask how a ring S is generated with respect to a ring R. As demonstrated earlier, $\mathbb{Z}$ is the initial object of **Ring**, and so in the arbitrary case $f : X \to S$ for some set X, we shall have a polynomial ring on the elements X over $\mathbb{Z}$. For those who are categorically inclined, try figuring out what is the category involved behind the scenes (this is similar to the construction for free groups in section 2.3.3)

***Proof* :**
The proof becomes easy to see once you realize a polynomial:

$$a_n x^n + a_{n-1}x^{n-1} + \cdots a_1 x + a_0$$

becomes

$$\varphi(a_n)s^n + \varphi(a_{n-1})s^{n-1} + \cdots \varphi(a_1)s + \varphi(a_0)$$

A particular case of this is when $\varphi : R \to R$ is an identity map. In this case, we may link the notion of a [formal] polynomial to a polynomial function:

Definition 9.2.4: Evalutation Map

Let R be a ring and let $\alpha \in R$. The natural ring homomorphism

$$\varphi : R[x] \to R$$

which acts as the identity on R and which sends x to α is called the *evaluation at* α and is often denoted ev_α.
We say that α is a *zero* (or root) of $f(x)$ if $f(x)$ is in the kernel of ev_α.

The evaluation map can come in quite handy in many proofs that will be seen later.

Example 9.2.1
Move this to somewhere else

1. Take $\mathbb{Q}[x, y]$ $(= (\mathbb{Q}[x])[y]$, i.e. a polynomial ring over $\mathbb{Q}[x]$ in $y)$. Is (x) a prime ideal? Take $\mathbb{Q}[x, y] \to \mathbb{Q}[x]$ by $p(x, y) \mapsto p(x, 0)$. This is the evaluation map at 0.

   This maps is clearly surjective, since for every $p(x)$, $p(x, 0)$ will map to it. This map is also clearly linear since

   $$\varphi(p(x, y) + q(x, y)) = \varphi\left(\sum_{i=1}^{n}(a_i)x^i + \text{y terms}\right) = \sum_{a_i}^{n} a_i = \varphi(p(x, y)) + \varphi(q(x, y))$$

   and since it's linear on every term, to check the multiplicative operation, it suffices to check on monomials, which is easy:

   $$\varphi(ax^n bx^m) = \varphi(abx^{n+m}) = abx^{n+m} = ax^n bx^m = \varphi(ax^n)\varphi(bx^m)$$

   and if y is aterm, y geos to zero so mulitplicativity for those terms is automatic. Therefore, φ is a ring homomorphism. It's kernel is clealry (y), meaning

   $$\mathbb{Q}[x, y]/(y) \cong \mathbb{Q}[x]$$

   and since $\mathbb{Q}[x]$ is an integral domain (y) is a prime ideal.

When working with functions, they are determined by the value at each point of evaluation. It must be stressed that in general, a polynomial is *not* determined by it's evaluation

Example 9.2.2: Polynomial Not Determined By Evaluation
[Polynomial Not Determined By EvaluationPolynomial Not Determined By Evaluation] Take $x^2 - x \in (\mathbb{Z}/2\mathbb{Z})[x]$. Does the evaluation of this polynomial uniquely determine $x^2 - x$?

9.2.2 *Formal Power Series

(A realisation i had that maybe should be integrated?: I just realised. So we know from complex that [complex] polynomial behaves like waves (since they are harmonic). We know that multiplication of polynomials introduces the operator of convolution. We know Fourier somehow gives you a notion

finding the fourier-coefficients, i.e. finding the waves. Well, if F is the fourier transform, the fact that $F(f * g) = F(f) \cdot F(g)$ feels like it makes more intuitive sense!)

Any polynomial is a finite linear combination of the terms x^n. If we allow for infinitely many terms, we call such an object a *formal power series*, and write it like so:

$$f(x) = \sum_{i=1}^{\infty} a_i x^i \qquad a_i \in R$$

The reader may ask what does it mean to take an infinite linear combination, as we have defined the $+$ notation to only take in two inputs, and if we recurse over the inputs we may have a finite linear combination. These are valid concerns which will be addressed in chapter 20, section 20.4. For now, we may consider $f(x)$ in the above equation as "actually" being an infinite tuple[12]

$$(a_0, a_1, ..., a_n) \in \prod_{i=0}^{\infty} R$$

however we should take this identification as simply insuring that $f(x)$ is well-defined; we shall not write $f(x)$ in the notation of tuples and from now on shall assume we can take infinite sums. The collection of all formal power series are certainly closed under termwise addition. We must make sure multiplication is still well-defined. For multipilcation to be well-defined, each coefficient must be well-defined. Fortunately, this is the case. If $p(x)$ and $q(x)$ where

$$\begin{aligned} p(x)q(x) &= \left(\sum_{i=0}^{\infty} a_i x^i\right)\left(\sum_{j=0}^{\infty} b_j x^j\right) \\ &= a_0 b_0 + (a_0 b_1 + a_1 b_0)z + (a_0 b_2 + 2a_1 b_1 + a_2 b_0)z^2 + \cdots \\ &= \sum_{n=0}^{\infty}\left(\sum_{i=0}^{n} a_i b_{n-i}\right) x^n \\ &= \sum_{n=0}^{\infty} (A_n * B_n)\, x^n \end{aligned}$$

where $A_n = \sum_{i=0}^{n} a_i$ and $B_n = \sum_{i=0}^{n} b_i$ and $A_n * B_n$ is equal to the the summand above and is called the *convolution* of the two finite sequences. Since addition and multiplication is well-defined, we have a ring:

Definition 9.2.5: Formal Power Series

The collection of all infinite R-linear combinations of $\{1, x, x^2, ...\}$ forms a ring known as the ring of *formal power series* over R, and is denoted $R[[x]]$

The term "formal" comes from how we often associate to an object $f \in R[[x]]$ the evaluation function. If we had a function with a domain and codomain for which $r \mapsto \sum_{i=0}^{\infty} a_i r^i$, we would call this function a *power series*. However, for such a function to be defined, we require the ability to unambiguously define summing infinitely many elements of a ring. Such considerations require the introduction of a

[12] this may push the concern in asking how $\prod_{i=0}^{\infty}$ is defined. For those with this concern, they may skip to section 20.4 or they may take for granted that we may always make a countable number of choices and remember the order

topology on our ring, something we shall not cover. We shall remark that it is very commmon in compex analysis to work with power series and choose $R \in \{\mathbb{R}, \mathbb{C}\}$ with the usual euclidean topology, leading to many very important results about complex-differentiable functions. There is one value we may unambiguously define without the need for any convergence considerations:

$$f(0) := a_0$$

Proposition 9.2.6: Formal Power Series Integral Domain

Let $R[[x]]$ be the formal power series over R. Then $R[[x]]$ is an integral domain

Proof **:**
proof here

Given that $R[[x]]$ is the extension of $R[x]$, it may see intriguing to think that these two rings share very similar properties. However, the "infinitude" of $R[[x]]$ will make the ring act rather differently. The following may be the most striking change when extending to infinite polynomials:

Lemma 9.2.7: Existence of More Units

Let $R[[x]]$ be the formal power series over a ring R. Then $(1 - x)$ has an inverse in $R[[x]]$

Proof **:**
Notice that

$$(1 - x)(1 + x + x^2 + \cdots + x^n + \cdots) = 1$$

and hence by definition that means that $1 - x$ has an inverse, namely $\sum_{n=0}^{\infty} x^n$.

By convention, we denote inverses using quotient notation, and hence we would write:

$$\frac{1}{1 - x} := \sum_{n=0}^{\infty} x^n$$

For those who have taken calculus, they may be familiar with an "analysis" version of this proof relying on absolute convergence. The above shows that this result is purely algebraic. In fact, we can fully classify all units in a formal power series

Proposition 9.2.8: Units In Formal Power Series

Let $f \in k[[x]]$. Then f is a unit (i.e. f has an inverse) if and only if $a_0 \neq 0$

Proof **:**
The condition is certainly necessary since if

$$g(x) = \sum_n b_n x^n \qquad f(x)g(x) = 1$$

then we must have $a_0 b_0 = 1$, so $a_0 \neq 0$. If conversely we have $a_0 \neq 0$, we shall show that the

formal power series $(a_0)^{-1}f(x) = f_1(x)$ has an inverse $g_1(x)$, and hence $(a_0)^{-1}g_1(x)$ is the inverse of $f(x)$. Decompose $f_1(x)$ into:

$$f_1(x) = 1 - u(x)$$

then substitute $u(x)$ for y in the proof of lemma 9.2.7 to get that $1 - u(x)$ has an inverse. But then we may conclude that $f(x)$ has an inverse, thus completing the proof.

Another question we may ask is the following: Given two polynomials $f(x), g(x)$, we may take $f(g(x))$, (replace x with $g(x)$). Can the same be done for power series? The following shows when this is possible:

Definition 9.2.9: Formal Derivative

Let $f \in k[[x]]$ be a formal power series. Then if $f(x) = \sum_{k=0}^{\infty} a_k x^k$, the formal derivative of f is the value

$$\frac{d}{dx}(f) = \sum_{k=0}^{\infty} k a_k x^{k-1}$$

we usually denote the derivative of $f(x)$ as $f'(x)$, and for higher derivatives we write $f^{(n)}(x)$ to represent taking recursively the derivative of f n times.

We may use the formal derivative recursively to find the coefficients of f. Namely, $f(0) = a_0$, $f'(0) = a_1$, ..., $f^{(n)}(0) = a_n$, and so forth.

Let f, g be power series. We may wonder when $g(f(z))$ is well-defined:

Proposition 9.2.10: Composing Power Series

Let f, g be formal power series. Then $g(f(z))$ is well-defined if $b_0 = 0$.

Proof **:**
When composing power series, we get:

$$a_0 + a_1(b_1 z + b_2 z^2 + \cdots) + a_2(b_1 z = b_2 z^2 + \cdots)^2 + \cdots$$

where each coefficient is an infinite sum, and so we must ask for their convergence conditions. One thing we can do to guarantee this is by setting $b_0 = 0$, which will guarantee that all the coefficients are finite sums and hence well-defined.

We often denote the composition as $g \circ f$.

Proposition 9.2.11: Associativity Of Composition

Let $f, g, h \in k[[x]]$ be power series. Then

$$(h \circ g) \circ f = h \circ (g \circ f)$$

Proof **:**
Henri Cartan Elementary Theory of Analytic functions p. 13

Define $I(x) = x$. Notice that for any $f \in k[[x]]$, $I \circ f = f \circ I = f$.

Proposition 9.2.12: Composition Has Inverse

Let $f \in k[[x]]$ and $I(x) = x$. Then a right inverse exists, $f \circ g$ if and only if $f(0) = 0$ and $f'(0) \neq 0$ (i.e. $a_1 \neq 0$). For g to also be a left-inverse, we require $g(0) = 0$, in which case

$$f \circ g = g \circ f = I$$

22: remark

(If you know some calculus) Notice how the condition $f'(0) \neq 0$ bears resemblence to the inverse function theorem

Proof **:**
We want to a g such that $f(g(z)) = z$. We want:

$$a_0 + a_1(b_1 z + b_2 z^2 + \cdots) + a_2(b_1 z = b_2 z^2 + \cdots)^2 = z$$

expanding and using the method of undetermined coefficients expanding the first two terms we get:

$$a_0 = 0 \qquad a_1 b_1 = 1$$

This shows us that $a_0 = 0$ and $a_1 = f'(0) \neq 0$ are necessary conditions. To show they are sufficient conditions, we can see that we may deduce all other coefficeints given these conditions.For example, we may deduce the coefficient of z^n as $a_0 + a_1 g(z) + \cdots + a_n g(z)^n$, so

$$a_1 b_n + P_n(a_2, ..., a_n, b_1, ..., b_{n-1}) = 0$$

Thus, we would start by doing $b_1 = 1/a_1$, and $b_2, b_3, ...$ are defined recursively (this resulting polynomials is called *Bells polynomial*). With our construction, we have that g satisfies $b_0 = 0$ and $b_1 \neq 0$. Thus, we can do the same thing to g, that is $g(f_1(w)) = w$. These inverses are equal:

$$f_1 = \mathrm{id} \circ f_1 = (f \circ g) \circ f_1 = f \circ (g \circ f_1) = f \circ \mathrm{id} = f$$

where we assumed the associativity of composition, something that can be checked. Finally, for the radius of convergence, we shall delay such considerations for now, noting that when we show that holomorphic functions are analytic we can use the inverse function theorem to estimate the values.

Example 9.2.3: Formal Power Series
Some classical examples of power series include:

1. $\exp(x) = e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

2. $\cos(x) =$
3. $\sin(x) =$
4. $\log(x) =$

9.2.3 Matrix Rings

As was briefly mentioned in the preliminary chapter (section 0.5), many structures can be "represented as matrices", that is, instead of studying all these structure as separate algebraic objects, they all follow the same rules as matrix rings.

Definition 9.2.13: Matrix Ring

Let R be a ring. Then $M_n(R)$ is the collection of all $n \times n$ matrices with coefficients in R. Addition is defined term-wise and multiplication is defined via matrix-multiplication as given in section 0.5. A matrix may be represented by:

$$(a_{ij})_{ij} \in M_n(R)$$

where a_{ij} represents the value at position ij of the matrix. The identity of $M_n(R)$ is denoted I_n.

In the above notation, matrix multiplication AB of two matrices can be represented as

$$(c)_{ij} = \left(\sum_{k=1}^{n} a_{ik} b_{kj} \right)_{ij}$$

Note that $M_1(R) \cong R$, and $M_2(R)$ is *not* commutative, and always has zero divisors, for:

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

R can be embedded into R by taking by taking mapping it to the diagonal:

$$a \mapsto (a)_{ii}$$

where the matrix $(a)_{ii}$ has zero in all other entries. It should be check that the subring of diagonal entries with all the same value is a subring, sometimes denoted $\iota(R) \subseteq M_n(R)$ to emphasize that $\iota(R)$ is an embedding, or RI_n. It should be check that $I_n = \iota(1)$. If R is commutative, then:

$$Z(M_n(R)) = RI_n$$

that is, the center of the matrix ring is the diagonals with the same coefficients. Slightly more generally, the subset of diagonal matrices also form a subring., espec Diagonal matrices are rather common, and so they are often denoted

$$\text{diag}(a_1, a_2, ..., a_n) = (a_i)_{ii}$$

Using the diagonal embedding, we can define a functor $M_n(-) : \mathbf{Ring} \to \mathbf{Ring}$ mapping R to the diagonal and φ to $M_n(\varphi)$ that maps $(a_{ij})_{ij} \mapsto (\varphi(a_{ij}))_{ij}$ (the reader should feel free to check this is indeed a functor). Thus, there is the following diagram:

$$\begin{tikzcd} R \arrow[r, "\varphi"] \arrow[d, hook] & S \arrow[d, hook] \\ M_n(R) \arrow[r, "M_n(\varphi)"'] & M_n(S) \end{tikzcd}$$

Thus, if $S \subseteq R$, then $M_n(S) \subseteq M_n(R)$.

Matrix rings shall appear in many context, notably in the case when $R = k$ is a field in the theory of vector-spaces and when R is a division ring in representation theory.

9.2.4 Group Rings

We have seen how to take a set and turn it into a group in many ways (ex. trivial group $T(X)$, permutation group S_X, free group $F(X)$). Of these methods the free-group construction gave us a way to lift a set map to a group map. If you have read section 2.3.5, **Set** and **Grp** form a free-forgetful adjoint, which tells us that than the "lifting" of set functions to group homomorphisms happened in a "natural way". We can repeat this process with groups and rings, forming a "free ring" with respect to groups. This "free ring" will not be free with respect to *sets*, and so will have a different flavour to that of free groups[13]. However, after going through free rings with respect to groups, the reader might start seeing why both constructions are called *free*.

Definition 9.2.14: Group Ring

Let R be a commutative ring and let G be any group, where $g_0 = e$ is the identity, with the operation written multiplicity. Then define the *group ring* $R[G]$ (or RG) of G with respect to R to be the set of all formal sums:

$$\sum_{g_i \in G} a_i g_i$$

where only finitely many a_i are non-zero where addition is done point wise and the product is defined per monomial as $(ag_i)(bg_j) = (ab)g_k$ where ab is performed in R and $g_i g_j = g_k$, and then extended this is extended to multiplication of the formal sums using convolution.

Since $g_0 = e$, we often write $a_0 g_0 = a_0$. In this way, we can identify $R \subseteq R[G]$, and R in as a subring of $R[G]$ has it's original ring operations. Similarly, taking $a_{ij} = \delta_{ij}$[14] where

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

we can identify $G \subseteq R[G]$, and G as a subgroup of $R[G]$ has it's original group operation. Even better, $G \leqslant R[G]^\times$. It is an easy exercise to check that $R[G]$ is indeed a ring, and that $R[G]$ is commutative if and only if G is commutative. Another simple exercise to check is that R (identified

[13] We will cover free rings later
[14] this is called the Kronecker delta

as a subring of $R[G]$) is in the center of $R[G]$ ($R \subseteq Z(G)$), and that $1 \in R$ (identified as a subring of $R[G]$) is the identity of $R[G]$

The reader may see that group rings are indeed the free objects with respect to the forgetful functor **Ring** $\to$ **Grp** by repeating the construction given for free groups: Take the collection of set maps $f : G \to R$ where only finitely many $f(g)$ are nonzero (such maps are said to have *finite support*[15]). Then two elements are added pointwise, $h \mapsto f(h) + g(h)$, and multiplication has the following clever interpretation:

$$h \mapsto \sum_{ab=h} f(a)g(b) = \sum_{a \in G} f(a)g(a^{-1}h)$$

Notice this sum is indeed finite since f, g have finite support.

Example 9.2.4: Group Rings

1. Let $R = \mathbb{Z}$ and $G = D_8$. Then, for example $3r^2 - s, rs + 9 \in \mathbb{Z}[D_8]$.
2. If G is any finite subgroup, and $S \subseteq R$ is a subring, then $S[G] \subseteq R[G]$ is a subring.
3. (If you know about the Hamiltonian's $\mathbb{H}$), then $\mathbb{R}[Q_8] \neq \mathbb{H}$, even though they both conains Q_8. Can you see why?
4. Let $R = \mathbb{Q}$ and $G = S_3$. Let $a = 3(2\ 3) - 4(1\ 2\ 3)$ and $b = 1/2(1\ 2) + 5(3\ 2\ 1)$. Compute:

$$a + b \qquad ab$$

Note that if G has an element of finite order, then the group ring always has zero divisors: let $g \in G$ such that $m = |g| > 1$. Then

$$(1 - g)(1 + ... + g^{m-1}) = 1 - g^m = 1 - 1 = 0$$

Exercise 9.2.1

1. Is $R[x]$ isomorphic to $\oplus_{i=1}^{\infty} R$? Is $R[[x]]$ isomorphic to $\prod_{i=1}^{\infty} R$?)
2. Let $f \big| g$. Does this imply $\deg(f) \big| \deg(g)$?
3. Let R be a **non**-commmutative ring. Define the appropriate ring so that for any set function $f : A \to S$, there is a unique map $\iota : A \to R$ to the ring $F : R \to S$ such that the diagram commutes. That is, define the free object in the general case so that it is left-adjoint to the forgetful functor **Ring** $\to$ **Set**.
4. Take the full sub-category **Poly-Ring** $\subseteq$ **Ring**. What is the inital object of **Poly-Ring**?
5. Let $p(x) \in R[x]$ be a polynomial such that for all $r \in R$,

$$p(r) = a_n r^n + \cdots + a_1 r + a_0 \neq 0$$

Show that there is an extension $S \supseteq R$ such that there exists $s \in S$ where $p(s) = 0$. In particular, there exists an S such that there exists a natural inclusion $\iota : R \hookrightarrow S$, $p(x)$ naturally

[15] For those who know some calculus/analysis, this is similar to how many smooth functions are defined on compact support

includes in S and some element $s \in S$ is a root to $p(x)$:

$$p(s) = a_n s^n + \cdots + a_1 s + a_0 = 0$$

6. Show that $R[\mathbb{N}] \cong R[x]$. Show that $R[\mathbb{Z}] \cong R[x, x^{-1}]$
7. Is $\mathbb{Z}[x, x^{-1}] \cong \mathbb{Z}[[x]]$

9.3 Ring Homomorphisms and Quotient Rings

In this section, we explore the structure of the quotient objects of rings, that is, if $\varphi : R \to S$ is a ring-homomorphism, we aim to find a way to talk about the structure of $\varphi(R) \subseteq S$ given structure for R. We first recall the definition of ring homomorphisms from the first section:

Definition 9.3.1: Ring Homomorphism Remainder

Let R and S be rings. Then:

1. *ring homomorphism* is a map $\psi : R \to S$ satisfying
   (a) $\varphi(a + b) = \varphi(a) + \varphi(b)$
   (b) $\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$
   (c) $\varphi(1_R) = 1_S$
2. The *kernel* of the ring homomorphism φ, denoted $\ker(\varphi)$ is the set of elements of R that map to 0 in S
3. A bijective homomorphism is an *isomorphism*.

If the context is clear, "homomorphism" will be used instead of "ring homomorphism". Also, if A, B are rings, then $A \cong B$ will always be interpreted as an isomorphism between the two rings.

Example 9.3.1

1. $\pi : Z \to \mathbb{Z}/2\mathbb{Z}$, here $\pi(x) = [x]$ is the natural projection map, is a ring homomrophism
2. $\varphi_n : Z \to \mathbb{Z}$, $\varphi_n(x) = nx$ is a ring-homomorphism.
3. $\varphi : \mathbb{Q}[x] \to \mathbb{Q}$, $\varphi(p(x)) = p(0)$ is a ring homomrophsim, here $\ker \varphi = \{a_n x^n + \cdots + a_1 x\}$. Notice that every elements in the kernel has an x. We can also think of this as $p(x) \cdot x$ for any $p(x) \in \mathbb{Q}[x]$. This way of looking at it will become the main way we look at the kernel of a ring homomorphism shortly.
4. $\mathbb{R}^2$ and $\mathbb{C}$ are *not* [ring] isomorphic! In $\mathbb{C}$, we have an element such that $a^4 = 1$, that is $i^4 = -1$. However, no such element exists in $\mathbb{R}^2$. Thus, even though these two are very similar (ex. similar dimension[a]), they *key* difference is the algebraic structre!
5. As shon earlier, if $S \subseteq R$ has a ring-structure, then if the inclusion map $i : S \to \mathbb{R}$ is a [ring] homomoprhism, then S is a subring.

6. The inclusion map $\mathbb{Z} \hookrightarrow \mathbb{Q}$ is a ring homomorphism. Is there a non-trivial homomorphim from $\mathbb{Q}$ to $\mathbb{Z}$?

7. Embedd $k[x] \hookrightarrow k[[x]]$ by mapping $k[x]$ to be all formal power series with only finitely many nonzero elements. Then by proposition 9.2.8, any polynomial $p(x) \in k[x]$ such that $p(0) \neq 0$ under the evaluation map is invertible in $k[[x]]$. Hence, there is a power series that can be identified by $1/p(x)$. Thus, we may define polynomial quotients in $k[[x]]$

$$\frac{p(x)}{q(x)} \qquad q(0) \neq 0$$

8. Is the derivative map $\frac{d}{dx} : k[[x]] \to k[[x]]$ defined by

$$\frac{d}{dx}\left(\sum_{k=0}^{\infty} a_k x^k\right) = \sum_{k=0}^{\infty} k a_k x^{k-1}$$

   a ring homomorphism? what is $\frac{d}{dx}(fg)$?

9. Let $k[[x]]$ and $k[[y]]$ be two formal power series over a field k. Let $T \in k[[y]]$ such that $b_0 = 0$. If we define $f \circ g$ to be (HERE). Then $f \mapsto f \circ T$ is a ring homomorphism from $k[[x]]$ to $k[[y]]$. What happens if $b_0 \neq 0$?

[a] If you did topology, these two with their standard topologies are homeomorphic

Just as the study of group-homomorphisms is equivalent to finding a normal subgroup H and studying G/H, so too will there be a link between ring-homomorphisms and a special subset I of R for which R/I represents the homomorphism. Just like for group-homomorphisms, the set I is based on the kernel of the ring-homomorphism, and R/I will be isomorphic to the image of φ:

Proposition 9.3.2: Image and Kernel of Ring Homomorphism

Let R, S be rings and $\psi : R \to S$ be a homomorphism.

1. The image of ψ is a sub-ring of S (as should be clear as it is one of the "properties" that are sought after for defining a homomorphism)

2. the kernel of ψ is, in general, *not* a subring of R. Rather, the kernel of ψ retains the abelian structure of R (i.e. is closed under adition an $a - b \in I$), and is closder under multiplication (left/right multiplication)[a] from elements of R (for all $r \in R$ and $i \in I$, $ri \in \ker(\psi)$ and $ir \in \ker(\psi)$).

[a] note that it was specified that it was both left and right multiplication because the multiplication is not necessarily commutative, while addition is

As an example of a kernel, take $\varphi : \mathbb{Z} \to \mathbb{Z}$, $x \mapsto 2x$. Then $\ker(\varphi) = 2\mathbb{Z}$, which is not a ring. Kernels are the more general object *rng*. Note that if we define a ring to not have an identity, then the kernel would indeed be a sub-ring of R. Some authors define rings to not have an identity for this very reason. However, as a consequence, almost every theorem will have to be prefaced with "let R be a ring with identity", meaning the study of ring theory will ultimately be the study of a special case of rings with identity, which begs the question whether broadening the definition to not have an

identity is worth the trade-off. However one sees it, it is simply a mathematical reality that kernels are different objects than rings (as the proposition suggests), and this author belives it is better to treat them as such[16]. Kernels will soon be defined as *ideals* (definition 9.3.3) since they have different properties from rings.

Proof **:**

1. This should be an easy exercise to make sure homomorphisms can be comfortably worked with

2. Let $I = \ker \varphi$ and take $a, b \in I$. Then

$$\varphi(a - b) = \varphi(a) - \varphi(b) = 0 \Rightarrow a - b \in I$$

showing that I is closed addition, and in particular is an abelian group.

Now take $r \in R$. We want to show that for any $i \in I$, $ri, ir \in I$. The key to this proof is the 1st property of property 9.1.5. Let $I = \ker \varphi$. Then

$$\varphi(ri) = \varphi(r)\varphi(i) = \varphi(r) \cdot 0 = 0 \Rightarrow ri \in I$$

and similarly

$$\varphi(ir) = \varphi(i)\varphi(r) = 0 \cdot \varphi(r) = 0 \Rightarrow ir \in I$$

Thus, I is closed under left-right multiplication of R.

23: Remark

(label these remarks in the future for ease of reference ringExtRemark)
As a consequence of kernels not being subrings, the concept of extensions as discussed in section 3.6 doen't apply so simply to rings! A group extension relied on the fact that both N and G/N were groups. Since ideals are not rings, the same concept does not apply for rings! There is an analogous concept of a "ring extension"[a], which generlize the notion of an extension by just requiring a ring extension S of R to be that there exists an injective ring homomorphism $\varphi : R \to S$ (similarly to how if $N \trianglelefteq G$ then we always have that $N \hookrightarrow G$). This notion of ring extension will be central to us in part V[b], but the "block-building" story that happened for groups does not play out in the same way way for rings. However, in part III, we will see that any ring R is a natural R-modules, and modules will have extensions defined on them [c]. Thus, we may study the extensions of rings in a group-theoretic way in the context of module theory.

[a] not to be confused with ring extensions that extend a ring by an abelian group as described in Nagata Masayoshi's book *Local Rings* (see wikipedia article)
[b] It will be the fundamental way we look at varieties and number fields
[c] The category of R-module is closed under kernels

This does not mean that we cannot define $R/\ker(\varphi)$. In fact, $R/\ker(\varphi)$ will be a well-defined ring, as we explore now. Note that since R is an abelian group, and φ is also always a group-homomorphism, $\ker \varphi$ is always be a normal group. If we label $I = \ker(\varphi)$ then R/I already has a group structure,

[16] The fact that the kernel is not a ring means that the category of **Ring** does not have kernel objects

that is:

$$(r+I)+(s+I)=(r+s)+I$$

is well defined. The question becomes if we can also add a multiplicative structure to the quotient group, that is:

$$(r+I)\cdot(s+I)=(rs)+I \qquad \text{(in equivalence class notation: } [x\cdot y]=[x]\cdot[y])$$

With I being the kernel of a ring homomorphism, this result is immediate (i.e. it is well-defined). For $i,k\in I$ and $r,s\in R$:

$$\varphi((r+i)(s+k))=\varphi(r+i)\varphi(s+k)=(\varphi(r)+\varphi(i))(\varphi(s)+\varphi(k))=\varphi(r)\varphi(s)+0+0+0=\varphi(r)\varphi(s)$$

where we used both the additive and multiplicative linearity condition. Labelling $[x]=x+I$, we call $(R/I,+_\sim,\cdot_\sim)$ with $+_\sim$ and $\cdot_\sim$ being defined through φ the *quotient ring* [17]. Thus, I being the kernel of a ring is a sufficient condition to make R/I into a ring. Is it a necessary condition? That is, can there be weaker structures that still preserves the multiplicative structure as shown above? For example, can I simply be a normal subgroup and always preserve the multiplicative structure? The answer is no:

Example 9.3.2: : Quotient group without preserving multiplicative property

Take $\mathbb{Q}/\mathbb{Z}$ as a quotient group. Since $\mathbb{Q}$ is abelian, $\mathbb{Q}/\mathbb{Z}$ is a well-defined quotient group. Let's say it was a well-defined quotient ring. Consider

$$\left(\frac{a}{n}+\mathbb{Z}\right)=(a+\mathbb{Z})\left(\frac{1}{n}=\mathbb{Z}\right)=(0+\mathbb{Z})\left(\frac{1}{n}+\mathbb{Z}\right)=[0]$$

so, every element greater than 1 is zero. Through this, we can make the rest of the elements of $\mathbb{Q}/\mathbb{Z}$ zero. Take $\frac{1}{n}$. Then

$$[0]=\left[0\cdot\frac{1}{n}\right][0]\left[\frac{1}{n}\right]=[1]\left[\frac{1}{n}\right]=\left[1\cdot\frac{1}{n}\right]=\left[\frac{1}{n}\right]$$

Which means that multiplication forced $(\frac{a}{b}+\mathbb{Z})=(0+\mathbb{Z})$, meaning that

$$(r+\mathbb{Z})(s+\mathbb{Z})\neq(rs+\mathbb{Z})$$

unless $(r+I)=[0]$ or $(s+I)=0$, meaning there is no salvageable sub-ring we can which would make this work.

So we must figure out what properties I requires for closure of multiplication. Let's say that I is a normal subgroup, and let's work with the cosets R/I. Let's say that multiplication of cosets is well defined. That would imply that

$$[r+i][s+j]=[rs] \qquad i,j\in I$$

that is, multiplication is independent of representation. If $r=s=0$, then that means that

$$(i+I)(j+I)\in I\Rightarrow ij\in I$$

[17] as before we'll delay defining the quotient object till we find the sufficient conditions on I independent of a ring homomorphism

meaning that I is closed under multiplication, making it a sub-ring. Next, what if $s = 0$ and r is arbitrary. Then that would lead to $[r][0] = [0]$ by proposition 9.1.5, which, written in terms of cosets, will be:

$$\begin{aligned}
&((r+i)+I)\cdot(j+I) = I \\
\Rightarrow &(rj+ij)+I = I \\
\Rightarrow &(ri)+I = I \\
\Rightarrow &rb \in I
\end{aligned}$$

and since r is any arbitrary element $r \in R$, then I must be closed under left multiplication of any element of R. Similarly if $r = 0$, meaning $is \in I$.

Conversely, let's say I is closed under multiplication from the left and right-side: $\forall i \in I$, $ri \in I$, $ir \in I$ $\forall r \in R$. Then

$$\begin{aligned}
(r+i)+I\cdot(s+j)+I &= (r+i)(s+j)+I \\
&= (rs+rj+is+ij)+I \\
&= (rs+I)+(rj+I)+(is+I)+(ij+I)
\end{aligned}$$

and since I is a sub-ring, $ij \in I$, and since I is closed under left and right multiplication, $rj, is \in I$, meaning we are left with:

$$= (rs)+I$$

meaning we are once again representative independent. Thus, we showed that left and right closure of I is a necessary *and* sufficient condition for the cosets defined by I to have a 2nd operation.

As with cosets, the equivalence classes induced by ring homomorphisms are given a special name:

Definition 9.3.3: Ideals

Let R be a ring, and let I be a subset of R, and let $r \in R$. Then:

1. $rI = \{ra | a \in I\}$, and $Ir = \{ar | a \in I\}$
2. a subset I of R is a *left ideal* if I is closed under *left* multiplication by elements from R ($rI \subseteq I$ for all $r \in R$)
3. Similarly, I is a *right* ideal if I is closed under *right* multiplication by elements from R ($Ir \subseteq I$ for all $r \in R$)
4. A subset I that is *both* a left and right ideal is called an *ideal* (or, for added emphasis, a *two-sided* ideal) of R

Example 9.3.3

1. Take $m \in \mathbb{Z}$. Then $m\mathbb{Z} \subseteq \mathbb{Z}$ is an ideal. Similarly for $\mathbb{Z}m$.
2. Take $m\mathbb{Z} \subseteq \mathbb{R}$. Then $m\mathbb{Z}$ is not not an ideal (since $\frac{1}{2m}mk \notin m\mathbb{Z}$).
3. In $\mathbb{Z}_4 \times \mathbb{Z}_4$, $\langle(1,1)\rangle$ is a sub-ring, but not an ideal
4. Take any product of two rings $R \times S$. Then $R \times 0 \subseteq \times S$ is an ideal, since for any $(a, 0) \in R \times 0$,

> $(R \times S) \cdot (a, 0) \subseteq R \times 0$. Notice that $R \times 0$ is also a subring of $R \times S$ meaning it contains an identity. This is the *only* exception in which a subring an ideal can contain an identity.

(A word somewhere on how ideals define the relatinships in rings, similarly to how normal groups did for rings. We do not use the same angle-braket notation $\langle\ \rangle$ as we've done for groups since this notation will be used for a different purpose (note to self: in constructing free-algebras. This genelizes better the angle-bracket notation since all ring's are subsumed as an algebra by being $\mathbb{Z}$-algebras))

24: Trivia

Why is it called an ideal? In number theory, people tried finding more information about polynomials like

$$x^2 - d = 0 \Rightarrow x = \sqrt{d}$$

And we would construct the field $\mathbb{Q}(\sqrt{d}) = \left\{a + b\sqrt{d} | a, b \in \mathbb{Q}\right\}$ which was called $\mathbb{Q}$ with $\sqrt{d}$ adjoint. This ring (in fact field) has a sub-ring $\mathbb{Z}[\sqrt{d}]$ and Q, making the diagram:

$$\begin{tikzcd}
\mathbb{Q}[\sqrt{d}] \arrow[d] \arrow[dr] & \\
\mathbb{Z}[\sqrt{d}] \arrow[dr] & \mathbb{Q} \arrow[d] \\
& \mathbb{Z}
\end{tikzcd}$$

So, with this construction, we have the question of developping the concept of prime number in $\mathbb{Z}$ to it in $\mathbb{Z}[\sqrt{d}]$. If you try to take the same idea (a element that has no divisors in the ring except the element and ± 1). But it turns that if you try to define an element of that, elements don't necessarily have a unique factorization!
Example
This is so fundamental that this was dropped as a good definition as prime element in $\mathbb{Z}[\sqrt{d}]$. Insead, the proper generalization of this concept will turn out to be ideals in $\mathbb{Z}$ which will be called *prime ideals*, and then look at their natural extension into $\mathbb{Z}[\sqrt{d}]$. The word "ideal" here comes from how these elements would be the *ideal numbers*. The word number will eventaully get dropped, leaving us with ideal.

25: Computational Shortcut

To prove that I is an ideal, we must show that I is non-empty, that it's closed under subtraction ($\forall r \in R$), closed under multiplication ($\forall r \in R$), and that it's a normal subgroup. If R has a 1, then you can replace "closed under subtraction" to "closed under multiplication" since $(-1)a = -a$ will cover the negative case

Another reason to treat ideals differently than rings is the following:

> Another way of how ideals are different form rings is that being a ideal is "ring-dependent". For example $(x) = \mathbb{R}[x]x$ is an ideal of $\mathbb{R}[x]$, but not an ideal of $\mathbb{C}[x]$, since $ix + 1$ notin $(x) = \mathbb{R}[x]x$. This points again to how ideals are different.

This discussion is summarized in the following proposition:

Definition 9.3.4: Quotient Ring

When I is an ideal of R. The ring R/I with the natural multiplicative operation is called the *quotient ring* of R by I.

Proposition 9.3.5: Quotient groups and Ring Homomorphism

Let R be a ring and let I be an ideal of I. Then the (additive) quotient group R/I is a ring under the binary operation:

$$(r+I)+(s+I)=(r+s)+I \qquad \text{and} \qquad (r+I)\cdot(s+I)=(r\cdot s)+I$$

for all $r, s \in \mathbb{R}$. Conversely, if I is any subgroup such that the above operations are well defined, then I is an ideal of R.

Just like for groups, we can ask if every ideal is the kernel of a ring homomorphism, essentially asking for the 1st isomorphism theorem for rings (i.e. the universal property of quotient rings). This turns out to be the case:

Theorem 9.3.6: The first isomorphism theorem for Rings

1. If $\varphi : R \to S$ is a homomorphism of rings, then the kernel of φ is an ideal of R, the image of φ is a subring of S and $R/\ker(phi)$ is isomorphic as a ring of $\varphi(R)$.
2. If I is any ideal of R, then the map

$$R \to R/I \qquad \text{defined by} \qquad r \mapsto r+I$$

is a surjective ring homomorphism with kernel I (this homomorphism is called the *natural projection* of R onto R/I). Thus every ideal is the kernel of a ring homomorphism and vice versa.

Proof **:**
This proof is similar to the fisrt isomorphism theorem (theorem 3.1.12), and we already did a good chunk of it in the preceeding paragarphs. The proof of the natural projection follows very similar lines to theorem 3.1.12.

As with groups, the bar notation $\overline{r}$ or r mod I notation is used to represent the equivalence class $[r]$.

Example 9.3.4

1. R and $\{0\}$ are ideals. An ideal is proper if $I \neq R$. The set with only zero is called the *trivial* ideal.
2. $n\mathbb{Z}$ is an ideal of $\mathbb{Z}$.
3. Take $R = \mathbb{Z}[x]$, and let I be the subset whose polynomials are of degree at least 2. Note

that I is closed under addition and multiplication, and thus an ideal. The cosets consist of the elements with the same 0th and 1st term, so

$$\overline{1+2x+x^2} = \overline{1+2x-5x^3} = \overline{1+2x+x^6+87x^{1231}}$$

Note that $\overline{x}\overline{x} = \overline{x^2} = \overline{0}$, thus R/I has a zero divisor, even though $\mathbb{Z}[x]$ does not .

4. We saw $\mathbb{Q}[x]/(x)$. Let's try another polynomial: $\mathbb{Q}[x]/(x^2-2)$. Since we can divide polynomials, we can divide any polynomial by x^2-2. We can in fact keep dividing it till we get a polynomial of lesser degree. We will actually then get that

$$\mathbb{Q}[x]/(x^2-2) = \{ax+b|a,b \in \mathbb{Q}\}$$

Note that x^2-2, so $x^2-2=0 \Rightarrow x = \pm\sqrt{2}$. So, since $[x^2-2]=[0]$, this will make

$$\mathbb{Q}[x]/(x^2-2) \cong \mathbb{Q}(\sqrt{2}) = \left\{x+y\sqrt{2}|x,y \in \mathbb{Q}\right\}$$

Or, more specifically, if we take the ring-homomorphism $\varphi : \mathbb{Q}[x] \to \mathbb{Q}(\sqrt{2})$, $p(x) \mapsto p(\sqrt{2})$, then This is a isomorphism.

5. If we take $k[x,y]$, then $k[x,y]/(x) \cong k[y]$

6. Let $G = \langle g \rangle$ be a cyclic group and k be any field. Define:

$$\varphi : k[G] \to k[x] \qquad \varphi(ag^k) = ax^k \qquad \forall a \in k$$

Show that:

$$k[G] \cong \frac{k[x]}{x^n-1}$$

Example 9.3.5

Let's say you want to figure out all integer solutions to

$$x^2+y^2=3z^2$$

Suppose such integers exist. Observe first that we may assume x, y and z have no factors in common, since otherwise we could divide through this equation by the square of this common factor and obtain another set of integer solutions smaller than the initial ones. This equation simply states a relation between these elements in the ring $\mathbb{Z}$. As such, the same relation must also hold in any quotient ring as well. In particular, this relation must hold in $\mathbb{Z}/n\mathbb{Z}$ for any integer n. The choice $n=4$ is particularly efficacious, for the following reason: the squares mod 4 are just 0^2, 1^2, 2^2, 3^2, i.e., 0, 1 (mod 4). Reading the above equation mod 4 (that is, considering this equation in the quotient ring $\mathbb{Z}/4\mathbb{Z}$), we must have

$$\begin{Bmatrix}0\\1\end{Bmatrix} + \begin{Bmatrix}0\\1\end{Bmatrix} = + \begin{Bmatrix}0\\3\end{Bmatrix} \qquad (\text{mod}4)$$

where the $\begin{Bmatrix}0\\1\end{Bmatrix}$ indicate that either a 0 or a 1 may be taken. The only possibility is if they're all even, but we assumed they weren't, thus there is no solution.

Note that this technique has limitation. The following equation has solution integers modulo any n, but no integer solutions (which is hard to prove).

$$3x^3 + 4x^3 + 5z^3 = 0$$

Eventually, we'll be using a similar technique to factor polynomial.

There is also the 2nd, 3rd, and 4th isomorphisms theorems. Each of these are quite easy to show: We already know they're true because R is an *additive group* (or *correspondence of groups*, in the case of the Fourth Isomorphism theorem). It only needs to be show that the group isomorphism is also a multiplicative map, and thus a ring-isomorphism.

Beforehand, we had $H, K \leqslant G$ and defined HK. Since we are now writing the group operation additively, we will convert our notation to saying $H + K$. If I, J are subrings of R, then since they're both normal (even abelian) subgroups, then $I + J$ is well-defined.

Theorem 9.3.7: 2nd, 3rd, 4th Isomorphism Theorem for Rings

(2) Let A be a subring and B be an ideal of R. Then $A + B$ is a subring of R, $A \cap B$ is an idea of A, and $\frac{A+B}{B} \cong \frac{A}{A \cap B}$

(3) Let I, J be ideals of R iwth $I \subseteq J$. Then J/I is an ideal of R/I and $(R/I)/(J/I) = R/I$.

(4) Let I be an ideal of F. The correspondence $A \leftrightarrow A/I$ is an inclusion preserving bijetion between the set of subrings A of R that contian I and the set of subrings of R/I. Furthermore, A (a subring containin I) is an ideal of R if and only if A/I is an ideal of R/I

***Proof* :**
The proof of each of these is an upgrade from the proof for the group version. Proving the group-isomorphism is immediate, then proving multiplicative linearity, it's an immediate consequence of the definition of the quotient ring.

(A word on simple ring, i.e. rings with no *two-sided* right-sided ideals? A word on $M_n(k)$ is simple, since no two-sided ideals (since two ided of ideals of $M_n(k)$ are of the form $M_n(I)$), but there exist left-sided ideals (ex. set of matrices with some fixed zero column) and right-sided ideals (set of with fixed zero row))

26: Remark

simpleCommRingIffFieldRmk
We can define a *simple ring* to be a ring whose only two-sided ideals are R and itself[a], which mirrors the definition of simple groups. Then fields *are* all simple commutative rings[b]. For the non-commutative case, it is harder to find simple rings. (see exercise ref:HERE for an example, or proposition 23.2.5).

[a] the reason for sticking to two-sided ideals is because we can quotient R by two-sided ideals to get a quotient ring, but we cannot quotient by by 1-sided ideals and get a quotient ring

[b] The story is a bit different for *rngs*; see comment in document

Exercise 9.3.1

1. Let R, S be rings and consider the product ring $R \times S$. What are the ideals of $R \times S$?
2. Let I be an ideal of the ring R and let $(I) = I[x]$ denote the ideal of $R[x]$ generated by I (the set of polynomials with coefficient I). Then

$$R[x]/(I) \cong (R/I)[x]$$

   In particular, if I is a prime ideal of R then (I) is a prime ideal of $R[x]$. Conclude that $\mathbb{Z}[x]/(n\mathbb{Z}[x]) \cong (\mathbb{Z}/n\mathbb{Z})[x]$
3. Show that the set of upper-triangular matrices is a subring.

9.4 Properties and Further Structures of Ideals

Since the kernel of ring homomorphisms are ideals, they form the core of much of ring theory. Recall how when working with groups, they can be thought of as $F(A)/N$ for an appropriate normal set N which represents the *relations* on the generators A. Ideals will play a role of similar importance: Ideals will represent the relations. There will be an analogous notion of free ring, however the study of generation of free rings shall be delayed to part III since the two constructions are identical.

This is not to say that thinking of ideals as adding relations is un-useful! For example, since $\mathbb{Z}$ is initial in rings, it is of great interest to see how what relations ideals of $\mathbb{Z}$ can give us, what properties do they have, and how they translate with ring homomorphisms. (We will focus a lot on this in number theory).

Furthermore, closure is not well-defined on subrings, that is the "closure" of a set $A \subseteq R$, that is the smallest largest set that makes $+$ and $\cdot$ well-defined, need not be a ring. Take for example $\{2\} \subseteq \mathbb{Z}$, then if $\overline{\{2\}}$ is the "ring closure" of $\{2\}$, then $\overline{\{2\}} = (2)$, which is not a ring. Another example is $\{1 + x\} \subseteq \mathbb{Z}[x]$. Then $\overline{\{1 + x\}}$ is not a ring, since $1 \notin \overline{\{1 + x\}}$[18], and it is not an ideal since $(x + 2)(x + 1) \notin \overline{\{x + 1\}}$[19]. It is a rng, and so it makes senes to define finitely generated rngs, however due to the reason stated in remark ref:HERE, we continue to focus on rings.

9.4.1 Operations and Generation on Ideals

The first thing we might think of is way of extending how to manipulates ideals through binary operations (in essense, we find different ways of combining what relations we impose on a ring). We give a list of common operations of ideals

[18]This is actually a bit difficult to prove, that no polynomial $p(y)$ exists such that $p(1+x) = 0$, and so no polynomial $q = p + 1$for which $q(x + 1) = 1$. The full answer will be given in part IV when discussing transcendental extensions.

[19]To see this, we require the fact that every polynomials factors uniquely into irreducibles, and so it's impossible that $p(x + 1) = (x + 2)(x + 1)$ for any $p(y)$. We will show this in chapter 11

Definition 9.4.1: Operations on Ideals

Let R be a ring and $I, J \subseteq R$ be (left, right, or two-sided) ideals. Then:

1. $I \cap J$ is the intersection of two ideals. More generally,
$$\bigcap_{i \in \mathcal{I}} I_i$$
is the intersection of ideals I_i indexed by $\mathcal{I}$.

2. Define the *sum* of I and J by $I + J = \{a + b | a \in I, b \in J\}$. More generally,
$$\sum_{i \in \mathcal{I}} I_i = \left\{ \sum_{i \in N} a_i \;\middle|\; N \subseteq \mathcal{I},\ |N| \in \mathbb{N},\ a_i \in I_i \right\}$$
is the sums of ideals I_i indexed by $\mathcal{I}$.

3. Define the *product* of I and J, denoted IJ, to be the set of all finite sums of elements of he form ab wit $a \in I$ and $b \in J$.
$$\left\{ \sum_{i=1}^{n} a_i b_i \middle| a_i \in I,\ b_i \in J \right\}$$
Note that we must take finite sums since the ideal must be closed under addition. More generally:
$$\prod_{i \in \mathcal{I}} I_i = \left\{ \sum_{i=1}^{m} a_{i_1} \cdots a_{i_n} \;\middle|\; a_{i_1}, ..., a_{i_n} \in \mathcal{I}, n, m \in \mathbb{N} \right\}$$

4. For any $n \geq 1$, define the nth power of I, denoted I^n to be the set consisting of all finite sums of elements of the form $a_1 a_2 \cdots a_n$ with $a_i \in I$ for all i¿ Equivalently, I^n is defined inductively by defining $I^1 = I$, and $I^n = II^{n-1}$ for $n = 2, 3, ...$ By convention $I^0 = R$.

5. Let R be commutative. Define the *ideal quotient* $(I : J)$ to be:
$$(I : J) := \{x \in R \mid rI \subseteq J\}$$

Note that if I, J are *both* ideals, then $I + J$ is an ideal (this set is closed under sums and $r(a + b) = ra + rb \in I + J$). It should be emphasized that the set IJ is the *finite sum* of elements of the form ab. This condition is required so that IJ is closed under addition, which is necessary for IJ to be an ideal. Just like groups and rings, the union of two ideals is not an ideal (for example, $2\mathbb{Z} \cup 3\mathbb{Z}$ does not contain 5, however $2 \in 2\mathbb{Z}$, $3 \in 3\mathbb{Z}$). Can you show that the quotient ideal $(I : J)$ is indeed an ideal? In some very lose way, if $a \in (x)$, then that means that x is a "factor" of a in the sense that $a = rx$ for some $r \in R$)

Example 9.4.1

1. Let $I = 6\mathbb{Z}$ and $J = 10\mathbb{Z}$. Then $I + J$ is the ideal of all integers of the form $6x + 10y$ for $x, y \in \mathbb{Z}$. Since 6 an 10 are divisible by 2, meaning that $6x + 10y = 2(3x + 5y)$, so that

$2\mathbb{Z} \supseteq 6\mathbb{Z} + 10\mathbb{Z}$. Conversely, we see that $2 = 6(2) + 10(-1)$, so $2x = 6(2x) + 10(-x)$, meaning that every element of $2\mathbb{Z}$ can be written of the form $6\mathbb{Z} + 10\mathbb{Z}$, so $2\mathbb{Z} \subseteq 6\mathbb{Z} + 10\mathbb{Z}$. Thus $2\mathbb{Z} = 6\mathbb{Z} = 10\mathbb{Z}$. We can actually generalize this to $n\mathbb{Z} + m\mathbb{Z} = d\mathbb{Z}$ where $d = (m, n)$, in the next section. Similarly, it is worth checking that $6\mathbb{Z}10\mathbb{Z} = 60\mathbb{Z}$.

2. Take $I \subseteq \mathbb{Z}[x]$ to be the ideal whose constant term is even. Then (2) and (x) are in subsets of I. Thus, $4 = 2 \cdot 2$ and $x^2 = x \cdot x$ are in I^2, as is their sum $x^2 + 4$. However, $x^2 + 4$ cannot be written as a product $p(x)q(x)$ of two elements of I.

3. the ideal $((0) : I)$ known as

4. Let $(2), (6) \subseteq \mathbb{Z}$. Then $((2) : (6)) = (3)$, showing how the quotient ideal generlizes some notion of division. *annihilator of* I and consists of all elements of R such that $rx = 0$ for all $x \in I$. Usually, for any ideal $I = (r)$ generated by a single element, we write $(r : I)$ instead of $((r) : I)$. Using this, can you deduce what $(0 : R)$ is equal to?

We can also ask oursevles about the generation of ideals:

Definition 9.4.2: Generating Ideals

1. (A) is the smallest ideal containing A. Then

$$(A) = \cap_{A \subseteq I} I$$

where I is an ideal containing A.

2. RA is he set of all finite sums of elements of the form ra between R and A.

$$\{r_1 a_1 + \cdots + r_k a_k | r_i \in R\,, a_i \in A\}$$

Similarly for AR and RAR.

3. An ideal generated by a single element is a *principal ideal.*

4. an ideal generated by a finite set is called a *finitely generated ideal.*

If R is commutative, then $RA = AR = RAR = (A)$.

27: Moment To Ponder

You might ask yourself, with addition and multiplication defined on the ideals, does the set of ideals form a ring? Try it out! Is this a way to study ideals? More on studying the ideal structure of a ring when studying Modules in part III (actually, wouldn't referencing Picard's group $\mathrm{Pic}(R)$ be a better reference here)

In the following chapter's, ideals generate by a singleton (a) will be of keen interest. As a prelude, notice that if $b \in (a)$, then $b = ra$ for some $r \in R$. Phrased differently, a divides b. Also, $b \in (a)$ if and only if $(b) \subseteq (a)$. In other words, being an element inside the ideal gives some idea of "divisibility" or "factorization".

Example 9.4.2

1. (0) and R are both principal ideals, since (0) is generated by 0, and $R = (1)$ (since we assumed R has an identity).

2. Every ideal of $\mathbb{Z}$ is a principal ideal

   Proof. We'll show that $(m) + (n) = gcd(m, n)$

   ($\Rightarrow$) Let $x \in (m) + (n)$. Let's show that $x \in (m, n)$. Since $x \in (m) + (n)$, then:

   $$x = ma + nb \qquad a, b \in \mathbb{Z}$$

   letting $p = (m, n)$, we see that

   $$ma + nb = (m'a + n'b)p = cp \in (m, n)$$

   where m', n' are the remainders, and $c = m'a + n'b$, showing that $cp \in (m, n)$

   ($\Leftarrow$) Let $p = gcd(m, n)$. Without loss of generality, let's say $m > n$. We will take advantage of the euclidean algorithm. First, start with:

   $$m = a_1 n - r_1$$

   where r_1 is the remaidner by dividng m by n and a is the number of times n can be divided into m. This is the extended euclidean algorithm covered in MAT240. We can work our way down till we get

   $$r_k = a_{k+1}p - 0$$

   where $p = gcd(m, n)$. Then, we can use Bézouts identity from the results to get that:

   $$p = k_1 m - k_2 n$$

   with appropriate k_1, k_2. This then implies that

   $$cp = ck_1 m - ck_2 n$$

   for any $c \in \mathbb{Z}$, and so cp can be written of the form $am + bn$, meaning $cp \in (m) + (n)$, as we sought to show.

   We can actually use this to show that *any* (A) is a principle ideal in $\mathbb{Z}$. Since $\mathbb{Z}$ is commutative, we can think of $(A) = RA$. Then we can do the same greatest common denominator trick to get down to our desied number such that $d\mathbb{Z}$ is equal to $\mathbb{Z}A$. We can use this to show that every ideal of $\mathbb{Z}$ are principal ideals. Since $\mathbb{Z}$ is also an integral domain, meaning it's also a domain, we will end up calling $\mathbb{Z}$ a *principal ideal domain.*

   However, it's not always true that if (a, b) can be reduced to (c) that a ring is a principal ideal domain. There is a sub-class of rings called *Bézout Domains* for which every pair of number can be reduced to a principal ideal, but the ring itself is *not* a principal ideal domain. More on that in chapter 10. □

3. Let $R = \mathbb{Z}[x]$. Then $(2, x)$ is *not* a principal ideal. Let

$$(2, x) = \{2p(x) + xq(x) | p(x), q(x) \in \mathbb{Z}[x]\}$$

that is, all polynomials with *even* constants (and hence, a proper ideals, since it doesn't containt 1). Let's say for some $a(x) \in \mathbb{Z}[x].(2, x) = (a(x))$. Since $2 \in (a(x))$, there must be some $p(x)$ such that $2 = p(x)a(x)$. By proposition 9.2.2, we know that degree $p(x)a(x) =$ degree $p(x)+$ degree$a(x)$, and since degree(2) is 0, then $a(x)$ and $p(x)$ must *both* be constant polynomials. Since 2 is a prime number, the only possibilities are $p(x), a(x) \in \{\pm 1, \pm 2\}$. If $a(x) = \pm 1$, then $1 \in (a(x))$, and so by proposition 9.4.6, $(a(x)) = \mathbb{Z}[x]$. But that's a contradiction to $(2, x)$ being proper. Thus, it must be that $a(x) = \pm 2$. But then, $x \in (\pm 2)$, and so $x = p(x)(\pm 2)$. But this is impossible; we'd need $\frac{1}{2}x$ to be $p(x)$, but $\frac{1}{2} \notin \mathbb{Z}$. Thus, $(2, x)$ is *not* a principal ideal. Thus $\mathbb{Z}[x]$ is *not* a principal ideal domain.

I like to thinke of elements of (p, q) for some $R[x]$ as elements in which every term has p or q as a factor. This is because when taking $\alpha p + \beta q$ for $\alpha, \beta \in R[x]$, you can have α or β be 0, meaning you just need elements which have p or q as a factor.

4. If we take $\mathbb{Q}[x]$ instead of $\mathbb{Z}[x]$, then this would be solved since $(2) = \mathbb{Q}[x]$ since $\frac{1}{2}2 = 1$, maning (2) has a unit, meaning $(2, x)$ has a unit. In fact, $\mathbb{Q}[x]$ will become a principal ideal domain because of this! Try playing around with the following ideals to see they are indeed principal

   - $(2, x)$
   - $(x, x^2 + 1)$
   - $(p(x), q(x))$ for $p, q \in \mathbb{Q}[x]$.

5. Take $R = k[x]$ and take $I = \{p(x) \in k[x] | p(q) = 0\}$ for some arbitrarry $q \in \mathbb{F}$. One can check that it is an ideal. More interestingly though, we know from high-school math that if $p(x) = 0$, then we can factor out $p(x) = (x - q)r(x)$ for any polynomial $r(x) \in \mathbb{F}[x]$. This will thus make $I = ((x - q))$, making these ideals principal ideals! However, if $I = \{p(x) \in \mathbb{F}[x] | p(a) = p(b) = 0, a \neq b\}$ for two arbitrary $a, b \in F$, $a \neq b$, then I is *not* a principal ideal.

As the intuition for ideals suggest, they are a sort of generalization of properties of $\mathbb{Z}$. For example, if $a \in R$ and (a) is an ideal, then $R/(a)$ is well-defined. Now, take $\overline{b} \in R/(a)$ and generate $(\overline{b})$. Then it should be verified that

$$(R/(a)/(\overline{b}) = R/(a, b)$$

Furthermore, notice that in $\mathbb{Z}$, the ideal (a, b) is equal to the ideal $(\gcd(a, b))$. For this reason, ideals capture some more general sense of divisibility for rings! Exploring the notion of divisibility further will be the goal of chapter 10.

Since ideals are so important, it is often the case that studying rings with simpler ideal structure are of much interest. Two particular cases are especially important due to the abundance of finiteness:

Definition 9.4.3: Noetherian rings

Let R be a commutative ring. Then if every ideal[a] $I \subseteq R$ is finitely generated, then R is called a *Noetherian ring*.

[a]Be sure not to mistake ideal for ring, for finitely generated rings have not been defined

Noetherian rings are particularly nice in that they are finite locally and globally. If we just require that the ring is finitely generated, it may be that a subring is non-finitely generated. Conversely. We shall have lots of experience with Noetherian rings later, for now it is useful to know these rings since since being Noetherian is stable under taking the quotient, and is also stable under the "converse" of taking the quotient (if you've covered section 3.6, the Noetherian condition is stable under extensions). For the next proposition, note that a finitely generated ideal is defined identically to a finitely generated ring (that it is generated by finitely many elements):

Proposition 9.4.4

let R be a Noetherian ring and I be any ideal of R. Then R/I is finitely generated. Conversely, if I and R/I is finitely generated, then R is finitely generated.

Proof **:**
Make it a guided exercise, since it is easy to follow

One of the simplest Noetherian rings are those for which every ideal is not only finitely generated, but also principal. If the ideal is furthermore an integral domain (to avoid pathologies from zero-divisors) then it is a principal ideal domain:

Definition 9.4.5: Principal Ideal Domain

Let R be an integral domain. Then if every ideal $I \subseteq R$ is a principal ideal, then R is called a *Principal Ideal Domain*.

As a simple example, $\mathbb{Z}$ is a PID (can you see why?). It should be noted that *every* ideal is principal, even those that may seem infinitely-generated. In fact, there are domain in which only finitely-generated ideals are principle (such rings are called *Bézout* rings).

Another way of looking at principal ideal domains is that they are "locally free". Here, locality is taken to mean that any proper submodule is free. Since R is an integral domain, it is torsion free, and since R is principle, if $I \subseteq R$ is a submodule (i.e. ideal), then

$$I = (a) \qquad a \in R$$

This fact shall make PID's very powerful, and in fact the idea that PID's have "local freeness" shall be further explore in chapter 16[20].

(TBD – somehow incorporate this?) *Principal Ideal Domains* (PIDs) will have the property of having "prime elements" well-defined and lend itself useful in many classification theorems (for example, study of Jordan Canonical Forms in chapter 14 will heavily rely on this type of ring). (The point is to introduce them later in the chapter on factorization)

[20]A projective module over a PID is in fact a free module

Exercise 9.4.1

1. Michael Atiyah has some excellent elementary exercises!! Some of his propositions are also great: the sequence of propositions on radical ideals can totally be exercises here! p. 6-9.
2. $(I \cap J)(I + J) = IJ$

9.4.2 Ideals and Units

As briefly mentioned, some ideals seem to be related to a notion of divisibility. This intuition more clear with the following theorem:

Proposition 9.4.6: units in Ideals and ideals in fields

1. $I = R$ if and only if I contains a unit
2. R is a division ring if and only if it's only left- and right-sided ideals are 0 and itself. If R is commutative, then if the only ideals of R are 0 and itself, then R is a field.

Proof **:**

1. one inclusion is easy. The key part to the other inclusion is that I must be closed under multiplication from all elements of R. If $I = R$ then I contains the unit 1. If I contains a unit u, then given the inverse $v \in I$ and any element $r \in R$

$$r = r \cdot 1 = r(uv) = (rv)u \in I$$

   Since I is an ideal.

2. If R is a division ring, then every element is a unit, and thus for any set A $RA, AR \subseteq R$ must contain units, and so $I = R$. Conversly, if the only left and right-sided ideals are R and 0, then for any $u \in R$, $Ru = uR = R$ (since it's the only ideal). Then $1 \in (u)$ meaning $1 = uv = vu$ for some $v \in R$. Since this holds for every element, then R is also a division ring. If R is also commutative, it is a commutative division ring, and thus a field.

Note: As a consequence, if $1 \in I$, then $I = R$, since 1 is a unit!! This means that in a proper ideal, $1 \notin I$, and so, they cannot be sub-rings with identity $1 \neq 0$.

It should be pointed out it was necessary to have the condition that both let and right-sided ieals are trivial, instead of just picking one or saying that only two-sided ideals are trivial.

Example 9.4.3: Two-Sided Ideals Trivial, but non-trivial left/right-sided ideals

Let k be a field and $M_2(k)$ the matrix ring over k. Then it should be checked that $M_2(k)$ has non-trivial left and right-sided ideals (think of columns and rows of $M_2(k)$). However, $M_2(k)$ has *no* two-sided non-trivial ideals.

Now, $M_2(k)$ is not a division ring: the element

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

does not have an inverse (as can be shown using some lienar algebra).

(make an exercise somewhere finding the ideals of $M_n(R)$, namely $M_n(I)$)

Since any two-sided ideal is a left and right ideal, then division ring's naturally have no non-trivial two-sided ideals.

Corollary 9.4.7: Division Ring and Ring-Homomorphism

If R is a division ring, then any non-zero ring-homomorphism $\varphi : R \to S$ to any other ring must be injective.

Proof **:**
Let φ be a ring-homomrophism. The kernel of φ is a two-sided ideal of R, but R is a divison ring, meaning the only ideals of R are 0 and R. Since φ is non-trivial, $\ker \varphi \neq R$, and so $\ker \varphi = 0$, making φ injective.

This might make one think that studying division rings, or more particularly fields, in terms of homomorphisms is less useful. However, we will *still* use ring-homomorphisms on fields, since mapping a field injectively into another field will turn out to be quite useful: we can study how fields *embedd* into one another (this will be called a *field extensions*, which will be defined in part IV, and an *integral extension* for more general rings like division rings, which will be defined in part V).

If a ring does only have 0 and R as two-sided ideals, we have the analogous concept to simple groups

Definition 9.4.8: Simple Rings

Let R be a ring. Then R is said to be simple of it's only [left/right/two-sided] ideals are the 0 ideal and itself

One might think that from here, there will be a similar story as in group theory: A lot of time and effort spent on classifying all rings (and finite rings) given this idea of a simple ring. It turns out though that the classification of finite rings are in fact much much easier: Every finite simple ring is isomorphic to the ring $M_n(\mathbb{F}_q)$ over a finite field of order q (i.e., intimately related to matrix rings)! In fact, the study of simple rings is the setting of representation theory! We'll cover simple ring in part VI

9.4.3 Maximal Ideals

Many times, we might want if an ideals is *as large as possible* in a ring. That is, if I is an ideal, then there does not exist another J such that $I \subseteq J$ unless $J = R$

Definition 9.4.9: Maximal ideal

An ideal M in an arbitrary ring S is called a *maximal ideal* if $M \neq S$ and the only ideals containing M are M and S.

Example 9.4.4: Ring with maximal ideal

Take $\mathbb{Z}$ and $6\mathbb{Z}$. Since $6 = 2(x)$, $6\mathbb{Z} \subsetneq 2\mathbb{Z}$. However, since 2 is a prime, we cannot repeat this same trick twice. Thus, $2\mathbb{Z}$ is a maximal ideal. In general, $p\mathbb{Z}$ for prime p is a maximal ideal

All rings have a maximal ring with the exception of a ring with the trivial ring structure:

Example 9.4.5: Ring without maximal ideal

Take any group with a non-maximal group (ex $\mathbb{Q}$) and take the trivial ring structure ($ab = 0$).

The only restriction on a ring that requires it to have a maximal ideal is that $1 \neq 0$. The way we will prove this will require a form of "super-induction". The use of this proof technique requires the axiom of choice. For those who want to have an idea of Zorn's lemma, see section A.3, and for those who want to see how it links to the axiom of choice, see appendix A

Proposition 9.4.10: Ring with $1 \neq 0$ have Maximal ideals

Let R be a non-trivial ring and $I \subsetneq R$ be a proper ideal. Then I is contained in a maximal ideal

Proof **:**

Since $I \subsetneq R$, $1 \neq 0$), and $1 \notin I$ (or else $I = R$), a maximal ideal can exist. To show that it is the case, we will us Zorn's Lemma (axiom A.2.3). Let $\mathcal{I}$ be the collection of all proper ideals of R containing I, and let it be ordered by inclusion. Then $(\mathcal{I}, \subseteq)$ is a partial order. If e take any chain $\mathcal{C} \subseteq \mathcal{S}$, then

$$J = \bigcup_{A \in \mathcal{C}} A$$

is the ideal which is the upper-bound for $\mathcal{C}$. For this to be the case, J must be an ideal.

1. J is non-empty, since any ideal must contain 0, thus $0 \in J$
2. Take $a, b \in J$ and consider $a - b$. Then for some $A, B \in \mathcal{C}$, $a \in A$ and $b \in B$. Since $\mathcal{C}$ is a chain, $A \subseteq B$ or $B \subseteq A$. Witout loss of generality, let's assume $A \subseteq B$ and say $a, b \in B$ (if e pick $B \subseteq A$ do $a, b \in A$). Then since B is an ideal, $a - b \in B$, so $a - b \in J$, showing that J is closed under subtraction
3. Take $a \in J$ and consider $r \cdot a$ and $a \cdot r$. Then since $a \in J$, a is in some ideal, let's say $a \in A$. Then $r \cdot a \in A$ and $a \cdot r \in A$, so $r \cdot a, a \cdot r \in J$, showing it's closed uner left and right multiplication

Thus, J is an ideal. It is further a proper ideal. If it ere not, then $1 \in J$, meaning for some ideal A, $1 \in A$. But then $A = R$, showing it is *not* a proper ideal, a contradiction to the definition of J. Thus, J must also be a proper ideal.

Therefore, by Zorn's Lemma, $\mathcal{I}$ must have a maximal element M, that is, if there is an element M' such that $M \subseteq M'$, then it must be that $M' \subseteq M$, so $M = M'$, which fulfills the definition of a maximal ideal, as e sought to show.

Note that we took the union of ideals within the proof. In general, if $I, J \subseteq R$, then $I \cup J$ is not an ideal (consider $2\mathbb{Z} \cup 3\mathbb{Z}$). The reason it worked for this proof is because the ideals are in a chain:

$$I_1 \subseteq I_2 \subseteq \cdots \subseteq I_n \subseteq \cdots$$

It should be pointed out that it is possible to have an infinitely long chain of the form $(x_1) \subsetneq (x_1, x_2) \subsetneq \cdots$ of which the union is not a maximal ideal, hence being maximal is a different condition that such a chain terminating in finite itterations, or even countably-infinite many iterations. This is part of the importance of using Zorn's lemma in the proof.

Since Zorn's Lemma was used, there is no way to computationally figure out what the maximal ideal is by just using this proposition. If a ring is also commutative, we can more easily characterize it with the structure of their quotient ring

Proposition 9.4.11: Maximal Ideals in Commutative Rings

Let R be a commutative ring with $1 \neq 0$. The ideal M is a maximal ideal if and only if the quotient R/M is a field

***Proof* :**
This follow's from the Correspondence Theorem and proposition 9.4.6. Let M be a maximal ideal so that there is no proper I such that $M \subsetneq I \subsetneq R$. Then by the Correspondence Theorem, the ideals I containing M correspond bijectively to the ideals of R/M, and since no proper ideal contains M, the only ideals of R/M are (0) and R/M, thus M is maximal if and only if the only ideals of R/M are (0) and R/M. By proposition 9.4.6, a ring is a field if and only if it's only ideals or (0) and itself. Thus, M is an ideal if and only if R/M is a field, as we sought to sho

The use of the Correspondence theorem in proposition 9.4.11 can be visualized with such a diagram:

$$\begin{array}{cc} R & R/M \\ | & | \\ M & 0 \end{array}$$

Notice that if R was not commutative, then left ideals are not necessarily right ideals, and so the theorem does not apply: if we pick any one-sided ideal (say a left-ideal), and perform proposition 9.4.10 to show the existence of a maximal left-ideal (and hence, we have a non-empty collection of maximal left-ideals), it is not guarnateed that the set of maximal left-ideals will contain any proper right-ideals[21]. A different approach must be taken to study R/I when R is not commutative and I is a maximal left/right ideal, which will lead to many fruitful results in part VI. This case will require some more module theory, and so will be left alone for now.

In the commutative case, this let's us study properties of R by studying the field R/M for all maximal ideal M. If R has a single maximal ideal (say $\mathfrak{m}$) then $R/\mathfrak{m}$ is called the *residue field*. An example of

[21] Indeed, we can modify example 9.4.3 to see that $M_2(k)/(0)$ is not a field, and hence a counter-example

a local ring would be $\mathbb{Z}/4\mathbb{Z}$, with single maximal ideal $\{\overline{0}, \overline{2}\}$. Such are called local rings for a reason; they show up intensively in commutative algebra, and will be central in chapter 21 in part V.

This last characterization of maximal ideals through quotienting for commutative rings will come to be very useful in the construction of some fields. For example this will be used to construct all finite fields by taking quotient of the ring $\mathbb{Z}[x]$ by maximal ideals. In a more specific case, we will be construction the "extension" of a field k by first defining a polynomial ring $k[x]$, and quotienting it by a maximal ideal $M \subseteq k[x]$. This will be the foundation of *field Theory*, and will be covered in part IV

Example 9.4.6

1. We've hand-waved a bit before how $p\mathbb{Z}$ is a maximal field. Since $\mathbb{Z}$ is commutative, we can now rigorously state this by noticing that $\mathbb{Z}/p\mathbb{Z}$ is a field
2. The ideal $(2, x)$ is a *maximal* in $\mathbb{Z}[x]$, since $\mathbb{Z}[x]/(2, x) \cong \mathbb{Z}/2\mathbb{Z}$, and $\mathbb{Z}/2\mathbb{Z}$ is a field
3. Let $R = \mathcal{F}([0, 1], \mathbb{R})$ be the ring of all functions from $[0, 1]$ to $\mathbb{R}$, and let M_a be the kernel of evaluation at a ($M_a = \{x \in [0, 1] | f(x) = a\}$). Show that M_a is maximal.

There are two more cool examples I'll get to when I see their importance

Note that a ring can have multiple maximal ideals. For example, $\mathbb{Z} \times \mathbb{Z}$ is a ring, and $I = ((1, 0), (0, 2))$ is an ideal generated by $(1, 0)$ and $(0, 2)$. Then $(\mathbb{Z} \times \mathbb{Z})/I \cong \mathbb{Z}/2\mathbb{Z}$, meaning I is a maximal ideal. Similarly, $J = ((2, 0), (0, 1))$ is an ideal and $(\mathbb{Z} \times \mathbb{Z})/J \cong \mathbb{Z}/2\mathbb{Z}$. Thus, both I and J are maximal. Note too that the two quotients need not be isomorphic (see exercise ref:HERE, thinking of $\mathbb{Z}$, with the fact that $\mathbb{Z}/3\mathbb{Z}$ and $\mathbb{Z}/5\mathbb{Z}$ are not isomorphic).

9.4.4 Prime Ideals

The first definition given of prime numbers is that a number is prime if and only if it's only divisors are 1 and itself, that is the only numbers such that $a|p$ are 1 and p. This definition works well for integers, but does it generalize well when we try to talk about divisibility over more general rings. This is due to a sneaky "global" assumption when working with primes, namely that an element is prime if *any* factorization that may include the prime element *has* to include the element as a factor. The following example should illustrate how this isn't necessarily the case:

Example 9.4.7: The Need For Globalness
[The Need For Globalness] Take $R = \mathbb{Z}[2i]$. Then $4 = (-2i)(-2i)$, and $-2i$ can no longer be factor[a]. However, we also have $4 = 2 \cdot 2$, and 2 can no longer factor (which in this case can be verified by taking a few cases). Thus, even though 2 was a factor of 4, we may split 4 differently, showing that there are "local" factorizations.

[a] up to multiplying by a unit, but factorizaiton is only defined up to unt, think $2 \cdot 3 = (-2) \cdot (-3)$

This subtly becomes very important, and shall be explored in depth in chapter 10.

To capture this idea, consider a prime number p. Then p is a prime number and if p divides ab, $a, b \in \mathbb{Z}$, then p divides a or p divides b. For example, if $3|15$, then either $3|3$ or $3|15$. If p was not a

prime number, this is not the case: $8|16[= 4 \cdot 4]$, but $8 \nmid 4$ and $8 \nmid 4$. The prime ideals for the integers are the sets that contain all the multiples of a given prime number

Definition 9.4.12: Prime Ideal

Assume R is commutative. An ideal P is called a ***prime ideal*** if $P \neq R$ and whenver the product ab of two element $a, b \in R$ is an element of P, then at least one of the a an b is an element of P.

Since R is commutative, another way of saying the condition is that if A, B where ideals and $AB \subseteq P$, then $A \subseteq P$ or $B \subseteq P$. The non-commutative case does not generalize well, however we discuss a remedy for this case in chapter 22[22].

Example 9.4.8: Prime Ideals

The origins of the idea come from generalizing the notion of primes in $\mathbb{Z}$, so we'll do the same. The prime numbers p generate the prime-ideals (p).

Let's say $ab \in (p)$, meaning $ab = rp$ for some $r \in \mathbb{Z}$ (note that $\mathbb{R}$ is commutative so $(p) = Rp$). Since $ab = rp$ that impleis that $p|ab$ (p divides ab): If you think about what it means for a number x to divide y ($x|y$), then that means there exists another number a such that $ax = y$. For example $3|6$, so $2\dot{3} = 6$.

Now, since $p|ab$, then $p|a$ or $p|b$. If $p \nmid a$ and $p \nmid b$, then ab wouldn't have it as a prime number under their factorization, but then $p \nmid ab$ – a contradiction. So, $p|a$ or $p|b$. But that would mean $r_a p = a$ or $r_b p = b$ for appropriate r_a or r_b. But then, $a \in (p)$ or $b \in (p)$.

Here's a concrete exmpale. Take $3\mathbb{Z}$, an ideal of $\mathbb{Z}$. (can think of $\varphi : \mathbb{Z} \to \mathbb{Z}/3\mathbb{Z},\ \varphi(x) = [x]$ to convince yourself it's an ideal). Now let's say $ab \in 3\mathbb{Z}$. That means that $3|ab$, meaning either $3|a$ or $3|b$. In either case, that would imply $a \in 3\mathbb{Z}$ or $b \in 3\mathbb{Z}$, showing that $3\mathbb{Z}$ is a prime ideal. On the other hand, if we have $6\mathbb{Z}$, Then $2 \cdot 3 \in 6\mathbb{Z}$, but $2 \notin 6\mathbb{Z}$ and $3 \notin 6\mathbb{Z}$, meaning $6\mathbb{Z}$ is *not* a prime ideal.

Note that this is an *inclusvie* or, not exlusive. If $p = 2$, and $a = b = 2$, then $2|4$, and both $2|2$ and $2|2$.

Another way I like to think about is is that we already know that if $a \in I$ and $b \in I$, then $ab \in I$. However, it's not clear the converse is true $ab \in I$, $a \in I$ and $b \in I$. This condition is a little too strong: Since $1 \cdot x \in I$ then $1 \in I$, meaning $I = R$. So what if we loosen the condition a bit , and say $ab \in I$ then $a \in I$ *or* $b \in I$. In other words, we can to a limited degree work backwards from *any* combination of elements $ab \in I$ to find one of them keeps needing to be in I. For example, if $abc \in I$ (say we braketed $(ab)c$) then either $ab \in I$ or $c \in I$. Since $ab \in I$, then either $a \in I$ or $b \in I$. This idea will come back in section 10.1, when we talk about factorizing elements in an integral domain, and will be the key idea to a couple of proofs[23].

(Is all I said earlier a fancy way to say that prime ideals are multiplicatively closed?)

This interpretation has some validity if we consdider the following scenario: Take a ring R, and a prime element $p \in R$. Then let's say there was an element $r \in R$ such that $r|p$. This would imply that $p = rr'$ for some $r' \in R$. Since the two values are the same $(rr') = (p)$, implying $rr' \in (p)$. Here, we get to use the property of p being prime: either $r' \in (p)$ or $r \in (p)$, meaning $p|r$ or $p|r'$. If $p|r$,

[22] namely, the ideal group is abelian for certain non-abelian

[23] for example, proposition 10.1.22 and ref:HERE

then we have $r|p$ and $p|r$, meaning $p = r$. Be careful here, this is not always true. If r, r' where unit, then we always have $r|r'$ and $r'|r$, but not awlays $r = r'$. For example, in $\mathbb{R}$, $1|2$, $2|1$ but $1 \neq 2$. However, it works here because HERE. So, we are forced to say $r' = u$, where u is a unit. For the othe way around, HERE.

Take $\mathbb{Z}$ again and let's say I was an ideal with this condition. then for some $x \in I$, $px \in I$ for every prime p. By our condition, $p \in I$. This will hold true for every prime-number, meaning I now has all but the identity. I'm not sure if this will always happen, but it's worth pondering.

Another thought: Take (x) and consider $ab \in (x)$. Then $ab = cx$ for appropriate $c \in R$. Then we no get the question $cx \in (x)$ too. This though is "obvious" since $x \in (x)$. But ab is not obvious. So the "representation" of the value that both ab and cx represents is important: it's a question about hether the constituent components of a represetnation is in there (a or b or c or x)

Here are some more example of prime ideals:

Example 9.4.9: Further Examples of Prime Ideals

1. Take $C([0,1])$, that is, all continuous function to/from $[0,1]$. Let $I \subseteq C([0,1])$ be all polynomials with roots at $\frac{1}{2}$ and $\frac{1}{3}$ ($f(\frac{1}{2}) = f(\frac{1}{3}) = 0$). Is I an ideal? Is it a prime ideal? Hint [a].

2. As an exercise to show that you can no longer "go backards", prove that if R is a commutative ring, and P is a prime ideal with no zero-divisors, then R is an integral domain. (Hint [b]).

[a] Think about two polynomials with single roots such that when multiplied, you get an element of I
[b] Assume there is a zero-divisor

This last example is important enough to be referenced many times, and so we make it a proposition:

Proposition 9.4.13: Prime Ideals And Integral Domains

Assume R is commmutative. Then the ideal P is a prime ideal in R if and only if the quotient ring R/P is an integral domain

Corollary 9.4.14: Maximal Ideal Then Prime Ideal

Assume R is commutative. Then every maximal ideal of R is a prime ideal

Proof **:**
If M is a maximal idea, then by proposition 9.4.11 R/M is a field. A field is an integral domain, so the corollary folllows from proposition 9.4.13

Example 9.4.10

1. The principal ideals generated by (p) where p is prime are both prime and maximal. Not all prime ideals are maximal: (0) is a prime ideal, but not maximal.

2. The ideal (x) is a prime ideal in $\mathbb{Z}[x]$ since $\mathbb{Z}[x]/(x) \cong \mathbb{Z}$, but not maximal since $\mathbb{Z}$ is not a field. Or, you can diretly see that $(x) \subseteq (2, x)$, meaning that (x) is not maximal.

Exercise 9.4.2

1. Let R, S be rings and $R \times S$ be the product ring. What are the prime ideals of $R \times S$(answer: $R \times P$ and $Q \times S$ for primes P and Q. In general, only one component can be the ideal)
2. Let $a, b \subseteq R$ be ideals. Show that $(a+b)(a \cap b) \subseteq ab$ and $ab \subseteq a \cap b$. Show that if $(a, b) = (1)$, then $a \cap b = ab$.
3. exercise on quotient ideals $(a : b)$
4. Let $f : R \to S$ be a ring homomorphism and $p \subseteq S$ a prime ideal. Show that $f^{-1}(p)$ is a prime ideal. If $m \subseteq S$ is maximal, is $f^{-1}(m)$ maximal? (Consider $R = \mathbb{Z}$ and $S = \mathbb{Q}$)
5. Perhaps, if possible at this stage how if if $a_1, a_2, ..., a_n$ are ideals and $\cap_i a_i$ is contained in the prime ideal p, then p contains one of the a_i, and if $p = \cap_i a_i$, then $p = a_i$ for on of them (shows how prime genelize numbers, so to speak).
6. Notice that we have defined what I^k is. Let's say we started with an ideal J; is there a rasonable way of defining $\sqrt{J}$? Define $\sqrt{J} = \{a \in R \mid \exists n \in \mathbb{N}_{>0},\ a^n \in J\}$. Show that $\sqrt{J}$ is an ideal.

9.5 Ring of Fractions: Localization

In this section, we'll take a moment to discuss how we can take any ring with no zero divisors, and extend the ring to one in which every element has a multiplicative inverse! More generally, given any non zero-divisor element $d \in R$ (where R will have to be commutative), we can construct a ring R_d which will where $d^{-1} \in R_d$. This concept was applied in the field of *algebraic geometry* for the purposes of looking at a ring "locally". What this means and how to think about this consruction "geometrically" will be discussed much later in part V in section 20.3. In this section, the full generality of this construction is not particularly useful. What's more interesting at the moment is that this gives us a way to add units to our ring!

We'll start with a practical example by using $\mathbb{Z}$ to construct $\mathbb{Q}$! Taking $\mathbb{Z}$, we can make an equivalence class

$$(a, b) \sim (c, d) \Leftrightarrow ad = bc$$

where $b, c \neq 0$. Then $[(a, b)]$ represents a fraction $\frac{a}{b}$, with the point of the equivalence relation is to make all fractions with similar divisors be the same:

$$\frac{1}{2} = \frac{2}{4} = \frac{3}{6} = \cdots$$

I'll stick with fraction notation for these equivalence relationships. We can see that we can define

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \qquad \text{and} \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$$

It should be checked that these operations are independent of representative. With this construction, we can identity $\mathbb{Z}$ with $\left\{\frac{a}{1} | a \in \mathbb{Z}\right\}$ [24], showing how $\mathbb{Z}$ is a sub-ring of $\mathbb{Q}$.

[24] meaning construct a isomorphism

With that intuition, we can ask about general commutative rings R. The problem comes when we allow $d = 0$ or be a zero divisor. Say $d \neq 0$ $bd = 0$. Then:

$$d = \frac{d}{1} = \frac{bd}{b} = \frac{0}{b} = 0$$

We'll avoid this by taking a set $R^\times$ which represents all the units of R

Theorem 9.5.1: Ring of Fractions

Let R be a commutative ring R (not necessarily with identity). Let D be any nonempty subset of R that does not contain 0, does not contain any zero-divisors and is closed under multiplication (i.e. $ab \in D$, for every $a, b \in D$)[a]. Then there is a commutative ring Q with 1 such that Q contains R as as a subring and every element of D is a unit in Q. The ring Q has the following additional properties

1. Every element of Q is of the form rd^{-1} for some $r \in R$ nd $d \in D$. In particualr if $D = R - \{0\}$ then Q is a field
2. (uniqueness of Q) The ring Q is the "smallest" ring containing R in which al elements of D become units. That is, if S is any commutative ring with identity and let $\varphi : R \to S$ be any injective ring homomorphism such that $\varphi(d)$ is a unit in S for every $d \in D$. Then there is an injective homomorphism $\Phi : Q \to S$ such that $\Phi|_R = \varphi$. In other words, any ring containing an isomorphic copy of R in which which all the elements of D become units must also contain an isomrophic copy of Q.

[a] such a set is called a *semi-group*

Proof **:**
We'll start like we started with $\mathbb{Z}$. Let $F = \{(r, d) | r \in R, d \in D\}$ (d in the 2nd component because we don't want zero-divisors in the bottom, and we need the denominator to be closed to define multiplication). Then, we define the equivalence relation

$$(r, d) \sim (s, e) \text{ if and only if } re = sd$$

It is easy to show that $\sim$ is an equivalence relation [a]

We commonly denote (r, d) by $\frac{r}{d}$. Now, let

$$Q = \left\{[(a, b)] \middle| a \in R, d \in D\right\}$$

be the set of equivalence relations. Note that $\frac{r}{d} = \frac{re}{de}$ for any $e \in D$, as we would expect.

We can define addition and multiplication as

$$\begin{aligned} [(a, b)] + [(c, d)] = [(ad + bc, bd)] \quad &\Rightarrow \quad \frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \\ [(a, b)] \cdot [(c, d)] = [(ac, bd)] \quad &\Rightarrow \quad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd} \end{aligned}$$

Then the following properties need to be checked. I'll show that addition is well-defined, and the rest is not hard to show that it follows:

1. The operations are well-defined (not dependent on representatives)

Proof. We'll show addition is well-defined. Let

$$\frac{a}{b} = \frac{a'}{b'} \ (ab' = a'd) \qquad \frac{c}{d} = \frac{c'}{d'} \ (c'd = cd')$$

then

$$\frac{ad+bc}{bd} = \frac{a'd' + b'c'}{b'd'}$$

which equivalently means:

$$\begin{aligned}
(ad+bc)(b'd') &= adb'd' + bcb'd' && \text{distributivity in } R \\
&= ab'dd' + cd'bb' && \text{commutativity in } R \\
&= ab'dd' + cd'bb' && ab' = a'b,\ cd' = c'd \\
&= (a'd' + b'c')(bd)
\end{aligned}$$

thus, the two are equivalent. A similar procedure works for multiplication □

2. Q is an abliean group under addition, where the aditive identity is $\frac{0}{d}$ for any $d \in D$ and the additive inverse of $\frac{a}{d}$ is $\frac{-a}{d}$
3. Multiplication is associative, distributive, commutative,
4. Q has an identity $\frac{d}{d}$ for any $d \in D$

Next, we show that R can indeed be embedded into Q. Let

$$i : R \to Q \qquad r \mapsto \frac{rd}{d}$$

Since $\frac{rd}{d} = \frac{re}{e}$, this function is independent of choice of d. Since d is closed under multiplication:

$$i(rs) = \frac{rsd}{d} = \frac{rs}{1} = \frac{r}{1}\frac{s}{1} = \frac{rd}{d}\frac{sd}{d} = i(r)i(s)$$

finally, for injectivity of the inclusion, notice that

$$i(r) = 0 \Leftrightarrow \frac{rd}{d} = \frac{0}{d} \Leftrightarrow rd^2 = 0 \Leftrightarrow r = 0$$

meaning the kernel is trivial, and i is injective.

Next, we show that every element $q \in Q$ can be written as rd^{-1}. Since $d = \frac{d}{1}$, then $d^{-1} = \frac{1}{d}$ since $\frac{d}{1}\frac{1}{d} = \frac{1}{1}$. Then by definition of an element in Q, we define an elemnent as $\frac{r}{d} \in Q$, and so we have that $q = rd^{-1}$.

Note that if $D = R - \{0\}$, then every element has a multiplicative inverse, and so that makes Q into a field.

Finally, we must show this extension is unique. To do this, let's say $\varphi : R \to S$ is an injective ring-homomorphism such that $\varphi(d)$ is a unit in S. In this way, S acts as "another" extension of R (since we shows that R can be imbedded into Q). Since $q = rd^{-1}$ for every $q \in Q$, extend φ to Φ by definining

$$\Phi(rd^{-1}) = \varphi(s)\varphi(d)^{-1}$$

for all $r, d \in R$. Note that $\varphi(d)^{-1}$ is defined since $\varphi(d)$ is a unit in S. This map is well-defined. If we have $rd^{-1} = re^{-1}$, then multiplying around we get $re = rs$, so by the property of ring homomrphisms $\varphi(r)\varphi(e) = \varphi(s)\varphi(d)$ and so

$$\Phi(rd^{-1}) = \varphi(r)\varphi(d)^{-1} = \varphi(s)\varphi(e)^{-1} = \Phi(se^{-1})$$

showing it is indeed well-defined. Checking Φ is a ring homomorphism is strait forward using commutativity of Q.

Last step is to show Φ is injective. There is a sneaky trick you can do to pull this off. Let's say $rd^{-1} \in \ker(\Phi)$. Then that means that $r \in \ker(\varphi)$ since we'd need $\Phi(rd^{-1}) = \varphi(r)\varphi(d)^{-1} = 0$ and $\varphi(d)^{-1}$ cannot be zero since d is a unit. However, $\ker(\varphi)$ is trivial, so $r = 0$, but then $rd^{-1} = 0d^{-1} = 0$, meaning Φ is also injective! Thus, we forced the S to have Q embedded inside of it, meaning *any* extension of R to S would inevitably have Q. This makes Q the smallest extension of R and the unique extension of R.

[a] Reflexivity and symmetry come immediately. For transitivity, combining the two equations appropriately and realise you can factor by $d \in D$, which is not zero or a zero-divisor, which will force the equation you want to be zero, forcing the equivalence for transitivity

We'll give Q a name:

Definition 9.5.2: Ring Of Fraction

Let R be a commutative ring, D be the appropriate subset and Q be the extension as in the above theorem. Then

1. The ring Q is called the *ring of fractions* of D with respect to R and is dneoted $D^{-1}R$
2. If R is an integral domain an $D = R - \{0\}$, Q is called the *field of fractions* or *quotient field* of R.

we often want all units of a ring R. We denote all units as $\mathbb{R}^{\times}$.

Here are some examples of ring of fractions

Example 9.5.1

1. the trivial example would a a field $\mathbb{F}$, which would be it's own field of fractions.
2. the construction we've done at the beginning is a special case of our construction. So, if we take $R = \mathbb{Z}$, and take $D = \mathbb{Z} - \{0\}$, then $D^{-1}\mathbb{Z} = \mathbb{Q}$
3. Let $R = \mathbb{Z}$ and take $D = \{1\}$. Then the quotient ring would be $\mathbb{Z}$ again!
4. $R = \mathbb{Z}$, and take $D = \{19, 19^2, ...\}$. This is a valid D, since it doesn't contain 0, doesn't contain zero divisors, and is closed under multiplicationso what is $D^{-1}\mathbb{Z}$?
5. The ring $2\mathbb{Z}$ is alsmot an integral domain; it doens't have an identity. However, $(R^{\times})^{-1}(2\mathbb{Z}) = \mathbb{Q}$ as well! If you think about it, being able to divide will give us the identity, as well as $\frac{1}{2}$, which will produce the rest
6. $Z[p]$ should be elaborated on

7. $\mathbb{F}[x]$ becomes $\mathbb{F}(x)$, that is, elements of the form $\frac{p(x)}{q(x)}$ where $q(x)$ is not the zero-polynomial (Since $\mathbb{F}[x]$ is an integral domain, it's the only value we would need to worry about). Notice the square-brakets became parenthesis. In this way, we might see the motivation to write $\mathbb{Q}(\sqrt{D})$, since we are allowed to divide those values as well! The difference here being that $x, x^2, x^3, ..., x^n, ...$ are all distinct elements, non of them being an element of their original field. However, this same chain doens't appear for $Q(\sqrt{D})$: $\sqrt{D}, \sqrt{D}^2 = D$. This is because $\sqrt{D}$ is not an arbitrary indeterminate value, giving us this relation, similar to what we saw with group-rings.
8. Take $\mathbb{Z}[i]$ and $D = R^\times = R - \{0\}$. Then the ring of fractions is $\mathbb{Q}[i]$! Notice it's not something like $\mathbb{Z}(i)$. This comes down to how multiplication is defined in this ring, which is much more akin to multiplying complex numbers.
9. what is the field of fractions of $\mathbb{Q}[\sqrt{2}]$?
10. What if we have $R = \mathbb{Z}/4\mathbb{Z}$ and $D = \{1, 3\}$?

Theorem 9.5.3: Universal Property Of Field Of Fractions

If $h : R \to F$ is an injective homomorphism from R into a field F, then there exists a unique ring homomorphism $g : Q \to F$ which extends h

Proof **:**
Use the uniqueness of the extension to show that $\Phi(Q)$ will be in any sub-field.

The universal property on $D^{-1}R$ will in fact still work if we generalize our construction to letting D have zero-divisors. The motivation for this generalization is something that (ref:HERE TBD). This topic will be exlpored in part V, section 20.3.

(Another perspective would be that we are giving a solution to all equations of the form $bx - a = 0$, for $a, b \in R$. Thus, for $\mathbb{Z}$, we are doing:

$$\frac{\mathbb{Z}[x_{1/2}, x_{1/3}, ..., x_{1/p}, ...]}{(2x_{1/2} - 1, 3x_{1/3} - 1, ...,)}$$

Though this way of looking at it seems more complicated, it fits better with the idea that we are adding relations to $\mathbb{Z}$. In particular, it shows that we are adding all *linear* relations (i.e. solutions) to $\mathbb{Z}$. We will in later section's show how we add non-linear solution)

28: Foreshadowing

By taking the "infinite" version of polynomials, we managed to make all the non-zero-divisors become units. In this section, another way of producing using by formally inverting them. These two processes shall seem very different, however they are related, and the connection will be made in section 20.4.

9.5.1 Constructing $\mathbb{R}$?

In ref:HERE, we mentioned how we can think of $\mathbb{N}$ being a free-monoid on 1 element, so in a sense, it was constructed from 1 element and the rules of a monoid. We then showed in ref:HERE how we can construct $\mathbb{Z}$ from $\mathbb{N}$. We just showed how we can construct $\mathbb{Q}$ from $\mathbb{Z}$. We have also give an exercise that constructs $\mathbb{C}$ from $\mathbb{R}$. The reader might feel like the natural next step up would be to construct $\mathbb{R}$ from $\mathbb{Q}$. There is in fact a construction of the real numbers (called the *Dedekind cut's construction* [25]). However, it turns out that there is no "algebraic" construction of the real numbers! That is, we can't start off with some set (say $\mathbb{Q}$) and extended using binary (more generally n-ary) operations with certain rules on creating new elements from old ones (like we did for $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$)!

This is a very lofty claim to be making. The fact that there isn't an algebraic construction of the reals is a fascinating result, and it will have to wait till we cover _____to fully appreciate. (ref:HERE).

Found this really interesting link!

```
https://math.stackexchange.com/questions/3761716/
is-there-an-algebraic-way-to-construct-the-reals
```

It seems to go down fundamentally that an algebraic construction tries to use the elements from current object to construct the bigger object *by means of finite-tuples* (because of the way we think of functions!!), so if we were to try to do $\mathbb{Q} \rightsquigarrow \mathbb{R}$, then it would be representing elements from $\mathbb{R}$ with finite tuples from $\mathbb{Q}$. But $\mathbb{Q}$ is countable, and the number of finite tuples from $\mathbb{Q}$ is countable, but $\mathbb{R}$ has uncountably many elements. Not sure about the subtlies yet, Löwenheim-Skolem's Theorem is beyond my comprehension right now, but I think this idea is good enough for now to appreciate!

Exercise 9.5.1

1. Let F be a field of characteristic 0. Show that there exists a ring R such that $\text{Frac}(R) = F$. (If F is not characteristic 0, then there are a couple of complications. The equivalent exercise is in ref:HERE (under valuations in comm alg, since it requires valuations))

9.6 Chinese Remainder Theorem

Let's say we have a ring R and a collection of ideals $I_1, I_2, ..., I_k$. Then given

$$a_1 \bmod I_1, \cdots, a_k \bmod I_k$$

does there exist a $a \in R$ such that

$$a \equiv a_1 \bmod I_1, \cdots, a \equiv a_k \bmod I_k$$

Think about it for a moment before looking at the answer

29: Taking a moment

aking a moment to think about the answer.

[25] See appendix ref:HERE

The answer depends on the choice of I_i. The reader should have tried this with $R = \mathbb{Z}$ and noticed it is not always possible, for example if $R = \mathbb{Z}$, $a \equiv 1 \bmod 2,\ a \equiv 2 \bmod 4$ does not have a solution.

The Chinese Remainder Theorem (CRT) is a way of finding an to this question. It's name sake most likely comes from a book *Sun Tzu Suan Ching*. It is said that it was originally developed either for accounting purposes or for dividing troops.

Before we get into the actual theorem, we'll first explore a direct application of Chinese Remainder Theorem:

Example 9.6.1: Concrete Use

simple form

$$\text{if } gcd(m,n) = 1 \text{ and } 0 \leq a < m\,, 0 \leq b < n$$

then

$$\exists!\ k,\ 0 \leq k < mn \text{ such that } k \equiv a \bmod m\,, k \equiv b \bmod n$$

Let's say: $m = 6$, $n = 5$, $a = 4$, $b = 2$. Then $k = 22$.

Proof. We have

$$k \equiv 4 \bmod 6,\ k \equiv 2 \bmod 5$$

so using the classical method, we first find

$$6 \cdot 5 = 30$$

from this, we find $30/6 = 5$ and $30/5 = 6$ to be the values such that and kno that we get the solution by solving the equivalent problem of:

$$k \equiv 5 \cdot 1 \bmod 30,\ k \equiv 6 \cdot 1 \bmod 30$$

meaning the solution is $(4 \cdot 5 \cdot 5) + (2 \cdot 6 \cdot 6) =$ □

When generalizing this result, we can no longer ask for numbers to be co-prime. The equivalent concept in rings are co-maximal ideals [26]

Definition 9.6.1: Comaximal Ideals

The ideals A and B of the ring R are said to be *comaximal* if $A + B = (1)$

The intuition on when to spot something is co-maximal is analogous to two numbers being prime. If $a, b \in \mathbb{Z}$ are prime, then by Bézout's identity, there exists numbers x, y such that $xa + yb = 1$, meaning $1 \in (a) + (b)$, meaning $(a) + (b) = \mathbb{Z}$ by proposition 9.4.6. In the next chapter, we'll generalize the concept of gcd for more general rings, meaning where the gcd can be defined, this theorem has a useful application [27]

[26] why not co-prime ideals? We'll show how prime and maximal ideals will match in special rings in chapter 22, section 22.2, here we would usually apply such a result. However, we stay general since we can prove it in more general contexts

[27] For example, we'll use this theorem when looking at polynomial rings over fields

Lemma 9.6.2: Characterizing Products of Coprime Ideals

Let $I_1, I_2, ..., I_n$ be coprime ideals. Then:

$$\bigcap_{i=1}^{n} I_i = \prod_{i=1}^{n} I_i$$

Proof **:**
We shall do induction so consdier first when $n = 2$ so that $I, J \subseteq R$ such that $I + J = (1)$. It is always the case that $IJ \subseteq I \cap J$ so we have to show that $I \cap J \subseteq IJ$ when $I + J = (1)$. Given $I + J = (1)$, there exists $a \in I$, $b \in J$, such that $a + b = 1$. Now, if $r \in I \cap J$, then

$$r = r \cdot 1 = r(a + b) = ra + ra \in IJ$$

Since $r \in J$ and $a \in I$, $ra \in IJ$, an similalry $rb \in IJ$ since $r \in I$ and $b \in J$.

Then the inductive step follows immediately.

Lemma 9.6.3: Recursive Comaximal Of Products

Let $I_1, I_2, ..., I_n \subseteq R$ be ideals such that $I_i + I_n = (1)$ for all $i \in \{1, ..., n-1\}$. Then $(I \cdots I_{n-1}) + I_n = (1)$

Proof **:**
By assumption, there exists $a_i \in I_k$ and $1 - a_k \in I_i$ such that

$$(1 - a_1) \cdots (1 - a_{n-1}) \in I_1 \cdots I_{n-1}$$

And

$$1 - (1 - a_1) \cdots (1 - a_{n-1}) \in I_n$$

since it is a comination of the elements $a_1, a_2, ..., a_{n-1} \in I_n$

Theorem 9.6.4: Chinese Remainder Theorem

Let $A_1, A_2, ..., A_k$ be ideals in R. The map

$$R \to R/A_1 \times \cdots \times R/A_k \qquad \text{defined by} \qquad r \mapsto (r + A_1, \cdots, r + A_k)$$

If for each $i, j \in \{1, 2, ..., k\}$ with $i \neq j$ the ideals A_i and A_j are comaximal, then this map is surjective and so

$$R/(A_1 A_2 \cdots A_k) = R(A_1 \cap A_2 \cap \cdots \cap A_k) \cong R/A_1 \times R/A_2 \times \cdots \times R/A_k$$

Proof **:**
Take

$$R \to \bigoplus_{i=1}^{n} R/A_i \qquad a \mapsto (a \bmod A_1, ..., a \bmod A_n)$$

Then this is easily checked to be a ring homomorphism with kernel $A = \bigcap_i A_i$. Assuming now the ideals are comaximal, we shall show surjectivity. We shall do a proof by induction. If $n = 1$ where done, so let for let's say

$$R/(A_1 \cdots A_{n-1}) \cong R/A_1 \times \cdots \times R/A_{n-1}$$

Then by lemma 9.6.3, $(A_1 \cdots A_{n-1}) + A_n = (1)$, and hence this is reduced to two ideals. By comaximality there exists $a_1 \in A_1$, $a_2 \in A_2 \cdots A_{n-1}$ such that $1 = a_1 + a_2$. Then writting

$$x = x_2 a_1 + x_1 a_2$$

we get that $x \equiv x_1 \bmod A_1$ and $x \equiv x_2 \bmod A_2 \cdots A_n$. But then $x \mapsto (x_1 \bmod A_1, x_2 \bmod A_2 A_3 \cdots A_{n-1})$, completing the proof.

A useful consequence to remember is that if there are elements $r_i \in R$ where $r_i \bmod A_i$, then there exists an element a

$$a \equiv r_i \bmod A_i$$

The ability to solve such a system mod A_i is sometimes called the *Weak Approximation Theorem.*

Corollary 9.6.5: Chinese Remainder Theorem For Integers

Let $R = \mathbb{Z}$ $a_1, a_2, ..., a_n \in \mathbb{Z}$ be elements such that $\gcd(a_i, a_j) = 1$ for all $i \neq j$. Then if $a = a_1 \cdots a_n$, The function

$$\varphi : \mathbb{Z}/a\mathbb{Z} \to \mathbb{Z}/a_1\mathbb{Z} \times \cdots \times \mathbb{Z}/a_k\mathbb{Z}$$

given in the natural way is an isomorphism

Proof **:**
This is an immediate consequene of the chinese remainder theorem since $\gcd(a, b) = 1$ if and only if $(a, b) = (a) + (b) = 1$

Exercise 9.6.1

1. Show that if $P + Q = R$, then for any $n, m > 0$, $P^n + Q^m = R$. Conclude the CRT can be generalized to ideals of the form $A_1^{n_1}, ..., A_k^{n_k}$.

9.7 *Monoidal Objects

Maybe have this here to be able to introduce R-algebras as the monoidal objects in the R-**Mod** Category?

10

Exploring Integral Domains: Factorization and Valuation

This chapter will assume basic number theory knowledge from the preliminary chapter (section 0.2)

Integral domains are a huge overarching category of rings, being in a meaningful way the generalization of integers. As they do not having zero-divisors makes it possible to talk about partially inverting elements (by proposition 9.1.14). There are many different consequences of this result especially in the right specializations, for example divisibility, size, and even the euclidean algorithm shall all be definable on the appropriate integral domains[1]. These topics all interact greatly with each other and there is no linear way of addressing these ideas. However, books are inherently written with a total ordering and we must learn content in some order, and hence this chapter has been written with the idea of exploration the notion of divisibility (which the reader can think of the best that can be done to check invertibility if an element is not a unit). As this exploration naturally progresses, all other pertinent notions about integral domains shall become natural questions and will be answered.

The following diagram sums up the results of this chapter:

[1]There is also a more geometric interpretation, which is that integral domains will correspond to "irreducible" geometric objects (think of how $R \times S$ is no longer integral. This idea is explore in chapter 21

Figure 10.1: Image form Alufi's textbook

At the end of this chapter, we shall explore how these abstract notions were in fact developed to solve number-theoretic problems. We shall see our first example of finding the solution to a Diophantine equation, namely

$$x^2 + y^2 = p$$

In other words, which primes can be written as sums of two squares. As equations are the fundamental way in which we determine the properties of objects, this will be the first glimpse into the properties of prime numbers. This exploration is greatly generalized in chatper 22, section 22.4.1.

> **Note**: Some results in this chapter apply beyond integral domains. It will be made extra-clear when that is.

Goal

In fields, every element has an inverse. This is not so for general rings, as we have seen when studying units and zero-divisors. If not every element is a unit, there is a new idea that is introduced when learning some basic number theory: *factorization*. In this way, we can "decompose" an element into some basic building blocks. This idea of factorizing is one the reader is already familiar with: any number $x \in \mathbb{Z}$ can be factored, even uniquely, into prime numbers:

$$x = up_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}$$

where $u \in \{1, -1\}$ and each p_i is a prime. Is there a generelizable notion of being factorizable in more general rings? Is this idea related to prime ideals? It turns out that some types of rings do admit an idea of factorization, or even some stronger conditions (like the euclidean algorithm being well-defined, or the factorization being unique), while other rings have a weaker condition (like the factorization not being unique or even existing). The first major focus of this chapter are these ideas.

The overarching idea in factorization is centered around the idea of divisibility like we have seen in $\mathbb{Z}$. For example:

$$2|8 \qquad 32|17728 \qquad \text{but} \qquad 3 \not| 8 \qquad 5 \not| 7$$

When an element is divisible, let's say $2\big|8$, then there exists some $x \in \mathbb{Z}$ such that $2x = 8$, that is, 8 got subdivided into other elements; in this way, factorization can be thought of as a "partial division" or "partial inverse" condition. This idea will be at the heart of our following exploration.

The result will give 3 new classes of rings and develop on 2 previously mentioned classes or rings which are very often studied.

1. *Atomic Domains* (or Domains with Factorization) are will be rings in which element can be factored into elements that can no longer be factored (they will be called irreducible elements), though not necessarily uniquely
2. *Noetherian domains* are similar to Unique Factorization domain as in every element can be factorzied, but it need not be unique. The reader may recall these as rings in which every ideal may be finitely generated.
3. *Unique Factorization Domains* (UFDs) will allow for every element to have a unique factorization, and let's us define the notion of there being a greatest common divisor between elements.
4. *Principal Ideal Domains* (PIDs) will be stronger than PIDs in that they allow for unique factorization and will have a more rigid structure. The reader may recall these as rings in which every ideal may be generated by a single element.
5. *Euclidean Domains* will have the property of the euclidean division algorithm applying to them to quickly find the greatest common divisor. These will allow for algorithmic decomposition of elements into prime elements.

Since factorization is heavily inspired by the properties of $\mathbb{Z}$, all rings in this chapter will be *integral domains* unless specified otherwise.

10.1 Factorization

We will first define more rigorously the notion of elements being factorizable

Definition 10.1.1: Divisibility

Let R be a commutative ring and $a, b \in R$ be elements of R. Then we say that a *divides* b (or b *is divisible by* a.) if there exists an element c such that

$$ca = b$$

Equivalently, $ca = b$ if and only if $b \in (a)$. Notationly, we write $a\big|b$ to mean "a divides b".

If R was not commutative, we can define left and right divisibility, though we do not cover the non-commutative case (see exercise ref:HERE for some exploration of the concept). The opposite of a being divisible by b is b being a *multiple* of a – this terminology is used much less often.

The equivalent condition with ideals are very insightful way of understanding divisibility. Here are the 4 equivlent condition's of divisibility for ease of reference:

1. $b|a$ if and only if

2. $a = bb'$ if and only if
3. $a \in (b)$ if and only if
4. $(a) \subseteq (b)$

It is unfortunate that the notation $b|a$ stuck, since you will have to swap at some point the order of the a and the b given that it's read "b divides a" and not "a is divided by b". It's a matter of getting used to.

The notation for divisibility using ideals is particularly insightful here; notice that if $u \in R$ is a unit, then for any element $r \in R$ $(r) = (ur)$. Both inclusion are easy and should be proven before continuing to read. Because of this property, it is immediate that $a\big|b$ if and only if $a\big|ub$, that is, if an element is divided (or not divided) by another, then multiplying it by a unit let's it remain divisible (or not divisible) by the other element. This is in fact something the reader is already is familiar with; notice that $6 = -1 \cdot -6$. Then $2\big|6$, and $2\big| -6$, but also $5 \nmid 6$ and $5 \nmid -6$. Thus, when we think of divisibility, it is important to think of a number being divisible "up to a unit":

Definition 10.1.2: Associative

Let R be a commutative ring and let $a, b \in R$ be two elements. Then we say that a and b are *associates* if the ideals generated by a and b are equal, that is:

$$(a) = (b)$$

It is important to check that this is equivalent to saying that a and b are associates if $a\big|b$ and $b\big|a$, showing how this links to divisibility. Notice that in the previous definition, R was not an integral domain, but just a commutative ring, which allows zero-divisors. This in fact breaks the intuition of associativity we just showed:

Example 10.1.1: associativity in non-integral domains

Let $R = \mathbb{Z}/6\mathbb{Z}$. Then $(\overline{2}) = (\overline{4})$, since

$$(\overline{2}) = \{\overline{2}, \overline{4}, \overline{0}\} = \{\overline{4}, \overline{2}, \overline{0}\} = (4)$$

Furthermore, $\overline{4} = \overline{2} \cdot \overline{2}$ but $\overline{2}$ is *not* a unit.

For this reason, restricting to integral domains let's us keep this "up to a unit" intuition for for divisibility using ideals:

Proposition 10.1.3: Associativity in Integral Domains

Let R be an integral domain. Then a and b are associates if and only if $a = ub$ for all units $u \in R$

Proof **:**

Let $(a) = (b)$, so that for some $r, s \in R$, $ra = b$ and $a = sb$. Then

$$sra = sb = a$$

and since the cancellation property holds for integral domains, we have $sr = 1$, showing that r

and s are units.

Conversely, if $a = ub$ for all units $u \in R$, then it's clear that $(a) \subseteq (b)$ since $u \cdot b = a$ and $(a) \subseteq (b)$ since $u^{-1} \cdot a = b$.

Incidentally, notice that in example 10.1.1, the units of $\mathbb{Z}/6\mathbb{Z}$ is only $\overline{5}$ and that $\overline{2} \cdot \overline{5} = \overline{4}$, meaning the previous proposition actually holds for $\mathbb{Z}/6\mathbb{Z}$. This is true for a special subset of commutative rings usually called *rings with harmless zero-divisors* (harmless since the intuition of divisibility is preserved). These are not covered in this book, but can be explored in (reference something: here). Since the intuition for associativity holds true for all integral domains, and most work being done which includes divisibility in rings focuses on integral domains (due to the cancellation property), we shall do the same restriction here and work over integral domains rather than generalize the following proofs into much more (arguably needlessly) complex arguments to account for the case of rings with harmless zero-divisors.

With a notion of divisibility up to unit established, we can start thinking of what are the building blocks, the "fundamental elements", into which we will factorize an element:

Definition 10.1.4: Irreducible and Prime Elements

Let R be an intergral domain

1. Let $r \in R$ be a nonzero element that is not a unit. Then r is called *irreducible* in R if whenever $r = ab$, at least one of a or b must be a unit in R. Otherwise, r is said to be *reducible*.

2. The nonzero element $p \in R$ is called *prime* in R if the ideal (p) generated by p is a prime ideal. In other words, a nonzero element p is a prime if it is not a unit and whenver $p|ab$ for any $a, b \in R$ then either $p|a$ or $p|b$.

Example 10.1.2: Irreducible and Prime elements

Let $R = \mathbb{Z}$, and take $6 \in \mathbb{Z}$. Then we can see that $2 \cdot 3 = 6$. Since 2 and 3 are both not units, that means that 6 is a reducible element in $\mathbb{Z}$. On the otherhand, 2 and 3 are not reducible elements. If there existed a a, b such that $2 = ab$, then by our knowledge that 2 is a prime number (or by some elementary number theory argument), the only number that divide it are 1 and 2, so it must be that either

$$1 \cdot 2 = 2 \quad \text{or} \quad -1 \cdot -2 = 2$$

In both cases, a is a unit, and so 2 is irreducible (similarly for 3). Furthermore, 2 and -2 are associates, since $2 = -1 \cdot -2$, i.e., the differ by a unit. in $\mathbb{Z}$, every n are associates to it's negative $-n$.

As pointed out in example 9.4.8, 2 is also a prime element since (2) is a prime ideal. In fact, in the previous exercise, we used the fact that 2 is a prime integer (which corresponds to the prime ideals of $\mathbb{Z}$) in order to show it's irreducible. Though 2 is both an irreducible and a prime element in $\mathbb{Z}$, it need not be the case in general integral domains that an element is both!

As an example, take $\mathbb{F}[x]$ to be the polynomials over a field, and take $F_{0x} \subseteq \mathbb{F}[x]$ to be all the polynomials such that the coefficient of x is zero. As an example of elements in F_{0x}:

$$ax^3 + b \in F_{0x} \qquad cx^3 + dx \notin F_{0x} \qquad \forall a, b, c, d \in \mathbb{F}$$

It is easy to check that F_{0x} is an integral domain (as you should check), and so we can try to find irreducible and prime elements. Take x^6. It has two factorizations: $x^6 = x^2x^2x^2$ and $x^6 = x^3x^3$. The elements x^2 and x^3 are irreducible, but not prime.

To show x^2 and x^3 are irreducible, let's say

$$x^2 = p(x)q(x)$$

then since we're working with polynomials over an integral domain, by proposition 9.2.2:

$$\deg(x^2) = \deg(p(x)) + \deg(q(x))$$

meaning $2 = \deg(p(x)) + \deg(q(x))$. Since both $p(x)$ and $q(x)$ cannot be one (since $ax \notin F_{0x}$, for all $a \in \mathbb{F}$) then one of them has to be of degree zero, but since x^2 is monic this implies it must be 1, meaning it's a unit, meaning x^2 is irreducible. Similarly for x^3. Notice that this means that x^6 had two decompositions into irreducible factors, meaning the decomposition was not unique.

Now, take (x^2), i.e. the ideal generated by x^2. Notice that $(x^3)(x^3) \in (x^2)$ because $x^4 \cdot x^2 \in (x^2)$. However, $x^3 \notin (x^2)$. If it were, then there would exist a polynomal $p(x)$ such that $p(x)x^2 = x^3$. By the degree formula, $p(x)$ would have to be of degree 1. However, no degree 1 polynomial exists in F_{0x}. Therefore, (x^2) is not prime! This harkens back to when we talked about prime-ideals, where being prime in some way means we can work backwards. This "working backwards" let's us guarantee we decompose our number into unique elements, as we'll soon see.

Though not all irreducible elements are prime, the converse is true, that is, being "globally divisible"' is a strong enough condition that it guarantee's irreducibility:

Proposition 10.1.5: Prime Then Irreducible In Integral Domain

Let R be an integral domain. Then prime elements in R are always irreducible

Proof **:**
Supose $(p) \subseteq R$ is a nonzero prime ideal, and that $p = ab$. Since (p) is a prime ideal, then either $a \in (p)$ or $b \in (p)$. If $b \in (p)$. Then $b = rp$ for some $r \in R$, and if $a \in (p)$ then $a = pr$ for some $r \in R$. Without loss of generality, let's say $a \in (p)$. Then

$$p = ab = prb$$

and since R is an integral domain, the cancellation property holds (proposition 9.1.14), and we get

$$1 = rb$$

implying r, b are units, and so b is a unit. The same argument applies if we chose b. Thus, p is irreducible.

Thus, being prime is stronger than being an irreducible. The natural question to ask now is whether all elements in an integral domain can be decomposed into irreducible or prime factors (if it is factored into prime factors, then automatically it is factored into irreducible factors). It might already be clear that some Integral domains we have explored have this property (ex. $\mathbb{Z}$, and if you're familiar with some manipulation of polynomials, then $\mathbb{F}[x]$). This is however not the case for all integral

domains:

Example 10.1.3: Domain with no Factorization

Let $R = \overline{\mathbb{Z}} \subseteq \mathbb{C}$, where $\overline{\mathbb{Z}}$ consists of all elements in $r \in \mathbb{C}$ such there exists a monic polynomial with integer coefficients

$$a_0 + a_1 x + a_2 x^2 + \cdots + x^n \qquad a_i \in \mathbb{Z}$$

such that

$$a_0 + a_1 r + a_2 r^2 + \cdots + r^n = 0$$

In other words, $\overline{\mathbb{Z}}$ is the set of elements which are the root of a monic polynomial with integer coefficients (i.e. $\mathbb{Z}[x]$). The ring $\overline{\mathbb{Z}}$ is called the *integral closure* of $\mathbb{Z}$ over $\mathbb{C}$ and will be extensively studied in chapter 22 (algebraic number theory). For now, we just need to show that it is indeed an integral domain, and there exists elements that do not factor.

Since $\overline{\mathbb{Z}}$ is a subring of $\mathbb{C}$, we can use the subring test, that is, see if $\overline{\mathbb{Z}}$ is closed under multiplication and passes the subgroup test. It should be clear that is passes the subgroup test, since for any $a, b \in \overline{\mathbb{Z}}$, there exists polynomials

$$a_0 + a_1 x + \cdots + x^n \qquad b_0 + b_1 x + \cdots + x^n$$

(where some coefficients might be zero to match the length) such that

$$a_0 + a_1 \cdot a + \cdots + a^n = 0 \qquad b_0 + b_1 \cdot b + \cdots + b^n = 0$$

then

$$(a_0 - b_0) + (a_1 - b_1)x + \cdots + x^n$$

is a polynomial that has root $a - b$. By a similar argument, $\overline{\mathbb{Z}}$ using the binomial theorem is closed under multiplication.

Note that $\overline{\mathbb{Z}}$ is an integral domain since $\mathbb{C}$ has no zero divisors (hence $\overline{\mathbb{Z}}$ can't have any either), however it is not a field. Notice that $2 \in \overline{\mathbb{Z}}$ (2 is the root of $x - 2$). However, $1/2 \notin \overline{\mathbb{Z}}$. If it were, then there would be a monic polynomial with root $1/2$. However, any monic polynomial would be of the form

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

if we plug in $1/2$, we get

$$\left(\frac{1}{2}\right)^n + a_{n-1}\left(\frac{1}{2}\right)^{n-1} + \cdots + a_1\left(\frac{1}{2}\right) + a_0$$

with a little more work (which I suggest the reader try out), this can be shown to never has a root (start with $x + b$ to see why it doesn't have a solution, and try to generalize from there).

However, $\overline{\mathbb{Z}}$ has elements (in fact all non unit elements) that do not factor! Take $r \in \overline{\mathbb{Z}}$. Then there exists integers $a_0, a_1, ..., a_{n-1} \in \mathbb{Z}$ such that

$$a_0 + a_1 r + a_2 r^2 + \cdots + a_n r^n = 0$$

Since $\overline{\mathbb{Z}} \subseteq \mathbb{C}$, the square root of any number is well-defined (see exercise ref:HERE (in Introduction to Rings)). Thus $\sqrt{r}$ is well-defined. Further, $\sqrt{r} \in \overline{\mathbb{Z}}$ since

$$a_0 + a_1 \sqrt{r}^2 + a_2 \sqrt{r}^4 + \cdots + a_n \sqrt{r}^{2n} = 0$$

Since $1/\sqrt{r} \notin \overline{\mathbb{Z}}$ (as similar argument to why $1/2$ is not in $\overline{\mathbb{Z}}$) it is not a unit, and hence $r = \sqrt{r}\sqrt{r}$ factors into two non-units. Now here is the key: we can repeat this process for $\sqrt{r}$, showing that $\sqrt{r} = \sqrt{\sqrt{r}} \cdot \sqrt{\sqrt{r}}$, and so on and so forth. This process never terminates, showing that $\overline{\mathbb{Z}}$ has elements that do not factor!

For more examples, see:

`https://math.stackexchange.com/questions/96790/integral-domain-that-is-not-a-factorization-domain`

For this reason, we must define a special subclass of integral domains which do admit a factorization into irreducibles for every element:

Definition 10.1.6: Atomic Domains

Let R be an integral domain. Then R is an *atomic domain* (or *domain with factorization*) if for every non-unit nonzero element $r \in R$, there exists irreducible elements $q_1, q_2, ..., q_n$ such that

$$r = q_1 q_2 \cdots q_n$$

that is, every element has a *factorization* (or *decomposition*) into irreducible

The name atomic is due to every element being decomposable into element that can no longer be subdivided, similar to atoms, where the word atom comes from the Greek *atomos* meaning "not cuttable". Note that the q_i's need not be distinct (for example, we will show that $\mathbb{Z}$ is an atomic domain, and that $4 = 2 \cdot 2$). Usually, we are more interested in domains where the factorization into irreducible elements is unique:

Definition 10.1.7: Unique Factorization Domain (UFD)

Let R be an integral domain. Then R is a *Unique Factorization Domain* (commonly abbreviated to UFD) if for every non-unit nonzero element $r \in R$, there exists irreducible elements $q_1, q_2, ..., q_n$ and unit u_r such that

$$r = u_q q_1 q_2 \cdots q_n$$

and if there is another set of irreducible element $q'_1, q'_2, ..., q'_m$ such that $r = u_{q'} q'_1 q'_2 \cdots q'_m$, then $n = m$, and the two products differ by a unit after some reshuffling (i.e. they are associates), that is, for appropriate $u \in R$

$$q_1 q_2 \cdots q_n = u q'_{\sigma(1)} q'_{\sigma(2)} \cdots q'_{\sigma(n)}$$

where $\sigma(i)$ is the appropriate permutation

The following corollary is stated for the sake of reference:

Corollary 10.1.8: Unique Factorization Domain $\Rightarrow$ Atomic Domain

Let R be a Unique Factorization Domain. Then R is an Atomic domain

***Proof* :**
By definition, every element in a Unique Factorization Domain has a factorization into irreducibles, and so is an Atomic Domain

Note that "unique" means unique up to unit, as we pointed out earlier. The factorization in both definitions could have had the unit dropped and just have been written $r = q_1, q_2, ..., q_n$ since it is a the result is an associate to $u^{-1}r$, and so it is harmless to drop the unit. However, I chose to keep the unit in the definition since it may lead to certain unfortunate mistakes when factoring concrete elements.

Example 10.1.4: Unique Factorization Domains

1. Every field is trivially a Unique Factorization Domain, since every nonzero element is a unit, so there are no elements for which the two properties must be verified. In some textbooks, fields are purposefully omitted as Unique Factorization Domains to omit mentioning them as exceptions in proofs. They are still considered Unique Factorization Domains in this book since they fit well in the diagram presented at the beginning of this chapter. However, in a sense, this means that in a field there is no "unique factorization", or at least it's vacuously true that all elements can be uniquely factorized since no nonzero elements are non-units.

2. As we saw in example 11.2.1, $F_{0x} \subseteq \mathbb{F}[x]$ which has x coefficient of 0 is not a Unique Factorization Domain. We saw that x^2 and x^3 are irreducible. If we take x^6, then has two factorizations: $x^6 = x^2x^2x^2$ and $x^6 = x^3x^3$, and hence the factorization is not unique.

3. The ring $\mathbb{Z}[2i]$ is *not* a Unique Factorization Domain. $4 = 2 \cdot 2 = ((-2i) \cdot 2i)$. The elements 2 and $2i$ are irreducible, but not associates, since $2 = 2 \cdot i$, but $i \notin \mathbb{Z}[2i]$. They are furthermore not prime, since $\mathbb{Z}[2i]/(2i) \cong \mathbb{Z}/4\mathbb{Z}$ which is not an integral domain (see proposition 9.4.13)

4. In the ring $\mathbb{Z}[i]$, 2 and $2i$ *are* associates. We have $i \in \mathbb{Z}[i]$, and $i \cdot i^3 = 1$ showing that i is a unit, and hence 2 and $2i$ are indeed associated. We'll soon show that $\mathbb{Z}[i]$ is also a Unique Factorization Domain (see section 10.3)

5. We shall prove in chapter 11 that if R is a UFD, then $R[x]$ is a UFD. Note how this contrasts with PID's ($\mathbb{Z}$ is a PID but $\mathbb{Z}[x]$ is not, see example 9.4.2[3]).

All of these are examples of Unique Factorization Domains. Atomic Domains which are not Unique Factorization Domains are rather difficult to find. Fortunately, there is an easy criterion that allows us to find (a subset of) common atomic domains

10.1.1 Example of Atomic Domains: Noetherian Domains

The goal of this section is to show that Noetherian domains are atomic domains, and that there exists noetherian domains which are not Unique Factorization Domain's. Let's first remember the definition of a Noetherian Domain, or more generally a Noetherian ring:

Definition 10.1.9: Noetherian Rings

Let R be ring. Then if every ideal $I \subseteq R$ is finitely generated, then R is called a *Noetherian Ring*. If R is an integral domain, we call R a *Noetherian Domain*

Most results that will be presented apply to Noetherian rings in general, and so we will work with Noetherian rings. The first thing to show is how to relate Noetherian rings to the idea of factorization. Emmy Noether discoverd in HERE a condition which let's us characterize this property[2]:

Proposition 10.1.10: Ascending Chain Condition

The following are equivalent:

1. R is a Noetherian ring
2. for every chain of ascending left-ideals of R

$$I_1 \subseteq I_2 \subseteq \cdots \subseteq I_n \subseteq \cdots$$

   there exists an number m such that for all $n > m$, $I_n = I_{n+1} = \cdots$
3. Every nonempty family of left-ideals of R partially ordered by inclusion has a maximal left-ideal

Note that replacing left by right or two-sided ideals will not break the theorem. If R is an integral domain, then there is no distinction between these concepts

Proof **:**

(1) $\Rightarrow$ (2) Let R be a Noetherian ring. If $I_1 = R$ we're done, so assume I_1 is a proper ideal of R, and consider any ascending chain of left ideals

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots \subseteq I_n \subseteq \cdots$$

Take

$$I = \bigcup_{i=1}^{\infty} I_i$$

which is an ideal for a similar reasoning as presented in the proof that every proper ideal is contained in a maximal ideal (see proposition 9.4.11). Since R is a Noetherian domain and $I \subseteq R$, then I is finitely generated. Let's say $I = (a_1, a_2, ..., a_n)$. Then these generators must each be in some ideal I_m, meaning that for all $n > m$, $I_{n+1} = I_{n+2} = \cdots$, showing that R respects the ascending chain condition

(2) $\Rightarrow$ (3) Let R respect the ascending chain condition on it's left ideals, and let $\mathcal{F}$ be a family of left-ideals of R. Pick some random $I_1 \in \mathcal{F}$. If I_1 is maximal in $\mathcal{F}$, then we'd be done, so let's say I_1 is not maximal in $\mathcal{F}$. Then there exists an I_2 such that $I_1 \subset I_2$. If I_2 is maximal in $\mathcal{F}$, then we're done. If not, then we can continue

[2] more on the generalization of it??

this process. If every element we pick is not maximal, we can create an infinitely long chain[a]

$$I_1 \subsetneq I_2 \subsetneq I_3 \cdots \subsetneq I_n \subsetneq \cdots$$

However, that contradicts the ascending chain condition, meaning for some n, the chain must stabilize. Thus, $\mathcal{F}$ must have a maximal element

(3) $\Rightarrow$ (1) Let $I \subseteq R$ be a left-ideal. Take $\mathcal{I}$ to be the set of all finitely-generated left-ideals contained in I. The set $\mathcal{I}$ is nonempty since $\{0\} \in \mathcal{I}$. It is thus a family of left-ideals of R and so by the 3rd condition, $\mathcal{I}$ admits a maximal,/Then if say I' [b]. If $I \neq I'$, then there exists an element $x \in I$ such that $x \notin I'$. But then $(I' \cup x)$ is a finitely generated left-ideal larger than I', contradicting the maximality of I'. Hence, $I = I'$, showing that I is finitely generated.

[a]technically, to create this infinitely long chain, we used the axiom of choice

[b]Note that Zorn's lemma can't be used here since it is not known whether every chain will have an upper bound in this set, their union potentially no longer being finitely generated

30: Remark

This proof will become a special case of the equivalent proof for modules in section 12.3.4.

For our purposes, this let's us think of Noetherian domains in terms of the ascending chain condition. You probably already see how this might start related to Factorization domains: If every element in the ring is finitely generated, it must be decomposable into a finite number of factors. This is essentially true:

Proposition 10.1.11: A.C.C. and Factorization

Let R be an integral domain, and let r be a nonzero nonunit element of R. Then if every ascending chain of *principal ideals*

$$(r) \subseteq (r_1) \subseteq (r_2) \subseteq \cdots \subseteq (r_n) \subseteq$$

stabilizes (there exists an m such that for all $l > m$, $(r_l) = (r_{l+1}) = \cdots$), then r can be factored into irreducible elements. In particular, if R respects the A.C.C. on principal ideals, then R is an atomic domain.

Re-writing these ideal using the divisibility notation, it may be more evident: if $r_1 | r$, $r_2 | r_1$, and so on eventually stabalizes, then r can be factored into irreducible elements.

***Proof* :**

Let's prove this using the contrapositive ($\neg B \Rightarrow \neg A$), so that we'll show if r does not factor into irreducible then there exists an ascending chain of principal ideals that does not stabalize.

Since r is not irreducible, then there exists two nonunit non zero element r_1, s_2 suc that $r = r_1 s_2$. Since r does not have a factorization, either r_1 or s_1 are reducible. Without loss of generality, let's say r_1 is reducible. Then $(r) \subseteq (r_1)$ since $s_2 \cdot r_1 = r$, and $r_1 = r_2 s_2$. The same being applied to r_1, we get $(r) \subseteq (r_1) \subseteq (r_2)$. Continuing on indefinitely, we get

$$(r) \subseteq (r_1) \subseteq (r_2) \subseteq \cdots \subseteq (r_n) \subseteq \cdots$$

which never stabilizes since r does not factor, as we sought to show.

Notice how similar this it to Noetherian rings: In a Noetherian ring, the ascending chain condition works for *all* left/right/two-sided ideals. Therefore, it certainly works when restricted to only it's *principal ideals*. We therefore get this immediate corollary

Corollary 10.1.12: Noetherian Domain $\Rightarrow$ Atomic Domain

Let R be a Noetherian Domain. Then it is an Atomic Domain

***Proof* :**
By proposition 10.1.10, R respects the ascending chain condition on any left ideals. Since any element $r \in R$ forms a principal ideal, any ascending chain of principal ideals must stabalize. Thus, by proposition 10.1.11 r must have a factorization. Since r was arbitrary, every element has a factorization, and so R is an Atomic Domain as we sought to show

It might seem like the converse is clearly true: If R is an atomic domain, then R respects the ascending chain condition on principal ideals. P.M.Cohn made this same assertion in 1968 during his research on rings[3], however it was ultimately proven to not be the case: there are element for which there are infinitely many factorizations. Though each factorization is finite, there are factorizations for every natural number (see exercise ref:HERE, maybe?). This theorem is still useful to our endeavour, since it let's us link a finite generation property (A.C.C. forces all principal ideals to be contained in at most a finitely generated ideal[4]) to a factorization condition (a condition for how to decompose elements into simpler elements).

Unfortunately, there isn't a simple counter-example I know of until we prove the famous Hilbert's Basis Theorem in section 20.2.1. For the reader who wants to know the counter-example, they can either skip to Hilbert's basis theorem on page 716 (you have everything you need to know to understand it now), or take on faith that "if R is a Noetherian Domain then $R[x_1, x_2, ..., x_n]$ is a Noetherian Domain". The following is a counter-example:

Example 10.1.5: Atomic Domain $\not\Rightarrow$ Noetherian Domain
Let $R = \mathbb{Z}[x_1, x_2, ..., x_n, ...]$ be a polynomial ring over an infinite number of variables (see section 9.2). Notice that $\mathbb{Z}$ is a Noetherian Domain: As we have shown, $\mathbb{Z}$ a Principal Ideal Domain (example 9.4.2, and so any ideal of $\mathbb{Z}$ is generated by one element, and so it is Noetherian. By Hilbert's Basis Theorem, $\mathbb{Z}[x_1, x_2, ..., x_n]$ is a Noetherian Domain as well.

Now, any polynomial $f \in R$ has finitely many terms by definition, and so is also an element of the subring $\mathbb{Z}[x_{\sigma(1)}, x_{\sigma(2)}, ..., x_{\sigma(n)}]$ where $\sigma(i)$ is some appropriate permutation function so that the terms of f are in the aforementioned subring. But then it is an element of a Noetherian Domain, which means it could be factored by proposition 10.1.11. Since $\mathbb{Z}[x_{\sigma(1)}, x_{\sigma(2)}, ..., x_{\sigma(n)}] \hookrightarrow R$, the same factorization happens in R, and since f was arbitrary, R is an Atomic Domain

However, R is not a Noetherian Domain. The ideal $(x_1, x_2, ..., x_n, ...)$ of all of the terms of R is not finitely generated (it does not respect the ascending chain condition).

[3] P.M. Cohn, Bézout rings and their subring; Proc. Camb. Phil.Soc. 64 (1968) 251–264
[4] assuming the Axiom of Choice

This means that in the image presented in the beginning of the chapter, the strict inclusion of Noetherian Domains in Atomic Domains is justified. Similarly, Being an Atomic Domain does not mean the factorization is unique

Example 10.1.6: Atomic Domain $\not\Rightarrow$ Unique Factorization Domain
Take the ring F_{0x} from example 10.1.2. Take any element $f \in F_{0x}$, and form a ascending chain of principal ideals. For the sake of contradiction, let's say the chain never stabilizes: $(f) \subsetneq (f_1) \subseteq (f_2) \subsetneq \cdots \subseteq (f_n) \subsetneq \cdots$. Then since $(f_i) \subsetneq (f_j)$, and every degree zero element is a unit (recall that $F_{0x} \subseteq \mathbb{F}[x]$ for a field $\mathbb{F}$) then the degree of f_j must be *less* than that of f_i. This means we have a decreasing chain:

$$\deg(f) > \deg(f_1) > \deg(f_2) > \cdots > \deg(f_n) > \cdots$$

Notice importantly that this chain *must* terminate since there are no polynomials of degree 0. Since f was arbitrary, every element has a factorization and so it is an Atomic Domain.

However, as we have shown in example 10.1.2, $x^6 = x^2x^2x^2 = x^3x^3$ are two factorizations of x^6 where x^2 and x^3 are not associates, showing it is not a Unique Factorization Domain.

Notice the ring F_{0x} is not a Noetherian Domain in the previous example. Using Hilbert's Basis Theorem, we can also find a Noetherian Domain which is not a UFD:

Example 10.1.7: Noetherian Domain $\not\Rightarrow$ Unique Factorization Domain
Let

$$R = \frac{\mathbb{C}[x, y, z, w]}{(xw - yz)}$$

be the polynomial ring over 4 variables quotiented by the ideal $(xw - yz)$ [a]. By Hilbert's Basis Theorem, $\mathbb{C}[x, y, z, w]$ is a Noetherian ring. By proposition 9.4.4, $\frac{\mathbb{C}[x,y,z,w]}{(xw-yz)} = R$ is a noetherian ring. It is also an integral domain, since $(xw - yz)$ is a prime ideal (see exercise ref:HERE. It is therefore a Noetherian Domain (and hence an Atomic Domain).

However, it is not a Unique Factorization Domain. The element x, y, z, w are all irreducible, and not associates. Since $xw - yz = 0$, we have $r = xw = yz$ has two distinct factorization, showing that R is not a Unique Factorization Domain.

[a] in chapter 21, we will see this ring be called the *quadratic cone in* $\mathbb{A}^4$

The converse is not true either: being a Unique Factorization Domain does not imply it is a Noetherian Domain. In fact, the ring $\mathbb{Z}[x_1, x_2, ..., x_n, ...]$ from example 10.1.5 is also a Unique Factorization Domain that is not a Noetherian Domain. This proof relies a theorem similar to Hilbert's Basis Theorem (in particular a generalization of theorem 11.2.3 on page 331). However, this theorem requires much longer build-up, and so for now the reader can take it on faith that such a ring is a counter-example (see exerise ref:HERE for a way to prove this using gcd domains).

Thus, we have shown (or claimed) that

$$\begin{array}{rcl@{\qquad}rcl}
\text{Noetherian Domain} & \Rightarrow & \text{Atomic Domain} & \text{Atomic Domain} & \not\Rightarrow & \text{Noetherian Domain} \\
\text{UFD} & \Rightarrow & \text{Atomic Domain} & \text{Atomic Domain} & \not\Rightarrow & \text{UFD} \\
\text{UFD} & \not\Rightarrow & \text{Noetherian Domain} & \text{Noetherian Domain} & \not\Rightarrow & \text{UFD}
\end{array}$$

This completes some of the layers in the diagram presented at the beginning of this chapter.

:exercise

Exercise 10.1.1

1. Let R be a finitely generated ring. Does that imply R is Noetheria
2. Show that $\mathbb{C}[x, y, z, w]/(xw - yz)$ is an integral domain.
3. Show that if $M \subseteq R$ is a maximal ideal then M is Noetherian. In fact, if the prime ideals are finitely generated, then R is Noetherian.

10.1.2 Unique Factorization Domains

We know continue exploring Atomic Domains in which the factorization of each element is unique (up to unit). As a reminder, here is the definition of a UFD:

Definition 10.1.13: Unique Factorization Domain (UFD)

Let R be an integral domain. Then R is a *Unique Factorization Domain* (commonly abbreviated to UFD) if for every non-unit nonzero element $r \in R$, there exists irreducible elements $q_1, q_2, ..., q_n$ such that

$$r = q_1 q_2 \cdots q_n$$

and if there is another set of irreducible element $q'_1, q'_2, ..., q'_m$ such that $r = q'_1 q'_2 \cdots q'_m$, then $n = m$, and the two products differ by a unit after some reshuffling (i.e. they are associates), that is, for appropriate $u \in R$

$$q_1 q_2 \cdots q_n = u q'_{\sigma(1)} q'_{\sigma(2)} \cdots q'_{\sigma(n)}$$

where $\sigma(i)$ is the appropriate permutation

Perhaps the most intuitive way of thinking about elements in Unique Factorization Domains is by linking to every element of the UFD a multiset (a set that allow's multiplicity) of irreducible factors. Units and 0 are both associated to empty multisets. The multiplication structure can also be thought of as forming a free abelian monoid structure with all prime elements as generators.

Lemma 10.1.14: Irreducible and Multisets

Let R be a Unique Factorization Domain and let $a, b, c \in R$ be nonzero elements of R. Then

1. $(a) \subseteq (b)$ if and only if the multiset of irreducible elements of a is contained in the multiset of irreducible elements of b
2. a and b are associates if and only if their corresponding multisets are equal
3. if $a = bc$, then the corresponding multiset of a is the union of the multisets of b and c

This lemma essentially shows that working with elements of a Unique Factorization Domain comes down to working with their corresponding multiset, making the analysis of their structure more

often than not being more set theoretical or number theoretical. As an example of this, Unique Factorization Domains make it very easy to define the concept of *greatest common divisor*.

Greatest Common Divisor

The set of sets of multi-sets thus has a natural lattice paralleled on rings given by the divisibility. In fact, meet and joins are well-defined[5]:

Definition 10.1.15: Greatest Common Divisors In Commutative Rings

Let R be a commutative ring and let $a, b \in R$ with $b \neq 0$. Then the *greatest common divisor* of a and b is a nonzero element d such that

1. $d|a$ and $d|b$,
2. if $d'|a$ and $d'|b$ then $d'|d$

A greatest common divisor of a and b will be denoted $gcd(a, b)$ or (a, b)

Notice that in the definition, R was a commutative ring, since divisibility was defined for commutative rings in general. It should be noted that the gcd of two elements is unique *up to associativity*. it is helpful to put the definition of gCD in ideal-interpretation of divisibility:

Definition 10.1.16: GCD Using Principal Ideals

Let R be a commutative ring and $(a), (b) \subseteq R$. Then $\gcd(a, b)$ is an ideal $I \subseteq R$ such that $(a, b) \subseteq I$ and I is the smallest ieal that contains (a, b). In other words

$$\gcd(a, b) = (a) + (b)$$

As an exercise, see that this does indeed match the previous definition of gcd, and test it in $\mathbb{Z}$. There do exist rings for which the gcd between two elements do not exist (in example 10.1.2, take the ring R_{0x} from the second part of the example. Prove that the $\gcd(x^5, x^6)$ does not exist, see exercise ref:HERE). For Unique Factorization Domains, every element has a gcd:

Proposition 10.1.17: GCD in Unique Factorization Domain

let R be a Unique Factorization Domain. Then for any $a, b \in R$, $\gcd(a, b)$ exists. Furthermore, given

$$a = up_1^{n_1} p_2^{n_2} \cdots p_k^{n_k} \qquad b = vp_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}$$

is the factorization of a and b, that is, every p_i is irreducible, u and v are associates, every p_i is not associates to p_j, and some n_i's or m_i's can be zero. Then the gcd (up to a unit) is:

$$\gcd(a, b) = p_1^{\min(n_1, m_1)} p_2^{\min(n_2, m_2)} \cdots p_k^{\min(n_k, m_k)}$$

[5]see section 0.3.1

***Proof* :**
This is an almost immediate consequence of using the factorization of a and b. Let d equal the above equation. Then clearly $d|a$ an $d|b$. If there is some c that also divides a and b ($c|a$ and $c|b$), then $a = ca'$ and $b = cb'$ for appropriate a' and b'. By lemma 10.1.14, c must be contained in both the multisets of a and b, and therefore

$$c = wp_1^{\ell_1} p_2^{\ell_2} \cdots p_k^{\ell_k}$$

where w is a unit and $\ell_i \leq n_i$ and $\ell_i \leq m_i$. Then $\ell_i \leq \min(n_i, m_i)$ for all i, and so it is clear that $(c) \subseteq (d)$, and by lemma 10.1.14 $c|d$, as we sought to show

31: Relation Between Lattice's and UFDs

It should be noted that there are integral domains which are not Unique Factorization Domain's but do have the concept of gcd defined for every element. The class of all integral domains which have the concept of gcd defined for every element is broader than Unique Factorization Domain's and are not always Noetherian, and so such domains need their own name and are called *gcd domains*. In a sense, gcd domains are a generalization of Unique Factorization Domains: A gcd domain is Unique Factorization Domain if and only if the gcd domain satisfies the ascending chain condition on principal ideals (similarly to how Atomic domains do). However, a gcd domain isn't always an Atomic domain, and hence fits a bit awkwardly in our current hierarchy. In essence, this is because gcd domains hold information on information between elements, while atomic domains hold element-wise information.
For future purposes, gcd domains do fit into an important hierarchy of domains: *normal domains* shall be introduced in chapter 20, definition 20.5.2, and all gcd are normal domains:

$$\text{normal domains} \supseteq \text{gcd domains} \supseteq \text{UFD's}$$

The definition of GCD's using ideals suggests that the GCD, when it exists, ought to be unique in general, and indeed it is:

Proposition 10.1.18: Uniqueness Of Greatest Common Divisor (In Integral Domain)

Let R be an integral domain. If (d) and (d') are both greatest common divisors of then $(d) = (d')$.

***Proof* :**
If d and d' are greatest common divisors, then they must each divide each other (condition 2 of definition 10.1.15), and so we'll get $(d') \subseteq (d)$ and $(d') \supseteq (d)$, meaning $(d') = (d)$,

Just like for the integers, we may also define the dual of the GCD, the least common multiple. In the following, the LCM will be defined in terms of principle ideals, the similar definition should be translated into elements:

Definition 10.1.19: Least Common Multiple (lCM)

Let R be a commutative ring and $(a), (b) \subseteq R$. Then the *least common muliple* is the ideal I such that $I \subseteq (a)$ and $I \subseteq (b)$ and is the largest such ideal. Namely

$$\operatorname{lcm}((a), (b)) = (a) \cap (b)$$

Properties of UFD's

As we have shown in example 10.1.2 and 10.1.7 being an irreducible element in an integral domain does not imply that the element is a prime element. This was because prime elements are "stronger" than irreducible in that they not only have to be able to be divisible, but they have to be "globally divisible" for any factorization. However, since factorization is unique in a Unique Factorization Domain, it should come as no surpise that there is no distinction between prime elements and irreducible elements:

Proposition 10.1.20: Prime If And Only If Irreducible In UFD

Let R be a Unique Factorization Domain. Then a nonzero element is a prime if and only if it is irreducible

Proof **:**
The fact that prime elements are irreducible was proven in proposition 10.1.5, so we'll show the converse. Let a be an irreducible element. Then a is not a unit, and if $a = bc$ then either b or c is a unit. Now, assume $bc \in (a)$ so that $d \cdot a = bc$. Then $(bc) \subseteq (a)$, and by lemma 10.1.14, the set of the irreducible factors of a (i.e. a itself) is contained in the set of irreducible factors of b and/or c. Therefore, it must be that a is a factor of b, or c, or both. If it's factors are within the factors of b, then $b \in (a)$, and if it's factors are within the factors of c, then $c \in (a)$. If in both, then $b, c \in (a)$. In any case, at least one of b or c are in (a), showing that (a) is prime.

Thus, if we see every irreducible element being prime as representing "global" irreducibility (or divisibility), then if we further assume the A.C.C. on principle ideals, we see that this ought to fully characterize UFDs. The following theorem demonstrate exactly this:

Theorem 10.1.21: Unique Factorization Domain Equivalence

Let R be an integral domain. Then R is a Unique Factorization Domain if and only if

1. The ascending chain condition for principal ideal holds in R
2. Every irreducible element is a prime element

Proof **:**

($\Rightarrow$) Let R be a Unique Factorization Domain. Then by proposition 10.1.20 every irreducible element is a prime element, and by corrollary 10.1.8 R is an atomic domain, and so by proposition 10.1.11 R respects the ascending chain condition

(this can also be proven directly using the corresponding multisets to elements)

($\Leftarrow$) Let R satisfy the A.C.C. for principal ideals and let every irreducible be prime. By proposition 10.1.11, R is an Atomic domain. It remains to show that R factors uniquely. To show this, let n be a nonunit and

$$n = p_1 p_2 \cdots p_n = q_1 q_2 \cdots q_m$$

be two factorizations of n into irreducibles. Then since $q_2 q_3 \cdots q_n \cdot q_1 = p_1 p_2 \cdots p_n$, we have $p_1 p_2 \cdots p_n \in (q_1)$. Since q_1 is irreducible, by the 2nd condition it is a prime, and so for some i, $p_i \in (q_1)$. Without loss of generality, we can say $p_1 \in (q_1)$ by just rearranging and relabelling the factors. Since $p_1 \in (q_1)$, then $u_1 \cdot q_1 = p_1$. Since p_1 is irreducible and q_1 is not a unit, necessarily u_1 is a unit. Then, we can re-write the equation as

$$u_1 q_1 p_2 \cdots p_n = q_1 q_2 \cdots q_m$$

Cancelling out q_1 from each side and replacing p_2 by $u_1 p_2$ produces us

$$p_2 p_3 \cdots p_n = q_2 q_3 \cdots q_m$$

from here, we can continue this process for the rest of the p_i's, where we will end up with

$$u_1 u_2 \cdots u_n q_1 q_2 \cdots q_m = q_1 q_2 \cdots q_m$$

leting $u = u_1 u_2 \cdots u_n$, we get $u q_1 q_2 \cdot q_n = q_1 q_2 \cdots q_n$. Note that $n = m$ since if not then at some point the remaininng part of the left hand side will be equal to 1, a contradiction to n not being a unit. Thus, the two factorization are equal up to associates, as we sought to show.

Notice how the uniqueness of the factorization allowed us to make the A.C.C. for principal ideals a defining feature of Unique Factorization Domains, in contrast to ho the A.C.C. for principal ideals was insufficient to define Atomic Domains.

This theorem is also very useful in easily remembering how Unique Factorization Domain's and Noetherian Domains are different: If we start with a domain with the A.C.C. for principal ideals (i.e. the types of Atomic domains we ill work with) strengthen the A.C.C. condition to work on all ideals, we get a Noetherian domain. Conversely, if we are working with a Noetherian domain and we loosen the A.C.C. condition to principal ideals but add that all irreducible are primes, then we get a Unique Factorization Domain.

10.1.3 Important UFD: Principle Ideal Domains

Principal Ideal Domains are stronger than Noetherian Domains in the sense that every ideal is generated by a single element (rather than finitely many element for Noetherian Domains). They are therefore immediately Atomic Domains by proposition 10.1.11. By theorem 10.1.21, if a Principal Ideal Domain also has that every irreducible element is prime, then it is automatically a Unique Factorization Domain. Are irreducible elements prime in a Principal Ideal Domain? The answer is yes; hence PIDs form a subclass of UFDs. Take a moment to think about why irreducible elements are prime in PIDs before proceeding to the next proposition (hint: can you prove prime ideals *must*

be maximal ideals in a PID?). From this proof, there will now be a new set of examples for Unique Factorization Domains similarly to how all Noetherian domains are an easy set of examples for Atomic domains

32

Taking a moment

As the hint suggests, we first prove the following important intermediary result:

Proposition 10.1.22: In PID, Prime Ideal iff Maximal Ideal

Let R be a Principal Ideal Domain and take a nonzero ideal $(r) \subseteq R$. Then (r) is a prime ideal if and only if it is a maximal ideal.

Intuitively, since

Proof **:**
Let R be a Principal Ideal Domain. If an ideal of R is maximal, by corollary 9.4.14 that ideal is prime. Thus we'll prove the converse.

Let $(p) \subseteq R$ be a prime ideal, and assume that $I \supseteq (p)$ is an ideal that contains (p). We must show that $I = (p)$ or $I = R$. Since R is a Principal Ideal Domain, there exists an $m \in R$ such that $(m) = I$. Since $(m) \supseteq (p)$, $p \in (m)$, and so $p = rm$. Since p is a prime element, then either $r \in (p)$ or $m \in (p)$. If $m \in (p)$, then $(m) = (p)$, fulfilling the $I = (p)$ case. If $r \in (p)$ then $r = sp = ps$ and so

$$p = rm = psm$$

And by the cancellation property of integral domains, $1 = sm$, and so $1 \in (m)$ implying that $(m) = R$, fulfilling the $I = R$ condition. That means that (p) fulfills the conditions of being a maximal ideal, as we sought to show.

The main idea is that any ideal requries only 1 element to repersent it, which as it will turn out will be key to why elements will factor uniquely.

Proposition 10.1.23: Prime If And Only If Irreducible In PID

Let R be a Principal Ideal Domain. Then a nonzero element is a prime if and only if it is irreducible. Along with the fact that Principal Ideal Domain are Noetherian Domains (and hence respecting the A.C.C. on principal ideals) we have

$$\text{PID} \Rightarrow \text{UFD}$$

Proof **:**
Since a Principal Ideal Domain is an integral domain, we've just shown in proposition 10.1.5 that prime elements are irreducible, so we'll show the converse.

Let's say p is an irreducible element of R. To show it's prime, we must how (p) is a prime ideal. Since we're working in a Principal Ideal Domain, by 10.1.22, it's equivalent to show that (p) is a maximal ideal. So, for the sake of contradiction, let's assume (p) is not maximal, and so there

exists a strictly larger ideal that is the maximal ideal containing (p) $(p) \subsetneq (m) \neq R$ for appropriate $m \in R$. Then

$$p = am$$

for some $a \in R$. Since p is irreducible, then either a or m is a unit in R. If M is a unit, then $m^{-1}m = 1 \in (m)$ implying $(m) = (1) = R$, but then it's not a maximal ideal, since maximal ideals are proper ideals. Conversely, if a is a unit, then by proposition p and m are associates, meaning $(p) = (m)$. But since (m) was maximal, that makes (p) maximal. Thus, (p), by proposition 10.1.22, (p) is a prime ideal.

Therefore, since all Principal Ideal Domain are Noetherian Domains and Noetherian respect the A.C.C. on principal ideals, by thereom 10.1.21, all Principal Ideal Domain are Unique Factorization Domains, as we sought to show.

There is also a very simple proof I like I found in Eisenbud: (Eisenbud) We know that if R is a Unique Factorization Domain, then it respects the A.C.C. on principal ideals. Use that to find a maximal ideal, which must be generated by one element since it's a Principal Ideal Domain, but then it must be that for one of the elements in the chain, it plateau's, and so it fulfills the ascending chain condition, and so it is a Unique Factorization Domain.

Corollary 10.1.24: $\mathbb{Z}$ is a UFD (Fundamental Theorem of Arithmetics)

The integers $(\mathbb{Z})$ as a ring are a Unique Factorization Domain.

Proof **:**
As shown in example ref:HERE (under defn of pid), $\mathbb{Z}$ is a Principal Ideal Domain, and so by proposition 10.1.22 is a Unique Factorization Domain.

The converse of this theorem is not true, that is if an integral domain is a Unique Factorization Domain, this does not imply that it is a PID, justifying the strict inclusion of the image at the very beginning. Recall from example 9.4.2 that the ideal $(2, x)$ from the ring $\mathbb{Z}[x]$ is not a principal ideal. However, the ring $\mathbb{Z}[x]$ is a Unique Factorization Domain. This is rather difficult to show directly: if we have any any polynomial $p(x) \in \mathbb{Z}[x]$, how can we guarantee it factors and that it's unique up to units, that is that $\mathbb{Z}[x]$ respects the ascending chian conditions on principal ideals and that all irreducibles are primes? On page 331, we shall show this result.

(Also something about begging the question now on finding qualities that allow for PIDs? For example, in $\mathbb{Z}$, we figured out it's a pid because for every number, we can make it a linear combination of other numbers. Is that so for every pid?)

([perhaps make this an exercise?] As a final comparison between UFDs and PIDs, the Chinese Remainder Theorem (CRT). Recall that in a pid, $\gcd(a, b) = 1$ if and only if $(a, b) = (1)$. This is *not* the case for UFD's, take for example the natural map

$$\mathbb{Z}[x] \to \mathbb{Z}[x]/(2) \times \mathbb{Z}[x](x)$$

Then this map is *not* surjective, though $\gcd(2, x) = 1$, even when the kernel, $(2x)$, is still the product of the two ideals.)

Proposition 10.1.25: Sufficient Condition for Greatest Common Divisor

If a and b are nonzero elements in the commutative ring R such that the ideal gnerated by a and b is a principal ideal (d), then d is a greatest common divisor of a and b

Proof **:**
p.274

This proposition explains the (a, b) notation for the gcd of a and b.

Note that an integral domain in which every ieal generated by 2 elements (a, b) is a principal ideal is called a *Bézout Domain* (more generally, if every *finitely generated* ideal is principal, then we have a Bézout Domain. For Principal Ideal Domains, every ideal is principal). Note that it's not a *necessary* condition.

For future reference, we also prove the following[6]:

Proposition 10.1.26: UFD And Maximal, Then PID

Let R be a Unique Factorization Domain where every prime ideal is maximal. Then R is a Principal Ideal Domain

Proof **:**
Let R be a Unique Factorization Domain where every prime ideal is maximal. First, note that if $I \subseteq R$ is prime, then I is principle. Indeed, for some $r \in I$, then (up to unit) we have $r = p_1, p_2, ..., p_n$. By primality, we have that $p_i \in I$. Since p_i is prime, by assumption it is maximal, and furthermore $(p_i) \subseteq I$. But then $(p_i) = I$.

Now, let's N be the set of all non-principal ideals. For the sake of contradiction let's say it's non-empty. Then N is partially ordered by inclusion and every chain has an upper bound: if $J = \bigcup_i J_i$, then if $J = (x)$, then $x \in J_i$ for some i and $(x) = J$; contradicting non-principality.

Hence, by Zorn's lemma there exists a maximal element $\mathfrak{n} \in N$. Then by what we pointed out earlier $\mathfrak{n}$ cannot be prime, or else it would be principal. Therefore there exists $a, b \in R$, $a, b \notin \mathfrak{n}$ but $ab \in \mathfrak{n}$. Consdier $\mathfrak{n} + (a)$ and $\mathfrak{n} + (b)$. By maximality of $\mathfrak{n}$ hese must be principal: $\mathfrak{n} + (a) = (u)$ and $\mathfrak{n} + (b) = (v)$. But then:

$$\mathfrak{n} = \mathfrak{n} + (ab) = (\mathfrak{n} + (a))(\mathfrak{n} + (b)) = (u)(v) = (uv)$$

but that's a contradiction to $\mathfrak{n} \in N$, showing that it must have been that $N = \emptyset$, completing the proof.

(also thought this was cool here)

Exercise 10.1.2

1. Let $R[x]$ be a Principal Ideal Domain. Then R is a UFD. Conclude that if R is not a field, then $R[x]$ is not a Principal Ideal Domain!

[6] This will be useful in number theory

Since $\mathbb{Z}$ is not a field, then $\mathbb{Z}[x]$ is not a Principal Ideal Domain! In example 9.4.2[3] it was manually shown that $(2, x)$ is not a principal ideal in $\mathbb{Z}[x]$; this theorem makes it easier to prove.

The converse is true to: If F is a field, $F[x]$ is a Principal Ideal Domain. In fact, it's even a Euclidean Domain. We'll cover this result when talking about polynomials over fields in section 11.1.

2. (Challenging) Let R be an integral domain. Then if every prime ideal is principle, R is a PID (In this way, UFDs are almost PIDs, however not all ideals are principle)
3. (If you know Complex Analysis) Show that the ring of entire functions is a gcd domain. Show that the ring of real analytic functions is a Bézout domain.

10.2 Euclidean Domains

So far, we have been exploring divisibility in terms of being able to directly factor elements. However, it is in general difficult to establish when an element is divisible by another (try it out yourself with relatively large elements of some of the more complicated UFDs that were presented). If an element is not divisible by another element is there some way to measure to what "degree" an element is not divisible by another? Can we quantify the lack of divisibility given by the gcd.

In the integers, there is the *euclidean algorithm* that gives the answer. The existence of an algorithm to find the GCD cannot be under-estimated: If given two integers with over a thousand digits, it will be impossible to find the greatest common divisor in a reasonable amount of time, that is, before the heat death of the universe (until a faster way of factorizing is found). The problem is about our initial choice for divisors. If I say R is a Unique Factorization Domain and $x \in R$, then how would you start finding it's prime factorization? If you have some a and b such that $ab = x$, How do you start decomposing them? Even for the integers, it is in general very difficult to find prime factors of an element x[8].

For the integers, Euclid in approximately the year 300BC discovered that there is a simple algorithm for finding the GCD that quantifies exactly how divisible two numbers are. The key observation is that if $a, b \in \mathbb{Z}$, where $a < b$ (notice the user of the total order, $<$, of $\mathbb{Z}$) then we may write

$$b = ka + r$$

where $r < a$. If $a|b$, then $r = 0$. If $a \nmid b$, then r is the "torsion" or "remainder" that is left-over. In terms of the GCD, if $a|b$, then $\gcd(a, b) = a$, and if $a \nmid b$, then $\gcd(a, b) < a$ and in the "worst case scenario" $\gcd(a, b) = 1$. The problem when generalizing this to general rings is that we require a notion of "size", namely it was crucial to use the fact that $a < b$ and it is imperative that $r < a$ to have a reasonable notion of such a factorization. In this section, we shall endavour to find a function, say $v : R \to \mathbb{N}_{\geq 0}$ that will allow us to assign a natural number to each ring element so that we may recapture this notion from the integers.

[8] RSA encryption is based off of the fact that prime factorization is a every difficult task if you don't know a priori what primes you are working with

Definition 10.2.1: [Ring] Valuation

Any function $v : R \to \mathbb{Z}$ with $v(0) = 0$ is called a *valuation* on the ring R. If $v(a) > 0$ for $a \neq 0$, then v is said to be a *positive valuation*.

Note: The notion of valuation shall be defined later in these notes on fields as *real [Archimedean] valuation* or *absolute value* (definition 19.2.1) and generalized to *G-valuation* (definition 20.6.1)[a]. The notion of a G-valuation is a generalization of the absolute value, however it is often presented with the axiom that $v(0) = \infty$. Valuations can just as well be defined with $v(0) = 0$, and hence the use of the term "valuation" here is justified.

[a] More specifically non-Archimedean valuations shall be generalized to G-valuations

Note: There are at least two common uses of the word "valuation" in contemporary abstract algebra: an *absolute-valuation* or more commonly an *absolute value* and a *G-valuation* (definition 20.6.1). These two concepts are non-compatible[a], notably G-valuations have axiom dictating that $v(0) = \infty$ while absolute values has an axiom dictating that $v(0) = 0$. The above use of the term valuation is closer to an *absolute value*.

[a] Though an absolute value may induce a G-valuation, this is shown when the concepts are introduced

If you know some analysis, you may wonder if the valuation is related to the one introduced on $\mathbb{R}$ or $\mathbb{C}$ (or normed vector spaces more generally). These concepts are indeed related, and we shall even see more explicit connection in chapter 19.

Returning to the issue at hand, we shall require a valuation N to be defined on an integral domain R that satisfies the following property:

Definition 10.2.2: Euclidean Valuation and Euclidean Domain

Let R be an integral domain and $N : R \to \mathbb{N}_{\geq 0}$ a valuation. Then v is a *euclidean valuation* on R if for any two elements a, b of R with $b \neq 0$ there exists elements q, r in R such that

$$a = qb + r \qquad \text{with } r = 0 \text{ or } N(r) < N(b)$$

The element q is called the *quotient* and the element r the *remainder* of the division if a euclidean valuation can be defined on a ring, then the ring is said to be a *Euclidean Domain* (or posses a *division algorithm*)

The main thing we can do in euclidean domains is run the division algorithm to keep "dividing" our

numbers

$$\begin{aligned} a &= q_0 b + r_0 \\ b &= q_1 r_0 + r_1 \\ r_0 &= 2_0 r_1 + r_2 \\ &\vdots \\ r_{n-1} &= q_{n+1} r_n \end{aligned}$$

Such a sequence will terminate since $N(b) > N(r_0) > \cdots > N(r_n)$, and it will eventually hit 0.

Example 10.2.1: Euclidean Domains

1. Every field with the euclidean valuation is a euclidean domain! In fact, any valuation will work!

   Proof. Let $x \in \mathbb{F}$ be an element of a field. Then since every element in a field has a unit, then
   $$x = xb + 0$$
   where $b = x^{-1}$.

   □

   Thus, we see that if a division rings was also commutative, then it'd clearly be a Euclidean domain.

2. In general, every unit will trivially satisfy this condition, since if $u \in R$ is a unit, then
   $$u = uu^{-1} + 0$$
   and so terminates instantly. So in a sense, Euclidean domains say that if we have elements that are *not* units, how many properties of divisible things can we put on them?

   Note importantly that euclidean domains must be domains (meaning no zero-divisors)! So oddities with multiplying not increasing the "size" of elements will not happen.

3. Ring of fractions in general are not euclidean domains, since they might have zero-divisors. So even though rings of fractions can be thought of as a way of enlarging a ring to make it's elements more divisible, it cannot do so if it has zero-divisors.

4. The integers $\mathbb{Z}$ are a euclidean domain with $N(a) = |a|$. The division algorithm is your usual euclidean algorithm. If you take $a, b \in \mathbb{Z}$, and $a < b$ then $b = aq + r$. We see that $N(r) = |r| < |b| = N(b)$. This will always be the case with $b \neq 0$! Thus, it is a Euclidean Domain.

5. If $\mathbb{F}$ is a field, then $\mathbb{F}[x]$ is also a Euclidean domain with $N(p(x)) = deg(p(x))$. The division algorithm is your usual polynomial division. The reason is simple for valuation: polynomial division *must* produce a remainder r with degree *less* than b, by definition! This makes $\mathbb{F}[x]$ into a Euclidean Domain. We'll prove this more rigorously in section 11.1.

6. Let
$$\mathbb{Z}[i] = \{a + bi | a, b \in \mathbb{Z}\}$$
And let
$$N(a + bi) = (a + bi)(a - bi) = a^2 + b^2$$
which is the usual valuation on $\mathbb{C}$. Intuitively, this is because we have subrings $\mathbb{Z}[i] \subseteq \mathbb{C}$. These are called the *Gaussian Integers*. We'll show that Gaussian integers ($\mathbb{Z}[i]$) is a Euclidean Domain.

   Consider $a, b \in \mathbb{Z}[i]$. We need to show that
$$a = kb + r \qquad N(r) < N(b)$$
for appropriate k, r. Since $\mathbb{Z}[i] \subseteq \mathbb{Q}(i)$, take
$$\frac{a}{b} = \frac{ac + bd}{c^2 + d^2} + \left(\frac{bc - ad}{c^2 + d^2}\right) = r + si$$
this gives us an idea of how close a is to b by having by looking at their fraction. Let p be the closest (or one of the 2 closest) integers to r and q be the closest (or one of the 2 closest) integers to s so that $|r - p|, |s - q| \leq \frac{1}{2}$. Then we'll show that
$$a = (p + qi)b + r$$
Whatis r? Notice that $t = (r - p) + (s - q)i$, so that $r = tb$, giving us $r = a - (p + qi)b$, showing that $r \in \mathbb{Z}[i]$. With this, we can calcualte the valuations quite easily! Notice that $N(t) = (r - p)^2 + (s - q)^2$ would at most
$$N(t) \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$
and so, by the multiplicity of this valuation (TBD move that part earlier so that I can use it)
$$N(tb) = N(t)N(b) \leq \frac{1}{2}N(b) \Rightarrow N(t) < N(b)$$
showing that $\mathbb{Z}[i]$ is indeed a Euclidean Domain.

7. Another class of Euclidean domains come from polynomial rings over fields. The proof is a little more elaborate and so we shall leave it till page 326.

Notice how central defining a valuation is to showing a ring is a euclidean domain. This algorithm *almost* produces a divisibility condition, making Euclidean domains "close" to fields. However, they are not fields, since $\mathbb{Z}$ is a euclidean domain. So instead of every element having an inverse, for every pair of elements (a, b), we can find a number(s) c such that we have a notion of $c|a$ and $c|b$. If every pair of numbers has this condition, we have a Euclidean domain!

(Add here a proof of how the gcd works for Euclidean Domains)

We have not seen many examples of Euclidean domains. The following is a class of rings that shall provide a testing ground for finding more Euclidean domains and will come up a lot in chapter 22:

Example 10.2.2: Rings of Integers of Quadratic Extenions

Let D be a squarefree integer (it's prime-factorization has exactly one factor for each prime that appears in it, so $10 = 2 \cdot 5$ is square free, but $18 = 2^2 \cdot 3$ is not). Then we can define $\mathbb{Z}[\sqrt{D}] = \left\{a + b\sqrt{D} | a, b \in \mathbb{Z}\right\}$, which is a ring, even a subring, of the quadratic field $\mathbb{Q}(\sqrt{D})$ (that we defined in the first examples). If we take $D \equiv 1 \bmod 4$, then consider the slighty larger subset

$$\mathbb{Z}\left[\frac{1+\sqrt{D}}{2}\right] = \left\{a + b\frac{1+\sqrt{D}}{2} | a, b \in \mathbb{Z}\right\}$$

is also a subring. With this, define

$$\mathcal{O} = \mathcal{O}_{\mathbb{Q}(\sqrt{(D)})} = \mathbb{Z}[\omega] = \{a + b\omega | a, b \in \mathbb{Z}\}$$

where

$$\omega = \begin{cases} \sqrt{D} & \text{If } D \equiv 2, 3 \bmod 4 \\ \frac{1+\sqrt{D}}{2} & \text{If } D \equiv 1 \bmod 4 \end{cases}$$

called the *ring of integers* in the quadratic field $\mathbb{Q}(\sqrt{D})$. The idea being that the elements of $\mathcal{O}$ of the field $\mathbb{Q}(\sqrt{D})$ have many properties analogous to those of the subring of integers $\mathbb{Z}$ in the field of rational numbers. Or, if we have a "field extension", which in a diagram we can think of as:

$$\begin{array}{ccc} \mathbb{F} & \supseteq & O_{\mathbb{F}} \\ \uparrow & & \uparrow \\ \mathbb{Q} & \supseteq & \mathbb{Z} \end{array}$$

where the arrows represent inclusions, then we want $\mathbb{F}$ to share properties that $\mathbb{Z}$ has Given an extension field $\mathbb{F}$ for $\mathbb{Q}$, we identity rings of integers $O_{\mathbb{F}}$ a subrging of F analogous to $\mathbb{Z} \subset \mathbb{Q}$

Note also that $\mathbb{Z}[\omega] = \mathcal{O}_{\sqrt{(D)}}$ is an integral domain, and that it's field of fraction is $\mathbb{Q}(\sqrt{D})$, giving us another way of thinking of ho $\mathbb{Z}[\omega]$ is sort of the "natural" integral domain to choose.

In the special case where $= -1$, we get $\mathbb{Z}[i]$, called the *Gausian integers*, used to study powers of numbers in the complex plane.

The reason for defining this is because we can define the concept of a *field norm*. Let $N : \mathbb{Q}(\sqrt{D}) \to Q$ by

$$N(a + b\sqrt{D}) = (a + b\sqrt{D})(a - b\sqrt{D}) = a^2 - Db^2 \in \mathbb{Q}$$

which is non-zero if $a + b\sqrt{D} \neq 0$. This gives an idea of "size" in the field $\mathbb{Q}(\sqrt{D})$. For instance, when $D = -1$, then the valuation of $a + bi$ is $a^2 + b^2$ (as we know it from Complex Analysis!), which is the square of the length of this complex number considered as a vector in the compelx plane.

We can see that N is multiplicative, that is, $N(\alpha\beta) = N(\alpha)N(\beta)$ for all $\alpha, \beta \in \mathbb{Q}(\sqrt{D})$ given by

$$N(a + b\omega) = (a + b\omega)(a - b\overline{\omega}) = \begin{cases} a^2 - b^2 D & \text{if } D \cong 2, 3 \bmod 4 \\ a^2 + ab + \frac{1-\sqrt{D}}{2} & \text{if } D \cong 1 \bmod 4 \end{cases}$$

where

$$\overline{\omega} = \begin{cases} -\sqrt{D} & \text{if } D \equiv 2, 3 \bmod 4 \\ \frac{1-\sqrt{D}}{2} & \text{if } D \equiv 1 \bmod 4 \end{cases}$$

It follows that $N(\alpha) \in \mathbb{Z}$, meaning we can relate properties of $\mathbb{Z}$ to the field norm!

For example, we can use this norm to characetrize the units of $\mathbb{Z}[\omega]$. If $N(\alpha) = \pm 1$, then $(a + b\omega)^{-1} = (a + b\overline{\omega})$, which is an element of $\mathbb{Z}[\omega]$, and so α is a unit. Suppose, conversely, that α is a unit in $\mathbb{Z}[\omega]$, so $\exists \beta$ such that $\alpha\beta = 1$. Then since N is multiplicative,

$$N(\alpha)N(\beta)N(\alpha\beta) = N(1) = \pm 1$$

and since $N(\alpha)$, $N(\beta)$ must be integers, they must be ± 1, and so

$$\text{the element } \alpha \text{ is a unit in } \mathcal{O} \text{ if and only if } N(\alpha) = \pm 1$$

Thus, the field norm also gives us an easy way to find the units of a given ring of integers.

We'll be using the field norm to how for certain values of D, $\mathbb{Z}[\sqrt{D}]$ will be a Euclidean Domain, and not for others (ex. we've already shown that $\mathbb{Z}[\sqrt{-1}] = \mathbb{Z}[i]$ is a Euclidean Domain).

Euclidean domain $\Rightarrow$ PID

As mentioned earlier, Euclidean Domains will be useful to factorizing elements in UFDs in which they are defined. Not all UFDs may have a valuation satisfying the euclidean algorithm, namely since being able to define such a va implies a good deal of structure on the ring. One important consequence of the existence of such a norm is that every Eucildean domain is a PID

Proposition 10.2.3: Euclidean Domain $\Rightarrow$ Pid

Given a Euclidean domain R, then every ideal of R is a principal ideal, meaning R is also a PID. In particular, if I is an ideal, then $I = (d)$, where d is any nonzero element of I of minimum norm

***Proof* :**
If I is the zero ideal, then we're done, so let $I \neq (0)$. This means that we can choose an element $d \in I$ such that $N(d)$ is a minimum element. This exists since $\{N(a) | a \in I\}$ has a minimum element by the well-ordering principle of $\mathbb{Z}$. We know show that $(d) = I$

($\subseteq$) it's clear that $(d) \subseteq I$ by the definition of ideals

($\supset$) Let $a \in I$ be any element of our ideal. Then since R is a euclidean domain, we can write a as :

$$a = qd + r$$

where $r = 0$ or $N(r) < N(d)$. Since $r = a - qd$, and $a, qd \in I$, then $r \in I$. But then, it's impossible that $N(r) < N(d)$ since $N(d)$ is already the minimal norm!! Thus $r = 0$, implying $a = qd$. But then $a = qd \in (d)$, as we sought to show

As usual, the contra-positive should be kept in mind: R is *not* an Euclidean Domain if it is *not* a PID.

Example 10.2.3: Implications

1. Since $\mathbb{Z}$ is a Euclidean domain, it is A Principal Ideal Domain. Note that we didn't show that PID $\Rightarrow$ Euclidean domain, we'll present a counter-example soon, so this was only a hint in the right direction
2. $\mathbb{Z}[x]$ is *not* a PID since $(2, x)$ is a non-principal ideal, and so not a euclidean domain.
3. $\mathbb{Q}[x]$ is a PID, and can be shown to be the euclidean domain with $N(p(x)) = \deg(p(x))$
4. $\mathbb{Z}[\sqrt{5}]$ is not a PID, and so not a Euclidean Domain

You might wonder if *all* PID's are euclidean domains. That need not be the case. To prove this, we'll first prove a another way of finding Euclidean domains:

Definition 10.2.4: Side Divisor

Let $\overline{R} = R^\times \cup \{0\}$ be the collection of units of R together with 0. An element $u \in R - \overline{R}$ is called a *universal side divisor* if for evry $x \in R$ there is some $z \in \overline{R}$ such that u divides $x - z$ in R, i.e., there is a type of "division algorithm" for u: every x may be written as $x = qu + z$ where z is either zero or a unit.

The existence of universal side divisors is weakening of the Euclidean condition, since we're saying that for any element $x \in R$, we can "divide" it by u and have a remainder of a unit z. These conditions are not as general as the conditions for the euclidean algorihtm (where z doesn't need to be constrained to be $N(z) = 1$). With this, we prove the following equivalent condition:

Proposition 10.2.5: Euclidean Domain Implies Universal Side Divisors

Let R be an integral domain that is not a field. If R is a euclidean domain, then there are universal side divisors in R

Proof **:**
Suppose R is Euclidean with respect to some norm N and let u be an element $R - \overline{R}$ (which is non-empty since F is not a field), of minimal norm. For any $x \in R$ write $x = qu + r$ where r is etiher 0 or $N(r) < N(u)$. In either case the minimality of u implies $r \in \overline{R}$. Hence u is a univerasl side divisor in R

With this, we can find

Example 10.2.4: : PID but not euclidean domain
We'll take the following ring of integers.

$$\mathcal{O}_{Q(\sqrt{-19}} = \mathbb{Z}\left[\frac{1+\sqrt{-19}}{2}\right]$$

with the field norm:

$$N(a + b\sqrt{-19}) = (a + b\frac{1+\sqrt{-19}}{2})(a - b\frac{1+\sqrt{-19}}{2}) = a^2 + ab + 5b^2$$

This ring is a PID; a fact we'll show in the next chapter. However, it is not a Euclidean domain, for there does not exist a universal side divisor.

We can find that $\overline{R} = \{0, \pm 1\}$ are the unit of R union 0. Now suppose $u \in R$ is a universal side divisor. Note that if $b \neq 0$, then

$$a^2 + ab + 5b^2 = (a + \frac{b}{2})^2 + \frac{19}{4b^2} \geq 5$$

and so the smallest values of N on R are 1 (for units ± 1) and 4 (for $a = \pm 2$). Let's say $x = 2$. Then, we have

$$2 = qu + z \Rightarrow 2 = 2u + 0 \qquad \text{or} \qquad 2 = u \pm 1$$

or, written in division form, we see that $u|2 - 0$ or $u|2 \pm 1$.

Let's say $z = 0$, so $2 = qu$. Then $N(2) = 4 = N(q)N(u)$. TBD (it's a tedious example)

It is tempting to think that PIDs are close to being Euclidean domains as they share a notion of gcd and lcm, unique factorization. The following measure the degree of seperation of a Principal Ideal Domain being a Euclidean Domain:

Definition 10.2.6: Dedikend-Hasse Valuation

Define N to be the *Dedekind-Hasse valuation* if N is a positive norm and for every nonzero $a, b \in R$ either a is an element of the ideal (b) or there is a nonzero elemtn of the ideal (a, b) of norm strictly smaller than the norm of b (i.e., either b divides a in R or their exists $s, t \in R$ such that $0 < N(sa - tb) < N(b)$

Note that R is Euclidean with respect to a positive norm N if it is always possible to satisfy the Dedekind-Hasse norm with $s = 1$, so this is indeed a weakening of our condition for Euclidean Domains.

Proposition 10.2.7: Principal Ideal Domain If And Only If Dedekind-Hasse Norm

The integral domain R is a PId if and only if R has a Dedekind-Hasse nrom

Proof **:**
p.281

Corollary 10.2.8: Defning Dedekind-Hasse Norm

let R be a Principal Ideal Domain. Then there exists a multiplicative Dedekin-Hasse norm.

Proof **:**
p.289

Example 10.2.5
Showing that $\mathbb{Z}[\sqrt{-19}]$ is a Principal Ideal Domain but not a Euclidean Domain.

10.3 Application: Sums of Primes and Gaussian Integers

Let us use the theory we've developed to get our first result about prime numbers, namely let us classify all prime numbers that can be written as the sum of two squares $p = x^2 + y^2$. [9]

Recall that the Gaussian integers are of the form:

$$\mathbb{Z}[i] \subseteq \mathbb{C}$$

which can be thought of as the smallest ring containing $\mathbb{Z}$ and i. The Gaussian integers have a pretty geometric picture as a lattice:

The Gaussian Integers $\mathbb{Z}[i]$

This gives us a geometric picture of this ring! For example, we may visualize principal ideals in this grid, for example the ideal $(2 + i)$ is the point:

[9] Naturally, a prime by definition cannot be the square of a natural number, however it can be $\bmod q$ under the right condition; this is called *quadratic reciprocity* and is covered in section 22.4.1

The Gaussian Integers $\mathbb{Z}[i]$ with $(2+i)$ highlighted

Recall how we introduced the field-norm on on quadratic integers, where $\mathbb{Z}[i] = \mathbb{Z}[\sqrt{-1}]$ is an example of quadratic integers. With the above visuals, it is more clear to see that the field norm has similar geometric information to the norm presented in real or complex analysis, namely:

$$N(a+bi) = (a+bi)(a-bi) = a^2 + b^2 = \|a+bi\|^2$$

Using the fact that the codomain of the field norm is $\mathbb{N}_{\geq 0}$, we can see that the norm is in fact a euclidean valuation

Lemma 10.3.1: Gaussian Integers Are Euclidean Domain

Let $R = \mathbb{Z}[i]$ and N be the field norm. Then N is a Euclidean valuation. Furthermore, N is multiplicative so that for every $z, w \in \mathbb{Z}[i]$,

$$N(zw) = N(z)N(w)$$

Proof **:**
Multiplicity is immediate as:

$$N(zw) = (zw)(\overline{zw}) = (z\overline{z})(w\overline{w}) = N(z)N(w)$$

The proof that it is a euclidean valuation is straight-forward if not a little verbose, and so is left as an exercise.

What could be more insightful is to understand using the lattice picture the significant of the field norm on $\mathbb{Z}[i]$ (this was demonstrated in Aluffi's textbook see cite:HERE). Take $z, w \in \mathbb{Z}[i]$ with $w \neq 0$ nd consider the lattice given by (w). Then either z lies on a point in this lattice, in which case $z = qw$ for appropriate q, or z is inside a "box" of the lattice. Pick some points qw that is a vertex of the box containing z, and set $r = z - qw$ so that $z = qw + r$. By visual inspection, we see that the length of r is ells than that of r, namely:

$$N(r) < N(w)$$

This trick shall not work for all quadratic fields, as we shall demonstrates after finishing elaborating on Gaussian integers.

It is a good exercise to show that if u is a unit then $N(u) = \pm 1$, from which we can conclude that the only possible units are $1, -1, i, -i$. We can use the norm to also detect irreducible elements. To see this, let' say $p \in \mathbb{Z}$ is a prime. Then since $\mathbb{Z}$ is an integral domain, p is irreducible, and so $p = ab$ if and only if a or b is a unit. Let's now take $\pi \in \mathcal{O}$ such that $p = N(\pi)$. If $\pi = \alpha \cdot \beta$, then:

$$p = N(\pi) = N(\alpha)N(\beta)$$

meaning either $N(\alpha)$ or $N(\beta)$ is a unit, meaning it's value is ± 1. Thus, we have that

$$\text{If } N(\alpha) \text{ is } \pm p \text{ for a prime } p \in \mathbb{Z}, \text{ then } \alpha \text{ is irreducible (even prime) in } \mathbb{Z}[i]$$

The converse is not necessarily true: we can have a prime in $\mathbb{Z}[i]$ but it's norm is not p. Fortunately, this is a feature of primes in $\mathbb{Z}[i]$ and demonstrates a deeper insight on the structure of primes in $\mathbb{Z}[i]$:

Lemma 10.3.2: Gaussian Integers Norm of Primes

Let $\pi \subseteq \mathbb{Z}[i]$ be prime where. Then

$$N(\pi) \in \{p, p^2\}$$

for appropriate prime $p \in \mathbb{Z}$.

Proof **:**
Let $\pi \in \mathbb{Z}[i]$ be a prime element, that is, (π) is a prime ideal in $\mathbb{Z}[i]$. As π is not a unit, $N(\pi) \neq 1$. Then $(\pi) \cap \mathbb{Z}$ is also a prime ideal: if for $a, b \in \mathbb{Z}$, we have $ab \in (\pi)$ and $ab \in \mathbb{Z}$[a] then either a or b is an element of (π), so a or b is in $(\pi) \cap \mathbb{Z}$. Thus, there exists a prime p such that $(p) = (\pi) \cap \mathbb{Z}$.

Now, take $(p) \subseteq \mathbb{Z}[i]$. Then $(p) \subseteq (\pi)$ so that $xp = \pi$ for some $x \in \mathbb{Z}[i]$ (which may be a unit, as p may be an associate of π). Then $N(p) = p^2$, and since the norm is multiplicative

$$N(\pi) \big| N(p) = p\overline{p} = pp = p^2$$

from which it can be immediately deduced that $N(\pi) \in \{p, p^2\}$, completing the proof.

[a] if $a, b \in \mathbb{Z}$, automatically $ab \in \mathbb{Z}$

There are prime elements in $\mathbb{Z}[i]$ that have norm p and norm p^2. Note that irreducible does not imply prime, as it was shown with $\mathbb{Z}[\sqrt{-5}]$, however if the same trick is done as in the above proof

it seems that irreducible elements of a quadratic ring are associated to prime elements in $\mathbb{Z}$ (see exercise ref:HERE). Under the right framework, this connection can be made more precise, which will be part of the aim of chapter 22.

Let us find some examples of primes with norms p and p^2. Let's take $\alpha = a + bi \in \mathbb{Z}[i]$. Then $N(\alpha) = \alpha\overline{\alpha} = a^2 + b^2 \in \{p, p^2\}$. If p is reducible in $\mathbb{Z}[i]$,then

$$p = \pi\pi' \qquad \Rightarrow \qquad N(p) = N(\pi)N(\pi') = p^2$$

Thus, p will factor in $\mathbb{Z}[i]$ into two irreducibles if and only if $p = a^2 + b^2$, that is, p is the sum of two integers squares. If p is not the sum of two integer squares, it will remain irreducible. Take or example $3 \in \mathbb{Z}[i]$. Then $N(3) = 9$, and if 3 factored we would need an element $a + bi$ such that $N(a + bi) = a^2 + b^2 = 3$ for natural numbers $a, b \in \mathbb{N}$. However, it can be quickly verified that no such numbers exist, and so 3 is prime in $\mathbb{Z}[i]$. For another example, take 2. Then as $2 = 1^2 + 1^2$, 2 factors in $\mathbb{Z}[i]$! In particular, $2 = (1 + i)(1 - i) = -i(1 + i)^2$ (note that $N(-i) = 1$ and so these two are associates). This will be the only case were conjugate irreducibles $a + bi$ and $a - bi$ can be associate. Any larger element will have $N(\pi') > 1$. For example, $N(5) = 25$, so we need to see if there is an element $a + bi$ such that $N(a + bi) = 5$, and indeed:

$$N(2 + 1 \cdot i) = 2^2 + (1)^2 = 5$$

and indeed $5 = (2 + i)(2 - i)$. This factorization picture can be seen more clearly when viewing this elements via their ideals in the lattice. In t he case of 3, there is no possible further refinement that can be done to the lattice, while for 5 we may further refine the lattice (without it being the entire lattice):

Let us now find a condition that shall let us immediately tell if a prime can split. First, any number in $\mathbb{Z}$ is congruent to $0, 1, 2, 3 \bmod (4)$. Therefore, the square of any number is congruent to either 0 or 1. Thus, an odd prime (so $p > 2$) in $\mathbb{Z}$ that is the sum of two squares must be congruent to 1 mod 4. Therefore, if $p \equiv 3 \bmod 4$ in $\mathbb{Z}$, then p is *not* the sum of squares, and remains irreducible (and thus prime) in $\mathbb{Z}[i]$.

Theorem 10.3.3: Fermat's Theorem on Sums Of Squares

The prime p is the sum of two integer squares, $p = a^2 + b^2$, $a, b \in \mathbb{Z}$ if and only if $p = 2$ or $p \equiv 1 \bmod 4$. Except for interchanging a and b or changing the sign of a and b, the representation of p as a sum of two squares is unique.
Furthermore, the prime elements in the Gaussian integers $\mathbb{Z}[i]$ are as follows:

1. $1 + i$ (which has norm 2)
2. the primes $p \in \mathbb{Z}$ with $p \equiv 3 \bmod 4$ (which have norm p^2)
3. $(a + bi)$, $(a - bi)$, the distinct irreducible factors of $p = a^2 + b^2 = (a + bi)(a - bi)$ for the primes $p \in \mathbb{Z}$ with $p \equiv 1 \bmod 4$ (both of which have norm p).

In other words, each prime $p \equiv 1 \bmod 4$ can be written as the sum of two squares:

$$p = a^2 + b^2$$

Proof **:**
The case for $1 + i$ has been done, so consider an odd prime $p \in \mathbb{Z}$. Then verify the following chain

of implications:

$$\begin{aligned} p \text{ splits in } \mathbb{Z}[i] &\Leftrightarrow \frac{\mathbb{Z}[i]}{p\mathbb{Z}[i]} \text{ is not an integral domain} \\ &\Leftrightarrow \frac{\mathbb{Z}/p\mathbb{Z}[x]}{(x^2+1)} \text{ is not an integral domain} \\ &\Leftrightarrow x^2+1 \text{ is not irreducible in } \mathbb{Z}/p\mathbb{Z}[x] \\ &\Leftrightarrow x^2+1 \text{ has a root in } \mathbb{Z}/p\mathbb{Z}[x] \\ &\Leftrightarrow \exists\, n \in \mathbb{Z} \text{ such that } n^2 \equiv -1 \bmod p \end{aligned}$$

Thus, the splitting of p has been reduced to solving the congruence equation $n^2 \equiv -1 \bmod p$. To verify this condition, recall that $(\mathbb{Z}/p\mathbb{Z})^*$ is as p is prime, and it has order $p-1$. As p is odd, $p-1$ is even, so there is a generator $g \in (\mathbb{Z}/p\mathbb{Z})*$ such that $|g| = 2\ell$ for some appropriate natural number ℓ. Then write $[n] = g^m$ for appropriate m. Now, certainly $[-1] \in (\mathbb{Z}/p\mathbb{Z})^*$ creates an (unique) order 2 subgroup, and g^ℓ also creates an order 2 subgroup, and so:

$$g^\ell = [-1]$$

Then $n^2 \equiv -1 \bmod p$ if and only if $g^{2m} = g^\ell$, that is if and only if

$$2m \equiv \ell \bmod 2\ell$$

Now this linear congruence relation is readily solvable: it is true if and only if ℓ is even, and so when $(p-1) = 2\ell$ is a multiple of 4, or shuffled around:

$$p \equiv 1 \bmod 4$$

giving us an exact condition for when p can be factored, equivalently when p can be written as the sum of two squares, completing the proof.

Lemma 10.3.4: Writing p as n^2+1

The prime number $p \in \mathbb{Z}$ divides an integer of the form n^2+1 if and only if p is either 2 or is an odd prime congruent to 1 mod 4.

***Proof* :**
By Fermat sum of square's theorem, we know that there exists an n such that $n^2+1 \equiv 0 \bmod p$, which his equivalent to saying p divides n^2+1.

Though this condition is equivalent to Fermat' sum of Square's Theorem, it is not as informative as $5 \big| 3^1+1 = 10$, however $3+i$ and $3-1i$ are *not* prime (for example, $3+i = (1+i)(2-i)$). There is however something very interesting that was demonstrated in this entire section: we have answered a purely number theoretic question about which primes can be expressible as sums of squares (a fact that should at first not be so obvious) by taking advantage of a lot of simple theoretical properties built-up about factorization and of cyclic groups!

Exercise 10.3.1

1. I think I can do an exercise finding solutions to $y^2 = y^3 - 1$ eh? I must've made a typo here, I'll revisit

11

Exploring Polynomial Rings

When first studying algebra, we often looked to find solutions to linear and quadratic equations:

$$ax + b = 0 \qquad ax^2 + bx + c = 0$$

This naturally generalizes to polynomial of higher degree. In pre-abstract algebra, solutions to such equations where usually limited to $\mathbb{R}$, probably $\mathbb{C}$. We have now developed the abstraction to consider the value (or root) of such a polynomial for an arbitrary ring R, that is for an arbitrary polynomial in $R[x]$, does there exist an element $r \in R$ such that

$$p(r) = 0$$

As is the usual strategy in algebra, we generalize a problem into an algebraic object, in this case the polynomial ring $R[x]$, and look for interesting properties of the objects and tools we can use to manipulate the object. Such polynomials rings are paramount to study. To give an idea of their importance:

1. they arbitrarily approximate continuous functions on compact neighborhoods
2. all complex differentiable function can be locally be represented as infinite polynomials (i.e. they are analytic)
3. All relations in rings are represented by polynomial (similarly to how all relations in a group is represented by a monomial)

The last point shows that polynomial rings are used to "locally" represent rings, namely through the universal property of [commutative] universal rings (see theorem 9.2.3). It may thus seem natural to consider studying polynomials rings themselves to see what properties we may deduce from them.

One necessary limitation of the exploration of polynomial rings in chapter is that we shall not look at finding solutions to polynomials, we shall rather look at the structure of polynomial rings. The

reason for this is that it is in fact a much more difficult task to find solutions requiring more build-up that starts venturing into the domain of algebraic geometry. The beginnings of this effort shall be presented in part V, chapter 21.

It must be noted that we shall further develop on poomial rings in chapter 20. This exploration is left to the further chapter since it is not necessary to cover some results until much later.

Note that this chapter shall use some exercises from exercise 9.3.1, most importantly:

Proposition 11.0.1: Ideals of Polynomial Ring wrt. R

Let $I \subseteq R$ be an ideal. Then:

$$(R[x])/(IR[x]) \cong (R/I)[x]$$

Proof **:**
exercise 9.3.1[3]

11.1 Polynomial Rings over Fields

We'll now explore polynomial rings with coefficients from a field $\mathbb{F}$. The first question we may ask if what type of ring is $\mathbb{F}[x]$? It's not a field, since we already saw that x is not invertible (meaning it's also not a division ring). The next strongest ring we know is Euclidean domains. If $\mathbb{F}[x]$ is a euclidean domain, there must be a norm we define, which we'll use $N(p(x)) = \deg(p(x))$ (the usual polynomial norm). Then $\mathbb{F}[x]$ is indeed a Euclidean Domain:

Theorem 11.1.1: Field Then $\mathbb{F}[x]$ Is Euclidean Domain

Let $\mathbb{F}$ be a field. The $\mathbb{F}[x]$ is a Euclidean Domain. In particular, if $a(x)$ and $b(x)$ are two polynomials in $\mathbb{F}[x]$ with $b(x)$ nonzero, then there are *unique* $q(x)$ and $r(x)$ such that

$$a(x) = q(x)b(x) + r(x_0) \qquad \text{with} \qquad r(x) = 0 \text{ or } \deg(r(x)) < \deg(b(x))$$

Proof **:**
First, the edge cases:

If $a(x)$ is the zero polynomial, then let $q(x) = r(x) = 0$. So let $a(x)$ be a nonzero polynomial. Let $b(x)$ be an arbitrary other polynomial. If $\deg(a(x)) < \deg(b(x))$, then let $q(x) = 0$ and $r(x) = a(x)$.

Now, take the nonzero polynomial $a(x)$ such that $\deg(a(x)) > \deg(b(x))$. We'll do a proof by induction on the degree of $a(x)$ ($n = \deg(a(x))$).

For visualization, we can denote $a(x)$ and $b(x)$ as:

$$a(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0 \qquad b(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_0$$

We want to apply the induction hypothesis. Let $a'(x) = a(x) - \frac{a_n}{b_m}x^{n-m}b(x)$, which purposely defined to eliminate the monomial of degree n from $a(x)$. Then by the induciton hypothesis:

$$a'(x) = q'(x)b(x) + r(x) \qquad r(x) = 0 \text{ or } \deg(r(x)) < \deg(p(x))$$

Then, we "reverse" the step of taking $a'(x)$ by doing $q(x) = q'(x) + \frac{a_n}{b_n}x^{n-m}$, which will give us

$$a(x) = q(x)b(x) + r(x) \qquad r(x) = 0 \text{ or } \deg(r(x)) < \deg(p(x))$$

showing we can factor two polynomials.

It remains to show that this factorization is unique. Suppose $q(x), r(x)$, and $q_0(x)$, $r_0(x)$ are both polynomials that satisfy the euclidean condition. Then

$$a(x) - q(x)b(x) \qquad a(x) - q_0(x)b(x)$$

are both of degree less than m. The difference between these two polynomials then:

$$b(x)(q(x) - q_0(x))$$

is also less than m. But the degree of product of polynomials over an integral domain is the sum of their degree ($m > \deg(b(x)(q(x) - q_0(x))) = \deg(b(x) + \deg(q(x) - q_0(x))$. So, $q(x) - q_0(x) = 0$, implying $q(x) = q_0(x)$. This then implies $r(x) = r_0(x)$, completing the proof of uniqueness

Corollary 11.1.2: Field Then $\mathbb{F}[x]$ PID And UFD

Let $\mathbb{F}$ be a field. The $\mathbb{F}[x]$ is a Principal Ideal domain, and a Euclidean Domain

***Proof* :**
By the previous theorem (theorem 11.1.1), we know $\mathbb{F}[x]$ is a euclidean domain. By 10.2.3, it is a Principal Ideal Domain, and since PID's are UFD's, it is also a Unique Factorization Domain

Example 11.1.1

1. $\mathbb{Q}[x]$ is a Principal Ideal Domain since $\mathbb{Q}$ is a field.
2. $\mathbb{Q}[x, y]$ is not a Principal Ideal Domain, since $\mathbb{Q}[x]$ is not a field (using the contra-positive).

One final fact we'll introduce as a result of what we talked about:

Corollary 11.1.3: Kernel Of Evaluation Map

Let $\mathbb{F}$ be a field and let α be an element of $\mathbb{F}$. Then the kernel of ev_α is the ideal $(x - \alpha)$

***Proof* :**
Let $I = \ker(\text{ev}_\alpha)$. By definition of ev_α, $x - \alpha \in I$. Now, since $\mathbb{F}$ is a field, $\mathbb{F}[x]$ is a Euclidean Domain, and thus a Principal Ideal Domain, so I is a principal ideal. Let's say $I = (p)$ for some $p \in \mathbb{F}[x]$. Then by definition, $x - \alpha = up$, meaning $p|x - \alpha$. Note that p cannot have degree higher than 2, since then $x - \alpha \notin I$. So, if p has degree one, then the two elements are associates, meaning $I = (p) = (up) = (x - \alpha)$. If p has degree 0, then it must be a constant. However, p has a root at α, so if p were a constant, it would have to be 0. But then $x - \alpha \notin (0)$ – a contradiction. Therefore $(x - \alpha) = \ker(\text{ev}_\alpha)$.

Here is also a fun result I want to put down:

Example 11.1.2: generalization of Famous infinite primes proof
Let F be a finite field. Then there are infinitely many prime elements in $F[x]$

Proof. Let's assume there is a finite amount. Then $p_1, ..., p_n$ is our list of irreducible polynomials. Note that $\deg(p_i) \geq 1$ for each polynomial. Then take $P = p_1 p_1 \cdots p_n + 1$. Then $\deg(P) \geq 1$, and so P is not a unit or 0. Therefore, P must be divisible by some p_i, meaning $p_i | P$, which implies $qp_i = P$, or equivalently $P \in (p_i)$. Next, by closure of ideals, $-p_1 p_2 \cdots p_n \in (p_i)$. Then, by closure under addition:

$$P - p_1 p_2 \cdots p_n = p_1 p_2 \cdots p_n + 1 - p_1 p_2 \cdots p_n = 1 \in (p_i)$$

meaning $up_i = 1$ for appropriate $u \in F[x]$, implying p_i is a unit! However, prime elements are *not* units. Since p_i was arbitrary, *no* p_i can divide P. But then, P will have non of the factor's in it's factorization, a contradiction to $\deg > 0$ and the p_i's being the only irreducibles! Therefore, P must be a new prime number!

Therefore, there is an infinite number of prime numbers. □

We finish off with an application of the Chinese remainder theorem with the new fact that $k[x]$ is a Euclidean domain, and hence a UFD:

Proposition 11.1.4: Chinese Remainder On Polynomials

Let $g(x)$ be a nonconstant monic element of $k[x]$ and let

$$g(x) = f_1(x)^{n_1} f_2(x)^{n_2} \cdots f_k(x)^{n_k}$$

be its factorization into irreducibles, where the $f_i(x)$ are distinct. Then we have the following isomorphism of rings:

$$k[x]/(g(x)) \cong k[x]/(f_1(x)^{n_1}) \times k[x]/(f_2(x)^{n_2}) \times \cdots \times k[x]/(f_k(x)^{n_k})$$

***Proof* :**
This follows from the Chinese Remainder Theorem (theorem 9.6.4), since $(f_i(x)^{n_i})$ and $(f_i(x)^{n_i})$ are comaximal if $f_i(x)$ and $f_j(x)$ are distinct (they are relatively prime in the Euclidean Domain $k[x]$, hence $gcd(f_i^{n_i}, f_j^{n_j}) = 1$, and so $p(x)f_i(x)^{n_i} + q(x)f_j(x)^{n_j} = 1$, so $1 \in (f_i^{n_i}) + (f_j^{n_j})$ implying $(f_i^{n_i}) + (f_j^{n_j}) = k[x]$ by propositin 9.4.6).

Exercise 11.1.1

1. Let $R[x]$ be a Principal Ideal Domain. Show that $R[x]$ is a Euclidean Domain.

11.2 Polynomial Rings over UFD's

We've seen R is an integral domain then $R[x]$ is an integral domain. We've also seen that if R is an integral domain, than we can embed R into a field of fraction F (see section 9.5). This means that $R[x] \subseteq F[x]$. Furthermore, since F is a field, then $F[x]$ is a euclidean domain. With this, we can do lots of our computation in $F[x]$ instead of $R[x]$, at the expense of allowing fractional coefficients. The question then arises if we can use information from $F[x]$ to gain information about $R[x]$!

For example, take a polynomial $p(x) \in R[x]$. Since $p(x) \in k[x]$, and $k[x]$ is a Unique Factorization Domain, then $p(x)$ factors in $k[x]$ into irreducibles. We can then ask if $p(x)$ will factor in $R[x]$. The answer is no in general, since $R[x]$ is not always a Unique Factorization Domain, since that would imply R is a Unique Factorization Domain since the coefficients of $R[x]$ must also factor uniquely. However, if R were a unique factorization domain, then this will actually imply that $R[x]$ is Unique Factorization Domain! Hence, factoring in $k[x]$ gives information on how we factor in $R[x]$. The way we'll prove this is first by uniquely factoring in $k[x]$, then "clear denominators" to obtain unique factorization in $R[x]$!

The first thing to do is make precise the notion of comparing factorization of a polynomial in $k[x]$ and a factorization in $R[x]$

Proposition 11.2.1: Gauss's Lemma

Let R be a Unique Factorization Domain with field of fraction F and let $p(x) \in R[x]$. If $p(x)$ is reducible in $k[x]$ then $p(x)$ is reducible in $R[x]$. More precisely, if $p(x) = A(x)B(x)$ for some non constant polynomials $A(x), B(x) \in k[x]$, then there are nonzero elements $r, s \in F$ such that $rA(x) = a(x)$ and $sB(x) = b(x)$, such that $a(x), b(x) \in R[x]$ and $p(x) = a(x)b(x)$ is a factorization in $R[x]$

A concrete case that can be kept in mind is with the polynomials $p(x), q(x) \in \mathbb{Q}[x]$, then $rp(x), sq(x) \in \mathbb{Z}[x]$ for appropriates $r, s \in \mathbb{Q}$. Note that it is possible that $q(x) \in \mathbb{Q}$, i.e. $q(x)$ is a constant polynomial[1].

Proof **:**
Let's say $p(x) \in k[x]$ and is reducible $p(x) = A(x)B(x)$. Then the coefficients of $p(x)$ are quotients from R. Since R is a Unique Factorization Domain, let's take d a number common betewen all denominators for all the coefficients of $A(x)B(x)$, so that we can obtain

$$dp(x) = a'(x)b'(x)$$

where $a'(x), b'(x)$ are now in $R[x]$ since we multiplied away the denominators with d! If d is a unit, then

$$a(x) = d^{-1}a'(x) \qquad b(x) = b'(x)$$

which completes the proof, so let's assume d is not a unit. Since $d \in R$, and R is a Unique Factorization Domain, then there is a unique facotirzation into irreducibles:

$$d = p_1 \cdots p_n$$

[1] In textbooks where the contra-positive of this statement is used, it is often specified that $p(x)$ is *primitive*, meaning for all principal prime $\mathfrak{p} \subseteq R$, $p(x) \notin \mathfrak{p}R[x]$; in the UFD case, this amounts to the gcd of the coefficients being 1

Since p_1 is an irreducible element, and R is a Unique Factorization Domain, then $(p_1) = p_1R[x]$ (since $R[x]$ is commutative $(a) = aR$) is prime (exercise), and so $R[x]/p_1R[x]$ is an integral domain (proposition 9.4.13) and $R[x]/p_1R[x] \cong (R/p_1R)[x]$ (exercise).

Now, reducing $dp(x) = a'(x)b'(x)$ modulo p_1, we get

$$\overline{0} = \overline{a'(x)} \cdot \overline{b'(x)}$$

Since $(R/p_1R)[x]$ is an integral domain, then either $\overline{a'(x)}$ or $\overline{b'(x)}$ is zero, say $\overline{a'(x)} = 0$. This would imply $p_1|a'(x)$, and so every coefficient of $a'(x)$ is divisible by p_1, meaning $\frac{1}{p_1}a'(x)$ also has coefficients from R! So, we can divide both sides by p_1 (that is d by p_1 and either $a'(x)$ or $b'(x)$ by p_1). But then, the factor d on the left hand side has one fewer irreducible element! We can continue this procedure till we cancel out all of the factors, leaving us with $p(x) = a(x)b(x)$, with $a(x), b(x) \in R[x]$, $a(x)$ and $b(x)$ being F-mutiples of $A(x)$ and $B(x)$!

Note that $A(x)$ and $B(x)$ aren't necessarily R-multiples, nor are they necessarily unique. For example, $x^2 \in Q[x]$ is reducible, and so reducible in $\mathbb{Z}[x]$. One factorization would be $A(x) = 2x$ and $B(x) = \frac{1}{2}x$. However, $\frac{1}{2} \notin \mathbb{Z}$: no integer multiple give's us a factorization of $A(x)$ and $B(x)$. What the theroem will tell us is that $\frac{1}{2}A(x) \in \mathbb{Z}[x]$, and $2B(x) \in \mathbb{Z}[x]$.

Since k is a field, the elements of R becomes units in $k[x]$ (the units in $k[x]$ being the non-zero constants). For example, $7x$ factors in $\mathbb{Z}[x]$ and produces two irreducible 7 and x, so $7x$ is *not* irreducible in $\mathbb{Z}[x]$. However, in $\mathbb{Q}[x]$, $7x$ is a unit 7 times an irreducible, so $7x$ *is* irreducible in $\mathbb{Q}[x]$. However, this is the *only* difference between the irreducible elements in $R[x]$ and $k[x]$

Corollary 11.2.2: Irreducibility In $R[x]$ And $k[x]$

Let R be a UFD, let $k = K(R)$ be its field of fractions, and let $p(x) \in R[x]$. Suppose the greatest common divisor of the coefficients of $p(x)$ is 1. Then $p(x)$ is irreducible in $R[x]$ if and only if it is irreducible in $K[x]$. In particular, if $p(x)$ is a monic polynomial that is irreducible in $R[x]$, then $p(x)$ is irreducible in $K[x]$.

Proof **:**
We'll proceed by the contrapositive: $p(x)$ is reducible in $R[x]$ if and only if $p(x)$ is reducible $K[x]$. The ($\Leftarrow$) direction is Gauss's Lemma, so let's go ($\Rightarrow$). Let's say $p(x)$ is reducible in $R[x]$ and the greatest comon divisor of the coefficients is 1. That means that when we reduce $p(x) = a(x)b(x)$, neither $a(x)$ or $b(x)$ can be constants! Note that polynomials which aren't constants in $K[x]$ are not units, meaning $p(x) = a(x)b(x)$ is also a factorization in $K[x]$, completing the proof!

This theorem is already incredibly powerful, it means that the task of finding the irreducible elements of $R[x]$ reduces to finding the irreducible elements of $k[x]$ (which we know is a euclidean domain) up to a multiple of R[2]. We shall later find more techniques for finding solutions to polynomials over fields, which by our above results immediately give us a way to translate these results for integral domains whose field of fraction is the given polynomial ring over a field.

[2]For those keen on doing algebraic geometry, this process of reducing the problem of factoring in $R[x]$ to factoring in $k[x]$ can be thought of as a form of using the stalk at 0 to find information about the elements of $R[x]$, that is it is taking advantage of local information for global results TBD

Yet another result is the following which tells us that extending a ring to form a polynomial ring preserves UFD!

Theorem 11.2.3: R UFD If And Only If $R[x]$ UFD

R is a Unique Factorization Domain if and only if $R[x]$ is a Unique Factorization Domain

***Proof* :**
If $R[x]$ is a Unique Factorization Domain, then the constants of $R[x]$ must uniquely factor, meaning R is a Unique Factorization Domain.

So let's suppose R is a Unique Factorization Domain, and take $p(x) \in R[x]$. If $p(x)$ is a constant, then it's uniquely factorizable, so let's say $\deg(p(x)) > 0$. Let d be the g.c.d. of the coefficients of $p(x)$, so that $p(x) = dp'(x)$, meaning the g.c.d. of the coefficients of $p'(x)$ is 1. Note that this factorization is unique up to a unit multiple of d (i.e. up to a unit in R by proposition 9.2.2), and since d can be factored uniquely in R (and these are also irreducibles in $R[x]$ too), it suffices now to prove that $p'(x)$ can be uniquely factored into irreducibles in $R[x]$.

Let k be the field of fractions of R. Then since $k[x]$ is a Unique Factorization Domain (proposition 11.1.1), $p(x)$ can be factored uniquely into irreducibles in $k[x]$. By Gauss's Lemma, such a factorization implies a factorization in $R[x]$ whose factors are F-multiples of the factors in $k[x]$. Since the g.c.d. of $p'(x)$ is 1, the g.c.d. of the coefficients of each of these factors in $R[x]$ must be 1. By corollary 11.2.2, each of these factors are irreducible in $R[x]$! Hence $p(x)$ can be written as the product of irreducibles in $R[x]$.

The uniqueness of the factorization of $p(x)$ follows from the uniqueness in $k[x]$. Suppose

$$p'(x) = q_1(x)q_2(x)\cdots q_n(x) = q_1'(x)q_2'(x)\cdots q_m'(x)$$

Since the g.c.d. of the coefficients of $p'(x)$ is 1, so to it is for both factorizations. By corollary 11.2.2, each $q_i(x)$ and $q_j'(x)$ is an irreducible in $k[x]$. By uniqueness of factorization on $k[x]$, $n = m$, and the up to rearrangement, the two factorizations are the same up to associates, that is $q_i(x) = uq_i'(x)$ for a unit $u \in k[x]$. It remains to show they are associates in $R[x]$. By proposition 9.2.2, the untis of $k[x]$ are the units of F, so

$$q_i(x) = \frac{a}{b}q_i'(x)$$

for $a, b \in R$. Thus, $bq_i(x) = aq_i'(x)$. Thinking about it another way, the g.c.d. of the coefficients of the polynomial on the left hand side is b, and a on the right hand side respectively. Since in a Unique Factorization Domain, the coefficients of a nonzero polynomial is unique up to units ($a = ub$ for a unit in R), then $q_i(x) = uq_i'(x)$, and so $q_i(x)$ and $q_i'(x)$ are associates in R as well, completing the proof!

Corollary 11.2.4: R UFD then Polynomial of several varaibles over R UFD

If R is a Unique Factorization Domain, then a polynomial ring in an arbitrary number of variables with coefficients in R is also a UFD

Proof **:**
For the finitely many case, apply induction on theorem 11.2.3, since $R[x, y] = R[x][y]$, and $R[x]$ is a Unique Factorization Domain. The general case follwos fromthe definition of a polynomial ring in an arbitrar ynumber of variables as the union of polynomial rings in finitely many variables.

Example 11.2.1

1. $\mathbb{Z}[x], \mathbb{Z}[x, y]$, and so on are Unique Factorization Domains. The ring $\mathbb{Z}[x]$ now gives a good example of a Unique Factorization Domain that is *not* a Principal Ideal Domain!
2. Similarly, $\mathbb{Q}[x]$, $\mathbb{Q}[x, y]$, and so forth are Unique Factorization Domains.

Finally, we'll ask if we can weaken the condition's and still get the same result. What if R was just an integral domain? Can we follow a similar construction? Generalizing in a different direction, if R is a Unique Factorization Domain, does that imply that $R[[x]]$ is a Unique Factorization Domain? The answer to both is in general no

Example 11.2.2: Counter-Example

1. Let $R = \mathbb{Z}[2i]$, and let $p(x) = x^2 + 1$. Then the field of fractions of R is $F = \mathbb{Q}[2i]$. The polynomial $p(x)$ factors uniquely into a product of two linear facrors in $k[x] : x^2 = (x-i)(x+1)$. so in particular $p(x)$ *is* reducible in $\mathbb{Q}[2i]$. However, Neither of these factors lies in $R[x]$ (because $i \notin R$) so $p(x)$ is *irreducible* in $R[x]$. In particualr, by corollary 11.2.2, $\mathbb{Z}[2i]$ is *not* a Unique Factorization Domain! If it were, then it should've been reducible!
2. It is a rather technical project to show that if R is a UFD, then $R[[x]]$ is a UFD. The somewhat canonical counter-example is

$$R = \frac{k[x, y, z]}{(x^2 + y^3 + z^7)}$$

See (citation commented) for an in-depth explanation[a].

[a] this paper does require some algebraic knowledge introduced in part V

Exercise 11.2.1

1. Let $x^5 - x + 1 \in \mathbb{Q}[x]$. Is this polynomial irreducible? (Hint: use Gauss's Lemma)
2. Recall that if R Is a UFD, it is not necessarily the case that $R[[x]$. Show that $\mathbb{R}[[x]]$ is a UFD

11.3 Irreducibility of Polynomials

Now that we've shown R, $R[x]$, $R[x, y]$ and so forth are Unique Factorization Domains if and only if R is a Unique Factorization Domain and that irreducibility over $R[x]$ is symmetrically reflected to irreducibility over $k[x]$, we now raise the question about finding irreducible elements in such a

ring (i.e. the prime elements – the "building blocks"). A polynomial $p(x)$ is reducible if we can find two polynomials of smaller degree such that $p(x) = a(x)b(x)$. A polynomial is irreducible if no such polynomials exist. In general this is difficult, which is why we look for nicer irreducibility criteria.

Throughout these proofs, we shall be taking advantage of the fact that the field of fractions is a UFD, hence the polynomial ring over the field is a UFD. We shall not even require the ring that we are using to construct the field of fraction to be a UFD, simply an integral domain. To demonstrate, consider the following key fact:

Proposition 11.3.1: Finding Linear Factors

Let R be an integral domain and let $p(x) \in R[x]$. Then $p(x)$ has a factor of degree one if and only if $p(x)$ has a root in F, i.e., there is an $r \in R$ such that $p(\alpha) = 0$

Proof **:**
Let $p(x) = (x - r)q(x)$ for some $r \in R$, $q(x) \in R[x]$. Then certainly

$$p(r) = (r - r)q(r) = 0q(r) = 0$$

Hence, r is a root. Conversely, if $p(r) = 0$ for some $r \in R$, take the field of fractions and consider $k[x]$. Then by the division algorithm in $k[x]$ we may write

$$p(x) = q(x)(x - a) + r$$

Since $p(r) = 0$, then r must be 0. Finally, multiply $q(x)$ by an appropriate α to clear the denominators so that $\alpha q(x) = Q(x) \in R[x]$, and hence we get

$$p(x) = Q(x)(x - r)$$

as we sought to show.

Corollary 11.3.2: Maximum Number Of Roots

[Maximum Number Of Roots] Let R be an integral domain and $p(x) \in R[x]$ of degree n. Then $p(x)$ has *at most* n roots.

Proof **:**
Let $k = K(R)$ be the field of fraction of R. Then the number of roots of $p(x) \in [R]x$ is certainly less than or equal to the number of roots in $p(x) \in k[x]$. Now, *a fortiori* k is a UFD, hence by theorem 11.2.3 $k[x]$ is a UFD, and so each root corresponds to an (irreducible) factor of degree 1. But then, f can have at most n factors of degree 1 by unique factorization, completing the proof.

Note that this statement can fail if R is not an integral domain!

Example 11.3.1: Polynomial With More Roots And Its Degree
[Polynomial With More Roots And Its Degree] Let $x^2 + x \in (\mathbb{Z}/6\mathbb{Z})[x]$. Show that this polynomial has 6 roots!

However, using proposition 11.3.1, we may show that polynomials over infinite integral domains are determined by their roots!

Corollary 11.3.3: Polynomials Detremiend By Evaluation

[Polynomials Detremiend By Evaluation] Let R be an infinite integral domain and $f, g \in R[x]$. Then $f = g$ if and only if

$$r \mapsto f(r) \qquad \text{and} \qquad r \mapsto g(r)$$

agree for all $r \in R$

Proof **:**
Two functions agree if every $r \in R$ is a root of $f - g$¡ but a nonzero polynomial over R cannot have infinitely many roots by corollary 11.3.2, hence they must be equal as we sought to show.

Continuing from this, we may ask about the irreducibility of degree 2 and 3 polynomials, which is an immediate consequence of the above results:

Proposition 11.3.4: Degree 2 Or 3 Polynomials

A polynomial of degree two or three over a field F is reducible if and only if it has a root in F.

Proof **:**
This follow's immediately from the previous proposition, see exercise ref:HERE

The reason for this is that if we have a degree higher than 4, then we can reduce to two irreducible polynomials of degree 2, like $(x^2 + 1)(x^2 + 1)$ in $\mathbb{Q}[x]$. For higher degree polynomials, we may have more solutions depending on what type of field we are working over. For example, we shall introduce in chapter 17, section 17.5.1, the existence of field for which only linear polynomials are irreducible ($\mathbb{C}$ being a canonical example of such a field). Yet another example is that any polynomial of degree ≥ 3 over $\mathbb{R}$ is reducible[3]. For more general rings, it is in general much more difficult to determine if a polynomial is reducible. There are however some tricks that help in arbitrary situations, especially if R is a UFD (for then we may use the gcd)

This next proposition let's us limit the number of "rational" solutions of a polynomial over a UFD.

Proposition 11.3.5: Rational Roots Test

Let R be a UFD and $k = K(R)$. Let $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdot + a_1 x + a_0$ be a polynomial of degree n with coefficients in the UFD R. If $r/s \in k$ with $gcd(r, s) = 1$, and r/s is a root of $p(x)$ ($p(r/s) = 0$), then r divides the constant term and s divides the leading coefficient of $p(x)$: $r|a_0$ and $s|a_n$. In particular, if $p(x)$ is a monic polynomial with integer coefficients and $p(d) \neq 0$ for all integers d dividing the constant term of $p(x)$, then $p(x)$ has no roots in $k[x]$.

[3]This is exercise ref:HeRE (closure of fields, 3rd exercise). An astute student would be able to work out this exercise

***Proof* :**
Let's say $p(r/s) = 0$. Then

$$p(r/s) = 0 = a_n(r/s)^n + a_{n-1}(r/s)^{n-1} + \cdots + a_1(r/s) + a_0$$

multiplying all by s^n, we get:

$$0 = a_n r^n + a_{n-1} r^{n-1} s + \cdots + a_1 r s^{n-1} + a_0 s^n$$

so

$$a_n r^n = s(-a_{n-1} r^{n-1} - \cdots - a_1 r s^{n-1} - a_0 s^{n-1})$$

thus, $s \big| a_n r^n$. $gcd(r, s) = 1$, then it must be that $s|a_n$. Doing the same trick for a_0:

$$a_0 s^n = r(-a_n r^{n-1} - a_{n-1} r^{n-2} - \cdots - a_1 s^{n-1})$$

then $r|a_0$.

If $a_n = 1$, then we'd have $s|1$, so $r/s = r/1 = r$. This means we only need to check the values r which divide a_0. If every value that dividse a_0 is not a root of $p(x)$, then $p(x)$ has no roots.

This gives us a really easy way of testing irreducibility of low-degree polynomials:

Example 11.3.2: Irreducibility Of Low-Degree Polynomials

1. The polynomial $p(x) = x^3 - 3x - 1$ is irreducible in $\mathbb{Z}[x]$. Since the polynomial is monic, by corollary 11.2.2, it must be reducible in $\mathbb{Q}[x]$. By the rational number test, we are given the candidates ± 1. Plugging both of them in, we get no roots of $\mathbb{Q}[x]$, and so by 11.3.4, this polynomial is irreducible, and so is irreducible in $\mathbb{Z}[x]$.
2. For any prime p, $x^2 - p$ and $x^3 - p$ are irreducible in $\mathbb{Q}[x]$. We have ± 1 and $\pm p$ as candidates. Both of them are not roots, and therefore, these polynomials are not reducible!
3. The polynomial $x^2 + 1$ is *reducible* in $(\mathbb{Z}/2\mathbb{Z})[x]$, namely $(1)^2 + 1 = 1 + 1 = 0$. It's factors are $(x + 1)(x + 1) = x^2 + 1$. The rational roots test applies since TBD.
4. The polynomial $x^2 + x + 1$ is *irreducible* in $(\mathbb{Z}/2\mathbb{Z})[x]$, since there are no roots.
5. Similarly for $x^3 + x + 1$ in $(\mathbb{Z}/2\mathbb{Z})[x]$

All the polynomials in the aforementioned examples are all of degree less than or equal to 3. This is because when factoring polynomials of degree greater than or equal to 4, it's possible that we get 2nd degree polynomials as irreducible factors (like $(x^2 + 1)(x^2 + 1)$ in $\mathbb{Q}[x]$, which has no factors).

We use another trick for these type of polynomials: We can use 11.0.1 ($R[x]/I[x] \cong (R/I)[x]$) to generalize the rational root test:

Proposition 11.3.6: Reduction Test

Let I be a proper ideal in the integral domain R and let $p(x)$ be a non-constant *monic* polynomial in $R[x]$. If the image of $p(x)$ in $(R/I)[x]$ cannot be factored in $(R/I)[x]$ into two polynomials of smaller degree, then $p(x)$ is irreducible in $R[x]$.

Proof **:**
For the sake of contradiction, let's supose $p(x)$ cannot be factored in $(R/I)[x]$, but $p(x)$ is reducible in $R[x]$. Then there must be two polynomials such that $p(x) = a(x)b(x)$. Note that both $a(x)$ and $b(x)$ must be monic.

Reducing the coefficients modulo I gives a factorization in $R[x]/I[x]$, but by 11.0.1, this is equivalent to $(R/I)[x]$ – contradiction our assumpiton.

The power here comes from the fact that R/I might have much simpler structure for factoring (ex. $\mathbb{Z}$ vs $\mathbb{Z}/n\mathbb{Z}$!) For example, in exercises 11.2.1, we may prove that $x^5 - x + 1$ using this trick rather quickly! So, If we reduce a polynomial into $(R/I)[x]$, and show it cannot be factored, then it is irreducible in $R[x]$ as well! However, the converse (if $p(x)$ is irreducible in $R[x]$, then $p(x)$ is irreducible in every $(R/I)[x]$) is not true.

Example 11.3.3: Limitation of reduction method
Take $x^4 + 1 \in \mathbb{Z}[x]$. This polynomial is irreducible in $\mathbb{Z}[x]$, (hint: recall that $\sqrt{2}$ is irrational), but is reducible for every prime. Verify this by considering cases of

$$p \equiv 1, 3, 5, 7 \bmod 8$$

since the case of $p = 2$ can quickly be manually verified.

So, if you get that a polynomial is reducible in $(R/I)[x]$, that unfortunately does not tell you it's reducible in $R[x]$.

A special case of proposition 11.3.6 that is quite useful is called Eisenstein's Criterion (or Eisenstein-Schönenmann Criterion, since Schönenmann discovered it first)

Proposition 11.3.7: Eisenstein's Criterion

Let P be a prime ideal of the commutative ring R and let $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ be a polynomial in $R[x]$ (where $n \geq 1$). Suppose $a_{n-1}, ..., a_0 \in P$, but $a_0 \notin P^2$. Then $p(x)$ is irreducible.

Proof **:**
Let's take $p(x)$ like in the proposition. For the sake of contradiction, let's say $p(x)$ is reducible: $p(x) = a(x)b(x)$, where $a(x)$ and $b(x)$ are non-constant.Then by proposition 11.3.6 (the contrapositive of it), it is reducible in $(R/P)[x]$. Since every non-leading coefficient is in P:

$$\overline{p(x)} = \overline{x^n} = \overline{a(x)} \cdot \overline{b(x)}$$

Since P is prime, R/I is an integral domain, and so the constant terms of $a(x)$ and $b(x)$ must be in P, or else we'll get $\overline{x} \cdot \overline{y} = \overline{0}$. But then, multiplying those terms means $a_0 \in P^2$, a contradiction!

For the purposes of this book, the following is the most immediate use of Eisenstein's Criterion:

Corollary 11.3.8: Eisenstein's Criterion For $\mathbb{Z}[x]$

Let p be a prime in $\mathbb{Z}$ and let $f(x) = x^n + a_{n-1}x^{n-1} + \cdots a_1 x + a_0$. If every a_i is divisible by p, but a_0 is not divisible by p^2, then $f(x)$ is irreducible in both $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.

***Proof* :**
This is proposition 11.3.7 and using corollary 11.2.2.

Here are some examples using Eisenstein's criterion, and some more tricks when finding irreducible polynomials:

Example 11.3.4: Irreducibility Tricks

1. When solving polynomials of degree 3 or less, it'd suggest to always start with checking roots: it's the simpliest criterion.

2. The polymoial $x^4 + 10x + 5$ in $\mathbb{Z}[x]$ is irreducible, if we take $p = 5$. Since $5|10$, $5|5$, but $5^2 \nmid 5$, then Eisntein's criterion holds.

3. For any integer a such that $p|a$ but $p^2 \nmid a$, then $x^n - a$ is irreducible in $\mathbb{Z}[x]$. If $a = p$ a prime, then $x^n - p$ is irreducible for all positive integers n, so for $n \geq 2$, the nth root of p are not rational numbers (i.e. has no roots in $\mathbb{Q}[x]$)

4. We cannot apply the same tricks we did before to the polynomial $x^4 + 1$ in $\mathbb{Z}[x]$, since nothing divides prime, and any proper ideal will not contain 1. Instead, we'll use a trick called *substitution*.

   Consider the polynomial $f(x+1) = (x+1)^4 + 1 = x^4 + 4x^3 + 6x^2 + 4x + 2$. On this polynimal, Eisntein's criterion works, since all non-leading coefficients are divisible by 2, and $2^2 \nmid a_0$. From this, it follows that $f(x)$ is irreducible, since any factorization of $f(x)$ will be a factorization for $f(x+1)$ (replace x by $x+1$ in the factorization!) Remember this trick for polynomials which seem at first glance to not get an application of Eisntein's criterion.

5. (cyclotomic polynomials) Perhaps the most famous application of Eisenstein's criterion is on polynomials of the following form:

   $$1 + x + x^2 + \cdots = x^{p-1} \in \mathbb{Z}[x]$$

   Eisenstein's criterion cannot be directly applied, however as before we can do the $x \mapsto x+1$ trick to get

   $$p(x+1) = 1 + (x+1) + \cdots + (x+1)^{p_1} = x^{p-1}\binom{p}{p-1}x^{p-1} + \cdots \binom{p}{2}x + \binom{p}{1}$$

   where $\binom{p}{1} = p$.It suffices to show that $p\Big|\binom{p}{k}$ for $1 \leq k \leq p-1$ to conclude the result.

6. On $\mathbb{Q}[x][X]$?

7. Notice that in the proof, we reduced to $(R/P)[x]$. If $R = \mathbb{F}$ was a field, then the only ideals are 0 and $\mathbb{F}$, meaning $(\mathbb{F}/P)[x] \cong 0[x] \cong 0$. Therefore, Einsteins criterion doesn't work will with such polynomials.

   For example, is $x^4 + 1$ reducible in $\mathbb{F}_5[x]$ ($\mathbb{F}_5$ is the field with 5 elements, and is isomorphic ot $\mathbb{Z}/p\mathbb{Z}$)? If if were over $\mathbb{Z}[x]$, we already mentioned how it is irreducible (using a tirck we which will be explained in chapter ref:HERE). It also has no roots, as can be verified by plugging in $0, ..., 4$.

   We can't use einstein's criteron, but since the degree is 4, we know if a factorization did exist, they would have to be of degree 2, that is

   $$x^4 + 1 = (x^2 + ax + b)(x^2 + cx + d) = x^4 + (a + c)x^4 + (b + ac + d)x^2 + (ab + bc)x + bd$$

   note that the coefficients of x^2 are one since the coeficient of x^4 is one. We can then deduce that

   $$a + c = 0, b + ac + d = 0, ad + bc = 0, bd = 1$$

   Manipulating a bit, we can get the following information:

   $$\begin{aligned} c &= -a \\ b + a(-a) + d &= 0 \\ ad + b(-a) &= 0 \\ a^2 &= b + d \\ ad &= ab \end{aligned}$$

   Let's take cases: let's say $a \neq 0$. Then $b = d$, $a^2 = 2b$, $b^2 = 1$. (TBD). So, we get a contradiction.

   Let's check $a = 0$. Then $062 = 0 = b + d$, implying $b = -d$, and $b^2 = -bd = -1 = 4$, so $b = 2$. implying $d = 2$ and $a = 0$, $c = 0$. No contradiction was found! Thus, we can plug in, and factor to get

   $$p(x) = x^4 + 1 = (x^2 - 3)(x^2 - 2) = (x^2 + 2)(x^2 + 3)$$

   Thus, $p(x)$ is reducible in $\mathbb{F}_5[x]$.

There are further irreducibiltiy algorithms that were programmed. One algorithm is Berlekamp Algorithm. there are others TBD.

Finally, we give a few important results that shall be very handy tools. The first tells us that factorization of a polynomial is invariant to field extensions, namely if $F \hookrightarrow E$ is an injective map, then the factor of $p(x)$ in E are the same factors (if not further reduced) in $E(x)$:

Proposition 11.3.9: Invariant Under Field Extensions

If you extend from a field F to E, then by the uniqueness of the euclidean algorithm, everything in $E[x]$ will still factor the same way as before (in $k[x]$): it cannot inroduce new solutions.

Proof **:**
exercise

This next one is of particular interest for the structure of the multiplicative group of a field!

Theorem 11.3.10: Subgroup of Multiplicative Group is Cyclic

A finite subgroup of the multiplicative group of a field is cyclic. In particular, if F is a finite field, then the multiplicative group $F^\times$ of nonzero elements of F is a cyclic group.

Another way of saying this is that $F^\times$ is locally cyclic (being "locally P" means that every finite subgroup has the property P)

Proof **:**
We will use the fundamental theorem for finitely generated abliean groups!

Let $A \subseteq F^\times$ be a finitely generated abelian group. Then:

$$A \cong \mathbb{Z}/n_1\mathbb{Z} \times \mathbb{Z}/n_2\mathbb{Z} \times \cdots \times \mathbb{Z}/n_k\mathbb{Z}$$

where $n_k \big| n_{k-1}, n_{k-1} | n_{k-2}, ..., n_2 \big| n_1$. Since n_k divides the order of each of the cyclic groups in the direct product, it follows that each direct factor contains n_k elements of order dividing n_k (by prop ref:HERE). If k were greater than 1, there would therefore be a total of more than n_k such elements. But then there would be more than n_k roots of the polynomial $x^{n_k} - 1$ in the field F, contradicting proposition 11.3.1!! Hence, $k = 1$, and the group is cyclic!

Corollary 11.3.11: $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic

Let p be a prime. The multiplicative group $(\mathbb{Z}/p\mathbb{Z})^\times$ of nonzero residue classes mod p is cyclic

Proof **:**
Since $\mathbb{Z}/p\mathbb{Z}$ is a field, apply proposition 11.3.10

These corollaries will be useful when proving Kronnecker-Weber's Theorem (theorem 22.5.3):

Corollary 11.3.12: Isomorphisms Of Multiplicative Groups

Let $n \geq 2$ be an integer with factorization $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n}$ in $\mathbb{Z}$, where $p_1, ..., p_n$ are distinct primes. Then we have the following isomorphsism of (multiplicative) groups:

1. $(\mathbb{Z}/n\mathbb{Z})^\times = (\mathbb{Z}/p_1^{\alpha_1})^\times \times (\mathbb{Z}/p_2^{\alpha_2})^\times \times \cdots \times (\mathbb{Z}/p_n^{\alpha_n})^\times$
2. $(\mathbb{Z}/2^\alpha\mathbb{Z})^\times$ is the direct product of a cyclic group of order 2 and a cyclic group of order $2^{\alpha-2}$, for all $\alpha \geq 2$
3. $(\mathbb{Z}/p^\alpha\mathbb{Z})^\times$ is a cyclic group of order $p^{\alpha-1}(p-1)$ for all odd primes p

Proof **:**
exercise (comment)

Exercise 11.3.1

1. Show that there are irreducible polynomial in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$ for every degree.
2. Let $f(x)$ be an irreducible monic polynomial of $\mathbb{Z}[x]$. Show that $f(x)$ mod p has a root for infinitely many primes. The opposite is true too, that there are infinitely many primes for which $f(x)$ mod p does *not* have a root, but as far as I am aware that requirse some analytic number theory[4]
3. Show that $t^2 + t + 1 \in \mathbb{F}_2[t]$ is irreducible. Use this to show that $\mathbb{F}_2[t]/(t^2 + t + 1)$ is a field with 4 elements.

[4] see CITE:CFT

Part III

Modules

When we explored groups, we saw that the kernel of groups were groups (in particular a special subset of a group called the normal subgroup), and if we have an abelian group, every abelian group is the kernel of some group.

The story for rings is more complicated. As we mentioned at the beginning of the part on rings, some authors include an identity in the definition, others do not. If we include an identity, then the kernel of ring homomorphisms are not rings. Even if we take exclude identities as part of the definition, notice that practically non of the ideals we talked about were previously mentioned rings, essentially because almost all rings we talk about have an identity, but if an ideal has an identity, it is the entire ring. The cokernels is also not as well-behaved (recall that $i : \mathbb{Z} \hookrightarrow \mathbb{Q}$ is an epimorphism). In this way, the kernel of rings and rings are very different objects, and so a lot goes into studying ideals as though they are a seperate object in ring theory. In particular, we do not focus a lot of attention of constructing rings based on other rings (similar to how we constructed polyabelian or poly-cyclic groups). Overall, the category of **Ring** and **Rng** are not the best behaved categories.

The concept of Modules can be thought of as a fix to this. Modules are more similar to abliean groups than they are to rings, for example if you have any sub-module, then the quotient will always be well-defined. Furthermore, all Rings can be thought as a module, making the study of module subsumes the study of rings. As we shall see,

We shall see that an R-module M will be an abelian group M with an ring action by R. The reason M is an abelian group and not just a set like in group actions is because it is a generalization of vector-spaces, not an addition to group actions. What I mean by this is that you have to think of the reasoning from the point of view that vector-spaces came first, and then there were many instances where the field acting on the group was too restrictive, which lead to module.

Over the 19th and 20th century, many mathematical objects were studied that all turn out to be modules. To those who want the punchline, I'll point out a few right now:

1. All abelian groups are $\mathbb{Z}$-modules
2. All linear transformations are $F[x]$-modules
3. All rings are modules (If R is a ring, then it is an R-module over itself)
4. All ideals are modules (if $I \subseteq R$ is an ideal, it is an R-module)
5. All vector-spaces over k are k-modules
6. All vector-bundles are projective modules

12

Introduction to Modules

A lot of the algebraic structures we've covered so far share a similar story: we ask how maps between them can be studied (leading us to explore sub-objects and quotient-objects), and we ask how to generate them. Modules will have the same story at the beginning. [Question: Why didn't group actions have the same story? no sub/quotient actions]. When we'll be building up modules, they'll more closely resemble groups and rings. In groups, simple groups and free groups played a bigger role than they did in ring theory. Same in Module theory (TBD).

12.1 Basic Definitions and Examples

Definition 12.1.1: [left]-Module

Let R be a ring (more generally a rng). A *left R-module* or a *left module over R* is a set M together with

1. a binary operation $+$ on M under which M is an **abelian group**, and
2. an action of R on M (that is, a map $R \times M \to M$) denoted by rm, for all $r \in R$ and for all $m \in M$ which satisfies
   (a) $(r+s)m = rm + sm$, for all $r, s \in R$, $m \in M$
   (b) $(rs)m = r(sm)$, for all $r, s \in R$, $m \in M$
   (c) $r(m+n) = rm + rn$ for all $r \in R$, $m, n \in M$

   If the rng R is a ring (i.e. has a 1), we impose the additional axiom

   (d) $1m = m$, for all $m \in M$

   And we sometimes call such a module a *unital* R-module.

Right modules are defined similarly. If the R ring is commutative, and M is an left $R-module$, we can turn it into a right $R-module$ by defining $rm = mr$. To distinguish when different rings may be acting on an abelian group M, the symbol $\curvearrowright$ (read *acts on*) may be used to show that a ring R is acting on M: $R \curvearrowright M$

Unless stated otherwise, the term "module" will reference "left-module" Note that for a module M, we usually denote what ring it is over (or what ring acts upon it) by saying it is an *[ring symbol here]-module* (ex. R-module, $F[x]$-module, $[M_{m\times n}(\mathbb{F})]$-module). We might also say R acts on M where R is the ring and M is the abelian group. We will always be working *rings* actions, however as noted in the definition, we can also work with *rng* actions. Note too that we will often say that "R acts on the module M" to mean that we define a ring-action of R on the abelian group M.

When working over rings, the distributivity condition forces M to be abelian: let $r = s = 1$:

$$(1+1)\cdot(m+n) = (1+1)\cdot m + (1+1)\cdot n = 1\cdot m + 1\cdot m + 1\cdot n + 1\cdot n = m + m + n + n$$

$$(1+1)\cdot(m+n) = 1\cdot(m+n) + 1\cdot(m+n) = 1\cdot m + 1\cdot n + 1\cdot m + 1\cdot n = m + n + m + n$$

which implies $m + n = n + m$ for all $m, n \in M$. Notice how this when working with group actions, the group acting on the set did not effect any strucutre of set, namely since the set has no additional structure. Be mindeful that when talking about a module, we will usually call M the module. For example we might say "Let $\mathbb{Z}$ act on $\mathbb{Q}$ by $r \cdot a = ra$; then $\mathbb{Q}$ has the following properties". When talking about $\mathbb{Q}$ here, we're talking about $\mathbb{Q}$ as a module, namely a $\mathbb{Z}$-module, not just as abelian group or a ring.

If R is commutative then we defined a two-sided action $r \cdot m = m \cdot r$. If R is not commutative, then defining this property will have problems with $2(b)$: Let R be non-commutative, let $r \cdot m = m \cdot r$, and take $r, s \in R$ such that $rs \neq sr$ and $rs \cdot m \neq sr \cdot m$ (an example of such a module will be presented

in the examples 12.1.1). Then:

$$(rs)m = m(rs) = (mr)s = (rm)s = sr(m) = (sr)m$$

however, $sr \cdot m \neq rs \cdot m$ so this statement is not necessarily always true. Note how it may be possible that even though $rs \neq sr$ that $rs \cdot m = sr \cdot m$, for example, the trivial module structure $r \cdot m = m$ for any $r \in R$ will produce this result, however as we'll soon see in example 12.1.1, every module will always induce a *faithful* module (the word faithful being defined equivalently to the group-action case[1]), and so making a distinction between left and right modules is very important over non-commutative rings.

We now have many operators to keep track of: the group operation of M, the two rings operations of R, and the ring action of R on M. When working with modules, we often focus on the ring-action, and use group and rings structures that were previously covered.

One might ask themselves if modules and group-rings might be similar concept, boths somehow being a ring "acting" on a group. If this suspicions crossed your mind, it is good idea to check how these two structure's differ (for example, if $|G| > 1$, recall a group-ring always has zero-divisors. Find an example in the list of examples bellow that does not have this property).

Since a module is an action on an abelian group, it would make sense if there is an equivalent to a permutation representation for module's as there is for group-actions. Recall that any group action can be seen as a group-homomorhpism:

$$\rho : G \to \mathrm{Sym}(S)$$

The exact same idea works for defining an R-module as an action. In order to do so, we need an a notion of an R-module homomorphism that perseves some of the structure. Before looking at the definition bellow, see if you can come up with the definition of an R-module homomorhpism.

Definition 12.1.2: R-module homomorphism

Let R be a ring and let M and N be R-module A map $\varphi : M \to N$ is an *R-module homomorphism* if it respects the R-module structure of M and N, i.e.

1. $\varphi(x + y) = \varphi(x) + \varphi(y)$ for all $x, y \in M$
2. $\varphi(r \cdot x) = r\varphi(x)$, for all $r \in R$, $m \in M$

Since M and N are both abelian groups, R-module homomorphisms are sometimes called *R-linear group homomorphisms*, where the map is R-linear since the ring actions passes through the homomorphism. Notice that both M and N both have the same ring acing on them: it is imperative that an R-module homomorphism be betewen R-modules.

Returning to the discussion on treating R-modules as an action, it is a good exercise to check ones understanding of the axioms of modules that the "axiomatic representation" of an R-module is equivalent to defining an R-module as a ring-homomorphism

$$\rho : R \to \mathrm{End}_{\mathbf{Ab}}(M)$$

where $\mathrm{End}_{\mathrm{Ab}}(M)$ is the collection of all group-homomorphisms from M to itself. Notice that $\mathrm{End}(M)$ is an [abelian] group because M is an abelian group, while the ring operation of composition, $\circ$,

[1] That is, there must exist an $m \in M$ st $r \cdot m = s \cdot m$ implies $r = s$

automatically satisfies the condition to make $\mathrm{End}_{\mathbf{Ab}}(M)$ a ring, which should be checked (for example, $\mathrm{id} : M \to M$ is an element of $\mathrm{End}_{\mathbf{Ab}}(M)$, hence it has a multiplicative identity). It will sometimes useful to define an R-module as a ring homomorphsim $R \to \mathrm{End}_{\mathbf{Ab}}(M)$, if we want to manipulate with the ring action on it (see exercise ref:HERE (how $R \times S$ can act on an R-module, easy to see with this definition))

As with rings and and with groups, we can define a category: given a ring R, the collection of R-modules and the collection of R-module homomorphisms forms a category:

Definition 12.1.3: R-Mod

The collection of all R-modules and R-module homomorphisms forms a category, which we'll denote R-**Mod** or $\mathbf{Mod}_R$.

This category has some of the nice properties that the other categories that were extensively worked in so far have:

Definition 12.1.4: Classical definition of Submodule

Let R be a ring and let M be an R-module. An R-submodule of M is a subgroup N of M which is closed under the action of ring elements, i.e., $rn \in N$, for $r \in R$, $n \in N$

Definition 12.1.5: Categorical Definition of Submodule

Let $N \subseteq M$. Then if there exists a R-module inclusion map $\varphi : N \to M$ that is also an R-module homomorphism, then N is a submodule of M

Proposition 12.1.6: The sub-module Criterion

Let R be a ring and let M be a R-module. A subset of N of M is a submodule of M if and only if

1. $N \neq \emptyset$, and
2. $x + ry \in N$ for all $r \in R$, and for all $x, y \in N$

Proof **:**
Easy to prove both ways (recall that M is an abelian group, and for the other way, simply set the values of r and x to some appropriate values)

Example 12.1.1: Modules

1. Let R be any ring. Then $M = R$ is a left R-module with the ring multiplication as the ring action: $r \cdot s = rs$. Take a moment to think about what are the submodules of R? The answer will be revieled in the next sentence. the sub-modules of M are the left-ideals of R! This is because ideals of R are closed under multiplication of R, and the axioms of modules require closure under "scalar multiplication" by the ring elements. Is R/I for any (two-sided) ideal of R also an R-module? Take a moment to think about it. The answer is that R/I

is indeed an R-module! We will soon show that the kernel of module-homomorphisms will be submodules, showing that the axioms of a module "fix" the category of **Ring** by being closed under kernels!

From now on, if we say that R is a [left/right] R-module over itself, it will be taken that the R-action on R is [left/right] multiplication by R, with a bias towards *left* R-modules if R is non-commutative and which side of the action was not specified. We will also sometimes say "let R have the natural R-module structure" to mean the one induced by the ring operation of R.

2. Let R be a ring with 1 and let $n \in \mathbb{Z}^+$. Then since R is a group, we can define the direct product:
$$R^n = \{(a_1, ..., a_n) | a_i \in R, \text{ for all } i\}$$
We can turn R^n to an R-module by defining addition in R^n component wise, and the action being componentwise:
$$r(a_1, ..., a_n) = (ra_1, ..., ra_n)$$
Such a module R^n is called a *free module* [a]. More on these in section 12.3.2. If $R = \mathbb{R}$, then we have the usual algebraic structure of $\mathbb{R}^n$ common in analysis and linear algebra.

3. If M is an R-module, and S is a sub-ring of R, then M is automatically an S-module (i.e. the set M is also a module under the ring action with S). Since if the module axioms hold for R, i.e. the ring being closed under the ring action and the rules holding, then it will still hold for a sub-ring. In this way $\mathbb{R}$ is an $\mathbb{R}$-module, a $\mathbb{Q}$-module, and a $\mathbb{Z}$-module with the ring action being defined by regular multiplication. Notice though that the $\mathbb{R}$-submodules, $\mathbb{Q}$-submodules, and $\mathbb{Z}$-submodules differ, showing that having a different ring act on $\mathbb{R}$ greatly changes its module structure.

4. Let M be an R-module. If $I \subseteq R$ is a two-sided ideal, is it always true that M is an R/I-module? Under what condition's is it an R/I-module? The following is an exploration of when this is the case: Let I be a 2-sided ideal of R such that $am = 0$, for all $a, \in I, m \in M$. It should be checked that I is indeed an ideal (it is usually denoted $\text{Ann}_R(M)$, and called the *annihilator* of M [over R]). Then we say that M is *annihilated* by I. In this situation, we can make M into an (R/I)-module by defining an action of the quotient ring R/I on M as follows: for each $m \in M$, and $(r + I) \in R/I$, let
$$(r + I)m = rm$$
This is well defined since for any $a \in I$, $(r + a)m = rm + am = rm$. Check that the axioms of a module still hold. Furthermore, The R/I action is a *faithful* action, since if $rx = sx$, then $r = s$. To see this, consider any $x \in M$. Then since M and R are groups, define the group-homomorhpism:
$$\varphi : R/I \to M \qquad \varphi(\overline{r}) = \overline{r} \cdot x$$
The function φ is indeed a group-homomorphism, since $\varphi(\overline{r} + \overline{s}) = (\overline{r} + \overline{s}) \cdot x = \overline{r} \cdot x + \overline{s} \cdot x = \varphi(\overline{r}) + \varphi(\overline{s})$. The kernel of this action is all elements such that $\varphi(\overline{r}) = \overline{r} \cdot x = (r + I) \cdot x = 0$. By definition of I,
$$I \cdot x = \{i \cdot x \mid i \in I\} = \{0\}$$
so it follows that if $\overline{r} \in \ker(\varphi)$, $\overline{r} = \overline{0}$, showing that $\ker(\varphi) = \overline{0}$ (Notice the parallell with the Orbit-Stabalizer, theorem 4.2.4)

If, conversely, if $I = \mathrm{Ann}_R(M)$, and we have defined an R/I-action on a module M, can we define an R-action on M? More generally, given two rings R and S acting on M, when can we say that the module structure they induce on M is "comparable"' in some way? One way two types of modules can be equivalent is if they are something called "Morita equivalence", something which will be explored throughout exercises and chapter 23 and 30[b])

5. What is and isn't a well defined R-module isn't always intuitive. Take for example the group ring $R = \mathbb{Z}[G]$ where $G = \{1, \sigma\}$ is a group of order 2. For the abelian group, take the module $M = \mathbb{Z}e_1 + \mathbb{Z}e_2$, where e_1, e_2 are generators for $\mathbb{R}^2$ (for example, $(1, 0)$ and $(0, 1)$ generate $\mathbb{R}^2$ as an abelian group), and the product $\mathbb{Z}e_i = \{ze_i | z \in \mathbb{Z}\}$.

   Define the action as:
$$\sigma(e_1) = e_1 + 2e_2 \qquad \sigma(e_2) = -e_2$$
   Then M is a $\mathbb{Z}[G]$-module! The big one to check is $r_1(r_2m) = (r_1r_2)m$ – the other conditions follow more easily. This is an example of how the limit on what can be defined as a module isn't easy, since an example like this one exists, checking when a module is defined can be quite useful.

6. Let R be any ring and let $F(A, R)$ be the set of all [set]-functions from A to R:
$$f \in F(A, R) \Rightarrow f : A \to R$$
   Then define the action by R to be
$$r \cdot f = rf \qquad r \cdot f(a) := rf(a) \qquad \forall a \in A$$
   since $f(a) \in R$ Band $r \in R$. Define the group operation to be the new function $f + g$ where
$$(f + g)(a) := f(a) + g(a) \qquad \forall a \in A$$
   Note that these functions are *not* R-module homomorphisms, since the set A has no structure! This R-module is is often dentoed as R^A (to symbolise all functions from A to R)[c].

   Very often in Analysis and Algebraic Geometry, the set A will be all continuous functions on a topological space X, denoted $C(X)$, and the ring R would be the real or complex numbers $\mathbb{R}$ or $\mathbb{C}$. The study of these spaces is done commonly in functional analysis and classical algebraic geometry, usually requiring some analysis knowledge. In this textbook, these types of R-module will come back when discussing free-modules in theorem 12.3.6.

7. Let X be any set and $P(X)$ the powerset.

8. (An example with a ring acting on a module in two different ways, non trivially)

[a] not dissimilar to free-groups we studied in 2.3.3

[b] For the categorically inclined, R-mod and S-mod are Morita equivalent if they are equivalent as categories. Are R-mod and R/I-mod equivalent?

[c] The reason for this notation in fact very interesting and is covered in ref:HERE (in the category theory chapter, categorification of natural numbers, see Emily Reil Category Theory in Context)

When we worked with group-actions, we can take any group G and make it act on any set S. This is not the case with modules (i.e. given some fixed ring R and abelian group M, there does not always exist a ring-homomorphism $\varphi : R \to \mathrm{End}_{\mathbf{Ab}}(M)$):

Example 12.1.2: Impossible Module
Consider $M = \mathbb{Z}$ and $R = \mathbb{Q}$. Is there always a $\mathbb{Q}$ action over $\mathbb{Z}$? If it was, then:

$$\frac{1}{2} \cdot 1 = z$$

must be an element of $\mathbb{Z}$ ($z \in \mathbb{Z}$). Then:

$$z + z = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1 = (\frac{1}{2} + \frac{1}{2}) \cdot 1 = 1 \cdot 1 = 1$$

as a direct consequence of the module axioms. However, this implies

$$z + z = 1$$

which is not possible for any $z \in \mathbb{Z}$ – and thus a contradiction! This is also an example where you can have a $\mathbb{Z} \subseteq \mathbb{Q}$ and a defined $\mathbb{Z}$-module, but there is no extension of M to a $\mathbb{Q}$-module for which the previous module structure is embedded within it (opposite to how if we had a $\mathbb{Q}$-module, it would automatically be a $\mathbb{Z}$-module)[a]

[a] There is in fact still some notion of extending a module to scalars from a super-ring, however it is not necessary that the resulting module will can be restricted back to the structure of the original ring, see chapter 15, section 15.3.1

12.1.1 Vector-spaces

If the ring is a field, we give a special name to such a module:

Definition 12.1.7: Vector Space

Let M be an R-module. If R is a field, then we call M a *vector-space over M*. We usually denote arbitrary fields $\mathbb{F}$ and arbitrrary $\mathbb{F}$-modules (i.e. vector-spaces) V.

Due to historical reasons $\mathbb{F}$-modules homomorphisms are called *linear transformations*. The motivation for this will be more clearly seen in chapter 13, but the gist is that when $\mathbb{F} = \mathbb{R}$, the geometric intuition of the images and kernels of linear transformations will all be hyper-planes in the domain and codomain. A simple example would be $f : \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ with $f(x) = (x, x)$. The resulting hyper-plane (simply called a line since it's 1 dimensional), looks like so:

If you are comfortable with linear algebra, then you are probably familiar of how simple the structure of vector-spaces are (if not, you may skip this remark: all that will be discussed here is covered in chapter 13). Essentially, a vector space V over k can be completely determined by it's dimension. For modules over a general ring R, this is sometimes far from the case: sometimes R will not even embed itself in V! However, there are ways in which we can systematically loosen the conditions on the ring k acting on V and get similar, but weaker, results. Thus, if it is intuitive to think of modules as generlizations of vector-spaces, which historically is how many mathematicians thought of modules, then these chapters on module theory can be thought of as what happens if the action on a vector-space no longer forces to give a basis.

Example 12.1.3: Vector Spaces

1. The most common example is when $R = \mathbb{R}$ and $M = R^n = \mathbb{R}^n$. This vector-space is commonly reffered to as *euclidean space* due to the classical geometrical properties that can be defined on it(ex. a notion of distance, length, angle, translation and rotation)

2. If I is a maximal ideal in a commutative ring, and $IM = 0$, then M is a vector space over R/I (since R/I is a field)

3. Take $\mathbb{F}[x]$ to be a polynomial ring over a field $\mathbb{F}$, an let $f(x)$ be any polynomial of degree $n \geq 1$. Then $V = \mathbb{F}[x]/(f(x))$ is another polynomial ring. Let $\mathbb{F}$ act on V by $a \cdot \overline{f} = \overline{af}$. This will be a vector-space. If $\mathbb{F}$ is a finite field of order p, then we can show that V has p^n.

4. (If you know some calculus). Consider $\mathbb{R}[x]$. Then it is a vector-space over $\mathbb{R}$ ($\mathbb{R}[x]$ is an abelian group, and the action is simply the multiplication in $\mathbb{R}[x]$, since $\mathbb{R} \subseteq \mathbb{R}[x]$). Recall that the derivative has the following two properties:
$$\frac{d}{dx}(p(x) + q(x)) = \frac{d}{dx}p(x) + \frac{d}{dx}q(x)$$
$$\frac{d}{dx}(cf(x)) = c\left(\frac{d}{dx}f(x)\right)$$

This shows that the derivative is in fact a linear transformation (i.e. $\mathbb{R}$-module homomorphism) from $\mathbb{R}[x]$ to itself![a]

5. If R is a ring where the only two-sided ieals are (0) and R, then R is called a *simple ring* (similarly to how simple groups are groups whose only normal subgroup is $\{0\}$ and the entire group). Then $Z(R)$ (the center of the ring) *has* to be a field (can you show this?). If $k = Z(R)$, then any simple ring is a vector-space over its center.

[a]For those who know some Linear algebra and Fourier Analysis, the "diagonalization" of the derivative is the Fourier transform of a function! For details, see the Gelfand Transformation of unbounded operators

If $R = \mathbb{F}$ is a field, we give a special name to sub-modules:

Definition 12.1.8: Subspace

If $W \subseteq V$ is a submodule of a vector space, then we call it a *subspace*

33: Remark

(If you know linear algebra) When determining where a linear transformation maps, it is sufficient to know where the basis maps to. General R-modules need not have a basis well-defined on them. However, *cyclic* modules will always have a basis, since $M = R \cdot x$ for some $x \in M$. Thus, R is always a cyclic R-module over itself since $R \cdot 1 = R$. General R-modules need not have a basis, and so R as an R-module a particularly nice module.
As a consequence of this, for any R-module homomorphism: $\varphi : R \to N$, it suffices to know where $\varphi(1)$ maps to. In general, it is much harder to determine a map $\varphi : M \to N$ given where only certain elements are mapped to.

Exercise 12.1.1

1. Let $\langle m \rangle = M$ be an R-module generated by a single element (i.e. if $x \in M$, $x = a_i m$ for some $a_i \in R$)[2]. Show that $\operatorname{Ann}(M) := \{m \in M \mid \exists r \in R \text{ such that } rm = 0\}$ forms an ideal of R.
2. Let M be as in the previous question. Show that $M/\operatorname{Ann}(M) \cong HERE$ (I.e. trying to show how we can study ideals in the context of modules, and how ideals are studied as R/I R-modules
3. (practice with categories). Let A be a set and $F(A)$ be the group freely generated by A (definition 2.3.19).
    1. Prove that the collection of all vector-spaces along with linear transformations form a category. Label it k-**Vect**.
    2. Show that $F(A)$ can have a vector-space structure over k put on it such that $F : \mathbf{Set} \to k\text{-}\mathbf{Vect}$ is a functor. (Hint: See the functor defined in section 2.3.5)
4. I feel like I should add this somehow link to Ab-enriched categories
5. What is the initial object of Mod_R (if any)?

[2]We'll explore module generation in section 12.3

12.1.2 Algebras

Recall that if S is a ring, S is automatically an abelian group. Given any ring homomorphism $\varphi : R \to S$, we can always define an R-module structure on S by defining

$$r \cdot s = \varphi(r)s$$

that is, the action is multiplication in S. The axioms of an R-module is immediate given the axioms of a ring and that φ is a homomorphism. As a simple example, let $\mathrm{id}_R : R \to R$. Then this is the R-module R over itself with the action being multiplication.

Notice that S is an R-module which also has a concept of multiplication of it's elements (since S is a ring). Since we can multiply, we may run into situations like the following: If $s_1 = r_1 \cdot x$ and $s_2 = r_2 \cdot y$, then:

$$s_1 s_2 = (r_1 \cdot x)(r_2 \cdot y)$$

Such an element is not necessarily equal to $(r_1 r_2) \cdot xy$ (similarly to how with rings, if the underlying set is not abelian, distributivity would not hold). This would be desirable, as in we would want $\varphi : R \to S$ to be a ring homomorpism that was also compatible with the R-module structure in the following way:

$$(r_1 \cdot x)(r_2 \cdot y) = (r_1 r_2) \cdot xy$$

Getting back to the homomorphism $\varphi : R \to S$ and the R-module structure on S, if we further require that R be *commutative* and that $\varphi(R) \subseteq Z(S)$ (the multiplicative center), then the left and right R-module structure will coincide (i.e. left and right multiplication produce the same module structure), and the ring operation of S will be compatible with that of R, that is

$$(r_1 r_2) \cdot (s_1 s_2) = \varphi(r_1 r_2)(s_1 s_2) = \varphi(r_1)\varphi(r_2)(s_1 s_2) = \varphi(r_1)s_1\varphi(r_2)s_2 = (r_1 \cdot s_1)(r_2 \cdot s_2)$$

Equivalently, we can axiomatize these properties (similarly to how we axiomatize group-actions or modules):

Definition 12.1.9: R-algebra

Let S be a ring, R a commutative ring. Then S is called an (associative) R-algebra if S is also an R-module satisfying:

$$r \cdot (xy) = (r \cdot x)y = x(r \cdot y)$$

Equivalently, S is an R-algebra if theer exists ring homomorphism $\varphi : R \to S$ where $\varphi(R) \subseteq Z(S)$. Furthermore:

1. If S is a division ring, then the R-algebra S is called a *division algebra*
2. If S is commutative, then the R-aglebras S is called a *commutative algebra*

Note that S has has a natural R-module structure by forgetting the multiplication so that the action

$$r \cdot a[= a \cdot r] = \varphi(r)a$$

is the R-module structure. Unless otherise stated, this will be the assumed R-module structure of an R-algebra S.

We added the word "associative" in the definition of an algebra, since there is a notion of an algebra that is not associative: some author's say the object in definition 12.1.9 an *associative algebra*, and define algebras more generally as follows: Take an R-module M, and equip it with a binary operation such that the binary operation is left and right distributive (for all $a, b, c \in M$, $a \cdot (b+c) = a \cdot b + a \cdot c$, an $(b+c) \cdot a = b \cdot a + c \cdot a$), and that it is compatible with scalars (for all $a, b \in R$ and $x, y \in M$, $(ab) \cdot (xy) = (a \cdot x)(b \cdot y)$), then M is called an R-algebra. If the binary operation is associative and an identity exists, then it would be a ring operation it will be called an associative algebra.

There is a good reason adding the prefix "associative" in the greater context of mathematics: a large domain of study in mathematics is the study of *lie groups* and *lie algebras*, which focus on understanding the structure of smooth manifolds. Notably, Lie algebras are *not associative*, satisfying instead a relation called the *Jacobi identity*. However, since they are non-associative, they are not *rings*. They are thus quite different objects from R-algebras as we've defined them, and so don't encompass our current discussion. Their study take up many books (references to are HERE) which the reader currently has the algebraic background to understand[3] (actually, I think more lin.alg is really needed). For the curious, we will briefly discuss Lie Algebra in chapter 25.

From now on, when we say algebra or R-algebra, we will mean an associative algebra as defined in definition 12.1.9. Algebra homomorphisms are defined similarly to how they've been defined before (can you figure it out without looking at the following definition?)

Definition 12.1.10: R-Algebra Homomorphism and Isomorphism

If A and B are R-algebra, then $\varphi : R \to S$ is an R-algebra homomorphism if φ is a ring-homomorphism and

$$r \cdot \varphi(x) = \varphi(r \cdot x)$$

If there exists an inverse function φ that is also an R-algebra homomrophism, then A and B are *algebra isomorphic*, or just *isomoprhic* to each other, and we write $A \cong B$ if the context permits.

With this definition, the collection of R-algebra's form the category R-alg. For those who are categorically inclined, another way of saying that an R-algebra adds the notion of multiplication to an R-module in a "consistent" way is by noticing that R-algebra are the monoidal objects in the category of R-mod: Since all rings are $\mathbb{Z}$-modules, the discussion we've done in section 9.7 shows that rings are the monoidal objects in $\mathbb{Z}$-Mod (Hence, the category $\mathbb{Z}$-Alg is the category **Ring**). It can be checked that R-algebra's are the monoidal objects in R-Mod.

If $R = k$ is a field, then using the terminology of vector-spaces, a k-algebra would be defining a way of multiplying vectors. It is non-trivial to always define a multiplication structure on an R-module (I suggest you try it on $\mathbb{R}^n$!). Even harder is to ask whether there is a "natural" way of creating a multipliction structure, or more broadly to make an R-module M big enough so that we can define an R-algebra in a "natural" way[4]. We will in fact define such a notion in section 15.4.

Example 12.1.4: R-Algebras

1. Let R be any ring. Then R is a $\mathbb{Z}$-algebra. Why is that? Notice the similarity between this and the fact that every abelian group is a $\mathbb{Z}$-module. Naturally, any R-algebra S also always has the trivial structure of $s \cdot r = r$.

[3] Though a good basis in differential geometry would go a long way for intuition

[4] natural meaning it satisfies the appropriate universal property

2. If S is commutative, then any subring of $R \subseteq S$ will define R-algebra on S (take $\iota : R \hookrightarrow S$). If S is not commutative, any subring of the center will define an R-algebra.

3. Let $R[x_1, x_2, ..., x_n]$ be the polynomial ring over several variables. Then it is an R-algebra through $r \cdot p(x) = rp(x)$.

4. Let $k[G]$ be a group ring over the field k. Then $k[G]$ is naturally a k-algebra.

5. If D is a divison ring, then $Z(D)$ is always a *field*. If $k = Z(D)$, then any division ring D has a k-algebra structure on it, which also implies it is a vector-space over k.

6. Let $R = k^n$ for some field k. Then R is a k-algebra, with multiplication done point-wise.

7. Let $R = M_n(k)$ over a field k. Then R is a k-algebra. More generally, If A is a k-algebra, then $M_n(A)$ is a k-algebra.

   Compare this example to the previous one, which we'll label $S = k^{n^2}$. As vector-spaces over k, $R \cong_k S$. However, as k-algebras, $R \ncong_{k\text{-}\mathbf{Alg}} S$, since the multiplication structure is different between the two (one has nilpotent elements, the other does not).

8. If A, B are k-algebra, so is $A \times B$. By the previous part, that implies

$$M_n(A) \times M_n(B)$$

   is also a k-algebra.

9. At the heart of real analysis is the $\mathbb{R}$-algebra $\mathbb{R}$, and at the heart of complex analysis is the $\mathbb{C}$-algebra $\mathbb{C}$. Note that $\mathbb{C}$ is also a $\mathbb{R}$-algebra.

10. Let $R = \mathbb{R}^2$. Then we know that $\mathbb{R}^2$ has point-wise multiplication, but we can also define the following:

$$\text{define complex number multiplication}$$

   then this also forms an $\mathbb{R}$-algebra. In fact, this is $\mathbb{R}$-algebra isomorphic to $\mathbb{C}$ (over $\mathbb{R}$)!

11. (If you know linear algebra) (example with defining multiplication in vector-spaces, Also thought the "Structure Coefficeints" section here was really cool! `https://en.wikipedia.org/wiki/Algebra_over_a_field`)

12. One might ask themselves how many $\mathbb{R}$-algebra's exist (i.e. classify all $\mathbb{R}$-algebras). A little more generally, one might ask how many finite-dimensional associatvive divison algebras there are, that is, algebra on which you can has division and asociativity, and the notion of finite-dimension is well-defined. It turns out that only $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$ exist, with dimension 1, 2, and 4! This is called *Frobenius' Theorem*. In this way, the real, compelx, and quaternions are kind of special. [a]

There are a large number of areas that algebras appear in, to name a notable few:

1. (is the cross product an R-algebra) Let define a $\mathbb{R}^3$-algebra with $(a, b, c) \cdot (x, y, z) =$

2. commutative algebras in algebraic geometry

3. Banach-Algebra's in real analysis

4. C^*-algebra or $*$-algebras more generally in functional analysis
5. Though Lie algebra's are not [associative] algebras, there are some interesting intersections between the two TBD

[a] https://math.stackexchange.com/questions/368456/frobenius-from-hurwitzs-theorem

Theorem 12.1.11: Universal Property of Free R-algebras

$R[A]$ is a free commutative R-algebra over a [finite set] A

Notice how this is similar to the universal property of commutative polynomial rings. In fact, that is a special case of this theorem by taking $R = \mathbb{Z}$

Proof **:**
Let S be any commutative R-algebra and $\alpha : A \to S$ be any set-function. Since S is an commutative R-algebra, it is accompanied by a function $f : R \to S$ defining its multiplication.

Since $A = \{a_1, a_2, ..., a_n\}$ is finite, $R[A] = R[x_1, x_2, ..., x_n]$. Similarly to how the universal property of polynomial rings was proven (theorem 9.2.3), we will extend $f : R \to S$ to $f_1 : R[x_1] \to S$ by mapping $f_1(x_1) = \alpha(a_1)$ and extending linearly. f_1 is still an R-algebra homomorphism since $f_1(n_0 + n_1x + n_2x^2 + \cdots + n_kx^k) = f(n_0) + f(n_1)\alpha(a_1) + f(n_2)\alpha(a_1)^2 + \cdots + f(n_k)\alpha(a_1)^k$, and Similarly for multiplying two polynomials (using the binomial theorem and both the additive and multiplicative linearity of f_1). We can again extend f_1 to f_2 by defining $f_2(a_2) = \alpha(a_2)$ so that $f_2 : (R[x_1])[x_2] \to S$ (usually written as $f_2 : R[x_1, x_2] \to S$) is an R-algebra homomorphism. Continuing inductively will get us an R-algebra homomorphism $F : R[x_1, x_2, ..., x_n] \to S$. Thus, picking $\iota : A \to R[x_1, x_2, ..., x_n]$ will show that $\alpha : A \to S$ clearly lifts uniquely to $F : R[A] \to S$, as we sought to show.

As a consequence of this, all R-algebras are quotient of polynomial rings!! That is, if S is a finitely generearted S algebra

$$R[x_1, x_2, ..., x_n] \twoheadrightarrow S$$

12.1.3 Abelian Groups as Modules

Some entire domains of study are done on particular modules; we have actually already studied a particular collection modules extensively! The most immediate example is that all rings are R-modules over themselves with the action defined by:

$$r \cdot x = rx$$

Then submodules are precisely the ideals of R, and so the study of all R-modules includes the study of the ring R.

In this section, we show another important classes of R-modules: let $R = \mathbb{Z}$, and let $M = A$ be *any* abelian group (finite or infinite), and write the operation of A as $+$. Make A into a $\mathbb{Z}$-module by defining the ring action as :

$$na = \begin{cases} a + a + \cdots + a \ (n \text{ times}) & \text{if } n > 0 \\ 0 & \text{if } n = 0 \\ -a - a - \cdots - a \ (n \text{ times}) & \text{if } n < 0 \end{cases}$$

Then M is a $\mathbb{Z}$-module. In fact, this ring action is the only possible ring action on M (i.e. abelian groups):

Proposition 12.1.12: Abelian Group iff $\mathbb{Z}$-module

Every abelian group is a $\mathbb{Z}$-module, and every $\mathbb{Z}$-module is an abelian group

Proof **:**
for $\Rightarrow$, we can always define the previous ring-action and so it's always a $\mathbb{Z}$-module

For $\Leftarrow$, it is simply a consequence of the existence of 1 and the distributivity property, i.e., the ring action $\mathbb{Z}$ performs on abelian groups is "stuck" since we must play out the consequence of the axioms.

By definition

$$n \cdot 1 = n$$

and therefore

$$n \cdot 2 = n \cdot (1 + 1) = n \cdot 1 + n \cdot 1 = n + n$$

and continuity inductively

$$n \cdot k = n \cdot (1 + \cdots + 1) = n + \cdots + n$$

k times. This result is exactly the ring action we defined earlier. And so, forced by the axioms of groups and modules, it must be that every $\mathbb{Z}$ action on an abelian foms this particular $\mathbb{Z}$-module.

This proposition has a very intersting consequence: for every R-module, $R \neq \mathbb{Z}$, then it always has at least one other module structure, namely the natural $\mathbb{Z}$-module structure! An intuitive way to remember this fact is that numbers as we have learnt them earlier in our life usually have a $\mathbb{Z}$-module structure ($\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$). The natrual numbers form an abelian monoid, so this intuition doesn't directly apply to $\mathbb{N}$, but they indirectly apply since the same proof works for ring-actions on monoids (TBD?).

This example can also make us think of the class of all abelian groups as a subclass of modules. Abelian groups are also a subclass of groups. However, not every group is a module (and vice versa). The following diagram I'm about to draw shouldn't be taken as a commutative diagram, but as an "analogy":

$$\begin{tikzcd} \textbf{Grp} & & R\textbf{-Mod} \\ & \textbf{Ab} \arrow[ul] \arrow[ur] & \end{tikzcd}$$

So in a sense, we are "switching gears" when studying modules. We are taking a different approach to the study of algebraic structures as compared to studying groups.

Exercise 12.1.2

1. Recall that for any $\alpha \in R$ for a ring R, $f_\alpha : R[x] \to R$ mapping $p(x) \mapsto p(a)$ is a ring-homomorphism. Is it an R-module homomorphism?

12.2 Homomorphism and Quotients of Modules

In this section we explore R-module homomorphisms and how the image of R-module homomorphisms can be represented by a particular quotient of the domain (as we've seen with groups and rings). As R-module homomorphsism are explored, they should be constrasted to rings in that they will have many nicer properties. They are in fact comparable to abelian groups; every subgroup of an abelian group is normal, and so too every submodule of a R-module will be "normal".

R-Module Homomorphisms

Definition 12.2.1: R-module homomorphism Terminology

Let R be a ring, M and N be R-module, and $\varphi : M \to N$ be a function. Then:

1. φ is called an *R-module homomorphism* if:

   (a) $\varphi(x + y) = \varphi(x) + \varphi(y)$ for all $x, y \in M$

   (b) $\varphi(r \cdot x) = r\varphi(x)$, for all $r \in R$, $m \in M$

   If furthermore there exists a φ^{-1} such that $\varphi \circ \varphi^{-1} = \varphi^{-1} \circ \varphi = \text{id}$ and φ^{-1} is an R-module homomorphism, then φ is called an *R-module isomorphism* and we write $M \cong_{R\text{-}\mathbf{Mod}} N$, or $M \cong N$ if it is unambiguous

   If $R = k$ is a field, then k-module homomorphisms are called *linear maps*, *linear functions*, or *linear transfomration*, and k-module isomorphisms are called *linear isomorphisms* (due to historical precedence)

2. Denote

   (a) $\ker(\varphi) = \{m \in M \mid \varphi(m) = 0\}$

   (b) $\text{im}(\varphi) = \varphi(M) = \{n \in N | \varphi(m) = n\}$

3. Define $\text{Hom}_R(M, N)$ be the set of all R-module homomorphisms from M into N, $\text{End}_R(M)$ the set of all R-module homomorphsism from M to itself, and $\text{Aut}_R(M)$ the set of all all R-module isomorhpisms from M to itself. A function $\varphi \in \text{End}_R(M)$ is called an *R-module enormophism*, and a function $\varphi \in \text{Aut}_R(M)$ is called an *R-module automorphism*.

Notice how it's key that the two structures be R-modules; If M is an R-module and N is an S-module, then $\varphi : M \to N$ is not an R-module or S-module homomrophsim unless either R has a compatible S-module structure of N has a compatible R-module structure that is preserved under φ (this is similar to how G-equivariant functions are betwen two G-sets)

Many of the properties demonstrated of group and ring homomorphisms apply for R-modules

Lemma 12.2.2: R-Module Homomorphism Equivalence

Let $\varphi : M \to N$ be a function. Then φ is an R-module homomorphism if and only if for every $r \in R$ and $x, y \in M$,

$$\varphi(x + ry) = \varphi(x) + r\varphi(y)$$

Proof **:**
exercise

Lemma 12.2.3: R-Module Isomorphism Equivalence

Let $\varphi : M \to N$ be an R-module homomorphism. Then φ is an R-module isomorphism if and only if φ is an R-module homomorphism and bijective

Proof **:**
Imitate the proof in proposition ref:HERE

Lemma 12.2.4: Composition Of R-Module Homomorphism/Isomorphism

Let M, N, L be R-modules. If $\varphi \in \mathrm{Hom}_R(L, M)$ and $\psi \in \mathrm{Hom}_R(M, N)$ then $\psi \circ \varphi \in \mathrm{Hom}_R(L, N)$

Proof **:**
exercise

Lemma 12.2.5: R-module Automorphism Group

Let $\mathrm{Aut}_R(M)$ be the such set of R-module automorphism on M. Then $(\mathrm{Aut}_R(M), \circ)$ forms a group under composition.

Proof **:**
exercise

Example 12.2.1: Module Homomorphisms

1. if a ring is an R-module over itself with the natural ring action, then R-module homomorphisms need not be ring homomorphisms (even from R to itself), and ring homomorphisms need not be R-module homomorphisms. For example, $\mathbb{Z}$ act on itself in the natural way, and let $\varphi(x) = 2x$. φ is not a ring homomorphism [a], but φ is an R-module homomorphism. For the example going the other way, $F[x]$ act on itself in the natural way, and let $\varphi : f(x) \mapsto f(x^2)$. Then φ is not an $F[x]$-module homomorphism [b], but it is a ring homomorphism.

2. Let R be a ring and $M = R^n$ be an R-module with the natural ring action. Then the projective map is naturally an R-module homomorphism.

3. If we take $\mathbb{R}$ to be an $\mathbb{R}$ module over $\mathbb{R}$, then $\mathbb{R}$ has no non-trivial submodules (since $\mathbb{R}$ is a field, it's only ideals are 1 and itself). We call such a module a simple or an irreducible module. Thus, the quotient structure of $\mathbb{R}$ as an $\mathbb{R}$-module is trivial. However, if we consider $\mathbb{R}$ as a $\mathbb{Q}$-module (which we can). Then $\mathbb{Q}$ is a submodule of $\mathbb{R}$ (since $\mathbb{Q}$ is clearly closed under the ring action of $\mathbb{Q}$)! $\mathbb{R}$ as a $\mathbb{Q}$-module will be an infinite-dimensional vector-space [c]. Therefore, it becomes meaningful in this situation to look at $\mathbb{R}/\mathbb{Q}$ under the $\mathbb{Q}$ action!!

4. It should be done as an exercice that if $\varphi : M \to N$ is an R-module homomorphism, then the image and kernel are both R-modules.

5. (p.346)

[a] $\varphi(xy) = 2xy \neq 2x2y = \varphi(x)\varphi(y)$
[b] $x^2 = \varphi(x) = \varphi(x \cdot 1) = x\varphi(1) = x$, a contradiction
[c] In fact, it will even be *uncountably* infinite-dimensional

Recall in remark HomGrpIfAbRmk that $(\mathrm{Hom}_{\mathbf{Set}}(A, G), +)$ (where $+$ is defined pointwise: $f + g(x) = f(x) + g(x)$) can be given a group structure if G is abelian, and is not a group if G was not abelian (i.e. it not important what the domain was for the structure of the hom-set). Since every R-module is always over an abelian group, $\mathrm{Hom}_R(M, N)$ is *always* an abelian group. In the following proposition we further the structure of $\mathrm{Hom}_R(M, N)$

Proposition 12.2.6: Properties of $\mathrm{Hom}_R(M, N)$

Let M, N and L be R-module. Then

1. $(\mathrm{Hom}_R(M, N), +)$ is an abelian group with $+$ being defined pointwise:

$$(\varphi + \psi)(m) = \varphi(m) + \psi(m) \qquad \forall \varphi, \psi \in \mathrm{Hom}_R(M, N),\ \forall m \in M$$

   In particular $\varphi + \psi \in \mathrm{Hom}_R(M, N)$

2. If R is a commutative ring, then for $r \in R$ define $r\varphi$ by

$$(r\varphi)(m) = r(\varphi(m)) \qquad \text{for all } m \in M$$

   Then $r\varphi \in \mathrm{Hom}_R(M, N)$, and with this action $\mathrm{Hom}_R(M, N)$ is an R-module.

3. With addition defined as above, $(\mathrm{End}_R(M), +, \circ)$ is a ring. When R is commutative, $\mathrm{End}_R(M)$ is an R-algebra.

***Proof* :**
The first statement was proven above remark HomGrpIfAbRmk.

For an R-action to be defined on $\mathrm{Hom}_R(M, N)$ we would need the following to be equal:

$$(rs\varphi)(m) = rs \cdot \varphi(m)$$
$$(r(s\varphi))(m) = r \cdot (s\varphi)(m) = sr \cdot \varphi(m)$$

showing R must be commutative. The other axioms follow quickly

Finally, $(\mathrm{End}_R(M), +, \circ)$ follow's the ring axioms. If R is commutative then for ring homomorhpism $f : R \to \mathrm{End}_R(M)$, since f is commutative, $f(R)$ is always abelian, and hence $f(R) \subseteq Z(\mathrm{End}_R(M))$, and so $\mathrm{End}_R(M)$ is always an R-algebra, completing the proof.

Note that $(\mathrm{Aut}_R(M), +)$ is *not* a group structure, since the identity *does not exist* (the zero homomorphism is not an automorphism)![5]

Like with groups and rings, we want to know what conditions are needed on a sub-module so that:

$$(x + N) + (y + N) = (x + y) + N \qquad r(x + N) = (rx) + N$$

are both well defined. Since Modules are already abelian groups, than M/N is already well-defined as an abelian group! What's left is to see what condition's are needed so that a submodule N can be used to define an R-module structure on the abelian group M/N. It turns out that like abelian groups, every sub-module is "normal"!

Proposition 12.2.7: sub-modules are "normal" sub-modules

Let R be a ring and let M be an R-module and let N be a submodule of M. The (additive, abeliean) quotient group M/N can be made into an R-module by definin an action of elements of R by

$$r(x + N) = (rx) + N, \qquad \text{for all } r \in R,\ x + N \in M/N$$

The natural projection $\pi : M \to M/N$ defined by $\pi(x) = x+N$ is an R-module homomorphism with kernel N.

Proof **:**
First, M/N is already an abelian quotient group. We want the ring action to be well defined, that is

$$x + N = y + N$$

which equivalently means that $x - y \in N$. Since N is a submodule, $r(x - y) \in N$, thus

$$rx + N = ry + N$$

therefore, the R-action on M naturally induces an action on M/N. The rest of the module axioms can now e easily verified.

Next, we must show the implication goes the other way, that is, if the elements of the quotient groups are formed by the cosets $x + N$, then N is the kernel of some R-module homomorphis (i.e. the natural projection). The natural projection map π described above is already the natural projection of the abelian group M into the abelian group M/N. It only remains to show that it is a R-module homomorphism:

$$\begin{aligned} \pi(rm) &= rm + N \\ &= r(m + N) && \text{(by definition of the action of } R \text{ on } M/N) \\ &= r\pi(m) \end{aligned}$$

[5]semi-groups are groups without an identity. Thus, $(\mathrm{Aut}_R(M)+)$ is a natural example of a semi-group(!), with a compatible group structure with $\circ$

Thus, the natural projection an R-module homomorphism, completing the proof.

As with groups and rings, R-modules also have the isomorphism theorems. Since R-modules are already abelian groups, half the proof is already done. What is left to check would be if the action still preserved under the isomorphisms. To define the 2nd isomorphism theorem, we define what it means to take sum of two modules:

Definition 12.2.8: Sum Of Modules

Let A, B be submodules of an R-module M. Then the *sum* of A and B is the set

$$A + B = \{a + b | a \in A, b \in B\}$$

As before, $A + B$ is the smallest module containing A and B (as you sould check)

Theorem 12.2.9: Isomorphism Theorems For Modules

1. Let M, N be R-modules and let $\varphi : M \to N$ be an R-module homomorphism. Then $\ker(\varphi)$ is a submodule of M and

$$M/\ker(\varphi) \cong \varphi(M)$$

2. Let A, B be submodules of the R-module M. Then

$$\frac{A+B}{B} \cong \frac{A}{A \cap B}$$

3. Let M be an R-module, and let A and B be submodules of M, with $A \subseteq B$. Then

$$(M/A)(B/A) \cong M/B$$

4. Let N be a submodule of the R-module M. There is a bijection beetween the submodules of M which contain N and the submodules of M/N. The correspondence is give nby $A \leftrightarrow A/n$, an for all $A \supseteq N$. The correspondence commutes with the processes of taking sums and intersection (i.e., is a lattice isomorphism between the lattice of submodules of M/N an dthe lattice of submodules of M which contain N

Proof **:**
Very similar to previous isomorphism proofs.

12.2.1 Extensions and Exactness

We conclude this section with a quick word on extensions and where they will show up. Similarly to groups (and rings), we may wonder under what conditions is a module M an extension of two modules A and B, that is $M \cong B/A$, or in terms of a short exact sequence:

$$0 \to A \to M \to B \to 0$$

In section 3.7, we showed that mathematicians managed to find the building blocks of finite group extensions, known as simple groups, and classify all finite simple groups, and in section 5.4, we showed that semi-direct products can be represeneted by split exact sequences. The classification of simple modules depends on the ring that's acting, R, and is much simpler since the underlying group is abelian. Similar to groups, there exist classification project for R-modules for some particular rings R, at least four of which will be covered in this book[6].

We also endeavour to study R-modules in which the extension property works nicely. This is further explored in part VI, chapter 23. Since all modules are abelian groups, the semi-direct product is simply the direct product. However, split exact sequences are still an important concept since it gives us a way of recognizing direct products using homomorphisms. A particularly interesting collection of modules are those for which every short exact sequence is a split exact sequence, or phrased differently that every short exact sequence splits. This endevaour is taken up in chapter 16, and explored in full length in part VII.

The vocabulary for a sequence of R-module homomorphisms that represent extensions is quite important, and so we define the term here for future use:

Definition 12.2.10: Exact Chain Complex

Let $M_0, M_1, M_2, ..., M_n, ...$ is a collection of R-modules $d_i : M_i \to M_{i-1}$ be a collection of R-module homomorphisms where

$$\cdots \xleftarrow{d_0} M_0 \xleftarrow{d_1} M_1 \xleftarrow{d_2} \cdots \xleftarrow{d_n} M_n \xleftarrow{d_{n+1}} \cdots$$

and $\ker(d_i) = \operatorname{im}(d_{i-1})$. Then the tuple $(M_i, d_i)_{i=-\infty}^{\infty}$ is called an *exact chain complex*, *exact chain*, or *exact sequence*. Often, we abbreviate and write $(M_\bullet, d_\bullet)$. If the arrows are pointing the opposite direction and $\ker(d_i) = \operatorname{im}(d_{i+1})$ then the resulting tuple $(M_i, d_i)_{i=-\infty}^{\infty}$ is called an *exact cochain*. If the chain has only finite length, (i.e. only finitely many of the modules are non-zero), then the resulting exact (co)chain is called a *finite exact (co)chain*.

(examples and some explanation here, important because it comes back in at least the next two chaptesr)

Exercise 12.2.1

1. Let M be an R-module. Prove that $\operatorname{Hom}_R(R, M) \cong M$
2. Let R^{op} be the opposite ring of R (see exercise ref:HERE), and let $\operatorname{End}_R(R)$ be all R-module homomorphisms from R to itself. Considering $(\operatorname{End}_R(R), +, \circ)$ as a ring show that $R^{\text{op}} \cong_{\mathbf{Ring}} \operatorname{End}_R(R)$ [Hint ; think of Yoneda's lemma]

12.3 Generation of Modules

We now extend the notion of generating algebraic structures to subsets of a module. We will start by generalizing the results of section 2.3.2: we take a subset of an R-module, and permit all possible

[6]When R is a Field, a Principal Ideal Domain, a Dedekind Domain, and Semi-Simple

combinations and R-actions of those elements to form an R-module, or equivaelntly we take the intersection of all R-submodules containing the elements of the set.

Then, as with groups and rings, the easiest way of creating a larger module from smaller ones is to take the direct sum or direct product. Due to their simplicity, it is worth taking the time to find out under what condition's can a module be decomposed into direct products or sums (just like for groups in section 5.2). We then take the approaching of trying to define the "most free" module we can put on an arbitrary set, similar to what we've done in section 2.3.4, that is we are looking to find the universal property of free-modules. As compared to groups, we will spend more time looking at minimal generating sets. Recall that for groups, it is possible that minimal generating sets can differ in cardinatliy. With modules, we will ask if there are modules where there exists a subset $A \subseteq M$ such that A generates M and if any other subset B generated A, then $|A| \leq |B|$. It will turn out that some modules have this property, some will be close, and other will not have this property at all.

After having tackled the question of generating modules "freely" and how to recognize free modules, we will then turn our attention to finitely generated modules, and modules where all submodules are finitely generated. Similalrly to how we can classify all finitely generated abelian groups, finitely generated and Noetherian modules have more "regularly", or equivalently have less "pathological" properties.

A new approach to studying structure's of objects we will take is considering the collection of homomorphisms on modules. As we saw in proposition 12.2.6, $\mathrm{Hom}_R(M, N)$ is always an abelian group, and if R is commutative it is always R-module. Thus, the "symmetries" of an R-module are themselves and R-module, meaning we would like to explore what properties M or N can have given the properties of $\mathrm{Hom}_R(M, N)$ have as an R-module, or even better that $\mathrm{End}_R(M)$ has as an R-algebra. It turns out that this exploration is very fruitful, and will be of key importance in the theory of representation (of finite groups, lie algebras, etc.). In fact, if this intrigues the reader, after reading chapter 13 and chapter 14, the reader can skip to part VI and read chapter 23 (only requiring knowledge from section 17.5 for the last 2 propositions of that chapter). A particular example of the above is when $N = R$ so that we consider $\mathrm{Hom}_R(M, R)$. This module is called the *dual module* of M, and will be very important in the study of M, even outside the field of algebra[7].

How does the generation of modules compare to the generation of groups and rings? With general groups, we saw that it was very difficult to show what the elements in an arbitrarily generated group look like (it is considered one of the hardest problems in mathematics, known as the *word problem for groups*). However, if G is abelian and finitely generated, then every element can be written in a very simple form due to the the fundamental theorem of finitely generated abelian groups (theorem 5.1.5) With rings, we had that the objects that we quotient by (ideals) were no longer rings, and so took the time to study generation of ideals, but we have not studied any structure theorems on rings.

With modules, the story of the exploration of module structures is much closer to that of abelian groups, namely because every module *is* an abelian group. One major advantage of studying module is that many of the previous structure's we've studied can be represented as modules: since all abelian groups are $\mathbb{Z}$-modules, the study of the structure of abelian group is subsumed in the study of the structure of modules (we prove the Fundamental Theroem of Finitely Generated Abelian Groups in greater generality in chapter 14), and since all rings are module over themselves, the study of the structure of modules is also the study of the structure of rings! Due to the fact that all module are abelian groups, we can always collect "like terms":

$$r_1a_1 + r_2a_2 + s_1a_1 = (r_1 + s_1)a_1 + a_2$$

[7] dual modules are central to Functional analysis and Harmonic Analysis

due to this, the order of the generating elements does not affect their sum! It becomes meaningful to ask if the ring acting on a module somehow determines some (or even all!) of the structure of the module. This question is a very meaningful question, and the answer will vary depending on what type of ring is acting on the module: chapters 13, 14, and 24 will each focus modules over particular types of rings and show how the answer to this question changes depending on what type of ring is acting on the module.

12.3.1 Direct Sum of Modules

We first update the terminology from the generation of groups and rings for the generation of modules:

Definition 12.3.1: Generating Modules

Let M be an left R-module and let $(N_i)_{i\in I}$ be the submodules of M where I is an indexing set.

1. The *sum* of $(N_i)_{i\in I}$ is the smallest submodule of M containing each N_i, and is the set of all finite sums of elements from N_i:

$$\sum_i N_i := \{a_1 + \cdots + a_k | a_i \in N_i \text{ for all } i\}$$

If I is finite with $|I| = k$, then it is usually written as

$$N_1 + \cdots + N_k$$

2. For any subset $A \subseteq M$, then RA is the set:

$$RA = \left\{r_1a_1 + \cdots + r_ma_m | r_1, ..., r_m \in R, a_1, ..., a_m \in A, m \in \mathbb{Z}^+\right\}$$

(where by convention $RA = \{0\}$ if $A = \emptyset$) is the smallest [left] submodule of M generated by the elements of A, an is called the *[left] submodule generated by* A. If A is finite the set $\{a_1, a_2, ..., a_n\}$ we shall write

$$Ra_1 + \cdots Ra_n$$

If N is a submodule of M (possibly $N = M$) and $N = RA$, for some subset, then A is called the *set of generators* or *generating set for* N, and we say N is generated by A and is sometimes written as $\langle A\rangle = N$.

3. A submodule N of M (possibly $N = M$) is *finitely generated* if there is some finite subset A of M such that $N = RA$.

4. A submodule N of M (possibly $N = M$) is *cyclic* if there exists an element $a \in M$ such that $N = Ra$, that is, if N is generated by one element

$$N = Ra = \{ra|\ r \in R\}$$

and is sometimes written as $\langle a\rangle = N$.

It is an easy starter exercise to show that the structures in definition 12.3.1 are indeed modules. If R is a *rng*, then it is not necessarily the case that $A \subseteq RA$.

Example 12.3.1: Example

1. Let $R = \mathbb{Z}$ and let M be any $\mathbb{Z}$-module. Take any $a \in M$. Then $\mathbb{Z}a$ is a cyclic sub-module. It is in fact a cyclic subgroup! More generally, A generates M as a $\mathbb{Z}$-module if and only if A generates M as an [abelian] group. That is, generating M through A as a group (using $+$ and $-$ in groups) will produce as many elements as generating M through A as a module (which also allows $\mathbb{Z}$-actions). Thus, the definition of a finitely generated $\mathbb{Z}$-module is identical to the definition of a finitely generated group (definition 5.1.3).

2. Let R be a ring and let $M = R$ be the [left] R-module of R acting on itself. Then R is finitely generated as a module, even if R is infinitely generated as a ring! Even better: it is cyclic, since $R \cdot 1 = R$. Thus, in terms of generation, R is the "simplest" R-module.

3. If R is an R-module over itslef, then $I \subseteq R$ is a finitely generated R-module if and only if it is a finitely generated *left-ideal*. If R was a right-module, then I would be a finitely generated *right-ideal*, and if R was commutative, I would be a finitely generated *two-sided ideal*.

4. Let M be a finitely generated R-module. Then it is not the case that every submodule of M is finitely generated! Take $P[x_1, x_2, ...] = P[X]$ (i.e. the polynomial ring over countably many coefficients) which is a $P[X]$-module over itself. Then $P[X]$ is cyclic as a $P[X]$-module, and hence finitely generaed, however $(x_2, x_4, x_6, ...) \subseteq P[X]$ is a $P[X]$-submodule of $P[X]$ (being a left-ideal), however, it is an *infinitely generated* left-ideal. Finitely generated modules in which all submodules are finitely generated will be called *Noetherian*.

5. Let R^k be an R-module with the action being pointwise multiplication. Then R^k is finitely generaed by
$$e_i = (\underbrace{0, ..., 0}_{i-1 \text{ times}}, 1, \underbrace{0, ..., 0}_{k-i \text{ times}})$$
hence, for any $(r_1, r_2, ..., r_k) \in R^k$,
$$(r_1, ..., r_k) = \sum_i r_i e_i$$

6. Here is a neat way of writing down M if $A = M/B$. Then $M \cong A' + B$, since any element in M is some element of A (namely A' is a set of coset representative when taking $b = 0$), and an element of B:
$$\frac{M}{B} = \frac{A' + B}{B} \cong A$$
Note that is in general smmimply a *subset* of M, we can define the operation on A' by taking:
$$a_1 + a_2 = a_3 \qquad a_3 \text{ is the appropriate representative from } [a_1 + a_2] \text{ that belongs in } A'$$
In this way, $A' \not\subseteq M$ for as an R-module since the binary operation is different. Note too that $M = A' + B$ does not imply that $M \cong A \oplus B$. For example, consider $A = \mathbb{Z}/6\mathbb{Z}$ with $M = \mathbb{Z}$, and $B = 6\mathbb{Z}$. Letting $A' = \{0, 1, 2, 3, 4, 5\}$, we get that as sets:
$$M = A' + B = \{0, 1, 2, 3, 4, 5\} + 6\mathbb{Z} = \mathbb{Z}$$

for example, $10 = 4 + 6$, and $101 = 5 + 96$. However, giving A' the group structure of $\mathbb{Z}/6\mathbb{Z}$:

$$\mathbb{Z} \ncong_R (\mathbb{Z}/6\mathbb{Z}) + 6\mathbb{Z}$$

For the right hand side has torsion elements, while the left does not. For the $+$ to be upgraded to a $\oplus$, the operation of the quotient must in some natural way be compatible with the operation in M, that is A must in not only be the quotient M/B, but must also be considered as $A \subseteq M$. This is explore in chapter 16.

Naturally, we would want to find when direct sums of modules behave like direct products, or more specifically external direct sums

Definition 12.3.2: Direct Product and Sum of Modules

Let $(M_i)_{i\in I}$ be a collection of R-modules where I is an indexing set. Then the collection of tuples $(m_i)_{i\in I}$ where $m_i \in M_i$ with addition and action of R defined componentwise is called the *direct product* of $(M_i)_{i\in I}$, denoted

$$\prod_{i\in I} M_i$$

If I is finite where $|I| = n$, then it is usually denoted:

$$M_1 \times \cdots \times M_k$$

If we take the subset of $\prod_{i\in I} M_i$ where only finitely many of the terms are nonzero, then this this set is called the *external direct sum*, and is written as:

$$\bigoplus_{i\in I} M_i$$

If I is finite where $|I| = n$, then it is usually denoted:

$$M_1 \oplus M_2 \oplus \cdots \oplus M_n$$

If I is infinite and $R \neq 0$, then $\bigoplus_{i\in I} M_i \subsetneq \prod_{i\in I} M_i$.

34: Remark

It might be tempting to say that:

$$M \cong N_1 \oplus N_2 \quad \Leftrightarrow \quad M = N_1 + N_2$$

however that is not the case: Take $\mathbb{Z}$ as a $\mathbb{Z}$-module over itself, $N_1 = 2\mathbb{Z}$ an $N_2 = (0)$. Then $2\mathbb{Z} + 0 = 2\mathbb{Z}$, and:

$$\mathbb{Z} \cong 2\mathbb{Z} \qquad \text{but} \qquad \mathbb{Z} \neq 2\mathbb{Z}$$

which we can see since, $1 \notin 2\mathbb{Z}$

If $M \cong A \times B$ for some modules M, A, and B, then for any $m \in M$, there exist unique pair of elements $a \in A, b \in B$ such that $m \mapsto (a, 0) + (0, b)$. Thus, any generating set (or more specifically minimal generating set) of M breaks down to a question of a generating set (or minimal generating

gset) of A and B. The key here is that any element $m \in M$ can be uniquely represented by elements a and b (namely since these two modules are isomorphic).

Let's say conversely that we want to find submodules $A, B \subseteq M$ for which $M = A + B$. If a module M is finitely generated, by Zorn's lemma there must exist a minimal nonnegative integer d such that M is generated by d elements (and no less)[8]. If M was not finitely generated, this need not be the case. If M can be generated by d elements, it need not be the case that there is a unique set of generating elements, nor the case that the R-linear combination of these elements be unique:

Example 12.3.2: Non-Unique Sets And R-Linear Combinations
here

It will turn out that we almost never get a unique minimal generating set [a]. Take for example R over itself. Let $u \in R$ be a unit of R. Since u is a unit, let $v \in R$ such that $uv = 1$. Then $R \cdot 1 = R \cdot (vu) = (Rv)u = R \cdot u$ (since $u \in R$, $uv = 1 \in Rv$, and from that we can show that $(Rv)u = Ru$).

[a] If you have already taken a course in linear algebra, this is equivalent to saying that there is no unique basis

Though unique minimal generated sets are rare to come accorss, uniqe R-linear combinations are much more common. For example, taking R over itself so that $\{1\}$ generated R as an R-module. Then every R-linear combination is unique, that is

$$r \cdot 1 = s \cdot 1 \Leftrightarrow r = s$$

We can see this by taking the homomorphism $\varphi : R \to M$, $r \mapsto r \cdot 1$ and look at the kernel. This property is very desirable, since it will allow us to show that M can split into a direct sum of submodules. If the R-linear combinations are unique, then we need only concern ourselves with any minimal generating set when considering the structure of such a module!

For the rest of the chapter, we identify equivalent properties to saying R-linear combinations are unique, and identify modules which have this property.

[8] Note that minimal doesn't mean smallest! There can be multiple different $d \in \mathbb{N}$ that satisfy this property, see the discussion in section 2.3

Proposition 12.3.3: Equivalent Conditions for Linear Combinations

Let $N_1, ..., N_k$ be (not necessarily distinct) submodules of R-module M. Then the following are equivelent

1. the map $\pi : \bigoplus_{i \in I} N_i \to \sum_i N_i$ defined by

$$(n_i)_{i \in I} \mapsto \sum_i n_i$$

is an R-module isomorphim. If $|I| < k$ is finite, we write:

$$N_1 + N_2 + \cdots + N_k \cong N_1 \times N_2 \times \cdots \times N_k$$

2. $N_j \cap (N_1 + N_2 + \cdots + N_{j-1} + N_{j+1} + \cdots + N_k) = 0$ for all $j \in \{1, 2, ..., k\}$
3. Every $x \in N_1 + N_2 + \cdots + N_k$ can be written *uniquely* in the form $a_1 + a_2 + \cdots + a_k$ with $a_i \in N_i$

***Proof* :**
For (1) implies (2). let's assume that the intersection is not empty. Then there exists some a_j such that

$$a_j = a_1 + a_2 + \cdots + a_{j-1} + a_{j+1} + \cdots + a_k$$

However, this implies that

$$\pi(0, ..., a_j, ..., 0) = \pi(a_1, a_2, ..., a_{j-1}, a_{j+1}, ..., a_k)$$

but

$$(0, ..., a_j, ..., 0) \neq (a_1, a_2, ..., a_{j-1}, a_{j+1}, ..., a_k)$$

contradiction injectivy, since π is an isomorphism

For (2) implies (3). Let's say there are two linear combinations

$$a_1 + a_2 + \cdots + a_k = b_1 + b_2 + \cdots + b_k$$

then we can isolate any jth components:

$$b_j - a_j = (b_1 - a_1) + \cdots + (b_{j-1} - a_{j-1}) + (b_{j+1} - a_{j+1}) + \cdots + (b_k - a_k)$$

which implies by our previous proposition

$$b_j - a_j \in N_j \cap (N_1 + N_2 + \cdots + N_{j-1} + N_{j+1} + \cdots + N_k) = 0$$

thus, $b_j - a_j = 0$ and so $b_j = a_j$ for every j, showing that the linear combination is unique!

For (3) implies (1), Notice that the function is surjective construction: for any element $a_1 + a_2 + \cdots + a_k$, the element $(a_1, a_2, ..., a_k)$ would map to it. What (2) does is show injectivity, since if two sums are the same, then the direct product of elements must also be the same!. This map is

also easily checked to be an R-module homomorphism. For example

$$\begin{aligned}\pi(r \cdot (a_1, a_2, ..., a_k) &= \pi(ra_1, ra_2, ..., ra_k) && (12.1)\\ &= ra_1 + ra_2 + \cdots + ra_k && (12.2)\\ &= r \cdot (a_1 + a_2 + \cdots + a_k) && (12.3)\\ &= r\pi(a_1, a_2, ..., a_k) && (12.4)\end{aligned}$$

where line 3 comes from induction on $r \cdot (m + n) = r \cdot m + r \cdot n$, and linearity comes easily too.

This show's us how we may split up a module into "direct products" which satisfy the universal property of direct products. We now introduce a notion that is slightly stronger than the external direct product and of sums of modules:

Definition 12.3.4: Internal Direct Sum

Let $(N_i)_{i \in I}$ be a collection of submodules of M modules respecting one of the 3 condition's in proposition 12.3.3. If:

$$\sum_i N_i = M$$

then we will replace the summation symbol $\sum_i$ with $\bigoplus_i$. If the indexing set is finite, then:

$$N_1 + \cdots + N_k = N_1 \oplus \cdots \oplus N_k$$

Thus, we are *overloading* the notation of $\oplus$ to represent the fact that the sum $\sum_i N_i$ is isomorphic to $\oplus_i N_i$ *and* that $\sum_i N_i = M$

We have already shown in remark where M is an R-module that is isomorphic to a direct sum of submodules $(N_i)_{i \in I}$ but $\sum_i N_i \neq M$, namely by taking $\mathbb{Z}$ as a $\mathbb{Z}$-module and $N_1 = 2\mathbb{Z}$, $N_2 = (0)$. Then:

$$\mathbb{Z} \cong 2\mathbb{Z} \oplus (0) \qquad \mathbb{Z} \neq 2\mathbb{Z} \oplus (0)$$

In chapter 16 and 23, we will further explore how to find when a module M can be decomposed into direct products.

To conclude, we re-phrase the above results in the language of universal properties. The properties of $M \oplus N$ can be described in terms of maps $\varphi : L \to M \oplus N$ where this map is uniquely determined by the maps $\varphi : L \to M$ and $\varphi : L \to N$, that is:

Proposition 12.3.5: Universal Property of Products, Modules

Let M, N be two R-modules. Then $M \oplus N$ as constructed above satisfies the universal property products: for any R-module L, if $\varphi :_M L \to M$ and $\varphi :_N L \to N$ are two R-module homomorphisms, then there exists a unique R-module homomorphism $\varphi : L \to M \oplus N$ such that the following diagram commutes:

$$\begin{tikzcd}
 & L \arrow[d, dashed, "\varphi"] \arrow[ddl, bend right, "\varphi_M"'] \arrow[ddr, bend left, "\varphi_N"] & \\
 & M \oplus N \arrow[dl, "\pi_M"'] \arrow[dr, "\pi_N"] & \\
M & & N
\end{tikzcd}$$

More generally, if M_i is a collection of R-modules for an indexing set I, for any R-module L and collection of R-module homomorphisms $\varphi_i : L \to M_i$, there exists a unique R-module homomorphism $\varphi : L \to \oplus_{i \in I} M_i$ such that all of the following diagrams commutes:

$$\begin{tikzcd}
L \arrow[d, "\varphi"'] \arrow[dr, "\varphi_i"] & \\
\oplus_{i \in I} M_i \arrow[r, "\pi_{M_i}"] & M_i
\end{tikzcd}$$

Proof **:**
The reader should do this exercise as a good work-through of how to show a universal property, and to be comfortable with commutative diagrams for proofs in later chapters.

12.3.2 Free Modules

Given a ring R and a an arbitrary abelian group A, we can always define the trivial module via the action $r \cdot a = 0$: let's label this module $T(A)$ and call it the trivial R-module. Just like trivial groups, trivial modules offer little in structure[9]. It would be desirable if we can find a better struture on A, in particular if A represents the generating set of some R-module M, then we would like to define a "free R-module" $F(A)$ (or more particularly $F^R(A)$ to remember the ring R) which represents an R-module with no relationship, and then find a R-submodule $N \subseteq F(A)$ such that $F(A)/N \cong M$ (i.e. the submodule N woule represent adding relationships to $F(A)$). This mirrors the development of the free group from section 2.3.3 and 3.2.

Let's say we were to define a R-module $F^R(A)$ on the generating set A of M with only the bare minimum relations to make it into an R-module ($1 \cdot m = m$, $(rs) \cdot m = r \cdot (s \cdot m)$, and so forth). Then similarly to free groups, we would want that for any element of $x \in F^R(A)$, which is some linear combination:

$$x = \sum_{a \in B \subseteq A} r_a \cdot a \quad \text{for some subset } B \subseteq A$$

[9]When dealing with groups, we also defined $\mathrm{Sym}(A)$. As a thought exercise, can you see why we cannot parallel the same idea with module? Hint: Think of $\mathbb{Z} \hookrightarrow \mathbb{Q}$

to be equal to 0 if and only if $r_a = 0$ for all a. This is similar to free groups, where any relation between generating elements can be move to the left hand side and the right hand side equals to the identity element. By proposition 12.3.3, this means that $F^R(A) = R^{|A|}$ is our desired free R-module. To see this more clearly, let's say that some $r_a \neq 0$. Then that means we have added some relation into $F^R(A)$, that is the function $\varphi : R^{|A|} \to F^R(A)$ would have a non-trivial kernel, going against the idea that $F^R(A)$ is a free-module.

Of course, if $F^R(A)$ is a free module on A, just like for free groups, any set-function from A to any R-module N, $f : A \to N$, should lift to an R-module homomorhpism $F : F^R(A) \to N$, i.e. there is a map $\iota : A \hookrightarrow F^R(A)$ that injects A into $F^R(A)$ and $f = F \circ \iota$. In particular, given any arbitrary map $f : A \to N$, no matter where f maps A, whether f is injective or not, we can find a map F such that $F|_A = f$, so F must map any element of the set A exactly as how f would map it, hence F is an extension of f that has the restriction of being an R-module homomorphism. Furthermore, N is any R-module, meaning any $f : A \to N$ can be lifted to $F : F^R(A) \to N$ (see how this is not the case for an arbitrary domain M, or check out section 2.3.3 for the parallel problem in group theory). A consequence of this result is that we can determine how $F^R(A)$ maps to an R-module N simply by determining where A maps to, that is we have the universal property of free modules.

35: On Recovering A

Of course, we would have to already know of the set A in order to do the above strategy. The notation $F^R(A)$ makes it seem obvious that we know about the set A, namely since we will soon construct $F^R(A)$ with respect to A, but it is not so obvious that if we are given a module M that we say is a free R-module, can we find a subset A such that $F^R(A) \cong_R M$. In fact, depending on the ring R acting on M, it will sometimes be possible to "recover" A, while other times it will be impossible[a]. We will explore this in chapter 13

[a] It will be possible to say there exists such a set using the axiom of choice, but it will be impossible to give a description of the elements

We now prove that $F^R(A) = R^{|A|}$ does indeed respect the universal property of free R-modules.

Theorem 12.3.6: Universal Property of Free Modules

For any set A, there exists a free R-module $F^R(A)$ on the set A where $F^R(A)$ satisfies the following *universal property*: If M is any R-module and $\varphi : A \to M$ is any map of sets, then there is a unique R-module homomorphism $\psi : F(A) \to M$ such that $\psi|_A = \varphi$, that is, the following diagram commutes

$$\begin{tikzcd} S \arrow[r, hook, "\text{incl.}"] \arrow[dr, "\varphi"'] & F(S) \arrow[d, "\psi"] \\ & M \end{tikzcd}$$

When A is a finite set $\{a_1, a_2, ..., a_n\}$, $F^R(A) = Ra_1 \oplus \cdots + \oplus RA_n \cong R^n$, and if A has any cardinality, then:

$$F^R(A) = \bigoplus_{a \in A} R_a$$

where A is treated as an indexing set.

***Proof* :**
if $A = \emptyset$, then let $F(A) = \{0\}$. Then this trivially satisfies:

$$\begin{tikzcd} \emptyset \arrow[r, hook, "\text{incl.}"] \arrow[dr, "\varphi"'] & \{0\} \arrow[d, "\Phi"] \\ & M \end{tikzcd}$$

since $\emptyset$ is included in any set, φ is the vacuous map, and $\Phi(0) = 0$ makes Φ into an R-module homomorphism which restircted to $\emptyset$ is itself again.

So let $A \neq \emptyset$. What we'll do is construct funtions with the domain as a set and the co-domain as the ring R. These functions will be the mechanism by which set A is given an R-module structure! Let $F^R(A)$ be the collection of all set functions $f : A \to R$ such that $f(a) = 0$ for all but finitely many $a \in A$:

$$F^R(A) = \{f \in \mathrm{Hom}_{\mathbf{Set}}(A, R) \mid f(a) \neq 0 \text{ for only finitely many } a \in A\}$$

Make $F^R(A)$ into an R-module by defining pointwise addition and pointwise multiplication for the action:

$$\begin{aligned} (f + a)(a) &= f(a) + g(a) \\ (rf)(a) &= r(f(a)) \end{aligned}$$

for all $a \in A$, $r \in R$, $f, g \in F^R(A)$. Notice how $F^R(A)$ is isomorphic to $R^|A|$, meaning this construction matches definition 12.3.7.

Let $\iota : A \to F^R(A)$ vis $a \mapsto f_a$, where f_a is the function which is 1 at a and zero everywhere else. We can thus identify a set $A \leftrightarrow A' \subseteq F^R(A)$ as the subset containing all of f_a. Notice that since any f has only finitely many non-zero elements, we can write any $f \in F^R(A)$ as

$$f = r_1 f_{a_1} + r_2 f_{a_2} + \cdots + r_n f_{a_n}$$

which by our identification of A' with A, represents:

$$r_1 a_1 + r_2 a_2 + \cdots + r_n a_n$$

where f takes on the value of r_i at a_i and is zero at all other elements of A. Even better, by construction, every element of $F^R(A)$ has a *unique* expression as such a formal sum, since by construction, non of the a_i have any relation to each other, there can't be a way of cancelling them out!

To establish the universal property, of $F^R(A)$, Let's say we have a function $\varphi : A \to M$, where M is an R-module. We need to show this function raises to an R-module homomorphism $\Phi : F^R(A) \to M$. Let Φ be defined by

$$\sum_{i=1}^{n} r_1 a_i \mapsto \sum_{i=1}^{n} r_i \varphi(a)$$

by the uniqueness fo the expression of the elements of $F^R(A)$ as linear combinations of a_i, we see that Φ is indeed a well dfiend R-module homomrophsim!!

check here

By definition, the restriction of $\Phi|_A$ is φ. For uniqueness of $F^R(A)$, since $F^R(A)$ is generated by A, once we know the values of an R-module homomorphism on A, tis value on every element of $F^R(A)$ are uniquely determined (by force of construction), so Φ is the unique extension of Φ to all of $F^R(A)$.

If A is finite ($\{a_1, a_2..., a_n\}$), proposition 12.3.3

$$F^R(A) = Ra_1 \oplus Ra_2 \oplus \cdots \oplus Ra_n$$

Since $R \cong Ra_i$ (under $r \mapsto ra_i$) by proposition 12.3.3 again, the direct sum is isomorphic to R^n.

Definition 12.3.7: Free Module

An R-module F is said to be *free* on the subset A of F if for every nonzero element x of F, there exists a unique nonzero elements $r_1, ..., r_n$ of R and unique $a_1, ..., a_n$ in A such that $x = r_1a_1 + \cdots r_na_n$, for some $n \in \mathbb{Z}^+$, which by proposition 12.3.3 implies that:

$$F = \bigoplus_{i \in I} R$$

If R is a commutative ring, the cardinality of A is called the *free rank* of F, or just *rank* if unambiguous.

Note that A need not be finite, but the linear combinations of elements of A need to be finite (notice that we say $a_1, ..., a_n \in A$, and not $a_1, ..., a_n, ... \in A$). From now on, we will write R^n to mean the free R-module $R \oplus R \oplus \cdots \oplus R$. So $(\mathbb{Z}/2\mathbb{Z})^3$ is a free $(\mathbb{Z}/2\mathbb{Z})$-module. R^1 is a free R-module of rank 1 (to be distinguished from R which we will treat as a ring). If the superscript is infinite, say $N = |\mathbb{N}|$ or $N = |\mathbb{R}|$, then R^N is considered an infinite direct sum, not a direct product. In fact, non-trivial infinite direct product of modules (modules of the form R^N for an N of infinite cardinality) are far from being free (see exercise ref:HERE[10]). Furthermore, the notion of rank is not necessarily well-defined if R is not commutative. We shall explore more of this in chapter 13.

Example 12.3.3: Free Modules

1. Let R be any ring with identity. Then R over itself is a free-module! The element 1 will always generated all of R since 1 is a unit (and so $ab \neq 0$ for any $b \neq 0$) and $r \cdot 1 = r$, and so it's a linearly independent generating set!

2. The "freeness" is dependent on the ring you choose, as well as the module (i.e., the abelian group). So $\mathbb{Z}/n\mathbb{Z}$ over itself is free (by the logic povided above), but not over $\mathbb{Z}$ (which we already know is a well-defined module), since

$$n \cdot 3 = 0$$

   In general, one criterion for free-ness is that $r \cdot a \neq 0$ for any r. This criterion is called being *torsion-free*, and if $r \cdot a = 0$, then r is said to be a *torsion element* of the R-module M. However, a module being torsion free does not imply the module is free.

[10] Inspired by the link here

3. Since $\mathbb{Z}/n\mathbb{Z}$ overt $\mathbb{Z}$ is not a free-module, it does not have a basis. If we try to choose $A \subseteq \mathbb{Z}/n\mathbb{Z}$ to be a basis, it will always have torsion elements. $\mathbb{Z}/n\mathbb{Z}$ over $\mathbb{Z}$ is also an example of a module *without* any linearly independent sets!

4. Let $\mathbb{C}[x, y]$ act on itself. Then $\mathbb{C}[x, y]$ is a free module. As stated in example 12.1.1, ideals of modules over themselves are sub-modules. Take $(x, y) \subseteq \mathbb{C}[x, y]$. Then

$$y \cdot x - x \cdot y = 0$$

showing that (x, y) is *not* free! However, (x, y) is certainly torsion-free, since $\mathbb{C}[x, y]$ is an integral domain.

Thus, free-ness is not preserved by taking sub-modules

5. Let $\mathbb{Z}$ act on itself. Then $\mathbb{Z}/n\mathbb{Z}$ is a quotient $\mathbb{Z}$-module as we've seen in section 12.2. However, as we've seen in the 2nd example, $\mathbb{Z}/n\mathbb{Z}$ over $\mathbb{Z}$ is not free.

Thus, free-ness is not preseved when taking quotient modules.

6. Let's say $S = \{1, 2, 3\}$, and $M = \mathbb{R}^3$ (over $\mathbb{R}$). Take any $f : S \to M$. For our concrete example, let's say:

$$f(1) = (\varphi, \sqrt{2}, 4) \qquad f(2) = (0, 0, 0) \qquad f(3) = (0, 5753, 9)$$

I chose particularily random values to show that f is indeed any set function map. Now if you consider $F^R(S)$ (over $\mathbb{R}$), we'd have some uncountable set of elements, which most importantly are all distinct! It's that last part that let's us say that

$$\psi : F^R(S) \to \mathbb{R}^3$$

will be the unique homomorphism where $\psi|_S = f$. Consider

$$\psi(x) = \psi(\sum_{i=1}^{3} r_i s_i) = \sum_{i=1}^{3} r_i \varphi(s_i)$$

This function is indeed well defined since every x is *unique* in it's represetnation from S, since if it wasn't, and we consider

$$x = y(\sum_{i=1}^{3} r_i s_i = \sum_{i=1}^{3} r_i' s_i)$$

then it is not true in general that

$$\psi(x) = \psi(y)$$

what let's use insure that this will always be true is precicely the linear independence, or free property, of the module!

Finally, this is indeed a [unique] homomorphism. it is easy to check that

$$\psi(x + ry) = \psi(x) + r\psi(y)$$

and the uniqueness will come from free modules elements having no relation to each other (if they did, you might be able to have a second homomrophism).

7. You can have a module with linearly independent sets, but that always fail to generate the module, meaning it is not a free module. For example, $\mathbb{Z}^{\mathbb{N}}$ has a great multitude of linearly independent sets (even infinite ones), however it is not a free-module. This is a rather complicated proof, and is relegated to exercise ref:HERE in chapter ref:HERE.

8. Take $\mathbb{C} \times \mathbb{C}$ over itself. This is free module (its basis is $(1, 1)$). Then $\mathbb{C} \times \{0\}$ is a submodule, but not a free module: Any basis of $\mathbb{C} \times \{0\}$ would consist of elements of the form $(0, a)$. If A is the colleciton of all elements in the 2nd component of a basis, then:

$$\sum_{a \in A} (a, 0) \cdot (0, a) = \sum_{a \in A} (0, 0) = (0, 0)$$

meaning a non-trivial linear combination summed to zero.

9. Let R be a Principal Ideal Domain acting over itself. Then any submodule N of R is an ideal, and since it's principal $N = (a)$ for some $a \in R$. Since $(a) = Ra$, it follows that N is torsion-free (see ref:HERE). Thus, any submodule of R is torsion-free. (TBD, not sure, trying to get to if R is not principal, then (x, y) is never torsion free. Maybe move this to Mod over PID sec? exercice 6 in section 12.1 dummit)

10. In chapter 13, we'll show that $\mathbb{R}$ is a $\mathbb{Q}$-vector space with unountable basis! Note that even though the basis is uncountalble, any element $x \in \mathbb{R}$ will still be a $\mathbb{Q}$-linear combination of *finitely many* basis elements.

11. On the other hand $\mathbb{Q}$ over $\mathbb{Z}$ is *not* a free module (i.e. $\mathbb{Q}$ is not a free abelian group). To see this, notice that for any $\frac{a}{b}, \frac{c}{d} \in \mathbb{Q}$, there exists $m, n \in \mathbb{Z}$ (namely $m = bc,\ n = -da$) such that $(bc)\frac{a}{b} + (-da)\frac{c}{d} = 0$, and hence any two elements are linearly independent, meaning if $\mathbb{Q}$ was a free $\mathbb{Z}$-module, it would have to be cyclic. As we've seen in section 2.3.4, the only free cyclic $\mathbb{Z}$-module is $\mathbb{Z}$. Can you see why $\mathbb{Z} \not\cong_{\mathbb{Z}} \mathbb{Q}$?

Take note of the fact that whether a module is free or not is highly dependent on what ring is acting on it. In $\mathbb{Z}$-**Mod**, $\mathbb{Z}/n\mathbb{Z}$ is certianly not a free (it is not a free group). However, we've just seen that $\mathbb{Z}/n\mathbb{Z}$ is a free-module over itself! The intuition here behind being "free" comes down to the ring-action being "free".

Corollary 12.3.8: Consequences Of Universal Property

1. If F_1 and F_2 are free modules on the same set A, there is a unique isomorphism between F_1 and F_2 which is the identity map on A

2. if F is any free R-module a free-generating set A, then $F \cong F^R(A)$. In particular, F enjoys the same universal property with respect to A as $F^R(A)$ does in theorem 12.3.6.

Proof **:**
exercise

36: foreShadowing

Though a free R-module has no relations, it is not the case that any submodule has no relation, that is the submodule of a free module *need not be free*! In other words, though it may at first seem counter-intuitive, the process of taking a submodule can add relations (see exercsie ref:HERE)! This contrasts to how the subgroup of a free group is always free (see appendix B). There are certain rings for which the R-submodule are always free, while there are other rings where submodules are far from being free (ex. some submodules will be torsion modules, which will be far away from being free). We will explore these concepts further in chapter 13 and 16.

Measuring degree of Freeness

As we've seen, not all modules are free modules. However, we may still like to find collection of elements that are linearly independent (i.e. that are "free" of relations). We explore these notions here.

Definition 12.3.9: [Free] Rank

For any ring R, the *rank* of an R-module M is the maximum number of R-linearly independent elements of M, that is if M has rank n, there exists elements $y_1, y_2, ..., y_n$ such that

$$r_1y_1 + r_1y_2 + \cdots + r_ny_n = 0 \quad \Leftrightarrow \quad r_1 = r_2 = \cdots = r_n = 0$$

If for all $n \in \mathbb{N}$, there exists a R-linearly independent set, then M is said to be of infinite rank.

Due to the importance of the notion of rank for modules in geometry, the rank is also sometimes called the *Betti Number*[11]. The rank of a module is similar to the size of the generating set of a module, and it will in fact match exactly the idea of rank for modules over fields, but this is not necessarily the case for modules over weaker rings:

Example 12.3.4: Different Rank and Size of Generators

Let $R = \mathbb{Z}[x]$ and let $M = (2, x)$ be an R-module. If we take $2 \in M$, then since R is an R an integral domain and M is an ideal of R, there does not exist a r such that

$$ra = 0$$

thus, $\{2\}$ is a linearly independent set from M, meaning it's at least rank 1. What if we two elements? Let's take $a, b \in M$. Since $a, b \in R$, in particular $-a \in R$

$$ba - ab = 0$$

meaning 2 elements will never be linearly independent. This will naturally generalize to n elements, so M is of rank 1.

However, M is generated by 2 elements, showing that the rank and the number of generators does not match.

[11] The kth Betti number loosely refers to the number of k-dimensional holes ona a topological surface

Sometimes, authors define the minimal set of generators to be the rank of the matrix, and rank as was defined above would be called the *free rank*. Since the size of the generating sets rarely comes up outside the context of free generating sets, this book decided to stick with "rank" to mean the maximal number of linearly independent elements.

Not that it is yet to be established if there is a unique maximal rank (just like in a tree, there may be many maximal heights). For modules over integral domains, the maximal height will alwaays be unique, and under appropriate finiteness conditions it shall also be unique, however there are counter-examples over more general rings; this shall be further explore in chapter 13.

Using the notion of rank, we can see how a module might fail to be free. In the "worst case scenario", we can single out elements fail to be linearly independent:

Definition 12.3.10: Torsion Element

Let M be an R-module and $m \in M$. Then m is called a *torsion element* if m is linearly *de*pendent, in particular there exists an $r \in R$ such that

$$r \cdot m = 0$$

The collection of all torsion elements of M is denoted $\mathrm{Tor}_R(M)$ or $\mathrm{Tor}(M)$ is the context is clear. An element that is not torsion is called *torsion-free*.

Elements that are torsion-free can be thought of as being "locally" free. Being is a property of modules, while being torsion-free is a property of elements. If every element has this property, then we give a name to such a module:

Definition 12.3.11: Torsion-Free Modules

Let M be an R-module. Then if $\mathrm{Tor}_R(M) = \{0\}$, then M is said to be *torsion-free* or to be a *torsion-free module*.

Being torsion free is an iteresting example of the fact atht it is notalwaays enough that every element have a proeprty for the global module to have the property. In this case, even if every element itself is free, this does not imply the module itself is free

Example 12.3.5: Torsion-free, But Not Free Module

Take $\mathbb{C}[x, y]$ over itself. Then since it's an integral domain, it is easy to see that it is torsion free and each submodule is torsion free. However, (x, y) is not free.

Looking instead at the ring action instead of the module element, we can also find all elements of the ring that fail to be linearly dependent with *all* module elements:

Definition 12.3.12: Annihilator

Let M be an R-module. Then the set

$$\mathrm{Ann}_R(M) = \{r \in R \mid r \cdot m = 0 \ \forall m \in M\}$$

is called the *annihilator* of M.

Recall that we have talked about this set in example 12.1.1(4) when discussing modules where the annihilator is trivial; such a module is given a name:

Definition 12.3.13: Faithful Module

Let M be an R-module. Then if $\mathrm{Ann}_R(M) = \{0\}$, M is said to be a *faithful R-module*, or that the ring R is acting faithfully on M

If R does not act faithfuully on M, when we may always induce a faithful action along the lines of example 12.1.1(4). The same cannot be said about non-torsion modules, instead we may ask if a module M can be split into a torsion-free part and a torsion part. Being trosion free is much more flexible than being free:

Proposition 12.3.14: Extensions And Submodules Of Torsion-Free Modules

Let M be an R-module. Then:

1. If M is torison-free, then any submodule is torsion free
2. the direct sum of torsion free modules is torsion-tree
3. if M is an extension of torsion-free modules, that is $0 \to A \to M \to B \to 0$ is a short exact sequence with A and B being torsion-free, then M is torsion-free.
4. If R is an integral domain M is free, then M is torsion-free (i.e. free modules over integral domains are torsion-free)

Proof **:**

1. This is immediate after realising $\mathrm{Tor}_R(N) \subseteq \mathrm{Tor}_R(M)$
2. this is also almost immediate givn a similar type of reasoning as in part (1)
3. If $m \in M$, where $r \cdot m = 0$, then $r \cdot \overline{m} = \overline{0}$, but since $r \in\in \mathrm{Tor}_R(B) = \{0\}$ $r = 0$, and so $\mathrm{Tor}_R(M) = 0$.
4. Since the direct of free modules is free, and the integral domain R as a module over itself is torsion free, M is torsion-free.

In chapter 14, we shall further explore the decomposition of the torsion and torsion-free part of a module over nicer enough rings so that the decomposition is possible.

If N is not a torsion module, then $\mathrm{Ann}(N) = 0$. It is possible that N is a torsion module, but $\mathrm{Ann}(N) = 0$:

Example 12.3.6: Torsion Module With No Annihilators

Take

$$Z = \bigoplus_k \mathbb{Z}/2^k\mathbb{Z}$$

then Z is a $\mathbb{Z}$-torsion module, since for every $z \in \mathbb{Z}$, we can take 2^k, where we pick k relative to

the largest nonzero compoenent. However,

$$\mathrm{Ann}(Z) = 0$$

Since if $r \in \mathbb{Z}$ was an annihilator, then $r \cdot z = 0$ for every z, but z can take any value of $\mathbb{N}$, leading us to a contradiction

If the module is finitely generated the annihilator will always be non-empty, since if we annihilate the generators, we annihilate the entire module.

Exercise 12.3.1

1. Let M be an R-module, and define $\mathrm{Ann}_R(M) = \{r \in R \mid r \cdot m = 0,\ \forall m \in M\}$.
    1. Show that $\mathrm{Ann}_R(M)$ is a two-sided ideal
    2. If $\mathrm{Ann}_R(M) = 0$, then M is said to be a faithful module. Show that all free modules are faithful, but there are faithful non-free modules
2. Let R be a commutative ring and an R-module over itself. Then show that an ideal $I \subseteq R$ is a free R-module if and only if it is principal and generated by a nonzero divisor.
3. Show that $\mathbb{Z}^{\mathbb{N}}$ (a direct product of countably many $\mathbb{Z}$ is *not* a free abelian group

12.3.3 Presentations and Resolutions

We claimed that all R-modules are a quotient of a free R-module, just how all groups are quotient of free groups. We will demonstrate this here, and then explore a bit furter how we can think of relations and presentations of modules:

Proposition 12.3.15: Constructing Modules using Free Modules

Any R-module M is the quotient of a free module $F^R(A)$ for appropriate set A:

$$F^R(A)/K \cong_R M$$

Proof **:**
Let M be an R-module and $\mathcal{F} = F^R(M)$. Define the map:

$$\varphi : \mathcal{F} \to M \qquad (r_i m_i)_{i \in M} \mapsto \sum_{i \in M} r_i m_i$$

where any element on the left hand side is a tulpe indexed by M where all but finitely many of the terms are nonzero, and hence the right hand side is well-defined. This map is evidentely surjective (since the tulpe of m in the first element maps to m), and so if K is the kernel of this map, by the first isomrophism theorem:

$$F^R(M)/K \cong_R M$$

as we sought to show.

Hence, the kernel of a free module plays the role of codifying all the relations in a module. When working with groups, we encoded these relations writting a group as either, for example:

$$\langle x, y \ : \ xy - yx = 0 \rangle \qquad \text{or} \qquad \frac{F(x,y)}{xy - yx}$$

The former notaiton being a convienient notation that can be easier to read before the introduciton of quotienting. In the next definition, we will define the presentaiton of a module. Going off of the fact that the kernel is the central object defining the relation, we incorporate this idea into the definition:

Definition 12.3.16: Presentation Of Module

Let M be an R-module. Then a *presentation* of M is a an exact sequence:

$$R^n \to R^m \to M \to 0$$

If n is finite (i.e. if $n \in \mathbb{N}$), then M is said to be *finitely presented*, or a *finitely presented module*.

Note that m and n need not be natural numbers, and that the above is *not* Equivalent to

$$0 \to R^n \to R^m \to M \to 0 \tag{12.5}$$

i.e. R^n need not map injectively into R^m. The easiest way to show this is if we classify all modules for which equation (12.5) holds, which turn out to be all modules over PID's, and hence modules that are *not* over PID's (for example, $\mathbb{Z}[x]$ over itself) cannot be written down in the form presented in equation (12.5). In chapter 14, we shall show that a module over a PID shall imply the existence of a chain as given in equation (eq:modOverPidAsPresentaiton), that is it is a sufficeint condition that R is a PID. It is easier to show that it is a necessary condition; we shall show this at the end of this seciton, for now we will assume the complete result to point out some interesting observations.

It is informative to the reader to see how presentations of modules compares to presentations of groups. Note that $\mathbb{Z}$ is a Principal Ideal Domain, and so for any [finitely generated?] abelian group (i.e. any $\mathbb{Z}$-module), we have the exact sequence:

$$0 \to \mathbb{Z}^n \to \mathbb{Z}^m \to G \to 0$$

Furthermore, any subgroup of a free group is free (see appendix B), and so more generally for any group G and appropriate sets A and B:

$$0 \to F(A) \to F(B) \xrightarrow{\varphi} G \to 0$$

where $F(A) = \ker(\varphi)$. Hence, the fact that there are modules that do not admit a chain like in equation (12.5) is a property that distinguishes groups from modules. In fact, there are even modules for which there is no finite length chain of free modules. If we interpret equation (12.5) as saying that the relations on a module is free, then if we extend the length by 1:

$$0 \to R^n \to R^m \to R^p \to M \to 0$$

can interpret the extra homomorphism as the relations of the realtions, where the "relations of the relations"' are free. Some modules do not a finite finite length chain of this sort!! We give a name to such a chain

Definition 12.3.17: Free Resolution

Let M be an R-module. Then if there exists a chain of free R-modules:

$$\cdots \to R^{n_2} \to R^{n_1} \to R^{n_0} \to M \to 0$$

Then M is said to have a *free resolution*. If there exists an n_k such that

$$0 \to R^{n_k} \to R^{n_{k-1}} \to \cdots \to R^{n_1} \to R^{n_0} \to M \to 0$$

Then M is said to have a *finite length free resolution*, of length k.

By definition, all free modules have a finite length free resolution of length zero since

$$0 \to R^{n_0} \to M \to 0$$

i.e. $R^{n_0} \cong M$. By the universal property of free modules, any R-module M has a free resolution (can you see why). The resolution is called a *free* resolution since we are using free-modules. If we use a different type of module, we will get a different type of resolution[12].

Example 12.3.7: Free Resolution

Let G be a finitely generated abelian group. Then by the fundamental theorem for such groups (theorem 5.1.5),

$$G \cong \mathbb{Z}^n \oplus \mathbb{Z}/n_1 Z \oplus \mathbb{Z}/n_2\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/n_k\mathbb{Z}$$

As pointed out, all free-modules have a resolution of length 1, namely

$$0 \to \mathbb{Z}^n \to \mathbb{Z}^n \to 0$$

On the other hand, any cyclic torsion group has a resolution of length 2:

$$0 \to \mathbb{Z} \xrightarrow{\cdot n} \mathbb{Z} \xrightarrow{\pi} \mathbb{Z}/n\mathbb{Z} \to 0$$

These results can be combined to show that G always has a resolution of length 2. Free resolution that must be at least of length 3 will be given in part VII.

Finally, as promised earlier, we shall show that it is a necesary conditions that R-modules admitting resolutions of length two must be over PID's:

Proposition 12.3.18: Free Resolution Of Length 2, Then Pid

Let R be an integral domain such that every R-module M admits a resolution of length 2. Then R is a PID. In particular, for any resolution:

$$R^n \to M$$

There exists an R^m such that

$$0 \to R^m \to R^n \to M$$

[12] In chapter 16, we will introduce projective, injective, and flat modules, and then in part VII, we will introduce projective, injective, and flat resolutions

***Proof* :**
Choose any $I \subseteq R$. Then R/I is an R-module and

$$R \to R/I$$

is an R-module homomorphism. By the property of R, the above homomorphism can be extended to a free-resolution of length 2. Since the kernel of the above homomorphism is I, we get:

$$0 \to I \to R \to R/I$$

where we must assume I is a free R-module. Since R is of rank 1, I cannot be greater than rank 1 (see exercise ref:HERE (above direct sums and homomorphisms)). But then, since I is free, I must be generated by this single element, making it free. Since this is true for an arbitrary ideal $I \subseteq R$, this implies that every ideal of R is principal making R a PID, as we sought to show.

12.3.4 Finitely Generated Modules and Noetherian Modules

As with many structures in mathematics, the "infinite" case (in our case infinitely generated objects) have different behavior than the finite case and require some different tools to analyze (possibly some analysis to use the notion of limits). We thus study the case where modules are finitely generated or all submodules are finitely generated.

Let R be a ring and let M be a left R-module. We first define what types of modules are *finitely generated*, and some consequences of being finitely generated,

Definition 12.3.19: Finitely Generated R-Module

Let M be an R-module. Then M is said to be finitely generated if there exists an $n \in \mathbb{N}_{>0}$ such that $M \cong_R R^n/K$ for some R-module K, that is there exists a map φ where

$$R^n \xrightarrow{\varphi} M \to 0$$

is exact so that $\ker(\varphi) = K$ and $R^n/K \cong_R M$.

Though we have defined finitely geneated modules, we usually require the stronger condition that all submodules of a finitely generated R-module to be finitely generated (which is not true for ever finitely generated modules). We will phrase this condition a little differently and show it's equivalent to the previous statement:

Definition 12.3.20: Noetherian R-Module and Ring

Let M be a left R-module Then:

1. M is said to be *neorthian R-module* or to satisfy the *ascending chain condition on submodules* (or A.C.C. on submodules) if there are no infinite increasing chains of submodules, i.e., whenever
$$M_1 \subseteq M_2 \subseteq M_3 \subseteq \cdots$$
is an increasing chain of submodules of M, then there is a positive integer m such that for all $k \geq m$, $M_k = M_m$ (so the chain becoems stationary at stage m: $M_m = M_{m+1} = M_{m+2} = \cdots$)

2. The ring R is said to be Neotherian if it is Neotherian as a left-module over itself, i.e., if there are no infinite increasing chains of left-ideals in R.

One can have an analogous definition of A.C.C. for right and two-sided ideals in (possibly noncommutative) ring R. For noncommutative rings, left and right A.C.C need not relate (you may be left A.C.C but not right, and vice-versa, see exercise ref:HERE)

Theorem 12.3.21: Equivalent Conditions To Noetherian

Let R be a ring and let M be a left R-module. Then the following are equivalent

1. M is a Noetherian R-module
2. Every nonempty set of submodules of M contains a maximal element under inclusion
3. Every submodule of M (including M) is finitely generated

Proof **:**
(1) $\Rightarrow$ (2). Let S be any collection of submodules. Let's pick some sub-module $M_1 \in S$. If S is maximal, then we're done. If not, then M_1 is not maximal, meaning there exists an M_2 such that $M_1 \subseteq M_2$. If M_2 is maximal, we're done. If not, pick M_3 in a similar fashion. Keep picking sub-modules M_i. If none of them are maximal, then by the axiom of choice, we can pick an infinitely long chain of submodules
$$M_1 \subseteq M_2 \subseteq \cdots \subseteq M_k \subseteq \cdots$$
which is a contradiction to the statement that M is Noetherian.

For (2) $\Rightarrow$ (3), Let M be a module. Pick $x \in M$. If $\langle x_1 \rangle = M$, then M is finitely generated. If $\langle x_1 \rangle = N_1 \subsetneq M$, then pick another element $x_2 \in M - N_1$. If $\langle x_1, x_2 \rangle = M$, we're done. If $\langle x_1, x_2 \rangle = N_2 \subsetneq M$, then pick another element $x_3 \in M - N_2$. Either this stabalizes, or we can craete an infinite chain:
$$N_1 \subseteq N_2 \subseteq \cdots \subseteq N_k \cdots$$
This is a collection of submomdules of M, and so by the 2nd condition there must exist a maximal module, say N. Since every module in this chain is finitely generated, N mut be finitely generated, and since it's maximal adding one more element to N, $N + x$ and taking it's closure, $\overline{N + x} = N'$ must make the new larger module N' equal to M. Hence, M is finitely genearted.

For (3) $\Rightarrow$ (1), Take any ascending chain of modules

$$N_1 \subsetneq N_2 \subsetneq \cdots \subsetneq N_k \subsetneq \cdots$$

then $N = \bigcup_{i=1}^{\infty} N_i$ is a module, and by assumption it is finitely generated, say by $\{x_1, x_2, ..., x_n\}$. By definition of N, for each x_i, there exists a j such that $x_i \in N_j$. Since N is the union of the modules N_j, it must be that at some finite distance into the chain, a module contains all the generators. But that means the chain stabalizes.

Note that if M is a Noetherian R-module, then R need not be Noetherian. If R is noetherian, then M is a Noetherian R-module? (Use the fact that the direct sum and quotient of a noetherian ring is noetherian, an the uni prop of free modules).[13]. In proposition 9.4.10, it was shown that rings where $1 \neq 0$ have maximal ideals. Since all rings are modules over themsleves, it might seem that all rings are Noetherian modules due to the second conditions. However, this is certainly not the case, consider $R[x_1, x_2, ..., x_n, ...]$ over itself and the chain $(x_1) \subsetneq (x_1, x_2,) \subsetneq \cdots$. The subtly lies in the fact that the maximal element must lie in the collection of submodules, while the maximal element in the previous example is the union of all the submodules.

Many textbooks introduce Noetherian modules as modules where each submodule is finitely generated. The reason the A.C.C. was chosen for this textbook was for it's greater applicability in many other fields. Notions of being "Noetherian" are pervasive in mathematics, but not all structure are "Noetherian" by having every substructure be finitely generated (sometimes, the notion of substructure is not even well-defined)[14]. The A.C.C. allows us to still consider those structures.

Since Principal Ideal Domain are *a fortiori* Noetherian rings, finitely generated modules over PID's are are also Noetherian, and all the results we discussed earlier apply to such modules (including PID's over themsleves).

Finite Generation vs. Finite Type

Contrast an finitely generated R-module M to a finitely generated R-algebra M:

$$R^{\oplus A} \twoheadrightarrow M \quad R[A] \twoheadrightarrow M$$

(more to say here)

Exercise 12.3.2

1. Is a subring of a Noetherian ring Noetherian?
2. (If you know some complex analysis) In this exericse, we try to show how the Noetherian property is not stable by taking subrings and show demonstrate a correspondence between being Noetherian and hwo the zero of functions are in some way "well-behaved". Consider

[13]This mirrors a more general theme in which a property of the underlying ring "becomes" the property of all modules over that ring. Can you see how we can (in a sense) call a field a "free ring"?

[14]See Noetherian topologies for an interesting example

the following chain of subrings:

$$\begin{gathered}\mathbb{R}[x] \\ \subseteq \\ \text{Analytic functions on } \mathbb{R} \\ \subseteq \\ \text{Analytic functions on } [0,1] \\ \subseteq \\ \text{Analytic functions on } (0,1) \\ \subseteq \\ \text{Analytic functions at } 0 \\ \subseteq \\ \text{Smooth functions near } 0 \\ \subseteq \\ \text{Former Power series}\end{gathered}$$

which of these rings are Noetherian, and which are not?

3. Show that a finitely generated R-module over a Noetherian ring R is itself Noetherian
4. Let M be an R-module over an integral domain R. Then if M is of rank n, any submodule $N \subseteq M$ can be of rank at most $n \leq m$ (hint: take the ring of fractions).

12.3.5 Direct Sums and Homomorphisms

For the first time in this book, we will pay closer attention to hom-sets. The structure of hom-sets of modules will allo us to determine many important properties of modules.

Proposition 12.3.22: Direct Sums And Homomorphisms

Let $(M_i)_{i\in I}$ and $(N_j)_{j\in J}$ be a collection of modules with I and J indexing sets. Then:

$$\mathrm{Hom}_R\left(\bigoplus_{i\in I} M_i, N\right) = \prod_{i\in I} \mathrm{Hom}_R(M_i, N)$$

and

$$\mathrm{Hom}_R\left(M, \prod_{j\in J} N_j\right) = \prod_{j\in J} \mathrm{Hom}_R(M, N_j)$$

Proof **:**
let $\varphi : \oplus_{i\in I} M_i \to N$. Then for any $(\alpha_i)_{i\in I} \in \oplus_{i\in I} M_i$, there are only finitely many non-zero entries. For ease of visualization, lkets represent the element $(\alpha_i)_{i\in I}$ as

$$(\alpha_1, \alpha_2, ..., \alpha_n, 0, ...)$$

This represetnation can be confusing on two terms: we have implicitly chosen an "order" of the elements and re-labelled the indices of the finite eleents accordingly, and it makes the set I look like it is *countable*, which is not necessarily the case. These two remarks should be kept in mind while working with this represetnation of the element.

Since there are n elements, write each of them as:

$$\alpha_i = \underbrace{0 + \cdots + 0}_{i-1 \text{ times}} + \alpha_i + \underbrace{0 + \cdots + 0}_{n-i \text{ times}}$$

Now, the finiteness condition was key here so that we may use linearity inductively, so that for any $\varphi \in \operatorname{Hom}_R\left(\bigoplus_{i \in I} M_i, N\right)$

$$\varphi((\alpha_i)_{i \in I}) = \varphi(\alpha_1) + \cdots + \varphi(\alpha_n)$$

Now, it is immediate that each function on the right hand side is naturally isomorphic to a function $\varphi_i \operatorname{Hom}_R(M_i, N)$. Labelling eahc of thes functions a φ_i, we get:

$$\varphi \to (\varphi_i)_{i \in I}$$

which is clearly an isomorphism.

The direct product case is done almost identically, except we don't need to worry about finite additivity, since for any

$$\varphi(\alpha) = (\beta_i)_{i \in I}$$

we can define:

$$\varphi_i(\alpha) = \beta_i$$

which will give us the desired isomorphism, completing the proof.

As a consequence of this, we can simplify:

$$\operatorname{Hom}_R\left(\bigoplus_i M_i, \bigoplus_j N_j\right) \cong \prod_{i,j} \operatorname{Hom}_R(M_i, N_j)$$

If you are comfortable with linear algebra, then in the case where both direct sums have only finitely many components (also known as *summands*), then if φ_{ji} represent's the component function taking in the ith component in the domain and is in the jth component of the domain, then: the there is an isomrophism mapping $\varphi \mapsto (\varphi_{ij})$ where the maps is defined as:

$$\varphi \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} = \begin{pmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1n} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2n} \\ \vdots & \vdots & \ddots & \cdots \\ \varphi_{n1} & \varphi_{n2} & \cdots & \varphi_{nn} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix}$$

Showing that there is a natural relation between $\operatorname{Hom}_R(\oplus_i^m M_i, \oplus_j^n N_j)$ and matrix rings, especially if $M_i = N_j$. More on that in chapter 24, section 23.

Notice how important linearity is to split up the homomorphsim in direct products in the *domain*. To make sure that no misconceptions are being formed, here is a counter-example when a function is not linear:

Example 12.3.8: Hom's And Products Not Commuting in Domain
Let $f(x, y) = xy$. Then $f(x, 0) = 0$ and $f(0, y) = 0$, showing this breakdown does not happen. Of course, this was a particualr example, not a proof of the non-existence of the isomorphism. However, we see with f that the output of f must always depend on both x and y, so any two functionts f_1 and f_2 solely dependning on the x or y variable cannot be used to reconstruct f using a direct product TBD.

On the other hand, if $f : M \to A \oplus B$ was some function, then by the universal property of the product we have that $f_a : M \to A$ and $f_b : M \to B$ exists. (Think more on why this exists but the other doesn't so nicely.. think about what it means for the uni prop of the co-product and how it relates to hom-sets.. TBD)

Exercise 12.3.3

1. Show that if M is an R-module, then $\mathrm{Hom}_R(R, M) \cong M$
2. Show that $\mathrm{Hom}_R(M \oplus M, R) \cong$(want to show the case where direct product becomes matrices! i.e. bringing out the direct product makes a $M_2(R)$ matrix, whatever should be there instead of R)
3. Let $\varphi : A \to B$ be an R-module isomorphism. Show that $\Phi : \mathrm{End}_R(A) \to \mathrm{End}_R(B)$ where $\Phi(f) = \varphi \circ f \circ \varphi^{-1}$ is a ring isomorphism, and if R is commutative an R-algebra isomrophism.
4. show that if $I \subsetneq R$ is a proper ideal, $I \cong S$ as an R-module isomorphism $\varphi(e) = 1$ a,d $e^2 = e$, then φ is in fact a ring-isomrophism, so I has the ring structure of S (though it cannot have the ring structure of R, since it cannot contain a unit).

13

Modules over Fields – Vector Spaces and Linear Algebra I

This chapter focuses on the study of modules where the ring are fields (i.e. [unital] commutative division ring). We'll use the letter k (from German Körperor) or the symbol $\mathbb{F}$ to represent an arbitrary field. As we've mentioned in definition 12.1.7, a module over a field is called a vector-space due to the historical precedence of the terminology which is now widely accepted in and beyond mathematics (for example, the term vector-space is used in physics, biology, and chemistry). This precedence of modules over fields comes from the fact that historically, fields were studied before general rings. Many problems in physics in the 19th century turned out to requires vector-spaces (ex. solutions to ODEs formed vector spaces, many important spaces of functions formed vector-spaces, the cartesian plane formed a vector space), leading to precedence in nomenclature for modules over fields. Mathematically, giving another name to modules over fields is not unreasonable: as we'll see, having a field act on a module is so restrictive (or powerful) that we will only need to keep track of two pieces of information to fully understand the structure and the classification of vector-spaces: the underlying field k, and a new number called the *dimension* of the vector-space[1]. No algebraic structure we've encountered requires so little data to be able to classify them (for example, the classification of finite groups required tens of thousands of paper's over almost a century, while the clasificaiton of vector-space is regularly taught in a 1st or 2nd year undergraduate class).

Due to the simplicity of vector-space structures, vector space homomorphisms (k-module homomorphisms) will imminently calssifiable. (A further word on this before talking a bit about dual theory and geometry)

As a consequence of the historical precedence of vector-spaces, a seperate vernacular was developed

[1]This is particularly true for finite dimensional vector-spaces. In the infinite dimensional case, there is usually more information present than in the finite dimensional case, and so we require more information to sufficiently classify infinite vector-spaces, see ref:HERE

Terminology over any Ring	**Terminology over Fields**
M is an R-module	M is a vector space over R
m is an element of M	m is a vector in M
N is a submodule of M	N is a subspace of M
M/N is a quotient module	M/N is a quotient space
M is a free module of rank n	M is a vector space of dimension n
M is a nonzero cyclic module	M is a 1-dimensional vector space
$\varphi : M \to N$ is an R-module homomorphism	$\varphi : M \to N$ is a (R) linear transformation
M and N are isomorphic as R-modules	M and N are isomorphic vector spaces
the subset A of M generates M	The subset A of M spans M
$M = RA$	each element of M is a linear combination of elements of A, i.e., $M = \mathrm{Span}(A)$

for modules over fields that is now considered standard:

13.1 Vector Spaces and their Properties

Vector spaces are famous for having a very simple structure: Given a vectorspace V over k, then V is isomorphic to k^N for appropriate N. In a linear algebra class, we say that every vector space has a basis. This N will be called the *dimension* of the vector-space. In many linear algebra classes, the attention is limited to $N \in \mathbb{N}$. However, this section will allow N to be of any cardinality (ex. $|\mathbb{N}|$ or $|\mathbb{R}|$). This is to show some of the power of vector-spaces: the (algebraic) structure doesn't get more complicated when dealing with infinite dimensions[2]

Before these results are shown, for those who gain an intuition via examples, here are a couple more example's of vector-spaces

Example 13.1.1: Vector Spaces
We already gave 4 examples in example 12.1.3. We give a few more here

1. We can generlize k^n as follows: notice that we can define k^n as

$$k^n = \{f \mid f : \{1, ..., n\} \to \mathbb{R}\}$$

   where addition is defined pointwise, and since k is commutative we can define a natural k-action on k^n (can you see it?). Any function $f \in k^n$ will tell us what value we should have in the component i. For example, if $f(1) = \pi$, $f(2) = 3$, an $f(3) = \sqrt{10}$, then this function is equivalent to:

$$(\pi, 3, \sqrt{10})$$

   Thus, We can consider:

$$k^{\mathbb{N}} = \{f \mid f : \mathbb{N} \to \mathbb{R}\}$$

   To be the $\mathbb{N}$-dimensional vector-space, or $k^{\mathbb{R}} = \{\mathbb{R} \mid f : \mathbb{R} \to \mathbb{R}\}$ to be the $\mathbb{R}$-dimensional vector space

2. Generalizing the above example, let k be a field, X a set, and $V = \mathrm{Hom}_{\mathbf{Set}}(X, k)$ (another common notation is k^X). Then we can make V into a vector-space via point-wise operation in the codomain: for any $f, g \in V$, $c \in k$, $f + g$ is defined to be:

$$(f + g)(x) = f(x) + g(x) \qquad \forall x \in X$$

   while fg and cf is defined as:

$$(fg)(x) = f(x)g(x) \qquad (cf)(x) = cf(x) \qquad \forall x \in X$$

   These definitions are possible since the image of any function in V lands in a field, where all the above operations are well-defined. If we take the collection of lienar transformation from V to k (where V is a vector space), then we often denote that as V^*. Note that since k is also a ring, we can also define $(fg)(x) = f(x)g(x)$, making $\mathrm{Hom}_{\mathbf{Set}}(X, k)$ into a k-algebra.

   Often, the set X is given its own properties (a popular example are X is a locally compact Hausdorff space, or X is a smooth manifold). Then we often try to deduce more information about X given the structure of hom-set $\mathrm{Hom}_{\mathbf{Set}}(X, k)$.

[2] On the other hand, the set-theoretical and analytic structure (if there is a topology present) gets more complicated, and arguably non-trivial, in infinite dimensions. For example, in infinite dimensions, surjective does not imply injective (similar to the proof for infinite sets), and there will exist linear transformations that are not continuous (for the appropriate topology).

3. Let $R = k[x]$ for some field x. Then $k[x]$ is a k vector-space with action $c \cdot p(x) = cp(x)$. More generally, if $R = k[x_1, x_2, ..., x_n, ...]$ is some polynomial ring over several variables, then R is a k vector-space with $c \cdot p(x_1, x_2, ..., x_n, ...) = cp(x_1, x_2, ..., x_n, ...)$.

4. Let $k = \mathbb{C}$ an $V = \mathbb{C}^n$. Then $\mathbb{C}^n$ is a complex vector space with the action:

$$z \cdot (z_1, ..., z_n) = (zz_1, ..., zz_n)$$

There is a second natural action, known as the *conjugate action*:

$$z \cdot (z_1, ..., z_n) = (\overline{z}z_1, ..., \overline{z}z_n)$$

More generally, if V is a vector-space over $\mathbb{C}$ (also known as a complex vector space), then $\overline{V}$ denote the same space with the element being conjugated before the action is applied.

5. (If you know some ODE's) Given a linear ordinary differential equation, the set of all solutions form's a vector-space. Take for example $\varphi : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ where:

$$f \mapsto f'' + f$$

Consider the subspace V where

$$V := \{f \in C^\infty(\mathbb{R}) \mid f'' + f \equiv 0\}$$

Then using elementary ODE theory, we find that:

$$V = \text{span}_{\mathbb{R}}(\{e^t, e^{-t}\})$$

The fact that all solutions to the ODE $f'' + f = 0$ is a linear combination of these two solutions is a remarkable property of *linear* ODE's. This simple property of linearity can be taken very far: in the world of PDE's (Parital Differential Equations), if you have something called a *linear PDE*, which means that if you have two solutions, then add them and/or scale the terms, the result is another solution. The power we get from this is that we can then see if this is generalizable to taking *limits* of some very simple solutions to the PDE. As a common example, let's say the solution is of the form $\cos(\alpha x)$ or $\sin(\beta x)$ for any α, β. Then under the right conditions[a], the collection of these (after being closed under addition) is *dense* in $L^2(S^1)$, meaning we can use limits to construct *any* $L^2(S^1)$ function and find another solution to our PDE. A common place this particular example is used is *Fourier Analysis*[b]

6. (If you know Differential Geometry) Let M be a smooth n-manifold and consider the collection of all $\varphi : C^\infty(M) \to \mathbb{R}$. Then the collection of all *derivations at* p[c] for some point $p \in M$, denoted $\text{der}_p(M) \subseteq \text{Hom}_{\mathbf{Smooth}}(C^\infty(M), \mathbb{R})$ forms a vector-space of dimension n, and is a common definition of the *tangent space* of the manifold M at the point p[d]

[a] See a textbook covering Fourier Analysis, ex. Pugh or Folland
[b] See EYNTKA Analysis (cite:HERE)
[c] linear function where $\varphi(fg) = \varphi(f)g + f\varphi(g)$
[d] See EYNTKA Differential Geometry, cite:HERE

Now, we shall develop a more sophisitaced notion of generating set:

Definition 13.1.1: Linear Independence And Basis

Let M be an R-module.

1. A subset $S \subseteq M$ is said to be *linearly indepenent*, or a set of *linearly independent elements*, if for any equation

$$a_1 m_1 + a_2 m_2 + ... + a_n m_n = 0$$

with $a_1, a_2, ..., a_n \in R$ and $m_1, m_2, ..., m_n \in S$, then $a_1 = a_2 = \cdots = a_n = 0$. If this is not the case, the elements are called *linearly dependent*. In other words, if $I = \{m_1, m_2, ..., m_n\} \subseteq M$, so that $\iota : I \to M$ is injective, then:

$$\begin{tikzcd} F^R(I) \arrow[r, "\varphi"] & M \\ I \arrow[u, "j"] \arrow[ur, "\iota"'] & \end{tikzcd} \tag{13.1}$$

commutes[a] and φ is injective. Note that I can be an arbitrary indexing set, that is for every $i \in I$, $m_i \in M$ corresponds to it, and then we would have an injection $\iota : I \to M$ which lifts to $\varphi : F^R(I) \to M$. If $R = k$, then the elements are called *linearly independent vectors*.

2. A *basis* for M is an *ordered set* $(a_1, a_2, ..., a_n)$ of lineraly independent elements which generates M (i.e. span M), that is if there exists $A \subseteq M$ such that:

$$\varphi : F^R(A) \cong M$$

is an isomorphism, i.e. the φ in equation (13.1) is both injective and surjective. In paticular, two bases will be considered different even if one is a rearrangement of the other (hence why it's ordered)

[a] The diagram commuting is the universal property of free modules

Note that the φ from the commutative diagram in equation (13.1) has as domain all *finite* combinations of elements of $\{m_i\}_{i \in I}$, so that when writing:

$$\sum_{i \in I} r_i m_i$$

only finitely many r_i are nonzero. If φ is injective, then $\sum_{i \in I} r_i m_i = 0$ if and only if $r_i = 0$ for all $i \in I$. Sometimes to emphasize the fact that it's ordered, a basis is called an *ordered basis*. In terms of indexing sets, the ordered elements are generalized to elements having different indexes. Notice how being linearly independent is different from proposition 12.3.3(3), where if $x \in \oplus_{i \in I} N_i$ implies there exists unique elements $v_i \in N_i$ such that $x = \sum_i v_i$. We are adding another requirement on $\bigoplus_i v_i$, namely that the function φ in the above definition is injective.

A module M having a basis is in fact a very restricting condition:

Lemma 13.1.2: Basis iff Free

Let M be an R-module. Then M is free if and only if it admits a basis, in paritcular $B \subseteq M$ is a basis if and only if the natural homomorphism $R^{\oplus B} \to M$ is an isomorphism

Proof **:**
This is immedaite from the definition: if $B \subseteq M$ is linearly independent and generates M, then $\varphi : R^{\oplus B} \to M$ is injective and surjective, and if $\varphi : R^{\oplus B} \to M$ is an isomorphism, then $\varphi(B)$ is a subset of M which generates it and is linearly independent.

Example 13.1.2: Linearly Dependent and Independent Sets

1. Let $M = \mathbb{Z}/n\mathbb{Z}$ be a $\mathbb{Z}$-module. Then:

$$n \cdot [1]_n = 0$$

but naturally $n \neq 0$. More generally, if R is a ring with a zero divisor, and $(z) \subseteq R$ is the sub-module generated by the zero-divor (z), then if $zz' = 0$, $z' \cdot z = 0$,showing that the singleton z is linearly dependent.

2. Let $R = \mathbb{C}[x, y]$ be a module over itself. This ring has no zero divisors, hence the above does not apply. However, that there are still linearly dependent elements, for example:

$$y \cdot x + (-x) \cdot y = xy - xy = 0$$

showing that x, y are *not* linearly independnet over $\mathbb{C}[x, y]$. However, if we consider $\mathbb{C}[x, y]$ over $\mathbb{C}$, then $\{x, y\}$ *are* linearly independent, since $\mathbb{C}[x, y]$ is the free $\mathbb{C}$-algebra. Hence:

$$F^{\mathbb{C}}(\{x, y\}) = \mathbb{C}^2 \to \mathbb{C}[x, y]$$

is an injective function. However, $\{x, y\}$ is not a basis since the above homomorphsim is not surjective, since nothing maps to xy. Is there any finite that will create a surjective map?

3. If R is any ring (which includes rings with torsion elements), then $(1) = R$, and $r \cdot 1 = r1$ if and only if $r = 0$. Hence,

$$R \mapsto R$$

is an injective homomorphism (even surjective, making it a basis). This shows that the choice of subset of the module effects whether the set if linearly independent or a basis.

Though R has a basis when over itself, show that $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z}$ over $\mathbb{Z}$ does *not* have a basis, or even a linearly independent set.

As we saw in the above example, some modules have linearly independent sets, some don't, and others have a basis. By a simple applications of Zorn's lemma and some manipulation of definition, we can show that every module has a maximally linearly independent set:

Lemma 13.1.3: Maximally Linearly Independent Set

Let M be an R-module and let $S \subseteq M$ be a lienarly independnet set. Then there exsits a maximal linearly independent subset of M containing S.

Proof **:**
This is an application of Zorn's lemma: Let $\mathcal{L}$ be the collection of linearly independent subset of M containing S ordered by inclusion. Since $S \in \mathcal{L}$, $\mathcal{L} \neq \emptyset$. If we take any chain $\mathcal{C} \subseteq \mathcal{L}$, and take $C = \cup\mathcal{C}$, then C is linearly independent, since it it were linearly dependent, then only finitely many linear element would be used in the equation, but then they all belong to some $C_i \in \mathcal{C}$, which by assumption is linearly independnet. Hence, by Zorn's lemma $\mathcal{L}$ contains a maximally linearly independent set containing S.

37: Trivia

The Axiom of Choice and the above lemma are in fact equivalent, hence it is impossible to prove the above statement without using Zorn's lemma.

If S is a maximally linearly independent set containing $\emptyset$, we shall simply say that S is maximally linearly independent. Note that being maximally linearly independent does not guarantee a unique size or that it generates the entire module, for exmaple $\{2\} \subseteq \mathbb{Z}$ is maximally linearly independent but does not generate $\mathbb{Z}$. In the special case where $R = k$ is a field so that we have a vector-space V, then a maximally linearly independent set is indeed a basis!

Theorem 13.1.4: Vector Space has Basis

Let V be a vector space over k and let B be a maixmally lienarly independnet set. Then B is a basis of V, hence V is a free k-module. Furthermore, if S is a liearly independent set, then S is contained in a basis B.

Proof **:**
We need to show that B generates V. Let $v \in V$, $v \notin B$. Then by maximality of B, $B \cup \{v\}$ is linearly dependent, hence there exists $v_1, v_2, ..., v_n$ and $c_0, c_1, ..., c_n$ such that

$$c_0 v + c_1 v_1 + \cdots + c_n v_n = 0$$

with not all $c_i = 0$. Note that $c_0 \neq 0$ or else we would get $c_1 v_1 + \cdots + c_n v_n = 0$ with not all $c_i \neq 0$, contradicting linearly independence of elements in B. Since k is a field, c_0 is a unit, hence:

$$v = (-c_0^{-1} c_1) v_1 + \cdots + (-c_0^{-1} c_n) v_n$$

But then, $v \in \operatorname{span}(B)$, showing that B generated V. By lemma 13.1.2, V is a free k-module.

Furthermore, if S is lienarly independent, then S is contained in a maximally linearly independent set, say B. But by our above reasoning B is a basis, hence S is contained in a basis, as we sought to show.

Hence, every vector-space is free. Before continuing our exploration of bases, a quick comment must be made: we can define the dual notion of *minimally spanning set* and have proved the same statment in the reverse direction namely:

Proposition 13.1.5: Vector space, Minimally Spaning then Basis

Let V be a k-vector space and let B be a minimally spanning set. Then B is a basis for V, making V a free k-module. Furthermore, if S is any spanning set,then S containins a minimal spanning set B.

Proof **:**
exercise

Corollary 13.1.6: Subspaces Of Vector-Spaces Are Free

Let V be a vector-space and $W \subseteq V$ be a subspace. Then W is free

Proof **:**
Let $W \subseteq V$ be a subspace. Since W is a k-module, by theorem 13.1.4, it is free.

Note that this is not true for arbitrary free-modules: there exists submodule of free modules which are *not* free[3]. In fact, since any ring R is a module over itself, then it is easy to find counter-examples; any polynomial rings over several variables will have non-free submodules, see exercise ref:HERE. Another simpler example is as follows

Example 13.1.3: Non Free Submodule
Take $\mathbb{Z}_6$ over itself, and consider the ideal (2). Show that it is not free[a].

[a]Hint: it's torsion

In later chapters, we shall see that submodules of free modules over slightly more general rings, namely PIDs, will still be free[4], however this generlization is limited in scope and does not further generlize over more general rings, as we shall discuss later.

The key in theorem 13.1.4 is that every element in a field is a unit; commutativity was not important. Thus, this same proof can work for division rings more generally[5]. Though this set is the maximal linearly independent set, since the partial order on $P(A)$ forms a lattice, there may be many different cardinalities of sets which form maximal generating sets (see exercise ref:HERE for an example where this happens for some rings and groups). We thus ask the question of when do we get bases with different cardinalites, or if there are conditions for which different basis always have the same cardinality. For vector-spaces, every basis *always* has the same cardinality! The following theorem is the key result for the proof:

[3]Though it is true for free groups, see appendix B
[4]see corollary 14.1.3
[5]And will be used in chapter 24, section 23

Theorem 13.1.7: Replacement Theorem

Assume $A = \{a_1, a_2, \cdots, a_n\}$ is a basis for V containing n elements and $B = \{b_1, b_2, ..., b_n\}$ is a set of linearly independent vectors in V. Then there is an order $a_1, a_2, ..., a_n$ st for each $k \in \{1, 2, ..., m\}$, the set $\{b_1, b_2, ..., b_k, a_{k+1}, a_{k+2}, ..., a_n\}$ is a basis of V. In other words, the elements $b_1, b_2, ..., b_m$ can be used to successively replace the elements of the basis A, and A will still be a basis.

***Proof* :**
Let the set A and B be as inthe statement of the theorem. We proceed with induction on k. If $k = 0$, we're done, so let's say $\mathcal{B}$ is a basis for V and

$$(b_1, b_2..., b_k, a_{k+1}, ..., a_n)$$

is a basis for V. Since it's a basis, it's a spanning set, and so for b_{k+1}:

$$b_{k+1} = \beta_1 b_1 + \cdots + \beta_k b_k + \alpha_{k+1} a_{k+1} + \cdots + \alpha_n a_n \tag{13.2}$$

Notice not all the α_i can be zero, since then b_{k+1} is a linear combination of elements in B, but B is linearly independent. Without loss of generailty, we can re-arrange the above equation so that $\alpha_{k+1} \neq 0$. Then re-arranging, we get:

$$a_{k+1} = \frac{1}{\alpha_{k+1}}(\beta_1 b_1 + \cdots + b_{k+1} + \alpha_{k+2} a_{k+2} + \cdots + \alpha_n a_n)$$

Thus, we get that:

$$\text{span}(\{b_1, ..., b_k, a_{k+1}, ..., a_n\}) = \text{span}(\{b_1, ..., b_{k+1}, a_{k+2}, ..., a_n)$$

it not suffice to show that that $\{b_1, ..., b_k, a_{k+1}, ..., a_n\}$ is linearly independent. Let's say:

$$\beta_1 b_1 + \cdots + \beta_{k+1} b_{k+1} + \alpha_{k+2} a_{k+2} + \cdots + \alpha_n a_n = 0 \tag{13.3}$$

By equation (13.2), we can make equation (13.3) a linear combination of elements of $(b_1, b_2..., b_k, a_{k+1}, ..., a_n)$:

$$(\beta_1 + \beta_{k+1}\beta_1')b_1 + \cdots + (\beta_k + \beta_{k+1}\beta_k')b_k + (\beta_{k+1}\alpha_{k+1}')a_{k+1} + \cdots + (\alpha_n + \beta_{k+1}\alpha_n')a_n = 0$$

In particular, we can multily this entire equation by $\beta_{k+1}(\alpha_{k+1}')^{-1}\beta_{k+1}^{-1}$ [a] so that the coefficient of α_{k+1} is β_{k+1}. Since this is a basis by the induction hypothesis, all the coefficients must be zero, including β_{k+1}. Thus, we get

$$\begin{aligned} 0 &= \beta_1 b_1 + \cdots + 0 b_{k+1} + \alpha_{k+2} a_{k+2} + \cdots + \alpha_n a_n \\ &= \beta_1 b_1 + \cdots + \beta_k b_k + \alpha_{k+2} a_{k+2} + \cdots + \alpha_n a_n \\ &= \beta_1 b_1 + \cdots \beta_k b_k + 0 a_{k+1} + \alpha_{k+2} a_{k+2} + \cdots + \alpha_n a_n \end{aligned}$$

where the last line is a basis by the induction hypothesis, and so all other coefficients must be zero, and so the set must be linearly independent, as we sought to show.

[a] The reason for writing the scalar like this is so show that the proof works even if we limit ourselves to division rings

Since the cardinality of the basis is independent of the choice of elements that make the basis, we can define it as an invariant on vector-spaces:

Definition 13.1.8: Dimension

Let V be a vector space over F. Then the cardinality of the basis is called the *dimension* of V, and is denoted $\dim_F(V)$ or $\dim(V)$ if context permits. If V is not finitely generated, then we say that V is infinite dimensional, and we write $\dim V = \infty$.

The key take-away of the Replacement Theorem is that there is a direct relation in cardinality between sets that are linearly independent and sets that are spanning:

Corollary 13.1.9: Steinitz exchange lemma

1. Suppose V has a finite basis with n elements. Any set of linearly independent vector has less than or equal to n elements. Any spanning set has greater than or equal to n elements
2. If V has some finite basis then any two bases of V have the same cardinality

Proof **:**

1. A be the basis. If a set is linearly independent, then we can use the replacement theorem to systematically replace the basis with the elements of the linearly indepennet set. Thus, it cannot have more than n elements, since a basis has maximum n elements that are linearly independent. If a set is a spanning set, then by corollary 13.1.9, the set contains a basis, say B. Thus, $|B| \leq |A|$.

   Since B is a linearly independent, we can use the replacement theorem to replace the elements of A with B. If $|B| < |A|$, that would contradict the minimality of the set A since A is a basis, and hence a minimal spanning set. Thus, it must be that $|A| \leq |B|$.
2. This is an immediate consequence of the first part, since both both conditions have to match.

As we saw in example 13.1.2, a singleton set with a single element which has torsion is always linearly dependent. In an integral domain, we have some more power:

Corollary 13.1.10: Module Linearity

Let R be an integral domain and let M be a free R-module of rank $n < \infty$. Then any $n+1$ elements of M are R-linearly dependent, i.e., for any $y_1, y_2, ..., y_{n+1} \in M$, there are elements $r_1, r_2 ..., r_{n+1} \in R$ not all nonzero such that

$$r_1 y_1 + r_2 y_2 + \cdots + r_{n+1} y_{n+1} = 0$$

Proof **:**
There is a trick similar to Gauss's lemma. We can embed R in its quotient field (since R is an integral domain, this is indeed an embedding), say k is its quotient field. Since M is a free module,

$$M \cong \underbrace{R \oplus R \oplus \cdots R \oplus R}_{n \text{ times}}$$

and so

$$M \subseteq \underbrace{k \oplus k \oplus \cdots \oplus k}_{n \text{ times}}$$

Since the latter is an n dimensional vector space over F, so any $n+1$ dimensional elements of M are F-linearly dependent. By clearing the denominators of the scalars (ex. multiplying through by the product of all the denominaotrs), we get an R-linear dependence realtion among $n+1$ elements of M.

An important consequence of the existence of a basis is that linear transformation $f : V \to W$ can be defined by first determining where a basis of V maps to, and then extending linearly (that is, if $v = \sum_{\beta_i \in \mathcal{B}} b_i \beta_{\in} W$, then $f(v) = f\left(\sum_{\beta_i \in \mathcal{B}} b_i \beta_i\right) = \sum_{\beta_i \in \mathcal{B}} b_i f(\beta_i))$. Note however that the choice basis is important, we shall get back to this in section 13.2, and that vector spaces may have many different bases (of the same cardinality by the replacement theorem):

Example 13.1.4: Different Basis
As we've defined in example 12.3.1the set

$$e_i = (\underbrace{0, ..., 0}_{i-1 \text{ times}}, 1, \underbrace{0, ..., 0}_{k-i \text{ times}}) \qquad 1 \leq i \leq n$$

is a basis for $\mathbb{Q}^n$. There are countably many different basis for $\mathbb{Q}^n$; can you think of a few? If you know about inner product spaces, can you think of an orthonormal [hamel][a] basis?

[a]The term orthonormal basis is usually reserved for *countable* linear combination for *Hilbert space*. Since the linear combination is not finite, it requires some analysis to define; this would be covered in a text on *Functional Analysis*. If you know about Hilbert spaces, then taking $\mathbb{C}[x]$ as the closure of $\mathbb{C}[x]$, is there an orthonormal basis for it?

13.1.1 Quotient Spaces and Subspaces

Now that we've established the struture of a vector-space, we can some questions about subspaces the structure of quotient spaces, with particular interest in finite vector-spaces to be able to take advantage of counting arguments. We already saw that subspace of vector spaces are free k-module.

Proposition 13.1.11: Dimension Of Quotient Space

Let V be a vector space over k and let W be a vector space. Then V/W is a vector space with $\dim V = \dim W + \dim V/W$. If one term is finite, all terms are finite

Proof **:**
Certainly V/W is a vector space being a quotient k-module. How would you prove $\dim V = \dim W + \dim V/W$? Why should this be intuitive?

Naturally, since we are working with quotients, we have an associated linear transformation The following is the "homomorphism-equivalent" of the previous proposition:

Corollary 13.1.12: Rank-Nullity Theorem

Let $\varphi : V \to W$ be a linear transfomration of vector spacees over k and $\dim(V) < \infty$. Then $\ker \varphi$ is a subspace of V, $\varphi(V)$ is a subpace of W, and

$$\dim_k(V) = \dim_k(\ker(\varphi)) + \dim_k(\varphi(V))$$

Proof **:**
follows from previous theorem

Since the dimensions are of the subspaces are invariant to basis, we can give a name to the dimension of the image and null space of a linear transformation

Definition 13.1.13: Rank And Nullity

If $\varphi : V \to W$ is a linear transformation of vector spaces over k, $\ker(\varphi)$ is called the *null space* of φ ,and the dimension of $\ker(\varphi)$ is called the *nullity* of φ. The dimension of $\varphi(V)$ is called the *rank* of φ. If $\ker(\varphi) = 0$, the transformatoin is said to be *nonsingular*.

The reasoning behind the name *nonsingular* is due to the fact that very often in (differentil) geometry, we associated to shapes a vector-space at each point (such an object is called a *vector bundle*) and if the curve "drops rank" this would correspond to a particular function having $\dim_k(\ker(\varphi)) > 0$[6]. For example, the curve $x^3 = y^2$ "drops rank" at $(0, 0)$.

Corollary 13.1.14: Properties of [finite dim.] Linear Transformations

Let $\varphi : V \to W$ be linear transfomration of vector spaces of the same finite dimenison. Then the following are equivalent

1. φ is an isomorphism
2. φ is injective, i.e., $\ker(\varphi) = 0$
3. φ is surjective, i.e. $\varphi(V) = W$
4. φ sends a basis of V to a basis in W

Notice that this proof is saying that if $f : V \to W$ is a k-homomorphism where $\dim_k(V) = \dim_k(W) = n$, then given f is injective, it is surejctive. This fails to be the case even for maps between free moduels over PIDs, for exmaple $f : \mathbb{Z} \to \mathbb{Z}$ mapping $z \mapsto 2z$ is an injective $\mathbb{Z}$-module homomorphism

[6] In particular, if $T_p f$ is the tangent map at p, then $T_p f$ would drop rank

between free $\mathbb{Z}$-modules, but is certainly not surjective! This also fails for infinite dimensional vectors spaces as will be demonsarted by a counter-example after the proof. Hence, the fact that it's sufficient to check injectivity shows how linear maps between finite vector-spaces are akin to maps between finite sets[7].

***Proof* :**
Counting dimension arguments

Note that finiteness of dimension is important in the previous corrollaries: these need not hold in infinite dimension

Example 13.1.5: Infinite Dimensinoal Linear Transformations
Take $V = \ell^2$, the set of all infinite sequences of real numbers satisfying the property:

$$\left(\sum_{i\in\mathbb{N}}^{\infty} |x_i|^2\right)^{1/2} < \infty$$

Then define $f : V \to V$ via:

$$(x_1, x_2, ..., x_n, ...) \mapsto (x_1, \frac{x_2}{2}, ..., \frac{x_n}{n})$$

Then this function is injective, but not surjective. Even more is true; V and f belong to some of the nicest infinite dimensional spaces and functions (V is a Hilbert space and f is a bounded self-adjoint linear operator), and hence this property is not easily rectified.

We now take a look at subspaces. We have already established that all subspaces of a vector-spaces are free. Another importnant property is that every subspace always has a *compliment*, that is if $W \subseteq V$, then there exists a $U \subseteq V$ such that $U \cap W = \{0\}$ and $W + U = V$. This is certainly *not* true for general R-modules, simply let $M = \mathbb{Z}/4\mathbb{Z}$ be a $\mathbb{Z}$-module and take $\{0, 2\} \subseteq \mathbb{Z}/4\mathbb{Z}$. Then we would require $U = \{0, 1, 3, 4\}$, which is not a subgroup. In fact, *no* finite abelian subgroup has this property[8]. However, we may always find such a compliment for subspaces of vector-spaces, given the space is finite dimensional:

Proposition 13.1.15: Existence Of Compliment

Let V be a finite dimensional vector space and $W \subseteq V$ be a subspace. Then there exists a subspace $U \subseteq V$ such that

1. $W \cap U = \{0\}$
2. $W + U = V$

If these two conditions are satisfied, we will write $W \oplus V$ and call it the direct sums (just like in definition 12.3.2).

[7] In fact, this parallels is runs more deeply than it may at first seem, see a text on K-theory for some interesting elaboration on this notion

[8] More particularly, no torsion $\mathbb{Z}$-module has this property. In the Category of **Grp**, there are finite groups that "spilt". See section 16.2 for a generalized notion of splitting and section 5.4.2 for identifying groups that split

Proof **:**
Let $U \subseteq V$ be a subspace. Choose basis $\{e_1, ..., e_k\}$. Then this basis can be extended to a basis $\{e_1, ..., e_k, ..., e_n\}$ for V. Define the map $\pi_U : V \to V$ given by:

$$\sum_i^n a_i e_i \mapsto \sum_i^k a_i e_i$$

This map is certainly a linear transformation, and $\text{im}(\pi) = U$. The kernel of this map is:

$$\ker(\pi_U) = \langle \sum_{i=k+1}^n a_i e_i | a_i \in k \rangle$$

Notice that $v \in \ker(\pi_U)$ if and only if $v \in \text{im}(I - \pi_U)$, namely:

$$v \in \ker(\pi_U) \Leftrightarrow \pi_U(v) = 0 \Leftrightarrow v - \pi_U(v) = v \Leftrightarrow v \in \text{im}(I - \pi_U)$$

I claim that $v = \pi_U(v) + (I - \pi_U)(v)$. Indeed if $v = \sum_i a_i e_i$:

$$\sum_i^n a_i e_i = \sum_i^k a_i e_i + \sum_{i=k+1}^n a_i e_i = \pi_U(v) + (I - \pi_U)(v)$$

and furthermore it is clear that $\ker(\pi_U) \cap \text{im}(\pi_U) = 0$, hence setting $W = \ker(\pi_U)$ and $V = \text{im}(\pi_U)$ we get:

$$V = \ker(\pi_U) \oplus \text{im}(\pi_U) = U \oplus W$$

as we sought to show.

The vector space being finite dimensional is key to the proof, as this does *not* generalize to general infinite dimensional vector-spaces (it can be thought of as characterizing *Hilbert spaces*).

Corollary 13.1.16: Finite Generation of Subspace

Let $W \subseteq V$ be a subspace of a finite dimensional vector space. Then W is finitely generated

Proof **:**
use proposition 13.1.15.

The map used in the above proof is important enough to be given a name:

Definition 13.1.17: Projection Map

Let $\pi : V \to V$ be a linear transformation $U \subseteq V$. Given $V = U \oplus W$, then the *projection map* (with respect to U) is the map $\pi_U : V \to V$ such that $\pi(v) = \pi(u + w) = u$.

Notice $\pi \circ \pi = \pi$, that is, applying the map twice gives back the same result: $\pi(\pi(v)) = \pi(v)$. This in fact characterizes such a map:

Proposition 13.1.18: Characterizing Projection Maps

Let π be linear operator over a [not necessarily finite dimensional] vector space such that $\pi \circ \pi = \pi$. Then π is a projection map.

***Proof* :**
Let $U = \text{im}(\pi)$ and $W = \text{im}(I - \pi)$. Then for any $v \in V$:

$$\pi(v) + v - \pi(v) = v$$

hence $U + W = V$. If $x \in U \cap W$. Since $x \in \text{im}(\pi)$, $x = \pi(y)$ for some $y \in V$, and so:

$$\pi(x) = \pi(\pi(y)) = \pi(y) = x$$

hence, $\pi(x) = x$. Then:

$$\begin{aligned}
x &= \pi(x) && x \in U \\
&= (I - \pi)(x) && x \in W \\
&= \pi((I - \pi)(x)) && (I - \pi)(x) = x \in U \\
&= \pi(x - \pi(x)) \\
&= \pi(x - x) && \pi(x) = x \\
&= \pi(0) \\
&= 0
\end{aligned}$$

Hence $x = 0$, and so $U \cap W = \{0\}$. With the above, we get that

$$V = U \oplus W = \ker(\pi) \oplus \text{im}(\pi)$$

as we sought to show.

Notice how the above proof did not rely on any structure of V (notably, it wasn't important that V had a basis). Thus, if $\varphi : M \to M$ is any R-module homomorphism such that $\varphi \circ \varphi = \varphi$, then

$$M = \ker(\varphi) \oplus \text{im}(\varphi)$$

For this reason, it is much more common to define projection maps as maps that satisfy the property of $\varphi^2 = \varphi$ (i.e., they are *idempotent*).

Exercise 13.1.1

1. Consider $\mathbb{R}[x]$ as an $\mathbb{R}$ module. Take $\mathbb{R}_n[x]$ to be all polynomial where $\deg(p(x)) \leq n$. Show that (evalutaion polynomials has $\left\{(x-a), (x-a)^2, ..., (x-a)^n\right\}$ as a basis). (This might at seem counter-intuitive at fisrt after doing some rings, but it's really cool!)
2. Let R be a commutative ring that is *not* a field. Show that not all R-modules are free, i.e. find a counter-example [Hint: Look back at the examples given in section 12.1]
3. (More challenging) Let R be any ring (not necessarily commutative). Show that if any left (or right) R-module is free, then R is a divison ring.

4. Show that $(x, y) \subseteq \mathbb{C}[x, y]$ is not a free $\mathbb{C}[x, y]$-submodule.
5. Show that any uncountable set of $\mathbb{C}[x]$ is linearly dependent. Any uncountable subset of $\mathbb{C}[x]$ is linearly dependent,
6. (Right shfit operator)

A lot of exercices in linear algebra are about determining if a set is linearly independent, spanning, (more?). These exercices are to practice these skills:

13.2 Linear Transformations and Matrix Representation

As mentioned at the beginning of the chapter, k-module homomorphisms are called linear transformation in the theory of vector-spaces. In this section, we will show how every linear transformation between finite dimensional vector-spaces can be represented as a matrix. Since all vector-spaces are free with an invariant basis number , then when defining a linear transformation, we only need to be concerned with where the basis get's mapped to. If $\varphi : V \to W$, where V and W are finite dimensional vector-spaces over k, and $\varphi \in \mathrm{Hom}_k(V, W)$. Let $\beta_V = \{v_1, v_2, ..., v_n\}$ and $\beta_W = \{w_1, w_2, ..., w_m\}$ be (ordered) bases for V and W respectively. Then we are only concerned where the v_i's map to, that is, what $\varphi(v_j) = w_j$ is. Since they map into the vector-space W, $w_j = \sum_{i=1}^m b_{ij} w_i$, $b_{ij} \in k$,

$$\varphi(v_j) = \sum_{i=1}^{m} b_{ij} w_i$$

Let $M_{\beta_V}^{\beta_W}(\varphi)$ be the $m \times n$ matrix whose ij entry is b_{ij} (i.e. the coefficients of w_i given v_j). This matrix is given a name:

Definition 13.2.1: Matrix Representation

Let $\varphi : V \to W$ be a linear transformation between two finite dimensional vector spaces with dimension n and M and basis β_V and β_W respecitvely. Then the *matrix representation* $M_{\beta_V}^{\beta_W}$ of φ (also written as $M_{\beta_V}^{\beta_W}(\varphi)$) is

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

where each entry ij is the coefficient of the ith basis element of β_W in the linear combination of where the element v_j maps.
If a matrix A is a matrix representation of a linear transformation φ, we say that A *represents* φ *with respect to the basis* β_V, β_W. Similarly, φ would be the linear transformation represented by A with respect to the bases β_V, β_W.

With the matrix representation of φ, we can re-write $\varphi(v) = w$ as

$$\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{13.4}$$

where $v = \sum_{i=1}^{n} a_i v_i$ and $\varphi(v) = \sum_{i=1}^{m} b_i w_i$. Sometimes, the following notation is used as a shorthand for (13.4): $[\varphi(v)]_{\beta_W} = M_{\beta_V}^{\beta_W} [v]_{\beta_V}$.

Example 13.2.1: Matrix Representation

1. Let $V = \mathbb{R}^3$, $W = \mathbb{R}^2$, and take $\beta_V = \{(1,0,0),(0,1,0),(0,0,1)\}$ and $\beta_W = \{(1,0),(0,1)\}$ (i.e., the standard basis for $\mathbb{R}^3$ and $\mathbb{R}^2$). Let

$$\varphi(x,y,z) = (4x + 2z, x - y + 5z)$$

Then $\varphi(1,0,0) = (4,1)$, $\varphi(0,1,0) = (0,-1)$, and $\varphi(0,0,1) = (2,5)$. Thus, the matrix representation of φ with respect to β_V and β_W is

$$\begin{pmatrix} 4 & 0 & 2 \\ 1 & -1 & 5 \end{pmatrix}$$

Theorem 13.2.2: Matrix Representation Correspondence

Let V and W be a finite dimensional vector space over F with dimension n and m, as well as basis β_V and β_W respectively. Then the map $\Psi : \mathrm{Hom}_F(V,W) \to M_{mn}(F)$, where $\varphi \mapsto M_{\beta_V}^{\beta_W}(\varphi)$ is a [vector-space] isomorphism.

Note that the operation on $\mathrm{Hom}_F(V,W)$ if point-wise function addition and $M_{mn}(F)$ is matrix addition.

Proof **:**
To show Ψ is F-linear, it comes down to the fact that every $\varphi \in \mathrm{Hom}_F(V,W)$ is F-linear. Since every column of $M_{\beta_V}^{\beta_W}$ represents $\varphi(v_i)$, and $(\varphi + \psi)(v) = \varphi(v) + \psi(v)$, then we get that the map Ψ is also F-linear.

The map Ψ is surjective, since given a matrix $A \in M_{mn}(F)$, the map φ that maps to it can be defined by reversing the process that was done in equation (13.4). This map is injective since if $A = B$, then $A = M_{\beta_V}^{\beta_W}(\varphi) = M_{\beta_V}^{\beta_W}(\psi) = B$. Since a basis uniquely defines a linear transformation and both representations are with respect to the same bases, it must be that $\varphi = \psi$, proving injectivity.

Thus, Ψ is an isomorphism, as we sought to show.

Note: Different choices of bases produces different isomorphisms. There is no "natural choice" in the same way that there is no natural choice for a basis for a vector-space

Corollary 13.2.3: Dimension Of $\mathrm{Hom}_F(V, W)$

The dimension of $\mathrm{Hom}_F(V, W)$ is $\dim(V) \cdot \dim(W)$

Proof **:**
Since $\mathrm{Hom}_F(V, W)$ is isomorphic to $M_{mn}(F)$, and $\dim(M_{mn}(F)) = mn$, then $\dim(\mathrm{Hom}_F(V, W)) = mn = \dim(V) \cdot \dim(W)$

We've shown that function addition is compatible with matrix addition. When it comes to function composition, the story becomes a little trickier, since we need the co-domain of one function to line-up with the domain of the next. Thus, to that end, let U, V, W be finite dimensional vector-spaces with dimension k, n, m and bases $\beta_U = \{u_1, u_2, ..., u_k\}$, $\beta_V = \{v_1, v_2, ..., v_n\}$, and $\beta_W = \{w_1, w_2, ..., w_m\}$ respectively. Let $\varphi : U \to V$ and $\psi : V \to W$ be two linear transformations. Their composition $\psi \circ \varphi$ is a linear fransformation from U to W, and so there exists a representation $M_{\beta_U}^{\beta_W}(\psi \circ \varphi)$ given by

$$(\psi \circ \varphi)(u_j) = \sum_{i=1}^{m} c_{ij} w_i \qquad c_{ij} \in F$$

Since φ and ψ are linear transfomation, they too have a matrix representation:

$$\varphi(u_j) = \sum_{p=1}^{n} a_{pj} v_p \qquad \psi(v_p) = \sum_{i=1}^{m} b_{ip} w_i$$

from this, we can find c_{ij} in terms of the coefficients a_{pj} and b_{ip}:

$$\begin{aligned}
(\psi \circ \varphi)(u_j) &= \psi\left(\sum_{p=1}^{n} a_{pj} v_p\right) \\
&= \sum_{p=1}^{n} a_{pj} \psi(v_p) \\
&= \sum_{p=1}^{n} a_{pj} \sum_{i=1}^{m} b_{ip} w_i \\
&= \sum_{p=1}^{n} \sum_{i=1}^{m} a_{pj} b_{ip} w_i \\
&= \sum_{i=1}^{m} \sum_{p=1}^{n} a_{pj} b_{ip} w_i
\end{aligned}$$

Thus, if we let $c_{ij} = \sum_{p=1}^{n} a_{pj} b_{ip}$, we get

$$= \sum_{i=1}^{m} c_{ij} w_i$$

Take a moment to look over this and think about how this would be represented as multiplying the matrix $M^{\beta}_{\beta_U} V(\varphi)$ and $M^{\beta_W}_{\beta_V}(\psi)$. In fact, the coefficents of $M^{\beta_W}_{\beta_U}(\psi \circ \varphi)$ are exactly what you get by multiplying the other two matrices! This is summed up in the following theorem:

Theorem 13.2.4: Matrix Representation And Composition

Let U, V, W be finite dimensional vector-spaces with bases β_U, β_V, β_W, of size k, m, n respectively. Let $\varphi : U \to V$ and $\psi : V \to W$ be two linear transformations. Then

$$M^{\beta_W}_{\beta_U}(\psi \circ \varphi) = M^{\beta_W}_{\beta_V}(\psi) M^{\beta_V}_{\beta_U}(\varphi)$$

Another way of seeing this theorem is by saying the following diagram commutes:

$$\begin{tikzcd}
M_{mp}(F) \times M_{mp}(F) \arrow[r] \arrow[d] & M_{mn}(F) \arrow[d] \\
\mathrm{Hom}_F(F^p, F^m) \times \mathrm{Hom}_F(R^p, R^m) \arrow[r] & \mathrm{Hom}_F(F^n, F^m)
\end{tikzcd}$$

Corollary 13.2.5: Matrix Multiplication Properties

Let $M_{mn}(F)$ be the set of $m \times n$ matrices over F. Then given appropriate matrices, matrix multiplication is associative and distributive

Note how this is different from the associativity and disctributivity of $GL_n(F)$. Namely, all the matrices of $GL_n(F)$ are square matrices. We have not yet proven that non-square matrices are associative and distributive, particularly because they do not always line up properly, and their result can pass from one set of matrices to another (ex. $A \in M_{ab}(F)$, $B \in M_{bc}(F)$, $AB = M_{ac}(F)$)

Proof **:**
We'll show the proof for associativity – the proof for distribuity uses the same technic. Let A, B, C be matrices such that $(AB)C$ and $A(BC)$ are defined. By theorem 13.2.4, (AB) represents $S \circ T$ for appropriate linear transformation S, T. Multiplying on the right side by C, $(AB)C$, the resulting matrix represents $(S \circ T) \circ R$ for appropriate R. Similalry, $A(BC)$ is represented by $S \circ (T \circ R)$. Since the composition of functions is assosiative, Then

$$\Psi((AB)C) = (S \circ T) \circ R = S \circ (T \circ R) = \Psi(A(BC))$$

and since Ψ is injective, $(AB)C = A(BC)$. A similar proof follows for distributivity.

Corollary 13.2.6: Matrix Representation Isomorphisms

1. If β is a basis of the n-dimensional vector space V, the map $\varphi \mapsto M_\beta^\beta(\varphi)$ is a ring isomorphism and a vector-space isomorphism of $\mathrm{Hom}_F(V, V)$ (i.e. $\mathrm{End}(V)$) onto $M_n(F)$:

$$\mathrm{End}(V) \cong_{\mathbf{Ring}} M_n(F) \qquad \mathrm{End}(V) \cong_{F\text{-}\mathbf{Vect}} M_n(F)$$

2. If we restrict to automorphisms, then $\mathrm{Aut}(V) \cong GL_n(F)$ as vector-spaces[a]. In particular

$$\mathrm{Aut}_{\mathbf{Vect}}(V) \cong_{F\text{-}\mathbf{Vect}} GL_n(F)$$

[a] $\mathrm{Aut}(V)$ is not a ring

Proof **:**
In both cases, we've already shown that it is isomorphic as F-vector-spaces, so we'll show they're isomorphic as rings

(1) Using corollary 13.2.5, we see that $M_n(F)$ is a ring (since multiplying two $n \times n$ matrices get's us an $n \times n$ matrix). We've shown this map is bijective in theorem 13.2.2. theorem 13.2.4 shows that this map also preserves the ring structure of $\mathrm{End}(V)$, showing that the map is a ring-isomorphism.

(2) This is the same proof as (1) (Dummit addded: since it sends units to units)

Finally, we convert the powerful result of corollary 13.1.14 into for matrices, in particular we showed that if f is injective between two finite dimensional vector space of the same dimension, then it is surjective. This result is important enough that the matrix representation of such an f should be given a name

Definition 13.2.7: Nonsingular Matrix

An $m \times n$ matrix A is called *nonsingular* if $Ax = 0$ with $x \in F^n$ implies $x = 0$

For some intuition beind the name, see remark ref:HERE (text under definition 13.1.13)

Proposition 13.2.8: Nonsingular Square Matrix

Let A be an $n \times n$ matrix. Then A is invertible if and only if A is nonsingular

Proof **:**
If $\dim_k(V) = \dim_k(W)$, then $V \cong_k W$, so without loss of generailty we may say $f : V \to V$ is a linear map from V to itself. Let's say A is invertible and $Ax = 0$. then

$$0 = Ax = A^{-1}Ax = x \quad \Rightarrow \quad x = 0$$

giving us that A is nonsingular.

Conversely, let A be nonsingular. Then A represents some linear transformation $\varphi : V \to V$ for some finite dimensional vector-space V with respect to basis β and γ for the domain and codomain.

Since $\varphi(x) = 0$ if and only if $x = 0$, then φ is injective, and so by corollary 13.1.14, φ is an isomorphism. Thus, there exists a linear transfomration φ^{-1} such that $\varphi^{-1} \circ \varphi = \text{id}$. By using theorem 13.2.2, let B be the representation of φ^{-1} with respect to γ and β. Using theorem 13.2.4 we have

$$AB = M_\beta^\gamma(\varphi)M_\gamma^\beta(\varphi^{-1}) = M_\gamma^\gamma(\varphi \circ \varphi^{-1}) = M_\gamma^\gamma(I) = I$$

thus $AB = I$. similarly $BA = I$. Thus, $B = A^{-1}$, as we sought to show.

(dummit here put's a throw-away definition and comment about the row rank and column rank, and that the rank of φ is the row rank of it's representation, and that the row and column rank match. Once I have a more solid grasp on whats going on here, I'll come back to it; for now my intuition relates this to kernel an co-kernel, or more generally injectivity/surjectivity test)

Exercise 13.2.1

1. We know that in general, if R is a ring and $x \in R$ is not a unit, that doesn't imply that x is a zero-divisor (ex. take the integers). Show that for the matrix ring $M_n(k)$, if $A \in M_n(k)$ is not a unit, then A is a zero-divisor.
2. Show that any polynomial (over a subfield of $\mathbb{C}$) of degree n is uniquely determined by $n+1$ points (You may freely use the fact that a polynomial of degree n can be broken down into a product of n linear terms).
3. (If you know some complex arithmetics) Let $f : \mathbb{C} \to \mathbb{C}$ be a $\mathbb{C}$-linear function. Recall that $\mathbb{C}$ is a 2 dimensional $\mathbb{R}$-vector space, and so f may be represented as a real matrix. Show that if $f(z) = \alpha z = (a + bi)z$, then the matrix representation of f in the standard basis will be of the form:
$$[f] = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

13.2.1 Change of Basis

While representing vectors in a vector space in terms of coordinates simplifies their representation, and linear transformations and their compositions can be seen as matrices and matrix multiplication, we are *always* dependent on a choice of basis (i.e. a choice of isomorphism $\varphi : k^n \to V$). As pointed out earlier, there is no "cannonical" choice for a basis – there are many choies of bases which will produce isomorphic vector spaces[9]. Thus, in this section we will explore how to change from one basis to another to be able to compare representations. We will be working with vector-spaces, but we will only be using properties of free-modules meaning the theory is true for free-modules in general.

Let V be an F-vector space of dimension n. Let β_A and β_B be two basis for V, meaning we have the two isomorphissm

$$F^{\oplus\beta_A} \xrightarrow{\varphi} V \qquad F^{\oplus\beta_B} \xrightarrow{\psi} V$$

Then we will get the following isomorphism:

$$F^{\oplus\beta_A} \xrightarrow{\psi^{-1}\circ\varphi} F^{\oplus\beta_B}$$

[9]There are ways to make it "cannonical" by augmenting the vector-space to have further structure which may naturally induce a basis (for example, an inner product). For such examples, see section 13.5

Then by theorem 13.2.2, there is a matrix $N_{\beta_A}^{\beta_B}$ that corresponds to $(\psi^{-1} \circ \varphi)$. In particular, the jth column of $N_{\beta_A}^{\beta_B}$ is the image of the jth element of A viewed as a column vector in $f^{\oplus \beta_B}$.
If $\beta_A = (a_1, a_2, ..., a_n)$ and $\beta_B = (b_1, b_2, ..., b_n)$ then

$$(\psi^{-1} \circ \varphi)(a_j) = \sum_{i=1}^{n} n_{ij} b_i$$

We give this matrix a name:

Definition 13.2.9: Change Of Basis

Let V be a F-vector space of dimension n, and let β_A and β_B be two bases corresponding to the isomorphisms φ and ψ. Then the *change of basis matrix* is the matrix representation of $(\psi^{-1} \circ \varphi)$.
We can represent this matrix as $N_{\beta_A}^{\beta_B}$ or $N_{\beta_A}^{\beta_B}(V)$

Corollary 13.2.10: Properties Of Change Of Basis

Let $N_{\beta_A}^{\beta_B}$ and $N_{\beta_B}^{\beta_C}$ be two change of basis for a vector-space V. Then

1. $(N_{\beta_A}^{\beta_B}) = N_{\beta_B}^{\beta_A}$
2. $N_{\beta_A}^{\beta_B} N_{\beta_B}^{\beta_C} = N_{\beta_A}^{\beta_C}$

Proof **:**
Writing out the isomorphism explicitely will immediately give the result, and so it is left as an exercise

With the change of basis for a vector space taken care of, we know check how to perform a change of basis for a matrix representation of a linear transformation. The idea behind this comes down to the following visual:

$$\begin{tikzcd}
F^{\oplus\beta_A} \arrow[dr, "\varphi_1"] \arrow[dd, "\nu_{\beta_A}^{\beta_B} = \varphi_2^{-1} \circ \varphi_1"'] & & F^{\oplus\beta_C} \arrow[dl, "\varphi_2"'] \arrow[dd, "\mu_{\beta_C}^{\beta_D} = \psi_2^{-1} \circ \psi_1"] \\
& V \arrow[r, "\alpha"] & W \\
F^{\oplus\beta_B} \arrow[ur, "\psi_1"'] & & F^{\oplus\beta_D} \arrow[ul, "\psi_2"]
\end{tikzcd}$$

Let N_A^B be the matrix representation of $\varphi_2^{-1} \circ \varphi_1$ and M_C^D be the matrix representation of $\psi_2^{-1} \circ \psi_1$. If we choose the basis β_A for V and β_C for W, then we have:

$$\varphi_2^{-1} \circ \alpha \circ \varphi_1 : F^{\oplus\beta_A} \to F^{\oplus\beta_C} \tag{13.5}$$

Let P be the matrix representation of $\varphi_2^{-1} \circ \alpha \circ \varphi_1$. If we choose the basis β_B for V and β_D for W, then we have:

$$\psi_2^{-1} \circ \alpha \circ \psi_1 : F^{\oplus\beta_B} \to F^{\oplus\beta_D} \tag{13.6}$$

now with some manipulation, we can transform equation 13.6 into equation 13.5 like so

$$\psi_2^{-1} \circ \alpha \circ \psi_1 = (\mu_{\beta_C}^{\beta_D} \circ \varphi_2^{-1}) \circ \alpha \circ (\varphi_1 \circ (\nu_{\beta_A}^{\beta_B})^{-1}) = \mu_{\beta_C}^{\beta_D} \circ (\varphi_2^{-1} \circ \alpha \circ \varphi_1) \circ (\nu_{\beta_A}^{\beta_B})^{-1}$$

And using the matrix correspondence, this is equivalent to

$$M_{\beta_C}^{\beta_D} P (N_{\beta_A}^{\beta_B})^{-1} = M_{\beta_C}^{\beta_D} P N_{\beta_B}^{\beta_A} = Q$$

This should be able to be read quite intuitively: $N_{\beta_B}^{\beta_A}$ converts our vector represented by a linear combination of vectors in β_B into linear combination of vectors in β_A, then P transforms the result and outputs it as a linear combination of vectors in β_C, and finally $M_{\beta_C}^{\beta_D}$ converst the result into a linear representation of vectors in β_D. W

We sum up these results into a proposition:

Proposition 13.2.11: Change Of Basis For Linear Transformation

Let $\varphi : V \to W$ be a linear transformation from two finite dimensional vector spaces, and let P be the matrix representation of φ with respect to bassi β_A and β_C for V and W respectively. Then the matrix representing linear transformation φ with respect to basis β_B and β_D is

$$M_{\beta_C}^{\beta_D} P N_{\beta_B}^{\beta_A}$$

with the notation given above.

More generally, if M and N are any invertible matrix, they represent a change of basis, so MPN would be a change of basis (given the appropriate dimensions).

Definition 13.2.12: Equivalence

Let $P, Q \in M_{mn}(F)$. Then we say P and Q are *equivalent* if they represent the same linear transformation $\varphi : V \to W$ up to a chocie of basis. If that's the case, there exists matrices M and N such that

$$Q = MPN$$

It should be shown that equivalence is an equivalence class, and hence there is an entire equivalence class of matrices that can represent a linear transformation, where any representative of the equivalence class equates to a choice of basis for the domain and codomain. This brings up an interesting question: is there a particular representative (i.e. choices of bases for P, Q) that will make it "easier" to see what the linear transformation is doing? This is a very interesting question, and in the next section shall be explored to great success, in particular, all linear transformation can be represented in "reduced form" and which can be algorithmically found.

13.2.2 Similarity

Before proceeding, there is one more important concept that must be covered, namely the special case where $V = W$ is the same vector-space:

$$\mathrm{Hom}_R(V, V) = \mathrm{End}_R(V)$$

that is, we shall look at linear transformations from V to itself. Such endomorphisms are in fact extremely common; they can be thought of as representing the "change in state" of a system. As the reader may surmise, this means they have applications all other the physical science, however they shall amply show up in pure mathematics. In fact, chapter 14 is essentially the theoretical build-up for studying such functions.

For this discussion, we shall tacitly assume $R = k$ is a field, however for most of the discussion having R simply be an integral domain shall work. The reader should keep in mind though that many results work in broader generality want to make this more the responsibility of the author rather than the reader

We shall simply start by asserting that the vector-space V in $\mathrm{Hom}_R(V, V)$ shall be considered to be the exact same space, not two isomorphic copies of these spaces as generally is considered. In this situation, an element $\alpha \in \mathrm{Hom}_R(V, V) = \mathrm{End}_R(V)$ is usually called a *linear operator* to emphasize the fact that the domain and codomain are "identical". With this in mind, there are new questions that could be asked, such as if there exists $v \in V$ such that $\alpha(v) = \lambda v$ for some $\lambda \in R$, or if a subspace $W \subseteq V$ maps into itself.

Another important consequence of identifying the domain and codomain is that when choosing a basis, the *same* basis must be chosen for both of them (as α is considered to be mapping V to itself). This means that there is a more specific notion of equivalence as only one matrix is chosen:

Definition 13.2.13: Similarity

Let $A, B \in M_n(R)$. Then these two matrices are *similar* if they represent the same linear operator $\alpha : V \to V$ up to choice of basis for V. Equivalently, two endomorphism α, β are similar if there exists an automorphism φ such that $\beta = \varphi \circ \alpha \circ \varphi^{-1}$
Similarity is an equivalence relation[a], and so we say $A \sim B$ if and only if A and B are similar.

[a] check if you are not convinced

An immediate consequence of staring at this diagram:

$$\begin{tikzcd}
R^n \arrow[dr, "\varphi"] \arrow[dd, "\nu_{\beta_A}^{\beta_B} = \varphi^{-1} \circ \varphi"'] & & R^n \arrow[dl, "\varphi"'] \arrow[dd, "\mu_{\beta_C}^{\beta_D} = \psi^{-1} \circ \psi"] \\
& V \arrow[r, "\alpha"] & V \\
R^n \arrow[ur, "\psi"'] & & R^n \arrow[ul, "\psi"]
\end{tikzcd}$$

we see that $A \sim B$ if and only if there exists an invertible $P \in \mathrm{GL}_n(R)$ such that

$$B = PAP^{-1}$$

which explains the definition for similarity of endomorphisms, though the definition for endomorphisms generalizes to non-finite cases. When working over a finite-dimensional vector spaces, then each endomorphisms has a matrix representation.

The reader may have noticed that $A \sim B$ if B is the conjugate of A via an invertible matrix. In the language of group theory, this shows that $\mathrm{GL}(V)$ naturally acts on $\mathrm{End}_R(V)$ via conjugation[10], hence $A \sim B$ iff A, B are in the same orbit in this action.

[10] Recall that conjugation is *the* natural group-action

Just like for equivalence, we may ask for when two matrices are similar, or if there is a nice representation in the orbit. This question is a bit trickier because the domain and codomain have are the same, and so there is less flexibility in choosing representatives. In chapter 14, we shall show how to find a nice basis to find a nice representation of α.

13.2.3 Gaussian Elimination: Finding a Nice Basis

Let $\varphi : V \to W$ be a linear transformation, and P be some matrix representation given by choosing a basis for V and W. Then P belongs to some equivalence class representing all possible representatives of φ. If we multiply P on the left or right by an invertible matrix, then by by proposition 13.2.11, P remains in the same equivalence class. It would thus be nice if given P, we may simplify P as much as possible to get an "nice" looking represetnative for φ. To do this, we will find nice sets of invertible matrices that do "elementary" operations to a marix which will allow the matrix P to be put in as simple a form as possible. In particular, we will find invertible matrice that will allow us to perfom the following "elementary" operations:

1. switching two rows (resp. columns) or P. For example, let's say we want to switch the 1st and 2nd row of a matrix. This can be done by multiplying on the right (resp. left) by:

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

   and the identity on the left (resp. right). In particular, we swapped the rows of the above matrix.

2. multiply all elements in a row (resp. column) of P by a unit of R by multiplying on the left (resp. right) by a matrix of the form:

$$\begin{pmatrix} u & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

   This particular matrix will scale the first row (resp. column) by a factor of u when multiplying on the left (resp. right).

3. add a multiple of one row (resp. column) to another row (resp. column) of P. For example, to add a multiple of the first row to the second,we would multiply on the right (resp. left) by:

$$\begin{pmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

   This particular matrix will add 3 copies of the 2nd row (resp. column) to the 1st row (resp. column) when multiplying on the left (resp. right).

It should be checked that these matrices are indeed invertible matrices (can you find their inverse?), and that the side on which the matrix is multiplie by effects either columns or rows. The following proposition summarizes the above discussion, generalizing it to general commutative rings R:

Proposition 13.2.14: Equivalence An Elementary Matrices

Let $P, Q \in M_{m,n}(R)$ where R is a commutative ring. Then if Q may be obtained from a sequence of elementary operations, then P an Q are equivalent.

Proof **:**
Verify this proposition.

Due to these matrices being the transition matrices between two equivalent matrices, we give matrices that perfom elementary operations a name:

Definition 13.2.15: Elementary Matrices

Elementary matrices are matrices such that when mutliplying a matrix from the left or right side,the reuslting matrix has had an elementary operation performed on it.

Notice that proposition 13.2.14 is not an "if and only if", simple an "if". This is because it is possible that P and Q are equivalent, but the invertible matrices that "connects" the two are not the product of elementary matrices. If R has certain extra properties, then it will be the case that all invertible matrices can be written as a product of elementary matrices, as we'll explore now. To make this easier to exlpore, let's give a name to all invertible linear matrices

Definition 13.2.16: General Linear Matrices

Let R be a ring and $M_n(R)$ the set of all matrices of of dimension n over R. Then:

$$\mathrm{GL}_n(R) := (M_n(R))^\times$$

that is, $\mathrm{GL}_n(R)$ is the set of all units of $M_n(R)$, i.e. the set of all invertible matrices of dimension n over R.

The word "dimension" is used to indicate the number of rows and columns (which are both equal in this case, and hence why a single number is named). The term "general" is used since there other set of invertible matrices commonly studied, usually given a different prefix ("special", "orthogonal", "orthonormal", "unitary", and so forth). This set of invertible matrices is called "general" since it ill encompass *all* of these other sets of invertible matrices. For who may have come acroess these terms before, notice that this set was already mentioned in definition 1.2.20. One thing that as dificult to prove direclty with matrices but easy to prove now is that $(\mathrm{GL}_n(R), \cdot)$ is associative: every invertible matrix represents a function, and function composition is associative, and hence $\cdot$ is an associative operation.

As mentioned, there can be matrices in $\mathrm{GL}_n(R)$ which are not a product of elementary matrices, that is, $\mathrm{GL}_n(R)$ need not be generated by elementary matrices. However, if $R = k$ is a field, then this is in fact the caes, and the "if" in proposition 13.2.14 becomes and "iff"

Proposition 13.2.17: General Linear Group Of Field

Let $R = k$ and $\mathrm{GL}_n(k)$ be the general linear group over a field. Then $\mathrm{GL}_n(k)$ is generated by elementary matrices.

Proof **:**

The idea is as follows: take advantage that all entries are units. Make a_{11} 1 and then eliminate all a_{i1} and a_{1j} (i.e. make them zero). Then you have made a $(n-1) \times (n-1)$ matrix ith a 1 in the top-left. Repeat this process until you get the identity:

$$I_n = M \cdot P \cdot N$$

Since M and N are invertible, we wget:

$$(NM)^{-1} = M^{-1} I_n N^{-1} = P$$

registering all the operations will produce you product of elementary matrices, an (NM) is still the prouct of invertible matrices, and so $(NM)^{-1} = P$ is the product of invertible matrices, completing the proof.

The above method can be used more generally to find the inverse of a matrix, and we only need to multiply *on the left hand side* or the *right hand side*. The method is common enough to be given a name: *Gaussian Elimination*. [11]. Since Gaussian elimination usually reffers to operations on the left, we may think of this as a form of "row"-equivalence. However, as we saw we may as easily multiply the elementary matrices only on the right and get "column"-equivalence. This shows us that the two forms of equivalence are the same. Some authors define the notion of *row rank* and *column rank* for matrices which correspond to the dimension of a vector-space spanned by the rows and column respectfully, and proceed to show that the row and column rank are the same number. For reference, we put the followign definition in a box:

Definition 13.2.18: [matrix] Rank

Let A be a matrix. Then the *row rank* is the dimension spanned by its rows, the *column rank* is the dimension spanned by its colun. By the above remark, the row-rank and the column-rank are equal, and so the single numer is called the *rank* of the matrix. If the rank is maximal, the matrix is said to be of *full rank*.

If a matrix is full rank, then the linear map it represents is surjective.

[11] Though Gauss has most likely invented the above method when working with matrices, it has been known under slightly different symbols and theory at least as far back as Han China

Corollary 13.2.19: Equivalence Of Matrices

Let $R = K$ be a field and $P \in M_{m,n}(R)$ be any m trix. Then P is equivalent to:

$$\left(\begin{array}{c|c} I_k & 0 \\ \hline 0 & 0 \end{array}\right)$$

for appropriate k, $1 \leq k \leq \min(m, n)$, here 0 is the represents that all of the entries are 0 (for the appropriate size).

Proof **:**
This is just combining the previous results.

Thus, two matrices that have a different value of r in the above matrix will *not* be equivalent. This also shows that there are only *finitely many* equivalence classes matrix representations of linear transformation's between two finite dimensional vector-spaces, in particular there is one for each dimension$1 \leq k \leq n$. This also show's that if we look at a linear transformation in the right way (i.e. pick the right basis), a linear transformation *always* looks like it some projection eliminating certain dimensions while embedding the rest of the space into the codomain! See proposition 13.1.15 to see how this intuition and the existence of compliments of subspaces work together.

To finish off this section, we can ask if the same logic can be done for the notion of similarity. Unfortunately, similarity is more complicated: since only one matrix is being picked (i.e. only one basis is being picked) the rows and column operations must be done "simultaneously". Not all hope is lost: using properties of PID's will show us how to find a good matrix representation for similar matrices in the next chapter.

Gaussian Elimination over Euclidean Domain

From the rings we explored earlier, the euclidean domain is a good start, and indeed with some cleverness a matrix over a euclidean domain can be reduced, though not intirely to an identity matrix in a sub-matrix of the original.

To demonstrate the idea, let R be a euclidean domain with valuation ν, and take any 2×2 matrix:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

then use elementary matrices to put the element with the lowest valuation at position $(1, 1)$. Then use the euclidean algorithm to find:

$$b = aq + r$$

where either $r = 0$ or $\nu(r) < \nu(a)$. Then do the appropriate elementary operation to get:

$$\begin{pmatrix} a & r \\ c & d - aq \end{pmatrix}$$

if $r = 0$, then we have eliminated one of the components. If $r \neq 0$, then repeat the same process, moving the term with the lowest valuation to the $(1, 1)$ position using elementary matrices. Since

each term has a finite valuation, it must be that eventually $r = 0$. With minor modifications, the same thing can be done in the $(2, 1)$ position, getting us:

$$\begin{pmatrix} a' & 0 \\ 0 & d' \end{pmatrix}$$

Next, notice that we may find an a', d' such that $a' \big| d'$. For if not, we may consider:

$$\begin{pmatrix} a' & d' \\ 0 & d' \end{pmatrix}$$

and repeat the process until we get to the same state, and keep repeating until either $d' = 0$ or $a' \big| d'$. Thus, we produce a 2×2 matrix where the diagonal divides each other.

From this, it is not too hard to see that this is easily generelizable to an $n \times n$ matrix:

Theorem 13.2.20: Smith Normal Form

Let R be a euclidean domain and $P \in M_{m,n}(R)$. Then P is equivalent to a matrix of the form

$$\left(\begin{array}{ccc|c} d_1 & \cdots & 0 & 0 \\ \vdots & \ddots & 0 & 0 \\ 0 & \cdots & d_k & 0 \\ \hline 0 & \cdots & 0 & 0 \end{array}\right)$$

where

$$d_1 \big| \cdots \big| d_k$$

The resulting matrix is called the *smith normal form.*

Proof **:**
exercise

Corollary 13.2.21: Finding Inverse Of Matrix

Let $A \in \mathrm{GL}_n(k)$. Then A^{-1} can be found by performing Gaussian Elimination.

Proof **:**
This is taking the strategy done in theorem 13.2.20 and keeping track of the elementary matrices used to produce the inverse. Multiplying all of them gives the desired result.

Using smith normal forms, we may prove that $\mathrm{GL}_n(R)$ is generated by elementary matrices. Unfortunately, there does exist a counter if R is a PID; there is exists a 2×2 counter-example over $\mathbb{Q}(\sqrt{19})$ (see cite:HERE, under books/math/papers/ the GL2 paper). On the other hand, we can put any matrix in $\mathrm{GL}_n(R)$ into smith normal form using the gcd and some non-elementary matrix operations. This shall be further explored in chapter 14.

13.2.4 Geometric Result: Systems of Linear Equations

Vector spaces and linear maps have part of their origins in geometry, where the term *linear* originates. The rigorous interpretation of a concept being "geometrical" is something that would be explored in a class on [differential] geometry, and so for the uses of this book geometrical shall mean "locally or globally looks like the solutions to equations":

$$\begin{aligned} f_1(\vec{x}) &= a_1 \\ f_2(\vec{x}) &= a_2 \\ &\vdots \\ f_n(\vec{x}) &= a_n \end{aligned}$$

Often, the constant is moved to the left modifying the equations to be:

$$\begin{aligned} f_1(\vec{x}) - a_1 &= 0 \\ f_2(\vec{x}) - a_2 &= 0 \\ &\vdots \\ f_n(\vec{x}) - a_n &= 0 \end{aligned}$$

This understanding of geometrical is in fact very potent, for example the reader may have already encountered many geometrical objects that are defined in this way, for example:

Figure 13.1: solution to $x^2 + y^2 - n^2 = 0$

For the particular explorations of this section, we shall show that vector spaces can be interpreted as the of solutions to *linear equations*:

Definition 13.2.22: Linear Equation and System of Linear Equations

Let V be a k-vector space of dimension n and $k[x_1, x_2, ..., x_n]$ be a polynomial ring over the field k with n variables. Then a *linear equation* is a polynomial, possibly over several variables, where the degree of each variable is at most 1:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

A collection of linear equations is called a *system of linear equations*.

Given a (system of) linear equation(s), we can define *linear functions*: if $\sum_i a_ix_i$ is a linear equation then we can define:

$$L_1 : k^n \to k \qquad L(x_1, x_2, ..., x_n) = a_1x_1 + a_2x_2 + \cdots a_nx_n$$

and similarly a system of linear functions is a collection of linear functions $L_1, ..., L_m$. Since these functions are linear, they can be represented by a matrix, namely given the standard basis:

$$A = [L_1] = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}$$

so that if $\vec{x}$ is an $n \times 1$ matrix with with x_i coefficients, then plugging in a value from k^n is represented as matrix multiplication by

$$A\vec{x} = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

Equivalently, any system of linear equation can be combined to form a linear equation:

$$L : k^n \to k^m \qquad L = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_m \end{bmatrix} \tag{13.7}$$

Let's now say we want to find the solution to the system:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n &= b_n \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots a_{mn}x_n &= b_m \end{aligned}$$

Then if $\vec{b} = (b_{k1})$, then what we are solving for is:

$$A\vec{x} = \vec{b}$$

If A is invertible and $n = m$, then the answer is simple:

$$\vec{x} = A^{-1}\vec{b}$$

where A^{-1} is found using corollary 13.2.21 (how does this relate properties of linear maps that have been explored earlier?). If A is not invertible or $n \neq m$, the next best thing that can be done is find its smith normal form giving us matrices M, N such that

$$M \cdot A \cdot N = \left(\begin{array}{ccc|c} d_1 & \cdots & 0 & 0 \\ \vdots & \ddots & 0 & 0 \\ 0 & \cdots & d_k & 0 \\ \hline 0 & \cdots & 0 & 0 \end{array}\right)$$

then if $\vec{c} = M\vec{b}$ and $y = (y_k)$ we get:

$$(M \cdot A \cdot N) \cdot \vec{y} = \vec{c}$$

and producing the system:

$$d_1 y_1 = c_1 \tag{13.8}$$
$$d_2 y_2 = c_2 \tag{13.9}$$
$$\vdots \tag{13.10}$$
$$d_r y_r = c_r \tag{13.11}$$
$$0 = c_{r+1} \tag{13.12}$$
$$\ddots \tag{13.13}$$
$$= c_m \tag{13.14}$$

where $d_i \big| c_i$. Conversely, if there is a set of linear equation of the form given in equation 13.8, then it has a solution of $d_i \big| c_i$.

Where this links to geometry is when considering the space of solution to systems of linear equations. First, notice that without loss of generality the linear equations may equal to 0, for this would simply translate the resulting set of solutions in the vector-space, hence it suffices to consider:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots a_{mn}x_n &= 0 \end{aligned}$$

That is, it suffices to consider the kernel of a system of linear equations, equivalently the kernel of L as defined in equation (13.7).

Now, to endeavor exploring the shape of the kernel of linear functions, let's start with $k = \mathbb{R}$, $L : \mathbb{R}^3 \to \mathbb{R}$, and always use the standard basis. It is easy to visualize $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3$, and higher

dimensional analogues can be fathomed. Naturally, if $L(x) = 0$, then $\ker = \mathbb{R}^3$. It is more interesting if L is non-trivial, which for linear functions implies L is surjective. Then by the Rank-Nullity Theorem $\dim \ker = 3 - 1 = 2$, i.e. it is a plane. A plane that is one-dimension less than the ambient space (in this case $\mathbb{R}^3$) is called a *hyper-plane*. Since it is a subspace, it must contain 0 and hence passes through the origin. Since the plane is two dimension, it is spanned by two vectors. Using Gaussian elimination, we can deduce it is spanned by two vectors:

$$\left\{ \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \right\}$$

Example 13.2.2: Finding A Hyperplane

Take $L(x, y, z) = 3x + 2y - 3z$. Find a basis for the kernel. The kernel when embedded in $\mathbb{R}^3$ looks like so:

Figure 13.2: $3x + 2y - 3z = 0$

Notice that since the image is one dimensional, the hyper-plane is defined by only one equation, however it is a two dimensional object. More generally, for a surjective linear function $L : \mathbb{R}^n \to \mathbb{R}$, the hyperplane shall be $(n - 1)$ dimensional and be defined by a single equation.

Next, let's consider the case where $L : \mathbb{R}^3 \to \mathbb{R}^2$, that is there are two linear equations $L_1 : \mathbb{R}^3 \to \mathbb{R}$ and $L_2 : \mathbb{R}^3 \to \mathbb{R}$. Let's again assume that L is surjective, or full-rank, for if L was one dimension less it would be equivalent to the case of $L : \mathbb{R}^3 \to \mathbb{R}$. Then by the Rank-Nullity theorem we have that $\ker L = 3 - 2 = 1$. Intuitively, this is because there are more restrictions for $L(x) = 0$, namely both L_1 and L_2 must equal 0. Geometrically, this can be thought of as the *intersection* of the kernel of L_1 and the kernel of L_2. Since L is full rank, the intersection is not parallel (the geometric word for this is that the intersection is *transversal*). More generally, $L : \mathbb{R}^n \to \mathbb{R}^2$ would have $\dim \ker L = n - 2$, and would be defined by two linear equations.

Extended the pattern further, given a full-rank linear function $L : \mathbb{R}^n \to \mathbb{R}^m$, the resulting plane given by the kernel would be $n - m$ dimensional and would represent the intersection of m hyper-plane of dimension $n - 1$, each new hyper-plane restricting the dimension of the kernel. The orientation of the kernel is completely dependent on the linear map in question. Nothing in these results required to work specifically with $k = \mathbb{R}$, it was simply convenient for visual purposes. These results can be summarized into the following:

Theorem 13.2.23: Classification Of Planes

Let V be a k-vector-space of dimension n. Then the sub-spaces of V correspond to the kernel of full-rank linear maps $L : V \to k^m$, where $1 \leq m \leq n$.

Proof **:**
Fill in the blanks from the above discussion

Given that L is surjective, notice that as m increases in size, the dimension of the kernel decreases in size; in a sense it is the "co-dimension" of the domain. This observation is not lost on mathematicians, and the notion of codimension is quite important in geometry: the dimension of an object represents the degree of freedom an object may have within a space (ex. a hyper-plane is $n-2$ dimensional) while the codimension represents the degree of restriction an object may have within a space (ex. a hyper-plane is 1 codimensional).

(Also, a geometric interpretation of first is iso theorem: when solving system of Lin. Eq, look for kernel. But notice the solution to $Ax + c = b$ is just an affine translation of the kernel! All elementary row operations fix the kernel and permute the cosets, when the elementary operations generate $\mathrm{GL}_n(R)$.)

Example 13.2.3: Exercices
Something with transpose. Perhaps even with conjugate? Get started with $\langle A, A*\rangle$ stuff?

Exercise 13.2.2

1. (from dummit and foote) (note that chagne of absis for a lienar transformation from V ot itself is the same as conjugation by some element of the group $GL(V)$ of nonsingualr linear transformation of V to V. In particular , the relation "similarity" is an equivalence relation whose equivalence classes are the orbits of $GL(V)$ acting by conjugation on $\mathrm{Hom}_F(V, V)$.)

13.2.5 Use Case: Reading Presentations

Recall from section 12.3.3 that a presentation of a [finitely generated] R-module can be represented as an exact sequence:

$$R^n \to R^M \to M \to 0$$

Using the above results, we can more easily read the presentation. Let $\varphi : R^n \to R^m$. Then by exactness, we have that $M \cong \mathrm{coker}(\varphi)$. Hence, if we find the image of φ, we found the relations of M, and and if R is a field of euclidean domain, this is task is simplified using Gaussian elimination:

Example 13.2.4: Finding A Presentation
Let $R = \mathbb{Z}$ and let φ be represented by the matrix:

$$\begin{pmatrix} 1 & 3 \\ 2 & 3 \\ 5 & 9 \end{pmatrix}$$

which means that $\varphi : \mathbb{Z}^3 \to \mathbb{Z}^2$ and this matrix determines an abelian group G (since abelian groups have free resolution of length at most 2 as mentioned in section 12.3.3). To find G, let us first simplify the above matrix. The reader shoul check that the matirx reduces to:

$$\begin{pmatrix} 1 & 0 \\ 0 & 3 \\ 0 & 0 \end{pmatrix}$$

Hence, G is isomorphic to the cokenrel of $\varphi : \mathbb{Z}^2 \to \mathbb{Z}^3$ that maps $(1, 0)$ to $(1, 0, 0)$ and $(0, 1)$ to $(0, 3, 0)$. In particular, since the image is $\mathbb{Z} \oplus 3\mathbb{Z} \oplus 0$, we get

$$G \cong \operatorname{coker} \varphi \cong \frac{\mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}}{\mathbb{Z} \oplus 3\mathbb{Z} \oplus 0} \cong (\mathbb{Z}/3\mathbb{Z}) \oplus \mathbb{Z}$$

Showing how a matix can uniquely represent an abelian group! In the next chapter, we shall show how a matrix can more generally represent a finitely generated PID.

38: Remark

Notice that we may eliminate first row and column if the leading term is 1 (or more generally a unit) since in the quotient the summand shall be completely quotiented.

13.3 Dual Spaces and Forms

Let's say you are studying some mathematical object, say the set of linear transformations. We may want to find numerical properties, for example quantify the size of a matrix, or give some numerical value to similarity, or the dimension of the image. We can define a function from $L(V)$ to a field, say $\mathbb{R}$, that shall give us this quanitifcation. The collection of all these functions are called the *functionals* on $L(V)$. Depending on what type of property we want, we will have different types of functionals. For example, when working in analysis, we shall often want our functional to be continuous or perhaps bounded in some way [12]. In Algebra, we shall want our functionals to be homomorphisms, and depending on what other property we want we shall would add more properties to our functionals (ex. we may want our functional to preserve solutions to a special set of polynomial equations).

The set of functionals on a vector space V can be thought of as containing all possible quantifiable "information" about a space (with the limits of the information provided by the properties of the functionals). Most of the time, the space of functionals is larger than the original space[13]. In some special cases, the spaces of all functionals is isomorphic to the space over which we are taking the functionals (as we shall see, all linear functionals on a finite dimensional vector space V is isomorphic to V). In this special case, we can think of our space being characterized by the property the functionals represent! This realization was the origin of *dual theory*, which we shall start exploring in this section. Dual theory is normally explored in earnest in Real and Functional analysis, however it shall show up in some important ways in chapter 21, and certain key functionals shall give a lot of information on linear transformations, namely the trace and the determinant.

[12] See Lebesgue Spaces for more details

[13] This is a guiding principle behind many generalizations in Mathematics, Distribution Theory being the prototypical example

13.3.1 Dual Spaces

In mathematics, there is frequently two "opposing" or "dual" ways of describing objects: Let's say X is our object of study

1. internally: Using properties internal to your object to describe that object. Often, maps *to* the object will give us internal information ($f : A \to X$ for some appropriate object A). This can may also be thought of as local information

2. externally: Using some exterior property (ex. be it by putting it an ambient space, or a sapce in which the object X "resonates" or "interacts" with[14]). Ofen, maps *from* the object will give us external information ($f : X \to A$ for osme appropriate object A). This can also often be thought as global information

As an example of this, when we were working with group theory, we found two way of looking at all possible groups:

1. Using quotients of free groups, we managed to make normal subgroups of free groups represent relations between elements. In the process of doing this, we defined a surjective homomorphism:

$$\varphi : F(A) \to G$$

2. Using Cayley's theorem, we manage to embed any group into a subgroup of S_A. In the process of doing this, we defined an invective map

$$\psi : G \to S_A$$

Notice how the first one defined us what's happening between elements in a sort of "local" way, while in the 2nd example we defined it by using properties of the larger group S_A and how elements of G embed into S_A. Another example of this principle can be seen with the universal property of products and coproducts of groups. In a sense, the universal property of products only requires the product group to have groups A and B "locally" in the most efficient way, leading to the direct product of A and B, while the corpoduct group require it to have A and B "globally" in the most efficient way, leading to the free product of A and B.

Often when studying an object X, we will study if from these two perspectives: from maps going *into* X and maps going *out of* X. One remarkable phenomena that occurs often in mathematics is that there is often a [contravariant] *dual object* in which the maps *into* and *out of* X are naturally associated to maps *out of* and *into* a dual object X^*. Note the reversal in the directions of the maps! It is often very fruitful to look at objects from the perspective of both X and X^* to gain more information.

To demonstrate a simple but powerful example of "dualness" in mathematics (or the pervasiveness of dualness), we require knowledge of fields outisde of algebra. Usually, I (the author of this book) would decide to leave out any such example to insure no prerequisites are required to learn the material. However, I have decided to break this rule here for those who know a bit of topology since it is an example so powerful that it changed my view of duality, and requires elementary topology knowledge. The following example is *not* necessary for understanding any material of this book

[14] If you know Fourier Analysis, think of phase space

and can be considered motivation for dual spaces and dual theory in general for those who know some topology (if you need a refresher on topology, see appendix C). In this example, we will take a moment and venture outisde of algebra and into the realm of analysis and and geometry to give an interesting example.We will give a common phenomena in which "algebraic" object have "geometric" objects as their dual. In particular, unlike the duality of a group in its represetnation in terms of the quotient of a free group or an embedding into S_G, both of which are algebraic ways of looking at groups, we will show that dual objects are sometimes found in other fields of mathematics, allowing us to connect these two fields of mathematics proof-strategies and techniques!

> First, define a *Boolean Algebra* to be a $\mathbb{F}_2$-algebra (where $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$). Since $\mathbb{F}_2$ is a field, a Boolean Algebra B is module-isomorphic to $\mathbb{F}_2^n$ for some cardinality n. For $0, 1$ in a Boolean algebra B:
>
> $$0+0=0 \quad 0+1=1+0=0 \quad 1+1=0 \qquad 0\cdot 0=0\cdot 1=1\cdot 0=0 \quad 1\cdot 1=0$$
>
> We usually also define the additional operation of $\neg$:
>
> $$\neg(x) = 1 - x$$
>
> If you have done some computer science or logic, then you might notice that 1 and 0 can represent the values of *true* and *false* respectively, while $+$ and $\cdot$ represent XOR and AND respectively (hence the name Boolean to this algebra). This algebra underlines most of the logic behind computer science, making us want to find properties of boolean algebras. One way we may study these properties is by looking for maps going *into* a boolean algebra B. Dually, by a result known as *Stone Duality*, we can instead study all the maps *out of* a totally disconnected[15] Hausdorff spaces B^*! Thus, the "dual" of Boolean Algebra is all continuous maps from B^* to $\mathbb{R}$! More generally, locally compact[16] topological spaces (which we can consider as a geometric object) are the dual of a type of algebra known as a C^*-algebras[17] (which we consider as algebraic objects): this is known as *Gelfand Duality*[18]. The study of these objects that are related to each other by "duality" is called *Dual Theory*.

Coming back to algebra and the study of vector-spaces, we will see the very beginnings of Dual Theory by studying the simplest case of dual objects, *dual vector-spaces*. The way we will define the dual of a vector-space will be through "external" means. When we defined vector spaces using a basis, that is a "local" property. In a sense, we showed that the map $f : F(A) \to V$ can define a vector space where A is a basis. What we'll now do is the *dual* equivalent: we'll define a vector space by mapping V into a different space.

The question arrises as to what space might we might we map to? When working with groups, it was not immediately obvious that S_A would be the right group to map. Fortunately for us, there is an "obvious" space to map to: for any vector-sapce over k, we are always given two algebraic objects, namely the abelian group V and the field k. Since k is always a vector-sapce over itself,we can consider linear maps $f : V \to k$! The collection of all such maps is called the *dual space* of V.

[15] A topological space in which the only connected sets are singletons

[16] a topological space in which for any point x, there exists a compact neighbourhood

[17] A Banach Algebra over $\mathbb{C}$ with a map known as an involution satisfying the properties of an adjoins

[18] This duality requires a good deal more topology and analysis than is part of the scope of this book. However, in part V, chapter 21.7, we will see the Zariski Duality which studies a geometric-algebraic duality without the need for much analysis or topology

Definition 13.3.1: Dual Space

Let V be a vector space over k. Then $V^* = \mathrm{Hom}_k(V, k)$ is the space of all linear transformations from V to f, and is called the *dual space* of V.

By proposition 12.2.6, $\mathrm{Hom}_k(V, k)$ is a vector-space over k (if you're not convinced, it is recommended to reprove this before continuing). The notation $V^\vee$ is also often used to define the dual space and will be used interchangibly. Elements inside the dual space V^* are sometimes called *covectors*, paralleling to how elements inside a vector space V is called *vectors*. In physics and analysis, there is seomtimes a different convention for how to write the evalutation of if $f \in V^*$ at $x \in V$: instead of writing $f(x)$, to mean "evaluate f at the point x", it is sometimes written $\langle f, x \rangle$.

Elements of V^* are given a name:

Definition 13.3.2: Linear Functionals

Let V^* be the dual space of V. Then $f \in V^*$ is called a *linear functional* or *linear form*.

Example 13.3.1: Linear Functional and Dual Spaces

1. (If you know some calculus) Perhaps the most famous linear functional is $\int_a^b __dx : L^1(\mathbb{R}) \to \mathbb{R}$, where $L^1(\mathbb{R})$ will denote the set of all (either Riemann or Lebesgue)[a] integrable functions $f : \mathbb{R} \to \mathbb{R}$ where $\int_{\mathbb{R}} |f| dx < \infty$ (so that the integral is well-defined). Note that $L^1(\mathbb{R})$ is a vector-space with addition and scalar multiplication defined pointwise since the codomain is $\mathbb{R}$:

$$\begin{aligned} f + g &\Leftrightarrow (f+g)(a) = f(a) + (g) \\ c \cdot f &\Leftrightarrow (c \cdot f)(a) = cf \end{aligned}$$

   Now, $\int_a^b __dx$ takes in a function $f : \mathbb{R} \to \mathbb{R}$, an gives th number $\int_a^b f dx \in \mathbb{R}$. By basic properties of integarls, we know that:

$$\begin{aligned} \int_a^b (f+g) dx &= \int_a^b f dx + \int_a^b g dx \\ \int_a^b cf dx &= c \int_a^b f dx \qquad \forall a \in \mathbb{R} \end{aligned}$$

   and hence $\int_a^b __dx$ is indeed a linear functional! Hence $\int_a^b __dx \in L^1(\mathbb{R})^*$, for any $a, b \in \mathbb{R}$[b]
   The study of Analysis focuses a lot on the study of the properties of these functionals, and would be covered in many Analysis courses. For textbooks, see ref:HERE.

2. Let $C([a, b])$ be the collection of all continuous maps from $[a, b]$ to $\mathbb{R}$, $f : [a, b] \to \mathbb{R}$, with addition and scalar multiplication defined pointwise. Then define: $\mathrm{ev}_x : C([a, b]) \to \mathbb{R}$ to be

$$f \mapsto f(x)$$

   Then it can be shown that ev_x is a linear transformation, called the *evaluation map*, and is thus a linear functional. More generally, as long as the codomain is a field $\mathbb{R}$, then the evaluation map is a linear functional.

3. A more advanced case of the above example occurs if we consider the set $C^1(\mathbb{R})$ of all differentiable functions[c] forms a vector-space with point-wise addition and scalar multiplicaiton. The operator $D_a : C^1(\mathbb{R}) \to \mathbb{R}$ takes in a differentiable function f and gives back $Df(a)$ (which in first year calculus is more commonly denoted $f'(a)$).

4. (If you know geometry or analysis) The following examples are not examples of dual spaces as we've defined them, but the generlization of them. In this generlization, the dual space we've defined would be called a *linear* dual space, or the dual space of in the category of $\mathbf{Vect}_k$. More generally, if X is an object in a category, then given a *dualizing* object D (often it is $\mathbb{R}$, $\mathbb{C}$, $\mathbb{C}^\times$, $\mathbb{Z}/n\mathbb{Z}$, or $\mathbb{R}/\mathbb{Z}$ depending on the context), then $\mathrm{Hom}_C(X, D)$ is the dual space. In the following example, let D be either $\mathbb{R}$ or $\mathbb{C}$. Let's say X is either a vector-space over $\mathbb{R}$ or $\mathbb{C}$ (you should have function spaces in mind) or a manifold over $\mathbb{R}$ or $\mathbb{C}$. Then $C_0(X)$, $C^k(X)$, $L^p(X)$, an so forth are all subspace of the dual space X^* in the appropriate category.

Another place in which [ring] funcitonals shall become of paramount importance is in chapter 21, particularly section 21.1 and 21.7 where we shall see define a duality between algebraic and geometric objects.

[a]For the sake of this example, it doesn't matter, though L^1 usually denote *Lebesuge* measurable functions where $\int |f| < \infty$

[b](if you know some analysis) one classical result known as Riesz Representation Theorem (for measures) is that all functionals $f \in L^1(\mathbb{R})^*$ will be of the form $f \mapsto \int fg$ where $g \in L^\infty(\mathbb{R})$ and the essential support of g is bounded.

[c]in particular, continuously differentiable functions. This distinction is of no importance for this example

In the above examples, all the vector-spaces are infinite dimensional. The vector-space being infinite dimensional will in fact greatly complicate the structure of dual spaces[19]. In general, more structure is added to a vector-space (ex. a norm) to have better "control" of infinite dimensional vector-spaces. Such discussion is left to a course on functional analysis[20]. We will mainly focus on the theory of finite dimensional vector spaces, since the results of this theory is in it of itself quite fruitful and provides a good basis for its study of dual spaces more generally.

Given a basis for V, it straight-forward to find a basis for $V^\vee$:

Definition 13.3.3: Dual Basis

Let $\beta = \{v_1, v_2, ..., v_n\}$ be a basis for the finite dimensional vector-space V. Then define $v_i^* \in V^*$ to be

$$v_i^*(v_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

for $1 \leq i, j \leq n$. Let β^* be the set of dual-basis elements

Since β^* has the word basis in it's name, you might guess that it is ineed a basis for the dual space:

Proposition 13.3.4: Dual Basis

Let $\beta*$ be defined as in definition 13.3.3. Then β^* is a basis for V^*

[19]This is one of the analytic properties that was discussed in the footnote in page 393

[20]Give references to books HERE, ex. Folland

Proof **:**
exercise

As a consequence, V^* has the same dimension as V

Proposition 13.3.5: Reflectivity Of Finite Dimensional Vector Spaces

Let V be a finite dimensional vector space. Then $V \cong V^*$.

Proof **:**
By proposition 13.3.4, since V and V^* are of the same dimension, $V \cong V^*$.

An intuitive way of viewing the reflectively product to reference the intuition presented in the introduction of this section mentioning how functional represent *information*. Each linear function provides some information about a vector-space (prototypical examples would be the projection functions $\pi_{\beta,i} : V \to k$ picking a component of a vector given a choice of basis β). The type of information provided is restricted by the properties of the functions on V, in this case the co-domain is always k, and the functional is always linear. In finite dimension, these functions can be completely characterized by elements in V by the existence of the above isomorphism. The choice of the isomorphism is still an important factor to consider, since there is not natural basis or extra information that allows for favoring one isomorphism over another.

Since the structure of V isomorphic to V^*, it would be interesting if each map $f : V \to W$ can be associated to an appropriate "dual" function f^* where composition is respected, hopefully giving the existence of the functor $(-)^*$ from Vect_k to itself. To start, there is a notion of a dual function:

Definition 13.3.6: Dual map (Transpose)

Let $f : V \to W$ be a linear transformation. Then the *transpose* of f, denoted f^*, is defined as:

$$f^* : W^* \to V^* \qquad f^*(\varphi) = \varphi \circ f$$

Unfortunately, $(-)^*$ can't so easily be defined as each choice of isomorphism $V \cong V^*$ for each vector-space $V \in \mathbf{Vect}_k$ may not be compatible, see exercise ref:HERE. Note that the transpose of a function insures that maps that used to "out of" the space V will now "go into" the space V^* (and similarly for maps that go into V). Thus, this definition matches with our goal for V^* to be the dual of V, and furthermore is independent of any choice of basis. The name transpose comes from the matrix representation of f^*

Proposition 13.3.7: Matrix Representation Of Dual Map

Let $f : V \to W$ be a linear transformation and $A = (a_{ij})$ be the matrix representation of f given some basis β and γ for V and W respectively. Then if $f^* : W^* \to V^*$ is the dual map of f, where W^* and V^* have as basis β^* and γ^*, then the change of basis matrix A^* is:

$$A^* = [f]_\beta^\gamma = [f^*]_{\gamma^*}^{\beta^*} = (a_{ji}) = A^t$$

Intuitively, in the matrix representation, each entry a_{ij} transposed with entry a_{ji}. For this reason, the map f^* is sometimes called the *transpose* of f.

Proof **:**
exercise

As a corollary, if A is a nonsingular map $f : V \to W$, then the matrix representation for the map $(f^{-1})^* : V^* \to W^*$ is $(A^{-1})^t$.

Similarly to how the kernel of a linear transformation f provides the needed information to deduce the structure of the image of f, we get the dual equivalent for f^*, given an appropriate "dual" for the kernel:

Definition 13.3.8: Annihilator Of Dual Map

$U \subseteq V$ be a subspace of V. Then the *annihilator* of U is:

$$U^o := \{f \in V^* \mid f(U) = 0\}$$

Note that $V^0 = \{0\}$ and $\{0\}^0 = V$. Another common notation is $U^\perp$, which is pronounced "U perp", which shows how U^0 is related to something called the orthogonal compliment of U (this will be covered in section 13.5)

Lemma 13.3.9: Properties Of Annihilator

Let $U \subseteq V$ be a subspace. Then:

1. $U^0 \subseteq V^*$ is a subspace
2. if $\dim(V) < \infty$, then $\dim(V) = \dim(U) + \dim(U^0)$

Proof **:**
exercise

Proposition 13.3.10: Properties Of Dual Maps

Let V and W be (not necessarily finite dimensional) vector space over k, $f \in \mathrm{Hom}_k(V, W)$ so that $f^* \in \mathrm{Hom}_k(W^*, V^*)$ is its dual map. Then:

1. $\ker(f^*) = (\mathrm{im}(f))^0$ (implying f^* injective if and only if f surjective)
2. $\mathrm{im}(f^*) \subseteq (\ker(f))^0$ is a subspace
3. If $\dim(V), \dim(W) < \infty$, then $\mathrm{im}(f^*) = (\ker(f))^0$ (implying f^* surjective if and only if f injective)

Proof **:**
(prove these instead of exercise? I generalize it in prop 13.5.26).

The first point in proposition 13.3.10 is particularly useful in that in infinite dimensions, f being injective *does not* imply f is surjective. In the right situation, checking the injectivity of a linear transformation is *much* easier than checking surjectivity, and so checking the injectivity of f^* is often advantageous.

The next topic we will cover might at first seem abstract, however it is an important demonstration of the concept of *naturality*. Since we have taken $\mathrm{Hom}_k(V, k)$ and have shown it is a vector-space, we can take the dual of it again:

Definition 13.3.11: Double Dual

Let V be a vector-space and V^* it's dual. Then the *double-dual* of V is the dual of V^*, denoted V^{**}

The double dual has the property of being *naturally isomorphic* to V. To explain the idea of a natural isomorphism, we will show how V is *not* naturally isomorphic to V^*:

Example 13.3.2: Double Dual Motivation

Let V be a finite-dimensional vector space over k and $V^* = \mathrm{Hom}(V, \mathbb{F})$. If $\beta = \{e_1, ..., e_n\}$ is a basis for V, and $\beta^* = \{e_1^*, ..., e_n^*\}$ is the dual basis for V^* where:

$$e_j^*(e_i) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Then the maps $e_i \mapsto e_i^*$ are bijective and linear, meaning V and V^* are isomorphic ($V \cong V^*$).

We can further consider $V^{**} = \mathrm{Hom}(\mathrm{Hom}(V, \mathbb{F}), \mathbb{F})$ being the set of all linear maps mapping all elements form $\mathrm{Hom}(V, \mathbb{F})$ to $\mathbb{F}$. We can in similar fashion make an isomorphism $V^* \cong V^{**}$ letting us conclude that $V \cong V^{**}$.

This whole construction started with a choice of basis form V, which need not be unique, and so it's not a "property" of V. However, we can still define an isomorphism between V and V^{**} *without* any reference to a basis. For any $v \in V$, we can define a map

$$(f \mapsto f(v)) : V^* \to \mathbb{F} \qquad (f \mapsto f(v))(g) = g(v)$$

that is, every functional $g \in V^*$ evaluates the value and gets us $g(v)$. This map is sometimes called the evaluation function. Let's switch up the notation for simplicity:

$$ev_v : V^* \to \mathbb{F} \qquad ev_v(g) = g(v)$$

where ev_v stands for evaluate at v. Now, defining map $V \to \mathrm{Hom}(V^*, \mathbb{F})$, $v \mapsto ev_v$ is an *isomorphism*. Bijectivity is left to check, for linearity, let φ be the function in question, so that for $k \in \mathbb{F}$ and $v, w \in V$:

$$ev(k \cdot v) = ev(kv) = ev_{kv} = k \cdot ev_v = kev(v)$$

$$ev(v + w) = ev_{v+w} = ev_v + ev_w = ev(v) + ev(w)$$

where kv represents the vector we get when applying the ring action on v. Note that we essentially used the lineary of the object we are mapping to in order to prove linearity.

Thus, we defined this isomorphism with *no* reference to a basis on V. We can't do the same trick for $V \to V^*$, since if we try to do $v \mapsto f(v)$, it is ambiguous which f we are talking about.

The intuition behind the double-dual having a natural isomorphism is that there is more information to work with; the reader should ponder on what extra information has been provided.

Prelude to Adjoints of Vector Space

Now that we have the basic structure of (finite dimensional) dual spaces, we may return to the discussion of defining V internally or externally. Let's say that you have been given $\mathrm{Hom}_k(V, k)$, but not the structure of V. Is the information in $\mathrm{Hom}_k(V, k)$ sufficient to recover the structure of V? The answer is almost: the following proposition shows how much information we can recover:

Proposition 13.3.12: Recovering V From it's Dual

Let V be some vector-space and V^* it's dual. Then if $f(v) = f(w)$ for all $f \in V^*$, then

$$v = w$$

Proof **:**
Exercise. This can be proven with and without the use of a basis for V.

Another way to bring in the intuition from the introduction is to take the dual of a functional $V \to k$, namely to think of elements of V as functions $k \to V$. The reader should check that all linear transformations $k \to V$ are of the form:

$$t \to tx$$

for some $x \in V$. Represent these transformations as $g_x : k \to V$. Taking the dual of g_x we get:

$$\begin{aligned} g_x^*(\varphi) &= \varphi \circ g_x \\ &\equiv \varphi(g_x(t)) && \forall t \in k \\ &= \varphi(tx) && \forall t \in k \\ &= t\varphi(x) && \forall t \in k \\ &\equiv g_{\varphi(x)} \end{aligned}$$

Hence g_x^* maps φ to $g_{\varphi(x)} : k \to k$ where $t \mapsto t\varphi(x)$. With the identification between elements of V and the linear transformations $k \to V$, we may interpret this as a functional $V^* \to k$ where:

$$\varphi \mapsto \varphi(x)$$

Given $V \cong V^*$, Then if β is the basis defining the isomorphism, each $\varphi \in V^*$ is associated to a vector $v \in V$, namely the vector for which:

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \end{pmatrix} \mapsto \sum_i v_i \beta_i$$

where $\sum_i$ is a finite sum. Then the linear functional $V \to k$ is defined by:

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \end{pmatrix} \mapsto \sum_i v_i x_i$$

where again $\sum_i$ is a finite sum since only finitely many v_i (and x_i) are nonzero Let's label this functional f_x. The element $x \in V$ may be interpreted, functionally, as $t \to tx$ or $v \to \sum_i v_i x_i$. Geometrically, the first function can loosely be thought of as defining a magnitude and direction. This *not* is a correct interpretation in infinite dimensional (in general) as we shall see in section 13.5, however it is valid in finite dimension[21]. The second function can is the "dual" notion, where instead of focusing on x, the focus is on how x manipulates every other element. The reader may want to ponder how this relates to the "internal" and "external" way of describing object mentioned at the beginning of section 13.3. This also shows that V always *injects* into V^*.

As mentioned, it is not always the case that $V \cong V^*$. This immediately falls apart if V is infinite dimensional. To demonstrate this, let us consider the duality presented above where each $x \in V$ is associated with the functional f_x. The key is that given any basis β for V, $[x]_\beta$ has only finitely many components. Hence, the following valid function $F \in V^*$ is not associated to any f_x:

$$[v]_\beta \to \sum_i v_i$$

where the sum is finite since v when represented as an infinite tuple must only have finitely many nonzero components. The functional F is associated to the sequence $(1, 1, ..., 1, ...)$, namely the sequence where 1 is in each position. No f_x can be associated to F since by construction f_x, when represented as an infinite tuple, must have all but finitely position zero. This immediately shows that V^* is much larger, and hence contains much more information than the elements of V can represent. In fact, V^* contains even more interesting functionals. Consider:

$$(v_1, v_2, ..., v_n, ...) \mapsto (0, v_1, ..., v_{n-1}, ...)$$

Notationally, this may be represented as $f(\beta_i) = \beta_{i+1}$ given a basis β for V. This is called he *right shift operator*. This linear transformation can only be defined since there are infinite dimensions, hence each component may shift over by 1.

There are in fact so many linear functionals on infinite dimensional vector spaces that the study of such spaces is almost always accompanied with additional structure to bring some more regularity to the elements. The core problem stems from the extra flexibility working with infinitely many dimensions entails. Hence, some form of "finiteness" is usually imposed via the introduction of a *norm*, or a particular powerful norm called an *inner product*. In the case where there is an inner product on a vector space V, the isomorphism $V \cong V^*$ is recovered where the isomorphism is an inner-product space isomorphism. The inner product gives even more information. As shown in exercise ref:HERE, $(-)^*$ is not a functor since there is no choice of basis that shall naturally allow for composition of linear transformations. In the case of there being an inner product, this issue will be resolved, namely for each matrix representation $[L^*] = [L]^t$, there shall exist a linear transformation L^t for which $L^t : W \to V$ is a linear transformation that shall make $(-)^*$ (resp. $(-)^t$ a functor! This new linear transformation L^t shall be called the *adjoint* of L. The exploration of these notions are very fruitful even in finite dimensions and hence is relegated to section 13.5.

need to incorporate this: this link

Exercise 13.3.1

1. Some things with infinite dimensions

[21] This interpretation holds true in Hilbert Spaces

2. Let $U \subseteq V$ be a subspace. Show that $U^{00} \cong U$. If $\widehat{(__)}$ is the natural embedding of V into V^{**}, then $\hat{U} = U^{00}$
3. Let U_1, U_2, V be finitely dimensional vector spaces. Show that:
   1. $(U_1 + U_2)^0 = U_1^0 \cap U_2^0$
   2. $(U_1 \cap U_2)^0 = U_1^0 + U_2^0$
   3. $U^{00} = U$
   4. If $W \subseteq V$, then $(V/W)^* = W^*$
4. Consider the category whose only object is a vector-space V and morphisms are all linear transformations from V to itself. Choosing a basis for V, show that $(-)^*$ is a contravariant-functor.

13.3.2 Multilinear Forms

Studying linear functionals allow us to give some scalar value to vectors. However, this is sometimes not enough: it is very common in mathematics that one of the two scenarios occur (we will see examples of these in this and coming section[22])

1. We want to study many functional on spaces simultaneously (ex. both V and V^*)

2. We require many copies of the same space to have enough "data" to define the desired functional

In this section, we elaborate on the theory allowing us to do just that, give some examples elaborating the above intuitions, and in the coming section we will see how these are of great use in defining some invariants for linear transformations and bringing in geometry into vector-spaces:

[22]There are also many examples in differential geometry, such as covariant/contravariant/mixed tensor fields on a manifold

Definition 13.3.13: Bilinear and Multilinear Maps

Let V and W be vector spaces over k. Then a map $\varphi : V \times W \to k$ is called *bilinear* if

1. for any $v \in V$, the map $\varphi(v, \cdot) : W \to k$ mapping $w \mapsto \varphi(v, w)$ is linear
2. for any $w \in W$, the map $\varphi(\cdot, w) : V \to k$ mapping $v \mapsto \varphi(v, w)$ is linear.

More generally, if $V_1, V_2, ..., V_n$ are vector-space sover k, then $\varphi : V_1 \times V_2 \times \cdots \times V_n \to k$ is *multilinear* (or n-linear if the number of variables is known) if for all i, $1 \leq i \leq n$, given fixed $v_1, v_2, ..., v_{i-1}, v_{i+1}, ..., v_n$, the map

$$\varphi(v_1, v_2, ..., v_{i-1}, \cdot, v_{i+1} ..., v_n)$$

is linear. Two important types of multilinear maps are:

- (Symmetric) for any $v_i, v_j \in V$:

$$\varphi(v_1, ..., v_i, ..., v_j, ..., v_n) = \varphi(v_1, ..., v_j, ..., v_i, ..., v_n)$$

- (Alternating) for any $v_i, v_j \in V$:

$$\varphi(v_1, ..., v_i, ..., v_j, ..., v_n) = -\varphi(v_1, ..., v_j, ..., v_i, ..., v_n)$$

39: WARNING

Since we are now dealing with multilinear maps, these are *no longer morphisms in the category of* Vect_k! Hence, multilinear maps are a different object of study. In chapter 15, we will show how we can study multilinear maps via linear algebra.

Example 13.3.3: Bilinear and Multilinear Maps

1. Let $\varphi : V \times V^* \to k$ be defined by $(v, f) \mapsto f(v)$. Then φ is bilinear
2. Let $V = \mathbb{R}^n$ be a vector-space over $\mathbb{R}$ and take $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ to be defined as:

$$\varphi\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}\right) = \sum_{i=1}^{n} x_i y_i$$

   Then φ is bilinear, and is called the *dot product* or *scalar product* of two vectors $x, y \in \mathbb{R}^n$
3. Let $\mathbb{R}$ be a vector-space over itself, and take $V : \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R} \to \mathbb{R}$. Then the map that multiplies all the elements:

$$V(x_1, x_2, ..., x_n) = x_1 x_2 \cdots x_n$$

   is a n-linear. As a paticualr example, consider $\varphi : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, and imagine any 3-tulpe (x_1, x_2, x_3) represents the length, width, and height of a shape. Then $V(x_1, x_2, x_3) = x_1 x_2 x_3$

will give the volume of a rectangular prism (or a cube if $x_1 = x_2 = x_3$). The fact that multiplication produces a natural multi-linear map means that multilienar maps show up a lot in geometry. it is for this reason that multilinear maps are sometimes called *volume forms*. We will see more of these in section 15.4.

4. (If you know some calculus) Let $V = C([0,1], \mathbb{R})$ be the set of all continuous functions from $[0,1]$ to $\mathbb{R}$. Then define $\varphi : V \times V \to \mathbb{R}$ to be:

$$\varphi(f,g) = \int_0^1 f(t)g(t)\,dt$$

Then φ is bilinear. This is the infinite dimensional equivalent of the dot product.

In the rest of this section, we will focus on Bilinear forms: the results can easily be generlized using induction to Mutilinear forms. The first thing worth trying to find is wether there is still a reasonable notion of matrix representaiton of bilinear forms, since they are no longer linear maps. The following definition and lemma give us jut that:

Definition 13.3.14: Bilinear Form Matrix Representation

Let V and W be finite dimensional vector spaces over k of dimension n and m and with basis β, γ respectively. Let $\varphi : V \times W \to k$ a bilinear form. Then define the *matrix representation* of φ to be:

$$[\varphi]_{\beta,\gamma} := (\varphi(\beta_i, \gamma_i))_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}}$$

Lemma 13.3.15: Bilinear Form Matrix Represetnation

Let V and W be finite dimensional vector spaces over k of dimension n and m and with basis β, γ respectively. Let $\varphi : V \times W \to k$ a bilinear form. Then

$$\varphi(v,w) = [v]_\beta^t [\varphi]_{\beta,\gamma} [w]_\gamma$$

Proof **:**
This is a direct computation, and so left as an exercise

(Should go over change of basis and look over orthogonality in terms of bilinear maps, and defining orthogonality)

Definition 13.3.16: Orthogonal

Let $\varphi : V \times V \to k$ be a bilinear form and $S \subseteq V$ a *subset*. Then the set:

$$\langle S \rangle^\perp := \{y \in V \mid \varphi(x,y) = 0\}$$

is called the *orthogonal compliment* of W. If $f(x,y) = 0$, then x is *perpendicular* to y, and we will write $x \perp y$. If $x \perp y$ for all $y \in S$, then $x \perp S$.

It should be checked that for any $S \subseteq V$, $S^\perp$ is always a subspace of V.

Exercise 13.3.2

1. Let $f : V \times W \to k$ be a bilinear form. Show that $f : V/W^\perp \times W/V^\perp \to k$ is non-degenerate.

13.4 Invariants of Linear Operators

In this section, we will go over to common functionals on $M_n(k)$ (equivalently, $\mathrm{End}_k(V)$ for an n-dimensional vector-space) that give more insight on linear transformations. A good chunk of the following theory works well over more general rings, and so we will sometimes work over a general commutative ring R.

13.4.1 Trace and Determinant

The first functional on $M_n(k)$ we will define is called the *trace* and will allow us to detect matrices are similar to each other (i.e. the two are related by a change of basis):

Definition 13.4.1: Trace

Let $M_n(R)$ be the collection of all $n \times n$ matrices over k. Then $\mathrm{tr} : M_n(R) \to R$ is the map:

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$$

The trace is studied for the following very useful properties:

Proposition 13.4.2: Properties of Trace

Let $A, B \in M_n(R)$. Then:

1. $\mathrm{tr}(A) = \mathrm{tr}(A^t)$
2. $\mathrm{tr}(AB) = \mathrm{tr}(BA)$

Proof **:**
computational exercise

As a consequence, if A and B are equivalent so that there exists a P such that $A = PBP^{-1}$, then:

$$\mathrm{tr}(A) = \mathrm{tr}(PBP^{-1}) = \mathrm{tr}(P(BP^{-1})) = \mathrm{tr}((BP^{-1})P) = \mathrm{tr}(BPP^{-1}) = \mathrm{tr}(B)$$

Hence the trace is invariant under change of basis, making it a good tool to study linear maps without needing to pick a basis

Definition 13.4.3: Trace Of Linear Map

Let $f : V \to W$ be a linear map and let β, γ be basis for V and W respectively. Then define $\operatorname{tr}(f)$ to be

$$\operatorname{tr}(f) := \operatorname{tr}([f]_\beta^\gamma)$$

Corollary 13.4.4: Trace Of Dual

Let $f : V \to W$ and $f^* : W^* \to V^*$ be the dual map of f. Then:

$$\operatorname{tr}(f) = \operatorname{tr}(f^*)$$

Proof **:**
exercise

(check this out to: geometric interpretation of trace)

The next map is a very common multilinear form that will be used extensively: this form will also be independnet of basis, but with the extra information giving a very clear geometric picture of how a linear transformation effects a 1-cube in $\mathbb{R}^n$:

Definition 13.4.5: Determinant

Let R be a ring and $A = (a_{ij}) \in M_n(R)$ a square matrix. then the *determinant* of A is a map $\det : M_n(R) \to R$:

$$\det(A) := \sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^n a_{i\sigma(i)} \in R$$

Proposition 13.4.6: Properties Of The Determinant

Let $A \in M_n(R)$. Then:

1. $\det(A) = \det(A^t)$, here $(a_{ij})^t = (a_{ji})$
2. if two rows or columns are equal (or equal if they are a unit multiple), then $\det(A) = 0$.
3. (multi-linearity)

Proof **:**
here

Determinents will have the incredible property of being able to detect nonsingualr matrice. To see this, let's take a moment study the determinent of serval maps of the form $f : \mathbb{R}^2 \to \mathbb{R}^2$ (HERE, show the geometric intuition)

With this geometric intuition, we will now proceed with how to capture it algebraically:

Proposition 13.4.7: Elementary Matrices And Determinant

Let A be a matrix then

1. (sitching rows) $\det(A') = -\det(A)$
2. (adding copies of one ro to another) $\det(A') = \det(A)$
3. (multiplying one ro by a unit c) $\det(A') = c\det(A)$

Proof **:**
here

Proposition 13.4.8: Invertability And Determinants

Let R be a commutative ring. Then:

1. $A \in M_n(R)$ is invertible if and only if $\det(A)$ is a unit in R
2. the determinante is a homomorphism $\det : \mathrm{GL}_n(R) \to R^\times$, that is, for any $A, B \in \mathrm{GL}_n(R)$:
$$\det(A \cdot B) = \det(A) \cdot \det(B)$$

Proof **:**
(if $R = k$ is a field)

(words here)

Proof **:**
(if R is a general commutative ring)

We will finish off this section an application of determinents to solve systems of linear equations. This will become particularly useful for a few proofs in the book[23].

Proposition 13.4.9: Cramer's Rule

Let R be a commutative ring, $A \in M_n(R)$, and $\det(R)$ be a unit of R. Let $A^{(j)}$ be the matrix obtined by replacing the jth column of A by the column vector b. Then:

$$x_j = \det(A)^{-1} \det(A^{(j)})$$

Proof **:**

(Also really want to put down the "matrix invertible if determant maps to units" result)

[23] In particular, proposition 20.5.3

The final topic covered in this section will be that of defining the *volume*:

Definition 13.4.10: Volume

Let V be an inner product space and $e = (e_1, e_2, ..., e_n)$ an orthonormal basis for V. Then if $v = (v_1, v_2, ..., v_n)$ is a set of vectors given by a change of basis matrix $v = Ae$, *volume* of v is

$$\mathrm{vol}(v) = |\det(A)|$$

The absolute value around $\det(A)$ is to eliminate orientation information. Geometrically, this can be seen as the volume of a parallellapided when $v_1, v_2, ..., v_n$ are linearly independent, namely if

$$P = \left\{ \sum a_i v_1 \;\middle|\; 0 \leq x_i < 1 \right\}$$

Then the intuitive notion of volume of P agrees with our definition. The case where $v = (e_1, e_2, ..., e_n)$ shows that the volume of a vector of unit lengths is always 1.

Exercise 13.4.1

1. Let $A \in M_n(k)$, $k \subseteq \mathbb{C}$. Show that $\det(\overline{A}) = \overline{\det(A)}$ where we define $\overline{A}$ component-wise to be $(\overline{A})_{ij} = \overline{a_{ij}}$ with $\overline{a_{ij}}$ being the complex conjugate ($\overline{x + iy} = x - iy$)
2. Let $A \in M_n(k)$. Define the *adjucate* of A to be $\mathrm{Ajd}(A) = C^T$ where C is the cofactor expansion matrix of A. show that if $\det(A) \neq 0$, then $A^{-1} = \det(A)^{-1}\,\mathrm{Adj}(A)$.
3. (If you konw multi-variable calculus) Treating $M_n(\mathbb{R})$ as $\mathbb{R}^{n^2}$, show that the dervative of the determinant map $\det : M_n(\mathbb{R}) \to \mathbb{R}$ at the identity I is the trace tr: $(\det)'(I) = \mathrm{tr}$. This gives a surprisingly interesting link between the two concepts, and is important in the beginning of the study of *Lie Algebras* of many matrix groups such as $\mathrm{GL}_n(k)$, $\mathrm{SL}_n(k)$, $O_n(k)$, and so forth, where $k \in \{\mathbb{R}, \mathbb{C}\}$.
4. More generally, show that $\det : M_n(\mathbb{R}) \to \mathbb{R}$ has a derivative at any point A $(D_A \det)(X) = \mathrm{tr}(\mathrm{Adj}(A^t)X)$.

***Resultant**

Before we leave dual spaces and linear functionals, we will introduce one very useful functional on $R[x]$ that allows us to detect when a polynomial has a root.

(Build up how the determinant can be used to see if two polynomials havea common root)

Theorem 13.4.11: Resultant

Let R be a commutative ring and $f(x), g(x) \in R[x]$ be two polynomials:

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \quad g(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0$$

Then if the determinant:

$$\operatorname{Res}(f,g) := \det \begin{pmatrix} a_n & 0 & \cdots & 0 & b_m & 0 & \cdots & 0 \\ a_{n-1} & a_n & \cdots & 0 & b_{m-1} & b_m & \cdots & 0 \\ a_{n-2} & a_{n-1} & \cdots & 0 & b_{m-2} & b_{m-1} & \ddots & 0 \\ \vdots & \vdots & \ddots & a_n & \vdots & \vdots & \ddots & b_m \\ a_0 & a_1 & \cdots & \vdots & b_0 & b_1 & \cdots & \vdots \\ 0 & a_0 & \ddots & \vdots & 0 & b_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & a_{n-1} & \vdots & \vdots & \ddots & b_1 \\ 0 & 0 & \cdots & a_0 & 0 & 0 & \cdots & b_0 \end{pmatrix}$$

is zero, then f and g have a comon root

Proof **:**
here

Definition 13.4.12: Resultant

The determinant in theorem 13.4.11 $\operatorname{Res}(f,g)$ is called the *resultant*.

(The answer on how to find a polynomial with roots of a^2 given a is a really cool application of the resultant: here)

13.4.2 Characteristic Polynomial and Eigenvectors

After having found the trace and determinant, the reader may wonder if there are more invariant functionals on $\operatorname{End}_k(V)$ for an n-dimensional vector space. As it turns out, these are both special cases of a collection of n invariant given by the coefficients of the following polynomial:

Definition 13.4.13: Characteristic Polynomial

Let M be a free R-module and let $\alpha \in \operatorname{End}_R(M)$. Let $I \in \operatorname{End}_R(M)$ denote the identity operator. Then the characteristic polynomial is defined to be:

$$P_\alpha(t) := \det(tI - \alpha) \in \operatorname{End}_R(M)$$

which when α has a basis representation A, then $P_\alpha(t)$ is an element of $R[t]$, R-polynomials with indeterminate t.

The origins of this seemingly complicated looking polynomial comes from its link to eigenvectors and eigenvalues (namely scalars and vector of a linear operator that satisfy $\alpha(v) = \lambda v$ for appropriate $\lambda \in R$ and $v \neq 0$) which shall be explored soon. Some important properties are established first:

Proposition 13.4.14: Properties Of The Characteristic Polynomial

Let $P_\alpha(t)$ be the characteristic polynomial of the linear operator $\alpha : M \to M$ for a free R-module of rank n. Then:

1. $P_\alpha(t)$ is monic of degree n
2. The coefficient of t^{n-1} term is $-\operatorname{tr}(\alpha)$
3. The constant term is $(-1)^n \det(\alpha)$
4. If α and β are similar, then $P_\alpha = P_\beta$

Proof **:**

1. Expanding the definition immediately reveals the result. Let $A = (a_{ij})$ be a representative of α given some choice of basis. Then:
$$P_\alpha(t) = \det \begin{pmatrix} t - a_{11} & -a_{12} & \cdots & -a_{1n} \\ t - a_{21} & t - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & t - a_{n2} & \cdots & t - a_{2n} \end{pmatrix}$$
2. Choose A as above. Then notice that when expanding $P_\alpha(t)$, the values of the t^{n-1} term only come from terms on the diagonal, namely we get:
$$\sum_i^n t \cdot t \cdot \cdots \cdot t \cdot a_{ii} \cdot t \cdot \cdots \cdot t \cdot t$$
giving us the desired result
3. Setting $t = 0$ and using multi-linearity gives the desired result
4. Assuming $\alpha \sim \beta$, there exists a ρ such that
$$\alpha = \rho \circ \beta \circ \rho^{-1}$$
Then by using properties of functions, we see that:
$$tI - \alpha = \rho \circ (tI - \beta) \circ \rho^{-1}$$
But then, the determinant of both sides are the same. Choosing a matrix representation P for ρ and B for $tI - \beta$ we see that by proposition 13.4.8
$$\det(PBP^{-1}) = \det(P)\det(B)\det(P)^{-1} = \det(B)\det(P)\det(P)^{-1} = \det(B)$$
hence
$$P_\alpha(t) = \det(tI - \alpha) = \det(\rho \circ (tI - \beta) \circ \rho^{-1}) = \det(tI - \beta) = P_\beta(t)$$
completing the proof.

Hence, the characteristic polynomial is of the form:

$$t^n - (\operatorname{tr}\alpha)t^{n-1} + \cdots + (-1)^n(\det\alpha)$$

As far as I am aware, the other coefficients of this polynomial don't have a special name like the trace and determinant, however on the other hand these coefficients are special since they always appear in any dimension (they are equal when $n = 1$ for trivial reasons, the determinant is zero if α is not surjective, and the trace is zero if α is a commutator, see exercise ref:HERE not done yet see this). Furthermore, note that all coefficients are invariant under similarity, meaning that if the trace or determinant between two linear operators differ, they are *not* similar! Unfortunately, the converse does not hold:

Example 13.4.1: Similar Matrices

Let:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

These each have a characteristic polynomial of $t^2 - 1$. However, they are not similar: Notice that A is the identity, and conjugating A by any invertible matrix P would give the identity, and hence never B.

As mentioned earlier, the characteristic polynomial naturally arises when studying eigenvector and values. Let us start by formally defining these terms:

Definition 13.4.15: Eigenvalue and Spectrum

Let M be a free R-module over a commutative ring R and let $\alpha \in \operatorname{End}_R(M)$ be a linear operator on M. Then a scalar $\lambda \in R$ is called an *eigenvalue* of α if there exists a nonzero $v \in M$ such that

$$\alpha(v) = \lambda v$$

The collection of all eigenvalues is called the *spectrum* of α.

Eigenvalues are extremely important having a wide range of application in quantum mechanics (the term *spectrum of an atom* comes from finding the spectrum of an endomorphism) functional analysis (K-theory has its origin in the spectrum of infinite linear operators), differential equations (in finding a basis to the solutions of an ODE), to number theory and Field theory (as we shall see later in the book), and so much more! Many even claim that the study of endomorphisms of vector-spaces is one of the fundamental concepts of mathematics or the sciences. In chapter 21, it shall be shown that any commutative ring will have a spectrum associated to (though by the definition of a spectrum of a ring, the connection shall not be obvious[24]).

It is easy to see that the eigenvalue is invariant under similarity, which the reader should attempt to prove before looking at the following: Letting α, β be similar using ρ ($\alpha = \rho \circ \beta \circ \rho^{-1}$) and $beta(v) = \lambda v$, then $\alpha \circ \rho = \rho \circ \beta$ and

$$\alpha(\rho(v)) = (\rho \circ \beta \circ \rho^{-1})(\rho(v)) = \rho \circ \beta(v) = \rho(\lambda v) = \lambda\rho(v)$$

[24]The curious reader may checkout EYNKTA functional analysis where this connection is fully fleshed out, as the connection is fundamentally a functional analysis result and would be too far out of the scope of an algebra textbook to be covered

since $\rho(v) \neq 0$ if $\lambda \neq 0$ (namely, it is invertible), we see that λ is an eigenvalue of α if and only if λ is an eigenvalue of β.

Now, as promised, the following shows why the characteristic polynomial is a natural concept based on its invariance in similarity:

Proposition 13.4.16: Eigenvalues And Characteristic Polynomial

Let $P_\alpha(t)$ be the characteristic polynomial of the linear operator α. Then λ is an eigenvalue if and only if λ is a root of $P_\alpha(t)$. In other words, the set of roots of $P_\alpha(t)$ is exactly the spectrum of α.

Proof **:**
Using the properties of determinants:

$$\begin{aligned}
\lambda \text{ is an eigenvalue} &\Leftrightarrow \exists v \neq 0 \text{ s.t. } \alpha(v) = \lambda v \\
&\Leftrightarrow \exists v \neq 0 \text{ s.t. } (\lambda I - \alpha)(v) = 0 \\
&\Leftrightarrow (\lambda I - \alpha) \text{ is not injective} \\
&\Leftrightarrow \det(\lambda I - \alpha) = 0 \\
&\Leftrightarrow P_\alpha(\lambda) = 0
\end{aligned}$$

as we sought to show.

Example 13.4.2: Eigenvalues And Characteristic Polynomials

1. If α is not injective, then 0 is an eigenvalue, otherwise 0 is not an eigenvalue.
2. Take
$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$
This matrix has a characteristic polynomial of $t^2 + 1$. If $R = \mathbb{R}$, then this characteristic polynomial has *no* roots. Similarly, the linear operator this matrix represents has no eigenvalues. Geometrically, this matrix can be seen as rotating $\mathbb{R}^2$ by $\pi/2$ degrees, which is why it has no eigenvalues over $\mathbb{R}$. If $R = \mathbb{C}$, then the polynomial has two roots, namely $i, -i$. The reader should verify that these are eigenvalues of a linear operator represented by this matrix. Intuitively, multiplying by i can be thought of as representing rotation by $\pi/2$ degree's which is why it is an eigenvalue.
3. Take
$$\begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$$
This matrix has a characteristic polynomial of $t^2 - 2$. If $R = \mathbb{R}$, then this polynomial has roots $\pm\sqrt{2}$, while if $R = \mathbb{Q}$, this polynomial has no roots, which equivalently implies that there is no eigenvalues when $R = \mathbb{Q}$. Geometrically, this transformation can be seen as reflecting the plane across a line passing through the origin with an irrational slope. Since

the slope is irrational, there is no nonzero irrational scalar to a vector, and hence cannot be a rational eigenvalue.

Geometrically, eigenvectors correspond to the linear operator scaling in the appropriate direction. In the simplest case, an n-dimensional linear operator α can be represented in a basis for which $[\alpha]$ just scales appropriately in each direction. We take a moment to explore these type of matrices.

Since roots have the notion of multiplicity (the number of roots sharing the same value), it natural to contemplate the multiplicity of eigenvalues and see what information this gives us:

Definition 13.4.17: Algebraic Multiplicity

Let λ be an eigenvalue of a linear operator α over a finitely generated free module. Then the *algebraic multiplicity* is its multiplicity as a root of $P_\alpha(t)$.

Example 13.4.3: Algebraic Multiplicity

Let M be a finitely generated free module over an integral domain R.

1. the identity operator I on M has a single eigenvalue, namely 1, which has algebraic multiplicity equal to the rank of M.
2. (give a non-trivial example)

Notice that the sum of the algebraic multiplicities is bounded by the rank of the free module M (resp. the dimension of the vector-space V). If R is *algebraically closed* (meaning every polynomial of degree n over R has exactly n roots[25]), then the sum of the algebraic multiplicities of the eigenvalues always equals the rank of M. The prototypical algebraically closed field is $\mathbb{C}$, the complex numbers.

The definition has been careful to call the multiplicity "algebraic". There is another notion of multiplicity which corresponds to the geometric behavior of multiplicity, namely there is a special subspace that is linked to each eigenvalue:

Definition 13.4.18: Eigenspace And Eigenvector

Let λ be an eigenvalue associated to the linear operator α over a finitely generated free R-module M. Then the sub-module

$$E_\lambda = \ker(\lambda I - \alpha)$$

is called the *eigenspace* associated to λ, and the elements of E_λ are called *eigenvectors* and represent the elements of M such that

$$\alpha(v) = \lambda v$$

Technically, as we have not limited to subspaces but sub-modules, it is more correct to call these spaces "eigen-modules", however the study of eigen-modules over vector-spaces greatly preceeds the generalization to modules, giving us our current vocabulary.

[25] See definition 17.5.1 and the lemma under it for more details

Definition 13.4.19: Geometric Multiplicity

Let E_λ be an eigenspace corresponding to an eigenvalue λ. Then the rank of E_λ is the *geomeric multiplicity* of λ.

It should be shown that the eigenspace is invariant under similarity, meaning it a special subspace that is invariant under α (we shall later call spaces that are invariant under an operator α-stable, where α is the operator in question).

Example 13.4.4: Geometric And Algebraic Multiplicity

Take

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Then the characteristic polynomial is $(t-1)^2$, and hence 1 is an eigenvalue with algebraic multiplicity 2. To find it's geometric multiplicity, by considering

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 + v_2 \\ v_2 \end{pmatrix}$$

we see that the right hand side is equal to $1v$ if and only if $v_2 = 0$, hence the eigenspace is

$$\operatorname{span} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

which is 1 dimensional, hence the geometric multiplicity is 1.

A result that should stand out to the reader given the above example is the following

Proposition 13.4.20: Algebraic And Geometric Multiplicity

Let λ be an eigenvalue associated to the linear operator α over a finitely generated free R-module M. Then the geometric multiplicity is always less than or equal to the algebraic multiplicity

Proof **:**

exercise

One of the main results comes in the algebraic multiplicities add up to the rank of the free module and when the algebraic multiplicity matches the geometric multiplicity. Then there exists a basis P making α a bunch of scalars, namely:

Theorem 13.4.21: Diagonalizable

Let α be a linear operator over a finitely generated free R-module M. Then if the algebraic multiplicities of the eigenvalues sum to n and the algebraic multiplicities of each eigenvalue equal to their respective geometric multiplicities, there exists an invertible matrix $P \in \mathrm{GL}_n(R)$ such that for any representation A of α:

$$A = P \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} P^{-1}$$

where $\lambda_1, \lambda_2, ..., \lambda_n$ are the eigenvalues of α, and α (resp. A) is said to be *diagonalizable*. If $m = V$ is a vector-space, α is diagonalizable if and only if there exists a basis for V consisting of eigenvectors.

Proof **:**
exercise.

Definition 13.4.22: Diagonalizable

Let α be a linear operator over a finitely generated free R-module M. Then if α admits a representation with only nonzero diagonal entries, α is said to be *diagonalizable*. Similarly, a matrix that is similar to a diagonal matrix is said to be *diagonalizable*.

If α is diagonalizable, then M can be expressed in the following useful decomposition:

Theorem 13.4.23: Spectral Decomposition

Let $\alpha : M \to M$ be diagonalizable. Then M admits the following decomposition into direct sums

$$M \cong \bigoplus_{\text{eigenvalue } \lambda} E_\lambda$$

Proof **:**
exercise

For future reference, let's give a name to the above decomposition:

Definition 13.4.24: Spectral Decomposition

Let $\alpha : M \to M$ be diagonalizable. Then the decomposition in the above theorem is called the *spectral decomposition* of M.

Given the spectral decomposition of M, α can simply be thought of as scaling each eigenspace E_λ by the eigenvalue, making the "action" of α on the space M completely transparent!

If the algebraic multiplicity does not sum to the rank of M or the geometric multiplicity does not match the algebraic multiplicity, then the matrix is *not* diagonalizable. If the algebraic multiplicity does not sum to the rank of M, then if we are working over field[26] we can take something called the *algebraic closure* of the field, which as the name implies means it is *alegbraically closed*. To rigorously go over the algebraic closure of a field is too much work at this stage, and so will just be taken for granted when it is brought up, the curious reader may checkout part IV chapter 17, section 17.5.1. With the reader's current knowledge, most of the section can be parsed.

If the geometric multiplicity does not match the algebraic multiplicity, then it is a harder to find a form. However, in the next chapter, we shall see that there exists a representation that is *almost* diagonalizable: it shall be diagonalizable with possibly a 1 in the super-diagonal (the position right above the diagonal). In fact, *all* square matrices will be able to have this almost diagonal form!

13.5 Geometry and Inner Product

As we saw in theorem 13.2.23, vector-spaces have a geometric interpretation by looking at the collection of kernels of linear transformations, namely there is a correspondence between k-dimensional subspace of an n-dimensional vector-space and the kernel of linear functions

$$\left\{k\text{-dimensional planes in } F^n\right\} \underset{\ker\varphi}{\overset{\text{sys. of eq.}}{\rightleftarrows}} \left\{\begin{matrix}\text{Kernels of linear functions}\\ \varphi : F^n \to F^{n-k}\end{matrix}\right\} \tag{13.15}$$

There is more geometric information that may be defined element-wise, for example, what is the size of a vector, distance between two vectors, or the angle between vectors (i.e. a notion of direction)? This information shall be captured by the *inner product* on a vector-space. It shall turn out that the inner product (and hence the notion of size and angle) cannot be defined on any type of vector-spaces, namely the underlying field may not allow for some of these notions to be defined (see exercise ref:HERE-ref:HERE). For example, for the notion of size, we need to have the notion of an element being bigger than other, and hence introduce an *order* (something not all fields exhibit). Another notion we need is for there to be "positive numbers" for we want all vectors to have positive length, as well as the notion of *square-root* (recall the Pythagorean Theorem). Of the fields we studied, $\mathbb{R}$ is a natural choice over which to define the inner product since many fields satisfying these properties may embed into $\mathbb{R}$, not to mention that $\mathbb{R}$ has other nice properties useful outside of algebra[27].

Though $\mathbb{R}$ is the natural choice for the above reason, $\mathbb{C}$ is in fact another natural choice of underlying field. The reason for this lies within [differential] geometry, and hence any reader unmotivated by such considerations can skip this paragraph. The real numbers are the natural geometric first choice as we may think of any shape as being a subset of $\mathbb{R}^n$ with extra properties[28]. As topologically, $\mathbb{C} \cong_{\mathbf{Top}} \mathbb{R}^2$, it may seem that we ought to stick to geometry over $\mathbb{R}$. However, the complex numbers have the additional property of being *naturally orientable*, namely by multiplying by i, we naturally turn counter-clockwise, while in $\mathbb{R}^2$ there is no reason to choose one orientation over the other (i.e. to orient $\mathbb{R}^2$ clock-wise or anti-clockwise). This in fact has far-reaching consequences for doing geometry over $\mathbb{R}$ as compared to doing geometry over $\mathbb{C}$ (for example, Hamiltonian mechanics is defined over *complex* manifolds due to the natural orientation information the complex numbers provide). Hence,

[26] more generally, over integral domains, since the ring of fractions of an integral domains is a field
[27] Most importantly outside of algebra, it is complete
[28] Namely, Whitney's Embedding theorem says any smooth n-manifold can be embedded in $\mathbb{R}^N$ for large enough N

the complex numbers and the real numbers offer different "geometries"[29]. The fact that there are these two "geometries" is reflected in the inner product being definable[30] *at least* over $\mathbb{R}$ or $\mathbb{C}$. As $\mathbb{R}$ and $\mathbb{C}$ are the only (up to isomorphism) complete fields with a complete Archimedean subfield[31] on which differentiation well-defined, there is no other natural choice of field to consider for the inner product, giving us *at most* $\mathbb{R}$ and $\mathbb{C}$.

The conclusion of the above paragraph is that we may define inner products over two fields, $\mathbb{R}$ and $\mathbb{C}$, producing two different geometries. Mathematically, these share many properties and hence the study of the inner product on these two fields shall be done in-tandem wherever possible.

13.5.1 Basic Properties and Definitions

Definition 13.5.1: Inner Product

Let V be a vector-space over k where k is either $\mathbb{R}$ or $\mathbb{C}$. Then define $\langle \cdot, \cdot \rangle : V \times V \to k$ to be

1. sesquilinear[a]: $\langle a + b, c \rangle = \langle a, c \rangle + \langle b, c \rangle$ and conjugate linear on the second term, and scalar's pass normally in the 1st, and conjugation through the second.
2. (positive definite) $\langle v, v \rangle \in (0, \infty)$ and $= 0$ if and only if $v = 0$
3. (conjugate symmetric) $\langle a, b \rangle = \overline{\langle b, a \rangle}$

[a]The term "sesqui" means $1\frac{1}{2}$, making it an apt choice of terminology

If $k = \mathbb{R}$, then the inner product becomes a symmetric bilinear form that is positive definite.

Example 13.5.1: Inner Product

In all these examples, let $k \in \{\mathbb{R}, \mathbb{C}\}$.

1. Let V be any vector-space over k with basis $\mathcal{B}$. Then define for any $v, w \in V$, we have:

$$v = \sum_{\beta \in \mathcal{B}} v_\beta \beta \qquad w = \sum_{\beta \in \mathcal{B}} w_\beta \beta$$

Since $\mathcal{B}$ is a basis, only finitely many terms in each summand is nonzero. Then the inner product between v, w is defined to be:

$$\langle v, w \rangle := \sum_{\beta \in \mathcal{B}} v_\beta \overline{w_\beta}$$

It should be checked that this is indeed a vector-space. This says that *every* vector-space has an inner product, and justifies our intuitive view of vector-spaces in Cartesian coordinates (since we can think of each dimension as being orthogonal).

Though this is insightful and let's us gain some intuition, it turns out that this inner product is limited to finite-dimensional vector-spaces. Once we work over infinite-dimensional spaces,

[29]If the reader knows differential geometry, prove that all complex manifolds are orientable, which is certainly not the case for real manifolds, ex. the mobius band

[30]This "definability" can be made rigorous using Model Theory

[31]for Archimedean fields, see chapter 19

this particular inner product will not be of much use as compared to some other ones. In particular, we will try to find an infinite dimensional equivalent to a basis (called an orthonormal basis). More on that in a course on functional analysis (for references, see ref:HERE)

2. As a particular example of the above, let $V = \mathbb{R}^n$. Then define:

$$\langle x, y\rangle = \sum_i x_i y_i$$

Then this is an inner product on $\mathbb{R}^n$, called the *dot product* or *scalar product*. As a simple variation of this example, if $r \in \mathbb{R}_{>0}$, then $\langle x, y\rangle' = r\langle x, y\rangle$ is also an inner product.

3. The complex dot product on $\mathbb{C}$ ($\langle x, y\rangle = x\overline{y}$) and real dot product on $\mathbb{R}^2$ behave differently, and hence these two are not isomorphic[a]. For example, in $\mathbb{C}$ $\langle a, b\rangle = 0$ if and only if a or $b = 0$, while in $\mathbb{R}^2$ there are many other possibilities, for example:

$$\left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = 0$$

This can be thought of as $\mathbb{C}$ having an inherent "memory" of the orientation which is reflected in the inner product.

4. (If you know some calculus) Let $V = \mathbb{C}[x]$ be the complex polynomials. Then treating them as functions, define:

$$\langle p(x), q(x)\rangle = \int_0^1 p(x)\overline{q(x)}dx$$

Verify that this is indeed an inner product. If the bounds of the are changed, the inner product gives different results, as the reader should check. Changing the bounds creates different inner products[b]

5. Take $V = M_n(k)$ and define:

$$\langle A, B\rangle = \operatorname{tr}(B^* A)$$

where $B^* = (\overline{B})^t$, that is $(B^*)_{ij} = (\overline{B})_{ji}$. Verify that this is an inner product; this inner product is called the *Frobenius inner product*.

[a] Similarly, these two share different geometries
[b] Though they may be isomorphic inner products, see definition 13.5.2

From now on, we shall assume that if V is finite dimensional then V has the inner-product given in example 13.5.1[1]. This shall be called the *standard inner product* (with respect to a basis β)

Definition 13.5.2: Isometry

Let V_1, V_2 be two inner product spaces with inner products $\langle -, -\rangle_1$ and $\langle -, -\rangle_2$. Then a linear function $f : V_1 \to V_2$ is called an *isometry* if for all $v, w \in V_1$

$$\langle f(v), f(w)\rangle_2 = \langle v, w\rangle_1$$

The reader should wonder whether it is possible to put different inner products on vector-spaces of

the same dimension (whether finite or infinite), which would demonstrate that inner product spaces cannot be characterized using only the dimension. This question shall be answered later, however it is a good type of question for the reader to come up for themselves and ponder.

Proposition 13.5.3: Properties Of Inner Product

Let V be an inner product space over k. Then

1. $\langle x, 0\rangle = \langle 0, x\rangle = 0$
2. $\langle x+y, x+y\rangle = \langle x, x\rangle + \mathrm{Re}(\langle x, y\rangle) + \langle y, y\rangle$
3. If $\langle x, y\rangle = \langle x, z\rangle$ for all x, then $y = z$

Proof **:**
exercise

As we saw in lemma ref:HERE, bilinear forms can be represented as a matrix. Given the inner product is symmetric bilinear when $k = \mathbb{R}$ and conjugate bilinear when $k = \mathbb{C}$, it is reasonable to ask if the matrix representation of an inner product has special properties, and indeed this is the case, most notably representable in finite dimension[32]:

Proposition 13.5.4: Matrix Representation Of Inner Product

Let V be a finite dimensional vector space over k, and choose a basis for V so that $V \cong k^n$. Then a function $\langle -, -\rangle : k^n \times k^n \to k$ is an inner product iff there exists a Hermitian positive-definite matrix $M \in M_n(k)$ such that:

$$\langle x, y\rangle = x^\dagger M y$$

where $x^\dagger$ is the row vector $(x)_{ij} = \overline{(x)_{ji}}$. A matrix is Hermitian positive-definite if:

1. (Hermitian) $M = \overline{M^t}$ (element wise, $m_{ij} = \overline{m_{ji}}$)
2. (Positive definite) $x^\dagger M x \in (0, \infty)$ for all $x \in k^n$

If $k = \mathbb{R}$, then M becomes a symmetric matrix: $M^T = M$.

Proof **:**
Also see this cool post!

Essentially, because any inner product admits an orthonormal basis, they are all equivalent.

The matrix above is important enough to be given a name

[32]The infinite dimensional case is covered in a course in Lie Groups or Functional Analysis

Definition 13.5.5: Hermitian Matrix and conjugate transpose

Let $A \in M_n(k)$. Then if $A = \overline{A^T}$, A is called *Hermitian*. The process of taking the conjugate and the transpose are called the *conjugate transpose* and is denoted

$$A^\dagger = \overline{A^T}$$

Some textbooks denote this as A^*; this books avoids this notation due to it already representing units of a ring.

It was claimed that the inner product will allow vector-spaces to have the notion of size, length, and angle defined on it. We start by defining the size of a vector:

Definition 13.5.6: Size Of Vector

Let V be an inner product space. Then the *size* or *norm* of a vector $v \in V$ is:

$$\|v\| := \sqrt{\langle v, v\rangle}$$

the norm can be treated as a functional $\| - \| : V \to \mathbb{R}_{\geq 0}$ which is called the *induced norm* with respect to the inner product.

Given the standard inner product space, we see that:

$$\|v\| = \sqrt{\sum_i^n v_i^2}$$

which should remind the reader of the Pythagorean theorem. If $n = 1$, then $\|v\| = |v|$, that is $\|v\|$ is the absolute value of v. Given an inner product, one may wonder what sort of relationship $\langle v, w\rangle$ and $\|v\|$, $\|w\|$ may have. The following is the most important identity relating these two:

Theorem 13.5.7: Cauchy-Schwartz Identity

Let V be an inner product space and let $\| - \|$ be the induced norm. Then:

$$|\langle v, w\rangle| \leq \|v\|\|w\|$$

where $| - |$ is the absolute value.

Proof **:**
If $\|y\| = 0$ then $y = 0$, and then $|\langle x, 0\rangle| = 0$, so assume $\|y\| \neq 0$. Then for all $t \in \mathbb{R}$

$$\begin{aligned} 0 &\leq \langle x - ty, x - ty\rangle \\ &= \|x\|^2 + 2t\,\mathrm{Re}\langle x, y\rangle + t^2\|y\|^2 \\ &= \|x\|^2 + \frac{2\,\mathrm{Re}\langle x, y\rangle^2}{\|y\|^2} + \frac{\mathrm{Re}\langle x, y\rangle^2}{\|y\|^2} \qquad t = \frac{\mathrm{Re}\langle x, y\rangle}{\|y\|^2} \\ \Rightarrow |\,\mathrm{Re}\langle x, y\rangle| &\leq \|x\|\|y\| \end{aligned}$$

To finish the proof, pick $\alpha \in \mathbb{C}$ with $|\alpha| = 1$ such that

$$|\operatorname{Re}(\alpha\langle x, y\rangle)| = |\langle x, y\rangle|$$

Notice this does not break the inequality, and so we can repeat the entire process but with α taken into account with t, and so:

$$|\langle x, y\rangle| \leq \|x\|\|y\|$$

as we sought to show

The norm $\| - \| : V \to \mathbb{R}_{\geq 0}$ can be treated as a functional which is characterized by the following properties:

Theorem 13.5.8: Norm

Let V be an inner product space. Then the norm $\| - \|$ satisfies the following 3 properties:

1. $\|v\| = 0$ iff $v = 0$
2. $\|cv\| = |c|\|v\|$
3. (triangle inequality) $\|v + w\| \leq \|v\| + \|w\|$

Proof **:**
exercise (for the triangle inequality, use Cauchy-Schwartz)

This notion is important enough to be given a name:

Definition 13.5.9: Norm

Let V be a vector space. Then a *norm* (on V) is a function $\| - \| : V \to \mathbb{R}_{\geq 0}$ such that

1. $\|v\| = 0$ iff $v = 0$
2. $\|cv\| = |c|\|v\|$
3. $\|v + w\| \leq \|v\| + \|w\|$

A tuple $(V, \| - \|)$ with a norm on is called a *norm space*.

The reader may ask if $(V, \| - \|)$ is a norm, then does this imply that there exists an inner product $\langle -, - \rangle$ on V such that $\| - \| = \sqrt{\langle -, - \rangle}$? Interestingly, the answer is no. Thus, all inner products give rise to norms, but not all norms give rise to inner products. This demonstrates that inner products enforce greater restrictions on vector-spaces. Norms that induce inner products satisfy the following identity:

Proposition 13.5.10: Polarization Identity

Let V be a normed vector-space. Then $\| - \|$ satisfies the *polarization identity*

$$\|v + w\|^2 + \|v - w\|^2 = \|v\|^2 + \|w\|^2$$

if and only if there exists an inner product $\langle -, - \rangle$ such that

$$\| - \|^2 = \langle -, - \rangle$$

Proof **:**
exercise

In the finite dimensional case, all norms satisfy the polarization identity, and hence these considerations are more interesting for infinite-dimensional vector-spaces. For the interested reader, the norm defined on the space of real continuous functions on $[0, 1]$ (denoted $C([0, 1])$) given by:

$$\|f\| = \sqrt{\int_0^1 f(x)dx}$$

Does not satisfy the polarization identity, and hence does not induce an inner product. Geometrically, the polarization identity tells us that norms that are "addition-invariant" are inner products, and all inner products satisfy this invariant property[33].

If $(V_1, \| - \|_1), (V_2, \| - \|_2)$ are two normed vector spaces, then they are considered isomorphic as normed vector-spaces if there exists a linear transformation $f : V_1 \to V_2$ such that for all $v \in V_1$, $\|f(v_1)\|_2 = \|v_1\|_1$. Such a function is called an *isometry*. The reader may have noticed that an inner-product isomorphism is already called an isometry. That is because when a normed-space V induces an inner product notions of isomorphism are identical:

Theorem 13.5.11: Isomorphisms For Norms And Inner Products

Let V_1, V_2 be a normed-spaces such that their norms induce an inner product. Then V_1, V_2 are isomorphic as normed spaces if and only if they are isomorphic as inner-product space. Hence, isomorphisms of normed spaces and inner product spaces are both called *isometries*

Proof **:**
Exercise.

This result emphasizes that inner product spaces can be thought of as special normed spaces, and gives credence in thinking that inner-products are normed spaces in which are addition-invariant. It is through the fact that inner-products are addition-invariant that the notion of angle is well-defined in inner-product spaces. To define angles in inner-product spaces, let us start by defining some terms. Given any vector v, we may *normalize* the vector v by scaling it by $1/\|v\|$:

$$\frac{v}{\|v\|}$$

[33] Similarly, metrics which are scalar invariant induce a norm

Any normalized vector has norm 1, or unit length. Given two vectors v, w, we may normalize them both to get them to unit length. Then we may define the *angle* between v, w by the following formula:

$$\cos(\theta) := \frac{\langle v, w\rangle}{\|v\|\|w\|}$$

If $v = w$, Then:

$$\cos(\theta) := \frac{\langle v, v\rangle}{\|v\|\|v\|} = \frac{\|v\|^2}{\|v\|^2} = 1$$

which can be interpreted as the two vector pointing in the same direction. If $v = -v$, then:

$$\cos(\theta) := \frac{\langle v, -v\rangle}{\|v\|\|v\|} = -\frac{\|v\|^2}{\|v\|^2} = -1$$

which can be interpreted as the two vector pointing in diametrically the opposite direction. Given the definition of $\cos(\theta)$, we see that:

$$-1 \leq \cos(\theta) \leq 1$$

Showing that this numbers gives some notion of correlation of direction between two vectors. If

$$\frac{\langle v, w\rangle}{\|v\|\|w\|} = 0$$

Then we give a name to these vectors:

Definition 13.5.12: Orthogonal and Orthonormal

Let $v, w \in V$ be two vectors in an inner product space. Then if

$$\frac{\langle v, w\rangle}{\|v\|\|w\|} = 0$$

we say that v and w are *perpendicular* or *orthogonal* and write $v \perp w$. A collection of pair-wise orthogonal elements is called an *orthogonal set*. If each vector is also of unit length, the set is called an *orthonormal set*.

Example 13.5.2: Orthonormal Sets

1. Let $V = k^n$. Then the standard basis is an orthonormal set, namely if $\{e_i\}_{i=1}^n$, then:

$$\langle e_i, e_j\rangle = \delta_{ij}$$

where

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

2. (If you know some Analysis) The following is the foundation of Fourier Analysis. Let $C(S^1)$ be the set of all complex continuous functions on the circle (more generally $L^2(S^1)$). We can parameterize the circle using the with coordinates representing the angle ranging in $[0, 2\pi)$ and hence this can be thought of as $C([0, 2\pi))$. Define $f_n : S^1 \to \mathbb{C}$, $f_n(t) = e^{int}$. Define

inner product

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f\overline{g}dx$$

similar to the inner product in example 13.5.1[4]. Then the collection $\mathcal{F} = \{f_n \mid n \in \mathbb{Z}\}$ is an orthonormal set (recall, or re-prove, that $\overline{e^{it}} = e^{-it}$).

Thus, the rather complicated vector-space of functions has an inner product with an orthonormal basis! If the reader is unimpressed (as admittedly I was at first), it is highly encouraged that they try to come up a qualitatively different orthonormal set.

Orthogonal vectors allow the generalization of a famous mathematical theorem:

Theorem 13.5.13: Pythagorean Theorem

Let $v, w \in V$ be vectors in the inner product space V such that $v \perp w = 0$. Then:

$$\|v + w\|^2 = \|x\|^2 + \|y\|^2$$

Proof **:**
exercise

A subset of particular interest for vector-spaces is a *basis*. A basis that is also an orthonormal set is called an *orthonormal basis*. Given any basis β for a finite-dimensional vector-space V, it can always induce an orthonormal basis:

Theorem 13.5.14: Grand-Schmidt Theorem

Let V be an n-dimensional inner product space, and suppose $(v_1, v_2, ..., v_n)$ is any ordered basis for V. Then there is an orthonormal ordered basis $(b_1, b_2, ..., b_n)$ satisfying the following condition:

$$\text{span}(b_1, b_2, ..., b_k) = \text{span}(v_1, v_2, ..., v_k)$$

for all $k \in \{1, 2, ..., n\}$

Proof **:**
Define the basis recursively by:

$$b_1 = \frac{v_1}{\|v_1\|}$$
$$b_i = \frac{v_i - \sum_{j=1}^{i-1}\langle v_i, b_j\rangle b_j}{\left\|v_i - \sum_{j=1}^{i-1}\langle v_i, b_j\rangle b_j\right\|} \qquad 2 \le i \le n$$

Since $v_1 \neq 0$ and $v_i \notin \text{span}(b_1, b_2, ..., b_{j-1})$ for each $j \geq 2$, the denominator is all nonzero. By construction, these vectors satisfy the condition in the theorem, and they are certainly orthonormal, completing the proof.

To generalize to infinite dimensions, important results from real analysis to show all [complete] inner

product spaces have an orthonormal basis. The reader shall see such proofs in a Real Analysis or Functional Analysis course[34]. Since an orthonormal basis is closely tied to the inner product, it is particularly easy to find the coefficients for a basis representation of a vector v given an orthonormal basis β using the inner product:

Theorem 13.5.15: Vector Representation With Orthonormal Basis

Let V be a finitely dimensional inner product space, β and orthonormal basis, and $w \in V$ an arbitrary vector. Then:

$$w = \sum_{v_i \in \beta} \langle w, v_i \rangle \cdot v_i$$

The coefficients $\langle w, v_i \rangle$ are called the *Fourier coefficients* of w (with respect to the inner product)

Proof **:**
exercise

Using the fact that every inner product space has an orthonormal basis also implies that the inner product defined on a vector-space is unique:

Theorem 13.5.16: Uniqueness Of Inner Product

Let V be a vector-space with two inner products: $\langle -, - \rangle_1$ and $\langle -, - \rangle_2$. Then there exists an isometry between these inner products

Proof **:**
Using these two inner products, define an orthonormal basis β, γ. Since they are the same size, define $f(\beta_i) = \gamma_i$. Show that f is an isometry.

This same result *does not hold* for normed spaces. Since all finite-dimensional normed spaces are inner-product spaces, this result is only visible in infinite dimensional vector-space. The reader taking a classs in an analysis will learn that the space of (real or complex) continuous functions admit infinitely many norms, only one of which induces an inner products[35].

The notion of orthogonality defined in definition ref:HERE lends itself well when introducing inner products:

Definition 13.5.17: Orthogonal Compliment

Let V be an inner product space and S a subspace. Then the *orthogonal compliment* of S is the set:

$$S^{\perp} := \{v \in V \mid \langle v, s \rangle = 0, \ \forall s \in S\}$$

The set is called S *compliment* or S *perp*.

[34] Or see EYNTKA Real Analysis or Functional Analysis
[35] Namely, the L^n norms where the L^2 norm induces an inner product

It is an exercise to show that all of the properties in exercises ref:HERE generalize. What is particular to inner products is the notion of the closest vector:

Proposition 13.5.18: Orthogonal Projection

here

Proof **:**
here (FIS p.366)

A final result before continuing to the next topic is a geometric characterization of inner product spaces:

Proposition 13.5.19: Complimentary Subspace Property

Let V be a [finite dimensional][a] inner product space and $W \subseteq V$ a subspace. Then there exists a unique $W \perp$ such that

$$V = W \oplus W^{\perp}$$

[a]The general proof of this fact does not require finite dimensions, however tools from analysis are needed to complete it

Proof **:**
Let $W^{\perp} := \{v \in V \mid \langle v, w \rangle = 0 \ \forall w \in W\}$. I claim $V = W \oplus W \perp$. Take an orthonormal basis for W and extend it to a basis for V and make it orthonormal via the grand-Schmidt process; label this basis β. Let $v \in V$. Then if $[v]_\beta = (v_1, ..., v_n)$, define:

$$w = (v_1, ..., v_m, 0, ..., 0) \qquad w' = (0, ..., 0, v_{m+1}, ..., v_n)$$

Certainly, $w \in W$, and since for each element $x \in W$, $\langle x, w' \rangle = 0$, $w' \in W^{\perp}$. Then as an immediate consequence $V = W + W^{\perp}$, and the reader should verify why $V = W \oplus W^{\perp}$, which would completing the proof.

It can be shown that the above property is another way of characterizing inner product space (see [Lindenstrauss and Tzafriri in 1971]).

13.5.2 Adjoint Operators in Inner Product Spaces

Recall that each linear transformation $f : V \to W$ is associated with a dual $f^* : W^* \to V^*$. It is not always the case that $V \cong V^*$, and furthermore $(-)^*$ is not a functor. When working with inner product spaces, it is always the case that $V \cong V^*$, and furthermore there shall be a map $f^\dagger : W \to V$ obtainable by the dual. This map shall be called the *adjoint* of f, its definition being the focus of this section. The key idea is that the inner product shall "pick" for us a nice basis which shall be compatible.

To start off, let us show that $V \cong V^*$ using the inner product:

Theorem 13.5.20: Riesz Representation Theorem

Let V be a finitely dimensional k-vector space[a]. Then the map:

$$R : V \to V^* \qquad R(v) = \langle -, v \rangle$$

mapping each v to the functional $\langle -, v \rangle$ ($R(v)(w) = \langle v, w \rangle$). Then R is an isomorphism. Equivalently, for each $f \in V^*$, there exists a unique $v \in V$ such that $f = \langle -, v \rangle$.

[a]The infinite-dimensional equivalent requires results from Real Analysis

Proof **:**
Linearity of R is given by:

$$R(v + w) = \langle -, v + w \rangle = \langle -, v \rangle + \langle -, w \rangle = R(v) + R(w)$$

For bijectivity, pick an orthonormal basis β for V and let $f \in V^*$. Then since V is a finitely dimensional vector space $f(x) = \sum_i a_i x_i \beta_i$ for appropriate a_i. Let:

$$v = \sum_i \overline{f(\beta_i)} \beta_i$$

Then for each β_j:

$$\begin{aligned} R(v)(\beta_j) &= \langle \beta_j, v \rangle \\ &= \langle \beta_j, \sum_i \overline{f(\beta_i)} \beta_i \rangle \\ &= \sum_i f(\beta_i) \langle \beta_j, \beta_i \rangle \\ &= \sum_i f(\beta_i) \delta_{ij} \\ &= f(\beta_j) \end{aligned}$$

Hence, as the functions agree on where they map the basis, $R(v) = f$, showing surjectivity. For injectivity, let's say $\langle -, v \rangle = \langle -, v' \rangle$. By proposition 13.5.3(3), $v = v'$, completing the proof.

The above proof required the vector-space to be finite-dimensional, however it does generalize to infinite dimensions with the appropriate tools from Real analysis or Functional analysis. All further results in this section are dependent on the Riesz Representation theorem, and hence since the infinite dimensional case of Riesz Representation Theorem cannot be proved here, many of the following results are limited to finite dimensions even though their proofs do not rely on the finiteness of the vector-space. Henceforth, if the vector-space has to be finite dimensional to invoke Riesz's Representation theorem, it shall be put in square brackets.

Recall that $f^* : W^* \to V^*$ is defined via $f^*(\varphi) = \varphi \circ f$. Using Riesz Representation Theorem, each φ is representable as a $\langle -, v \rangle$ for a unique $v \in V$, and hence we get:

$$f^*(\langle -, v \rangle) = \langle f(-), v \rangle$$

Now, f^* is defined on the dual. The goal is to define from f^* a function $f^\dagger : W \to V$. To start off, we focus on the case of $f : V \to V$. The key insight is that using the uniqueness of Riesz Representation Theorem, we may define a function $f^\dagger : V \leftarrow V$ such that

$$f^*(\langle -, v\rangle) = \langle f(-), v\rangle \overset{!}{=} \langle -, f^\dagger(v)\rangle = \mathrm{id}(\langle -, f^\dagger(v)\rangle)$$

which moves the responsibility of the function f^* on the dual spaces to the spaces!

Theorem 13.5.21: Existence Of Linear Adjoint

Let V be a [finite dimensional] inner product space an $f : V \to V$ a linear operator. Then there exist a unique linear operator $f^\dagger$ such that

$$\langle f(v), w\rangle = \langle v, f^\dagger(w)\rangle \qquad \forall v, w \in V$$

Proof **:**
Pick any $w \in V$. Define $g_y(x) = \langle f(x), y\rangle$. g_y is certainly linear, hence by Riesz Representation Theorem we have the there exists a unique y' such that:

$$\langle f(v), w\rangle = g_w(v) = \langle v, w'\rangle$$

Define $f^\dagger(w) := w'$. The function $f^\dagger$ is linear since for any v:

$$\begin{aligned}\langle v, f^\dagger(cx_1 + x_2)\rangle &= \langle f(v), cx_1 + x_2\rangle \\ &= \overline{c}\langle f(v), x_1\rangle + \langle f(v), x_2\rangle \\ &= \overline{c}\langle v, f^\dagger(x_1)\rangle + \langle v, f(x_2)\rangle \\ &= \langle v, cf^\dagger(x_1) + if(x_2)\rangle\end{aligned}$$

thus, by proposition 13.5.3(3), $f^\dagger(cx_1 + x_2) = cf^\dagger(x_1) + if(x_2)$ showing linearity. For uniqueness, if $g : V \to V$ is another linear operator satisfying

$$\langle f(v), w\rangle = \langle v, g(w)\rangle \qquad \forall v, w \in V$$

then $\langle v, f^\dagger(w)\rangle = \langle v, g(w)\rangle$ for all v, and hence $f^\dagger(w) = g(w)$, showing $f^\dagger = g$, completing the proof.

Definition 13.5.22: Adjoint Operator

Let $f : V \to V$ be a linear operator on a [finite dimensional] inner product space. Then the *adjoint operator* is the unique linear operator $f^\dagger : V \to V$ satisfying:

$$\langle f(-), -\rangle = \langle -, f^\dagger(-)\rangle$$

Note that since $\langle v, w\rangle = \overline{\langle w, v\rangle}$ we have:

$$\langle v, T(w)\rangle = \overline{\langle T(w), v\rangle} = \overline{\langle w, T^*(v)\rangle} = \langle T^*(v), w\rangle$$

Proposition 13.5.23: Matrix Representation Of Adjoint

Let V be a [finite dimensional] inner product space with orthonormal basis β. Then if $f : V \to V$ is a linear operator,

$$[f^\dagger]_\beta = [f]_\beta^*$$

that is, $[f^\dagger]_\beta$ is the conjugate transpose of $[f]_\beta$

***Proof* :**
Let $A = [f]_\beta$, $B = [f^\dagger]_\beta$, and $\beta = \{\beta_1, \beta_2, ..., \beta_n\}$. Then by theorem 13.5.15 it can be deduced that:

$$B_{ij} = \langle f^\dagger(\beta_j), \beta_i \rangle = \overline{\langle \beta_i, f^\dagger(\beta_j) \rangle} = \overline{\langle f(\beta_i), \beta_j \rangle} = \overline{A_{ji}}$$

which gives the desired result.

Corollary 13.5.24: General Adjoint

Let $f : V \to W$ be a linear transformation between finite dimensional vector spaces. Then there exists a unique linear transformation $f^\dagger : W \to V$ such that the following diagram commutes:

$$\begin{tikzcd} W^* \arrow[r, "f^*"] \arrow[d, "R_W"'] & V^* \arrow[d, "R_V"] \\ W \arrow[r, "f^\dagger"'] & V \end{tikzcd}$$

where R_W and R_V are the Riesz Isomorphisms for V and W respectively. Furthermore, $f^\dagger$ satisfies the properties of the adjoint

***Proof* :**
Define

$$f^\dagger = R_V^{-1} \circ f^\dagger \circ R_W : W \to V$$

which is unique by theorem 13.5.20 and commutes by definition. TO show this matches with the adjoint, let

$$\widehat{f} = \begin{pmatrix} 0 & 0 \\ f & 0 \end{pmatrix} \in \mathrm{End}_k(V \oplus W)$$

By theorem 13.5.21, there is a unique function $\hat{f}^\dagger$ such that

$$\langle \widehat{f}(x_1, x_2), (y_1, y_2) \rangle = \langle (x_1, x_2), \widehat{f}^\dagger(y_1, y_2) \rangle$$

By definition of $\widehat{f}$, we see

$$\widehat{f}^\dagger = \begin{pmatrix} 0 & f^\dagger \\ 0 & 0 \end{pmatrix}$$

Hence we have that $f^\dagger$ is a function with domain and codomain $f^\dagger : W \to V$. It follows that

$$\langle f(x), y \rangle = \langle x, f^\dagger(y) \rangle$$

and by theorem 13.5.21, f and $f^\dagger$ respect all the desired properties, completing the proof.

Proposition 13.5.25: Properties Of Adjoint

Let V be a [finite dimensional] inner product space, let f, g be linear operators on V and $c \in k$. Then:

1. $(f+g)^\dagger = f^\dagger + g^\dagger$
2. $(cf)^\dagger = \overline{c} f^\dagger$
3. $(g \circ f)^\dagger = f^\dagger \circ g^\dagger$
4. $(f^\dagger)^\dagger = f$
5. $I^\dagger = I$

In other words, $(-)^\dagger$ is an [additive][a] idempotent contravariant functor

[a] an additive functor is on that respects the abelian group structure of $\mathrm{Hom}_k(V, V)$, see definition 26.0.1

Proof **:**
exercise

With the definition of the adjoint operator, proposition 13.3.10 can be appropriately modified:

Proposition 13.5.26: Range and Kernel Of Adjoint

Let V be a inner product space and $f : V \to V$ a linear operator. Then:

1. $\ker(f^\dagger) = \mathrm{im}(f)^\perp$
2. $\mathrm{im}(f^\dagger)^\perp = \ker(f)$. If V is finite dimensional, $\mathrm{im}(f^\dagger) = \ker(f)^\perp$

Hence, $f^\dagger$ is injective if and only if f is surjective.

Proof **:**
exercise. Reference proposition 13.3.10 if needed.

The above can be thought of as the geometric characterization of the adjoint. Recall that each subspace $W \subseteq V$ of an inner product space has a complimentary subspace $W^\perp$ (proposition 13.5.19). Then the adjoint is the natural way to switch back and forth between the two "viewpoints". In other words, proposition 13.5.26 demonstrate that the adjoint can be thought of as transforming a space into its "dual" without the need of passing through the dual-space.

Proposition 13.5.27: Eigenvectors Of Adjoint

Let f be a linear operator over a *finite dimensional* vector space. Then if f has eigenvectors, so does $f^\dagger$. If λ is a eigenvalue of f, then the corresponding eigenvalue of $f^\dagger$ is $\overline{\lambda}$

Proof **:**
Since f has an eigenvector v, $f(v) = \lambda v$, hence $f(v) - \lambda v = 0$. Now:

$$0 = \langle 0, w\rangle = \langle (f - \lambda I)(v), w\rangle = \langle v, (f - \lambda I)^\dagger(w)\rangle = \langle v, (f^\dagger - \overline{\lambda} I)(w)\rangle$$

Thus, the [nonzero] eigenvector v is orthogonal to the range of $f^\dagger - \overline{\lambda} I$, and so $f^\dagger - \overline{\lambda} I$ cannot be surjective. Since the vector-space is finite-dimensional, it cannot be injective, and hence has a kernel and eigenvector, completing the proof.

The infinite dimensional case requires a more careful examination of eigenvectors of infinite dimensional linear operators, and hence is left for a class in functional analysis[36]

Proposition 13.5.28: Shur's Lemma For Matrices

Let f be a linear operator on a *finite dimensional* inner product space where the characteristic polynomial splits (equiv. V is a $\mathbb{C}$-vector space). Then there exists an orthonormal basis β for V such that $[f]_\beta$ is upper-triangular

Note that it was not specified that the geometric and algebraic multiplicities align, hence it is not guaranteed to be upper-triangular.

Proof **:**
FIS. p. 370

13.5.3 Self-Adjoint and Normal Operators

By proposition 13.5.25, $(-)^\dagger$ maps the identity to the identity and is an idempotent operator. This may seem somewhat familiar to the reader who has experience with the complex numbers, namely for any $z \in \mathbb{C}$, $\overline{1} = 1$ and $\overline{\overline{z}} = z$. As the complex numbers are a very important field, drawing comparisons between the ring $\mathrm{End}_k(V)$ (equivalently, $M_n(k)$ for $k \in \{\mathbb{R}, \mathbb{C}\}$) and $\mathbb{C}$ may give insight into the structure of this rather complicated [non-commutative] ring! As matrices play central roles in various parts of mathematics (particularly for this book, Lie Theory and Representation Theory), gaining insight in the structure of $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$ is of great value. Since $\mathbb{R}$ and $\mathbb{C}$ naturally embeds into $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$ (respectfully) via $z \mapsto zI$, these rings can be thought of as extensions of the complex numbers[37]. For the rest of this section, we shall stick to $k = \mathbb{C}$.

When working with complex numbers, the complex-conjugate has the following properties of interest:

1. $z\overline{z} = \overline{z}z$
2. $z = \overline{z}$ if and only if $z \in \mathbb{R}$.
3. $z\overline{z} \in [0, \infty)$
4. $z = a + bi$ for real $a, b \in \mathbb{R}$.

[36] In particular, this generalizes to compact normal operators
[37] This extension gives rise to the Morita equivalence for finite dimensions and the strong Morita Equivalence for C^*-algebras, telling us exactly which ring properties are preserved after the extension

How many of these have analogies for the adjoint operator? The reader should take a moment to see if the first condition is satisfied before reading further

40: Moment

Taking a moment

The following example will be illustrative:

Example 13.5.3: Commuting Adjoint Operator
Take the following matrices (show they do not commute)

Hence, it is not generally the case that $ff^\dagger = f^\dagger f$. If f is a diagonalizable operator, then this is certainly the case. Are there other operators that satisfy this property? The reader should again ponder for a to come up with an example or attempt a proof for the lackthereof

41: Moment

Taking a moment

The following example will be illustrative:

Example 13.5.4: Commuting Adjoint Operator II
give an example

Hence, let us start by labeling these types of operators

Definition 13.5.29: Normal Operators

Let V be a [finite dimensional] complex inner product space. Then if f satisfies

$$ff^\dagger = f^\dagger f$$

the operator f is said to be a *normal operator*.

Is the set of all normal operators a subspace of $M_n(\mathbb{C})$ (as a $\mathbb{C}$-vector space)? This question shall be answered later in the section. For the moment, we may consider normal operator as our closest analogy of the matrices that extend the complex numbers. Certainly $\mathbb{R} \cdot I$ (resp. $\mathbb{C} \cdot I$) are normal, and by 13.5.4 this set is larger than the identity matrices. If $f = \overline{f}$, would it be possible to think of f as being the equivalent of the real numbers in $\mathbb{C}$? This will turn out to indeed be the case, namely it shall be all diagonalizable matrices with real eigen-vectors, and furthermore any normal operator f shall be decomposable into $f = A + Bi$. Starting by giving operators of the form $f = \overline{f}$ a name

Definition 13.5.30: Self-adjoint Operators

Let V be a [finite dimensional] complex inner product space. Then if f satisfies

$$f = f^\dagger$$

the operator f is said to be a *self-adjoint operator*.

Note that not all Normal operators are normal, as operators satisfying $f^\dagger = -f$ are normal but not self-adjoint. The matrix representation of a self-adjoint operator is often called a *Hermitian matrix* (and if $f^\dagger = -f$, the matrix is called an *anti-Hermitian matrix*).

Proposition 13.5.31: Properties Of Normal Operators

Let f be a normal operator on a finite dimensional vector space. Then:

1. $\|f(v)\| = \|f^*(v)\|$ for all $v \in V$
2. for all $c \in k$, $f - cI$ is normal
3. if v is an eigenvector of f, v is an eigenvector of f^* with eigenvalue $\overline{\lambda}$
4. If λ_1, λ_2 are distinct eigenvalues of f with corresponding eigenvectors v_1, v_2, then v_1, v_2 are orthogonal.

Proof **:**
FIS p.371

(words here? Note that the field must be complex because we need the polynomial to split)

Theorem 13.5.32: Eigen-basis For Normal Operators

Let f be a complex linear operator on a *finite dimensional* inner product space. Then f is normal if and only if there exists an orthonormal basis for f consisting of eigenvectors of f. here[a]

[a] For infinite dimensions, the operator must be *compact*

Proof **:**
here

Example 13.5.5: Infinite Dimensional Normal Operator

The counter example being the translation on $C(S^1)$. If S is an orthonormal basis, indexed by $\{f_i\}_{i\in\mathbb{Z}}$, then define:

$$R(f) = f_1 f \qquad L(f) = f_{-1} f$$

Then:

$$R(f_n) = f_{n+1} \qquad L(f_n) = f_{n-1}$$

Prove that R and L are normal and have no eigenvectors.

L

(words for the following)

Proposition 13.5.33: Properties Of Self-Adjoint Operators

Let f be a normal operator on a finite dimensional vector space. Then:

1. Every eigenvalue is *real*
2. if V is a $\mathbb{R}$-vector space, then the characteristic polynomial of f splits.
3. There exists an orthonormal basis consisting of eigenvectors of f

Proof **:**
FIS p.373

We now focus on the equivalent of the "unit circle" for linear operators:

Definition 13.5.34: Unitary/Orthogonal Operators

Let V be a [finite dimensional] inner product space over $k \in \{\mathbb{R}, \mathbb{C}\}$. Then if $\|f(x)\| = \|x\|$ and f is surjective[a] for all $x \in V$, f is called a:

1. *unitary operator* if $k = \mathbb{C}$
2. *orthogonal operator* if $k = \mathbb{R}$

[a] This is guaranteed in finite dimensions

Proposition 13.5.35: Complex Matrix Representation

Let $M \in M_n(\mathbb{C})$. Then there exists Hermitian matrix A and anti-Hermitian matrix B such that:

$$M = \frac{1}{2}(A + A^*) + \frac{1}{2}(B - B^*)$$

Proof **:**
later

Exercise 13.5.1

1. (Grand schmit in countable Hilbert Space?)
2. (Show that the orthogonal compliment property characterizes Hilbert spaces? Limit it to fin dim?)
3. (Mention that all Hilbert spaces embed into $\ell^2(U)$. Make it an exercise for fin dim)

14

Modules Over PID's: Linear Algebra II

Modules over fields have some of the simplest possible structure, namely since they are always free so that there always exist a short exact sequence:

$$0 \to R^n \to M \to 0$$

Using definition 12.3.17, every module over a field admits a free resolution of rank 1. It is natural to want to study modules that admit a free resolution of rank 2, namely if R is a ring and M a finitely generated module, then if

$$R^n \to M \to 0$$

then there exists an R^m such that

$$0 \to R^m \to R^n \to 0$$

As was foreshadowed in section 12.3.3, all [finitely generated] PID's admit free resolutions of rank 2, and hence shall be the focus of this chapter. Having a free resolution of rank 1 (i.e. being free) implied the existence of basis, even over PID's matrix representation of linear transformations admitted smith normal forms. A free resolution of rank 2 shall offer weaker results, however they shall still be incredibly powerful introducing us to *Jordan Cannonical forms* and *Rational Cannonical Forms*. These forms shall be a new important way of representing linear transformation using linear maps that will give more information, namely ref:HERE.

The reader may naturally ask if it is interesting to further push the degree of the free resolution and explore modules that admit a free resolution of degree 3, or 4, or n. While Noetherian modules always admit a finite length resolution (exercise), most interesting examples of such modules become a good deal more complex and require a build-up in algebraic geometry and Homological algebra. For the reader who is interested, these are linked to the notion of *krull dimension* of a commutative ring[1].

[1]The link is not perfect, often some more conditions must be added to the ring such as R needing to be a local rings

The essential concepts understanding this link shall be covered in this book (namely, chapter 21 section 21.5 and chapter 28 section 28.3), and further reading can be done in cite:HERE.

Goal

The goal of this section is to show finitely generated modules over PID's have the following structure:

$$M \cong R^k \oplus R/(a_1) \oplus \cdots \oplus R/(a_m)$$

We'll define the terms a_m and how to find k soon. Essentially, M shall be split into a part which is "as linearly dependent as possible", and the rest of M will be the elements which cannot be linearly independent (over R). These "rest of elements" of M can be identified with the simplest quotient rings of R, in particular a cyclic module.

Since PID's are UFD's, we will be able to find a 2nd representation given appropriate prime elements:

$$M \cong R^k \oplus R/(p_1) \oplus \cdots \oplus R/(p_n)$$

where p_i will turn out to represent prime elements of R (and the k will be the same).

A consequence of this would be the proof of the Fundamental Theorem of finitely Generated abelian groups (theorem 5.1.5). Since $\mathbb{Z}$ is a finitely generated Principal Ideal Domain, if M is a $\mathbb{Z}$-module (and so an abelian group), it will be a consequence of the following theoretical build-up that

$$M \cong \mathbb{Z}^k \oplus \mathbb{Z}/n_1\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/n_k\mathbb{Z}$$

Another consequence will be the ability to classify all linear maps from a vector-space to itself (up to similarity). This means that the structure of $M_n(k)$ and $\mathrm{GL}_n(k)$ will be eminently understandable, having huge consequence in representation theory (part VI) and many areas beyond algebra (Differential Geometry, K-theory, Optimization, ODEs, and so forth). In particular, if V is a vector space over $\mathbb{F}$ and T is a linear transformation, then like in example ref:HERE, we can form an $F[x]$-module over V where representing T with the action:

$$p(x) \cdot v = (a_n x^n + a_{n-1}x^{n-1} + \cdots a_1 x + a_0) \cdot v = a_n T(v)^n + \cdots + a_1 T(v) = a_0$$

then we can use the fact that $F[x]$ is a Principal Ideal Domain to decompose T into a nicer representation. Since every linear transformation can be represented as a matrix, then we can also decompose matrices in this fashion (which will be called Rational Canonical form and Jordan Canonical Form).

Note that finiteness is an important condition: we shall go over an example where if you have an infinitely generated Principal Ideal Domain over a module, then the decomposition is not unique, among other issues [2]

14.1 Over PID, Submodule of Free Module is Free

(I like this build-up: `https://kconrad.math.uconn.edu/blurbs/linmultialg/modulesoverPID.pdf`)

[2] wikipedia: Structure theorem for finitely generated modules over a principal ideal domain, Non-finitely generated modules

(also downloaded a pdf called moduleoverPID)

Remember that a sub-module of a free module need not be free. For example, if we take $\mathbb{C}[x, y]$ over itself, then it is free with rank 1, but the sub-module (ideal) (x, y) is not free. If it were of rank 1[3], then (x, y) would be generated by a single element, but it has been shown earlier this is not possible.

However, if R is a PID, then just like for modules over fields, submodules of free modules over a PID are free. For free modules, this was because every module over a field is free, however there do exist modules over PID's that are not free (consider any torsion abelian group). To prove this result, we shall show that if F is a free module over a PID and $M \subseteq F$, we can iteratively split off a cyclic direct summand $\langle y \rangle$ from $M \subseteq F$, which shall result in separating the free part from the torsion part. Furthermore, y can be chosen in such a way that it is a scalar multiple of a basis element of F: $y = ax$ for appropriate $a \in R$; think of submodules of $\mathbb{Z}^n$ as a canonical example of this behavior:

Lemma 14.1.1: Splitting Modules Over Pid

Let R be a PID, F a finitely generated free module over R, and $M \subseteq F$ be a nonzero submodule. Then there exists elements $a \in R$, $x \in F$, $y \in M$, and submodules $F' \subseteq F$ and $M' \subseteq M$ such that

1. $y = ax$
2. $M' = F' \cap M$
3. $F = \langle x \rangle \oplus F'$, $M = \langle y \rangle \oplus M'$

Proof **:**
Since $F \cong R^n$, it is perhaps tempting to think that we can our factor may be one of the components of R^n, however this implicitly chooses a particular representation (i.e. a basis for the isomorphism $\cong$). Since we don't know that M is free yet, we cannot know how to pick a basis representation of F that matches well with a basis representation of M, we cannot pick a basis β for F in such a way that the free submodule M is a projection. In that case, the splitting in the lemma would be obtained by choosing a component, or in terms of functions using the projection map:

$$\pi(r_1, r_2, ..., r_n) = r_1$$

Instead, we shall employ the following trick that has been used in theorem ref:HERE to make it choice-independent: we shall take the see of all possible homomorphism $\varphi : F \to R$ and use Zorn's lemma to choose a maximal element.

Thus, define the following collection of ideals

$$\mathscr{F} = \{\varphi(M) \mid \varphi \in \mathrm{Hom}_R(F, R)\} = \{\varphi(M) \mid \varphi \in F^*\}$$

This set contain non-trivial functions since the projections $\pi_k \in \mathscr{F}$ and M, F are nonzero. Thus since R is a PID, namely it is Noetherian, by theorem 12.3.21 there exists a $\alpha : F \to R$ such that $\alpha(M)$ is maximal. Since $\mathscr{F}$ is nontrivial, it must be that $\alpha(M) \neq 0$. Since R is a PID, there exists an $a \in R$ such that $\alpha(M) = a$. Thus, there exists a $y \in M$ such that $\alpha(y) = a$.

We shall see that this y and a are the desired values. To find the desired x, we shall show that a

[3] Why can't the rank be ≥ 1? hint: use field of fractions

divides $\varphi(y)$ for all $\varphi \in F^*$. To see this, let b_φ be the element such that

$$(b_\varphi) = (a, \varphi(y))$$

Then there exists $r, s \in R$ such that

$$b = ra + s\varphi(y)$$

Using this, define the homomorphism $\psi := r\alpha + s\varphi$. Since $a \in (b)$, $\alpha(M) \subseteq (b)$, and by maximality $\psi(M) \subseteq \alpha(M)$, hence $\alpha(M) = \psi(M)$. But then

$$(a) = (b) = (a, \varphi(y))$$

namely, $\varphi(y) \in (a)$, hence a divides $\varphi(y)$. Now, since F is a finitely generated free module, $F \cong R^n$. Given this representation of F, we may write:

$$y = (s_1, s_2, ..., s_n)$$

and define the projection $\pi_i : F \to R$, $\pi_i(y) = s_i$. Since $\pi_i \in \mathscr{F}$, a divides s_i for each s_i, namely there exists $r_1, ..., r_n$ such that $s_i = ar_i$. Letting x be the element associated to the element

$$(r_1, ..., r_n)$$

in the isomorphism, we get that

$$y = ax$$

Given us our desired x.

Let's now show that x, y have the desired properties. First, notice that

$$a = \alpha(y) = \alpha(ax) = a\alpha(x)$$

Since R is an integral domain, $\alpha(x) = 1$. Next, let $F' = \kappa(\alpha)$ and $M' = F' \cap M$. Notice that for every $z \in f$,

$$z = \alpha(z)x + (z - \alpha(z)x)$$

which by linearity:

$$\alpha(z - \alpha(z)x) = \alpha(z) - \alpha(z)\alpha(x) = \alpha(z) - \alpha(z) = 0$$

Thus,

$$z - \alpha(z)x \in \ker \alpha$$

implying that $F = \langle x \rangle + F'$. Next, take $rx \in \langle x \rangle \cap F'$. Then $\alpha(rx) = 0$, and by linearity $r\alpha(x) = 0$, but $\alpha(x) = 1$ so $r = 0$, and so $\langle x \rangle \cap F' = 0$, which by definition implies:

$$F = \langle x \rangle \oplus F'$$

Giving part of the 3rd claim. Finally, if $z \in M$, then $a \big| \alpha(z)$, and indeed

$$\alpha(z) \in \alpha(M) = (a)$$

Writing $\alpha(z) = ca$, we get $\alpha(z)x = cax = cy$. Splitting z as we've done before, notice that:

$$z - \alpha(z)x = z - cy \in M \cap F' = M'$$

and so, repeating the same logic as before, we get:

$$M = \langle y \rangle \oplus M'$$

completing the last part of the proof.

With this, we may prove one of the main results:

Theorem 14.1.2: Submodule Of Free Module Over PID

Let R be a PID, let F be a free R-module, and let $M \subseteq F$. Then M is free

Proof **:**
If $M = 0$, we are done, so let $M \neq 0$. by lemma 14.1.1 there exists a $y_1 \in M$ and a $M^1 \subseteq M$ such that

$$M = \langle y_1 \rangle \oplus M^1$$

If $M^1 = 0$, we are done. Otherwise, since $M^1 \subseteq F$, we may keep applying lemma 14.1.1 producing a y_2 and M^2. Since the set $\{y_1, y_2, ..., y_m\}$ is linearly independent and F has finite rank, this process must terminate, leaving us with a $M^m = 0$ for some $m \leq n$, at which point:

$$M = \langle y_1 \rangle \oplus \langle y_2 \rangle \oplus \cdots \oplus \langle y_m \rangle$$

as we sought to show.

This result did note require the divisibility result we have just done. Using it, we can get a more precise result giving us something similar to a smith normal form for PIDs:

Corollary 14.1.3: Basis For Submodules Of Free Modules

Let R be a PID, F a finitely-generated free R-module, and $M \subseteq F$. Then there exists a basis $(x_1, 2, ..., n)$ for F and nonzero elements $a_1, a_2, ..., a_m \in R$ with $m \leq n$ such that

$$(a_1 x_1, ..., a_m x_m)$$

is a basis for M. Furthermore, $a_1 | a_2 | \cdots | a_m$

Proof **:**
By theorem 14.1.2 M is free and by inductively applying lemma 14.1.1 the basis for M can be chosen so that it is of the form in the lemma.

To show the coefficients divide each other, it suffices to show that $a_1 \big| a_2$ since then the argument can be repeated inductively. Recall that (a_1) is maximal among the ideals $\varphi(M)$, $\varphi \in F^\vee$.

Corollary 14.1.4: If PID, Then Free Resolution Of Length 2

Let R be a PID. Then every finitely generated R-module M and every surjective homomorphism:

$$R^{m_0} \xrightarrow{\pi_0} M \to 0$$

there exists a free R-module R^{m_1} and a homomorphism $\pi_1 : R^{m_1} \to R^{m_0}$ such that the sequence

$$0 \to R^{m_1} \xrightarrow{\pi_1} R^{m_0} \xrightarrow{\pi_0} M \to 0$$

is exact. Along with proposition 12.3.18, we get that R is a PID if and only if every R-module admits a free resolution of length 2.

***Proof* :**
Since R^{m_0} is an R-module over a PID, $\ker(\pi_0)$ is is free by theorem 14.1.2, and since R^{m_0} is finitely generated, $\ker(\pi_0) \cong R^{m_1}$ for some $m_1 \in \mathbb{N}$. Hence choose the natural embedding $\pi_1 : \ker(\pi_0) \hookrightarrow R^{m_0}$ giving the desired exact sequence, completing the proof.

Since all modules over PID's admit a free resolution of length 2, they are characterized by a matrix (up to equivalence), namely if

$$0 \to R^m \xrightarrow{\varphi} R^n \xrightarrow{\pi_0} M \to 0$$

Then $M = \operatorname{coker}(\varphi)$, and by exactness we know that φ injects into R^n (or, if the reader rather not use exactness, φ injects since $R^m = \ker(\pi_0)$). Then by corollary 14.1.3, given a basis $(x_1, x_2, ..., x_n)$ of R^n, there exists nonzero elements $a_1, ..., a_m \in R$ where $a_1 \big| a_2 \big| \cdots \big| a_m$ and $(a_1x_1, a_2x_2, ..., a_mx_m)$ forms a basis for $R^m \cong \ker(\pi_0)$. Hence, M is presented by the matrix:

$$\begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_m \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

By remark (readingPresentationsAndUnits),any row/column which contains an a_i that is a unit may be removed. Without loss of generality, let's say that $a_1, a_2, ..., a_m$ are non-units. Then:

$$M \cong (R/a_1) \oplus \cdots (R/a_m) \oplus R^{(n-m)}$$

Notice what just happened: a finitely generated module over M is isomorphic to a rather simple structure, that is such modules lend themselves to a classification theorem! The following theorem summarizes this result and proves a couple more important facts:

Theorem 14.1.5: Invariant Factor Form of Modules over PID's

Let R be a Principal Ideal Domain and let M be a finitely generated R-module. Then:

1. M is isomorphic to a direct sum of finitely many cyclic modules. More precisely,

$$M \cong_R R^r \oplus R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_m)$$

for some integer $r \geq 0$ and nonzero element $a_1, a_2, ..., a_m$ of R which are not units in R and which satisfy the divisibility relations

$$a_1 \mid a_2 \mid \cdots \mid a_m$$

2. M is torsion-free if and only if M is free

3. In the decomposition of (1)

$$\mathrm{Tor}(M) \cong_R R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_m)$$

in particular M is a torsion module if and only if $r = 0$ and in this case the annihilator of M is the ideal (a_m).

In ref:HERE, we shall prove the uniqueness of this decomposition up to isomorphism, allowing us to say it is *the* direct sum of finitely many cyclic modules.

***Proof* :**
Let's first state more succinctly the above analysis to prove the existence of part (1). Let M be a finitely generated module over a PID R, say $x_1, x_2, ..., x_n$ is a generating set of minimal cardinality. Then given R^n with basis $b_1, b_2, ..., b_n$:

$$\pi : R^n \to M \qquad \pi(b_i) = x_i$$

is a surjective R-module homomorphism with kernel $\ker(\pi)$ which is free by theorem 14.1.2. Then by corollary 14.1.3 choose a basis $y_1, y_2, ..., y_n$ for r^n such that $a_1y_1, a_2y_2, ..., a_my_m$ is a basis for $\ker(\pi)$ with $a_1 \big| a_2 \big| \cdots \big| a_m$. Then:

$$M \cong R^n / \ker(\pi_=(y_1R \oplus y_2R \oplus \cdots \oplus y_nR)/(a_1y_1R \oplus a_2y_R2 \oplus \cdots \oplus a_my_mR))$$

which the reader may verify (exercise ref:HERE) that is isomorphic to:

$$R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_m) \oplus R^{n-m}$$

This immediately gives us the other two conclusions:

(2) Since $R/(a_i)$ is torsion and the direct product of torsion modules is torsion, then M is torsion free if and only if $M \cong R^n$, which implies it's free (hence being "element-wise" free implies being "globally" free)

(3) This follows from the fact that the direct product of torsion modules is torsion, hence

$$\mathrm{Tor}_R(M) \cong_R R/(a_1) \oplus \cdots \oplus R/(a_m)$$

In order to prove the uniqueness of this decomposition, we shall show how we can systematically "embed" parts of M into a vector-space and take advantage of the invariant basis number (IBN) of finitely dimensional isomorphic vector spaces, that is if two vector-spaces are isomorphic they are the same dimension.

First, since R is a PID the annihilator of M is a principal ideal. We start by finding the annihilator of M when it is a torsion module:

Lemma 14.1.6: Annihilator of Module Over PID

Let M be a finitely generated torsion module over a PID R with decomposition:

$$R/(a_1) \oplus \cdots R/(a_m)$$

with $a_1 \big| \cdots \big| a_m$. Then $\mathrm{Ann}_R(M) = (a_m)$.

Proof **:**
$0 \neq r \in \mathrm{Ann}_R(M)$ so that

$$0 = r \cdot (1, ..., 1) = (r, ..., r)$$

which by the given decomposition implies:

$$r \equiv 0 \bmod (a_m) \qquad \Leftrightarrow \qquad r \in (a_m)$$

Hence, $\mathrm{Ann}_R(M) \subseteq (a_m)$. Conversely, if $r \in (a_m)$ and $y \in M$, decomposing y so that

$$y = (y_1, ..., y_m) \qquad y_i \in R/(a_i)$$

since $r \in (a_m) \subseteq (a_i)$, we see that for each a_i:

$$ry_i = 0$$

hence $ry = 0$, and so $r \in \mathrm{Ann}_R(M)$, as we sought to show.

Next, since R is a UFD, we may find a decomposition, we have $a_m = up_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$. Since $\gcd(p_i^{k_i}, p_j^{k_j}) = 1$, by exercise ref:HERE (rings, under CRT)

$$(p_i^{k_i}) + (p_j^{k_j}) = (1)$$

Hence, by the Chinese Remainder Theorem: $R/(a_m) \cong R/(p_1^{k_1)}) \oplus \cdots \oplus R/(p_r^{k_r})$ Doing this to every a_i gives the following decomposition:

Definition 14.1.7: Elementary Divisor Form Of Modules Over PID's

Let M be a finitely generated torsion module over a PID R. Then given the above decomposition, the isomorphism:

$$M \cong R^r \oplus R/(p_1^{k_1}) \oplus \cdots \oplus R/(P_s^{k_s})$$

is called a[a] elementary divisor form of M with each (not necessarily unique) p_i being primes are R and are called the *elementary divisors*.

[a] we can't yet say "the"

Given M is torsion, since M can be decomposed into modules of the form $R/(p_k^{k_i})$, we may group together these modules to form a decomposition:

$$M \cong N_1 \oplus \cdots N_s$$

where the elements of N_i are annihilated by a power of a distinct prime p_i.

Theorem 14.1.8: Primary Decomposition Theorem

Let R be a PID and let M be a nonzero torsion R-module (not necessarily finitely generated) with nonzero annihilator a. Suppose the factorization of a into distinct primer powers in R is

$$a = up_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n}$$

and let $N_i = \{x \in M | p_i^{\alpha_i} x = 0\}$, $1 \leq i \leq n$. Then N_i is a submodule of M with annihilator $p_i^{\alpha_i}$ and is the submodule of M of all elements annihilated by some power of p_i. Thus, we get

$$M = N_1 \oplus N_2 \oplus \cdots \oplus N_n$$

If M is finitely generated, then each N_i is the direct sum of finitely many cyclic modules whose annihilators are divisors of $p_i^{\alpha_i}$

***Proof* :**
The finite case has already been proved. For the general case, modify the CRT to modules and take advantage of the fact that ideals of the form $(p_i^{k_i})$ are coprime (since R is a PID) to get the desired result

We give a name to the N_i modules

Definition 14.1.9: p_i-Primary Components

The submodules N_i in the previous theorem are called *p_i-primary components* of M.

With this, we come to the key result:

Lemma 14.1.10: Embedding Module over PID into Vector Space

Let R be a PID and let p be a prime in R. Let F denote the field $R/(p)$. Then:

1. if $M = R^t$, then $M/pM \cong F^t$
2. if $M = R/(a)$ where a is a nonzero element of R, then
$$M \cong \begin{cases} F & \text{if } p \text{ divides } a \text{ in } R \\ 0 & \text{if } p \text{ does not divides } a \text{ in } R \end{cases}$$
3. if $M = R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_k)$ where each a_i is divisible by p, then $M/pM \cong F^k$

Proof **:**

1. Let $M = R^t$. Define the map $\varphi : R^t \to (R/(p))^t$ by:
$$(r_1, ..., r_t) \mapsto (r_1 \bmod (p), ..., r_t \bmod (p))$$
Then certainly $\ker(\varphi) = pR^t = pM$, φ is surjective, and thus $M/pM \cong F^t$
2. Let $M = R/(a)$. Let $\pi : R \to R/(a)$ be the quotient map. Then
$$\pi((p)) = pM = p(R/(a)) = ((p) + (a))/(a)$$
Since R is a PID,
$$(p) + (a) = \gcd(a, p)$$
If p divides a, then $\gcd(a, p) = p$, and is 1 otherwise. Applying the appropriate isomorphism theorems gives the desired result.
3. Let $M = R/(a_1) \oplus \cdots \oplus R/(a_k)$. This follows the same procedure as theorem 14.1.5 with the modification that each a_i is divisible by p, giving us $M/pM \cong F^k$

Theorem 14.1.11: Fundamental Theorem of Uniqueness of Factors

Let R be a Principal Ideal Domain. Then two finitely generated R-modules M_1 and M_2 are isomorphic if and only if they have the same rank and the same list of invariant factors (resp. list of elementary divisors).

Proof **:**
If M_1 and M_2 have the same rank and list of invariant factors (resp. Elementary divisors), then they are clearly isomorphic, hence consider the converse: assume $M_1 \cong_R M_2$. Then any isomorphism must map torsion to torsion, and hence $\operatorname{Tor}_R(M_1) \cong_R \operatorname{Tor}_R(M_2)$. Then:
$$R^{t_1} \cong_R M_1/\operatorname{Tor}_R(M_1) \cong_R M_2/\operatorname{Tor}_R(M_2) \cong_R R^{t_2}$$
that is $R^{t_1} \cong_R R^{t_2}$. Choosing any prime $p \in R$, we may apply lemma 14.1.10 $F^{t_1} \cong_F F^{t_2}$ (can

you see why this is an F-module isomorphism where $F = R/(p)$?). Then as we saw before, $t_1 = t_2$, giving the invariant rank.

Hence, we may quotient by the free-part of each module and show that $\mathrm{Tor}_R(M_1) \cong_R \mathrm{Tor}_R(M_2)$ implies the invariant factors (resp. elementary divisors) are the same. Starting by showing the elementary divisors are the same, it suffices to show that for any fixed prime p, the elementary divisors which are a power of p are the same for M_1 and M_2. Hence, we may reduce to the case where $M_1 \cong_R M_2$ have annihilators which are a power of p.

We do this by induction on the power of p for the annihilator of $M - 1$. If the power is 0, then $M_1 = M_2 = 0$ and hence they are clearly isomorphic. Thus ,assume M_1 (thus M_2) have nontrivial elementary divisors. Say that the elementary divisors of M_1 and M_2 are:

$$M_1 : \underbrace{p, p, ..., p}_{a \text{ times}}, p^{k_1}, p^{k_2}, ..., p^{k_s}$$
$$M_2 : \underbrace{p, p, ..., p}_{b \text{ times}}, p^{r_1}, p^{r_2}, ..., p^{r_t}$$

Then the submodules pM_1, pM_2 would eliminate the p elementary divisors and reduce each k_i, r_i by 1 (i.e. they would have power $k_i - 1$ and $r_i - 1$). Since $M_1 \cong_R M_2$, $pM_1 \cong_R pM_2$. By the induction hypothesis, the elementary divisors of pM_1 are the same as those of pM_2, that is $s = t$ and $k_i - 1 = r_i - 1$, and hence $k_i = r_i$. What's left is to show that $a = b$. By what we've just shown, we have the following isomorphism

$$M_1/pM_1 \cong_R M_2/pM_2$$

which without the fact that the elementary divisors of pM_1 and pM_2 are the same would not necessarily be true (recall that $\mathbb{Z} \cong_{\mathbb{Z}} 2\mathbb{Z}$ but $\mathbb{Z}/2\mathbb{Z} \ncong_{\mathbb{Z}} 2\mathbb{Z}/2\mathbb{Z}$). Then by lemma 14.1.10 we have $F^{a+s} \cong_F F^{b+t}$ implying $a + s = b + t$ which since $s = t$ we get $a = b$, giving us that the elementary divisors match!

Finally, for the invariant factors, invariant factors $a_1 \big| \cdots \big| a_m$ and $b_1 \big| \cdots \big| b_n$ for M_1 and M_2 respectively, take the set of elementary factors for M_1 and notice that a_m is the product of the largest of the prime powers among these elementary divisors, a_{m_1} is the product of the largest prime powers once the factors of a_m have been removed, and so forth. Doing the same procedure to the set of b's, we see that these must produce the same elements, and hence the invariant factors must match, completing the proof.

Finally, we can now talk about *the* invariant factor form and *the* elementary divisor form of a finitely generated R-module M for R a PID.

Corollary 14.1.12: Computation Shortcut For Invariant Factors

Let R be a Principal Ideal Domain and let M be a finitely generated R-module. Then:

1. The elementary divisors of M are the prime power factors of the invariant factors of M.
2. The largest invariant factor of M is the product of the largest of the distinct prime powers among the elementary divisors of M, the next largest invariant factor is the product of the largest of the distinct prime powers among the remaining elementary divisors of M, and so on

Proof **:**
Fallows by the uniqueness given from 14.1.11

Corollary 14.1.13: The Fundamental Theorem of f.g. Abelian Groups

Let G be a finitely generated group G. Then:

$$G \cong \mathbb{Z}^r \times \mathbb{Z}_{n_1} \times \cdots \mathbb{Z}_{n_s}$$

for some integers $r, n_1, ..., n_s$ with the following conditions:

1. $r \geq 0$ and $n_j \geq 2$ for all j, and
2. $n_{i+1} | n_i$ for $1 \leq i \leq s-1$

Proof **:**
Take $R = \mathbb{Z}$ in theorem 14.1.5. Note that the invariant factors are listed in reverse order in Chapter 5 for ease of computation.

Summary

Any finitely generated module over a PID is fully characterised by a number $n \in \mathbb{N}$ and a choice of nonunits ideals $(a_1), ..., (a_m)$ st

$$(a_1) \supseteq (a_2) \supseteq \cdots \supseteq (a_m)$$

If M is torsion, (a_m) is the annihilator of M. The elements $a_1, a_2, ..., a)m$ may be combined to a single ideal that holds all the information about he module, the ideal has a name:

Definition 14.1.14: Characteristic Ideal

Let M be a fg module over a PID. Then the collection of invariant factors forms an ideal:

$$(a_1, ..., a_m)$$

known as the *characterstic ideal* of M.

The name "characteristic ideal" is given due to its link to the characteristic polynomial explored in section 13.4.2.

14.2 Application: Linear Algebra revisited

With the classification established, it can be put to good use in better understanding linear operators between vector spaces. Recall in section 13.2.2 that it was mentioned that chapter 14 is essentially the theoretical build-up for studying linear operators. We now make good on this promise. In this section, we define a $k[x]$-module structure on vector-spaces given by a linear operator α, and we shall define a new invariant similar to the characteristic polynomial known as the *minimal polynomial*.

To define this new module, we'll have to juggle lot of components. Let k be a field, let x be an indeterminate and let R be the polynomial ring $k[x]$. Let V be an k-module (i.e. a vector space over k). Notice that we're using the same field for V and for $k[x]$. Finally, let α be a linear operator on V, that is a linear operator from V to V where we consider the domain and codomain to be the same object. We shall extend the k-module V (i.e. a vector-space) into an $k[x]$-module using α. To accomplish this, we'll take advantage of nice properties of linear operators. First, define

$$\alpha^0 = I \qquad \alpha^n = \alpha \circ \alpha \circ \cdots \circ \alpha \qquad (n \text{ times})$$
$$a_n\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 \in \mathrm{End}_k(V)$$

where I is the identity map from V to V and $\circ$ is function composition (which is well-defined since α maps to V). We'll now define the ring action of $k[x]$-module (i.e. how polynomials act on elements in V). Let $p(x) \in k[x]$ be a polynomial. Then define an action of the ring element $p(x)$ on the module element b by:

$$\begin{aligned} p(x) \cdot v &= (a_n\alpha^n + \cdots + a_1\alpha + a_0)(v) \\ &= a_n\alpha^n(v) + \cdots + a_1\alpha(v) + a_0 v \end{aligned}$$

That is, $p(x)$ acts by substituting the linear operator α for x in $p(x)$ and applying the resulting linear operator to v). One should check that these rules satisfy the module axioms that make V into an $k[x]$-module. Note that k is a subring of $k[x]$, and this new module with constant polynomials acting on it is simply the structure of the vector space. Thus the definition of $k[x]$-module *extends* the action of k to an action of the larger ring $k[x]$.

The key is that $k[x]$ is a PID when k is a field (recall theorem 11.1.1 and its corollaries) and so we have:

$$V \cong_{k[x]} (k[x])^n \oplus (k[x]/(a_1)) \oplus \cdots \oplus (k[x]/(a_m))$$

Since V is finite dimensional, the action $k[x] \curvearrowright V$ must have rank 0, since if it had rank $n \geq 1$, then by the invariant factor form V would contain a subspace that would be k-isomorphic to copies of $k[x]$, which as a k-module:

$$k[x] \cong k \oplus k \oplus \cdots \oplus k \oplus \cdots$$

that is, $k[x]$ is an infinite dimensional k-vector space. Thus,

$$V \cong_{k[x]} (k[x]/(a_1)) \oplus \cdots \oplus (k[x]/(a_m))$$

Since the action of $k[x]$ is completely dependent on α, the above proposition may be interpreted as there being a special basis for V that will make the action of α on V rather transparent!

Example 14.2.1: Actions of this Module

1. If $\alpha = 0$, that is, the function maps every element to the identity in V (recall that V is a(n) [abelian] group). Then

$$p(x)v = a_n \cdot 0 + \cdots + a_1 \cdot 0 + a_0 v = a_0 v$$

   Thus, with such a linear operator, the $k[x]$-module structure is just like the k-module structure.

2. Another simple example is $\alpha = I$, that is, the identity. Thus

$$p(x)v = a_n v + \cdots + a_0 v = (a_n + \cdots + a_0)v$$

   Or simply the sum of coefficient of $p(x)$ act on v

3. Here's a more specific example. Take V to be the affine n-space k^n and let α be the "shift operator"

$$\alpha(x_1, ..., x_n) = (x_2, x_3, ..., x_n, 0)$$

   Since V could be a vector-space, let e_i be the usual basis vectors. Then applying this linear operator to only the basis, we get:

$$\alpha^k(e_i) = \begin{cases} e_{i-k} & \text{if } i > k \\ 0 & \text{if } i \leq k \end{cases}$$

   So that if $m < n$

$$(a_m x^m + a_{m-1} x^{m-1} + \cdots + a_0)e_n = (0, 0, ..., 0, a_m, a_{m-1}, ..., a_0)$$

4. under what condition is $W \subseteq V$ a $k[x]$-submodule of V with the $k[x]$-action? Since $k[x]$ extends the actions of k, it is necessary that W be a subspace. By the subspace criterion we have that it is necessary that $\alpha(W) \subseteq W$. Not all subspaces satisfy this conditions, if they do they are called α-stable subspaces.

Now, notice that we may take the annihilator of the $k[x]$-module V, that is all elements $p(x) \in k[x]$ such that $p(\alpha) = 0$. Since $k[x]$ is a PID, the annihilator is a principal ideal. This ideal shall be important in contexts where R need not be a field, and hence we define it more generally:

Definition 14.2.1: Annihilator Ideal and Minimal Polynomial

Let M be a finitely generated free R-module where R is an integral domain, and $\alpha \in \mathrm{End}_R(M)$. Then the ideal:

$$\mathfrak{M}_\alpha = \mathrm{Ann}_R(M)$$

is called the *annihilator ideal* of α. If K is the field of fraction of R, then α can be viewed as an element of $\mathrm{End}_K(K^n)$ and $\mathfrak{M}_\alpha^{(K)} \subseteq k[x]$ is the annihilator of α over K, where since $k[x]$ is a PID we get:

$$\mathfrak{M}_\alpha^{(K)} = (m_\alpha) \qquad m_\alpha \in k[x]$$

the resulting polynomial is called the *minimal polynomial* of α

Recall that $k[x]$ is a PID if and only if k is a field, hence it is natural to work in the setting of $R = k$ is a field to use the above module structure and the classification theorem. The term "minimal polynomial" is because it is the smallest polynomial (with respect to divisibility) that solves the equation $f(\alpha) = 0$. The reader may ask if the characteristic polynomial shares the same property, that is if

$$P_\alpha(\alpha) = 0$$

This shall indeed be the case and shall be called the *Cayley-Hamilton Theorem* and will be proven in due time[4] (the reader may try to use definition 14.1.14 to find a connection between these two polynomials and hence a proof of the above equation). The annihilator ideal shares the same nice invariance properties as the characteristic polynomial:

Lemma 14.2.2: Annihilator Ideal And Similarity

Let α be similar to β. Then $\mathfrak{M}_\alpha = \mathfrak{M}_\beta$

Proof **:**
Let $\rho \in \mathrm{GL}_R(M)$ such that $\beta = \rho \circ \alpha \circ \rho^{-1}$. Then since when composing conjugates:

$$\beta^k = \rho \circ \alpha^k \circ \rho^{-1}$$

we get that for any $f(x) \in R[x]$,

$$f(\beta) = \rho \circ f(\alpha) \circ \rho^{-1}$$

Hence, it is clear that $f(\alpha) = 0$ if and only if $f(\beta) = 0$, as we sought to show.

The converse is not necessarily, the case: if $\mathfrak{M}_\alpha = \mathfrak{M}_\beta$, it does not imply that $\alpha \sim \beta$:

Example 14.2.2: Same Annihilators, Not Similar
(put a matrix in JcF that makes this evident. Also ask if then we can condlue that the char poly and min poly is enough to deduce similarity for dimension $n \leq 3$)

On the other hand, going back to example 13.4.1, we see that $x - 1$ is the annihilator of the

[4] the curious reader may see exercise ref:HERE to prove this result with judicious application of Cramer's rule

identity matrix, while it is not the annihilator of:

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Coming back to the $k[x]$-module on V, we may use it to finally have a converse to establishing two linear operators are similar:

Lemma 14.2.3: Similarity And New Module Structure

Let α, β be linear opereators on a free R-module M. Then the corresponding $R[x]$-module structures on M are isomorphic if and only if $\alpha \sim \beta$

Proof **:**
Let M_α, M_β be the corresponding $R[x]$-modules given by α, β respectively. Assume $\alpha \sim \beta$ so that there exists a $\rho \in \mathrm{GL}(M)$ such that

$$\beta = \rho \circ \alpha \circ \rho^{-1} \Rightarrow \beta \circ \rho = \rho \circ \alpha$$

I claim that $\rho : M_\alpha \to M_\beta$ is an $R[x]$-linear map. It is certainly a R-linear map, hence to see if it $R[x]$ linear consider the action of x in M_α and M_β. Then:

$$\rho(x \cdot v) = \rho \circ \alpha(v) = \beta \circ \rho(v) = x \cdot \rho(v)$$

which then may be generalized to show that it is indeed $R[x]$-linear. Furthermore, ρ is certanly an invertible $R[x]$-linear map (being bijective and an $R[x]$-module homomorphism) and hence $M_\alpha \cong_{R[x]} M_\beta$, as we sought to show.

Corollary 14.2.4: Similarity And Isomorphism Classes

Let α, β be two linear operators on a free R-module M. Then there is a one-to-one correspondence between the similarity classes of R-linear operators of a M and the isomorphism classes of of $R[x]$-module structures on M.

Proof **:**
an immediate consequence of the above. If M has finite rank, then it is a bijection.

Exercise 14.2.1

1. Show that the $k[x]$-module on V can be constructed using the universal property of polynomials

14.3 Rational and Jordan Canonical Form

From now on, $R = F$ is a field and V is an n dimensional vector-space over F. We shall use the above results to find anice representation of a lienar operator α, namely we shall find a matrix

representation for the action of x $R[x] \curvearrowright V$.

Take the following decomposition of V:

$$V \cong k[x]/(a_1(x)) \oplus k[x]/(a_2(x)) \oplus \cdots \oplus k[x]/(a_m(x))$$

where $a_1(x), a_2(x), ..., a_m(x)$ are polynomial in $F[x]$ of degree at least 1 with the divisibility condition

$$a_1(x)|a_2(x)|\cdots|a_m(x)$$

These invariant factors are determined up to units, which in $F[x]$ is F, meaning we can safely say they are all monic. Since the annihilator of V is the ideal $(a_m(x))$ (lemma 14.1.6), we get:

$$a_1(x)|a_2(x)|\cdots|m_A(x)$$

we will make a proposition of this for easy reference:

Proposition 14.3.1: Finding Minimal Polynomial

The minimal polynomial of $m_T(x)$ is the largest invariant factor of V. All the invariant vactors of V divide $m_T(x)$

***Proof* :**
direct application of (theroem 14.1.5)

Be mindful that since everything divides $m_A(t)$, $a_i(x)|m_A(x)$, then $(m_A(x)) \subseteq (a_i(x))$, so it's the *smallest* ideal of all the invariant factors. This should make sense, as since everything divides it, all of the other ideals should be bigger. Be careful then when thinking of the word "minimal", for it's actually the *biggest* polynomial relative to all the invariant factors. It is smallest when it comes to being the *smallest* polynomial which can annihilate V over $F[x]$ (i.e $m_T(T) \equiv 0$) All the other invariant factors will not manage to do this.

To calculate V with respect to $F[x]/(a_i)$ we will do the following:

For every $F[x]/(a_i)$, we can treat it as a vector-space over $F[x]$, and get a basis

$$1, \overline{x}, \overline{x}^2, ..., \overline{x}^{k-1}$$

where $a(x) = x^k + b_{k-1}x^{k-1} + \cdots + b_1x + b_0$ is any monic polynomial in $F[x]$ and $\overline{x} = x \bmod (a(x))$.

Next recall that:

$$V \cong F[x]/(a_1(x)) \oplus F[x]/(a_2(x)) \oplus \cdots \oplus F[x]/(a_m(x))$$

Applying T on the left-side get's transalte to multiplying by x on the right-side.

$$T(V) \cong x \cdot F[x]/(a_1(x)) \oplus x \cdot F[x]/(a_2(x)) \oplus \cdots \oplus x \cdot F[x]/(a_m(x))$$

we we concentrate on on $F[x]/(a_i(x))$ and it's basis we get

$$1 \mapsto \overline{x},\ \overline{x} \mapsto \overline{x}^2,\ \overline{x}^2 \mapsto \overline{x}^2,\ \cdots \overline{x}^{k-1} \mapsto \overline{x}^k = -b_0 - b_1\overline{x} - \cdots - b_{k-1}\overline{x}^{k-1}$$

where we got the last value since

$$a(\overline{x}) = 0 \Rightarrow \overline{x}^k + b_{k-1}\overline{x}^{k-1} + \cdots + b_1\overline{x} + b_0 = 0$$

With respect to the aformentioned basis, the linear transformation T (i.e. the matrix multiplication by x) is therefore

$$\begin{pmatrix} 0 & 0 & \cdots & \cdots & \cdots & -b_0 \\ 1 & 0 & \cdots & \cdots & \cdots & -b_1 \\ 0 & 1 & \cdots & \cdots & \cdots & -b_2 \\ 0 & 0 & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \cdots & 1 & -b_{k-1} \end{pmatrix}$$

If we we do this for $F[x]/(m_A(x))$, then we can already see that if we do $m_A(V)$, we will annihilate V

Example 14.3.1: Sanity Check

Let's say $m_T(x) = x^5 + x^4 + 5x^3 - 2x^2 + 3$. Then, we we just consider $F[x]/(m_T(x))$, and think of $T(V)$ as multiplying the basis of $F[x](m_T(x))$ by x, we get the matrix:

$$\begin{pmatrix} 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

If we do $m_T(V)$ we get (test)

$$\begin{aligned} &\begin{pmatrix} 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & -1 \end{pmatrix}^5 + \begin{pmatrix} 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & -1 \end{pmatrix}^4 + 5\begin{pmatrix} 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & -1 \end{pmatrix}^3 - 2\begin{pmatrix} 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & -1 \end{pmatrix}^2 \\ &+ 3\begin{pmatrix} 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

that is, m_T indeed annihilated V, as we sought to show. This doesn't tell us yet what the other invariant factors should be, or even how to find the minimal polynomial, but it does confirm that matrices of this type do what we would hope they do.

we give this matrix a name:

Definition 14.3.2: Companion Matrix

Let $a(x) = x^k + b_{k-1}x^{k-1} + \cdots + b_1 + b_0$ be any monic polynomial in $F[x]$. The *companion matrix* of $a(x)$ is the $k \times k$ matrix whith 1's down the first subdiagonal, $-b_0, -b_1, .., -b_{k-1}$ down the alst column and zeros elsewhere. The companion matrix of $a(x)$ will be denoted $C_{a(x)}$

We can do this process to each cyclic module $F[x]/(a_i(x)$. Let β_i be the elements of V that map to the basis of $F[x]/(a_i(x))$. Then by definitin, the linear transformation T acts on β_i by the companion matrix for $a_i(x)$, as we have seen by how the right hand side is multiplies by x. If we take the union of every β_i,$\beta = \cup_i \beta_i$, then this forms a basis for V (with over $F[x]$) since the sum on the RHs is a direct sum, and with respect to this basis the linear transformation T has as matrix the *direct sum* of the companion matrices, which by ref:HERE [5] we saw can be represented by:

$$\begin{pmatrix} C_{a_1(x)} & & & \\ & C_{a_2(x)} & & \\ & & \ddots & \\ & & & C_{a_m(x)} \end{pmatrix}$$

As we can see, this matrix is determined by the invariant factors of $F[x]$-module V, in fact is uniquely determined by theorem 14.1.11. Thus, we can give this matrix a name:

Definition 14.3.3: Rational Cannonical Form

1. A matrix is said to be in *rational canonical form* if it is the direct sum of companion matrices for monic polynomials $a_1(x), a_2(x)..., a_m(x)$ of degree at least one with $a_1(x)|a_2(x)|\cdots|a_m(x)$. The polynomial $a_i(x)$ are called the *invariatn factors* of the martrix. Such a matrix is also said to be a *block diagonal* matrix with blocks the companion matrice for $a_i(x)$

2. A *rational canonical form* for a linear transformation T is a matrix represetning T which is in rational canonical form

Note that we proved that the RCF is unique, meaning if

1. we are given monic polynomial $a_1(x), a_2(x), ..., a_m(x)$ of degree at least 1 such that $a_i(x)|a_{i+1}(x)$ for all i

2. We have a linear transformation $T : V \to V$, with a matrix representation A

3. we have a basis β of V such that $[A]_\beta$ is the direct sum of companion matrices

4. Then V is the direct sum of T-stable subspaces D_i, one for each $b_i(x)$, such that the amtrix of T on each D_i is the compantion matrix of $b_i(x)$. Some more words

which prove the following result:

[5] showing this eventually using the universal property?

Theorem 14.3.4: Rational Canonical Form

Let V be a finite dimensional vector sapce over the field F and let T be a linear transformation of V

1. There is a basis for V with respect to which the matrix T is in rational canonical form, i.e., is a block diagonal matrix whose diagonal blocks are the companion matrices for monic polynomials ${}_1(x), a_2(x), ..., a_m(x)$ of degree at least one with $a_1(x)|a_2(x)|\cdots|a_m(x)$
2. The factors $a_1, ..., a_m$ are called the *invariant factors* of α.

Corollary 14.3.5: RCF And Similarity

Let $\alpha, \beta : V \to V$ be two linear on a finitely dimensional vector space V. Then $\alpha \sim \beta$ if they have the same RCF.

Proof **:**
combine everything we did earlier.

why it's called rational: if we go in a bigger field, the values don't change. In this way it's rational Here is a theorem linking to that

Corollary 14.3.6: RCF Over Extended Fields

Let A and B be two $n \times n$ matrices over a field F and suppose F is a subfield of the field K ($F \subseteq K$, or sometimes written K/F) Then

1. The RCF of A is the same whether it is computed over K or over F. The minimal and characteristic polynomials and the invariatn factors of A are the same whether A is considered as a matrix over F or as a matrix over K.
2. The matrices A and B are similar over K if and only if they are similar over F, i.e., there exists an invertible $n \times n$ matrix p with entries from K such that $B = P^{-1}AP$ if and only if there exists an (in general different) invertible $n \times n$ matrix Q with entries from F such that $B = Q^{-1}AQ$

Proof **:**
Let M be the rational canonical form of a matrix A over the field F ($\subseteq K$). Since M is in rational canonical form, by 14.3.4, it is unique. Thus, if we were to extend the field to K, then we will get the same matrix. Thus, the invariant factors are the same as well. In particular, since the minimal polynomial is the largest invariant factor of A, it also does not depend on the field over which A is viewed.

Lemma 14.3.7: Characteristic Polynomial For Block Matrix

Let $a(x) \in F[x]$ be any monic polynomial

1. The characteristic polynomial is the product of the invariant factors:

$$P_\alpha(x) = a_1(x)a_2(x)\cdots a_m(x)$$

2. If M is the block diagonal matrix

$$M = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k \end{pmatrix}$$

given by the direct sum of matrices $A_1, A_2, ..., A_k$, then the characteristic polynomial of M is the product of the characteristic polynomial of $A_1, A_2, ..., A_k$

Proof **:**
A direct application of the determinant formula.

This gives the promised Cayley-Hamilton Theorem:

Proposition 14.3.8: Cayley-Hamilton Theorem

Let A be an $n \times n$ matrices over the field F. Then the minimal polynomial of A divides the characteristic polynomial of A.

Proof **:**
This is immediate by the above lemma.

This will come back in Field theory where finding a polynomial which is used to represent an extension will exactly be the polynomial found using Cayley-Hamilton (build up, and eventually ref:HERE into field theory)

14.3.1 Converting a square matrix to Rational Canonical Form

Let A be an $n \times n$ matix with entries from a field F

1. First, diagonalize the matrix $xI - A$ over $F[x]$ (same process as described when finding Eigenspaces in ref:HERE), while keeping track of the process so that you can apply them to step (2) (you can do them on the same time). The matrix should reach the Smith Normal from form theorem ref:HERE

   Let $d_1, d_2, ..., d_m$ be the degrees ofthe monic nonconstant polynomials $a_1(x), ..., a_m(x)$ that are in the respective diagonals

2. Let P' be an $n \times n$ identity matrix. apply the operations from (1) to P'.
3. Once you have diagonalize $xI - A$ to the Smith Normal Form, the first $n - m$ columns of the matrix P' should be 0, and the remaining m columns of P' are nonzero. For each $i = 1, 2, ...m$, multiply the ith nonzero column of P' successively by $A^0 = I, A^1, A^2, ..., A^{d_i - 1}$, where d_i is the integer in (1) aboce and use the resulting column vectors (in this order) as the next d_i column of an $n \times n$ matrix P¿ Then $P^{-1}AP$ is in rational canonical form (whose diagonal blocks are the companion matrices for the polynomials $a_1(x), ..., a_m(x)$ in (1))

Example 14.3.2: Finding Rational Canonical Form

1. Let the following 3×3 matrices be over $\mathbb{Q}$:

$$A = \begin{pmatrix} 2 & -2 & 14 \\ 0 & 3 & -7 \\ 0 & 0 & 2 \end{pmatrix} \quad B = \begin{pmatrix} 0 & -4 & 85 \\ 1 & 4 & -30 \\ 0 & 0 & 3 \end{pmatrix} \quad C = \begin{pmatrix} 2 & 2 & 1 \\ 0 & 2 & -1 \\ 0 & 0 & 3 \end{pmatrix}$$

We can see that each matrix has the same characteirstic polynomial

$$c_A(x) = c_B(x) = c_C(x) = (x-2)^2(x-3)$$

Since the minimal and characteristic polynomial have the same roots, the only possibilities for the minimal polynomials are $(x-2)(x-3)$ or $(x-2)^2(x-3)$. Since there are only two options, we it will either be one minimal polynomial or the other. Thus, with some computation, we see that $(A - 2I)(A - 3I) = 0$, $(B - 2I)(B - 3I) \neq 0$ (the $1, 1$-entry is nonzero) and $(C - 2I)(C - 3I) \neq 0$ (the $1, 2$-entry is nonzero). Thus, we get

$$m_A(x) = (x-2)(x-3), \qquad m_B(x) = m_C(x) = (x-2)^2(x-3)$$

since other invariant factors would need to divide $m_B(c)/m_C(x)$, there must be no other invariatn facors for B and C. Since the invariant factors for A divide the minimal polynomial and have as a product the characteristic polynomial, we see that A has for invariant factors the polynomials, $x - 2, (x-2)(x-3) = x^2 - 5x + 6$. (For 2×2 and 3×3 matrices the determination of the characteristic and minimal polynomials determine all the invariant factors, due to _____).

Thus, we can conclude that B and C are similar, and both are not simialr to A. The rational canonical forms are(note $(x-2)^2(x-3) = x^3 - 7x^2 + 16x - 12$)

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & -6 \\ 0 & 1 & 5 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 12 \\ 10 & 0 & -16 \\ 0 & 1 & 7 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 12 \\ 1 & 0 & -16 \\ 0 & 1 & 7 \end{pmatrix}$$

2. In example (1), we managed to conclude the rational canonical form from the minimal polynomials, which we can always do for 3×3 matries and lower. However, that is not the case for matries of higher degree. Before doing such an example, let's practice by going the long way on matrix A from example (1).

   (a) We first start by reducint $xI - A$ in $\mathbb{Q}[x]$

$$xI - A = \begin{pmatrix} x-2 & 2 & -14 \\ 0 & x-3 & 7 \\ 0 & 0 & x-2 \end{pmatrix}$$

which we can do by

$$here$$

so, our final result is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & x-2 & 0 \\ 0 & 0 & x^2-5x+6 \end{pmatrix}$$

This gives us the invariant factors of $x-2$ and $x^2-5x+6=(x-2)(x-3)$ for this matrix. Now, let V be a 3 dimensional vector space over $\mathbb{Q}$ with standard basis e_1, e_2, e_3 and let T be the corresponding linear operator (which defines the action of x on V):

$$\begin{aligned} xe_1 &= T(e_1) = 2e_1 \\ xe_2 &= T(e_2) = -2e_1 + 3e_2 \\ xe_3 &= T(e_3) = 14e_1 - 7e_2 + 2e_3 \end{aligned}$$

taking the row operations from earlier

$$here$$

and applying them to this basis we'll get

$$here$$

with a final result of

$$[-e_1 - (x-3)(e_2e_1), e_3 + 7(e_2 - e_1), e_2 - e_1]$$

and pplying the result of T from eraliaer, we'll get

$$[0, -7e_1 + 7e_2 + e_3, -e_1 + e_2]$$

which correspond to the elements 1, $x-2$, and x^2-5x+6 we found earlier. Since the first invariant factor has degree 0, we can ignore it. So, we have $f_1 = -7e_1 + 7e_2 + e_3$, and $f_2 = -e_1 + e_2$. These are the $\mathbb{Q}[x]$-module generators for the two cyclic factors of V in its invariant factor decomposition as $\mathbb{Q}[x]$-modules. For the corresponding $\mathbb{Q}$-vector space bases for theset two factros, since $x-2$ has degree 1, and x^2-5x+6 has degree 2, we get f_1, f_2, xf_2 as a basis, where $xf_2 = T(f_2) = -4e_1 + 3e_2$, giving us

$$P = \begin{pmatrix} -7 & -1 & -4 \\ 7 & 1 & 3 \\ 1 & 0 & 0 \end{pmatrix}$$

which we can use to find P^{-1} and get

$$P^{-1}AP = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & -6 \\ 0 & 1 & 5 \end{pmatrix}$$

which is the same result we had in (1).

(b) We can also convert strait into Rational cannonical form. Given the operation for the diagonalization of A, we use them on the identity matrix:

$$here$$

Since we have $d_1 = 1$ and $d_2 = 2$, we can take the 2nd row the following for P:

$$\begin{pmatrix} -7 \\ 7 \\ 1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} A \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -4 \\ 3 \\ 0 \end{pmatrix}$$

which gives the same P as before.

Exercise 14.3.1

1. Let$A \in M_n(k)$ be a square matrix. Then A is similar to its trnaspose (hint: is transpose similarity invariant?)

14.4 Jordan Canonical Form

The RCF was found by taking using the invariant factors. Let us now find a form based on the elementary divisors when the characteristic polynomial factors completely. If the characteristic polynomial does not factor completely, then the field can be enlarged so that every polynomial factors completely (as mentioned in section 13.4.2 the avid reader may checkout section 17.5.1 for more details on algebraic closure). As we've outlined earlier, the similarity of two linear transformations is independent of the base field, and hence this pursuit does lead to a useful matrix form for deducing similarity.

Given a $\alpha \in \mathrm{End}_k(V)$¡ we may obtain the elementary divisor decomposition of the correspondign $k[x]$-module to be:

$$V = \bigoplus_{i,j} \frac{k[x]}{p_i(x)^{r_{ij}}}$$

Then it is an immediate consequence that the characteristic polynomial $P_\alpha(x)$ is equal to the product:

$$\prod_{i,j} p_i(x)^{r_{ij}}$$

This gives us the following results:

Lemma 14.4.1: Elementary Divisors And Characteristic Polynomial

Let α be a linear operators on a finitely dimensional vector space V, and assume its characteristic polynomial $P_\alpha(x)$ factor completely so aht

$$P_\alpha(x) = \prod_i^s (x - \lambda_i)^{m_i}$$

where λ_i are the distinct eigenvalues and m_i are teh algebraic multiplicities. Then:

$$p_i(x) = (t - \lambda_i) \qquad m_i = \sum_j r_{ij}$$

And the minimal polynomial is:

$$m_\alpha(x) = \prod_{i=1}^s (x - \lambda_i)^{\max_j\{r_{ij}\}}$$

***Proof* :**
This is an exercise in putting together what was covered so far.

With this, let's start finding the matrix representation of α. Just like before, let's tart with the cyclic factor, $F[x]/((x-\lambda)^k)$. We'll purposefully be choosing our basis so that our representation is particularly simple:

$$(\overline{x} - \lambda)^{k-1}, (\overline{x} - \lambda)^{k-2}, ..., (\overline{x} - \lambda), (\overline{x} - \lambda^)0 = 1$$

this is indeed a basis, since expanding out, simplifying, we get these elements to be a linear combination of $\overline{x}^{k-1}, \overline{x}^{k-2}, ..., \overline{x}, 1$, meaning they do form a basis. Then, we can associate the linear operator T to x, and get

$$\begin{cases} (x-\lambda)^{r-1} \mapsto x(x-\lambda)^{r-1} = \lambda(x-\lambda)^{r-1} + (x-\lambda)^r = \lambda(x-\lambda)^{r-1} \\ (x-\lambda)^{r-2} \mapsto x(x-\lambda)^{r-2} = (x-\lambda)^{r-1} + \lambda(x-\lambda)^{r-1} \\ \vdots \\ 1 \mapsto x = (x-\lambda) + \lambda \end{cases}$$

which in matrix form is be represented as:

$$\begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}$$

where the blank spots are all zero. We give a name to this matrix

Definition 14.4.2: Jordan Block

The $k \times k$ matrix with λ along the main diagonal and 1 along the first superdiagonal depicted avoce is called the $k \times k$ *elementary Jordan matrix with eigenvalue* λ, or the *Jordan block of size* k *with eigenvalue* λ

As before, we may combine all of these to get a full representation of α:

Definition 14.4.3: Jordan canonical Form (JCF)

Let α be a linear operator on a finitely dimensional vector space V. Then it's *Jordan Canonical Form* (JCF) is the combination of all the Jordan blocks:

$$\left(\begin{array}{c|c|c} J_{\lambda_1,(r_{1,j})} & & \\ \hline & \ddots & \\ \hline & & J_{\lambda_s,(r_{s,j})} \end{array}\right)$$

where $(x - \lambda_i)^{r_{ij}}$ are the elementary divisors of α.

42: An ambiguity in Uniqueness

While the RCF is is completely unique up to isomorphism, since there is no unique way of arranging the eigenvalues, there is no unique way of arranging the Jordan blocks (unlike the invariant blocks which are arranged by the divisibility of the invariant factors)

An interesting proeprty the JCF gives us is that it distinguishes algebraic and geometric multiplicity of eigenvalues:

Proposition 14.4.4: Geometric An Algebraic Multiplicity Of Eigenvalues Using JCF

Let α be a linear operator on a finitely dimensional vector space V. Then the geometric multiplicity of an eigenvalue λ of α equals the number of Jordan blocks corresponding to λ in the Jordan canonical form of α, and the algebraic multiplicity equals the number of times the eigenvalue shows up in the Jordan canonical form.

Proof **:**
The fact that the algebraic multiplicity of the eigenvalue corresponds to the number or times it shows up in the JCF is immediate, hence let's show the number of Jordan blocks equasl the geometric multiplicity. Let:

$$J =: \begin{pmatrix} \lambda & 1 & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}$$

be a Jordan block and $v = \begin{pmatrix} v_1 \\ \vdots \\ v_r \end{pmatrix}$ be an eigenvector corresponding to λ. Then:

$$\lambda \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_r \end{pmatrix} = J \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_r \end{pmatrix} = \begin{pmatrix} \lambda v_1 + v_2 \\ \lambda v_2 + v_3 \\ \vdots \\ \lambda v_r \end{pmatrix}$$

which solving gives us:

$$v_2 = \cdots = v_r = 0$$

Hence, the eigenspace of λ is generated by:

$$v = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

completing the proof.

Corollary 14.4.5: Diagonalizability And Minimal Polynomial

Assume the characteristic polynomial of $\alpha \in \mathrm{End}_k(V)$ factors completely over k. Then α is diagonalizable iff the minimal polynomial of α has no multiple roots

Proof **:**
exercise

Naturally, the diagonalizability of a matrix also depends on the base field, for example the rotation by $\pi/2$ degrees for $\mathbb{R}^2$ is not diagonalizable, while it is diagonalizable over $\mathbb{C}$

14.4.1 Changing From one Canonical Form to Another

see D& F

14.4.2 Computation: Convert to JCF

15

Tensor Products Of Modules: Linear Algebra III

In this chapter we'll introduce a new way of constructing modules by defining a new operation: the tensor product. The chapter begins with a motivation for this new construction, and defines the tensor product. The chapter then explores important properties of the tensor product, and finishes off by constructing some important algebraic structures, namely different tensor algebras, that are very important in geometry.

15.1 What's a Tensor Product?

On a very high level, the tensor products let us take multi-linear maps and transform them in linear maps which, as these things usually are in algebra, is done in a "natural" way (i.e. with as little extra information as possible given the context). Thus, multi-linear map from chapter 13, section 13.3, 13.4.1, and 13.5 (including the trace, determinant, multilinear forms, inner product, and so forth) can be suitably transformed to be a linear map, allowing us to use all the results of linear maps we have covered the previous sections![1] The way we will do this is by "replacing" the domain of multi-linear maps ($M_1 \oplus M_2 \oplus \cdots \oplus M_n$) to form a new algebraic object that will be suitable for linear maps. This new domain will be denoted $M_1 \otimes M_2 \otimes \cdots \otimes M_n$.

Thus, looking at tensor products very generally, and quite loosely![2], they let you define a way to "formally multiply" two modules A, B, denoted $A \times B$ in such a way that if $\varphi : A \times B \to C$ is a bilinear map, we may "raise" it to a linear map $\Phi : A \otimes B \to C$. We already encountered something

[1] For those who were motivated by the geometric picture of linear maps in section 13.2.4, this also means that tensor products strongly related to constructing higher-dimensional spaces from lower dimensional spaces

[2] We are omitting distinctions between commutative and non-commutative, (R, S)-bimodules, and so forth

that sounds like we multiplied two modules, that is the product of two modules $M \times N$. We showed that product of modules exists and can be defined on modules: if M and N are modules, then $M \times N$ is a module. What will distinguish $M \times N$ and $M \otimes N$ is that $M \times N$ simply "holds" the information of M and N. We can see this through the universal property of products: for any module L,

$$\begin{tikzcd}
 & L \arrow[d, dashed, "\varphi"] \arrow[ddl, bend right, "\varphi_M"'] \arrow[ddr, bend left, "\varphi_N"] & \\
 & M \oplus N \arrow[dl, "\pi_M"'] \arrow[dr, "\pi_N"] & \\
M & & N
\end{tikzcd}$$

which is also the coproduct for R-modules (for the reason as with abelian groups) [3]. This case is different, you need to have a bilinear map. As an example, consider $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ given by $(x, y) \mapsto xy$ (i.e. multiplication). Then this map is multilinear, meaning our conventional tools of linear algebra does not apply to it. We want to somehow "raise" $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R} \otimes \mathbb{R}$ so that multi-linear maps such as the one just presented become linear and hence is open to all of our tools from linear algebra. In more algebraic language, we would want the $\mathbb{R} \otimes \mathbb{R}$ to respect the following commutative diagram:

$$\begin{tikzcd}
M \otimes N \arrow[r, "\psi"] & L \\
M \times N \arrow[u, "i"] \arrow[ur, "\varphi"'] &
\end{tikzcd}$$

where φ is multi-linear, ψ is linear, and $\varphi = \psi \circ \iota$. In this way, the map $\mathbb{R} \otimes \mathbb{R} \to \mathbb{R}$, $x \otimes y \mapsto xy$ will be a *linear map*. In the special case of the multiplication operation above, we made multiplication (the binary operation of the ring) something we can study using linear algebra! Thus, a mnemonic for remembering a tensor product is that tensor products let's us "multiply" elements of modules and keep linearity. The result of this multiplication will be inside a new module $\mathbb{R} \otimes \mathbb{R}$. Though this mnemonic can be helpful, be sure not to loose sight of the original more theoretical interpretation of tensor products, since there are many multilinear maps that do not arise as multiplication (see example 13.3.3). However, the case where the multilinear map does arrise as a form of multiplication (especially in a geometric sense, like how the area of a shape is given by multiplying its sides) is very rich in content and will be one of the two most prominent form of multilinear map that will show up due to their usefulness. In section 15.4, we will will make the intuition that the tensor product is multiplication more rigorous by showing how we can actually take any R-module M, and naturally create an R-algebra $\mathcal{T}(M)$ from it.

Since we can see tensor products as a form of multiplying module elements, the next question becomes what are the properties we will want this new "multiplication" operation to have. When we do regular multiplication in a ring, one of the first thing's we do is have it be distributive

$$(a + b)c = ac + bc \qquad a(b + c) = ab + ac$$

We might also want it to be associative, however we shall see that associativity shall be a provable property in the construction. Next, we have that we are working with module elements, meaning there is a notion of *scalars*. Let's for $c \in R$ and $n, m \in M$, let's write down cn, cm instead of $c \cdot n, c \cdot m$.

[3] However, this is *not* the coproduct of commutative and non-commutative R-algebras! The shall be constructed using the tensor product

We'll use the $\cdot$ symbol for now to represent "multiplying" module elements. How would we want scalars to interact with our new multiplication? Similar to the motivation for algebra in section 12.1.2, if we have $n \cdot m$, then multiplying either n or m by a scalar should have the same result:

$$cn \cdot m = n \cdot cm$$

which is analogous to algebras (a structure where we already dealt with definig module element multiplication). So the scalar should be able to move around freely. In fact, these are the "weakest" conditions that we need to add to define multiplication. All other conditions (like associativity, inverse, etc) are not needed to a new operation. Thus, when we try to define the tensor product, we'll be mindfull of these two properties.

15.1.1 Constructing the Tensor Product

There are many ways of construction the tensor product in different levels of generality. For this construction, we start with the tensor product of two modules over a commutative ring before moving on to modules over general rings.

Let M, N be R modules and $\varphi : M \times N \to L$ a bilinear map. We seek a structure $M \otimes N$ so that $\psi : M \otimes N \to L$ is a linear map where ψ is induced by φ and $M \otimes N$ is an R-module. If we can naturally include $\iota : M \times N \to M \otimes N$, then it is natural to ask for $\varphi = \psi \circ \iota$, that is:

$$\begin{tikzcd} M \otimes N \arrow[r, "\psi"] & L \\ M \times N \arrow[u, "\iota"] \arrow[ur, "\varphi"'] & \end{tikzcd}$$

If furthermore, $(M \otimes N, \iota)$ is the unique pair that makes the above diagram commute, then we have a universal proeprty, that is the tensor product is represents multilinear maps. The following construction $M \otimes N$ is enlightening enough to deserve (TBD).

For $M \otimes N$ to be an R-module, we would require that it is an abelian group with an R-action. Consider $F^R(M \times N)$, the R-module with $M \times N$ as the free-generators Notice that this group can be *huge*, if $M = \mathbb{R}^2$ and $N = \mathbb{R}^3$, then the set $M \times N$ consists of all *generators*! Since $F^{\mathbf{Ab}}(F \times N)$ is free, it has no relations. As we mentioned earlier, we want the distributive law to hold:

$$(m_1 + m_2, n) = (m_1, n) + (m_2, n) \tag{15.1}$$
$$(m, n_1 + n_2) = (m, n_1) + (m, n_2) \tag{15.2}$$
$$\tag{15.3}$$

Furthermore, we want scalars to commute:

$$(rm, n) = (m, rn) \tag{15.4}$$

To accomplish this, for every $m, m_1, m_2 \in M$,$n, n_1, n_2 \in N$, and $r \in \mathbb{R}$, take the set of equations:

$$(m_1 + m_2, n) - (m_1, n) - (m_2, n) \tag{15.5}$$
$$(m, n_1 + n_2) - (m, n_1) - (m, n_2) \tag{15.6}$$
$$(rm, n) - (m, rn) \tag{15.7}$$

These equations form an abelian group, label it K. Then $F^R(M \times N)/K$ is a new abelian group in which we have imposed these relations, namely equation (15.1), (15.2), and (15.4) are all true in $F(M \times N)/K$! If you are unconvinced, verify by hand that:

$$[(m_1 + m_2, n)] = [(m_1, n)] + [(m_2, n)]$$

By equation (15.4), $[(rm, n)] = [(m, rn)]$. This will be key in ensuring our map will be *linear*. The R-module $F(M \times N)/K$ is given a name:

Definition 15.1.1: Tensor Product (Over Commutative Ring)

Let M, N be R-module where R is commutative. Then the *tensor product* of M and N is defined to be:

$$M \otimes_R N := F^R(M \times N)/K$$

where the subscript R is dropped if the context permits.

The cosets $[(m, n)] \in M \otimes_R N$ are denoted $m \otimes_R n$; such an element is called a *simple tensor*. There is no reason for every element of $M \otimes N$ to be a simple tensor (there will be some special tensor product in which this is the case, but is a special case rather than expected structure). However, they do form the generators of $M \otimes_R N$, and hence if two R-module homomorphisms agree on pure tensors, they are equal.

Theorem 15.1.2: Universal Property Of Tensor Product (Over Commutative Rings)

Let M, N be R-modules over a commutative ring R and let $\varphi : M \times N \to L$ be a R-bilinear map. Then there exists a unique linear map $\psi : M \otimes N \to L$ such that $\varphi = \psi \circ \iota$, that is, the following diagram commutes:

$$\begin{tikzcd} M \otimes N \arrow[r, "\psi"] & L \\ M \times N \arrow[u, "\iota"] \arrow[ur, "\varphi"'] & \end{tikzcd}$$

Proof **:**
Let $\varphi : M \times N \to L$ be a bilinear map. Consider $M \times N$ is just a set and φ as a set-map, by the universal properties of free R-modules there exists an R-module homomorphism $\tilde{\psi}$ such that

$$\begin{tikzcd} F^R(M \times N) \arrow[r, "\tilde{\psi}"] & L \\ M \times N \arrow[u, "\iota"] \arrow[ur, "\varphi"'] & \end{tikzcd}$$

Notice that for any $x \in F(m, n)$ such that $x \in K$ given K as defined above:

$$\psi(\iota(x)) = \varphi(x) = 0$$

which follows by R-bilinearity of φ. Hence, by the universal property of kernels, there exists a

unique map ψ such that:

$$\begin{tikzcd}
M \otimes_R N \arrow[r, "\psi"] & L \\
F^R(M \times N) \arrow[u] \arrow[ur, "\tilde{\psi}"] & \\
M \times N \arrow[u, "\iota"] \arrow[uu, bend left, "\otimes"] \arrow[uur, "\varphi"'] &
\end{tikzcd}$$

giving us our unique R-module homomorphism, completing the proof.

Hence, we have defined the tensor product over R-modules with over a commutative ring R. We now generlize to R-modules over a non-commutative ring. Since the module are non-commutative, we shall need to specify whether we are working over *left* or *right* modules. This is key, for it changes how we may wish to define our action, in particular the equation:

$$(rm, n) = (m, rn) \tag{15.8}$$

Now depends on whether we have a left or right action. To emphasize the non-commutative nature, we shall consider the case where M is a *right* R-module and N a *left* R-module, while keeping in mind that M and N may be either left or right modules. In the case just mentioned, it will be required that:

$$(mr, n) = (m, rn) \tag{15.9}$$

Furthermore, it is no longer clear what the R-module structure may be[4]. However, it is always an abelian group. Thus, inspired by our earlier equations, we define the following generalizations of a R-bilinear map:

Definition 15.1.3: R-balanced map

Let M, N be R-modules, and let L be an abelian group (written additively). Then a map $\varphi : M \times N \to L$ is called R- *balanced* or *middle linear with respect to* R if

$$\begin{aligned}
\varphi(m_1 + m_2, n) &= \varphi(m_1, n) + \varphi(m_2, n) \\
\varphi(m, n_1 + n) &= \varphi(m, n_1) + \varphi(m, n_2) \\
\varphi(m, rn) &= \varphi(mr, n)
\end{aligned}$$

Note how this is more powerful than being $\mathbb{Z}$-biliner since the scalar r can jump between the two modules, however since L is not an R-module it makes no sense to say the map is R-bilinear. The simplest example of R-balanced maps are R-bilinear maps, satisfying all the above property and homogeneity of the scalar.

Naturally, we may want a universal property for R-balanced maps similar to the above. As it happens, $M \otimes_R N$ is already the solution, as we shall now demonstrated. First take the equation (15.1), (15.2), and (15.8) to be the generators of a group K' and define:

$$F^{\mathbf{Ab}}(M \times N)/K' = \mathbb{Z}^{\oplus M \times N}/K'$$

[4]Why is not as simple as defining $r \cdot (m, n) = (mr, n)$?

I claim that $F^R(M \times N)/K = M \otimes_R N$ is isomorphic to the above as groups. This will allow us to readily generlize theorem 15.1.2. Indeed, every element of $M \otimes_R N$ can be written as

$$\sum_i m_i \otimes n_i$$

Hence, we have a surjective group homomorpihsm

$$\mathbb{Z}^{\oplus M \times N} \twoheadrightarrow M \otimes_R N$$

It should be clear (or verified) that the kernel of the above map is generated by elements of the form

$$\begin{aligned} (m_1 + m_2, n) - (m_1, n) - (m_2, n) \\ (m, n_1 + n_2) - (m, n_1) - (m, n_2) \\ (mr, n) - (m, rn) \end{aligned}$$

which are the generators of K', hence as groups:

$$F^{\mathbf{Ab}}(M \times N)/K' \cong F^R(M \times N)/K$$

Hence, we see that $M \otimes_R N$ is also the solution to the following universal propert

Theorem 15.1.4: Universal Property Of Tensor Product (Over Non-Commutative Rings)

Let M, N be R-modules over a (not necessarily commutative) ring R, L an abelian group, and and let $\varphi : M \times N \to L$ be an R-balanced map. Then there exists a unique group homomorphism $\psi : M \otimes_R N \to L$ such that $\varphi = \psi \circ \iota$, that is, the following diagram commutes:

$$\begin{tikzcd} M \otimes N \arrow[r, "\psi"] & L \\ M \times N \arrow[u, "\iota"] \arrow[ur, "\varphi"'] & \end{tikzcd}$$

Proof **:**
See the proof of theorem 15.1.2 and modify appropriately.

Hence, we have a tensor product defined over arbitrary R-modules. However, the consequence is that we have lost the module structure on $M \otimes_R N$. We finish this section by showing we may still define a module structure provided a slight augmentation in structure. Again, let M be a right R-module and N a left R-module (remembering again that this choice was arbirary). We shall augment N to be the following new type of structure:

Definition 15.1.5: (S, R)-Bimodule

Let R and S be rings. An abelian group M is called an (S, R)-bimodule if M is a left S-module, a right R-module, and $s(mr) = (sm)r$ for all $s, \in S$, $r \in R$, and $m \in M$.

Example 15.1.1: Bimodules

1. Let $f : R \to S$ be any ring homomorhpism. Then S has a natural (S, R)-bmodule structure with $s \cdot r = sf(r)$.
2. As a special case of the last example, let $I \subseteq R$ be a two-sided ideal. Then R/I is a $(R/I, R)$-bimodule.
3. If R is commutative, any R-module has a compatible left and right R-module structure, and hence is always a natural (R, R)-module. This is called the *standard* R-module structure on M.

Hence, to define a module structure on $M \otimes_R N$, we can see that if M is a (S, R)-bimodule, we can define on each simple tensor the S-action:

$$sm_i \otimes n_i = (sm_i) \otimes n_i$$

and extend linearly. The proof of which is very similar to our earlier construction. Notice too that this is a natural (R, S)-bimodule, and if R is commutative then we have the standard R-bimodule structure, recovering the definition from the commutative case. Thus, we have a general notion of tensor product

Definition 15.1.6: Tensor Product

Let M, N be two (left or right) R-modules. Then their tensor product is

$$M \otimes_R N := F^R(M \times N)/K$$

as defined above. If M is furthermore a (S, R)-bidmodule. Then the above is an S-module

Example 15.1.2: Tensor Products

1. Natrually, for any $M \otimes_R N$, $m \otimes 0 = 0$ and $0 \otimes n = 0$
2. Let $\mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/3\mathbb{Z}$. Since $\overline{a} = \overline{3a}$

   $$\overline{a} \otimes b = \overline{3a} \otimes \overline{b} = \overline{a} \otimes \overline{3b} = \overline{a} \otimes \overline{0} = \overline{0}$$

   And since this is true for any simple tensor, it is true for the finite sum of simple tensors, and so $\mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/3\mathbb{Z} = 0$. However, If we have $\mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/2\mathbb{Z}$, then the story is different. We can show that $1 \otimes 1 \neq 0$. Take the map

   $$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} \qquad (a, b) \mapsto ab$$

   then this is a bilinear $\mathbb{Z}$-map. Thus $\Phi : \mathbb{Z}/2\mathbb{Z} \otimes \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ is the corresponding $\mathbb{Z}$-homomorphism sending $1 \otimes 1 \mapsto 1$. Since 1 is non-zero, $1 \otimes 1$ is non-zero (contrapositive of $v = 0$ then $\varphi(v) = 0$ for group-homomorphisms). In fact, that map is an isomorphism (on can readily check injectivity and surjectivity), thus $\mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/2\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z}$
3. More generally, show that:

   $$\mathbb{Z}/m\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/d\mathbb{Z}$$

   where $d = gcd(a, b)$.

4. Show that $\mathbb{Q}/\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Q}/\mathbb{Z} = 0$
5. Show that $\mathbb{Q} \otimes_{\mathbb{Q}} \mathbb{Q}$ and $\mathbb{Q} \otimes_{\mathbb{Z}} \mathbb{Q}$ are iso as 1-dimensional vector spaces over $\mathbb{Q}$. Show that $\mathbb{C} \otimes_{\mathbb{C}} \mathbb{C}$ and $\mathbb{C} \otimes_R \mathbb{C}$ are *not* isomorphic (show their dimensions as $\mathbb{R}$-modules don't match)

Example 15.1.3: Tensor Products in Calculus

I'll take a moment here to point out how some other non-algebra text-books define tensor properties (which is common in physics and sometimes analysis). Most of the time in physics, you'll be working over many vector-spaces, usually over the same field $\mathbb{F}$, sometimes of varying dimensions. In this context, the very often implicitly use the universal property of tensors to define the multilinear map

$$T(x_1, ..., x_n)$$

to *be* a k-tensor. and if you have two k-tensors (i.e. multilinear maps), they define the tensor product as

$$(T \otimes S)(x_1, ..., x_n, x_1', ..., x_m') = T(x_1, ..., x_n) \cdot S(x_1', ..., x_m')$$

The point of looking at tensors as multi-linear maps is of interest in analysis since we can then bring in concepts like continuity, differentiability, or C^r and C^∞, that is functinos that are r-times continuously differentiable or smooth.

Proof. here □

Next, we want to show that we can keep "adding" tensors products to a tensor, that is, make sense of $M \otimes N \otimes L$, and more generally an n-fold tensor product $M_1 \otimes M_2 \otimes \cdots \otimes M_n$, unambiguously whenever it is defined:

Theorem 15.1.7: Associativity of the Tensor Product

Suppose M is a right R-module, and N is an (R, T)-bimodule, and L is a left T-module. Then there is a unique isomorphism

$$(M \otimes_R N) \otimes_T L \cong M \otimes_R (N \otimes_T L)$$

of abelian groups such that $(m \otimes n) \otimes l \mapsto m \otimes (n \otimes l)$. If M is an (S, R)-bimodule, then this is an isomorphism of S-modules.

Proof **:**
exercise

Corollary 15.1.8: Associativity for tensor products *with* a commutative ring

Suppose R is commutative and M, N, and L, are left R-modules. Then

$$(M \otimes N) \otimes L \cong M \otimes (N \otimes L)$$

as R-modules for the standard R-module structure on M, N, L.

Naturally, we can extend the notation of biliniar maps to R-multiliniar maps, and we can then extend the notion of making multiliniar maps into linear maps (i.e. R-module homomorphisms)

Definition 15.1.9: Multi-linear maps

Let R be a commutative ring and let $M_1, ..., M_n$ and L be R-module with the standard R-module structures. A map $\varphi : M_1 \times \cdots \times M_n \to L$ is called n-multiliniear over R (or simply multilinear if n and R are clear from the context) if it is an R-module homomorphism in each component when the other compohent entries are kept constant, i.i., for each i

$$\begin{aligned}\varphi(m_1, ..., m_{i-1}, m_i + r'm_i', m_{i+1}, ..., m_n) \\ = r\varphi(m_1, ..., m_i, ..., m_n) + r'\varphi(m_1, ..., m_i', ..., m_n)\end{aligned}$$

for all $m_i, m_i' \in M_i$ and $r, r' \in R$. WHen $n = 2$ (respectively, 3) one says φ is biliniear (respectively trilinear) rather than 2-multilinear (or 3-multilinear)

Theorem 15.1.10: universal property for multilinear maps

Let R be a commutative ring and let M_1, .., M_n, L be R-modules. Let $M_1 \otimes M_2 \otimes \cdots \otimes M_n$ denote any bracketing of the tensor product of these module. Then if $phi : M_1 \times \cdots \times M_n \to L$ is an n-multilinear map, there is a unique R-module homomorphism $\psi : M_1 \otimes \cdots \otimes M_n \to L$ such that $\varphi = \psi \circ \iota$:

$$\begin{tikzcd} M_1 \times \cdots \times M_n \arrow[r, "\iota"] \arrow[dr, "\varphi"'] & M_1 \otimes \cdots \otimes M_n \arrow[d, "\psi"] \\ & L \end{tikzcd}$$

This final result will be important for section 15.3

Theorem 15.1.11: The "Tensor Product" of two Homomorphisms

Let M, M' be right R-module, let N N' be left R-module, and suppose $\varphi : M \to M'$, and $\psi : N \to N'$ are R-module homomorphisms.

1. There is a unique gruop homomorphism, denoted by $\varphi \otimes \psi$, mapping $M \otimes_R N$ into $M' \otimes_R N'$ such that $(\varphi \otimes \psi)(m \otimes N) = \varphi(m) \otimes \psi(n)$ for all $m \in M$ and $n \in N$

2. If M, M' are also (S, R)-bimodules for some ring S and φ is also an S-module homomorphism, then $\varphi \otimes \psi$ is a homomorphism of left S-module. In particular, if R is commutative then $\varphi \otimes \psi$ is alwasy an R-module homomorphism for the standard R-module structure

3. if $\lambda : M' \to M''$ and $\mu : N' \to N''$ are R-module homomorphisms then

$$(\lambda \otimes \mu) \circ (\varphi \otimes \psi) = (\lambda \circ \varphi) \otimes (\mu \circ \psi)$$

Exercise 15.1.1

1. Let $\varphi : R \to S$ be a ring homomorphism, and M and R-module. Show that $S \otimes_R M$ has a

natural S-module structure on it. (there may be some functorial way of going between these?? Some adjoint between R-mod and S-mod?)

2. Show that the tensor product is the *coproducts* of commutative R-algebras.
3. Let M be an R-module over an integral domain and $K = K(R)$ it's field of fractions. Show that the rank of M is equal to the dimension of $K \otimes_R M$ as a K-vector space.
4. Show that if $f : A \to M$ and $g : B \to M$ are linear, the map $f + g : A \times B \to M$ is linear and not bilinear. This maps shows that the product and coproduct of commutative R-modules match, and is not the same as the tensor product.

15.2 Properties of Tensor Products

With an intuition for tensor product and a proper construction, we can move on to the properties of tensor products. Really make sure that you understood the last section before you move on – no point in basing the next section on faulty ideas.

Theorem 15.2.1: Tensors products of direct sums

Let M, and M' be right R-modules and let N, N' be left R-modules. Then there are unique group isomorphisms

$$(M \oplus M') \otimes_R N \cong (M \otimes_R N) \oplus (M' \otimes_R N)$$
$$M \oplus (N \otimes_R N') \cong (M \otimes_R N') \oplus (M \otimes_R N')$$

such that $(m, m') \otimes n \mapsto (m \otimes n, m' \otimes n)$ and $m \otimes (n, n') \mapsto (m \otimes n, m \otimes n')$ respectively. If M, M' are also (S, R)-bimdoulde,s then these are isomorphisms of left S-modules. In particular ,if R is commutative, these are isomorphisms of R-modules

Proof **:**
exercise

Corollary 15.2.2: Arbitrary direct sums commute with tensor products

$$\left(\bigoplus_{i \in \mathcal{I}} M_i\right) \otimes N = \bigoplus_{i \in \mathcal{I}} (M_i \otimes N)$$

Proof. exercise □

Corollary 15.2.3: Tensor of Vector-spaces

Let V, W be vector-spaces over F. Then (look at exercice 12 proof).

***Proof* :**
transcribe this image

Let V be a vector space over the field F and let V and V' be nonzero elements of V.

Consider that $B = \{e_i\}_I$ be a basis for V. Every simple tensor of $V \otimes_F V$ can be written in the form $\sum vi \otimes e_i$. For all $i \in I$, $F-$module injectin $\varphi_i : V \to V \otimes_F V$, $V \to V \otimes e_i$

By universal property of direct sums, an $F-$module homomorphism ψ : $\oplus_i V \to V \otimes_F V$, $\psi(V_i) = \sum V_i \otimes e_i$.
Now define $\varphi : V \times V \to \oplus_e V$, $\varphi\left(v \sum a_i e_i\right) = (va_i)$
φ is well defined because every element of V is a finite linear combination of e_i, and φ is bilinear.
We have an $F-$module homomorphism.
$\Phi : V \otimes_F V \to \oplus_e V$, $\phi(V \otimes e_i) = (e_i(V)_i$ where l_i denotes the ith natural injection $V \to \oplus_i V$.
So Φ and ψ are mutual inverses

$$\begin{aligned}(\phi \circ \psi)(V_i) &= \phi\left(\sum V_i \otimes e_i\right) \\ &= (V_i)\end{aligned}$$

$$\begin{aligned}\left(\psi \circ \phi)(V \otimes \left(\sum a_i e_i\right)\right) &= \psi(Va_i) \\ &= \sum Va_i \otimes e_i \\ &= V \otimes \sum a_i e_i\end{aligned}$$

Suppose that $V \otimes V' = V' \otimes V$
Consider that $V = \sum a_i e_i$ and $V' = \sum b_i e_i$
Then $\phi(V \otimes V') = \phi(V' \otimes V)$
From this we get $(Vb_i) = (V'a_i)$ $\forall j \in I$
If $(Vb_i) = 0$, then $Vb_i = 0$ $\forall_i$, as $V \neq 0$ then $b_i = 0 \, \forall_i \in I$ $w = 0$ and this is contradiction.

$Vb_i = wa_i \neq 0 \, \forall_i \in I$
$v = b_i^{-1} a_i w$, so v and v' are linearly depndent
Suppose $v = av'$ so $v \otimes v' = v' \otimes v$
If $v = av'$

$$\begin{aligned}v \otimes v' &= v \otimes av \\ &= v_a \otimes v \\ &= v' \otimes v\end{aligned}$$

Now we get $v \otimes v' = v' \otimes v$ in $V \otimes_F V$ if and only if $v = av'$

Figure 15.1: type up!

pretty cool

This last corollary will be very useful when proving

$$\mathrm{Hom}_R(\oplus R_i, R) \cong \prod \mathrm{Hom}_R(R_i, R)$$

with another fact from the book

It can be more directly applied to generalise the extensions of scalars:

Corollary 15.2.4: Extension of Scalars of Free Modules

The module obtained from the free R-module $N \cong R^n$ by extension of scalars from R to S is the free S-module S^n, i.e.

$$S \otimes_R R^n \cong S^n$$

as left S-module

Proof. follows very fast from the theorem and last corollary. □

Corollary 15.2.5: tensor product of commutative free modules

Let R be a commutative ring and let $M \cong R^s$ and $N \cong R^t$ be free R-module with basis $m_1, ..., m_s$ and $n_1, ..., n_t$ respectively. Then $M \otimes_R N$ is a free R-module of rank st with basis $m_i \otimes n_j$, $1 \leq i \leq s$ and $1 \leq j \leq j$, i.e.

$$R^s \otimes_R R^t \cong R^{st}$$

Note: By the previous exercice, we see that this will generalise well. That is, the tensro product of two free modules of arbitrary rank over a commutative ring is free.

We've dealt with commutative modules quite often. So we can naturally enquire if such a commutative structure between two commutative R-module will nicely transfer. The answer is yes:

Proposition 15.2.6: Commutative R-modules and tensor products

Suppose R is a commutative ring and let M, N be left R-modules, considered with the standard R-module structures. Then there is a unique R-module isomorphism

$$M \otimes_R N \cong N \otimes_R M$$

Proof. not too bad. p.374 □

After doing some excercices, I've gained a second intuition behind tensor products. Many tensor products between similar structure (ex R a commutative ring, I, J ideals, and $R/I \otimes_R R/J$) will share and/or be isomorphic to other modules. In the aforementioned example $R/I \otimes_R R/J \cong R/(I + J)$. I don' know why yet, but I'm sure there's some more intuition here! Another example (17) in the book, where somehow, a bilinar map with I an ideal of a ring, when it became and R-module homomorphism, with $I \otimes_R I$, included a derivative!

Tensor Product of Algebras

We'll be extensively dealing with R-algebras when we'll come accorss tensor algebras, so having an early intuition how R-algebras behave under the tensor product is a good to have early

Theorem 15.2.7: Tensor product of R-algebras

Let R be a commutative ring and let A and B be R-algebras. Then the multiplication $(a \otimes b)(a' \otimes b') = aa' \otimes bb'$ is well defined and makes $A \otimes_R B$ into an R-algebra

intuition

Proof. p.374 □

Example 15.2.1: Tensoring Algebras

1. the example elaborated was $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$

15.3 Functors Between Mod and Scalar Extension

Let N be an R-module, and $R \subseteq S$. We want to find a suitable S-module in which we can *extend* the scalars acting on N in the "natural way". As the previous example showed, sometimes, this space will be smaller (will smaller always imply 0?) so the "extension of scalars" can result in smaller spaces.

(don't forget this: Richard Borcherd)

To extend the scalars, it means that we want the following relations to be well defined:

$$(s_1 + s_2)n = s_1 n + s_2 n \tag{15.10}$$
$$s(n_1 + n_2) = sn_1 + sn_2 \tag{15.11}$$
$$s_1(s_2 n) = (s_1 s_2)n \tag{15.12}$$

and in particular for that last relation:

$$s_1(rn) = (s_1 r)n \tag{15.13}$$

The reason for the restriction of the second S element to elements $r \in R$ will be made clear soon.

To define these relations, there's a very interesting construction! First, consider every elements from $S \times N$ as a distinct element to generate a free $\mathbb{Z}$-module. So for every $s, n \in N$, (s, n) is an element that is one of the generators for our free $\mathbb{Z}$-module. We choose it to be a module over $\mathbb{Z}$ since every $\mathbb{Z}$ module is abelian group, this is naturally a free abelian group. Since N and S are abelian groups, it is natural to define the free-group as a free abelian group, that is, the elements of the sums commute. All possible sums and "inverses" of (s_i, n_i) is an element of the free $\mathbb{Z}$-module. Addition of elements is defined as creating a linear combination of the two elements:

$$(s_1, n_1) + (s_1, s_2) = \sum_{i=1}^{2} (s_i, n_i)$$

And add the zero element as usual in the construction of a free group. Let this free $\mathbb{Z}$-module be denoted M. Note that this free module has nothing in common with N except that it was part of the elements that were the generators for it (with the elements of S being the other part of the generators for the (s, n) elements). Note that this module can be absolutely humongous! If $N = \mathbb{R}^3$, an $S = \mathbb{R}$, then realize that every

$$(s_i, n_i) \qquad s_i \in \mathbb{R},\ n_i \in \mathbb{R}^3$$

are *generators* of the free group – that is the set $\mathbb{R} \times \mathbb{R}^3$ is the set of generators! All possible finite linear combinations of these elements produce the free element.

This module is without relation. What we'll do to impose a relation on it, and show that this new set will be a submodule of this free $\mathbb{Z}$-module. To construct the quotienting module, we'll construct a sub-module of M which is generated by the elements of the form

$$(s_1 + s_2, n) - (s_1, n) - (s_2, n) \tag{15.14}$$
$$(s, n_1 + n_2) - (s, n_1) - (s, n_2) \tag{15.15}$$
$$(sr, n) - (s, rn) \tag{15.16}$$

For all $s_1, s_2 \in S,\ n, n_1, n_2 \in N$, $r \in R$, where rn refers to the R-module structure already present in N. Let K represent the set using the three equations as the generators. The notice that K will represent all possible ways of getting the "zero" element if we were extending the scalars of S to N! If we already new K was a S-module, it is clear that

$$(s_1 + s_2)n - s_1n - s_2n = s_1n + s_2n - s_1n - s_2n = s_1n - s_1n + s_2n - s_2n = 0$$

This is exactly what we're trying to do here. That means that if we consider M/K, that is, quotient M by K, this particular sub-module of our free $\mathbb{Z}$-module (that is our free abelian group) represents "zero", which is exactly how we think about elements of a quotient modules (and quotient objects more generally)!

In fact, we can already conclude that M/K is well-defined as a *quotient group* (or a quotient $\mathbb{Z}$-module), since M is abelian! It is an easy matter to check that K is a subgroup (i.e. $\mathbb{Z}$-submodule) of M. You can use the map $\iota : M \to M/K$, $(s, m) \mapsto [(s, m)]$. Our map ι

The resulting quotient is denoted $S \otimes_R N (= M/K)$ (where R can be dropped if contexts permits), and is called the *tensor product of S and N over R*. Let $s \otimes n$ denote the coset containing (s, n) in $S \otimes_R N$ (a coset represented by a single element is called a *simple tensor*). Note that since $[x + y] = [x] + [y]$, then $[\sum_{i=1}^n (s_i, n_i)] = \sum_{i=1}^n [(s_i, n_i)]$ and so

$$\left[\sum_{i=1}^n (s_i, n_i)\right] = \sum_{i=1}^n s_i \otimes n_i$$

Note that not every tensor can be written as a simple tensor, since they might belong to different equivalences classes.

Examining the elements of this quotient group, it should be clear that the following relations are satisfied:

$$(s_1 + s_2) \otimes n = s_1 \otimes n + s_2 \otimes n \tag{15.17}$$
$$s \otimes (n_1 + n_2) = s \otimes n_1 + s \otimes n_2 \tag{15.18}$$
$$sr \otimes n = s \otimes rn \tag{15.19}$$

For example, to show the first equation is true,

$$(s_1 + s_2) \otimes n = [(s_1 + s_2, n)] = (s_1 + s_2, n) + K$$

and so

$$\begin{aligned}
&[(s_1 + s_2, n) - (s_1 + s_2, n + (s_1, n) + (s_2, n))] \\
&= [(s_1, n) + (s_2, n)] \\
&= [(s_1, n)] + [(s_2, n)] \\
&= s_1 \otimes n + s_2 \otimes n
\end{aligned}$$

$$\Rightarrow (s_1 + s_2) \otimes n = s_1 \otimes n + s_2 \otimes n$$

And similarly for other 2 equalities. The trick used to figure this out was in a sense a "fudge the zero" trick. Note none of these relations are true axiomatically! They were all constructed. There isn't a ring action being defined here – these are just relations elements of the quotient group $S \otimes_R N$ satisfy. These relations will become extremely useful when we *do* show that $S \otimes_R N$ can be an S-module.

To reach that goal of defining an S-module, we have to prove that that $S \otimes_R N$ can have an S ring action. Define the ring action in the following way:

$$s \cdot \left(\sum_{i=1}^{n} s_i \otimes n_i \right) = \sum_{i=1}^{n} ss_i \otimes n_i$$

For this ring action to be well defined, we must show that it's independent of any particular representative of the coset. To do this, we'll use a nice mathematical trick. First, notice that for any $k \in K$, applying the S ring action on k will map the element to another generator of K (notice that the right and side of the next equation is another generator of K):

$$s \cdot ((s_1 + s_2, n) - (s_1, n) - (s_2, n)) = (ss_1 + ss_2, n) - (ss_1, n) - (ss_2, n)$$

Thus, K is closed under the ring action. With this knowledge, we'll be able to prove any element in $S \otimes_R N$ with the ring action applied to it is independent of representation. Let

$$\sum_{i=1}^{n} s_i \otimes n_i = \sum_{i=1}^{n} s_i' \otimes n_i'$$

be two representation of some element in $S \otimes_R N$. Then by properties of quotient groups,

$$\sum_{i=1}^{n} (s_i, n_i) - \sum_{i=1}^{n} (s_i', n_i') \in K$$

But, elements in K are closed under the S ring action, and so

$$\sum_{i=1}^{n} (ss_i, n_i) - \sum_{i=1}^{n} (ss_i', n_i') \in K$$

which by the properties of quotient groups implies that

$$\sum_{i=1}^{n} ss_i \otimes n_i = \sum_{i=1}^{n} ss_i' \otimes n_i'$$

showing that the ring action is independent of the representation of elements of $S \otimes_R N$. This statement closed under the ring action S is precisely because every element in K has the first element of the "tuple" to be an $s \in S$.

With this, checking the rest of the ring actions follows easily. For example:

$$\begin{aligned}
(s+s')(s_i \otimes n_i) &= ((s+s')s_i) \otimes n_i && \text{by definition of } S \text{ ring action} \\
&= (ss_i + s's_i) \otimes n_i && S \text{ ring property} \\
&= ss_i \otimes n_i + s's_i \otimes n_i && \text{property of } S \otimes_R N \text{ discussed earlier} \\
&= s(s_i \otimes n_i) + s'(s_i \otimes n_i) && \text{by definition of } S \text{ ring action}
\end{aligned}$$

This means that the relations shown earlier are also well-defined, producing the properties we sought after to extend the scalars! This means that we've completed our diagram from earlier:

$$\begin{array}{ccc}
R & \subseteq & S \\
N & \xrightarrow{\varphi_R} & S \otimes_R N
\end{array}$$

15.3.1 Universal property of $S \otimes_R N$

Given $S \otimes_R N$, we would how it relates to other S-modules, in relation to how N related to other S-modules. In particular, if L is any S-module, and

$$N \xrightarrow{\varphi} L$$

is any R-module homomorphism (since $R \subseteq S$, this would be well defined), then can we find an appropriate ψ such that

$$S \otimes_R N \xrightarrow{\psi} L$$

is an S-module homomorphism. That is, we want to find an appropriate S-module homomorphism from $S \otimes_R N$ to L given an R-module homomrophism from N to L. What "appropriate" here means is if we restrict ψ to $1 \otimes N$, then if N is imbedded in $S \otimes_R N$, $\psi|_R = \varphi$. TBD

In the best scenario, ψ will always exist given any N, S and L, and be unique given φ. [5]. It will turn out we can do exactly such a construction, essentially bijecting the set R-module homomorphism from N to L, to the set of S-module homomorphism from $S \otimes_R N$ to L (surjectivity comes form existence, injectivity from uniquenes, and S-module homomorphism will come because of our construction of $S \otimes_R N$). In other words, we are aiming for this diagram:

[5] Notice these requirements and the definition of initial/final objects in 1.4.4, and of the proof that $\mathbb{Z}$ is initial in **Ring** (proposition 9.1.17) are similar? This is no coincidence: $S \otimes_R N$ is the initial object in the appropriate category. More on this in ref:HERE

$$\begin{tikzcd} S \otimes_R L \arrow[r, "\psi"] & L \\ N \arrow[u, "\iota"] \arrow[ur, "\varphi"'] & \end{tikzcd}$$

where ι is an appropriate map relating N to $S \otimes_R N$ TBD

To relate N with $S \otimes_R N$, there is a natural map $\iota : N \to S \otimes_R N$ given by $n \mapsto 1 \otimes n$ (where we first map $n \in N$ to $(1, n)$, then pass it through the quotient). Since $1 \otimes rn = r \otimes n = r(1 \otimes n)$, then it is easy to see that ι is a R-module homomorphism. Since we passed it into the quotient group, this function is not injective in general. Thus, $S \otimes_R N$ need not always contain (an isomorphic copy of) N. Given this relation, between N and $S \otimes_R N$, can we find ψ? That is the context of this next theorem:

Theorem 15.3.1: Universal property of Scalar Extension

Let R be a subring of S, let N be any left R-module and let $\iota : N \to S \otimes_R N$ be the R-module homomorphism defined by $\iota(n) = 1 \otimes n$. Suppose that L is any left S-module (hence also an R-module) and that $\varphi : N \to L$ is an R-module homomorphism from N to L. Then there is a unique S-module homomorphism $\psi : S \otimes_R N \to L$ such that φ factors through ψ, i.e., $\varphi = \psi \circ \iota$ and the diagram

$$\begin{tikzcd} S \otimes_R N \arrow[r, "\Phi"] & L \\ N \arrow[u, "\iota"] \arrow[ur, "\varphi"'] & \end{tikzcd}$$

commutes. Conversly, if $\psi : S \otimes_R N \to L$ is an S-module homomorphism then $\varphi = \psi \circ \iota$ is an R-module homomorphism from N to L. In other words, the above diagram commutes.

Notice the key words: *any* left R-module N, *any* S-module L, *unique* S-module homomorphism where φ factors. We're essentially stating there is a way to *universally* find a ψ that will work for any N and φ and φ!

Proof **:**
Since our construction being based on free modules, this proof heavily relies on the universal property of free-modules. Suppose that $\varphi : N \to L$ is an R-module homomorphism to the S-module L. Consider the product $S \times N$ as a set. Then by the universal property of free modules (theorem 12.3.6), there exists a $\mathbb{Z}$-module homomorphism ψ from the free $\mathbb{Z}$-module $F(S \times N)$ to L that sends each generator of (s, n) to $s\varphi(n)$: $\psi((s, n)) = s\varphi(n)$

$$\begin{tikzcd} F(S \times N) \arrow[r, "\psi"] & L \\ S \times N \arrow[u, "\iota"] \arrow[ur, "f"'] & \end{tikzcd}$$

Since φ is an R-module homomorphism, the generators of the subgroup of K of $F(S \times N)$ defined in our construction of the tensor-product will map to 0! That is, we can define $\Phi : S \otimes_R N \to L$,

by $\Psi([(s,n)]) = s\varphi(n)$, which is well defined since if $(s,n) \in K$,

$$\Phi : F/K = S \otimes_R N \to L \qquad \Phi([(s,n)]) = 0$$

We need to show that Φ is the unique S-module homomorphism that we get that extends φ.

To show it's an s module, let $s' \in S$. Then

$$s'\Phi(s \otimes n) = s'(s\varphi(n)) = (s's)\varphi(n) = \Phi((s's) \otimes n) = \Phi(s'(s \otimes n))$$

for any $s' \in S$. Furthermore, Φ is already additive by it's construction through φ. Therefore, Φ is indeed an S-module homomorphism, showing that such a function exists!

To show uniquness, note that the module $S \otimes_R N$ is generated as an S-module by the elements of the form $1 \otimes n$ (since $s \cdot 1 \otimes n = s \otimes n$, and so we can have the left component as 1). Then, we have that $\Phi(1 \otimes n) = \varphi(n)$, meaning the value of Φ is determined from φ. But that means that Φ is a direct consequence of how φ works by definition – changing Φ would mean it no longer coresponds to φ. Therefore, it must be unique.

From this, it is immediate that if $\Phi : S \otimes_R L \to L$ is an S-module homomorphism, then

$$\varphi(n) = \Phi \circ \iota(n) = \Phi(\iota(n)) = \Phi(1 \otimes n) = \varphi(n)$$

showing the composition works too!

The key take-away from this theorem is that analyzing the module $S \otimes_R N$ can always be done through the original module N!

An interesting question to ask is will N map injectively into it's S-module extension:

Corollary 15.3.2: When does ι map invectively

Let $\iota : N \to S \otimes_R N$ be an R-module homomorphism from the previous theorem. Then $N/\ker\iota$ is the unique largest quoitent of N that can be embedded in any S-module. In particular, N can be embedded as an R-submodule of some left S-module if and only if ι is injective (in which case N is isomorphic to the R-submodule $\iota(N)$ of the S-module $S \otimes_R N$).

Proof. HERE □

Example 15.3.1

1. For any ring R, and any left R-module, we have $\mathbb{R} \otimes_R N \cong N$. This follows by taking the identity map.
2. If we have $a \otimes_R 0 \in A \otimes_R B$, then $a \otimes_R 0 = 0$, since

$$a \otimes_R 0 = a \otimes_R (0 + 0) = a \otimes_R 0 + a \otimes_R 0 \Rightarrow a \otimes_R 0 = 0$$

   same holds for $0 \otimes_R a$.
3. Let $R = \mathbb{Z}$ and $S = \mathbb{Q}$, and let A be any finite abliean group of order n, and consider $\mathbb{Q} \otimes_{\mathbb{Z}} A$.

Then given any simple tensor $q \otimes a$

$$q \otimes a = \frac{q}{n} n \otimes a = \frac{q}{n} \otimes na = \frac{q}{n} \otimes 0 = 0$$

and since this is true for any arbitrary simple tensor, it's true for arbitrary tensor, and so

$$\mathbb{Q} \otimes_{\mathbb{Z}} A = 0$$

We can also see that the map $\iota : A \to \mathbb{Q} \otimes_{\mathbb{Z}} A$ is the zero map. (words here)

4. Let N be a fintely generated free R-module, so that $N \cong R^n$. Then (words here)

$$S \otimes_R R^n \cong S^n$$

5. In the special case where N is a vector-space, we get SAME THING
6. maybe this one? (rep theory build-up)

Example 15.3.2: Exercies
btw re: induced representations the idea is ... Well, think of Z_3, which sits as a subgroup of S_3 of index 2. So if you have a 10-dimensional representation of Z_3, you can build a 20-dimensional representation of S_3 using tensor products So let Z_3 act on a 10-dimensional vector space V over a field K. This makes V into a module over the "group ring" $K[Z3] = K[x]/(1 - x^3)$. But since Z_3 is contained in S_3, this makes $K[S_3]$ into a module over $K[Z_3]$. So now you have two modules over $K[Z_3]$: namely V and $K[S_3]$. Take their tensor product. Ta-da! induced representation. Ok just wanted that to be written down somewhere.

15.4 Tensor Algebras and Graded Algebras

For the rest of this section, let R be any commutative, and the left/right actions of R for each R-module are the same. We're in general interested in the special case when $R = F$ is a field, but the basic construction holds in general. Recall how it was briefly mentioned how tensor products are in a sense "multiplying", or forming the "product" of two vectors $m_1 \otimes m_2$, where the value is not in M but in $M \otimes M$. For this to actually from an algebraic structure, this must be "closed" under multiplication, that is elements of the form:

$$m_1 \otimes m_2 \otimes m_3 \qquad m_1 \otimes m_2 \otimes m_3 \otimes m_4,$$

and all finite sums of such products should be defined. Such an algebraic structure based on M would in a sense "add" the multiplying operation, making our abelian group M into a ring, and thus making our R-module M into an R-algebra $\mathcal{T}(M)$, where $\mathcal{T}(M)$ is the structure we're going to rigorously define that makes this multiplication operation well-defined in a way that will be "universal" with respect to the rings containing M (that is, with respect to the homomorphic images of M)

For each integer $k \geq 1$, define

$$\mathcal{T}^k(M) = \underbrace{M \otimes_R M \otimes_R \cdots \otimes_R M}_{k \text{ times}}$$

and set $\mathcal{T}^0(M) = R$. The elements of $T^k(M)$ are called k-tensors . Now define

$$\mathcal{T}(M) = R \oplus \mathcal{T}^1(M) \oplus \mathcal{T}^2(M) \oplus \cdots = \bigoplus_{k=0}^{\infty} T^k(M)$$

Thus, every element of $\mathcal{T}(M)$ is a finite linear combination of k-tensors for various $k \geq 0$. We identify M with $\mathcal{T}^1(M)$ (since $\mathcal{T}^1(M) = M$), so that M is an R-submodule of $\mathcal{T}(M)$. This way, there is a way in which it makes sense to say that $\mathcal{T}(M)$ "extends" M. We'll show that $\mathcal{T}(M)$ does indeed "extend" M and add multiplication structure on the elements:

Theorem 15.4.1: $\mathcal{T}(M)$ is an R-algebra

If M is any R-module over a commutative ring R, then

1. $\mathcal{T}(M)$ is an R-algebra containing M with multiplication defined by the map

$$(m_1 \otimes \cdots \otimes m_i)(m_1' \otimes \cdots \otimes m_j') = m_1 \otimes \cdots \otimes m_i \otimes m_1' \otimes \cdots \otimes m_j'$$

and extended to sums via the distributive law. With respect to this multiplication

$$\mathcal{T}^i(M)\mathcal{T}^j(M) \subseteq \mathcal{T}^{i+j}(M)$$

2. (*universal property of Free Algebras*) If A is any R-algebra and $\varphi : M \to A$ is an R-module homomorphism, then there is a unique R-algebra homomorphism $\Phi : \mathcal{T}(M) \to A$ such that $\Phi|_M = \varphi$:

$$\begin{tikzcd} \mathcal{T}(M) \arrow[r, "\Phi"] & A \\ M \arrow[u, "\iota"] \arrow[ur, "\varphi"'] & \end{tikzcd}$$

The functor $M \mapsto \mathcal{T}(M)$ sending R-modules to R-algebras is left-adjoint to the forgetful functor from R-algebras to R-modules.

Proof **:**
The proof for 1 follows easily through computation, and the proof for (2) follows from the universal property of algebras with some additional computational checks.

The functor $\mathcal{T}(-)$ shall induce from φ for each $\mathcal{T}^k$ the map:

$$T^k(\varphi) : m_1 \otimes \cdots \otimes m_k \mapsto \varphi(m_1) \otimes \cdots \otimes \varphi(m_k)$$

Note that the multiplication of the tensor algebra is *not* the same as that defined in ref:HERE that is $(a \otimes b)(a' \otimes b') = aa' \otimes bb'$, since M is not an algebra.

Definition 15.4.2: Tensor Algebra

Let M be an R-module. Then the R-algebra $\mathcal{T}(M)$ is called the *tensor algebra* of M

Note that a tensor algebra is generated by an R-module M with $\mathcal{T}^0(M) = R$ and $\mathcal{T}^1(M) = M$. Similarly, every R-module M has a unique R-algebra $\mathcal{T}(M)$. Note also that in most cases, $\mathcal{T}(M)$ is infinite dimensional.

Proposition 15.4.3: $\mathcal{T}^k(V)$ basis and dimension

If we take $M = R^n$ a finite rank free R-module with basis $\mathcal{B} = \{v_1, ..., v_n\}$, Then the R-tensors

$$v_{i_1} \otimes v_{i_2} \otimes \cdots \otimes v_{i_k} \qquad \text{with } v_{i_j} \in \mathcal{B}$$

are minimal generators of $\mathcal{T}^k(M)$ over k. In particular $\dim_k(\mathcal{T}^k(V)) = n^k$

***Proof* :**
This follows immediately from ref:HERE

With the last theorem and proposition, we can see that $\mathcal{T}(V)$ may be regarded as the *noncommutative polynomial algebra over* R in the (noncommuting) variables $v_1, ..., v_n$.

Example 15.4.1: Tensor Algebras

1. Let $R = \mathbb{Z}$ and $M = \mathbb{Q}/\mathbb{Z}$ Then we've shown that $(\mathbb{Q}/\mathbb{Z}) \otimes_R (\mathbb{Q}/\mathbb{Z}) = 0$. Thus $\mathcal{T}(\mathbb{Q}/\mathbb{Z}) = \mathbb{Z} \oplus (\mathbb{Q}/\mathbb{Z})$. Addition is pointwise and

$$(r, \overline{p})(s, \overline{q}) = (r, \overline{p}, s, \overline{q}) = (rs, \overline{p}, 0, \overline{q})$$

   The ring $(\mathbb{Z} + \mathbb{Q}[x])/(x)$ is isomorphic to $\mathcal{T}(\mathbb{Q}/\mathbb{Z})$.

2. Let $R = \mathbb{Z}$ and $M = \mathbb{Z}/n\mathbb{Z}$. Then $(\mathbb{Z}/n\mathbb{Z}) \otimes_{\mathbb{Z}} (\mathbb{Z}/n\mathbb{Z}) = \mathbb{Z}/n\mathbb{Z}$ (shown in an excerice). Thus, $\mathcal{T}^i(M) \cong M$ for all $i > 0$ and so

$$\mathcal{T}(\mathbb{Z}/n\mathbb{Z}) = \mathbb{Z} \oplus (\mathbb{Z}/n\mathbb{Z}) \oplus (\mathbb{Z}/n\mathbb{Z}) \oplus \cdots$$

   Show that $\mathcal{T}(\mathbb{Z}/n\mathbb{Z}) \cong \mathbb{Z}[x]/(nx)$.

3. What is $\mathcal{T}(\mathbb{Q})$ where $\mathbb{Q}$ is a module over itself? By the remark above this example, this is not a polynomial ring for the "indeterminates" don't commute

The reader may have noticed that each tensor algebra has a structure similar to that of polynomial rings, namely that there is a similar notion to "degree" or "grading". Making this idea into a definition shall be very fruitful, for there shall be quite a few rings that will share this property that is not necessarily a polynomial ring:

Definition 15.4.4: Graded Ring/Algebra

Let R be a commutative ring.

1. A ring R is called a *graded ring* if it is the direct sum of additive subgroups: $R = R_0 \oplus R_1 \oplus R_2 \oplus \cdots$ such that $R_i R_j \subseteq R_{i+j}$ for all $i, j \geq 0$. The elements of R_k are said to be *homogeneous of degree* k and R_i is called the *homogeneous component of* R *of degree* k.

2. An ideal I of a graded ring R is called a *graded ideal* or *homogeneous ideal* if $I = \bigoplus_{k=0}^{\infty}(I \cap R_k)$.

3. A *graded-ring homomorphism* $\varphi : R \to S$ between two graded rings is a ring-homomorphism that conserves the grading structure on R and S, i.e., if $\varphi(R_k) \subseteq S_k$ for $k \in \mathbb{N}$.

4. a *graded module* $M = \bigoplus_i M_i$ is a module over a graded ring $R = \oplus_j R_j$ such that

$$R_j M_i \subseteq M_{i+j}$$

and a *graded algebra* is a graded ring R over a graded ring S.

Note that since $S_0 S_0 \subseteq S_0$, then S_0 is closed under multiplication, and since a graded ring is a ring, then the ring axioms also hold in S_0, thus S_0 is a sub-ring of S, and thus as discussed in the first section, this makes S into a S_0-module. Furthermore, if S_0 is contained in the center of S and contains the identity of S, then S is an S_0-algebra. Note also that it is not necessarily the case that $S_i \subseteq S_j$, $i < j$. Take $R[x]$. then $S_k = (x^i)$. Then $(x)(x) \subseteq (x^2)$ but clearly $(x) \not\subseteq (x^2)$. However, it could also be that $S_k = \prod_i^k (x^i)$ so that $S_i \subseteq S_j$, $i < j$, since that doesn't break the inclusion rule of $S_i S_j \subseteq S_{i+j}$.

Example 15.4.2: Graded Rings and Algebras

1. For any ring R, $R[x]$ is is naturally a garded algebra since:

$$R[x] \cong R \oplus (x) \oplus (x)^2 \oplus \cdots$$

More generally, $R[x_1, x_2, ..., x_n]$ is graded since:

$$R[x_1, x_2, ..., x_n] \cong R \oplus \langle x_1, x_2, ..., x_n \rangle \oplus \langle x_1^2, x_2^2, ..., x_n^2 \rangle \oplus \cdots$$

Show that any ideal of I is graded if and only if it is generated by homogeneous polynomials (polynomials for which each monomial adds up to the same degree). Use this to show that not every ideal is graded (ex. $(1 + x) \subseteq \mathbb{Z}[x]$)

2. For any ideal $I \subseteq R$, we may define the R-module:

$$\bigoplus_{i \in \mathbb{N}} I^l = R \oplus I \oplus I^2 \oplus \cdots$$

which naturally has $I^i I^j \subseteq I^{ij}$. This algebra is called the *Rees* Algebra of R with respect to I and is denoted $\text{Rees}_R(I)$[a].

[a] Rees Algebras correspond to the notion of *Blow ups* in Algebraic Geometry

Graded ideas are distinguished since they give rise to new graded structures via quotienting:

Proposition 15.4.5: Quotient of Graded Ring by Graded Ideals

Let S be a graded ring, let I be a graded ideal in S and let $I_i = I \cap S_k$ for all $k \geq 0$. Then S/I is naturally a graded ring whose homogeneous component of degree K is isomorphic to S_k/I_k. Hence if $\varphi_i : S_i \to T_i$ is the restriction of φ to S_i (which is well-defined since φ is graded), then $\ker \varphi = \oplus_i \ker \varphi_i$.

***Proof* :**
This is very similar to other such proofs. The map

$$\bigoplus_{k=0}^{\infty} S_k \to \bigoplus_{k=0}^{\infty} (S_k/I_k)$$

$$(..., s_k, ...) \mapsto (..., s_k \text{ mod } I_k, ...)$$

is surjective with kernel $I = \bigoplus_{k=0}^{\infty} I_k$ and defines an isomorphism of graded rings.

It is often useful to be able to find graded ideals. Recall a multi-variate polynomial is homogeneous if the sum of the degree's of each of it's monomial add up to the same number (ex. $x^2 + xy + y^2$). More generally, an element $r \in R$ of a graded ring R is called *homogeneous* [of degree i] are the nonzero elements of S which are given by the natural inclusion map $S_i \hookrightarrow S$. Then the product of two homogeneous elements is homogeneous (i.e., if r_i and r_j are of degree i, j respectively, then $r_i r_j = r_{i+j}$ for some element of degree $i + j$, given it is not zero).

Proposition 15.4.6: Characterizing Graded Ideals

Let I be an ideal of the graded ring $S = \bigoplus_i S_i$. Then the following are equivalent:

1. I is a graded ideal
2. if $s \in S$ and $s = \sum_i s_i$ is the decomposition of s int homogeneous elements $s_i \in S_i$, then $s \in I \Leftrightarrow s_i \in i$ for all i
3. I admits a generating set consisting of homogeneous elements
4. I is the kernel of a graded homomorphism

***Proof* :**
exercise

Using this, it can be shown that $(y - x^2)$ is not graded, while $(y, y - x^2)$ is graded for $k[x, y]$.

The next two examples of graded algebras are very important and shall represent two of the most common types of algebra's the reader will encounter[6]. We'll construct the next two algebras "globally"

[6]They shall show up extensively in differential geometry

and "locally": it will be shown that they're quotient of the tensor algebra, and also show that they are a special while the sub-structure construction giving local structure.

Proposition 15.4.7: Arithmetic of Homogeneous Ideals

Let $I, J \subseteq R$ be homogeneous ideals. Then:

1. $I + J$ is a homogeneous ideal
2. IJ is a homogeneous ideal
3. $I \cap J$ is a homogeneous ideal
4. $\sqrt{I} = \left\{ r \in R \mid r^k \in K \right\}$ is a homogeneous ideal (called the radical of I[a]

[a] This shall be in chapter 21

Proof **:**
exercise

15.4.1 Symmetric and Exterior Algebras

Symmetric algebras are the natural abelianization of tensor algebras (recall section 5.2).

Definition 15.4.8: Symmetric Algebra

the *symmetric algebra* of an R-module M is the R-algebra obtained by taking the quotient of the tensor algebra $\mathcal{T}(M)$ by the ideal $\mathcal{C}(M)$ generated by all elements of the form $m_1 \otimes m_2 - m_2 \otimes m_1$, for all $m_1, m_2 \in M$. The symmetric algelbra $\mathcal{T}(M)/\mathcal{C}(M)$ is denoted $\mathcal{S}(M)$.

$$\mathcal{S}(M) = \frac{\mathcal{T}(M)}{\mathcal{C}(M)} \qquad \mathcal{C}(M) = \{m_1 \otimes m_2 - m_2 \otimes m_2 \mid m_1, m_2 \in M\}$$

By construction, $\mathcal{S}(M)$ is a commutative ring. It is also easy to show that $\mathcal{C}(M)$ is a graded ideal since the ideal is generated by homogeneous tensors, hence we have $\mathcal{S}^k(M) = \mathcal{T}^k(M)/\mathcal{C}^k(M)$. Since $\mathcal{C}(M) \cap M = 0$, M can be identified with $\mathcal{S}^1(M) = M$ in $\mathcal{S}(M)$, hence this is indeed a [commutative] extension of M. Similarly, $\mathcal{S}^0(M) = R$. For the other $\mathcal{S}^k(M)$:

Theorem 15.4.9: Properties of Symmetric Algebras

Let M be an R-module over the commutative ring R and let $\mathcal{S}(M)$ be its symmetric algebra

1. The R-algebra $\mathcal{S}^k(M)$ is isomorphic to:

$$\frac{\mathcal{T}^k(M)}{\langle (m_1 \otimes m_2 \otimes \cdots \otimes m_k) - (m_{\sigma(1)} \otimes m_{\sigma(2)} \otimes \cdots \otimes m_{\sigma(k)}) \mid m_i \in M,\ \sigma \in S_k \rangle}$$

2. The R-algebra $\mathcal{S}^k(M)$ satisfies the universal property of *symmetric k-multilinear maps*, that is any symmetric k-multilinear map to a module N uniquely factors through a map from $\mathcal{S}^k(M)$ to N.

3. The R-algebra $\mathcal{S}^k(M)$ satisfies the universal property of *commutative R-algebras*, that is if A is a *commutative* R-algebra and $\varphi : M \to A$ is a R-module homomorphism, then there is a unique lift to an R-algebra homomorphism $\Phi : \mathcal{S}(M) \to A$

4. The map $\mathcal{S}^k(-)$ is functorial

Note that symmetric maps may be characterized by maps $\varphi : M_1 \times M_2 \times \cdots \times M_k \to N$ satisfying the condition:

$$\varphi(m_{\sigma(1)}, m_{\sigma(2)}, ..., m_{\sigma(k)}) = \varphi(m_1, m_2, ..., m_k)$$

***Proof* :**

Part (2) and (3) should be verified by the reader. For part (1), notice that the k-tensors in $\mathcal{C}^k(M)$ in the ideal $\mathcal{C}(M)$ are finite sums of tensors of the form:

$$m_1 \otimes \cdots \otimes m_{i-1} \otimes (m_i \otimes m_{i+1} - m_{i+1} \otimes m_{i+2}) \otimes m_{i+3} \otimes \cdots \otimes m_k$$

which corresponds to the transposition $(i\ i+1)$ in S_n. Conversely, recall that S_n is generated by elements of the form $(i\ i+1)$.

For (4), note that the map $\mathcal{T}^k(\varphi)$ his maps generators of each of the homogeneous components of $\mathcal{C}(M)$ to themselves. Thus, φ induces a R-module Homomorphisms on the quotients:

$$\mathcal{S}^k(\varphi) : \mathcal{S}^k(M) \to \mathcal{S}^k(N) \qquad \text{and} \qquad \bigwedge^k(\varphi) : \bigwedge^k(M) \to \bigwedge^k(N)$$

and, each of these maps is a ring homomorphism (and hence they're also R-algebra homomrophisms)

Corollary 15.4.10: Dimension of $\mathcal{S}(V)$

$$\dim_F(\mathcal{S}^k(V)) = \binom{k+n-1}{n-1}$$

***Proof* :**
exercise

The next example of may at first seem less natural than the tensor algebra and symmetric algebra, however it is in many ways more important than those two graded algebras. The exterior algebra is an algebra that very naturally lends itself to geometry, where elements of an exterior algebra give a notion of *orientable volume* (orientable in this context means there is "chirality" of left/right-handedness). If the reader has done some differential geometry, or intends to do some differential geometry, then *differential forms*[7] is an exterior algebra.

Definition 15.4.11: Exterior Algebras

the *exterior albegra* of an R-module M is the R-algebra obtained by taking the quotient of the tensor algebra $\mathcal{T}(M)$ by the ideal $\mathcal{A}(M)$ generated by all elements of the form $m \otimes m$ for $m \in M$:

$$\bigwedge(M) = \frac{\mathcal{T}(M)}{\mathcal{A}(M)}$$

The exterior algebra $\mathcal{T}(M)/\mathcal{A}(M)$ is denoted by $\bigwedge(M)$ and the image of $m_1 \otimes \cdots \otimes m_k$ in $\bigwedge(M)$ is denoted by $m_1 \wedge \cdots \wedge m_k$. The $\wedge$ is called the *wedge product*.

As before, $\mathcal{A}(M)$ is generated by homogeneous elements, and hence is a graded ideal. Thus $\bigwedge^k(M) = \mathcal{T}^k(M)/\mathcal{A}^k(M)$. We can also identity R and M in the same way as last time. The R-module $\bigwedge^k(M)$ is called the kth *exterior power of* M. The multiplication of two elements in the exterior algebra is called the *wedge* (or *exterior*) *product*.

The exterior algebra does satisfy a universal property. The exterior algebra is the representation of a type of map called *alternating*. A map:

$$\varphi : M_1 \times M_2 \times \cdots \times M_k \to N$$

is called *alternating* if

$$\varphi(m_1, m_2, ..., m_k) = 0 \qquad \text{whenever} \qquad m_i = m_j \text{ for some } i \neq j$$

A common alternative definition of an alternating map is a map $\varphi : M_1 \times M_2 \times \cdots \times M_k \to N$ satisfying

$$\varphi(m_{\sigma(1)}, m_{\sigma(2)}, ..., m_{\sigma(k)}) = (-1)^\sigma \varphi(m_1, m_2, ..., m_k)$$

where $(-1)^\sigma$ represents the parity of σ (whether it is even or odd). However, this definition breaks down over rings of characteristic 2:

Lemma 15.4.12: Equivalence Of Alternating Map

Let $\varphi : M^k \to N$ be a multilinear map. If φ is alternating, then

$$\varphi(m_{\sigma(1)}, m_{\sigma(2)}, ..., m_{\sigma(k)}) = (-1)^\sigma \varphi(m_1, m_2, ..., m_k)$$

If $2 \in R$ is a unit, then the converse holds true as well.

[7] the fundamental algebraic object giving information about the (oriented) volume of a manifold at each point that is sufficient to define integration on manifolds

***Proof* :**
It suffices to show that interchanging any two factors switches the sign. Since switching two factors will leave the rest of the function invariant, it suffices to reduce to the case where $k = 2$. Then:

$$\begin{aligned} 0 &= (m + m') \wedge (m + m') \\ &= (m \wedge m) + (m \wedge m') + (m' \wedge m) + (m' \wedge m') \\ &= (m \wedge m') + (m' \wedge m) \\ &= (m \wedge m') + (m' \wedge m) \end{aligned}$$

implying $(m \wedge m') = -(m' \wedge m)$.
For the converse, if 2 is a unit, then again reducing to the $k = 2$ case:

$$\varphi(m_1, m_2) = -\varphi(m_2, m_1) \quad \Rightarrow \quad \varphi(m, m) = -\varphi(m, m) \quad \Rightarrow \quad 2\varphi(m, m) = 0$$

Since 2 is a unit,

$$\varphi(m, m) = 0$$

giving the desired result.

Using this, we can show the wedge product is *anti-commutative* on M:

$$m \wedge m' = -m' \wedge m$$

for every $m, m' \in M$. Note that this is not true in general for $a, b \in \bigwedge(M)$, that is

$$ab = -ba$$

is not always true

Proposition 15.4.13: Properties of Exterior Algebras

Let M be an R-module over the commutative ring R and let $\bigwedge(M)$ be its exterior algebra.

1. Th k^{th} exterior power, $\bigwedge^k(M)$, of M is equal to $M \otimes \cdots \otimes M$ (k factors) modulo the submodule generated by all elements of the form

$$m_1 \otimes m_2 \otimes \cdots \otimes m_k \qquad \text{where } m_i = m_j \text{ for some } i \neq j$$

   In particular,

$$m_1 \wedge \cdots \wedge m_k = 0$$

   if $m_i \; n_j$ for some $i \neq j$

2. (universal property for alternating multilinear maps) If $\varphi : M \times \cdots \times M \to N$ is an alternating k-multilinear map then there is a unique R-module homomorphism $\Phi : \bigwedge^k(M) \to N$ such that $\varphi = \Phi \circ \iota$ where

$$\iota : M \times \cdots \times M \to \bigwedge^k(M)$$

   is the map defined by

$$\iota(m_1, ..., m_k) = m_1 \wedge \cdots \wedge m_k$$

3. the operation $\wedge(-)$ is functorial

Proof **:**
These proofs are similar to those for the symmetric algebra and is left as an exercise (see exercise ref:HERE)

Example 15.4.3: Exterior Algebras

1. If V is a 1 dimensional vector space over k with basis element v, then $\bigwedge^k(V)$ consists of finite sums of elements of the form $a_1 v \wedge \cdots \wedge a_k v$, which will be $a_1 \cdots a_k (v \wedge \cdots \wedge v)$, and since $v \wedge v = 0$, it follows that

$$\bigwedge^0(V) = F$$
$$\bigwedge^1(V) = V$$
$$\bigwedge^2(V) = 0 \qquad\qquad i \geq 2$$

   Thus we know that for 1 dimensional vector space, we'll always have

$$\bigwedge(V) = F \oplus V \oplus 0 \oplus \cdots$$

2. Let's not take V to be a 2 dimensional vector space over k. Thus, any element in $\bigwedge^k(V)$ is

of the form

$$(a_1 v + a_1' v') \wedge \cdots \wedge (a_k v + a_k' v')$$

We already know what happens for the 0 and 1 valued wedge products. for the 2-valued wedge products, we get

$$\begin{aligned}(av + bv') \wedge (cv + dv') &= ac(v \wedge v) + ad(v \wedge v') + bc(v' \wedge v) + bd(v' \wedge v') \\ &= (ad - bc) v \wedge v'\end{aligned}$$

and for further dimensions, we'd get $\bigwedge^i(V) = 0$, since there will be some wedging of either v or v' with itself. Thus, it follows that

$$\bigwedge(V) = k \oplus V \oplus k(v \wedge v') \oplus 0 \oplus \cdots$$

More generally, given an n-dimensional vector space with basis $\{e_1, ..., e_n\}$, we get that the basis for $\bigwedge^k(V)$ is:

$$e_{i_1} \wedge \cdots \wedge e_{i_k} \qquad 1 \leq i_1 \leq \cdots \leq i_k$$

where $k \leq n$ and $\bigwedge^K(V) = 0$ for $K > n$. As a result, $\bigwedge(V)$ can be thought of as the collection of all oriented parallelotopes. The following is a visual in the first three dimensions

HERE

3. In the special case of $\dim_k(V) = 3$, the wedge product $v \wedge w$ is called the *cross product* due its historic significance in physics. If $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is a basis for V, then:

$$v \wedge w = (v_1 w_2 - v_2 w_1)(\mathbf{e}_1 \wedge \mathbf{e}_2) + (v_2 w_3 - v_3 w_2)(\mathbf{e}_2 \wedge \mathbf{e}_3) + (v_1 w_3 - v_3 w_1)(\mathbf{e}_1 \wedge \mathbf{e}_3)$$

where the basis $(\mathbf{e}_1 \wedge \mathbf{e}_2), (\mathbf{e}_2 \wedge \mathbf{e}_3)(\mathbf{e}_1 \wedge \mathbf{e}_3)$ can be identified with the basis for V to think of this result as giving back a vector in V. In physics, this is denoted $v \times w$. The fact that the dimension line up in this way is special, which is how come this particular wedge product has been singled out by physics.

4. If the reader knows some differential geometry, then the differential form vector-bundle where each vector-space is an exterior algebra, namely if M is a manifold then:

$$\bigwedge_{\mathcal{O}_M} \Omega^1_M$$

is the differential form with exterior algebras at each point.

5. The inclusion map $\varphi : I \hookrightarrow R$ of the ideal (x, y) into the ring $R = \mathbb{Z}[x, y]$, both considered as R-modules, induce the map

$$\bigwedge^2(\varphi) : \bigwedge^2(I) \to \bigwedge^2(R)$$

but since $\bigwedge^2(R) = 0$ and $\bigwedge^2(I) \neq 0$, the map cannot be injective.

By the previous example, we see that if M is a finitely generaed module, then $\bigwedge M$ is a finitely

generated module. If M is a finite dimensional vector-space, then it's dimension can easily be counted:

Corollary 15.4.14: dimension of $\bigwedge(V)$

Let M be a finitely generated module. Then $\bigwedge M$ is finitely generated. If M is a finitely dimensional vector-space, then : $\dim_F\left(\bigwedge^k(V)\right) = \binom{n}{k}$.

***Proof* :**
If M is finitely generated, then for $N > n$ we see that the wedging of elements shall have at least one repeated generator, and hence

$$\bigwedge^N M = 0 \qquad M > n$$

If M is a vector-space, the key is to generalize the above example, namely it comes from counting sequences $1 \leq i_1 \leq \cdots i_k \leq n$. Let $\{e_1, ..., e_n\}$ be a basis for M and define the map:

$$\varphi_I : M^k \to R$$

where $I = (e_{i_1}, ..., e_{i_k})$ is a k-tuple and

$$\varphi_I(e_{j_1}, ..., e_{j,k}) = \begin{cases} (-1)^\sigma & \text{if } \exists \text{ permutation } \sigma \text{ s.t. } \sigma(j_k) = i_k \text{ for all } k \\ 0 & \text{otherwise} \end{cases}$$

and then extending multi-linearly. Since permutations are bijective, if such a σ exists it is unique, and since M is free each element $m \in M$ is expressed as a unique linear combination of e_i's, and hence φ_I is well-defined. Furthermore, φ_I is alternating, and hence factors through:

$$\overline{\varphi_I} : \bigwedge^k M \to R$$

This map is certainly surjective. For injectivity, if $n = k$ then $e_1 \wedge \cdots \wedge e_k$ is linearly independent since it is not a torsion element (as the reader should verify). If $n \geq 2$ assume:

$$m = \sum_{1 \leq i_1 < \cdots < i_k \leq n} a_{i_1, ..., i_k} e_{i_1} \wedge \cdots \wedge e_{i_k} = 0$$

By a counting argument, it can be deduced that there is an even number of elements in the linear combination. Hence for any specific $I = (i_1, ..., i_k)$:

$$a_I = \overline{\varphi_I}\left(\sum_{1 \leq i_1 < \cdots < i_k \leq n} a_{i_1, ..., i_k} e_{i_1} \wedge \cdots \wedge e_{i_k}\right) = \overline{\varphi_I}(0) = 0$$

giving the linear independence of the elements, as we sought to show.

Using this result, we see that if M is a free R-module of rank n, then

$$\bigwedge^n M \cong R$$

with the isomorphism:

$$\varphi_{1\ldots k}(e_{i_1}, \ldots e_{i_k}) = \begin{cases} \pm 1 & \text{the indices } i_1, \ldots, i_k \text{ are distinct} \\ 0 & \text{otherwise} \end{cases}$$

where the sign is determined by the sign of the given permutation of the bass elements. The simple alternating maps φ_I are also useful when characterzing extension of endomorphisms $\varphi : V \to V$

Corollary 15.4.15: Simple Alternating Maps Values

Let V be a finitely dimensional F-vector space of dimension n, φ an endomorphism, and A a matrix representation of φ. If $a_i \in F^k$ represents the columns of A, then:

$$\varphi_{1,\ldots,k}(a_1, \ldots, a_k) = \det(A)$$

In other words:

$$\bigwedge^n (\varphi)(w) = \det(A)(w)$$

***Proof* :**
This follows from the properties of the determinant, namely the simple alternating maps are multilinear and alternating.

To end off this section, there are a few more interesting properties of the functors $\mathcal{S}(-)$ and $\bigwedge(-)$ that will be of particular interest for the reader who will reach part VII. In part VII, we shall be very interested in whether functors "preserve exactness", namely if

$$0 \to L \to M \to N \to 0$$

is a short exact sequence of R-modules and $\mathscr{F}$ is a [say covariant][8] functor, is:

$$0 \to \mathscr{F}(L) \to \mathscr{F}(M) \to \mathscr{F}(N) \to 0$$

exact. In other words, does the functor $\mathscr{F}$ translates the notion of building up a modules from simpler modules into the new category, or after whatever transformation $\mathscr{F}$ brought about? This is a fascinatingly deep question which leads to many amazing results, the beginnings of which shall be explored in part VII. For now, consider what happens after applying $\mathcal{S}(-)$. The reader should verify that if $M \to N \to 0$ is a surjective map, then $\mathcal{S}(M) \to \mathcal{S}(N) \to 0$ remains surjective. However, it is not necessarily the case that if $0 \to L \to M$ is injective then $0 \to \mathcal{S}(L) \to \mathcal{S}(M)$ is injective! To fix this, notice that L can be identified as a subset of M, which can be identified as a subset of $\mathcal{S}(M)$ since $M \cong \mathcal{S}^1(M)$. Hence we have the ideal $L \cdot \mathcal{S}(M) \subseteq \mathcal{S}(M)$. Then the reader should verify that:

$$0 \to L \cdot \mathcal{S}^{*-1}(M) \to \mathcal{S}^*(M) \to \mathcal{S}^*(M) \to 0$$

[8] If it is contravariant, simply consider the opposite category

is a short exact sequence, showing that $\mathcal{S}(-)$ does not preserve exactness, however with a little tweaking we can still "recover" an exact sequence.

For $\bigwedge(-)$, we get a something slightly more complicated. For the general case of a finitely generated R-module, the more general tools of homological algebra are required, so for this model case consider $M = R^n$. By definition, R is a free R-module over itself if R does not contain zero-divisors, however R/I is certainly not free (over R). We may ask if instead R/I has a *free resolution*, that is an exact sequence such that:

$$\cdots \to F_3 \to F_2 \to F_1 \to F_0 \to R/I \to 0$$

If R was a PID, then as we explored in chapter 14 such a free resolution always exists. In the more general case this is not necessarily the case. However, if $I = (a_1, ..., a_n)$ and we have the property that a_1 is not a zero divisor, a_2 is not a zero divisor $\bmod (a_1)$, a_3 is not a zero-divisor mod (a_1, a_2), and so on, then we may form a free resolution. An ideal I where the elements can be arranged into such a sequence is called a *regular ideal*, and the sequence of elements is called a *regular sequence* (these shall be very important in algebraic geometry, generalizing the notion of "independence" and (if you know differential geometry) can detect "transversal intersection"). To form a free resolution of R/I, take $F = R^n$. Then define the R-module homomorphism $d_r : \bigwedge^r(F) \to \bigwedge^{r-1}(F)$ given by:

$$d(e_{i_1} \wedge \cdots \wedge e_{i_r}) = \sum_{j=1}^{r} (-1)^{j-1} a_{i_j} e_{i_1} \wedge \cdots \wedge \widehat{e_{i_j}} \wedge \cdots \wedge e_{i_r}$$

where $\widehat{e_{i_j}}$ represents that this element is omitted in the wedge product. Then the reader should verify that:

$$d_{r-1} \circ d_r = 0 \qquad \Leftrightarrow \qquad \mathrm{im}(d_r) \subseteq \ker(d_{r-1})$$

In fact, since I is a regular ideal, we have equality in the above (which the reader is highly encouraged to prove!). If I were not regular, we still get the above equation, but we would not get equality. We thus get:

$$0 \to \bigwedge^n(F) \xrightarrow{d_n} \bigwedge^{n-1}(F) \xrightarrow{d_{n-1}} \cdots \xrightarrow{d_2} \bigwedge^1(F) \xrightarrow{d_1} \bigwedge^0(F) = R \to R/I \to 0$$

which is is exact when I is a regular ideal. The above sequence is called the *Koszul complex*. This shall be paramount in detecting intersections of polynomials (and more generally "transversal" intersections of schemes). The preliminaries of these concepts shall further be explored in chapter 28 .

15.4.2 Symmetric and Alternating Tensors

We've defined the symmetric and exterior algebras by taking quotients. We'll now see how we can sometimes do these constructions as sub-algebras of $\mathcal{T}(M)$ by showing how there are *symmetric* and *alternating* tensors, which ones identified we can use to generate the respective sub-algebras. The advantage of this is that we can how these structures are present in $\mathcal{T}(M)$ giving us more insight in the algebra of tensors. However, this could make certain computations difficult. This approach of constructing symmetric and alternating tensors is reveleant in differential geometry, and so it is worth taking a moment to look at this.

For any R-module M there is a natural left group action of the symmetric group S_n on $M \times M \times \cdots \times M$ (n factors) given by permuting the factors:

$$\sigma(m_1, ..., m_k) = (m_{\sigma^{-1}(1)}, ..., , m_{\sigma^{-1}(k)})$$

where we take σ^{-1} instead of σ is to insure this is a *left* group action. The above map is certainly R-multilinear and hence efined an R-linear left group action on $\mathcal{T}^k(M)$.

Definition 15.4.16: Symmetric and alternating n-tensors

Let M be an R-module. Then

1. An element $z \in \mathcal{T}^n(M)$ is called a *symmetric* n-tensor if $\sigma z = z$ for all $\sigma \in S_k$
2. An element $z \in \mathcal{T}^n(M)$ is called a *alternating* n-tensor if $\sigma z = \epsilon(\sigma) z$ for all $\sigma \in S_n$ where $\epsilon(\sigma)$ is the sign of σ.

Certainly the collection of all symmetric and alternating n-tensors forms a submodules of $\mathcal{T}^n(M)$. Furthermore, $\mathcal{C}^n(M)$ and $\mathcal{A}^n(M)$ are stable under S_n, and hence the action can be reduced to $\mathcal{S}^n(M)$ and $\bigwedge^n(M)$, and in particular, this action is trivial on $\mathcal{S}^n(M)$. Let's now see when the quotient modules and the collection of symmetric/alternating tensors are isomorphic. The key is that $n!$ (the number defined to be 1 added to itself $n!$ times) is a unit in R. Define the following two functions from $\mathcal{T}^n(M)$ to itself:

$$Sym(z) = \sum_{\sigma \in S_n} \sigma(z) \qquad Alt(z) = \sum_{\sigma S_n} \epsilon(\sigma)\sigma(z)$$

Then the image of z in the maps is symmetric and alternating respectively. If z is already symmetric (resp. alternating), then it image in Sym (resp. Alt) is $n!z$. Hence, if $n!$ is a unit:

$$\frac{1}{n!} Sym(z) = z \qquad \frac{1}{n!} Alt(z) = z$$

Then the maps $\frac{1}{n!} Sym$ and $\frac{1}{n!} Alt(z)$ becomes surjective R-module homomorphisms from $\mathcal{T}^n(M)$ to the submodule of symmetric (resp. alternating) tensors.

Proposition 15.4.17: Characterizing Symmetric And Alternating Tensors

Let M be an R-module $n!$ a unit. Then:

$$\frac{1}{n!} Sym : \mathcal{S}^k(M) \to \{\text{symmetric } n\text{-tensors}\} \qquad \frac{1}{n!} Alt : \bigwedge^k(M) \to \{\text{symmetric } n\text{-tensors}\}$$

are isomorphisms

***Proof* :**
exercise

The reader is invited to check that $\frac{1}{n!} Sym$ and $\frac{1}{n!} Alt$ are in fact projections, and hence they each split $\mathcal{T}^n(M)$ into a direct product of the image and kernel. This way of defining an element by its "average" shall become important when disusing *Maschke's Theorem* (theorem 24.1.8 when doing representation theory

Exercise 15.4.1

1. Let $v_1, ..., v^n \in \bigwedge^1(V)$ be linearly dependent vectors. Show that $v_1 \wedge \cdots \wedge v_n = 0$.

2. Let k be any field where $-1 \neq 1$ an let V be a finite dimensional vector space over k. Show that $V \otimes_k V \cong \mathcal{S}^2(V) \oplus \bigwedge^2(V)$.

15.5 Monoidal Categories

Monoidal Categories use the tensor notation and are interesting; should I add this here?

see: this

16

Extension Problem – Prelude to Homological Algebra

Note: This chapter is not necessary for part IV and V, and so this chapter can be comfortably skipped until part VII. However, these concepts do show up in a couple of places as "Easter eggs", and some objects of study shall be shown to be projective/injective/flat.

In this chapter, we take a closer look at the extension of R-modules. Recall that in section 3.6, we saw that a group G is an extension of a group A by B[1] if

$$G \cong_{\mathbf{Grp}} B/A$$

Similarly, an R-module M is an extension of a module A by B if

$$M \cong_R B/A$$

In this chapter, we build up the necessary tools to get an idea of all possible extensions of R-modules B by A. The result shall be a group known Ext group and is denoted $\mathrm{Ext}^1_R(M, N)$. Some parts of the proof of the existence of $\mathrm{Ext}^1_R(M, N)$ shall be relegated to part VII without any issue, the heart of the argument shall presented in this chapter.

We shall tackle this problem by first tacking some modules that have "trivial" extensions. In section 12.3, we explored when a module can be decomposed into a direct product. As with groups, the direct product of two modules is the simplest kind of extensions there is, namely if $B = M \oplus N$,

[1] For historical reasons, extensions in group theory usually switch the order of which group is extending which. If $G \cong_{\mathbf{Grp}} B/A$, then a classical group theorist would say B is an extension by A

the algebraic structure of the quotient $(M \oplus N)/N$ is "compatible" with the algebraic structure of $M \oplus N$:

$$\frac{M \oplus N}{N} \cong M \subseteq M \oplus N$$

In this chapter, we start by exploring modules P where if M is an extension of P by any N, then M must always be a direct summand of P and N:

$$M \cong P \oplus N \qquad \Rightarrow \qquad \frac{P \oplus N}{N} \cong P$$

In other words, the structure of P forces the simplest possible extension. It is worth remembering that this is not always the case.

Example 16.0.1: Trivial And Non-trivial Extensions

Take the following $\mathbb{Z}$-modules: $P = \mathbb{Z}/6\mathbb{Z}$, $M = \mathbb{Z}$ and $N = 6\mathbb{Z}$. Then $P \cong M/N$. However, it is certainly not the case that $M \cong P \oplus N$, for the right-hand side has torsion, while the left-hand side does not. On the other hand, if P is a vector-space then P certainly has this property, for if $P \cong M/N$, then picking a basis for M, we get k^n, and quotienting by some $N \subseteq M$ such that $N \cong k^m$ $(m \leq n)$:

$$k^{n-m} \cong P \cong M/N \qquad m \leq n$$

and so $M \cong k^{n-m} \oplus k^m \cong k^n$. This is a bit of a trivial example for vector spaces are (as an algebraic object) completely characterized by their dimension. Is there then some weaker property a module can satisfy that will force each extension to be a projection (i.e. if P is an extension of A by B, then A must "trivially" contribute so that $A \cong P \oplus B$)? The answer is yes, and we shall demonstrate in this chapter.

Modules P which force trivial extensions are called *projective* (for reasons that shall be made clearly shortly). As the reader may start to imagine, if an R-module P is projective, then R-module homomorphisms either to or from any module may be interesting (similarly to how linear transformations between vector spaces had a very simple structure). Since M, N need not be projective, it is interesting to ask about R-module homomorphisms to or form P which are not projective themselves. Though not as simple as vector spaces, it turns out that a projective R-module P and the structure of $\mathrm{Hom}_R(P, M)$ is intimately related. This realization is at the origins of the theory if *homological algebra*, which this chapter can be considered a prelude to. To get an idea of what homological algebra is, let's first develop an analogy.

> Let's say you have a company with multiple departments. Each department is defined by what they do: You can have a department of finance, a department of Public Relations, a department for cashiers, department of logistics and so on. Given a certain department, it might need to talk to some department, or it might not. For example, let's say the department of Public Relations needs to talk to the department of cashiers (for example, to deal with customers issues) but it does not need to talk to the department of logistics (since logistical workers will not have interactions with the public as employees of the company). Given this image, we can imagine a directed graph, where each vertex represents a department and each edge represents whether the department needs to be in touch with another department. So far, we have defined departments by what they do "internally" (ex. work with cashiers or logistical workers) and have given that they can interact with other departments. Let's now say that it is your job to create a department

> which must be able to communicate with all departments. Notice that the defining feature of this department is not something "internal", but by the fact that it relates to every other department – something "relational".

If we replace the term department with R-module and "need for interaction" with homomorphism, then we have a similar idea of how to define an R-module *homologically*[2]: We can define an R-module by how it is related to other R-modules. We can define an R-module by how it relates to only certain R-modules (we will do this extensively in this chapter) or we can do it for how it is related to all R-modules. We have actually already done this: a free-modules F has the property that for any module M, there exists a homomorphism homomorphism $F \to M$ given by where a set of generators for F map to, hence if S is a set of generators for F, $\mathrm{Hom}_{\mathbf{Set}}(S, M) \equiv \mathrm{Hom}_R(F, M)$. This property characterizes free-modules, and hence a weaker property shall characterize projective modules, as we shall explore in section 16.2

To fully build-up to the study of projective modules and the Ext group, we will require to build up our categorical tool kit a little more. We have introduced the notion of functors and categories, we shall now introduce functor that are compatible with extensions. This shall be very beneficial, for within this context we shall see that exploring projective modules is just one part in exploring two more types of module known as an *injective modules* and *flat modules*. Injective modules will be the dual of projective modules, having properties that can be thought of as being the opposite of projective modules. Flat modules shall be motivated by the fact that any currying of Hom's can be uncurried by using the tensor product:

$$\mathrm{Hom}_R(M, \mathrm{Hom}(N, L)) \cong_{\mathbf{Ab}} \mathrm{Hom}_R(M \otimes_R N, L)$$

16.1 Exact Sequences

We generlize now the results of section 3.6:

Definition 16.1.1: Exact Sequences

Let X, Y, Z be R-modules.

1. The pair of homomorphisms $X \xrightarrow{\alpha} Y \xrightarrow{\beta} Z$ is said to be *exact* (at Y) if $\alpha(X) = \ker(\beta)$ (image $\psi = \ker \varphi$)
2. A sequence $\cdots \to X_{n-1} \to X_n \to X_{n+1} \to \cdots$ of homomorphisms is said to be an *exact sequence* if it is exact at every X_i between a pair of homomorphisms.

[2]The origins of the word "homological" may also bring some clarity to this analogy. the morpheme "hom" comes from the word "homomorphism". Thus, "homological" can be thought of as "homomorphism-like" (the same way external can be though of as exterior-like). When we study homological algebra, we are thus studying R-modules through how they interact with homomorphisms.

Proposition 16.1.2: Injective and Surjective maps in Exact Sequences

Let A, B and C be R-modules over some ring R. Then

1. The sequence $0 \to A \xrightarrow{\psi} B$ is exact (at A) if and only if ψ is injective
2. The sequence $B \xrightarrow{\varphi} C \to 0$ is exact (at C) if and only if φ is surjective.

Proof **:**

1. Take α to be the R-module from 0 to A. Then since it's a R-module, it can only be $0 \mapsto 0$. Thus $0 = \text{image } \alpha = \ker\psi$, and so ψ must be injective.
2. Similarly, take β to be the R-module from $C \to 0$. Then it can only be that $\beta(C) = 0$, and so
$$\text{image}(\varphi) = \ker(\beta) = C$$
which makes it clear that φ is surjective.

Example 16.1.1: Exact Sequences

Recall the automorphism group $\text{Aut}(G)$. We mentioned in section 4.5.1 that the center of a group G and the outer-automorphisms $\text{Out}(G)$ are a sort of "dual" of each other. This can be more easily visuallized using exact sequences. Define the map $\sigma : G \to \text{Aut}(G)$ mapping $\sigma(g) = (h \mapsto hgh^{-1})$. Then we have: with $Z(G)$ on one end and $\text{Out}(G)$ on the other! We can see that $f : G/Z(G) \to \text{Inn}(G)$ is an isomorphism, and that they are on the opposite ends of this sequence!

Using the results of proposition 16.1.2, we can find the "shortest" path from 0 to 0 that is exact. If we take $0 \to A \to 0$, then 0 would have to be the zero module. If we take:

$$0 \to A \xrightarrow{f} B \to 0$$

then f would have to be both injective and surjective, i.e. always bijective, and hence A and B are isomorhpisms. Thus, the shortest sequence of exact moduels that produce more intersting results must be of length 3:

Corollary 16.1.3: Short Exact Sequence

The sequence $0 \xrightarrow{\psi} B \xrightarrow{\varphi} C \to 0$ is exact if and only if ψ is injective, φ is surjective and $\psi(A) = \ker(\varphi)$, i.e., B is an extension of C by A.

short exact sequence are not limited to modules – we can define them for groups as was done in section 3.6.

Definition 16.1.4: Short Exact Sequences

Let A, B, C be R-modules (or groups). Then if

$$0 \to A \xrightarrow{\varphi} B \xrightarrow{\psi} C \to 0$$

is an exact sequence, we call it a *short exact sequence*, and we say that "B is an extension of C by A".

43: Group Extension Nomenclature

Due to historical reasons, when we talk about extension of groups, it will sometimes be said that B is an extension of A by C. This terminology is backwards to how short exact sequences are usually studied now, but it is still used commonly enough in the literature of classical group theroy

Example 16.1.2: Short Exact Sequences

1. Given modules A and C we can always form their direct sum $B = A \oplus C$ and the squence

$$0 \to A \xrightarrow{\iota} A \oplus C \xrightarrow{\pi} C \to 0$$

where $\iota(a) = (a, 0)$ and $\pi(a, c) = c$ is a short exact sequence. IN particualr, it follows that there always exists at least one extension of C by A.

2. As a special case, Consider

$$0 \to \mathbb{Z} \xrightarrow{\ell} (\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}) \xrightarrow{\pi} \mathbb{Z}/n\mathbb{Z} \to 0$$

where $\ell(z) = (z, 0)$. and $\pi(z, [z]) = [z]$, giving us an extension of $\mathbb{Z}/n\mathbb{Z}$ by $\mathbb{Z}$. Interestingly, this is not the unique extension of $\mathbb{Z}/n\mathbb{Z}$ by $\mathbb{Z}$. Consider:

$$0 \to \mathbb{Z} \xrightarrow{n} \mathbb{Z} \xrightarrow{\pi} \mathbb{Z}/n\mathbb{Z} \to 0$$

where $n(\mathbb{Z}) = nz$, and $\pi(z) = [z]$ is the natural projection. showing that $\mathbb{Z}$ is also an extension of $\mathbb{Z}/n\mathbb{Z}$ by $\mathbb{Z}$. Note that $\mathbb{Z}$ and $(\mathbb{Z}, \mathbb{Z}/n\mathbb{Z})$ are not isomorphic (even though A and C are), giving us two "inequivalent" extension of $\mathbb{Z}/n\mathbb{Z}$ by $\mathbb{Z}$!

3. Let's say we have a B, and C. Then we can choose an R-module homomorphism $\varphi(B) \to C$ and then we'd have

$$0 \to \ker\varphi \xrightarrow{\ell} B \xrightarrow{\varphi} \text{image}\varphi \to 0$$

would be a short exact sequence. If φ is surjective, then

$$0 \to \ker\varphi \xrightarrow{\ell} B \xrightarrow{\varphi} C \to 0$$

would, of course, be a short exact sequence.

4. Here's an important particular example of the last example. . Let M be an R-module, S be a set of generators for M so that $M = RS$. Let $F(S)$ be the free-module generated by S. Then
$$0 \to K \xrightarrow{\ell} F(S) \xrightarrow{\varphi} M \to 0$$
is a short exact sequence, where φ is the unique homomorphism which is the identity on S, and $K = \ker\varphi$, ℓ is the inclusion map.

5. The give an example with two "inequivalent" extensions for groups, Take the Klein 4-group represented as the cyclic group Z_2 of order two. Then
$$herep.380$$
are two extensions.

16.1.1 Relating Short Exact Sequences

As seen in example 2 and 5, we can have more than one "unique" extensions of modules (or groups). The natural problem then would be to know when an extension is "the same" and when it we have two different copies:

Definition 16.1.5: homomorphism of short exact sequence and equivalent sequences

Let $0 \to A \to B \to C \to 0$ and $0 \to A' \to B' \to C' \to 0$ be two short exact sequence of modules

1. A *homomorphism of short exact sequences* is a triple α, β, γ of module homomorphisms such that the following diagram commutes
$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "\varphi"] \arrow[d, "\alpha"] & B \arrow[r, "\psi"] \arrow[d, "\beta"] & C \arrow[r] \arrow[d, "\gamma"] & 0 \\ 0 \arrow[r] & A' \arrow[r, "\varphi'"] & B' \arrow[r, "\psi'"] & C' \arrow[r] & 0 \end{tikzcd}$$
the homomorphism is an *isomorphism of short exact sequence* if α, β, γ are all isomorphisms, in which case the extension of B and B' are said to be isomorphic.

2. The two exact sequences are called *equivalent* if $A = A'$ and $C = C'$ and there is an isomorphism between them as in (1) that is the identity maps on A and C. in this case, the corresponding extensions B and B' are said to be *equivalent* extensions.

You get some very interesting results if you can say two short exact sequence are isomorphic. Lets' say you have two isomorphic short exact sequence, so that $\beta : B \to B'$ is an isomorphism. Then β restricts to an isomorphism on A, that is $\beta|_{\varphi(A)} \equiv \alpha$, and induces an isomorphism on C, that is, given some right-inverse R (which exists since ψ is surjective, then $\beta|_{R(C)} \equiv \gamma$! If two short exact sequence are equivalent, then these two restrictions are both identity maps!

Isomorphic short exact sequence give us all possible "equivalent" short exact sequence (and indeed we can make an equivalence relation, since isomorphisms are invertible). If we have two "equivalent" short exact sequence, it is stronger than two short exact sequence. This occurs whne ψ' maps to C' in an "equivalent" way to ψ. This equivalent short exact sequence also make sure that the maps "agree" as well.

Example 16.1.3

1. Let m, and n be integesr greater than 1. Assume n divides m $n|m$, and let $k = \frac{m}{n}$. Define a map from the exact sequence of $\mathbb{Z}$-module in example 2 from the preceding set of examples. Then:

$$\begin{tikzcd}
0 \arrow[r] & \mathbb{Z} \arrow[r, "n"] \arrow[d, "\alpha"] & \mathbb{Z} \arrow[r, "\pi"] \arrow[d, "\beta"] & \mathbb{Z}/n\mathbb{Z} \arrow[r] \arrow[d, "\gamma"] & 0 \\
0 \arrow[r] & \mathbb{Z}/k\mathbb{Z} \arrow[r, "l"] & \mathbb{Z}/m\mathbb{Z} \arrow[r, "\pi'"] & \mathbb{Z}/n\mathbb{Z} \arrow[r] & 0
\end{tikzcd}$$

   where α and β are natural projection, γ is the identity, l maps a mod k to na mod m ,and π' is the natural projection of $\mathbb{Z}/m\mathbb{Z}$ onto it's quoitnet $(\mathbb{Z}/m\mathbb{Z})(n\mathbb{Z}/m\mathbb{Z})$. THis can be checked to be an homomorphism of two short exact sequences.

2. Take again $0 \to \mathbb{Z} \xrightarrow{n} \mathbb{Z} \xrightarrow{\pi} \mathbb{Z}/n\mathbb{Z} \to 0$, and another short exact sequence where every module is the image of a function that maps the module to itself by $x \mapsto -x$. Then this is an isomorphism of two short exact sequence. These are not equivalent short exact sequences since the isomorphisms are not identity maps.

3. From the previous set of examples (following the definition of short exact sequence), the two short exact sequences from example (2) are *not* Isomorphic (the extension module is not isomorphic). Similarly for the previous example (5)

4. Here are two equivalent short exact sequence

$$\begin{tikzcd}
0 \arrow[r] & \mathbb{Z}/2\mathbb{Z} \arrow[r, "\psi"] \arrow[d, "\mathrm{id}"] & \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \arrow[r, "\varphi"] \arrow[d, "\beta"] & \mathbb{Z}/2\mathbb{Z} \arrow[r] \arrow[d, "id"] & 0 \\
0 \arrow[r] & \mathbb{Z}/2\mathbb{Z} \arrow[r, "\psi'"] & \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \arrow[r, "\varphi'"] & \mathbb{Z}/2\mathbb{Z} \arrow[r] & 0
\end{tikzcd}$$

   (p.382)

5. Here are two isomorphic sequences which are *not* equivalent:

$$0 \to \mathbb{Z}/2\mathbb{Z} \xrightarrow{\psi} \mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \xrightarrow{\varphi_i} \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \to 0$$

   (p.382)

You might've noticed that the homomorphism/isomorphisms between short exact sequence seem to

somewhat influence each other. If you've seen this, than that is a valid observation, and is formalised in the following lemma:

Proposition 16.1.6: Short Five Lemma

Let α, β, γ be homomorphism of short exact sequence and consider

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "\psi"] \arrow[d, "\alpha"] & B \arrow[r, "\varphi"] \arrow[d, "\beta"] & C \arrow[r] \arrow[d, "\gamma"] & 0 \\ 0 \arrow[r] & A' \arrow[r, "\psi'"] & B' \arrow[r, "\varphi'"] & C' \arrow[r] & 0 \end{tikzcd}$$

Then

1. if α and γ are injective, then so if β
2. if α and γ are surjective), then so if β
3. if α and γ are isomorphisms, then so if β (and then the two sequence are isomorphic)

Note: This result also holds proofs for short exact sequence of groups (not neccessarily abelian) as the proof will demonstrate

Proof **:**
I'll do the proof for injectivity. Surjectivity is similar, and isomorphism will naturally come from that. I'll break down the proof in multiple implications. The idea is that we want to show that if $\beta(b) = 0$ then $b = 0$. I'd recommend to draw an image while looking through this proof to trace where all the arrows are going:

$$\begin{aligned} \text{Consider } \beta(b) = 0. &\Rightarrow \varphi'(\beta(b)) = 0 \\ \text{By commutativity} &\Rightarrow \gamma(\varphi(b)) = 0 \\ \text{Since } \gamma \text{ is inj} &\Rightarrow \varphi(b) = 0 \\ \text{written differently} &\quad b \in \ker(\varphi) \\ \text{by exactness} &\Rightarrow b \in \psi(A) \Rightarrow b = \psi(a) \end{aligned}$$

line (19) gives us some crucial information, which is why we've done (18). With (19), we can use exactness to get (21). That last line might feel redundant, we already know there should be $a \in A$ such that $\psi(a) = b$. But ψ is only injective, we didn't know *for sure* that there was $a \in A$ that mapped to a $b \in B$ such that $\beta(b) = 0$. Which means we've gained information. This is great, since we can use the properties of the diagram yet again:

$$\text{By commutativity} \Rightarrow \psi'(\alpha(a)) = \beta(\psi(a)) = \beta(b) = 0 \tag{16.1}$$

$$\text{Crucially, since } \psi', \alpha \text{ inj} \Rightarrow a = 0 \tag{16.2}$$

$$\text{Therefore} \quad b = \psi(a) = \psi(0) = 0 \tag{16.3}$$

and so $\varphi(b) = 0 \Rightarrow b = 0$, showing that β is injective.

Another common way of relating two short exact sequences is by combining them to form a long exact sequence:

Proposition 16.1.7: Mini Snake Lemma

Let α, β, γ be homomorphism of short exact sequence and consider

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r, "\psi"] \arrow[d, "\alpha"] & B \arrow[r, "\varphi"] \arrow[d, "\beta"] & C \arrow[r] \arrow[d, "\gamma"] & 0 \\ 0 \arrow[r] & A' \arrow[r, "\psi'"] & B' \arrow[r, "\varphi'"] & C' \arrow[r] & 0 \end{tikzcd}$$

Then there exists an exact sequence:

$$0 \to \ker(\alpha) \xrightarrow{\overline{\psi}} \ker(\beta) \xrightarrow{\overline{\varphi}} \ker(\gamma) \xrightarrow{d} \operatorname{coker}(\alpha) \xrightarrow{\overline{\psi}'} \operatorname{coker}(\beta) \xrightarrow{\overline{\varphi}'} \operatorname{coker}(\gamma) \to 0$$

where $\overline{\varphi}, \overline{\psi}$ are restriction, $\overline{\varphi}', \overline{\psi}'$ are induced by φ' and ψ', and d is a homomorhpisms called the *boundary homomorphism*.

Proof **:**
To define the boundary homomorphism, start by taking $y \in \ker(\gamma)$. Then since φ is surjective, we have there exists an x such that $\varphi(x) = y$. By commutativity:

$$\varphi'(\beta(x)) = \gamma(\varphi(x)) = 0$$

and so $\beta(x) \in \ker(\varphi') = \operatorname{im}(\psi')$, meaning there exists a x' such that $\psi'(x') = \beta(x)$. Thus, define $d(x')$ to be the image of x' in $\operatorname{coker}(\alpha')$. With this, the verification that this is indeed an exact sequence follows naturally through diagram chasing as was done in proposition 16.1.6.

The proposition is called the "mini" snake lemma since it is a shorter version of the more general snake lemma that will be presented in chapter ref:HERE (homAlg).

(I think I should also include the "famous 5 lemma", important for tensor products, see this Richard Borcherd video)

16.1.2 Splitting Sequences

Coming back to the earlier example of $0 \to A \xrightarrow{\iota} A \oplus C \xrightarrow{\pi} C \to 0$. _____(p.383). When this occurs, we'll say the sequence splits. This will become very important when we talk about *projective modules* soon:

Definition 16.1.8: short exact sequence that *splits*

1. Let R be a ring and let $0 \to A \xrightarrow{\psi} B \xrightarrow{\varphi} C \to 0$ be a short exact sequence of R-modules. The sequence is said to *split* if there is an R-module complement to $\psi(A)$ in B. In this case, up to isomorphism, $B = A \oplus C$ (more precisely, $B = \psi(A) \oplus C'$ for some submodule C', and C' is mapped isomorphically onto C by $\varphi(C') \cong C$)
2. same for groups

In either case, the extension B is said to be a *split exentsions* of C by A

Since in a sense, the question of whether an extension siplits is the question of the existnece of a complemnt to $\psi(A)$ in B isomorphic (by φ) to C, so the notion of a split extension may equivalently be phrased in the language of homomorphisms

Proposition 16.1.9: Identifying splits

The short exact sequence $0 \to A \xrightarrow{\psi} B \xrightarrow{\varphi} C \to 0$ of R-modules is split if and only if there is an R-module homomorphism $\mu : C \to B$ such that $\varphi \circ \mu$ is the identity map on C.
Similarly, the short exact sequence $1 \to A \xrightarrow{\psi} B \xrightarrow{\varphi} C \to 1$ of groups is split if and only if there is a group homomorphism $\mu : C \to B$ such that $\varphi \circ \mu$ is the identity map on C

Proof. This follows directly from the definitions: if μ is given, define $C' = \mu(C) \subseteq B$ and if C' is given, define $\mu = \varphi^{-1} : C \cong C' \subseteq B$. □

The functionμ is quite important, and so is given a name:

Definition 16.1.10: section and splitting homomorphism

With notation as in the previous proposition, any set map $\mu : C \to B$ such that $\varphi \circ \mu = \mathrm{id}$ is called a *section* of φ. is μ is a homomorphism as in the previous proposition, then μ is called a *splitting homomorphism* for the sequence.

1. naturally, $0 \to A \xrightarrow{l} A \oplus C \xrightarrow{\pi} C \to 0$ has the evident splitting homomorphism $\mu(c) = (0, c)$
2. Take our two short exact sequence from (2) from example 5.1. Clearly the first one splits, but the second one doesn't since there is no nonzero map frmo $\mathbb{Z}/n\mathbb{Z}$ to $\mathbb{Z}$.
3. Example 5.1 (5) is not a split extension
4. However, D_8 can be a split extensionof Z_2 by Z_4. Namely

$$1 \to \mathbb{Z}_4 \xrightarrow{l} D_8 \xrightarrow{\pi} Z_2 \to 1$$

namely,

$$1 \to \langle r \rangle \xrightarrow{l} D_8 \xrightarrow{\pi} \langle s \rangle \to 1$$

to understand this one, we'd need semi-direct products, which I've skipped learning at this stage.

Proposition 5.2 showed that if $0 \to A \xrightarrow{\psi} B \xrightarrow{\varphi} C \to 0$ is a splitting sequence, then there exists a splitting homomorphism $\mu : C \to B$

Proposition 16.1.11: the other side of the squence

Let $0 \to A \xrightarrow{\psi} B \xrightarrow{\varphi} C \to 0$ be a short exact sequence of modules (respectively for groups). Then $B = \psi(A) \oplus C'$ for some submodule C' of B with $\varphi(C') \cong C$ if and only if there is a homomorphism $\lambda : B \to A$ such that $\lambda \circ \psi$ is the identity map on A.

Proof. here □

Note: to check. If two sequence are isomorphic, one splits if and only if the other splits? Does it need to be equivalent to hold?

16.2 Projective Modules

We turn back to short exact sequence, and repeating the *extension problem*: Given L and N, what R-module's M exist such that

$$0 \to L \to M \to N \to 0$$

is exact, i.e, what are the possible extensions of N given L? The answer is non-trivial, and will require a lot of build-up culminating in defining a group $\mathrm{Ext}^i_R(L, N)$ which will be isomorphic to the set of all possible extensions of N by L (given an appropriate group-structure on the extensions M_i). In this section, we will do the first part of this build-up by finding modules that would only have trivial extensions: $\mathrm{Ext}^i_R(-, P) = 0$.

Suppose we have a short exact sequence of R-modules $0 \to L \to M \to N \to 0$; we shall constantly reference to this short exact sequence. We know that for any R-module D, we can define the functor $\mathrm{Hom}_R(D, -)$ and $\mathrm{Hom}_R(-, D)$. The first thing we will ask is when are $\mathrm{Hom}_R(D, -)$ and $\mathrm{Hom}_R(-, D)$ *exact functors*, that is, Given a short exact sequence like the one earlier, when is

$$0 \to \mathrm{Hom}_R(D, L) \to \mathrm{Hom}_R(D, M) \to \mathrm{Hom}_R(D, N) \to 0$$

exact, or when is

$$0 \to \mathrm{Hom}_R(N, D) \to \mathrm{Hom}_R(M, D) \to \mathrm{Hom}_R(L, D) \to 0$$

exact (notice that for $\mathrm{Hom}(__, D)$, the order is flipped since $\mathrm{Hom}(__, D)$ is a contravariant functor). We will start by studying $\mathrm{Hom}(D, __)$ being applied to the short exact sequence.This functor can be visualized in the following [not necessarily commutative] diagram:

$$\begin{tikzcd}
 & & D \arrow[dl] \arrow[dr] & & \\
0 \arrow[r] & L \arrow[r] & M \arrow[r] & N \arrow[r] & 0
\end{tikzcd}$$

Given a homomorphism $L \to M$, do we have a homomorphism from $\mathrm{Hom}_R(D, L)$ to $\mathrm{Hom}_R(D, M)$:

$$\begin{tikzcd} D \arrow[d, "f"] \arrow[dr, dashed, "F?"] & \\ L \arrow[r, "\psi"] & M \end{tikzcd}$$

the answer to this is quite easy: map f to $\psi \circ f$:

$$\begin{tikzcd} D \arrow[d, "f"] \arrow[dr, "F=\psi\circ f"] & \\ L \arrow[r, "\psi"] & M \end{tikzcd}$$

Or, written another way, there exists a function:

$$\psi' : \mathrm{Hom}_R(D, L) \to \mathrm{Hom}_R(D, L) \qquad f \mapsto f \circ \psi \ [= f']$$

Furthermore, if $\psi : L \to M$ is injective, that is $0 \to L \to M$ is exact, then $\psi' : \mathrm{Hom}_R(D, L) \to \mathrm{Hom}_R(D, M)$ is injective, that is $0 \to \mathrm{Hom}_R(D, L) \to \mathrm{Hom}_R(D, M)$ is exact (we'll prove that in just a moment). However, it is *not* the case that if we have a homomorphism $\varphi : M \to N$ and a homomorphism $f : D \to N$ that there will be an $F : D \to M$ that maps to it, that is, in the following diagram:

$$\begin{tikzcd} D \arrow[d, dashed, "F?"] \arrow[dr, "f"] & \\ M \arrow[r, "\varphi"] & N \end{tikzcd}$$

the homomorphism F doesn't necessarily have to exist

Example 16.2.1: Counter-Example

Take the $\mathbb{Z}$-module homomorphism $\pi : \mathbb{Z} \to \mathbb{Z}2\mathbb{Z}$ where $\pi(z) = \overline{z}_2$ is the natural projection, and let $D = \mathbb{Z}/2\mathbb{Z}$. Take $\mathrm{id}_{\mathbb{Z}/2\mathbb{Z}} : D \to \mathbb{Z}/2\mathbb{Z}$. As we've seen before, there exists only the trivial homomorphism from D to $\mathbb{Z}$, thus we are forced to have the following diagram:

$$\begin{tikzcd} \mathbb{Z}/2\mathbb{Z} \arrow[d, "0"] \arrow[dr, "\mathrm{id}_{\mathbb{Z}/2\mathbb{Z}}"] & \\ \mathbb{Z} \arrow[r, "\pi"] & \mathbb{Z}/2\mathbb{Z} \end{tikzcd}$$

this diagram is not commutative, since $\pi \circ 0 \neq \mathrm{id}_{\mathbb{Z}/2\mathbb{Z}}$. Since 0 must map to 0, we are left with no possible maps $g : \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ such that $\mathrm{id}_{\mathbb{Z}/2\mathbb{Z}} = \pi \circ g$.

Phrasing this observation in terms of the exact sequence, if $\varphi : M \to N$ is surjective, that is $M \to N \to 0$ is exact, then it need not be the case that $\varphi' : \mathrm{Hom}_R(D, M) \to \mathrm{Hom}_R(D, N)$ is surjective, that is $\mathrm{Hom}_R(D, M) \to \mathrm{Hom}_R(D, N) \to 0$ need not be exact.

We flesh out these observation in the following proposition:

Proposition 16.2.1: Left-exactness of $\mathrm{Hom}(D, -)$

Let D, L, and M be R-modules and let $\Psi L :\to M$ be an R-module homomorphism. Then the map

$$\psi' : \mathrm{Hom}_R(D, L) \to \mathrm{Hom}_R(D, M) \qquad f \mapsto f \circ \psi \ [= f']$$

is a homomorphism of abelian groups. If ψ is injective, then ψ' is injective, i.e.,

$$\text{if } 0 \to L \xrightarrow{\psi'} M \text{ is exact,}$$

$$\text{then } 0 \to \mathrm{Hom}_R(D, L) \xrightarrow{\psi'} \mathrm{Hom}_R(D, M) \text{ is also exact}$$

More generally, if

$$0 \longrightarrow L \xrightarrow{\psi} M \xrightarrow{\varphi} N \longrightarrow 0$$

is exact, then

$$0 \longrightarrow \mathrm{Hom}_R(D, L) \xrightarrow{\psi'} \mathrm{Hom}_R(D, M) \xrightarrow{\varphi'} \mathrm{Hom}_R(D, N)$$

is exact for all D, i.e., $\mathrm{Hom}(D, __)$ is [covariant] *left-exact*

***Proof* :**
Let $f, g \in \mathrm{Hom}_R(D, L)$ and we have $\mathrm{Hom}_R(D, M)$,

$$\psi'(f + g) = \psi(f + g) = \psi \circ f + \psi \circ g = \psi'(f) + \psi'(g)$$

showing that ψ' is a homomorphism, even when we replace M with any other R-module.

Now, Let's show left-exactness. Suppose that $0 \to L \xrightarrow{\psi} M$ is exact. That would imply that ψ is injective. By proposition ref:HERE (Monomorphism), ψ is a monomorphism. That implies that if we have $f, g \in \mathrm{Hom}_R(D, M)$ where $\psi'(f) = \psi'(g)$, that is, $\psi \circ f = \psi \circ g$ then $f = g$, showing that ψ' is also injective.

Next, we need to show exactness at $\mathrm{Hom}_R(D, M)$, that is $\mathrm{im}(\psi') = \ker(\varphi')$.

($\subseteq$) Let $F \in \mathrm{im}(\psi')$, that is there is an $F' \in \mathrm{Hom}_R(D, L)$ such that $\psi'(F') = F$, i.e. $(\psi \circ F')(d) = F(d)$ for all $d \in D$. we want to show that $\varphi(F) = 0$, that is $(\varphi \circ F)(d) = 0$. By surjectivity, we have $\varphi(F(d)) = \varphi(\psi(F'(d))$. Since $\mathrm{im}(\psi) = \ker(\varphi)$, we have that $\varphi \circ \psi = 0$, and so $\varphi(F(d)) = \varphi(\psi(F'(d)) = 0$ for any $d \in D$. Thus, it follows that φ' maps F to the zero function, that is $\varphi'(F) = 0$, and so $F \in \ker(\varphi')$.

($\supseteq$) Let $F \in \ker \varphi'$, that is $F : D \to M$ such that $\varphi \circ F = 0$ (the zero homomorphism from D to M. Then for any $d \in D$, $F(\varphi(d)) = 0$, so $F(d) \in \ker(\varphi)$. Then since M is exact in $0 \to L \to M \to N$, then $F(d) \in \mathrm{im}(\psi)$. So, there exists some $l \in L$ such that $F(d) = \psi(l)$. Since ψ is injective, l is unique. Thus, we can define a map $F' : D \to L$, $F'(d) = l_d$. It is a good exercise to verify that F' is also a homomorphism, which shows that $F' \in \mathrm{Hom}_R(D, L)$.

In terms of a commutative diagram:

$$\begin{tikzcd} & D \arrow[dl, "F' : F'(d) = l_d"'] \arrow[d, "F"] & \\ L \arrow[r, "\psi"] & M \arrow[r, "\varphi"] & N \end{tikzcd}$$

Finally, since $(\psi \circ F')(d) = \psi(l_d) = F(d)$, we have that $\psi'(F') = F$, meaning $F \in \operatorname{im}(\psi')$

Thus, $\ker(\psi') = \operatorname{im}(\varphi')$, showing that the sequence is exact at $\operatorname{Hom}_R(D, M)$ as well, as we sought to show.

Hence, when there is a module P that makes $\operatorname{Hom}_R(P, -)$ an exact functor,[3] we have a special kind of module:

Definition 16.2.2: Projective module

An R-module P is called *projective* $\operatorname{Hom}(P, __)$ is an exact additive functor.

This is a good theoretical definition, but it doesn't lend itself well to computationally. The following theorem gives a nice characterization of projective modules:

[3] which we already demonstrated in example 16.0.1

Theorem 16.2.3: Equivalencies for Projective Modules

Let P be an R-module. Then the following are equivalent

1. For any R-modules L , M, N, if

$$0 \to L \xrightarrow{\psi} M \xrightarrow{\varphi} N \to 0$$

is a short exact sequence, then

$$0 \to \mathrm{Hom}_R(P, L) \xrightarrow{\psi'} \mathrm{Hom}_R(P, M) \xrightarrow{\varphi'} \mathrm{Hom}_R(P, N) \to 0$$

is a short exact sequence.

2. For any R-modules M and N, if $M \xrightarrow{\psi} N \to 0$ is exact, then every R-module homomorphism from P into N lifts to an R-module homomorphism into M, i.e., given $g \in \mathrm{Hom}_R(P, N)$ there is a lift $F \in \mathrm{Hom}_R(P, M)$ making the following diagram commute

$$\begin{tikzcd} & P \arrow[dl, dashed, "F"'] \arrow[d, "f"] & \\ M \arrow[r, "\psi"'] & N \arrow[r] & 0 \end{tikzcd}$$

3. If P is a quotient of the R-module M , then P is isomorphic to a direct summand of M, i.e., every short exact sequence $0 \to L \to M \to P \to 0$ splits

4. P is a direct summand of Free R-modules.

Note: In the 4th implication, it is **direct summand**, not direct sum, of free R-modules! A direct summand is any subpart of a direct sum (including the sum itself). If $A \oplus B \oplus C$ is a direct sum,

$$A,\ B,\ C,\ A \oplus B,\ A \oplus C,\ B \oplus C,\ A \oplus B \oplus C$$

are its direct summand.

As a concrete example, take $\mathbb{Z}_6$ as a free module over itself. Then $\mathbb{Z}_2 \oplus \mathbb{Z}_3 \cong \mathbb{Z}_6$, so $\mathbb{Z}_2$ is a projective $(\mathbb{Z}/6\mathbb{Z})$-module, since it's a direct *summand* of a free $(\mathbb{Z}/6\mathbb{Z})$-module. It is *not* a free $\mathbb{Z}/6\mathbb{Z}$ module, since $\mathbb{Z}/2\mathbb{Z}$ cannot be isomorphic to a direct sum of $(\mathbb{Z}/6\mathbb{Z})^n$ as an $\mathbb{Z}/6\mathbb{Z}$-module

Proof **:**

$(1 \Leftrightarrow 2)$ One implies two comes immediately from the reasoning we talked about earlier

$(2 \Rightarrow 3)$ Since P is a quotient of an R-module M, then we have a short exact sequence $0 \to L \to M \to P \to 0$. By (2), we have that any map from P to P lifts to a map from P to

M. Let $\mathrm{id} : P \to P$ lift to $f : M \to P$:

$$\begin{tikzcd} & P \arrow[dl, dashed, "F"'] \arrow[d, "\mathrm{id}"] & \\ M \arrow[r, "\varphi"] & P \arrow[r] & 0 \end{tikzcd}$$

thus, $f \circ \varphi = \mathrm{id}$, implying that φ splits, i.e. $M = N \oplus P$

($3 \Rightarrow 4$) Since P is an R-module, it is a quotient of a free R-module (corollary ref:HERE), meaning we have the following short exact sequence

$$0 \longrightarrow K \longrightarrow F(S) \longrightarrow P \longrightarrow 0$$

for appropriate generating set S (for example, take $S = P$) and kernel K. Then by (3), we have $F(S) = K \oplus P$, showing P is a direct summand of a free R-module

($4 \Rightarrow 2$) Let P be a direct summand of a free R-module for some set S, that is $F(S) = P \oplus K$. Let f be the homomorphism from P to N, and φ be the homomorphism from M to N:

$$\begin{tikzcd} & P \arrow[d, "f"] & \\ M \arrow[r, "\varphi"] & N \arrow[r] & 0 \end{tikzcd}$$

we will make a map from P to M by invoking the universal property of free R-modules. Since P is the direct summand of a free R-module, let π be the natural projection from $F(S)$ to P. Since there is a map from P to f, we thus have a set map from S to N through $s \mapsto f \circ \pi(s) = n_s \in N$. Since φ is surjective, let $m_s \in M$ be an element such that $\varphi(m_s) = n_s$. Thus, there is a map from S to M. Through this, the following diagram commutes:

$$\begin{tikzcd} S \arrow[r, dotted, "\iota"] \arrow[dr, dotted, "s \mapsto m_s"'] & F(S) = P \oplus K \arrow[d, "\pi"] & \\ & P \arrow[d, "f"] & \\ & M \arrow[r, "\varphi"] & N \arrow[r] & 0 \end{tikzcd}$$

Thus, by the universal property of free R-modules, there exists an R-module homomorphism F' from $F(S)$ to M where $F'(s) = m_s$:

$$\begin{tikzcd} S \arrow[r, dotted, "\iota"] \arrow[dr, dotted, "s \mapsto m_s"'] & F(S) = P \oplus K \arrow[d, "\pi"] \arrow[dl, dashed, "F'"] & \\ & P \arrow[d, "f"] & \\ & M \arrow[r, "\varphi"] & N \arrow[r] & 0 \end{tikzcd}$$

Note that this diagram is indeed commutative since $\varphi \circ F'(s) = \varphi(m_s) = n_s = f \circ \pi(s)$, and the functions are determined by where s is mapped to, it follows that $\varphi \circ F' = f \circ \pi$.

Now, define $F : P \to M$, $F(p) = F'((p,0)) = (F \circ \iota)(p)$ where ι is the natural injective map $\iota : P \to P \oplus K$. Since ι and F' are R-module homomorphisms, so is F. We want to show that we can draw an arrow from P to M in our diagram (i.e., that F would be commutative), so we must show that $\varphi \circ F = f$. Indeed

$$(\varphi \circ F)(p) = (\varphi \circ F')((p,0)) \overset{!}{=} (f \circ \pi)((p,0)) = f(p)$$

where the $\overset{!}{=}$ equality comes from the previously demonstarted commutativity. Thus, we can put an arrow from P to M. Simplifying our diagram, we get:

$$\begin{tikzcd} & P \arrow[dl, dashed, "F"'] \arrow[d, "f"] & \\ M \arrow[r, "\varphi"] & N \arrow[r] & 0 \end{tikzcd}$$

showing that every homomorphism f from P to N lifts to a homomorphism P to M, proving $4 \to 2$

Notice how powerful projective modules are: given any surjective map $P \to N$, we can make N "bigger" and define the map $P \to M$ where $\varphi : M \twoheadrightarrow N$. In this way, we can see where the terminology "projective" comes from, since we can interpret M as the space where all the fibers $m + \ker(\varphi)$ have not yet been "projected" onto N (i.e. that φ has yet to add "extra relation", hence extra restrictions), so P can always naturally maps to the "total space M" and factor through the "projection" φ. It further helps that the structure of projective modules are simply direct summands which are the natural candidates (and as we have just proven only candidates) to have this projective property.

Thinking of projective modules in terms of generators and relations, recall that direct sums can be interpreted as the trivial extensions. Each direct summands of a free R-module thus can be thought as appending the least amount of new relations. In this way, projectives are the natural generalization from free-modules in the more general case where the ring R that is acting on the abelian group M may have non-trivial structure, but still behaves "freely".

As mentioned, free modules are projective. The converse is not the case. Take the ring $R = (\mathbb{Z}/2\mathbb{Z}) \oplus (\mathbb{Z}/2\mathbb{Z})$ over itself. Any ring over itself is a free module. Since $(\mathbb{Z}/2\mathbb{Z})$ is a direct summand of R, it is projective. However it is not free for:

$$(0,1) \cdot (1,0) = 0$$

that is, the R-module $(\mathbb{Z}/2\mathbb{Z}) \oplus 0$ has an element that is linearly dependent. Hence:

$$\text{free} \Rightarrow \text{projective}, \ \text{projective} \not\Rightarrow \text{free}$$

Summing this up in a corollary

Corollary 16.2.4: free modules are projective

Let F be a free R-module. then F is a projective module.

Proof **:**
This is an immediate consequence of the 4th implication of theorem 16.2.3. In particular, Since F is free, then

$$F = \bigoplus_{i \in \mathcal{I}} R$$

and so F is it's own direct summand.

We will show a partial converse after a couple of examples of projective modules. In some cases, the following characterization of projective modules is useful[4]:

Proposition 16.2.5: Dual Basis Theorem

Let A be an R-module. Then A is projective if and only if there is a set of elements $a_i \in A$ and a set of R-maps $f_i : A \to R$ such that the sum

$$\sum f_i(a) a_i$$

if finite and equal to a for each element $a \in A$

Proof **:**
Since all projective modules are a direct summand of a free module, Let F be a free R-module with basis (e_i) where $\varphi : F \to A$ is the natural projection. Let $a_i = g(e_i)$. Then A is projective if and only if there is an R-map $f : A \to F$ such that $\varphi \circ f = \text{id}$. Given an f, we can write $f(a) = \sum r_i e_i$ and defien $f_i(a) = r_i$, so that the f_i's are the R-maps that we sought. Conversely, given such f_i, we can define f by $f(a) = \sum f_i(a) e_i$, completing the proof.

Corollary 16.2.6: Hom Characterization of of Projective Module

Let

$$\varphi : \text{Hom}_R(A, R) \otimes_R A \to \text{Hom}_R(A, A) \qquad \varphi(f \otimes b)(a) = f(a)b$$

Then A is finitely generated and projective if and only if φ is an isomorphism.

Proof **:**
exercise, use the above proposition

Example 16.2.2: Projective Modules
Recall lemma lemma ref:HERE that proves $\oplus$ and Hom commutes. This allows for the construction of some interesting projective modules, as shall be left for the reader to to play around with.

1. If F is a field, then F-modules (i.e. vector spaces over F) are free (theorem 13.1.4), and hence projective
2. More generally, any ring R over itself is a free R-module or rank 1, and so is projective. We

[4]This characterization will be needed in some results in number theory, chapter 22

can also see this by recalling that for any R-module $\mathrm{Hom}_R(R, M) \cong_{\mathbf{Ab}} M$ so the short exact sequence of R-modules maps to a short exact sequence of abelian groups.

3. As an intersting case, take $R = \mathbb{Z}$ as a $\mathbb{Z}$-module. By the 2nd condition of theorem 16.2.3, If we have $M \xrightarrow{\varphi} N \to 0$ to be an exact sequence, and a homomorphism $f : \mathbb{Z} \to N$, then it can be lifted to $F : \mathbb{Z} \to M$. Notice that f is completely determine by where 1 maps to (i.e. given $f(1) = n_1$, we can determine all other values f maps to by linearity). Then taking any m_{n_1} such that $\varphi(m_{n_1}) = n_1$ we can define $F(1) = m_{n_1}$, as our extension of f.

   If instead we look at the 3rd condition of theorem 16.2.3 , then any short exact sequence $0 \to M \xrightarrow{\psi} N \xrightarrow{\varphi} \mathbb{Z} \to 0$ splits. Since φ is surjective, let $n_1 \in N$ such that $\varphi(n_1) = 1$. Take $f : \mathbb{Z} \to N$ where $f(x) = x \cdot n_1$. Then f is a homomorphism since $f(x + y) = (x + y) \cdot n_1 = x \cdot n_1 + y \cdot n_1 = f(x) + f(y)$, and furthermore $\varphi \circ f = \mathrm{id}$. To show this, notice that

$$\varphi(n_p) = p = \underbrace{1 + 1 \cdots + 1}_{n \text{ times}} = \underbrace{\varphi(n_1) + \varphi(n_1) + \cdots \varphi(n_1)}_{n \text{ times}} = p \cdot \varphi(n_1)$$

   so

$$\varphi(f(p)) = \varphi(n_p) = p \cdot \varphi(n_1) = p \cdot 1 = p$$

   thus f is a section homomorphism, showing that our short exact sequence is indeed a split exact sequence, so that $0 \to L \to M' \oplus \mathbb{Z} \to \mathbb{Z} \to 0$. Since L is in the kernel of the map from $M' \oplus \mathbb{Z}$ to $\mathbb{Z}$, $M' = L$.

4. Since free $\mathbb{Z}$-modules have no nonzero element of finiter order, there does not exist a finite submodule of a free $\mathbb{Z}$-module (i.e. free abelian group). Thus, no finite abelian group is a projective $\mathbb{Z}$-module. As an example, we can take $\mathbb{Z}/n\mathbb{Z}$ ($n \geq 2$) from before, and for practice show that it fails the 1st condition of theorem 16.2.3. To do so, we must have a short exact sequence that when passed through $\mathrm{Hom}(\mathbb{Z}/n\mathbb{Z}, __)$ is no longer exact. To do this, we can simply take

$$0 \to \mathbb{Z} \xrightarrow{\cdot n} \mathbb{Z} \xrightarrow{\pi} \mathbb{Z}/n\mathbb{Z} \to 0$$

   from before. The first two terms become 0 since there are no non-trivial homomorphisms from a torsion group to a torsion-free subgroup. For the third term, we saw in exercise ref:HERE that $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/n\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}) \cong \mathbb{Z}/n\mathbb{Z}$, so we get

$$0 \to 0 \xrightarrow{\cdot n'} 0 \xrightarrow{\pi'} \mathbb{Z}/n\mathbb{Z} \to 0$$

   which is not exact at $\mathbb{Z}/n\mathbb{Z}$

5. Similarly to the previous reasoning, since $\mathbb{Q}/\mathbb{Z}$ is a torsion $\mathbb{Z}$-module, it is not a submodule of a free $\mathbb{Z}$-module, and so is not projective. We should therefore be able to find a short exact sequence that doens't split, and indeed

$$0 \to \mathbb{Z} \to \mathbb{Q} \to \mathbb{Q}/\mathbb{Z} \to 0$$

   does not split, since $\mathbb{Q}$ (a torsion-free module) does not contain a submodule to $\mathbb{Q}/\mathbb{Z}$

6. A little harder to show is that $\mathbb{Q}$ is not a projective $\mathbb{Z}$-module. It is left as an exercise to show that

$$0 \to \mathbb{Q} \to \mathbb{Z}^I \to \mathbb{Q} \to 0$$

does not split (if $\iota : \mathbb{Q} \to \mathbb{Z}^I$ is the injective map and $\pi_i : \mathbb{Z} \to \mathbb{Q}$ is the projection for the ith component, then $\pi_i \circ \iota : \mathbb{Q} \to \mathbb{Z}$ is a $\mathbb{Z}$-module homomorphism; what can you deduce?). If we accept that all module over PID that are projective must be free (see theorem ref:HERE), that would imply $\mathbb{Q}$ is projective iff it is free. But the only proper subset of lienarly independent elements of $\mathbb{Q}$ are singletons, which certainly don't span $\mathbb{Q}$.

7. Let L, M, N be any free R-module and L^*, M^*, N^* be their duals (ex. $M^* := \mathrm{Hom}_R(M, R)$). Then if:
$$0 \to L \to M \to N \to 0$$
is a short exact sequence, then:
$$0 \to N^* \to M^* \to L^* \to 0$$
is a short exact sequence, i.e. $(_)^*$ is a contravariant exact functor. This comes from the fact that $(_)^*$ is contravariatn and that L, M, N are free, and hence projective.

8. Given two projective modules P_1 and P_2, then $P_1 \oplus P_2$ is also projective.

9. (Dummit gives an example with $R = M_n(F)$, but I'm not yet sure what he means)

As mentioned before the examples, there is a strengthening of conditions that we can do that will guarantee that a projective module as actually free:

Proposition 16.2.7: [fg] Projective Modules Over Pid'S

Let M be a finitely generated projective module over a Principal Ideal Domain R. Then R is a free-module.

Proof **:**
By the classification of finitely generated modules over PID's, M is free iff it is torsion free. Since M is projective, it is a direct summand of a free module, and a free module is torsion-free, M must be torsion free. But then M is torsion-free, hence free.

The importance of the finite generation is important, without it the result does not hold (exercise ref:HERE). There shall be a few other times where projective modules over the right type of ring imply the module is free ,for example projective module over *local ring* (theorem ref:HERE) but these considerations are better off within their own context.

16.3 Injective Modules

It is important to take a moment and consider the dual question of $\mathrm{Hom}_R(-, D)$. The categorical preliminaries mentioned how the theory of such modules is the same as that for projective modules. However, when concretely interpreting modules for which $\mathrm{Hom}_R(-, D)$ is exact, they turn out to be more subtle R-modules. This warrants proper exploration. Applying $\mathrm{Hom}_R(-, D)$ will result in flipping the arrows in our diagram: If $0 \to L \xrightarrow{\psi} M \xrightarrow{\varphi} N \to 0$ is a short exact sequence of R-modules

then, the result short exact sequence of abelian groups (if it exists) would be

$$0 \longleftarrow \mathrm{Hom}_R(L, D) \longleftarrow \mathrm{Hom}_R(M, D) \longleftarrow \mathrm{Hom}_R(N, D) \longleftarrow 0$$

or if we flip our diagram around:

$$0 \longrightarrow \mathrm{Hom}_R(N, D) \longrightarrow \mathrm{Hom}_R(M, D) \longrightarrow \mathrm{Hom}_R(L, D) \longrightarrow 0$$

In other words, instead of considering the maps *from* an R-module D into L or N , we take maps from L or N *to* D.

$$\begin{tikzcd} 0 \arrow[r] & L \arrow[r] \arrow[dr] & M \arrow[r] & N \arrow[r] \arrow[dl] & 0 \\ & & D & & \end{tikzcd}$$

We know explore the property of $\mathrm{Hom}(__, D)$ applied to short exact sequences. Like before, one of the directions come immediately: if we have $M \to N$ and we have a homomorphism from N to D the functor $\mathrm{Hom}(__, D)$ will produce:

$$\varphi' : \mathrm{Hom}_R(N, D) \to \mathrm{Hom}_R(M, D) \qquad f \mapsto f' = f \circ \varphi$$

or as a commutative diagram:

$$\begin{tikzcd} M \arrow[r, "\varphi"] \arrow[d, dashed, "F"] & N \arrow[dl, "f"] \\ D & \end{tikzcd}$$

here $F = f \circ \varphi$, and as before, we will have that:

$$\text{if } M \xrightarrow{\varphi} N \to 0 \text{ is exact,}$$

$$\text{then } 0 \to \mathrm{Hom}_R(N, D) \xrightarrow{\varphi'} \mathrm{Hom}_R(M, D) \text{ is exact.}$$

which we shall shortly prove. However, it will (again) not be the case that if we have $L \to M$ and a homomorphism form L to D, then there will exist a homomorphism from L to M that commutes with the homomorphism from L to D:

$$\begin{tikzcd} M \arrow[r, "\varphi"] \arrow[d, "f"] & N \arrow[dl, dashed, "F?"] \\ D & \end{tikzcd}$$

Example 16.3.1: Counter Example

Take $f : \mathbb{Z} \xrightarrow{\cdot 2} \mathbb{Z}$ and let $D = \mathbb{Z}/2\mathbb{Z}$. And consider

$$\begin{tikzcd} \mathbb{Z} \arrow[r, "\cdot 2"] \arrow[d, "\pi"] & \mathbb{Z} \arrow[dl, dashed, "F?"] \\ \mathbb{Z}/2\mathbb{Z} & \end{tikzcd}$$

Then if there existed a map $F : \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$, it must be that $F \circ (\cdot 2) = \pi$, that is $F((\cdot 2)(x) = \pi(x)$. However $F((\cdot 2)(x)) = F(2x) = 0$ for all x, so it must be $F = 0$. However, $0 \circ (\cdot 2) \neq \pi$, so the lift F does not exist.

This counter-example also answers question (here) (ex:QuotientGrpExercise) on whether we can reverse the prosess of taking $f : A \to M/N$ to a map $F : A \to M$ where $f = \pi \circ F$: the answer is no in general.

Proposition 16.3.1: $\mathrm{Hom}(__, D)$ and left-exactness

Let D, L, M, and N be R-modules. If

$$0 \to L \xrightarrow{\psi} M \xrightarrow{\varphi} N \to 0 \text{ is exact}$$

then the associated sequence

$$0 \to \mathrm{Hom}_R(N, D) \xrightarrow{\psi'} \mathrm{Hom}_R(M, D) \xrightarrow{\varphi'} \mathrm{Hom}_R(L, D) \text{ is exact}$$

That is, $\mathrm{Hom}(__, D)$ is a *left-exact* contravariant functor.

***Proof* :**
This proof is similar to proposition 16.2.1, with the execption of using the dual concept in certain cases (for example, instead of using properties of monomorphisms, we would use properties of epimorphisms)

Thus, similarly to how $\mathrm{Hom}(D, __)$ is always left-exact, but not necessarily exact, $\mathrm{Hom}(__, D)$ is left-exact, but not necessarily exact. We can thus ask ourselves if there are any modules Q such that $\mathrm{Hom}(__, Q)$ is exact. If there is, we will give such a module a name:

Definition 16.3.2: Injective Modules

An R-module Q is caleld *injective* if $\mathrm{Hom}(__, Q)$ is an exact additive functor.

The reason for this name will come after we find out some more properties of injective modules

Theorem 16.3.3: Injective Modules Equivalences

Let Q be an R-module. Then the following are equivalent:

1. For any R-modules L, M, and N, if

$$0 \to L \xrightarrow{\psi} M \xrightarrow{\varphi} N \to 0$$

is a short exact sequence, then

$$0 \to \mathrm{Hom}_R(N, Q) \xrightarrow{\psi'} \mathrm{Hom}_R(M, Q) \xrightarrow{\varphi'} \mathrm{Hom}_R(L, Q) \to 0$$

is also a short exact sequence

2. For any R-modules L and M, if $0 \to L \xrightarrow{\varphi} M$ is exact, then every R-module homomorphism from L into Q lifts into an R-module homomorphism of M into Q, i.e., given $f \in \mathrm{Hom}_R(L, Q)$ there is a lift $F \in \mathrm{Hom}_R(M, Q)$ making the following diagram commute

$$\begin{tikzcd} 0 \arrow[r] & L \arrow[r, "\varphi"] \arrow[d, "f"] & M \arrow[dl, dashed, "F"] \\ & Q & \end{tikzcd}$$

3. If Q is submodule of the R-module M then Q is a dirrect summand of M, i.e., every short exact sequence $0 \to Q \to M \to N \to 0$ splits.

Proof **:**

$(1 \Leftrightarrow 2)$ This comes from proposition 16.3.1.

$(2 \Leftrightarrow 3)$ Let $0 \to Q \xrightarrow{\psi} M \to N \to)$ be a short exact sequence, and take $\mathrm{id}_Q : Q \to Q$ to be homomorphism to be lifted. Then by (2), there exists an F such that $F \circ \psi = \mathrm{id}_Q$, so F is a splitting homomorphism. The other way follows a proof similar to the projective case.

The 3rd equivalence shows that the R-module Q must map injectively into an isomorphic copy of Q in M ($Q \hookrightarrow Q \times M'$), which explains the name injective.

When studying projective modules, we had a 4th equivalence of having the projective R-module P be a direct summand of a free R-module. Such a simple characterization doesn't happen for injective modules, and hence their classification is more difficult. To give the reader a better understanding of their structure, a result known as *Baer's Criterion* shall show how to determine an R-module is injective based on the ring R (the same way that an R-module is projective if it is a direct summand of $R^{\oplus S}$), and the classification of injective $\mathbb{Z}$-modules shall give a family of injective modules so that the reader has a healthy set of concrete examples to keep in mind. We shall see that most interesting examples of injective modules come from taking a $\mathbb{Z}$-module Q and applying appropriate functors to get injective R-modules.

Starting off with the following important characterization of injective R-modules:

Theorem 16.3.4: Baer's Criterion

Let Q be an R-module

1. The module Q is injective if and only if for every left ideal I of R, any R-module homomorphism $g : I \to Q$ can be extended to an R-module homomorphism $G : R \to Q$.
2. If R is a P.I.D. then Q is injective if and only if $rQ = Q$ for every $r \in R$. In particular, a $\mathbb{Z}$-module is injective if and only if it is divisible ($rQ = Q$ for all $r \in Q$). When R is a P.I.D., quotient modules of injective modules are again injective.

***Proof* :**

1. To prove the "only if" direction, let $f : I \to Q$ be an R-module homomorphism from the [nonzero] idea l I of R. Then since Q is injective and $0 \to I \to R$ exact, we know there is a lift from I to R such that $F : R \to Q$ is an R-module homomorphism.

   Conversely, suppose that for any $f : I \to Q$, we can lift this to $f : R \to Q$. We will show that Q is injective by proving the 2nd equivalence of theorem 16.3.3, that is, if $0 \to L \to M$ is exact and $f : L \to Q$ is an R-module homomorphism, then there is a $F : M \to Q$ that extends f, that is

$$\begin{tikzcd} 0 \arrow[r] & L \arrow[r, hook, "\varphi"] \arrow[d, "f"] & M \arrow[dl, dashed, "F?"] \\ & Q & \end{tikzcd}$$

   To do this, we will use Zorn's lemma. Let $\mathcal{L}$ be the collection of lifts of f (f', L'), where $f' : L' \to Q$ and a submodule L' of M containing [an isomorphic copy of] L. Then we can define the ordering $(f', L') \leq (f'', L'')$ if $L' \subseteq L''$ and $f' = f''$ on L'. This is clearly a partial order since $\subseteq$ is a partial order on the submodules of M. Furthermore, every chain has an upper bound (if $(L_1, f_1) \leq (L_2, f_2) \leq \cdots$ is a chain, take $\mathfrak{L} = \cup_i L_i$ and let $f : \mathfrak{L} \to Q,\ f(l) = f_i(l)$ where we choose the f_i $l \in L_i$). Since $(L, \mathrm{id}_L) \in \mathcal{L}$, this partial order is nonempty, and so by Zorn's lemma (theorem A.2.3), there exists a maximal element $(M', F) \in \mathcal{L}$ (i.e. if for some $(L', f') \in \mathcal{L}$ $(M', F) \leq (L', f')$, then $(M', F) = (L', f')$). It suffices to show now that $M' = M$

   For the sake of contradiction, let's say there is a nonzero element $m \in M$ where $m \notin M'$. Take $I = \{r \in R \mid rm \in M'\}$ (i.e. all element $r \in R$ such that rm will be contained in M'). Notice I is a left ideal of R: $0 \in I$, if $a, b \in I$, then $a + b \in I$, and for all $r \in R$, $ra \in I$. Furthermore, we can define the map $g : I \to Q$ by $g(x) = F(xm)$. Thus, by our assumption, there exists a lift of g to $G : R \to Q$ (so $G|_I = g$). From this we can construct an R-module larger than M' that also lifts

   Take the submodule $M' + Rm \subseteq M$ and define $F' : M = +Rm \to Q$, $F'(m' + rm) = F(m') + G(r)$. To show it's well defined, take $a, b \in M' + Rm$ and consider $a = b$. Then since $a = m_1' + r_1 m$ and $b = m_2' + r_2 m$, we have

$$m_1' + r_1 m = m_2' + r_2 m \quad \Rightarrow \quad (r_1 - r_2)m \overset{!}{=} m_2' - m_1' \in M'$$

   by definition of I, $(r_1 - r_2) \in I$ so

$$G(r_1 - r_2) = g(r_1 - r_2) = F((r_1 - r_2)m) \overset{!}{=} F(m_2' - m_1')$$

$$\Rightarrow F(m_1) = G(r_1) = F(m_2) + G(r_2) \Rightarrow F'(a) = F'(b)$$

thus F' is well-defined. Furthermore, F' is also a lift (that is $F' \circ \varphi = f$) since L map injectively into M' so the following diagram commutes:

$$\begin{tikzcd} L \arrow[r, hook, "\varphi|_{M'+Rm}"] \arrow[d, "f"] & M' + Rm \arrow[r, hook, "\iota"] \arrow[dl, "F'"] & M \\ Q & & \end{tikzcd}$$

If we restict $M' + Rm$ to M', then $F'|_{M'} = F$. Thus, $(M' + Rm, F')$ is a larger element than (M', F), but that contradicts maximality of (M', F). Thus, it can only be that $M' = M$, completing the proof

2. Let R be a Principal Ideal Domain and let M be an R-module. If M is injective, then to prove that M is divisibe, that is, for any $x \in M$ and any $n \in R$, there exists a $y \in M$ such that $n \cdot y = x$. choose an arbitary $x \in M$ and $n \in R$. Then $(n) = I$ is an ideal of R. Thus, if we have an R-module homomorphism $f : (n) \to Q$, where $f(n) = x$ which is extended linearly. This function is well-defined and an R-module homomorphism since for any $m \in I$, $m = m'n$, so $f(m) = f(m'n) \overset{\text{def}}{=} m' \cdot f(n) = m' \cdot x$ and $f(a+b) = f((a'+b')n) = (a'+b') \cdot x$. Thus, there is an extension to an R-module homomorphism $F : R \to Q$. Since $1 \in R$, $F(10$ is well-defined Now, notice that

$$x = f(n) = F(n) = n \cdot F(1)$$

If we let $F(1) = y$, we get $x = n \cdot y$. Since x and y ere arbitrary, $rQ = Q$ (notice that the fact that R is a Principal Ideal Domain as not used)

For the converse, the proof is very similar to (1) up to using Zorn's lemma to find (M', F) and defining the [now two sided][a] ideal to consstruct the larger module based off of an element $m \in M$ but $m \notin M'$. At this point, we can use the fact that R is a Principal Ideal Domain to say that there is an element $r \in R$ such that $(r) = I$. Note that if $r = 0$, then $m \in (0) = \{0\}$, but $m \neq 0$, so $r \neq 0$. No, notice that $F(rm)$ is defined since $rm \in M'$, but $F(m)$ is not defined since $m \notin M'$. Since M is divisible, we can have more control over what $F(rm)$ is. In particular, by chosing $x = F(rm)$ and $n = r$, by divisibility of M there exists a y such that

$$F(rm) = r \cdot y$$

From this, we can extend the homomorphism to $\overline{F} : M' + (r)m \to Q$ by

$$\overline{F}(m' + nm) = F(m') + n \cdot y$$

To check that this is indeed a R-module homomorphism, it is the same as for (1) so instead we will check that it is well-defined. It suffices to show that nm indeed maps to $n \cdot y$. Since $n \in (r)$, $n = n'r$, so

$$\overline{F}(nm) = \overline{F}(n'rm) = n' \cdot \overline{F}(rm) = n'r \cdot y = n \cdot y$$

showing it is indeed well-defined, and so we again have a larger module than the maximal module M', a contradiction.

[a] Since R is a Principal Ideal Domain, it is commutative

We can generalize the last result to integral domains, that is, if R is an integral domain and M is an R-module such that $rM = M$, then M is also injective.

Example 16.3.2: Injective Modules

1. $\mathbb{Z}$ is not an injective $\mathbb{Z}$-module, since $\mathbb{Z}$ is a Principal Ideal Domain, but not divisible for given 1 and 2, there does not exist a number x such that $2x = 1$. We can also see this by saying that the exact sequence

$$0 \longrightarrow \mathbb{Z} \xrightarrow{\cdot 2} \mathbb{Z} \xrightarrow{\pi} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0$$

   does not split. Notice that $\mathbb{Z}$ is a free $\mathbb{Z}$-module, so free $\not\Rightarrow$ injective

2. $\mathbb{Q}$ is an injective $\mathbb{Z}$-module, since for any $z \in \mathbb{Z}$, $z\mathbb{Q} = \mathbb{Q}$. It follows that any short exact sequence

$$0 \longrightarrow \mathbb{Q} \longrightarrow M \longrightarrow N \longrightarrow 0$$

   splits. Proving this without directly referencing Baer's criterion is a good exercise [hint: use a similar argument creating a partial order and use Zorn's lemma to find a (M', F). Then use properties of $\mathbb{Q}$ (in particular, $\mathbb{Q}$ is locally cyclic, meaning any subgroup generated by finitely many elements is cyclic) to show that $M' = \mathbb{Q}$]

3. $\mathbb{Q}/\mathbb{Z}$ is an injective $\mathbb{Z}$ module, being the quotient of an injective module (note that where the quotient need not be itself injective).

4. Is $\mathbb{Q} \oplus \mathbb{Q}/\mathbb{Z}$ an injective $\mathbb{Z}$-module? More generally, if A and B are injective R-modules, is $A \oplus B$ an injective R-module? (see exercise ref:HERE)

5. As can be deduced from chapter 14, no finitely generated abelian group is divisible, hence no finitely generated $\mathbb{Z}$-module is injective

6. Let $R = F$ be a field. Then every F-module is injective! (split exact sequence if and only if basis can be extended, and we proved basis can be extended)

7. (final example in dummit p. 397)

16.3.1 Example: Divisible Groups

The category $\mathbb{Z}$-Mod can be thought as the simplest R-Mod category, as though the $\mathbb{Z}$-action is just a formality (which in some strict sense is the case). Hence, classifying injective $\mathbb{Z}$-modules is a natural starting point. Since $\mathbb{Z}$-modules are abelian groups, for this section we shall switch over from $\mathbb{Z}$-module terminology to abelian group terminology

Definition 16.3.5: Divisible Group

Let G be an abelian group. Define $nG := \{ng \mid g \in G\}$. If

$$nG = G, \ \forall n \in \mathbb{N}_+$$

Then G is called a *divisible group*

If we wanted to use the multiplication notation for G, we would write $G^n = G$. Note also that n need not be unique (as we will see in the examples)

Example 16.3.3: Divisible Groups

1. The cannonical example is $\mathbb{Q}$ since for any $\frac{p}{q} \in \mathbb{Q}$ and $n \in \mathbb{N}_+$, Take $\frac{p}{nq} \in \mathbb{Q}$, then

$$\frac{p}{nq} + \frac{p}{nq} + \cdots + \frac{p}{nq} = \frac{p}{q}$$

   and since n and $\frac{p}{q}$ as arbitarry, $\mathbb{Q}$ is divisible

2. The group $\mathbb{Q}/\mathbb{Z}$ is divisible. This can be done "manually" by showing that for any choice of n and $x \in \mathbb{Q}/\mathbb{Z}$, there will be a $y \in \mathbb{Q}/\mathbb{Z}$ such that $ny = x$, however Baer's Criterion proves this alraedy. Note that $\mathbb{Q}/\mathbb{Z}$ is completely torsion, and so torsion groups can also be divisible. Note that if the group is torions, it is no longer necessary that the solution to the equation $x = ny$ is unique, that is, there can exist to distinct y and y' such that $x = ny = ny'$.

3. Let $\mathbb{Z}_{p^\infty} = \{[x] \in \mathbb{Q}/\mathbb{Z} \mid \exists\, n \in \mathbb{N} \text{ such that } p^n[x] = 0\}$ [a]. This is indeed a group, since

   (a) $[0] \in \mathbb{Z}_{p^\infty}$
   (b) if $[y] \in \mathbb{Z}_{p^\infty}$, then for some n, $p^n[y] = 0$, so $p^n[(-y)] = -(p^n[y]) = [-0] = [0]$
   (c) For $[x], [-y] \in \mathbb{Z}_{p^\infty}$, take n and m such that $p^n[x] = 0$ and $p^m[-y] = 0$. Then since the elements of $\mathbb{Q}/\mathbb{Z}$ are commutative, $0 = 0 \cdot 0 = p^n[x]p^m[-y] = -p^{n+m}[xy]$

   and so $\mathbb{Z}_{p^\infty}$ is a group. This group is divisible. Choose some $[x] \in \mathbb{Z}_{p^\infty}$ and $n \in \mathbb{N}_+$. Let's first write n as $n = p^k m$, here m is coprime to p. Next, since $[x] \in \mathbb{Z}_{p^\infty}$, there exists an $l \in \mathbb{N}$ such that $p^l[x] = 0$. Thus, let $\frac{a}{p^l}$ represent $[x]$, here a is a integer coprime to p.

   Now, since m is coprime to p, m is invertible in $\mathbb{Z}/p^l\mathbb{Z}$, that is there exists an m' such that $mm' = 0 \bmod p^l\mathbb{Z}$. Let $b = am'$ so that $\frac{b}{p^l}$ represents a number $z \in \mathbb{Z}_{p^\infty}$, and notice that $m\frac{b}{p^l} \equiv \frac{a}{p^l} \bmod \mathbb{Z}$, so $mz = x$ in $\mathbb{Z}_{p^\infty}$.

   Finally, take $y \in \mathbb{Z}_{p^\infty}$ be the element that is represented by $\frac{b}{p^l}$ so that $p^k y = z$. Then

$$ny = mp^k y = mz = x$$

   showing that $\mathbb{Z}_{p^\infty}$ is indeed divisible, as we sought to show

   Note that an equivalent way of defining $\mathbb{Z}_{p^\infty}$ is

$$Z := \left\{ \left[\frac{a}{p^n}\right] \in \mathbb{Q}/\mathbb{Z} \;\middle|\; a, n \in \mathbb{Z},\ n \geq 0 \right\}$$

   It is left as an exercise to prove this equivalence.

4. Similarly to the last example, $\mathbb{Q} \oplus \mathbb{Q}$ and $\mathbb{Q} \oplus \mathbb{Z}_{p^\infty}$ is divisible which will be proven in the next proposition.

5. No finite group can be divisible, which is easy to show (can you divide an element by its order?)

[a] Fun fact, this is $\varinjlim \mathbb{Z}/p^k\mathbb{Z}$

Note that Baer's criterion that $\mathbb{Q}/\mathbb{Z}$ is divisible, and it is easy to see that $\mathbb{Q} \oplus \mathbb{Q}$ are divisible. The direct sum can be extended finitely through the same procedure, and arbitrarily using Zorn's lemma. For the direct sum, slightly more is true: if $G \oplus H$ is divisible, then G and H are divisible. Easiest way to verify this is by considering if one of the were not divisible – say G was not divisible. So then there exists some $n \in \mathbb{N}_+$ and $x \in G$ where there is no $y \in G$ where $ny = x$. But then there exists no $z \in G \oplus H$ such that $nz = (x, 0)$ – contradiction.

With some idea of what a divisible subgroup is, we can ask ourselves if every group has a divisible subgroup, noting that it is certainly not the case that all subgroups of a divisible group is divisible (think $\mathbb{Z} \subseteq \mathbb{Q}$). Since $\{0\} \leqslant G$ is trivially a divisible subgroup, we can define the following:

Definition 16.3.6: Divisible Subgroup

Let G be a group. Then $dG \leqslant G$ is the subgroup of G generated by all divisible subgroup of G (i.e. the sum of all divisible groups [a]). If $dG = 1$, then we say that G is reduced.

[a] If we used multipliciton notation, it would be the internal direct product of all divisible groups

Thus, every group has a divisible subgroup. As a fun exercise, show that any homomorphism $f : G \to H$ sends dG to dH, and so the operator d is actually functorial.

Lemma 16.3.7: Splitting any Abelian Group

Let G be a group and dG be its divisible subgroup. Then

$$G \cong dG \oplus R$$

where R is the reduced subgroup of G

Proof **:**
Since dG is a divisible subgroup, the following short exact sequence splits

$$0 \to dG \to G \to G/dG \to 0$$

Thus, $G \cong dG \oplus H$, here H is the compliment of dG. We just need to show that H is reduced. Notice that $dH \leqslant H \leqslant G$, so dH is contained in dG. But if that's the case, then since H is the compliment of dG, $dH \leqslant dG \cap H = 0$, showing that $dH = 0$, and so H must be reduct.

Lemma 16.3.8: Torsion Group Of Divisible Group

Let G be a divisible group and $\operatorname{Tor}(G)$ be its torsion subgroup. Then $\operatorname{Tor}(G)$ is divisible. Furthermore, $G \cong \operatorname{Tor}(G) \oplus G/\operatorname{Tor}(G)$.

***Proof* :**
We'll first show that $\operatorname{Tor}(G)$ is divisible. Take $x \in \operatorname{Tor}(G)$ and choose any $n \in \mathbb{N}_+$. Then since $x \in G$ and G is divisible, there exists a $y \in G$ such that $n \cdot x = y$. Since x is torison, there exists an m st $m \cdot x = 0$. Thus:

$$(mn) \cdot y = m \cdot x = 0$$

showing that $y \in \operatorname{Tor}(G)$, and hence $\operatorname{Tor}(G)$ is divisible. Thus, $G = \operatorname{Tor}(G) \oplus H$. It is clear from here that $H = G/\operatorname{Tor}(G)$[a], since all non-torsion elements must be contained in the compliment of $\operatorname{Tor}(G)$, and so

$$G \cong \operatorname{Tor}(G) \oplus G/\operatorname{Tor}(G)$$

as we sought to show

[a] note that $G/\operatorname{Tor}(G)$ is divisible being the quotient of a divisible group

Finally, we present some structure theorems for $\operatorname{Tor}(G)$ and $G/\operatorname{Tor}(G)$

Lemma 16.3.9: Structure Of $G/\operatorname{Tor}(G)$

Let G be a group and consider $G/\operatorname{Tor}(G)$. Then $G/\operatorname{Tor}(G)$ is a $\mathbb{Q}$-vector-space

***Proof* :**
It is easy to see that $G/\operatorname{Tor}(G)$ is divisible, and by lemma 16.3.9, $G/\operatorname{Tor}(G)$ is torsion-free. The fact that it's torsion-free is important, elements will be uniquely divisible, that is, if y and y' both satisfy the condition that $x = n \cdot y = n \cdot y'$, then $n \cdot y - n \cdot y' = n \cdot (y - y') = 0$, but then $y - y'$ would be torsion, so the only possibility is that $y - y' = 0$, so $y = y'$. From this fact, we can define an action from $\mathbb{Q} \to \operatorname{Aut}_{\mathrm{Ab}}(G/\operatorname{Tor}(G))$ which takes

$$\frac{a}{b} \mapsto (x \mapsto y)$$

where $by = ax$. This is easily seen as a well-defined $\mathbb{Q}$-action, and hence a $\mathbb{Q}$-vector space, as we sought to show

As a consequence, $G/\operatorname{Tor}(G) \cong \mathbb{Q}^I$ where I is an inexing set

Lemma 16.3.10: Structure Of $\operatorname{Tor}(G)$

Let G be a group and consider $\operatorname{Tor}(G)$. Then $\operatorname{Tor}(G)$ is a direct sum of p-groups. Furthermore, each of these direct sums is isomorphic to $\mathbb{Z}_{p^\infty}$ for appropriate prime p.

***Proof* :**
Take $g \in \mathrm{Tor}(G)$. Since g is a torsion element, let

$$|g| = n = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

where each p_i is a distinct prime number and $n_i > 0$. We will make g the sum of elements from $\mathrm{Syl}_{p_i}(G)$.

Let $q_i = n/p_i$. Then it is clear that each q_i are pair-wise coprime as each p_i are pair-wise coprime. Thus, for appropriate h_i, we get

$$h_1 q_1 + h_2 q_2 + \cdots + h_k q_k = 1 \tag{16.4}$$

Furthermore, let $g_i = (h_i q_i) g$. Notice that each $g_i \in \mathrm{Syl}_{p_i}(G)$, and

$$g = g_1 + g_2 + \cdots + g_k$$

showing that every element can be split into a sum of elements from the sylow p subgroups. Since this is true for any element, and all the elements are independent (by equation 16.4), G can be split into a direct sum of Sylow subgroup, that is, primary p-groups.

For the second part of this theorem, we will first prove a quick lemma: If G is a divisible p-group, then G splits with $\mathbb{Z}_{p^\infty}$

Proof. let G be a divisible p-group, and take an element $x_1 \in G$ of order p. Then since G is divisble, choose $x_2 \in G$ such that $px_2 = x_1$. Continuing this way, we get a sequence

$$(x_1, x_2, x_3, ...)$$

where $px_{i+1} = x_i$. Let $Z = \langle x_1, x_2, ... \rangle$. Then Z is isomorphic to $\mathbb{Z}_{p^\infty}$. Since the group is abelian, it suffices to map generators bijectively map to generators. In our case, notice that

$$x_r \mapsto [1/p^r]$$

is our desired isomorphism. This should be clear by looking at the 2nd definition of $\mathbb{Z}_{p^\infty}$ in example 16.3.3[2]. Since Z is a divisible subgroup, it is a summand of G: $G = \mathbb{Z}_{p^\infty} \oplus G'$. □

To show that all of G splits into direct summands of $\mathbb{Z}_{p^\infty}$ for appropriate p, we will use Zorn's lemma [a]. Let $\mathcal{S}$ be the set of all subgroups of subgroup of G isomorphic to $\mathbb{Z}_{p^\infty}$, and let $\mathcal{T} \subseteq P(S)$ be the set of all subgroups for which the sum of the subgroups is a direct sum. Then $\mathcal{T}$ is clearly partially ordered by inclusion and every chain has an upper bound (take all the summands together). So, by Zorn's lemma, $\mathcal{T}$ has a maximal element T. Let's say $H = \oplus T_i$ is the direct sum of the groups of T. Since H is the direct sum of divisible groups, it is divisible, so H splits G, that is $G = H \oplus K$ for some compliment K. If K is a non-trivial, then K is a divisible torsion subgroup, and so by our lemma, K splits with $\mathbb{Z}_{p^\infty}$. Let's label this subgroup P, so $K = P \oplus K'$. But then, $T \subsetneq T \cup \{P\} \in \mathcal{T}$, showing that T is not maximal – a contradiction. Thus, it must be that $K = \{0\}$, and that $G = H$, as we sought to show

[a] which, tangentially, I'm starting to see as a stronger version of induction

With all the pieces solved, we simply put them together for the last theorem

Theorem 16.3.11: Classification Of Divisible Groups

Let G be a divisible group. Then

$$G \cong \left(\bigoplus_{p \in P} (\mathbb{Z}_{p^\infty})^{I_p} \right) \oplus \mathbb{Q}^I$$

here P is the set of primes uniquelely determiend by G, and I_P as ell asll I are indexing sets uniqueley determined by G

Proof **:**
We simply combine all the results we've seen so far. By lemma 16.3.8, $G \cong \operatorname{Tor}(G) \oplus G/\operatorname{Tor}(G)$, by lemma 16.3.9, $G/\operatorname{Tor}(G) \cong \mathbb{Q}^I$ for appropriate indexing set I, and by lemma 16.3.10, we have that $\operatorname{Tor}(G)$ is the direct sum of primary p-groups, in particular $\mathbb{Z}_{p^\infty}$. Combining this all get's our desired result.

16.3.2 Injective Hull

Injective modules have some properties that are dual to projective modules. For example, all projective modules are free modules, and hence every module is a quotient of a projective module, which furthermore satisfies a universal property.There is a dual result for injective modules! We will first proof a build-up lemma:

Lemma 16.3.12: $\mathbb{Z}$-module Submodules of Injective Modules

Every $\mathbb{Z}$-module is a submodule of an injective $\mathbb{Z}$-module

Proof **:**
Let M be a $\mathbb{Z}$-module, and let A be a set of $\mathbb{Z}$-module generators for M so that $F(A)$ is the free $\mathbb{Z}$-module generated by A. Then by the universal property of free modules there exists a surjective $\mathbb{Z}$-module homomorphism $F(A) \twoheadrightarrow M$. Let $K \subseteq F(A)$ be the kernel of thei homomorphism so that

$$M \cong_{\mathbb{Z}\text{- Mod}} F(A)/K$$

Now, let $Q(A)$ be the free $\mathbb{Q}$-modules over A (and *a fortiori* a $\mathbb{Z}$-module). Since $Q(A)$ is a vector-space, it is a free $\mathbb{Q}$ module, so $Q(A) = \mathbb{Q}^{\oplus A}$. Furthermore, $\mathbb{Q}$ is divisible by ref:HERE, so is $Q(A)$, and so is injective. Notice as a $\mathbb{Z}$-module, $F(A)$ is a submodule of $Q(A)$. As a consequence K is a submodule of $Q(A)$. Thus, $Q(A)/K$ is a well-defined $\mathbb{Z}$-module quotient, and $F(A)/K$ is a submodule of $Q(A)/K$. Note that $Q(A)/K$ is still a $\mathbb{Z}$-module by theorem 16.3.4. Then

$$M = F(A)/K \subseteq Q/K$$

showing that M is a submodule of an injective module, finishing the proof.

Lemma 16.3.13: Inducing Injective R-modules

Let $f : S \to R$ be a homomorphism of commutative rings, and Q and injective S-module. Then $\mathrm{Hom}_R(S, Q)$ is an injective R-module

Proof **:**
This uses the adjuncation I haven't yet presented! I ought to present this in the tensor chapter
Define:

$$f^!(M) := \mathrm{Hom}_R(S, M) \qquad f^*(M) = M \otimes_R S \qquad f_*(M) = \text{restriction of scalars}$$

BY adjunction, the followingare isomorphism as functors R-Mod $\to$ R-Mod:

$$\mathrm{Hom}_R(-, f^!(Q)) \cong \mathrm{Hom}_S(f_*(-), Q)$$

Since f_* is exact, $\mathrm{Hom}_S(-, Q)$ is exact (by hypothesis), $\mathrm{Hom}_R(-, f^!(Q))$ is exact, as needed.

Corollary 16.3.14: Inducing Injectives Using Divisible Groups

Let R be a commutative ring and D a divisible group. Then $\mathrm{Hom}_{\mathbb{Z}}(R, D)$ is injective in R-Mod.

Proof **:**
direct consequence of above.

Now generalizing to general injective R-modules:

Lemma 16.3.15: R-module Submodules of Injective Modules

Let R be a ring and let M be an R-module. Then M is contained in an injective R-module.

Proof **:**
Since:

$$M \cong \mathrm{Hom}_R(R, M) \hookrightarrow \mathrm{Hom}_{\mathbb{Z}}(R, M) \subseteq \mathrm{Hom}_{\mathbb{Z}}(R, D)$$

then since we found D in lemma 16.3.12, the proof is complete.

Definition 16.3.16: Injective Resolution

Let M be an R-module. Then a chain:

$$M \to Q^0 \to Q^1 \to \cdots$$

where each Q^0 is injective is called an *injective resolution*.

Finally, we can prove a proper dual to the universal property of free modules:

Theorem 16.3.17: Universal Property Of Injective Hulls

HERE

Proof **:**
here

(Something here relates to representation theory in D&F that sounds interesting)

(also can mention generalizing the result for divisible groups, we get every injective R-module is a direct summand of $\mathrm{Hom}_{\mathbb{Z}}(R, \mathbb{Q}/\mathbb{Z})^S$ for some set S. This can be re-framed in terms of injective hulls)

16.3.3 Duality on R-modules

Thinking of that Richard Borcherds Video. (Rings 12 Duality and injective modules) (also Aluffi in Hom and Duals, good stuff! Particularly poignant is that dualizing kills torsion for modules over integral domains)

(Duality is a special case of $\mathrm{Hom}_R(-, R)$, $(-)^\vee =:= \mathrm{Hom}_R(-, R)$)

In section 13.3, we defined the dual of a vector space V to be $V^* = \mathrm{Hom}_k(V, k)$. We found that $V \cong V^*$ and that $V \cong V^{**}$ in a natural way. Does this duality continue to work over more general modules? It is easy to see it works for free modules. For a projective module P, since there always exists a Q such that $P \oplus Q$ is free, then it is easy to check that $(P \oplus Q)^* = P^* \oplus Q^*$, and we can use this to show the isomrophism $P \mapsto P^{**}$ is natural.

It is natural to ask if this pattern will continue for non-projective modules. Take for example $R = \mathbb{Z}$ and $M = \mathbb{Z}/2\mathbb{Z}$. Then computing, we get:

$$\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z}, \mathbb{Z}) = 0$$

and so, duality fails since there are no non-trivial maps from $\mathbb{Z}/2\mathbb{Z}$ into $\mathbb{Z}$. However, if we replace the codomain $\mathbb{Z}$ with $\mathbb{Q}/\mathbb{Z}$, we get that:

$$\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/2\mathbb{Z}, \mathbb{Q}/\mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z}$$

Showing we recovered a form of duality! The module $\mathbb{Q}/\mathbb{Z}$ here is called the *dualizing module*. In the dual theory of finite dimensional vector space, the dualizing module is always the underlying field, but sometimes in more general module theory we require a more general module. More generally, we can take:

$$\mathrm{Hom}_{\mathbb{Z}}(Z/n\mathbb{Z}, \mathbb{Q}/\mathbb{Z}) \cong \mathbb{Z}/n\mathbb{Z}$$

Note that this isormophism is not natural, since for a general cyclic group C_n, there is no natural choice of generator. However, the double-dual is still naturally isomorphic. Hence, we have a theory of dual's for all abelian groups! The key for the dualizing to work was the presense of roots. We could have used S^1, $\mathbb{R}/\mathbb{Z}$, or $\mathbb{C}^\times$ as the dualizing modules too.

Example 16.3.4: Dualizing
Let $(\mathbb{Z}/8\mathbb{Z})^\times = \{1, 3, 5, 7\}$ be the units of $\mathbb{Z}/8\mathbb{Z}$ and let $\mathbb{C}^*$ be the dualizing module. Then $((\mathbb{Z}/8\mathbb{Z})^\times)^*$ will have 4 elements. To visualize these elements better, we will create a table, where the elements will be labeled χ_1, χ_3, χ_5, χ_7 due to historical precedence and it's connection to

Character Theory[a]:

	1	3	5	7
χ_0	1	1	1	1
χ_3	1	1	-1	-1
χ_5	1	-1	1	-1
χ_7	1	-1	-1	1

[a]We will character theory in more detail in chapter 24, section 24.2. In character theory, following elements of the dual are also called *Dirichelt Characters*

(The rest is in this video: commented)

16.4 Flat Modules

(geometric picture) `https://math.stackexchange.com/questions/1502957/intuition-for-when-modules-are`

The $\mathrm{Hom}(D, __)$ and $\mathrm{Hom}(__, D)$ functor's are very natural ones to analyze since they exist for any category, and the $\mathrm{Hom}(D, __)$ let's us define the derived functors which leads us to defining Ext groups. Tensoring an exact sequence (i.e. $D \otimes __$) might seem not as intuitive of an object to study. With $\mathrm{Hom}(D, __)$ and $\mathrm{Hom}(__, D)$, there were relatively types of objects which will preserve exact sequences, and allowed us to categories different structures like PID's. However, they are in fact intimitely linked to $\mathrm{Hom}(D, __)$ and $\mathrm{Hom}(__, D)$: $D \otimes __$ is *left-adjoint* to $\mathrm{Hom}(D, __)$. To undertsand what this means, it is productive to point out an example where we have already proved something earlier in the book is a left-adjoint, but haven't explicitely called it a left-adjoint: Free functors are left-adjoint to forgetful functors. What this means is

$$HERE$$

Similarly (Antoher example here)

(More words on adjoint)

Theorem 16.4.1: Tensor and Hom Adjoint

Let R and S be rings, let A be a right R-module, let B be an (R, S)-bimoudle, and let C be a right S-module. Then there is an isomorphism of abelian groups

$$\mathrm{Hom}_S(A \otimes_R B, C) \cong_{\mathbf{Ab}} \mathrm{Hom}_R(A, \mathrm{Hom}_S(B, C))$$

intuition

Proof **:**
We will show a homomorphism and inverse homomorphism exist. Let $\varphi : A \otimes_R B \to C$ be a homomorphism. From φ, we will show how to define a new map $\Phi \in \mathrm{Hom}_R(A, \mathrm{Hom}(B, C))$. Fix an $a \in A$ and defien the map $\Phi(a) : B \to C$, $\Phi(a)(b) = \varphi(a \otimes b)$. Notice that

$$\Phi(a)(b + b') = \varphi(a \otimes (b + b')) = \varphi(a \otimes b + a \otimes b') = \varphi(a \otimes b) + \varphi(a \otimes b') = \Phi(a)(b) + \Phi(a)(b')$$

and that

$$\Phi(a)(b \cdot s) = \varphi(a \otimes (b \cdot s)) = \varphi((a \otimes b) \cdot s) = \varphi((a \otimes b)) \cdot s = \Phi(a)(b) \cdot s$$

showing that $\Phi(a)$ is a right S-module homomorphism. With $\Phi(a)$ defined, we can define $\Phi : A \to \mathrm{Hom}_S(B, C)$ mapping a to $\Phi(a)$. Thus, let $f : \mathrm{Hom}_S(A \otimes_R B, C) \to \mathrm{Hom}_R(A, \mathrm{Hom}_S(B, C))$, $f(\varphi) = \Phi$. This map is indeed a group homomorphism, since

$$f(\varphi + \varphi') = \Phi + \Phi' \equiv (\Phi + \Phi')(a) = \Phi(a) + \Phi'(a) \equiv f(\varphi) + f(\varphi')$$

for all $a \in A$. Thus, f is a well-defined homomorphism.

Conversely, take $\Phi : A \to \mathrm{Hom}_S(B, C)$ to be a an R-module homomorphism in $\mathrm{Hom}_R(A, \mathrm{Hom}_S(B, C))$. We will use the universal property of tensor products. Take $\varphi' : A \times B \to C$, $\varphi'((a, b)) = \Phi(a)(c)$. Notice that

$$\varphi'((a + b, c)) = \Phi(a + b)(c) = \Phi(a)(c) + \Phi(b)(c) = \varphi'((a, c)) + \varphi'((b, c))$$

Since Φ is a homomorphism. Similarly $\varphi'((a, b + c)) = \varphi'((a, b)) + \varphi'((a, c))$ since $\Phi(a)$ is a homomorphism! Furthermore,

$$\varphi'((a \cdot r, b)) = \Phi(a \cdot r)(b) \overset{!}{=} \Phi(a)(r \cdot b) = \Phi'((a, r \cdot b))$$

where $\overset{!}{=}$ is true since B has a left R-module structure and Φ is an R-homomorphism. Thus, by the universal property of tensor prodcuts, there exists a map $\varphi : A \otimes_R B \to C$. Thus, $g(\Phi) = \varphi$ is a well-defined map, and is immediately seen to be a homomorphism by our previous resoning.

Thus, $\cong_{\mathbf{Ab}}$ well-defined, as we sought to show.

With this in mind, we see how there is some value in studying $D \otimes __$ as (More when I get an intuition on Adjoints)

(Old stuff with intuitions)

So I've thought of another intuition from math overflow from here:

flat modules and fibers

I found another source which is a teacher trying to figure out how to intuitively introduce flat and projective modules:

Mathoverflow Professor

He claims that for for a commutative ring R, and an ideal I of R, and the short exact sequence

$$0 \to I \to R \to R/I \to 0$$

and if S is a flat-module, then $I \otimes_R S = IS$.

In the last two section, we've looked at mapping short exact sequence $0 \to L \to M \to N \to 0$ to short exact sequence of $\mathrm{Hom}_R(P, -)$ or the dual $\mathrm{Hom}_R(-, Q)$. We'll now do the same with a new functor:

$$D \otimes_R - : X \to D \otimes_R X$$

which ia a *covariant functor* from the category of left R-modules to the category of abelian groups (respectively, to the category of left S-modules when D is an (S, R)-bimodule.

Note: If we consider the dual:

$$- \otimes_R D : X \to X \otimes_R D$$

it is also a covariant functor (as I think you can eventually check), and will produce the same results, with minor alterations. Thus, we'll only consider the first case.

So let's see if we already know something about this functor. Given a short exact sequence,

$$0 \longrightarrow L \longrightarrow M \longrightarrow N \longrightarrow 0$$

will the resulting sequence after applying the functor be exact:

$$0 \longrightarrow D \otimes_R L \longrightarrow D \otimes_R M \longrightarrow D \otimes_R N \longrightarrow 0$$

The answer is no. We've already seend examples where the map

$$1 \otimes \varphi : D \otimes_R L \to D \otimes_R M$$

induced by injective map $\varphi : L \hookrightarrow M$ is no longer injective (take $\mathbb{Z} \hookrightarrow \mathbb{Q}$ of Z-modules induces the zero map form $\mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z} = \mathbb{Z}/2\mathbb{Z}$ to $\mathbb{Z}/2\mathbb{Z} \otimes_R \mathbb{Q} = 0$). On the other hand, it can be shown that for any surjective map $M \twoheadrightarrow N$, we'll have an induced surjective map (TBD p.398).

As the two previous times, if we have an short exact sequence, will applying the functor make the 3rd term exact, or with our new vocabulary from category theory, be *right exact*? The answer will be yes:

Theorem 16.4.2: $D \otimes_R -$ is right exact

Suppose that D is a right R-module and L, M, and N are left R-modules. If

$$0 \to L \xrightarrow{\psi} M \xrightarrow{\varphi} N \to 0 \qquad \text{is exact}$$

then the asociated sequence of abelian groups

$$D \otimes_R L \xrightarrow{1 \otimes_R \psi} D \otimes_R M \xrightarrow{1 \otimes_R \varphi} D \otimes_R N \to 0 \qquad \text{is exact}$$

If D is an (S, R)-bimodule then the previous equation is an exact sequence of left S-modules. In particular, if $S = R$ is a commutative ring, then the prvious equation is an exact seuqence of R-modules with repesct to the standard R-module structures. The map $1 \otimes \psi$ (is there a mistake in dummits book? he sais $1 \otimes \varphi$) is not in general injective

Proof. p.399 □

As before, we ask when the functor $D \otimes_R -$ is *exact*. As you've shown in an exercice, it is quite easy to show that tensor products commute with [arbitrary] direct sums. So if the short exact sequence splits, then by theorem ref:HERE, it follows that the the:

$$0 \to D \otimes_R L \xrightarrow{1 \otimes \psi} D \otimes_R M \xrightarrow{1 \otimes \varphi} D \otimes_R N \to 0$$

also a split short exact sequence. However, I don't think it *needs* to split for the functor to be exact. The following propositions clarifies this point:

Proposition 16.4.3: conditions for $D \otimes_R -$ to be exact

Let A be a right R-module. Then the following are equivalent

($\Rightarrow$) For any left R-modules L, M, and N, if

$$0 \to L \xrightarrow{\psi} M \xrightarrow{\varphi} N \to 0$$

is a short exact sequence, then

$$0 \to A \otimes_R L \xrightarrow{1 \otimes_R \psi} A \otimes_R M \xrightarrow{1 \otimes_R \varphi} A \otimes_R N \to 0 \qquad \text{is exact}$$

($\Rightarrow$) For any left R-modules L and M, if $0 \to L \xrightarrow{\psi} M$ is an exact sequence of left R-modules (i.e. $\psi : L \to M$ is injective) then $0 \to A \otimes_R \xrightarrow{1 \otimes \psi} A \otimes_R M$ is an exact sequecne of abelian groups (i.e. $1 \otimes \psi$ is injective)

Proof. already outlined earlier □

Definition 16.4.4: Flat-Modules

A right R-module A is called *flat* if it satisfies either of the two equivalent conditions of the previous proposition

Note: It is actually a *right*-module. If the functor was reversed, i.e. $- \otimes_R D$, then it would be a *left*-module.

So now with a basic look into flat modules, why bring them up when we were looking at $\mathrm{Hom}_R(D, -)$ and the dual equivalent? Well as it turns out,

$$\text{free} \Rightarrow \text{projective} \Rightarrow \textit{flat}$$

Corollary 16.4.5: free and projective modules are flat

Free modules are flat; more generally, projective modules are flat

Proof. p.400 Not actually a difficult proof to follow. Once you prove flat, projective follows easily. □

here are multiple examples to get comfortable:

Example 16.4.1: Flat Modules

1. Since $\mathbb{Z}$ is a projective $\mathbb{Z}$-module, it is also flat. Since $\mathbb{Z}/2\mathbb{Z}$ is not a projective $\mathbb{Z}$-module, then it is not flat.

2. The $\mathbb{Z}$-module $\mathbb{Q}$ module is a flat $\mathbb{Z}$-module. TODO
3. The $\mathbb{Z}$-module $\mathbb{Q}/\mathbb{Z}$ is injective, but not flat
4. the direct sum of flat modules is flat. Thus $\mathbb{Q} \oplus \mathbb{Z}$ as a $\mathbb{Z}$-module is flat. Because of this, this module isneither projective nor injective.

We've briefly touched on why we studied the tensor products. There's in fact a deeper relation between the tensor product and Hom:

(adjointness used to be here)

this let's us do certain proofs much quicker.

Corollary 16.4.6: tensor product of projective modules

If R is commutative then the tensor product of two projective R-modules is projective

Proof. p.402 □

16.4.1 Flat Extensions

Bring in Richard Borchard!

16.5 Ext and Tor

It is finally time to look into using the above information to find some information about the number of extensions tah can be done between two modules. As of this stage of development, not enough tools are present to fully prove this result. However, I believe that introducing these ideas at this stage will give the reader a concrete appreciation for the exploration of part VII, considering that large theoretical build-up necessary to get the pay-off of homological algebra. In particular, the *derived functor* and *derived-category* that shall be studied in part VII is often touted to be rather illusive. Thus, This section is meant to serve as both an enticement to homological algebra as well as a sort of "loan" of interesting ideas that will help motivate going through the more abstract setting in which homological algebra will be developed.

Let $0 \to A \to B \to C \to 0$ be a short exact sequence. Then we have the following exact sequences after applying the 3 main functors we have been studying:

$$0 \to \mathrm{Hom}_R(A, D) \to \mathrm{Hom}_R(B, D) \to \mathrm{Hom}_R(C, D)$$

$$0 \to \mathrm{Hom}_R(D, C) \to \mathrm{Hom}_R(D, B) \to \mathrm{Hom}_R(D, A)$$

$$A \otimes_R D \to B \otimes_R D \to C \otimes_R D \to 0$$

each of these are one-sided exact. Focusing on the tensoring, we shall define a module that will fix exactness which will be denoted $\mathrm{Tor}_i^R(C, D)$:

$$\mathrm{Tor}_1^R(C, N) \to A \otimes_R D \to B \otimes_R D \to C \otimes_R D \to 0$$

In the case where N is flat, then $\mathrm{Tor}_1^R(C,N)=0$, thus giving an short exact sequence. One may ask if :

$$0 \to \mathrm{Tor}_1^R(C,N) \to A \otimes_R D \to B \otimes_R D \to C \otimes_R D \to 0$$

However this is not necessarily the case. Instead, there are two other R-modules that take place in the diagram:

$$\mathrm{Tor}_1^R(A,D) \to \mathrm{Tor}_1^R(B,D) \to \mathrm{Tor}_1^R(C,N) \to A \otimes_R D \to B \otimes_R D \to C \otimes_R D \to 0$$

In some cases, we get that one of these are zero, and the long exact sequence terminates. Other times, this keeps going, with:

$$\cdots \to \mathrm{Tor}_R^2(C,D) \to \mathrm{Tor}_R^1(A,D) \to \mathrm{Tor}_R^1(B,D) \to \mathrm{Tor}_R^1(C,D) \to A \otimes_R D \to B \otimes_R D \to C \otimes_R D \to 0$$

The collection of $\mathrm{Tor}_i\, R(-,D)$ are known as *derived functors* for the tensor. They shall serve as the modules that "capture" the lack of exactness after tensoring by a module. But what is $\mathrm{Tor}_R^i(-,D)$? Concretely, for any R-module M, to find $\mathrm{Tor}_R^i(M,D)$, first find a free resolution of M[5]

$$\cdots \to R^{\oplus S_2} \to R^{\oplus S_1} \to R^{\oplus S_0} \to M \to 0$$

Getting rid of M and tensoring the sequence and noting that $R^{\oplus S} \otimes_R N \cong N^{\oplus S}$:

$$\cdots \xrightarrow{d_3} N^{\oplus S_2} \xrightarrow{d_2} N^{\oplus S_1} \xrightarrow{d_1} N^{\oplus S_0} \xrightarrow{d_1} 0$$

where at each module, we have $\mathrm{im}(d_{i+1}) \subseteq \ker(d_i)$, but not necessarily $\mathrm{im}(d_{i+1}) = \ker(d_i)$. Now define:

$$\mathrm{Tor}_R^i(M,N) = \frac{\ker(d_i)}{\mathrm{im}(d_{i+1})} := H_i(M_\bullet \otimes_R N)$$

The is invited to check that these will indeed create a long exact sequence. This same procedure could have been done with $D \otimes_R -$ which will produce a naturally isomorphic result. Furthermore, *any* choice of free resolution would produce the same result! This means that the choice made when choosing the free resolution will be independent of the given Tor values, and hence we are free to choose the simplest resolution we like. This is a rather powerful result, and will be proved in part VII[6]. Notice that $\mathrm{Tor}_R^0(M,N) = M \otimes_R N$, hence "recovering" the dropped M.

The same procedure with the appropriate alterations can be done with $\mathrm{Hom}_R(-,D)$ and $\mathrm{Hom}_R(D,-)$, the former using a projective resolution and the latter an injective resolution. These produce the followed derived functors for hom:

$$\mathrm{Ext}_R^i(D,-) \qquad \mathrm{Ext}_R^i(-,D)$$

which fit into the following long exact sequence:

$$\begin{tikzcd}
0 \arrow[r] & \mathrm{Hom}_R(D,A) \arrow[r] & \mathrm{Hom}_R(D,B) \arrow[r] & \mathrm{Hom}_R(D,C) \arrow[dll, "\delta_0"] \\
 & \mathrm{Ext}_R^1(D,A) \arrow[r] & \mathrm{Ext}_R^1(D,B) \arrow[r] & \mathrm{Ext}_R^1(D,C) \arrow[dll, "\delta_1"] \\
 & \mathrm{Ext}_R^1(D,A) \arrow[r] & \mathrm{Ext}_R^1(D,B) \arrow[r] & \mathrm{Ext}_R^2(D,C) \arrow[r] & {}
\end{tikzcd}$$

[5]In fact, a flat resolution will do, but such a technicality is of no importance until the proper theory of homological algebra

[6]see section ref:HERE

and:

$$\begin{tikzcd}
0 \arrow[r] & \mathrm{Hom}_R(C,D) \arrow[r] & \mathrm{Hom}_R(B,D) \arrow[r] & \mathrm{Hom}_R(A,D) \arrow[dll, "\delta_0'"] \\
& \mathrm{Ext}^1_R(C,D) \arrow[r] & \mathrm{Ext}^1_R(B,D) \arrow[r] & \mathrm{Ext}^1_R(A,D) \arrow[dll, "\delta_0'"] \\
& \mathrm{Ext}^1_R(C,A) \arrow[r] & \mathrm{Ext}^1_R(B,D) \arrow[r] & \mathrm{Ext}^2_R(A,D) \arrow[r] &
\end{tikzcd}$$

for suitable homomorphisms δ_i, δ_i' that shall be defined in part VII. Notice that the "same" Ext functor was used in both cases, and indeed we shall see that $\mathrm{Ext}^i_R(-,-)$ is compatible with the $\mathrm{Ext}^i_R(D,-)$ and $\mathrm{Ext}^i_R(-,D)$. As before, if $D = P$ is projective (resp. $D = Q$ is injective), we get that this long exact sequence terminates immediately, namely:

$$\mathrm{Ext}^1_R(P,-) = 0 \qquad \mathrm{Ext}^1_R(-,Q) = 0$$

With this build-up, we may finally state what we alluded to: the group $\mathrm{Ext}^1_R(M,N)$ is in bijection with the set (of equivalence classes) of all extensions of M by N:

(Aluffi exercises! Not sure if I should put the solution here, p.556)

16.6 Summary

There's a lot in this section, and it's useful to put the pertinent functors and examples in one place. Each funtor $\mathrm{Hom}_R(A,-)$, $\mathrm{Hom}_R(-,A)$ and $A \otimes_R -$, map left R-modules to abelian groups; the function $- \otimes_R A$ maps right R-modules to abelian groups. WHen R is commutative all four function map R-module to R-modules.

1. Let A be a left R-module. The functor $\mathrm{Hom}_R(A,-)$ is covariant and left exact; the module A is projective if and only if $\mathrm{Hom}_R(A,-)$ is exact (i.e. also right exact)
2. Let A be a left R-module. The functor $\mathrm{Hom}_R(-,A)$ is contravariant and right exact; the module A is injective if and only if $\mathrm{Hom}_R(-,A)$ is exact (i.e. also left exact)
3. Let A be a left R-module. The functor $A \otimes_R -$ is [covariant and right exact; the module A is flat if and only if $A \otimes_R]-$ is exact (i.e. also left exact)
4. Let A be a left R-module. The functor $- \otimes_R A$ is covariant and right exact; the module A is flat if and only if $- \otimes_R A$ is exact (i.e. also left exact)
5. projective module are flat. The $\mathbb{Z}$-module $\mathbb{Q}/\mathbb{Z}$ is injective, but not flat. The $\mathbb{Z}$-module $\mathbb{Z} \oplus \mathbb{Q}$ is flat, but neither projective of injective.

Part IV

Fields

When studying classical algebra, a lot of time and energy is dedicated to solving equations. Famously, for any equation of the form $ax^2 + bx + c = 0$ has solution

$$x = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

which was known since at least the time of ancient egypt (though not in the current algebraic form[7])After a long gap, _____discoverd the equation for the cubi equation in terms of x by using an "imaginary" number i as a trick to solve the problem of $\sqrt{-n}$ coming up (it will take some more time before i is considered a number and not simply a tool to solve equations). Using this equation, a solutoin for quartic equation with respect to x was also found. The mission was on in finding a solution to the quintic equation. However, over 200 years, no such equation was found. Abel and _____have in fact proved that only using $+, -, \times, \div$ and $\sqrt{\ }$, it is *impossible* to iscolate x!

This is the origin of the study of field theory and Galois theory, and will be the driving motivation in this part of the book, however the study of fields have in modern times moves past this and are now used as a primary tool in the study of algebraic number theory. In chapter 22, we will use the language of field theory as the language of algebraic number theory.

(Another thing! This is a thing that makes primes "special". The field operation of reversing squaring by yourself does not surject onto the rationals, namely since it misses *every prime power*, and hence misses any number is not a perfect square. It misses prime's because the square root of a prime number is *irrational*, showing that in some fundamental way, you cannot go "backwards" with prime numbers!)

[7]It was a geometric proof involving fitting squares of different sizes together

17

Field Theory

A reminder that a field F is a unital commutative division ring (definition 9.1.7). Another way of thinking of a field is that it's a set with two binary operations both acting like abelian groups. Yet another way of seeing them is that $F^\times = F - \{0\}$ is an abelian group. In this chapter, we'll be mostly thinking about fields as unital commutative divsion rings, in particular, we're going to devlop how the trivial ideal structure of fields (proposition 9.4.6) lets us uniquely define some extensions of a field. We will denote $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$ and usually use the letters k and K to reffer to fields.

Furthermore, the complex numbers $\mathbb{C}$ will be referenced for it's nice algebraic properties (ex. $i^2 = 1$ or $z = a + bi$ for any $z \in \mathbb{C}$). If the reader is unfamiliar with these notion, there are many resources online explaining these concepts[1].

17.1 Properties of Fields

First, all fields are rings, and so we can form the category **Fld** with fields as objects and ring homomorphisms as morphisms. We will use the term field homomorphsisms sometimes since the objects of study are fields with the knowledge that they are just ring homomorphisms. In general, the term ring homomorphism will be reserved for when rings will be brought into the discussion as will sometimes happen. Since the objects are fields, it is important to remember that all homomorphisms must be *injective* (see corollary 9.4.7)

[1] A famous book explaining these notions is *Althor's Complex Analysis*, in which the first chapter explores complex numbers

Proposition 17.1.1: Field Homomorphisms

Let $\varphi : k \to K$ be a ring homomorphism of fields. Then φ is either identically 0 or is injective, so that the image of φ is either 0 or isomorphic to k. Since the set 0 is not a field (by definition, $0 \neq 1$), φ in **Fld** can only be injective.

Proof **:**
This is proposition 9.4.6 (2) updated so that we only care about fields.

Thus, If a homomorphism between fields $\varphi : k \to K$ exists, $\varphi(k) \cong k$. In this way, K can be thought of as extending k:

Definition 17.1.2: Field Extension

Let k and K be fields. If there exists a homomorphism $\varphi : k \to K$ (which as we've shown must be injective), then we say that K *extends* k, or K is a *field extension* of k. We will denote this by $k \subseteq K$, K/k, or

$$\begin{array}{c} F \\ | \\ K \end{array}$$

the field from which we are extending is sometimes called the *base field* or *ground field*.

Note that since fielsd are commutative, field extension $\varphi : k \to F$ makes F a k-algebra.

Example 17.1.1: Field Extensions

1. For any field F, F/F is trivially a field extension, since the identity map is a field homomorhpism.

2. Perhaps the most famous field extension is that of $\mathbb{C}/\mathbb{R}$, since $\mathbb{R}$ is directly embedded into $\mathbb{C}$. A perhaps easy mistake to make when first learning about field extensions is that $\mathbb{R}^2/\mathbb{R}$ is not a field extension since $\mathbb{R}^2$ is simply not a field. On the other hand, $\mathbb{C}/\mathbb{Q}$ is also a field extension, since the natural inclusion map $\iota : \mathbb{Q} \to \mathbb{C}$ is a field homomorphism. Similarly, $\mathbb{R}/\mathbb{Q}$ is a field extension.

3. In example 9.1.6, we saw that $\mathbb{Q}(\sqrt{D})$ is a field where D is not a perfect square (i.e. there does not an element $d \in \mathbb{Q}$ s.t. $d^2 \in \mathbb{Q}$). For a concrete examples, let's say $D = 2$. Then $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ is a field extension. If $D = -1$, then we usually write $\sqrt{-1}$ as i, and we get that $\mathbb{Q}(i)/\mathbb{Q}$ is a field extension. We also have that $\mathbb{C}/\mathbb{Q}(i)$ is a field extension, with the map $\iota : \mathbb{Q}(i) \to \mathbb{C}$ given by mapping $a + bi \mapsto a + bi$. Notice that the map $a + bi \mapsto a - bi$ is also a valid field homomorphism, showing that there need not be a unique map.

   We will soon find a way to define $\mathbb{Q}(\sqrt{D})$ without needing to reference set-builder notation.

4. Let k be a field and $k(x) = \mathrm{Frac}(k[x])$ be the field of fractions of $k[x]$. Then $k(x)/k$ is a field extension.

The notation K/k may be read as "K over k". It does *not* represent the quotient K by k. This can be consider an overload of notation, but since the kernel of a ring homomorphism between fields must *always* be trivial, it is safe to see that K/k would not represent a quotient of the field K by k.

Another thing we can observe is that every field homomorphism $\varphi : F \to K$ can be extended into a ring homomorphism $f^{\varphi} : F[x] \to K[x]$ through

$$\sum_i a_i x^i \mapsto \sum_i \varphi(a_i) x^i$$

the fact that this is a ring homomorphisms comes straight from the properties of φ. Thus, we can consider this process to be a functor $f^{--} : \mathbf{Fld} \to \mathbf{Ring}$.

If F/k is any field extension, then we can naturally make F into a *vector-space over* k by making k act on F by multiplication (since $k \subseteq F$, this action is well-defined). Since every field has a prime subfield, every field is at least a vector-space over it's prime-subfield. The dimension of the vector-space is given a name:

Definition 17.1.3: Degree

The *degree* (or *relative degree* or *index*) of a field of extension F/k, denoted $[F : k]$, is the dimension of F as a vector-space over k (i.e. $[F : k] = \dim_k(F)$). The extension is said to be *finite* if $[F : k]$ is finite and is said to be *infinite* otherwise.

Definition 17.1.4: Finite Extension

Let F/k be a field extension. Then F/k is called a *finite extension* if $[F : k]$ is finite, and an *infinite extension* otherwise.

In example 17.1.1, $\mathbb{C}/\mathbb{R}$ is an extension of degree 2, the basis of $\mathbb{C}$ over $\mathbb{R}$ consiss of $\{1, i\}$, and hence a finite extension. The extension $k(x)/k$ is an extension if infinite degree: If it were finite, then any rational function can be written as a sum of a finite number of polynomials up to a constant $c \in k$, but then this collection must have a highest degree term, say n. Then any polynomial of degree $n + 1$ cannot be written as a linear combination of the basis – a conradiction. Hence, it is an infinite extension. Note too that as was demonstrated in theorem 13.1.4, every vector-space has a basis, and so any element $x \in F$ can be represented as a finite linear combination:

$$a_1\beta_1 + a_2\beta_2 + \cdots + a_n\beta_n \qquad a_i \in k,\ \beta_i \in \mathcal{B}$$

We will soon show how to find such a basis.

Since F/k can also be represented as a vector-space over k, we might ask if any field homomorphism/isomorphism can is also a linear transformation. The answer is yes, if we are a bit careful. Let's say that F/k and L/k are two fields extending k, and $\varphi : F \to L$ is a field homomorphism/isomorphism. We will look for a condition under which φ is also a k-vector space homomorphism. Clearly, linearly comes for free. For homogeneity of the scalar, we see that for any $r \in k$ and $m \in F$:

$$\varphi(rm) = \varphi(r)\varphi(m) \overset{!}{=} r\varphi(m)$$

where $\overset{!}{=}$ is the condition we need so that φ is a k-vector space homomorphism. In particular, this conditions say's that for any element $r \in k$, $\varphi(r) = r$. This is stronger than saying φ maps

k to itself (i.e. $\varphi(k) = k$), rather it is saying that φ is the *identity* on k. Since such field homomorphisms/isomorphsisms are also vector-space homomorphisms, they are given a name:

Definition 17.1.5: Fixed Homomorphisms

Let F/k and L/k be two field extensions of k and $\varphi : F \to L$ be a field homomorphism (resp. isomorphism). Then if φ is the identity on k, φ is said to be a *k-homomorphsm* (resp. *k-isomorphism*). If there exists a k-isomorphisms $A \to B$, then we'll say that A is k-isomorphic to B and write $A \cong_k B$.

Note that a fixed homomorhism is in fact a k-algebra homomorphism: if L/k, then there is a natural k-module structure on L given by multiplication.

As was shown in section 13.2, a linear transformation is determined by where it maps it's basis to. Hence, if we have a k-homomorphism, then can deduce where the element map (or resp. construct a k-homomrophism) by finding where the basis maps to (resp. by only choosing where the basis maps to). Even better, this linear transformation respects the multiplicative structure of F (that is $\varphi(ab) = \varphi(a)\varphi(b)$), that if a basis β can be generated multiplicatively and additively from a set S, $\langle S \rangle = \beta$, then a k-homomorphism is only determined by where it sends S. Even better, this makes any k-homomorphism a k-algebra homomrophism, and a field F that extends k (F/k) is a k-algebra.

One more word of warning is in order: Let's say we find two field extensions A/k, B/k such that $A \cong_k B$. Then this implies that $\dim_k A = \dim_k B$, however it does not imply $A = B$. This might seem obvious, however it is an mistake that is easy to make after working with field extensions for awhile (this nuance will be particularly important in theorem 17.5.12).

In most of this and the next chapter, we will focus on finite extensions. Some results for extensions of infinite degree will arise as a natural generalization of the results of finite extensions, however finiteness is used in some key proofs to construct bijections (most importantly, in proposition 18.1.8, TBD). We will touch on infinite extensions in TBD (don't yet know when they will become central, i.e. when transcendental degrees will come in play).

Characteristic

The following is is perhaps the simplest invariant of a field k (that is, if $\varphi : k \to K$ is a homomorphism, the property we're about to discuss is preserved). Every field has a multiplicative identity $1 \in k$. Since k is a field, it is also a ring, and hence there exists a unique ring homomorphism from $\mathbb{Z}$ to k mapping 1 to 1 (see proposition 9.1.17):

$$\mathbb{Z} \to k,\ n \mapsto \begin{cases} \underbrace{1 + 1 \cdots + 1}_{n \text{ times}} & n > 0 \\ -(\underbrace{1 + 1 \cdots + 1}_{n \text{ times}}) & n < 0 \\ 0 & n = 0 \end{cases}$$

We will use multiplication notation $n \cdot 1$ as a shorthand [2]. Recall $\mathrm{im}(\mathbb{Z})$ of the natural homomorphism is always the smallest subring of k (exercise ref:HERE (ex:ringCatExercise)). Recall further from exercise ref:HERE (ibid) the following definition:

[2] Note that we can also have defined it as a $\mathbb{Z}$-module on k

Definition 17.1.6: Characteristic

The *characteristic* of a field F, dentoed $ch(F)$, is the order of $\operatorname{im}(\mathbb{Z})$. If $|\operatorname{im}(\mathbb{Z})| < \infty$, we say that F has a *finite characteristic*, and characteristic 0 otherwise. In other words, it is defined as the smallest positive integer n such that $n \cdot 1 = 0$ if such a n exists, and is defined to be 0 otherwise

If n exists, then n needs to be prime. If it weren't, then

$$n = n \cdot 1 = (a \cdot 1)(b \cdot 1) = 0$$

meaning either $a \cdot 1$ or $b \cdot 1$ must be zero-divisors – a contradiction since a field has no zero-divisors! Notice that we only need to have an integral domain for the previous proposition, and that the characteristic of the integral on an integral domain will be the same as it's field of fractions. Since $\mathbb{Z}/p\mathbb{Z}$ is a field (since $\mathbb{Z}/p\mathbb{Z}^\times = \mathbb{Z}/p\mathbb{Z} - \{0\}$, see exercise ref:HERE (ex:introToRings)). However, $\mathbb{Z}$ is not a field. Using theorem 9.5.1, we know the smallest field containing $\mathbb{Z}$ would be $\mathbb{Q}$! In both cases, the field this map injects to must be the *smallest* field that contains 1 (which can also be thought of as the field *generated* by 1 in k)

Definition 17.1.7: Prime Subfield

The *prime subfield* of a field k is the subfield k generated by the multiplicative identity 1_F of k. It is (isomorphic to) either $\mathbb{Q}$ (if $\operatorname{ch}(k) = 0$) or k_p (if $\operatorname{ch}(k) = p$).

Note: A shorthand of writing the characteristic of a field is the following: if $p \cdot 1 = 0$, then it is sometimes written $p = 0$ to mean a field with characteristic p.

Notice that for a field k of characteristic n ($n \geq 0$), $k/\mathbb{F}_p$ (resp. $k/\mathbb{Q}$) is always a field extension, since we can always map $\mathbb{F}_p$ or $\mathbb{Q}$ to k through the identity map.

Note that for fields of characteristic 0, field homomorphism are automatically a vector-space homomorphism over $\mathbb{Q}$: linearity holds automatically, and for any scalar $c \in \mathbb{Q}$:

$$\varphi(cx) = \varphi(c)\varphi(x) \overset{!}{=} c\varphi(x)$$

where $\overset{!}{=}$ comes from how a field homomorphism must map $\mathbb{Q}$ to itself.

The claim at the beginning of this section should now be checked, namely that if $\operatorname{ch}(k) = n$, $n \geq 0$, and $\varphi : k \to K$ is a homomorphism, then $\operatorname{ch}(K) = n$ (see exercise ref:HERE). Since charactersitic is preserved, we can take the subcategory of $\mathbf{Fld}_0$ or $\mathbf{Fld}_p$ where we restric to fields of characteristic p or 0. We won't really do this, however it is inisghtful that doing so will now make these categories have initial objects, namely the respective prime subfield (can you prove this? Can you see why $\mathbf{Fld}$ has no initial objects?). More generally, since homomorphisms can only be injective, we can take $\mathbf{Fld}_k$ to be the category ring homomorphisms as morphisms and all fields that extend k as objects! In most of the following studies, we will fix a field k and study extensions from it, and so field theory can be thought of as studying particular $\mathbf{Fld}_k$ categories

Torsion in Fields

As we have shown in theorem 11.3.10, $F^\times$ is locally cyclic. We recall the theorem here for convinience:

Theorem 17.1.8: $F^\times$ is Locally Cyclic

A finite subgroup of the multiplicative group of a field is cyclic. In particular, if F is a finite field, then the multiplicative group $F^\times$ of nonzero elements of F is a cyclic group.

Note that this does not mean that for every element $a \in F$, there exists a power n such that $a^n = 1$. For exmaple, in $\mathbb{R}$, no power of 2 will result in $2^n = 1$. If this is the case for an element, we will give it a name:

Definition 17.1.9: Root of Unity

Let F be a field and $r \in F$. If there exists an $n \in \mathbb{N}$ such that $r^n = 1$, then we call r an *nth root of unity*.

Example 17.1.2

In $\mathbb{C}$, all elements of the form $e^{\frac{2\pi i k}{n}}$ for $0 \leq k \leq n-1$ are nth roots of unity. In fact, they are all the roots of unity of $\mathbb{C}$. The real numbers $\mathbb{R}$ have only 2 roots of unity of degree 1 and 2 respectively: ± 1.

Notice that not all $e^{\frac{2\pi i k}{n}}$ are generators for the cyclic subgroup they belong to. The generators of these cyclic subgroups are given a name:

Definition 17.1.10: Primitive Root Of Unity

Let F be a field. Then a *primitive nth root of unity* is a generator of the cyclic multiplicative subgroup of all nth roots of unity.

We will return to the study of the roots of unity later; they will be the "building blocks" for [finitely generated] abelian groups in field theory (and Representation theory)[3]

Exercise 17.1.1

1. (Do some exercises showing relations beween roots of unity and other identites? Ex. I think $\zeta_8 + \zeta_8^7 = \sqrt{2}$. Important later. Also elaborate on $\mathbb{C}$ for all future purposes, since it comes up a lot)
2. (Analysis Problem) Show that the roots of unity are dense in $S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$.

Generating Fields

Let's say $k \subseteq F$ is a subfield of F, and take some subset $A \subseteq F$. We can construct a field using k and A in the same way as we have done all the way back in section 2.3.2: Take all subfield of F containing k and A, and intersect them. Similarly to section 2.3.2, this is equivalent to "closing" the field k with A by taking all possible combinations. There is always at least exists the field K with

[3] Remember by the fundamental theory of finitely generated abelian groups that any finitely generated abelian group is the direct product of a free part $\mathbb{Z}^r$ and a torsion part? Notice how the roots of unity can represent any cyclic subgroup?

this which contains this property. Since the intersection of fields is a field (this should be done as an exercise), then we can intersect all fields with this property to get the smallest such subfield.

Definition 17.1.11: Field Generated By Elements Over F

Let F/k be a field extension and S be a set of elements of F. Then the smallest subfield of F containing both k and the elements of S is denoted $k(S)$ and is called *the field generated by S over k*. If S is finite so that $S = \{\alpha_1, \alpha_2, ..., \alpha_n\}$, we usually denote $k(S)$ by $k(\alpha_1, \alpha_2, ..., \alpha_n)$ and call it the field *generated by $\alpha, \alpha_2, ..., \alpha_n$ over k*.

If we have a field extension F/k, and F can be generated by a single element, then we call that a simple extension:

Definition 17.1.12: Simple Extension

Let F/k be a field extension. Then if there exists an element $\alpha \in F$ such that $F = k(\alpha)$, we say that F is a *simple extension*, and call α the *primitive element* of the extension. The field $k(\alpha)$ is sometimes called "k adjoin α".

Be mindful to not confuse the terms "finite generation over k", "finite type k-algebra" [4], or "finite degree over k" (for example, $k(t)/k$ is a simple extension, but certainly not finite as was discussed under the definition of degree).

A similar definition works if we broaden to rings: instead of taking the intersection of all fields containing α and F, we take all rings containing α and F. It is easy to generlize the proof that the intersection of fields is a field to rings. We will usually denote this smallest ring $F[S]$, and if $S = \{\alpha_1, \alpha_2, ..., \alpha_n\}$ is finite, then $F[S] = F[\alpha_1, \alpha_2, ..., \alpha_n]$.

Proposition 17.1.13: Characterizing $F(S)$ and $F[S]$

Let K/F be a field extensions, $S \subseteq K$, and consider both $F(S)$ and $F[S]$. Then:

1. $F[S]$ can be characterized as the set of all finite linear combinations with coefficients in K of finite products of powers of elements of S, that is, polynomials with variables taken from S
2. $F(S)$ can be characterized as the set of all $ab^{-1} \in K$ with $a, b \in F[S]$ with $b \neq 0$. Furthermore, $F(S) \cong \mathrm{Frac}(F[S])$

Note that in the "polynomials" mentioned above, the varaibles S are *not* indeterminates, since they might have relations in K (for example, it might be that for some $s \in S$, $s^2 + 1 = 0$)

Proof **:**
These proofs are similar to those of section 2.3.2 with appropriate modifications for the added operation of multiplication and inversing.

[4] It is important to point out that there is actually a deep connection between those two concepts which we'll explore in part V. As foreshadowing, try seeing if you can think of a finitely generated field extension F/k that is not a finite-type k-algebra (or vice-versa). The answer to this question is given by Zariski's Lemma

To close off this section, we introduce a generalization to notion of the smallest field K/k generated by some elements to the smallest field generated with all the elements of another field:

Definition 17.1.14: Composite Field

Let K_1 and K_2 be two subfield of K. Then the *composite field* of K_1 and K_2, denoted K_1K_2, is the smallest subfield of K containing both K_1 and K_2. Similarly, the composite of any collection of subfields of K is the smallest subfield containign all the subfields.

Note that like in section 2.3.2, we can describe K_1K_2 as the intersection of all subfields contianing K_1 and K_2 (or any number of subfields).

Example 17.1.3: Composite Fields

The composite field $\mathbb{Q}(\sqrt{2})\mathbb{Q}(\sqrt[3]{2})$ is $\mathbb{Q}(\sqrt[6]{2})$, since thie field contains both these fields, and conversely any field containing both $\sqrt{2}$ and $\sqrt[3]{2}$ contains their quotient, namely $\sqrt[6]{2}$

Exercise 17.1.2

1. Sometimes, we want to talk about field extensions without an underlying field. Let E, F be two fields of the same characteristic, say p. Then $E \otimes_{\mathbb{F}_p} F$ (resp. $E \otimes_{\mathbb{Q}} F$) as a $\mathbb{F}_p$ (resp. $\mathbb{Q}$)-algebra is a field if and only if $[EF : \mathbb{F}_p] = [E : \mathbb{F}_p][F : \mathbb{F}_p]$ (resp. $[EF : \mathbb{Q}] = [E : \mathbb{Q}][F : \mathbb{Q}]$)

17.2 Field Extensions

We now continue our exploration of field extensions. As we saw in proposition 17.1.13, any field K/F such that $K = F(S)$ is the field of fractions of all "polynomials" with variables S and coefficients F. In particular, these polynomials have "solutions" or "roots" characterizing their relations. Similarly to how generators in a group where characterized by a monomial, (for example, D_4 is characterized with the generators a, b satisfying the monomails $ab^{-1}ab = 1$ and $a^2 = b^4 = 1$), generators in fields are characterized by polynomial relationships (for example, $\mathbb{C}/\mathbb{R}$ is characterized by being generated by $\mathbb{R}$ and i where $i^2 = -1$). Just like how any relation of elements within a group can be shifted so all the elements are on the left hand side (ex. writing $a^{-1}bab = 1$ instead of $ab = b^{-1}a$), we can do the same with polynomial relationships (ex. $i^2 + 1 = 0$). In proposition 17.2.12, we will show why it only matters to characterize polynomials, and not rational polynomials (i.e. elements of $p(S)/q(S)$)

As we've seen in chapter 11, every polynomial $p(x) \in F[x]$ can be broken down into irreducible polynomials. We will start finding an easy way of extending any field k which will give us a large class of field and field extensions to work with using irreducible polynomials (recall that $x^2 + 1$ has no solutions in $\mathbb{R}$, but does in $\mathbb{C}$, so is there an extension from $\mathbb{R}$ to $\mathbb{C}$?). We will find that the method used to construct fields will almost create "universal fields" (the same way free groups are sort of the universal groups since any group is the quotient of a free group). This endevaour to create a universal property will be very close in obtaining a universal property for fields, however it will fail in one important way. Due to that failure, it will be important carry around some more information to take into account the lake of universality. In fact, the lack of universal property may be seen less as a failure and more as a peek into the further structure that underly field extensions. Figuring all of this out is the goal of this section

Field Extensions using Polynomial

If we take our field to be $\mathbb{R}$, then $x^2 + 1 = 0$ does not have a solution in $\mathbb{R}$. The question then is whether there is a larger field K for which $x^2 + 1 = 0$ *has* a solution. More generally, given any irreducible polynomial $p(x) \in k[x]$, does there exists a field K such that K contains a root of $p(x)$[5]?. The answer is yes! We actually have all the tools already to answer this question!! I suggest to take a moment and try to find the larger field K which satisfies this!!

44

Taking a moment

Theorem 17.2.1: Kronecker's Theorem

Let k be a field and let $p(x) \in k[x]$ be an irreducible polynomial. Then there exists a field K containing an isomorphic copy of k in which $p(x)$ has a root. Identifying k with this isomorphic copy through i shows that there exists an extension of k in which $p(x)$ has a root. Furthermore, if L/k is a field extension in which $p(x)$ has a root, then there exists a k-homomorphism $j : K \to L$ such that

$$\begin{tikzcd} K \arrow[r, hook, "j"] & L \\ k \arrow[u, hook, "\iota"] \arrow[ur, hook, "i"'] & \end{tikzcd}$$

commutes. Due to this, the field K produced in the following construction is sometimes called the *Kronecker field for* $p(x)$ *over* k.

If in the Kronecker construction we use a *reducible* polynomials, then the resulting object is no longer a field (it contains zero-divisors), but remains to be a k-algebra.

Proof **:**
Consider

$$K = k[x]/(p(x))$$

Since $k[x]$ is a Principal Ideal Domain (corrollary 11.1.2) and since $p(x)$ is irreducible (so prime by proposition 10.1.23), then $p(x)$ is maximal (proposition 10.1.22). Then $k[x]/(p(x))$ is a field by proposition 9.4.11. If we take the cannonical projection $\pi : k[x] \to K(= k[x]/(p(x)))$, then we can restrict it to F ($\pi|_F : F \to K$) to get a homomorphism from F to K. Since this cannot be identically 0 or else K would be the zero field, so by proposition 17.1.1, F maps injectively into K, so $\varphi(F)(\cong F)$ is an isomorphic copy of F contained in K. We identify F with its isomorphic image in K and view K as a *subfield* of K. If $\overline{x} = \pi(x)$ denotes the image of x in the quotient K, then

$$\begin{aligned} p(\overline{x}) &= \overline{p(x)} && \text{since } \pi \text{ is a homomorphism} \\ &= p(x) \bmod (p(x)) && \text{in } k[x]/(p(x)) \\ &= 0 \end{aligned}$$

[5]We can assume that $p(x)$ is irreducible in $F[x]$, since if any factors of $p(x)$ has a root, then so does $p(x)$

so that K does indeed contain a root of the polynomial $p(x)$! Thus, K is an extension of F in which the polynomial $p(x)$ has a root

For the second assertion, let's say L/k has a root to $p(x)$, say $p(\alpha) = 0$. Then the evalutation map $\text{ev}_\alpha : k[x] \to L$ (which we have proved is a ring-homomorphism in theorem 9.2.3), $g(x) \mapsto g(\alpha)$ must map $p(x)$ to 0. Hence, $p(x) \in \ker(\text{ev}_\alpha)$ meaning $(p(x)) \subseteq \ker(\text{ev}_\alpha)$, and by maximality of $(p(x))$ $(p(x)) = \ker(\text{ev}_\alpha)$. Since ideals correspond one-to-one with homomorphisms, we have a homomorphism:

$$j : \frac{k[x]}{(p(x))} \to L$$

and since $\frac{k[x]}{(p(x)} \cong K$, we have:

$$j : K \to L$$

Furthermore, for any $r \in k$, $j(r) = r$. Thus j clearly commutes in the aformentioned diagram, completing the proof.

From now on, when we talk about irreducible polynomial, we will specifically mean *non-linear irreducible polynomials*, since linear irreducible polynomials produce trivial field extensions (i.e. k being extended to sovle for $x - \alpha \in k[x]$ is just k) unless stated otherwise.

The field extension to the Kronecker field K/k is *almost* universal with respect to finding roots to the polynomial $p(x)$. If it were universal, then if there was another field extension F/k such that F has a root of the irreducible polynomial $p(x) \in k[x]$, there would exist a *unique* homomorphism $j : F \to K$ such that the diagarm commutes. However, the homomorphism j need not be unique.

Example 17.2.1: Lack Of Uniqueness

Take $K = \mathbb{Q}[x]/(x^2 - 2)$ and $\mathbb{R}$. Then the polynomial $x^2 - 2$ has a root in $\mathbb{R}$, hence we know there exists a homomorphism:

$$K = \mathbb{Q}[x]/(x^2 - 2) \hookrightarrow \mathbb{R}$$

However, there exists two homomorphisms mapping K to $\mathbb{R}$. Since K extends $\mathbb{Q}$, any homomorphism can be represented by where the roots map: we can map $1 \mapsto 1$ and $\overline{x} \mapsto \sqrt{2}$, or map $1 \mapsto 1$ and $\overline{x} \mapsto -\sqrt{2}$. Both of these are valid homomorphisms, and so adjoints are not universal with respect to roots.

The universality can be recovered if we keep track of which root we want to map to (ex. $\sqrt{2}$), however we would have to make sure that there is only 1 homomorphism given some root (how do we know in more general cases that there can't be 2 homomorphisms mapping $\overline{x}$ to the same root?). This concern is solved soon in theorem 17.2.13.

Since the aforementioned diagram commutes, notice that for any $r \in k$:

$$r \overset{!}{=} i(r) = j(\iota(r)) = j(r)$$

where ι is the natural inclusion into K, and the image of i is identified with the domain, justifying the $\overset{!}{=}$. Thus, j must *always* fix the base field, i.e. j is a k-homomorphism/isomorphism.

You might be concerned about *which* root of an irreducible polynomial does $\overline{x}$ represent in the previous theorem. For example, $x^2 - 2$ is irreducible over $\mathbb{Q}$, with $\pm\sqrt{2}$. So which root should we pick to represent $\overline{x}$? In fact, we can pick $\overline{x}$ to be represented *any* root of this polynomial: the field extending to contain $\sqrt{2}$ or $-\sqrt{2}$ will be isomorphic. More generally, if $p(x)$ is any irreducible

polynomial, then $k[x]/(p(x))$ can be representing *any* choice of the possible roots of $p(x)$[6] and extending k to contain that root. To make the choice concrete, we can choose a particular root of $p(x)$, say α (which we define by saying α satisfies some polynomial equation), and isomorphically identify $K \xrightarrow{\sim} k(\alpha)$ (which we will show in more detail in theroem 17.2.4).

Note that to find the root of a polynomial, we required that polynomial to be irreducible. This technique will not obtain us a root of $(x^2-2)(x^3-2) = x^6 - 5x^2 + 6$, namely becaue $(x^6 - 5x^2 + 6)$ is not even a prime ideal (the resulting quotient contains zero-divisors). In section 17.5.2, we will explore how to find extensions for which *any* polynomial has all of its roots, essentially by finding the roots of every irreducible polynomial and progressively extending our field by adjoning roots.

Note that we haven't shown a way to "find" the root. For example, if we were looking at irreducible polynomial over $\mathbb{R}$ and we pick a root $\alpha \in \mathbb{C}$ that satisfies the polynomial equation, we don't have a computational way of finding it:

Example 17.2.2: Existence, but not construction

Let's say say F is a field extension $F/\mathbb{Q}$. Consider the irreducible polynomial $x^3 - 3x - 1$. over $\mathbb{Q}$ (it has no rationals roots by prosposition 11.3.4). Thus, $F = \mathbb{Q}[x]/(x^3 - 3x - 1)$ is a field which has a root! The fact that all the roots of $x^3 - 3x - 1$ are real (and hence being able to identify F with a subfield of $\mathbb{R}$) can be considerd as taken for granted for now; we will have the tools to show that by the end of section 18.1.1. Visually, if we plot $x^3 - 3x - 1$, notice that all the roots are real, and hence $\mathbb{Q}$ can only by roots of $x^3 - 3x - 1$ which are real.

Let's say α is the root, so $\alpha \in \mathbb{R}$. We might hope that we can find which α in $\mathbb{R}$ satisfies the polynomial given that we know α is a root, however there is no *algebraic* way of finding this root! Using Newton-Raphsons approximation method, we can find that it is a real number $x \approx 1.97938....$ Finding these roots is a sub-branch of numerical analysis known as *root-finding algorithms*

For most common roots, we have already developped a symbol to represent them, namely $\sqrt[n]{m}$ is the root of $x^n - m$. More generally, if we extend $\mathbb{Q}[x]/(x^n - m)$ to solve another polynomial of a similar form, say $x^k - l$:

$$\frac{(\mathbb{Q}[x]/(x^n - m))[y]}{(y^k - l)}$$

we can "nest" as many square roots as we want. Since adding and multiplying is closed in a field, any extension of such polynomials produce elements of the form:

$$\sqrt[3]{\frac{3 + \sqrt[9]{40}}{\sqrt[8]{\sqrt{2}\sqrt{10}}}}$$

However, most more complicated polynomials are not given a special symbol, a notable exception being the golden ratio φ which is the root of $x^2 - x - 1$[7]. In section 18.2.4, we will further explore extensions which only add radical elements.

With the extension field through an irreducible polynomial constructed, it would be advanatagous to be able to easily work with it's elements. Since K is a sufield of k, we can describe K as a k-vector-space!

[6] What it means to "choose" a root here is a bit ambiguous, since we have yet to say there exists a field that makes $p(x)$ factor into linear terms. Such a field is constructed in section 17.5.1

[7] Though, this is not much of an exception since $\varphi = \frac{1+\sqrt{5}}{2}$

Theorem 17.2.2: Basis For Kronecker Field

Let $p(x) \in F[x]$ be an irreducible polynomial of degree n over the field F and let K be the field $F[x]/(p(x))$. Let $\theta = x \bmod (p(x)) \in K$. Then the elements

$$1, \theta, \theta^2, ..., \theta^{n-1}$$

are a basis for K as a vector space over F, so the degree of the extension is n, i.e., $[K : F] = n$. Thus

$$K = \left\{a_0 + a_1\theta + a_2\theta^2 + \cdots + a_{n-1}\theta^{n-1} | a_0, a_1, ..., a_{n-1} \in F\right\}$$

consists of all polynomials of degree $< n$ in θ. In particular, all field extensions of this form have finite degree.

Proof **:**
Take $K = F[x]/(p(x))$, and consider $a(x) \in F[x]$ to be any polynomial. Then, by the Euclidean algorithm, we get that

$$a(x) = q(x)p(x) + r(x) \qquad q(x), r(x) \in F[x] \text{ with } \deg(r(x)) < n$$

then in K, we get that $[a(x)] = [r(x) + q(x)p(x)] = [r(x)]$, so any polynomial can be represented by it's residue polynomial, which will always be of degree $< n$. Thus, $1, \theta, \theta^2, ..., \theta^{n-1}$ (the image of $1, x, x^2, ..., x^{n-1}$) span the quotient. It remains to show they are linearly independent to conclude they are a basis.

If they weren't, then for some $b_i \in F$ (not all 0)

$$b_0 + b_1\theta + \cdots + b_{n-1}\theta^{n-1} = 0$$

that is

$$b_0 + b_1x + \cdots + b_{n-1}x^{n-1} \equiv 0 \bmod (p(x))$$

meaning $p(x) \big| b_0 + b_1x + \cdots + b_{n-1}x^{n-1}$. However, that's a contradiction, since the degree of $p(x)$ is higher than the polynomial on the right hand side (which is at most $n-1$)! Thus, these must be linearly independent. Since they also span, they fulfill the condition of a basis.

Since a basis is generated as a k-algebra by θ, any k-homomorhpism $\varphi : k(\theta) \to F$ is determined by where it maps θ, making the map particularly easy to determine if we determine where θ maps to (the criterion for which we do in corollary 17.2.15).

As a reminder of arithmetics in $F[x]/(p(x))$:

Corollary 17.2.3: Shortcut On Arithmetics In K

Let K be as in theorem 17.2.1, and let $a(\theta), b(\theta) \in K$ be two polynomials of degree $< N$ in θ. Then addition in K is defined simply bu usual polynomial addition and multiplicatin in K is defined by

$$a(\theta)b(\theta) = r(\theta)$$

where $f(x)$ is the remainder (of degree $< n$) obtaiend after dividing the polynomial $a(x)b(x)$ by $p(x)$ in $F[x]$

***Proof* :**
An exercice in quotient ring of polynomials. Addition multiplication is done as in any quotient ring:

1. addition is usual addition of polynomials, since it doesn't change the degree
2. For multiplication, one can multiply the terms, then find a representative in the coset with degree $< n$

Note that what is described in corrollary 17.2.3 works in any quotient field, including over polynomials which are *not* irreducible. However, the result is not a field – not even an integral domain [8].

Example 17.2.3: Field Extension Through Irreducible Polynomials

1. Let $F = \mathbb{R}$ and $p(x) = x^2 + 1$. Then we know that $x^2 + 1$ does not have a root in $\mathbb{R}$. Thus, consider:
$$K = \mathbb{R}[x]/(x^2 + 1)$$
is an extension field (of degree 2) that contains a root to this polynomial. Let $\theta = \overline{x}$. By theorem 17.2.2, $K/\mathbb{R}$ has degree 2 and any element $t \in K$ can be written as:
$$t = a + b\theta \qquad a, b \in \mathbb{R}$$
If you are comfortable with the arithmetics of $\mathbb{C}$, show that:
$$\mathbb{R}[x]/(x^2 + 1) \cong \mathbb{C}$$
2. Let $F = \mathbb{Q}$ and $p(x) = x^2 + 1$. Then by substitution with Eisntein's Criteron, $x^2 + 1$ is irreducible, hence $K = \mathbb{Q}[x]/(x^2 + 1)$ is a field extension of $\mathbb{Q}$. Since the base field is now $\mathbb{Q}$, any element $t \in K$ is of the form:
$$t = a + b\theta \qquad a, b \in \mathbb{Q}$$
3. Let $F = \mathbb{Q}$ and $p(x) = x^2 - 2$. Then $x^2 - 2$ is irreducible by Eisenstein's, so $F = \mathbb{Q}[x]/(x^2 - 2)$ is a field extension of $\mathbb{Q}$. Letting $K = \mathbb{Q}[x](x^2 + 1)$, show that:
$$F \ncong K$$
that is, extending by different roots of different irreducible polynomial produces distinct fields.

[8] Recall that if $p(x) = a(x)b(x)$, $a(x)$, $b(x)$ not being units, then in $F[x]/(p(x))$, $\overline{a(x)}, \overline{b(x)}$ are both nonzero, but $\overline{a(x)} \cdot \overline{b(x)} = \overline{0}$, meaning $F[x]/(p(x))$ has zero-divisors

4. Let $F = \mathbb{F}_2$ (the finite field with 2 elements) and $p(x) = x^2 + x + 1$. Then $p(x)$ is irreducible since $0^2 + 0 + 1 = 1^1 + 1 + 1 = 1$. Hence, $K = \mathbb{F}_2[x]/(x^2 + x + 1)$ is irreducible of degree 2. Can you make the multiplication table for this field?

Classifying Simple and Transcendental Extensions

Let's now turn our attention to classifying field extensions, given $k \subseteq K$ where we consider k the base field. If $A \subseteq K$, then we already have a notion of field extensions: $k \subseteq k(A)$. We will focus on the special cae where $A = \{\alpha\}$, that is, simple extension with $\alpha \in K$. As an example of this, imagine taking $\mathbb{R} \subseteq \mathbb{C}$, $\alpha = i$, and considering $\mathbb{R}(i)$. Then it is a known fact that $i^2 + 1 = 0$. We might ask if in general, given some $\alpha \in K$, can we find a polynomial that sovles it? In this way, we can characterize the field $k(\alpha)$ without needing to resort to taking the intersection of all subfields of K containing F and α. The answer is in fact, half the time. All simple extensions can be characterized by either being isomorphic to the rational functions over the base field or in terms of a quotient of a polynomial ring, giving us an easy classification of simple extensions:

Theorem 17.2.4: Classifying Simple Extensions

Let K/k be a field, $\alpha \in K$, and $k(\alpha)$. Consider $\mathrm{ev}_\alpha : k[x] \to k(\alpha)$ mapping $f(x) \mapsto f(\alpha)$. Then if ev_α is injective, $k(\alpha)$ is an infinite extension and $k(\alpha) \cong k(x)$, and if ev_α is not injective, then there exists a polynomial $p(x)$ such that $k[x]/(p(x)) \cong k(\alpha)$. Furthermore, $p(x)$ is the polynomial of *minimal degree* that has α as a root.

Note that in this theorem, we're saying that *any* field over K extending k in which $p(x)$ contains a root in K and is irreducible in k contains a subfield isomorphic to the extension of k as we constructed in theorem 17.2.1 and that this field is (up to isomorphism) the smallest extension of k containing such a root.

Proof **:**
Take the natural homomorphism

$$\mathrm{ev}_\alpha : F[x] \to F(\alpha)(\subseteq K) \qquad a(x) \mapsto a(\alpha)$$

which maps the polynomial $a(x)$ to the polynomial evaluated at α. Note that $\mathrm{ev}_\alpha(F[x])$ is an integral domain, since $F(\alpha)$ is integral domain.

If ev_α is injective, then $F[x]$ can be identified with an subring of $F(\alpha)$ that is an integral domain. By the universal property of field of fraction (theorem 9.5.3), ev_α uniquely extends to a homomorphism

$$\varphi : F(x) \to F(\alpha)$$

Since $\varphi(F(x))$ is a field containing F and α, $\varphi(F(x)) \supseteq F(\alpha)$, and since φ is a function $\varphi(F(x)) \subseteq F(\alpha)$, hence $\varphi(F(x)) = F(\alpha)$. To find the degree $[F(\alpha) : F]$, notice that $1[= \alpha^0], \alpha, \alpha^2, ...$ are all linearly independent over F in $F(\alpha)$ (since $1[x^0], x, x^2, ...$ are all linearly independent all linearly indepndent over F in $F(x)$), hence any basis must at least contain countably many linearly independent elements, and hence the degree must be infinite.

On the otherhand, if ev_α is not injective, then $\ker(\mathrm{ev}_\alpha) \neq 0$ and ev_α is clearly surjective. Since ev_α is surjective and $F(\alpha)$ is a field, $\ker(\mathrm{ev}_\alpha)$ must be a maximal ideal, and since $F[x]$ is a euclidean domain (and hence a PID), any polynomial of minimal degree in $\ker(\mathrm{ev}_\alpha)$ can represent the ideal,

that is $(m_{\alpha,k}(x)) = \ker(\mathrm{ev}_\alpha)$ for some minimal degree polynomial $m_{\alpha,k}(x) \in \ker(\mathrm{ev}_\alpha)$. In fact, this minimal polynomial must be unique up to unit: if there where two minimal polynomials where $(m_{\alpha,k}) = (m'_{\alpha,k})$, then they must be equal up to unit, i.e. up to a multiple of k. Hence, we have that

$$\frac{k[x]}{(m_{\alpha,k}(x))} \cong F(\alpha)$$

By theorem 17.2.2, $[F(\alpha) : F] = \deg m_{\alpha,k}(x)$, and so the extension is finite, completing the proof.

This theorem motivates the following vocabulary:

Definition 17.2.5: Algebraic and Transcendental Elements

Let F be a field and let K be an extension of F. Then:

1. The element $\alpha \in K$ is said to be *algebraic* over F if α is a root of some nonzero polynomial $f(x) \in F[x]$.
2. If α is not algebraic over F (i.e., is not the root of any nonzero polynomial with coefficeints in F) then α is said to be *transcendental over F*.

Dealing with more general transcendental extensions is a straight-forward generlization of the previous thoerem; it naturally generlizes for $k(\alpha_1, ..., \alpha_n) \subseteq K$ where each α_i is transcendental. We first define the following term for the corollary:

(are all monomorphisms injective? I replaced the word monomorphism with injective in my notes for definition 17.2.6)

Definition 17.2.6: Algebraically Independent

Let R be a k-algebra. Then $\alpha_1, ..., \alpha_n \in R$ are *algebraically independent* over k if the k-algebra homomorphism ev_k (i.e., the evaluation map) which fixes k:

$$\mathrm{ev}_k : k[x_1, ..., x_n] \to R \qquad \mathrm{ev}_k(x_i) = \alpha_i$$

is injective. The set is called *algebraically dependent* otherwise.

Corollary 17.2.7: Classifying Finitely Generated Transcendental Extensions

Let K/k be a field extensions such that $K = k(\alpha_1, ..., \alpha_n)$ where each element is trancsendental. Then if $k(x_1, ..., x_n)$ is the field of rational functions over the indeterminates ${x_i}^a$, then, the set $\alpha_1, ..., \alpha_n$ is algebraically independent, and:

$$k(x_1, ..., x_n) \cong_k k(\alpha_1, ..., \alpha_n)$$

[a] sometimes called *formal variables*

The case for arbitrarily many transcendental elements is the same proof with more book-keeping on notation.

***Proof* :**
This is an immediate generalization of theorem 17.2.4

This shows that in terms of fields, algebraically independent sets don't offer much to our analysis of classification; they act more like a basis with essentially no relation[9]. This is not to say that they don't offer any use in Galois theory, where we will correspond group for transcendental extensions, but in terms of field theory their structure is very simple. Thus, we will take more time stuyding algebraically dependent sets, especially since we can use theorem 17.2.1 to study study irreducible polynomials that aquire a root.

Going back to the minimal polynomial that we found in the proof,

Definition 17.2.8: Minimal Polynomial

Let $k(\alpha)$ be a finite simple extension as in theorem 17.2.4, and $m_{\alpha,k}(x)$ be the polynomial given in the proof. Then $m_{\alpha,k}(x)$ is called the *minimal polynomial* of α over k .

it can be shown that $m(x)$ is unique *up to base field* by taking advanage of the euclidean algortihm for $F[x]$ (see exercise ref:HERE), and we have an easy way of linking minimal polynomial between fields:

Proposition 17.2.9

If L/F is an extension of fields and α is algebaric over both F and L, then $m_{\alpha,L}(x)$ divides $m_{\alpha,F}(x)$ in $L[x]$.

***Proof* :**
Since both $m_{\alpha,L}(x)$ and $m_{\alpha,F}(x)$ are polynomials in $L[x]$, we can divide $m_{\alpha,F}(x)$ by $m_{\alpha,L}(x)$ by each other. If we consider:

$$m_{\alpha,L}(x) = m_{\alpha,F}(x)g(x) + r(x)$$

then since α is a root:

$$0 = m_{\alpha,L}(\alpha) = m_{\alpha,F}(\alpha)g(\alpha) + r(\alpha) = 0g(x) + r(\alpha)$$

which implies $r(\alpha) = 0$, and has degree smaller than $m_{\alpha,F}(x)$ – a contradiction to the minimality of the degree of $m_{\alpha,F}(x)$ unless $r(x) = 0$. Thus:

$$m_{\alpha,L}(x) = m_{\alpha,F}(x)g(x) \qquad g(x) \in L[x]$$

However, since $m_{\alpha,L}(x)$ is the minimal polynomial over $L[x]$ that has a root α, it must be that $g(x) = l \in L$. However, if we proceed with the same proof dividing $m_{\alpha,F}(x)$ by $m_{\alpha,L}(x)$:

$$m_{\alpha,F}(x) = m_{\alpha,L}(x)g(x) \qquad g(x) \in L[x]$$

then this does not contradict minimality, showing that $m_{\alpha,L}$ divides $m_{\alpha,F}$ in $L[x]$.

to obtain (using the euclidean algorithm):

$$m_{\alpha,F}(x) = m_{\alpha,L}(x)g(x) + r(x)$$

[9]In fact, we will define the notion of *transendental basis* to be the largest algebraically independent set of a ring, which will be key to defining a notion of "dimension" for varities. More on that in part V, chapter 21

where $\deg r(x) < \deg m_{\alpha,F}(x)$. Then:

$$0 = m_{\alpha,L}(\alpha) = m_{\alpha,F}(\alpha)g(\alpha) + r(\alpha) = 0g(x) + r(\alpha)$$

which implies $r(\alpha) = 0$, and has degree smalelr than $m_{\alpha,F}(x)$ – a contradiction to the minimality of the degree of $m_{\alpha,F}(x)$ unless $r(x) = 0$. Thus:

$$m_{\alpha,L}(x) = m_{\alpha,F}(x)g(x) \qquad g(x) \in L[x]$$

completing the proof.

It is worth noting that the miniaml polynomial will always be irreducible: If it was reducible, then $m_{\alpha,k}(x) = g(x)h(x)$ would produce a polynomial where either $g(\alpha) = 0$ or $h(\alpha) = 0$, contradicting the minimality of $m_{\alpha,k}(x)$. To find the minimal polynomial we can do (ex. 20 in the book), it is surprisngly linked to finding a particular characteritic polynomial!!

Proposition 17.2.10: Finding Minimal Polynomial

Let F/k be a finite field extension. Then there exists a minimal polynomial $m_{\alpha,k}$ such that $m_{\alpha,k}(\alpha) = 0$.

Proof **:**
Let F/k be a finite field extension.

1. For any $\alpha \in F$, prove that α acting by left multiplication on F is an k-linear transfomration of F, label it L_α.
2. Show α is a root of the characteristic polynomial for L_α. This gives an effective procedure for determining an equation of degree n satisfied by an element α in the extension of k of degree n.

Alternitavely, we can use what find a Gröbner bases(see section 20.2.1)

Example 17.2.4: Minimal Polynomials

1. The minimal polynomial for $\sqrt{2}$ over $\mathbb{Q}$ is $x^2 - 2$ and $\sqrt{2}$ is of degree 2 over $\mathbb{Q}$: $[\mathbb{Q}(\sqrt{2}), \mathbb{Q}] = 2$
2. The minimal polynomial for $\sqrt[3]{2}$ over $\mathbb{Q}$ is $x^3 - 2$ and $\sqrt[3]{2}$ is of degree 3 over $\mathbb{Q}$: $[\mathbb{Q}(\sqrt[3]{2}), \mathbb{Q}] = 3$
3. similarly, for any $n > 1$, the polynomial $x^n - 2$ is irreducible (by Eisenstein's Criterion).
4. The minimal polynomial and the degree of an element α are dependent on the base field! For example, if $F = \mathbb{R}$, then the element $\sqrt[n]{2}$ has degree 1, with minimal polynomial
$$m_{\sqrt[n]{2},\mathbb{R}}(x) = x - \sqrt[n]{2}$$
5. the minimal polynomial for the golden ratio φ is $x^2 - x - 1$.
6. In general, most algebraic elements do not have a special symbol representing them, and so we would have to say that some $\alpha \in K$ has minimal polynomial (say) $x^5 - x - 1$

As a result of theorem 17.2.4, it also makes sense to ask the context of the existence of an irreducible polynomial what is the degree of an α over F: given $p(x)$ is the polynomial for which α is a root, then it's the degree of $\dim_F F[x]/(p(x)) = \dim_F F(\alpha)$. We may write $\deg \alpha$ (or $\deg_F \alpha$) for the degree of α (over F). If $\deg \alpha = 1$, then $\alpha \in F$.

Combining the classification of finite simple extensions with corrollary 17.2.3, we get to easily represent $F(\alpha)$

Corollary 17.2.11: Representing $F(\alpha)$

Suppose in theorem 17.2.4 that $p(x)$ is of degree n. Then

$$F(\alpha) = \left\{a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_{n-1}\alpha^{n-1} | a_0, a_1, ..., a_{n-1} \in F\right\} \subseteq K$$

Proposition 17.2.12: Finite Simple Extensions Elements

Let $k(\alpha)/k$ be a finite simple extension. Then:

$$k[\alpha] = k(\alpha)$$

Proof **:**
exercise

This justifies why we only need to care about polynomials, and not rational polynomials.

Example 17.2.5: Simple Extension

1. In $\mathbb{Q}(\sqrt{2})$, $\dim_{\mathbb{Q}} \mathbb{Q}(\sqrt{2}) = 2$, and we can write $\mathbb{Q}(\sqrt{2}) = \left\{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\right\}$. We could also have chosen the other root of $x^2 - 2$, namely $-\sqrt{2}$. However, these two fields are isomrophic: $\mathbb{Q}(\sqrt{2}) \cong \mathbb{Q}(-\sqrt{2})$, showing they are algebraically indistinguishable.
2. In $\mathbb{Q}(\sqrt[3]{2})$, $\dim_{\mathbb{Q}} \mathbb{Q}(\sqrt[3]{2}) = 3$, and we can write $\mathbb{Q}(\sqrt[3]{2}) = \left\{a + b\sqrt[3]{2} + c(\sqrt[3]{2})^2 \mid a, b, c \in \mathbb{Q}\right\}$.
3. Over $\mathbb{R}$, $x^3 - 2$ has no more solutions than the last example. However, over $\mathbb{C}$, there are two more solutions:
$$\zeta_2 = \sqrt[3]{2}\left(\frac{-1 + i\sqrt{3}}{2}\right) \qquad \zeta_3 = \sqrt[3]{2}\left(\frac{-1 - i\sqrt{3}}{2}\right)$$
Thus, $\mathbb{Q}(\zeta_2)$ and $\mathbb{Q}(\zeta_3)$ are field extensions.

Thus, we have fully classified all simple extensions of a field and dealt with transcendental extensions. In following sections, we will show how to classify fields generated by finitely many elements (ex. $k(\alpha, \beta)$) and infinitely many elements. Before continuing into that endeavour, isomorphisms between fields should be given some more attention to be more cear on how to classify fields.

The fields in example 17.2.5 need not be isomorphic (as fields) if the roots solve different polynomials, meaning different irreducible polynomials will produce different extensions:

Example 17.2.6: $\mathbb{Q}(\sqrt{2}) \ncong \mathbb{Q}(\sqrt{5})$
Take $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{5})$. Let's say φ is an isomorphism. Then we know that

$$\varphi(1) = 1$$

since 1 is the only element of order 1 in both fields. Then:

$$\varphi(\sqrt{2}) = a + b\sqrt{5}$$

for some non-zero value in the codomain. Then

$$\varphi(2) = a^2 + 2ab\sqrt{5} + 5b^2$$

but we also know that

$$\varphi(1) = 1 \qquad \varphi(2) = \varphi(1+1) = \varphi(1) + \varphi(1) = 1 + 1 = 2$$

so

$$2 = a^2 + 2ab\sqrt{5} + 5b^2 \qquad a, b \in \mathbb{Q}$$

If $a = b = 0$, $2 = 0$, a contradiction. If $a, b \neq 0$, then:

$$\frac{2 - a^2 - 5b^2}{2ab} = \sqrt{5}$$

implying that $\sqrt{5}$ is a rational number, a contradiction. If $a = 0, b \neq 0$, then:

$$\frac{\sqrt{2}}{\sqrt{5}} = b$$

but b is a rational number, a contraidction. If $a \neq 0, b = 0$, then:

$$\sqrt{2} = a$$

but a is a rational number, a contradiction. Thus, these two fields cannot be isomorphic.

However, notice that the extension of to $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(-\sqrt{2})$ were isomorphic, and so was all the extensions of $\mathbb{Q}$ by roots of $x^3 - 2$:

$$\mathbb{Q}(\sqrt[3]{2}) \cong \mathbb{Q}\left(\sqrt[3]{2}\left(\frac{-1 + i\sqrt{3}}{2}\right)\right) \cong \mathbb{Q}\left(\sqrt[3]{2}\left(\frac{-1 - i\sqrt{3}}{2}\right)\right)$$

even though those 3 fields are not equal. In particular, that means that non of these fields contain *all* of the roots of $x^3 - 2$. One of the large motifs in the following study of extensions is to find extensions in which for any irreducible (or more generally any polynomial) will contain *all* of the roots.

17.2.1 Isomorphism of Fields Extensions

As we have seen in theorem 17.2.1, If we know that an irreducible polynomial $p(x) \in k[x]$ has a root in a field F. Then there exists a [not necessarily unique] isomorphism φ such that the following

diagram commutes:

$$\begin{tikzcd} k[x]/(p(x)) \arrow[r, "\varphi"] & F \\ k \arrow[u, hook, "\iota"] \arrow[ur, hook, "\iota"'] & \end{tikzcd}$$

The fact that φ is not unique gives raise to several interesting results. As we have shown in example 17.2.1, $\mathbb{Q}(\sqrt{2})$ has two homomorphisms into $\mathbb{R}$. In fact, since $-\sqrt{2} \in \mathbb{Q}(\sqrt{2})$, we have that:

$$\begin{tikzcd} \mathbb{Q}(\sqrt{2}) \arrow[r, "\varphi"] & \mathbb{Q}(-\sqrt{2}) \\ \mathbb{Q} \arrow[u, hook, "\iota"] \arrow[ur, hook, "\iota"'] & \end{tikzcd}$$

If we extend the bottom to be the identity map form k to k, then we get a square:

$$\begin{tikzcd} \mathbb{Q}(\sqrt{2}) \arrow[r, "\sim"] & \mathbb{Q}(-\sqrt{2}) \\ \mathbb{Q} \arrow[u] \arrow[r, "\mathrm{id}"] & \mathbb{Q} \arrow[u] \end{tikzcd}$$

Showing the fields with the two roots of $x^2 - 2$ are isomorphic. In fact,$-\sqrt{2} \in \mathbb{Q}(\sqrt{2})$, and so this is in fact a field automorphism (though in general it doesn't have to be, like $\mathbb{Q}(\sqrt[3]{2})$ and $\mathbb{Q}(\sqrt[3]{2}\zeta_3)$). We have claimed earlier than two fields that have two different roots of the same irreducible polynomial adjoined to them are algebraically indistinguishable. This shows it in the case for $x^2 - 2$ over $\mathbb{Q}$, and the following theorem shows the generlizatoin of this result. Note how by keeping track of the root we have in fact recovered a sense of uniqueness:

Theorem 17.2.13: Extending Isomorphisms to Adjoint Fields

Let $k_1 \subseteq F_1 = k_1(\alpha_1)$, $k_2 \subseteq F_2 = k_2(\alpha_2)$ be simple extensions and let $\varphi : k_1 \to k_2$ be an isomorphism of fields. Let $p(x) \in k_1[x]$ be an irreducible polynomial and let $p'(x) \in k_2[x]$ be the irredcibble polynomial obtained by applying the map φ to the coefficients of $p(x)$. Let α_1 be a root of $p(x)$ (in some extension of k_1) and let α_2 be a root of $p'(x)$ (in some extension k_2). Then there is an isomorphism:

$$\sigma : k_1(\alpha_1) \to k_2(\alpha_2) \qquad \alpha_1 \mapsto \alpha_2$$

mapping α_1 to α_2 and extending φ, .i.e., such that σ restricted to F is the isomorphism φ.

In other words, if we have

$$\begin{tikzcd} k_1(\alpha) & k_2(\beta) \\ k_1 \arrow[u, hook] \arrow[r, "\sim"] & k_2 \arrow[u, hook] \end{tikzcd}$$

and $\iota(m_{\alpha,k_1})(x)) = m_{\beta,k_2}(x)$ (i.e. it maps the coefficients of the first minimal polynomial to the

second one), then we can *uniquely* extend ι to an isomorphism j between the two adjoint fields:

$$\begin{tikzcd} k_1(\alpha) \arrow[r, dashed, "\sim"] & k_2(\beta) \\ k_1 \arrow[u, hook] \arrow[r, "\sim"] & k_2 \arrow[u, hook] \end{tikzcd}$$

such that $j(\alpha) = \beta$. Note that j being an extension of ι means the elements of k_1 are fixed in j.

***Proof* :**
Let $\varphi : k_1 \to k_2$ be an isomorphism of fields. The map φ induces a ring isomorphism:

$$f^{\varphi} : k_1[x] \to k_2[x]$$

defined by applying φ to the coefficeints of a polynomial in $k_1[x]$. Let $p(x) \in k_1[x]$ be an irreducible polynomial, and let $p'(x) \in k_2[x]$ be the polynomial obtained by applying the map φ to the coefficients of $p(x)$, i.e., the image of $p(x)$ under f^{φ}. The isomorphism f^{φ} maps the maximal ideal $(p(x))$ to the ideal $(p'(x))$, so the ideal is also maximal, which shows that $p'(x)$ is also irreducible in $k_2[x]$.

Taking the quotinet by these ideals of the map, we get the following isomorphism of fields

$$k_1[x]/(p(x)) \to k_2[x]/(p'(x))$$

By theorem 17.2.4, the left side is isomorphic to $k_1(\alpha_1)$, and the right side to $k_2(\alpha_2)$, so composing these maps will produce a map for σ! It should also be clear now that the restriction of this isomorphism to k_1 is φ.

This isomorphism will come back big time when we address galois theory. It is sometimes represented in as in the following diagram:

$$\begin{tikzcd} \sigma : & k_1(\alpha_1) \arrow[r, "\cong"] \arrow[d, dash] & k_2(\alpha_2) \arrow[d, dash] \\ \varphi : & k_1 \arrow[r, "\cong"] & k_2 \end{tikzcd}$$

If we let $\mathrm{id} : k \to k$ be our k-isomorphism, and $m_{\alpha,k}(x)$ be our minimal polynomial with roots α, β, then $\varphi : k(\alpha) \xrightarrow{\sim} k(\beta)$ is also an isomorphism, showing that the roots of a minimal polynomials are algebraically indistinguishable. Furthermore, we can upgrade Kronecker's Theorem so that it gives us a universaly property: instead of L being a field that has a root of $p(x)$, upgrade that to $\alpha \in L$ being a root of $p(x)$. Then there exists a *unique* homomorhpism mapping $\overline{x}$ to α (with uniqueness established by the above theorem).

Through this theorem, we get another important insight. If we have a non-trivial finite field extension $k(\alpha)/k$ which has roots of a minimal polynomial $m_{k,\alpha}(x)$, then if any other roots of $m_{k,\alpha}(x)$ are in $k(\alpha)$, there is a k-automorphisms of $k(\alpha)$ (can you see why a k-automorphism is determined by where it sends the roots?). This means we can find elements of the group $\mathrm{Aut}_k(k(\alpha))$ (where the subscript means we fix k), and hence the structure of $k(\alpha)$ [10]. More conretely, $k(\alpha)/k$ represents adding a

[10] Recall that we already used the automorphism group to deduce information about groups, for example $\mathrm{Out}(G) \cong_{\mathbf{Grp}} \mathrm{Aut}(G)$ if and only if G is abelian

new element α to the field k which satisfies an algebraic relation $m_{\alpha,k}(\alpha)$ (a polynomial). Finding information about the new field $k(\alpha)$ by looking at the k-automorphisms is essentially studying the properties of the polynomial $m_{\alpha,k}(x)$.

Even better, we will show that this characterizes *all* k-automorhpisms, giving us a complete description of these symmetries. These symmetries are actually at the heart of studying fields in galois theory. When Galois originally formulated his theory, he looked different ways to permute the roots and noticed that there are often limitation to the permutations of the roots since the roots may be related to each other[11]. For example if α is a root of $x^3 - x - 1$, then the other two roots are:

$$\frac{1}{1-\alpha} \qquad 1 + \frac{1}{\alpha}$$

Similarly, if ζ is a root of $x^n - 1$, then the other roots of $x^n - 1$ are $\zeta^2, \zeta^3, ..., \zeta^{n-1}, 1$. Thus, we take some time to define our terms more precisely, and elaborate further what the automorphisms look like:

Definition 17.2.14: Automorphism and fixed Element

Let K be a field and $\mathrm{Aut}(K)$ be the collection of automorphisms (isomorphisms from k to itself). Then:

1. An auotmorphism $\sigma \in \mathrm{Aut}(K)$ is said to *fix* an element $\alpha \in K$ if $\sigma\alpha = \alpha$. If F is a subset of K, then an automorphism σ is said to *fix* F if it fixes all the elements of F

$$\sigma a = a \qquad \forall a \in F$$

2. If F/k be an extension of fields, then $\mathrm{Aut}(F/k)$ or $\mathrm{Aut}_k(F)$ is the collection of all automorphsism of F which fix k.

the previous theorem gives us the following link between $\mathrm{Aut}_k(F)$ and F given that F is finite and simple:

Corollary 17.2.15: Order Of Fixed Automorphism Group

Let F/k be a simple finite extension (so $F = k(\alpha)$ for some $\alpha \in F$), and let $m_{k,\alpha}(x) \in k[x]$ be the minimal polynomial of α. Then $|\mathrm{Aut}_k(F)|$ is equal to the number of *distinct* roots of $m_{k,\alpha}(x)$ in F. In particular,

$$|\mathrm{Aut}_k(F)| \leq [k(\alpha) : k] \qquad |\mathrm{Aut}_k(F)| = [k(\alpha) : k]_d$$

where $[k(\alpha) : k]_d$ represents the number of *distinct* roots $k(\alpha)$ contains of the minimal polynomial $m_{k,\alpha})(x)$. Equality holds if and only if $m_{k,\alpha}(x)$ factors over F as a product of *distinct linear* polynomials

This result will be called the *fixed root criterion* when we cover Galois theory.

[11] Later we shall see that the roots are almost never related to each other, that is most Galois groups are S_n, but it is a good insight to notice that the roots of small polynomials are often related

***Proof* :**
Let $\sigma \in \mathrm{Aut}_k(F)$ be an automorphism fixing k (so for any $t \in k$, $\sigma(t) = t$). Then since every element $t \in F = k(\alpha)$ can be written as a polynomial $t = f(\alpha) \in k[\alpha]$ (recall that $k(\alpha)$ is finite, and so has a basis of the form $\alpha, \alpha^2, ..., \alpha^n$ for some appropriat n), and σ fixes k, all the elements are determined by where α maps to, that is $\sigma(\alpha)$. Then since σ is an automorphism (hence ring homomorphism), given

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

is the minimal polynomial with coefficients from k so that

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0$$

then, applying any automorphism σ to this polynomial will get us

$$\begin{aligned} \sigma(\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0) &= \sigma(0) \\ \sigma(\alpha)^n + a_{n-1}\sigma(\alpha)^{n-1} + \cdots + a_1\sigma(\alpha) + \sigma(a_0) &= 0 \qquad \sigma \text{ fixes } a_i \text{ for each } i \end{aligned}$$

since $a_i \in F$ for each i, showing that $\sigma(\alpha)$ is also a root of the minimal polynomial, i.e. σ must map α to some other root of $m_{k,\alpha}$.

As a consequence, there can be at most $\deg m_{k,\alpha}(x)$ automorphisms.

Conversely, for any root β of $p(x)$ in $k(\alpha)$, $\beta \in k(\alpha)$, we can take the isomorphism $\mathrm{id} : k \to k$, which by theorem 17.2.13, there exists a unique lift to a root $k(\alpha) \to k(\beta)$ that fixes k. Since $k(\alpha) \cong k(\beta)$, then we can use this isomrophism to define an automorphism from $k(\alpha)$ to $k(\alpha)$, showing us that $\mathrm{Aut}_k(F)$ must have at least as many elements as there are roots of the minimal polynomial in F. If F contains all the roots of $p(x)$ and each root of $p(x)$ is distinct, then $|\mathrm{Aut}_k(F)| = [F : k]$, completing the proof.

Interpreting the above example as a group actions on the set of roots R_F of $m_{\alpha,k}(x)$ in F, $\mathrm{Aut}_k(F) \curvearrowright R_F$, we have that $\mathrm{Aut}_k(F)$ acts *faithfully and transitively* on the set of roots of $p(x)$ in F (the group $\mathrm{Aut}_k(F)$ can be identified with some transitive subgroup of the symmetric group acting on the roots of $p(x)$ in F). This correlation hints at the connection beween fields and groups at the heart of the galois theory, since any collection of automorphisms will always form a group[12]. The main impediment to directly translating this result of more general [finite] field extensions[13] is that the minimal polynomial might not factor into *linear* and *distinct* terms over the simple extension $k(\alpha)/k$[14]. For example, in $\mathbb{Q}(\sqrt[3]{2})$, the polynomial $x^3 - 1$ doesn't fully factor. The next two sections will encompass finding field extensions in which these conditions are met. In particular, we shall take some time to find fields in which:

1. All polynomials factor (splitting fields)
2. If $f(x) \in k[x]$ is irreducible, then in F/k, either $f(x)$ factors completely or $f(x)$ remains irreducible (normal extensions)
3. Given a polynomial $f(x)$, F/k factors completely (Splitting field)

[12] If you are not convinced, now is good time to check this
[13] infinite extension are more finicky to deal with, and are relegated to chapter ref:HERE
[14] It is possible that a irreducible polynomial factors linearly, but the factors are not distinct, see section 17.6

4. If $f(x)$ is irreducible in k, then in F/k, $f(x)$ factors into linear terms of degree 1 (Separable extensions)

As a final word on relating isomorphisms and extensions, note that if $\dim_k A = \dim_k B$, this does not imply $[B : A] = 1$ or that $A \cong_k B$:

Example 17.2.7: Nuance On Dimension

Let $A = \mathbb{Q}(\sqrt{3})$ and $B = \mathbb{Q}(\zeta_3\sqrt{3})$. Then the two are of the same dimension, both extend $\mathbb{Q}$ and $A \cong_k B$, but $[B : A]$ is not even well defined. To visualize:

$$\begin{tikzcd} \mathbb{Q}(\sqrt{3}) \arrow[rr, phantom, "\cong_k"] & & \mathbb{Q}(\zeta_3\sqrt{3}) \\ & \mathbb{Q} \arrow[ul] \arrow[ur] & \end{tikzcd}$$

Exercise 17.2.1

1. Let $\mathrm{ch}(k) = n$, $n \geq 0$, and $\varphi : k \to K$ be a homomorphism. Then $\mathrm{ch}(K) = n$.
2. Show that the intersection of any collection of subfields of a field is a field. Is the same true for the union? What if the fields where given a partial order through inclusion?
3. (personal idea, not sure if valid yet) Want to give another intuition using the fact that $R[x]$ is a free monoid, so that in a sense, we can get a "presenation" when we do $R[x]/p(x)$. If $p(x)$ is also irreducible then we get a field as a result.
4. Let $N \subseteq M_{2\times 2}(\mathbb{R})$ be a subset where

$$N = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \;\middle|\; a, b \in \mathbb{R} \right\}$$

   Then $N \cong \mathbb{C}$.
5. Show that If Let E/k and F/k be field extensions. Show that if E/k is finite, then so is EF/F. Show that if E/k and F/k are finite, so is EF/F.
6. Suppose that $f(x) \in k[x]$ where $f(x)$ factors into $f(x) = p_1(x)p_2(x)\cdots p_n(x)$ where $p_i(x) \neq p_j(x)$. Show that $k[x]/(f(x))$ is a product of fields (Hint: Chinese Remainder Theroem). If one of the terms had multiplicity greater than 1, would the result still be a product of fields?
7. (If you know Calculus) The $\mathbb{R}$-algebra $\mathbb{R}[x]/(x^2)$ is sometimes the *dual numbers*. Let any element of this degree 2 extension of $\mathbb{R}$ be written in the form $a + b\epsilon$. If $p'(x)$ is the derivtive of $p(x)$, show that $p(a + b\epsilon) = P(a) + bP'(a)\epsilon$, that is, the derivitve "automatically" shows up.
8. Let m, n be square free and $n \neq m$. Show that $\mathbb{Q}(\sqrt{n}) \ncong_{\mathbb{Q}} (\sqrt{m})$.
9. Let ζ_n be a root of unity where n is odd. Show that $\mathbb{Q}(\zeta_n) \cong_{\mathbb{Q}} \mathbb{Q}(\zeta_{2n})$

17.3 Algebraic Extensions

We'll now focus some attention on the algebraic elements of an extension. We first give name to elements for which all elements are algebraic or non-algebraic:

Definition 17.3.1: Algebraic and Transcendental Extensions

Let F be a field and let K be an extension of F. Then:

1. The element $\alpha \in K$ is said to be *algebraic* over F if α is a root of some nonzero polynomial $f(x) \in F[x]$.
2. If α is not algebraic over F (i.e., is not the root of any nonzero polynomial with coefficeints in F) then α is said to be *transcendental over* F.
3. The extension K/F is said to be *algebraic* or an *algebraic extension* if every element of K is algebraic over F.
4. The extension K/F is said to be *transcendental* or a *transcendental extension* if every element of K not in F is transcendental over F.

Note that if an element α is algebraic over a field F, then it is algebraic over *any* extension field L of F (if $f(x)$ having α as a root has coefficients in F, then it also has coefficients in L). Note too that the polynomial for which α must be a root need not be irreducible, however it is an immediate consequence that if α is the root of a polynomial $p(x) = (a_1(x))^{n_1} \cdots (a_k(x))^{n_k}$, it is a root for some irreducible polynomial $a_i(x)$.

Example 17.3.1: Algebraic extensions

1. Take $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$. Then every element in $\mathbb{Q}(\sqrt{2})$ is a solution to a polynomial over $\mathbb{Q}$, thus $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ is algebraic: any element $a + b\sqrt{2}$ is the root of $\left(\frac{x-a}{b}\right)^2 - 2$. If you take $\mathbb{Q}(\sqrt{2})$ as a vector-space over $\mathbb{Q}$, then we see that 1 and $\sqrt{2}$ generate $\mathbb{Q}(\sqrt{2})$. Note that $\sqrt{2}$ is not in $\mathbb{Q}$, but but $\sqrt{2}$ is algebraic over $\mathbb{Q}$).

   More generally, is it true that any finite simple extension $F(\alpha)$ of F is an algebraic extension? Is every element of $F(\alpha)$ the root of a polynomial with coefficients in F?

2. $\mathbb{C}$ as an $\mathbb{R}$-vector space is generated by 1 and i. Both of these are root to polynomials over $\mathbb{R}$, in particular $x - 1$ and $x^2 + 1$. It is a famous result that any $c \in \mathbb{C}$ is a solution of a polynomial over $\mathbb{R}$ (i.e., with coefficients from $\mathbb{R}$). Thus, $\mathbb{C}/\mathbb{R}$ is algebraic.

   The degree of $[\mathbb{C} : \mathbb{R}] = 2$

3. Consider $\mathbb{C}$ again. For any $\alpha \in \mathbb{C}$, is there a polynomial $p(x) \in \mathbb{C}[x]$ for which $p(\alpha) = 0$? In a sense, we're asking if $\mathbb{C}$ is algebraic over itself. [a]

4. For a moment, take for granted that there is no polynomial in $\mathbb{Q}$ such that π is a root (TBD? appendix?). Then π is *transcendental* over $\mathbb{Q}$, and so $\mathbb{R}/\mathbb{Q}$ is *not* an algebraic extension.

   Here is an idea to ponder: What is $[\mathbb{R} : \mathbb{Q}]$? The answer will be given in example 17.3.5

[a] The answer is yes! We call a field such that every element of it has a root over itself *algebraically closed*

In these examples, the algebraic extensions all had a finite degree. Is it possible to have an algebraic extension with infinite degree? If K is algebaric over F, then every element of K is a roof a polynomial over F? What about the converse: If K/F is finite, is K algebraic over F? (take a moment to ponder here)

Proposition 17.3.2: degree of Algebraic Extension

The element α is algebraic over F if and only if the simple extension $F(\alpha)/F$ is finite. More precisely, if α is an element of an extension of degree n over F, then α satisfies a polynomial of degree at most n over F and if α satisfies a polynomial of degree n over F, then the degree of $F(\alpha)$ over F is at most n.

Proof **:**
If α is algebraic over F (so for some $p(x) \in F[x]$, $p(\alpha) = 0$), then the degree of the extension $F(\alpha)/F$ is the degree of the minimal polynomial, making it an extension of finite degree at most n.

Conversly, let α be an element of degee n over F (ex. $[F(\alpha) : F] = n$). Then by 17.2.2, we know that $n + 1$ elements will be linearly dependent:

$$c_0 + c_1\alpha + c_2\alpha^2 + \cdots + c_n\alpha^n = 0$$

with $c_0, c_2, ..., c_n \in F$ which are not all 0. Thus, α must be the root of some non-zero polynomial with coefficients from F (of degree $\leq n$), showing α is algebraic over F.

Corollary 17.3.3: Finite Extension Then Algebraic Extension

If the extension K/F is finite, then it is algebaric

Proof **:**
let K/F be finite and take any $\alpha \in K$. Consider the simple extension $F(\alpha)$. Consider K as an F-vector space, then since the extension is finite, $\dim_F K = n$. Now, $F(\alpha) \subseteq K$ would be a subspace of K ($F(\alpha)$ also being an F-vector sapce), so

$$\dim_F F(\alpha) \leq \dim_F K \qquad \text{or} \qquad [F(\alpha) : F] \leq [K : F]$$

showing the extension is finite, and so by proposition 17.3.2, α is algebraic over F. since α was an artibrary element of K, this means that K is algebaric over F, as we sought to show.

Though not important to us in this chapter, we also have the following characterisation of an algegraic extension more useful in algebraic number theory:

Corollary 17.3.4: Algebraic Extension and Algebras

K/F is algebraic if and only if every sub F-algebra is a field

Proof **:**
exercise

Note that the converse of corrollary 17.3.3 is not true: If K is algebraic over F, this does not mean that K/F has finite degree

Example 17.3.2: Algebraic and infinite degree
We require corrollary 17.3.10 which is introduced in a couple of pages, but we can still state the case with the added assumption here to get the idea:

Take $\mathbb{C}/\mathbb{Q}$, and take the set $\overline{\mathbb{Q}}$ to be the set of all algebraic elements over $\mathbb{Q}$. Then by the aforementioned corrollary, this is a subfield of $\mathbb{C}$. Since $\sqrt[n]{2}$ is algebraic over $\mathbb{Q}$ (root of the polynomial $x^n - 2$), we have

$$[\mathbb{Q}(\sqrt[n]{2}) : \mathbb{Q}] = n$$

for every n. Since $\mathbb{Q}(\sqrt[n]{2}) \subseteq \overline{\mathbb{Q}}$, it follows that $\overline{\mathbb{Q}}$ is an algebraic extension of infinite degree.

However, a weaker version of this direction does have a converse, and we will build up to it now. Let's first show that the finite extension of a finite extension remains a finite extension (if you like, the composition of finite extensions is a finite extension)

Theorem 17.3.5: Multiplicity of Finite Extensions (Tower Law)

Let $k \subseteq E \subseteq F$ be fields. Then

$$[F : k] = [F : E][E : k]$$

that is, extension degree multiply each other, where if one side of the equation is infinite, so is the other.

Proof **:**
The infinite case is clear, so we will deal with the finite case. First, if F is finite dimensional over k, then the subspace E is also finite dimensional. Furthermore, any linear combination of elements of F over the field k naturally is a linear combination over the larger field E. Thus, if $[F : k]$ is finite, so is $[F : E]$ and $[E : k]$.

Conversely, suppose $[F : E]$ and $[E : k]$ are both finite. Let $(e_1, e_2, ..., e_m)$ be a basis for E over k and $(f_1, f_2, ..., f_n)$ be a basis for F over E. I claim that

$$(e_1 f_1, e_1 f_2, ..., e_1 f_n, e_2 f_1, ..., e_m f_n)$$

is a basis for F over k.

Let $g \in F$. Then for appropriate $d_i \in E$,

$$g = \sum_j d_j f_j$$

Next, for each d_j, for appropriate $c_{ij} \in k$, we have

$$d_j = \sum_i c_{ij} e_i$$

Thus, we have:

$$g = \sum_i \sum_j c_{ij} e_i f_j$$

Therefore, the products $e_i f_j$ span F, showing that F must at most be finite dimensional.

To show they form a basis, assume

$$\sum_{i,j} \lambda_{ij} e_i f_j = 0$$

Then:

$$\sum_j \left(\sum_i \lambda_{ij} e_i \right) f_j = 0$$

which, since the the elemnts f_j are linearly independent over E, we have that $\sum_i \lambda_{ij} e_i = 0$. But then, the elements e_i are linearly indepenent over e_i,and so each $\lambda_{ij} = 0$, showing that the original linear combination is also linearly independent, as we sought to show.

Notice how this corollary mirrors Lagranges theorem with respect to the order of groups. This parrallel is actually quite a deep one which will capture our attention in the next chapter. The following corollary deepends this connection:

Corollary 17.3.6: Degree Extension Divisibility

Suppose L/F is a finite extension and let K be any subfield of L contianing F, $F \subseteq K \subseteq L$. Then $[K : F]$ divides $[L : F]$

Proof **:**
Since

$$[L : F] = [L : K][K : F]$$

this result holds.

This result gives us a similar power that Lagrange's theorem gives us

Example 17.3.3: Use Of Tower Law

1. We can quickly rule out that $\sqrt{2} \notin \mathbb{Q}(\sqrt[n]{2})$ if n is odd. To see this, if $\sqrt{2} \in \mathbb{Q}(\sqrt[n]{2})$, then $\mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt[n]{2})$. Then by what we've just shown:

   $$[\mathbb{Q}(\sqrt[n]{2}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt[n]{2}) : \mathbb{Q}(\sqrt{2})][\mathbb{Q}(\sqrt{2}) : \mathbb{Q}]$$

   However, that would imply that an even number divides an odd numbers – a contradiction.

2. More generally, we can conclude the following: If K/k is an extension of odd degree and $\alpha \in K$, then α can be written as a polynomial in α^2, that is $\alpha = p(\alpha^2) = \sum_i k_i (\alpha^2)^i$ for $k_i \in k$.

   To see this, we first notice that through simple set theory we must have

   $$k \subseteq k(\alpha^2) \subseteq k(\alpha)$$

and that $k(\alpha)/k(\alpha^2)$. To find the degree d of this extension, notice that over $k(\alpha^2)$, we get that $x^2 - \alpha^2 \in k(\alpha^2)[x]$, so $d \leq 2$. On the other hand, d must divide $[k(\alpha) : k]$ which is odd, forcing us to have $d = 1$. Thus $k(\alpha^2) = k(\alpha)$, and so $\alpha \in k(\alpha^2)$.

3. Let's say K/k and $[K : k]$ is prime. Then K contains *no* fields that contain k. This is similar to how all finite groups of prime order are simple.

4. Take $\mathbb{Q} \subseteq \mathbb{R}$ and consider $\mathbb{Q}/(x^3-3x-1) \cong \mathbb{Q}(\alpha)$ as in example 17.2.2. Since $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$, and $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$, and $2 \nmid 3$, then $\mathbb{Q}(\sqrt{2}) \not\subseteq \mathbb{Q}(\alpha)$, so $\sqrt{2} \notin \mathbb{Q}(\alpha)$! It is much more difficult to directly show $\sqrt{2}$ cannot be written as a $\mathbb{Q}$-linear combination of $1, \alpha, \alpha^2$.

5. More in Dummit and Foote?

Continuing to strive towards a partial converse, we define the following generlization of a simple extension. The key feature about finitely generated fields is that they can be decomposed into simple extensions:

Lemma 17.3.7: Recursive Definition Of Extension Field

$F(\alpha, \beta) = (F(\alpha))(\beta)$, i.e., the field generated over F by α and β is the field generated by β over the field $F(\alpha)$ generated by α.

***Proof* :**
This proof takes complete advantage that by definition, field extensions are the *smallest* field containing F and α (or α and β)

The field $F(\alpha, \beta)$ is the smallest field to contain F, α, β, so $F(\alpha, \beta) \subseteq (F(\alpha))(\beta)$. Conversely, $F((\alpha))(\beta)$ is the smallest field to contain $F(\alpha)$ and β. Since $F(\alpha) \subseteq F(\alpha, \beta)$ and $\beta \in F(\alpha, \beta)$, $(F(\alpha))(\beta) \subseteq F(\alpha, \beta)$, implying

$$F(\alpha, \beta) = (F(\alpha))(\beta)$$

With this, we can break down any finitely generated:

$$K = F(\alpha_1, \alpha_2, ..., \alpha_k) = (F(\alpha_1, \alpha_2, ..., \alpha_{k-1}))(\alpha_k) = ((\cdots F(\alpha_1))(\alpha_2)) \cdots (\alpha_n)$$

Thus, we get the following chain of extensions:

$$F = F_0 \subseteq F_1 \subseteq F_2 \subseteq \cdots \subseteq F_k = K$$

where

$$F_{i+1} = F_i(\alpha_{i+1}) \qquad i = 0, 1, ..., k-1$$

Importantly, this reduces everything down to lots of simple extensions! Now, suppose that the elements $\alpha_1, \alpha_2, ..., \alpha_k$ are algebraic over F of degrees $n_1, n_2, ..., n_k$

$$[F(\alpha_1) : F] = n_1, \ [F(\alpha_2) : F] = n_2 \cdots [F(\alpha_k) : F] = n_k$$

By proposition 17.2.9, we can replace F with F_i. Then these can all be characterized in terms of simple extensions, meaning we can use proposition 17.3.2. The relative extension degree $[F_{i+1} : F_i]$ is equal to the degree of the minimal polynomoial of α_{i+1} over F_i, which is at most n_{i+1} (and is equal

to n_{i+1} if and only if the minimal polynomial of α_{i+1} over F remains irreducible over F_i). So, by the Tower Law:

$$[K : F] = [F_k : F_{k-1}][F_{k-1} : F_{k-2}] \cdots [F_1 : F_0] \leq [F(\alpha_1) : F][F(\alpha_2) : F] \cdots [F(\alpha_k) : F] \tag{17.1}$$

And so $[K : F]$ must be a *finite* extension of degree less than or equal to $n_1 n_2 \cdots n_k$! We sum up this entire discussion of finite degrees in the following theorem:

Theorem 17.3.8: Degree of Algebraic Extension

The extension K/F is finite if and only if K is generaed by a finite number of algebraic elements over F. More precisely, a field generated over F by a finite number of algebraic elements of degrees $n_1, n_2, ..., n_k$ is algegraic of degree $\leq n_1, n_2, ..., n_k$

Before, we hade

$$\alpha \text{ algebraic over } F \Leftrightarrow F(\alpha)/F \text{ finite}$$

and when we try to generlize, we loose one side of the implication

$$K \text{ algebraic over } F \Leftarrow K/F \text{ finite}$$

with the counter example 17.3.2 for $\Rightarrow$.

We now restricted the hypothesis to give us the if and only if again:

$$K \text{ generated by finite number of algebraic elements over } F \Leftrightarrow K/F \text{ finite}$$

Proof **:**
Let the extension K/F be finite, so $a_1, a_2, ..., a_n$ forms a basis for K over F. Then since $F(a_i) \subseteq K$ as a F-vector-space, $[F(a_i) : F] \leq [K : F]$, meaning it has finite degree, and so by 17.3.2, a_i is algebraic. Since a_i was arbitrary, this means that every a_i is algebraic. Thus K is generated by a finite number of algebraic elements over F.

For the converse, let's say K is generated by a finite number of algebraic elements over F. Let

$$\alpha_1, \alpha_2, ..., \alpha_n$$

all be algebraic over F and generate K. Since they generated K, it is clear that $K = F(\alpha_1, \alpha_2, ..., \alpha_n)$. Then by what we've seen in equation (17.1), we have that K/F is finite.

Using what we've found, it is easier to compute the degree of non-simple extensions:

Example 17.3.4: Computing Degree of non-simple Extension

1. The extension $\mathbb{Q}(\sqrt[6]{2}, \sqrt{2})$ is $\mathbb{Q}(\sqrt[6]{2})$ since $\sqrt{2} \in \mathbb{Q}(\sqrt[6]{2})$. Another way of showing this is that the degree d of $\sqrt{2}$ over $\mathbb{Q}(\sqrt[6]{2})$ is 1. Thus, even though the degree of $\sqrt{2}$ is 2 and the degree of $\sqrt[6]{2}$ is 6, the degre of $\mathbb{Q}(\sqrt[6]{2}, \sqrt{2})$ is 6, not 12.

2. Consider $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. We'll proceed to find the degree of this extensions over $\mathbb{Q}$. Since $\sqrt{3}$ is

of degree 2 over $\mathbb{Q}$, then the degree of the extension $\mathbb{Q}(\sqrt{2},\sqrt{3})/\mathbb{Q}(\sqrt{2})$ is at most 2 ($\sqrt{3}$ can at most add another dimension), and *is* 2 if and only if x^2-3 is irreducible over $\mathbb{Q}(\sqrt{2})$. To see if it is, let's suppose it is reducible. Since x^2-3 is of degree 2, by proposition 11.3.4, it is irreducible if and only if it has a root, that is, if $\sqrt{3}\in\mathbb{Q}(\sqrt{2})$.

If $\sqrt{3}\in\mathbb{Q}(\sqrt{2})$, $\sqrt{3}=a+b\sqrt{2}$. From here, it's a lot of manipulation (squaring both sides and checking cases). It will turn out that every possibility wil make either $\sqrt{2}$, $\sqrt{3}$ or $\sqrt{6}$ rational – a contradiction.

Thus, $\deg_{\mathbb{Q}(\sqrt{2})}\mathbb{Q}(\sqrt{2},\sqrt{3})=2$, so

$$\deg_{\mathbb{Q}}\mathbb{Q}(\sqrt{2},\sqrt{3})=4$$

and we can find the basis for it by "closing" the set $1,\sqrt{2},\sqrt{3}$ to produce $1,\sqrt{2},\sqrt{3},\sqrt{6}$, giving us

$$\mathbb{Q}(\sqrt{2},\sqrt{3})=\left\{a+b\sqrt{2}+c\sqrt{3}+d\sqrt{6}|a,b,c,d\in\mathbb{Q}\right\}$$

Corollary 17.3.9: Algebraic Elements are Closed

Suppose α and β are algebraic over F. Then $\alpha\pm\beta$, $\alpha\beta$, α/β (for $\beta\neq 0$, are all algebraic (since α/β is algebraic $1/\alpha$ is also algebraic)

***Proof* :**
All these elements are part of $F(\alpha,\beta)$, which is finite over F by theorem 17.3.8, and so are algebraic by corollary 17.3.3 (in particular, $F(\alpha+\beta)$ is a simple extension, and since $\alpha+\beta\in F(\alpha,\beta)$(and $F\subseteq F(\alpha,\beta)$), we have that $F(\alpha+\beta)\subseteq F(\alpha,\beta)$ by definition, so it is a finite extension, so by corollary 17.3.3 every element will be a root of a polynomial over F).

Corollary 17.3.10: Algebraic Elements Form a Field

Let L/k be an arbitrary extension. Then the collection of elements of k that are algebraic over F form a subfield K of F.

***Proof* :**
By the previous corrollary (17.3.9), we get that the collection is closed under $+$ and $\times$. From here, it is easy to verify any extra axiom we want.

Example 17.3.5: Algebraic Closure

1. Consider the extension $\mathbb{C}/\mathbb{Q}$, and let $\overline{\mathbb{Q}}$ represent the subfield of $\mathbb{C}$ containing all elements that are algebraic over $\mathbb{Q}$. Then $\sqrt[n]{2}\in\overline{\mathbb{Q}}$ for all $n\in\mathbb{N}$ (taking the positive roots), so $[\overline{\mathbb{Q}}:\mathbb{Q}]\geq n$ for every $n>1$. Thus, $\overline{\mathbb{Q}}$ is an *infinite* algebraic extension of $\mathbb{Q}$, called the field of *algebraic numbers*. We have now proved what we have assumed in example 17.3.2.

2. Take $\mathbb{R}$ and the subfield $\overline{\mathbb{Q}}$ consisting of all algebraic elements over $\mathbb{Q}$, and consider $\mathbb{R}\cap\overline{\mathbb{Q}}$. This set is countable: $\mathbb{Q}$ is countable, and $\mathbb{Q}[x]$ consists of the countable union of polynomials

of degree n ($S_0 \cup S_1 \cup \cdots$ where S_n contains all polynomials of degree n), so $\mathbb{Q}[x]$ is countable. $\overline{\mathbb{Q}}$ must have a cardinality at least less than $\mathbb{Q}[x]$, and so is also countable.

Since $\mathbb{R}$ is uncountable, we see that there must be elements in $\mathbb{R}$ which are *not* algebraic (i.e. transcendental) over $\mathbb{Q}$. In fact, the vast majority of elements are transcendental. The subfield $\overline{\mathbb{Q}} \cap \mathbb{R}$ is a *proper* subfield of $\mathbb{R}$.

Though the majority of elements[a] are transcendental, it is extremely difficult to show that an element *is* transcendental. For example, it was proved that $\pi = 3.1415926...$ was transcendental in 1880, and $e = 2.171828...$ is transcendental in 1870. Proving these are transcendental is very involved. Even combining these elements is tricky to prove: e^π was proven to be transcendental in 1934, however as of writing this book, it is unknown whether π^e is transcendental. In fact, it is also not known whether $\pi + e$ or πe are transcendental.

We actually have a small list of numbers which we *know* to be transcendental, even though the vast majority of them are transcendental [b]. As of writting this, there are approximately 24 classes of known transcendental numbers, including numbers like e^a for an algebraic number a, a^b where a is albebraic $\neq 0, 1$ and b is irrational (ex. $2^{\sqrt{2}}$), $\sin(a)$ for an algebraic number a, and so on (commented here link to wikipedia, since it was not liking some characters in it)

[a] In terms of probability, 100% of the elements

[b] If you've taken some statistics/measure theory, then if $\mathbb{A}$ are the algebraic numbers over $\mathbb{Q}$, $\int \chi_{\mathbb{A}} = 0$, i.e., the chances of picking an algebraic number from the set of real number is 0!

We can extend the results of theorem 17.3.8 a bit further to algebraic extensions (finite or infinite)

Theorem 17.3.11: Composing Algebraic extensions

Let $k \subseteq E \subseteq F$ be field extensions. Then F/k is algebraic if and only if F/E and E/k are algebraic.

Proof **:**
If F/k is algebraic, then *a fortiori* F/E is algebraic (If $\alpha \in F$ are all roots of some polynomial $k[x]$ and $k \subseteq E$, then they are certainly roots in $E[x]$). Since $E \subseteq F$, *a priori* E/k.

For the other way, assume E/k and F/E are algebraic, and consider $\alpha \in F$. Since F/E is algebraic, there exists a $p(x) \in E[x]$ such that

$$p(\alpha) = \alpha^n + e_{n-1}\alpha^{n-1} + \cdots + e_1\alpha + e_0 = 0$$

We thus have that α is algebraic over the subfield $k(e_0, ..., e_{n-1}) \subseteq E$. Hence, by theorem 17.3.8

$$k(e_1, ..., e_{n-1}) \subseteq k(e_1, ..., e_{n-1}, \alpha)$$

is a finite extension. Furthermore, by theorem 17.3.8 again, we have that

$$k \subseteq k(e_1, ..., e_{n-1})$$

is also a finite extension, since each e_i is in E, so is algebraic over k by assumption that E/k is algebaric. Hence, by the Tower Law,

$$k \subseteq k(e_1, ..., e_n, \alpha)$$

is a finite extension as well, and hence α is algebraic over k. Since α was arbitrary, F is algebraic over k, as we sought to show.

To close off this section, we show how our previous discussion applies to composite fields:

Proposition 17.3.12: Degree Of Composite Field

Let K_1 and K_2 be two finite extensions of a field F contained in K. Then

$$[K_1K_2 : F] \leq [K_1 : F][K_2 : F]$$

with equality if and only if an F-basis for one of the fields remains linearly independend over the over field. If $\alpha_1, \alpha_2, ..., \alpha_n$ and $\beta_1, \beta_2, ..., \beta_n$ are bases for K_1 and K_2 over F respectively, then the elements $\alpha_i\beta_j$ for $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$ span K_1K_2 over F

***Proof* :**
exercise

summing up the result in a diagram:

$$\begin{tikzcd}
 & K_1K_2 \arrow[dl, dash, "\leq n"'] \arrow[dr, dash, "\leq m"] & \\
K_1 \arrow[dr, dash, "n"'] & & K_2 \arrow[dl, dash, "m"] \\
 & F &
\end{tikzcd}$$

Corollary 17.3.13: Degree of Composite Field and Equality

Suppose that $[K_1 : F] = n$, $[K_2 : F] = m$ in proposition 17.3.12, where n and m are relatively prime $(n, m) = 1$. Then $[K_1K_2 : F] = [K_1 : F][K_2 : F] = nm$

***Proof* :**
In general, the extension is divisible by n and m since K_1 and K_2 are subfields of K_1K_2 (so divisible by its least common multiple). Since $(m, n) = 1$, then $[K_1K_2 : F]$ is divisible by nm. Since by proposition 17.3.12, $[K_1K_2 : F] \leq nm$, we get $[K_1K_2 : F] = nm$

Exercise 17.3.1

1. (practice quadratic extensions here, see Dummit and Foote example)
2. Show that If Let E/k and F/k be field extensions. Show that if E/k is algebraic, then so is EF/F. Show that if E/k and F/k are algebraic, so is EF/F.

17.4 Historical Detour: Straightedge and Compass Construction

For millennia in mathematical history, mathematicians as far back as ancient greeks tried to construct shapes or compute numbers through geometric means by using a straight edge and compass. For example, using a straight edge and compass, there is a way to compute sums, products, quotient (which would have been known as ratios to the ancient greeks), and square-roots. Geometry, mathematicians have figured out how to double the area of the square, take exactly half the angle, and navigate using the stars [15]. These all conformed with the rules of Euclid's geometry, and so straight-edge and compass constructions seemed like a powerful algorithm for geoemtry and arithmetics. However, a few problems were open since at least since the times of the ancient greeks around 400 BC, namely:

1. Given a square, using a straight-edge and compass, can we find a circle with the same area

2. Given a cube, using a straight-edge and compass, can we find another cube that's double it's volume

3. Given an angle, using a straight-edge and compass, can we triscect the angle (i.e. split the angle into 3 equa angles)

These problems don't have much use practically. However the simpliciy in their statement, the ubiquity of using rulers and compasses in many fields of engeneering, navigation, and so on, as well as the belief that the foundation of the world's geometry is inherently euclidean (in particular, that geometric reality followed euclidean rules and were provable using euclid's simple axioms), made these problems the central attention of a huge plethora of mathematicians for over 2 millennia.

With the new tools of field theory, we can characterize the types of numbers a straight-edge and compass can give us:

1. using a straight-dge construction, the 4 basic operations $(+, -, \times, \div)$ can be derivd. Thus, it will form a field.

2. the compass constructions allows the performance of square-roots. Translating this into the cartesian plane, it will be the intersction of a straight-line and a circle, which will sometimes gives polynomials which will have solutions *not* inside the base field, and so we would use an algebraic extension. Such polynomials will always be of degree 2. Hence, if we keep trying to extend these fields, we will always get multiples of 2 as the degree of the extension

From this comes the theorem that feels so mind-blowing that I can state after literal millenia of people trying to solve this problem:

[15] and an additional instrument to measure the angle between the person and the horizon, and the north star

Theorem 17.4.1: Straightedge And Compass Imposibilities

The 3 classical greek straight-edge and compass problems:

1. **Doubling the square** Using only straight edge and compasss, construct a cube with precisely twice hte volume of a given cube
2. **Triscecting an Angle** Given a straightedge and compass, triseced any given angle θ
3. **Squaring the circle** Given a straightedge and compass, construct a square whose area is precisely the area of a given circle.

are all answered in the negative

Proof **:**
Comes down to the aforementioned intuition and a bit more: Dummit and Foote p. 553

17.5 Closures of Fields

Through theorem 17.2.1, we saw that if we have a irreducible polynomial $f(x) \in k[x]$, we can find a larget field which gives $f(x)$ a root: $f(x) = f'(x)(x - \alpha)$. However, we have only established result for 1 polynomial, and for only 1 root. It would be nice if the set of all roots of every polynomials formed a field, and even better if that field existed for any field k and was unique (up to isomorphism). In such a field, every polynomial would *fully factor* or *split* into linear terms. Through corrollary 17.3.10, we found that given K/k the set of algebraic elements $\alpha \in K$ over k forms a field, namely the set of elements from the larger field K which are algebraic over k. In example (17.3.5) (1), we where implicitely using the fact that every polynomial over $\mathbb{C}$ reduces to linear terms, allowing us to say that $\overline{\mathbb{Q}}$ to be algebraically closed. It would be much better if the definition would be independent of any larger field, that is, find a field $\overline{k}$ such that the only irreducible polynomials in $\overline{k}[x]$ are of degree 1. Furthermore, if $\overline{k}$ existed for every field and was unique, we can call it *the* algebraic closure of k. This is the first thing proven in this section. Then, we take a closer look at fields for which only some subset $S \subseteq k[x]$ have polynomials that split. These fields will be at the heart of Galois theory, for as we saw in corrollary 17.2.15, the group of k-automorphism $\mathrm{Aut}_k(F)$ depends on the number of roots of a polynomial in F. Thus, we take some time constructing such a field.

17.5.1 Algebraic Closure

We start off this section by reminding the reader what an algebraically closed field is:

Definition 17.5.1: Algebraically Closed

Let K be a field. Then K is called algebraically closed if every polynomial in $K[x]$ factors into a product of linear terms.

There are many equivalent ways of defining algebraic closure:

Lemma 17.5.2: Algebraic Closure Equivalences

Let K be a field. then the following are equivalent:

1. K ia algebraically closed
2. every maximal ideal in $K[x]$ is of the form $(x - c)$ for all $c \in K$
3. There does not exist nontrivial algebraic extensions of K
4. If L/K is some field extension, and $\alpha \in L$ is algebraic over K, then $\alpha \in K$

Proof **:**
exercise

These show that being an algebraically closed field K means that there are no more algebraic extensions L/K, besides trivial ones. Given some field k, there can be many field extensions that are algebraically closed. For example, $\overline{\mathbb{Q}}/\mathbb{Q}$, $\mathbb{C}/\mathbb{Q}$, and $\mathbb{C}(t)/\mathbb{Q}$ are all field extensions of $\mathbb{Q}$ to algebraically closed fields. However, the first extensions seems in some way more simple or "minimal" than the other in that it is defined through only algebraic elements (recall form example (17.3.5)(1) that all the elements of $\overline{\mathbb{Q}}$ are algebraic over $\mathbb{Q}$), while $\mathbb{C}$ contains transcendental elements (we've mentioend in example (17.3.5)(2) how π and e are not algebraic).

Definition 17.5.3: Algebraic Closure

Let k be any field. Then an *algebraic closure* of k is a field $\overline{k}$ where $\overline{k}/k$ is an *algebaric extension* and $\overline{k}$ is algebraically closed.

Soon, the statement can be re-written with "the algebraic closure" and "the field" instead of "an algebraic closure" and "a field". The field $\overline{k}$ will be the smallest field in which every nonconstant polynomial $p(x) \in k[x]$ will split into linear factors, and since $\overline{k}$ is algebraically closed, the subfield of $\overline{k}$ that contains all the roots of all non-constant polynomials in $k[x]$ is $\overline{k}$ itself. To show the existence of such a field, we first prove that there exists a field extension where every polynomial in $k[x]$ has *at least* one root, and then recursively apply the theorem to get a long chain of field extensions, each containing at least one more root to every polynomial.

Lemma 17.5.4: Adding One Root

Let k be a field. Then there exists a field K such that every non-constant polynomial of $k[x]$ admits at least 1 root.

The following proof is credited to Emil Artin

Proof **:**
Let $\mathcal{F}$ be a set that is in bijection with all non-constant monic polynomials $F \subseteq k[x]$: $\mathcal{F} = \{t_f\}_{f \in F}$, and let $k[\mathcal{F}]$ be a polynomial ring with indeterminates $t_f \in \mathcal{F}$. Let $I \subseteq k[\mathcal{F}]$ be the set of all polynomials generated by all $f(t_f)$ (that is, the polynomial f that takes in the indeterminate t_f that it is in bijective correspondence).

I claim that I is a proper ideal of $k[\mathcal{F}]$. Indeed, if it weren't, then we would have:

$$1 = \sum_i^n a_i f(t_{f_i})$$

with $a_i \in k[\mathcal{F}]$. The identity cannot be possible, for that would lead to imply $1 = 0$. Let's take $k \subseteq F$ where each polynomial $f_1(x), ..., f_n(x)$ has a root $\alpha_1, ..., \alpha_n$ (apply theorem 17.2.1 n times). Then viewing the previous identity to be in $F[\mathcal{F}]$ we get:

$$1 = \sum_i^n a_i f(t_{f_i}) = \sum_i^n a_i f(\alpha_i) = \sum_i^n a_i \cdot 0 = 0$$

a contradiction.

Now, since I is proper, it is contained in a maximal ideal m (proposition 9.4.10). Thus, we have the field extension:

$$k \subseteq K := \frac{k[\mathcal{F}]}{m}$$

By construction every non-constant monic (and so every nonconstant) polynomial $f(x) \in k[x]$ must have a root in K, namely $\overline{t_f}$

Note that we had to use Zorn's lemma to find the maximal ideal m. Very commonly, a statement using Zorn's lemma is equivalent to it (ex. instead of using Zorn's lemma to prove the existence of a maximal ideal, we can use the existence of a maximal ideal to prove Zorn's lemma). We might ask if the existence if the previous result is also equivalent to Zorn's lemma, however it turns out that this is not the case (it is apparently due to the *compactness theorem of first order logic*, which is known to be weaker than the axiom of choice)

We can now find a field K_1 in which every polynomial in $k[x]$ has at least 1 root. We can continue this process by applying the theroem on K_1, and defining the field K_2 for which every polynomial of K_1 has at least 1 root in $K_1[x]$ (and every polynomial of degree 2 in $k[x]$ must not split into linear factors). Continuing this process, we get a chain:

$$k \subseteq K_1 \subseteq K_2 \subseteq \cdots$$

Now, take $L = \bigcup_i K_i$[16]. L can easily be turned into a field through the operation of $a + b$ and $a \cdot b$ where we choose K_i in which $a, b \in K_i$.

Lemma 17.5.5

Let L be defined as above. Then L is algebraically closed

***Proof* :**
Let $f(x) \in L[x]$ be a nonconstant polynomial. Then since $f(x)$ has finitely many coefficients, there exists some field K_i for some i which has all the coefficients of $f(x)$. Thus, $f(x)$ has a root in $K_{i+1} \subseteq L$: $f(x) = f'(x)(x - \alpha)$. Since $f'(x) \in L[x]$ (the coefficients of $f'(x)$ must be in some K_j, $j \geq i$), this process repeats until f is completely split into linear terms. Since this is true for arbitrary $f(x) \in L[x]$, L is algebraically closed, completing the proof.

[16]This is an example of a *direct limit*

Thus, for any field k, there exists an algebreaically closed field L such that L/k. Note that L need not be algebraic closure, since L/k need not be an algebraic extension. However, with an algebraically closed field, we can very simply construct an algebraic closure of k:

Theorem 17.5.6: Existence Of Algebraic Closure

Let k be a field and L/k be a field extension where L is algebraically closed. Then

$$\overline{k} := \{\alpha \in L \mid \alpha \text{ is algebraic over } k\ \}$$

is an algebraic closure of k

Note how this mimicked $\overline{\mathbb{Q}}$ in example (17.3.5)(2) assuming $\mathbb{C}$ is algebraically closed.

Proof **:**
First, by corrollary 17.3.10, $\overline{k}$ is a field, and *a prior* $\overline{k}$ is an algebraic extension of k. To show that k is algeraically closed, let α be algebraic over $\overline{k}$. Then

$$k \subseteq \overline{k} \subseteq \overline{k}(a)$$

is a composition of albegraic extensions, hence $\overline{k}(a)$ is algebraic (theorem 17.3.11), thus α is algebraic over k. But then by definition of $\overline{k}$, $\alpha \in \overline{k}$. Thus, by lemma 17.5.2$\overline{k}$ is algebraically closed, completing the proof

Next, we strive for uniqueness. As we establisehd in theorem 17.2.1, being a field extension is not a universal property, and so the best we an strive for is the existence of a homomorphism, though not necesarrily a unique one. We first establish a lemma:

Lemma 17.5.7: Embedding Into Algebraic Extension

Let L/k be any field extension where L is algebraically closed over k. Let F/k be an algebraic extensions. Then there exists a [not necessarily unique] homomorphism $i : F \to L$ fixing k.

Note that *a-priori*, L need not have any relation to F (ex. we don't know if F is a subset of L, they might only intersect on k).

Proof **:**
This is a Zorn's lemma argument similar to Baer's Criterion (theorem 16.3.4). Let S be the set of homomorphisms of the form

$$i_K : K \to L$$

The set S is nonempty since $i_k : k \to L \in S$. Define a patial orer on S $i_K \preceq i_{K'}$ if $K \subseteq K' \subseteq L$ and $i_{K'}|_K = i_K$. To apply Zorn's lemma, we need to show that every chain has an upper bound. Let $C \subseteq S$ be a chain, and let K_C be the union of all the domains of elements $i_K \in C$. Then K_C is clearly a field (lemma ref:HERE), and for every $\alpha \in K_C$ define $i_{K_C}(\alpha)$ to be $i_K(\alpha)$ for appropriate K where $\alpha \in K$. Then i_{K_C} is clearly well-defined, not being dependent on the choice of K, and so we have a homomorphism $i_{K_C} : K_C \to L$ where $i_K \preceq i_{K_C}$ for all $i_K \in C$, showing that i_{K_C} is indeed an upper bound for C. Thus, by Zorn's lemma, Z has an upper bound $i_G : G \to L$ where $k \subseteq G \subseteq L$.

I claim that $G = F$, which will prove the statement. For the sake of contradiction, let's assume this is not the case, so that there exists an element $\alpha \in F \setminus G$. Consider $G(\alpha)/G$. Since $\alpha \in F$ is algebraic over k, α is algebraic over G, and since $\alpha \notin G$, it is the root of an irreducible polynomial $g(x) \in G[x]$. Let $H = i_G(G)$. Then as discussed in the first pages of section 17.2 i_G induces a ring homomorphism

$$f^{i_G} : G[x] \to H[x]$$

Let $h(x) = f^{i_G}(g(x))$. Then $h(x)$ is irreducible over H, and since L is algebraically closed, it must have a root $\beta \in L$.

We are now setup in a situation to apply theorem 17.2.13: we have an isomorphism $i_G : G \to H$ and two simples extensions $G(\alpha), H(\beta)$ with α and β being roots of $g(x)$ and $h(x)$ respectively (with $h(x) = i_G(g(x))$). Thus, i_G lifts to an isomorphism:

$$i_{G(\alpha)} G(\alpha) \to H(\beta) \subseteq L$$

which sends α to β. Then $i_G \preceq i_{G(\alpha)}$, contradicting the maximality of i_G.

Thus, it must be that $G = F$, and so $i_F : F \to L$ is a field homomorphism, as we sought to show.

The fact that the isomorphism is not unique will become important in representing algebraic closures. We first finish off by showing that the algebraic closure is indeed unique up to isomorphism (given a fixed field k)

Theorem 17.5.8: Uniqueness Of Algebraic Closure

Let $k \subseteq \overline{k_1}$ and $k \subseteq \overline{k_2}$ be two algebraic closures of k. Then $\overline{k_1} \cong \overline{k_2}$ where the isomorphism extends the identity on k, or equivalently k_1 is isomorphic to k_2 as extensions. In terms of a commutative diagram, in the following

$$\begin{tikzcd} k_1 \arrow[rr, "\sim"] & & k_2 \\ & k \arrow[ul, hook] \arrow[ur, hook] & \end{tikzcd}$$

there always exists an isomorphism from k_1 to k_2

the second condition for the isomorphism is so that the two fields represent the algebraic closure over the common field k. Sometimes, such a isomorphism (resp. homomorphism) is called a k-isomorphism (resp. k-homomorphism)

***Proof* :**
Since $\overline{k_1}/k$ is algebraic and $\overline{k_2}$ is algebraically closed over k, by lemma 17.5.7, there exists a homomorphism $i : \overline{k_1} \to \overline{k_2}$. This homomorphism is injective, being a field homomorphism, and surjective since otherwise, $\overline{k_1} \subseteq \overline{k_2}$ would be a non-trivial algebraic extension, contradicting lemma 17.5.2 since $\overline{k_1}$ is algebraically closed. Thus, i must be an isomorphism as we sought to show.

The fact that the isomorphism is not necessarily unique has some important consequences: There is no "cannonical" way of representing the algebraic closure. At first, it might seem harmless that the

isomorphism is not unique, as long as the two field are isomorphic. However, this means there is no real cannonical "choice" when it comes to representing algebraic closures.

Take for example $\overline{\mathbb{R}}$, which we know to be $\mathbb{C}$. We can write $\overline{\mathbb{R}}$ as $\mathbb{R}[i]$, which for many mathematicians is "the" algebraic closure of $\mathbb{R}$. However, if we go over to an electrical engineer, we will see that they have also constructed an algebraic closure for $\mathbb{R}$ to represent electrical charge and circuits where they use the letter j (since i is already used for something else), and write it as $\mathbb{R}[j]$. We might think it to be harmless to make the isomorphism $1 \mapsto 1$ and $i \mapsto j$, however the electrical engineers have decided that electrical charge should be represented by a -1. Does this means we should instead have

$$1 \mapsto 1 \qquad i \mapsto -j?$$

Mathematically, there is no good reason to pick one representation over another. In other words, if one person constructs a splitting field L for k and another person constructs a algebraic closure L' for k, then we know that we get:

$$\begin{tikzcd} k_1 \arrow[rr, "\sim"] & & k_2 \\ & k \arrow[ul, hook] \arrow[ur, hook] & \end{tikzcd}$$

however, there is no "natural" way to associate an element of L to an element of L'. This often means that we should keep track of the construction of the splitting field we have done, since if we don't we might not be able to communicate to other mathematicians who have done a different construction (how can you compare your work if you can't correspond the results you have come up with with the other mathematicians).

The idea there there is no natural choice of isomorphism can be made more rigorous if we ask it in the language of category theory. If we map every field to it's algebraic closure $k \to \overline{k}$, does this define a functor? In order for it to define a functor, we would need that every map $k \to l$ of field defines a map of algebraic closures

$$\begin{tikzcd} k \arrow[r] \arrow[d] & \overline{k} \arrow[d, dashed] \\ l \arrow[r] & \overline{l} \end{tikzcd}$$

However, it is not clear what map should be chosoen here[17]. There is in fact a way to reduce this ambiguity, which I don't fully undertand yet intuity, but it was mentioned in Richard Borcherds video: here.

17.5.2 Splitting Field and Normal Extensions

Having established that every field has an algebraic closure, meaning we can unambigurously talk about roots of polynomials, we now return to analyzing algebraic extensions. We start off by zoom in on fields that only split some subset of polynomials $S \subseteq k[x]$. In the following sections, we will be working over the case where S is finite [18]. Since S is finite, say $S = \{f_1(x), ..., f_n(x)\}$, then having each $f_i(x)$ split is equivalent to having $f_1(x)f_2(x) \cdots f_n(x)$, thus, we usually have $S = \{f(x)\}$ be one polynomial

[17] perhaps you can force through a construction using the axiom of choice, but it is certainly not a natural choice
[18] The infinite case of this I'm not yet sure is not studied TBD

Definition 17.5.9: Splitting Field

Let k be a field and $f(x) \in k[x]$ be a polynomial where $\deg(f(x)) = d$. Then a *splitting field* for $f(x)$ over k is a minimal extension F/k such that

$$f(x) = c\prod_{i}^{d}(x - \alpha_i)$$

i.e. $f(x)$ splits completely in $F[x]$, and $f(x)$ does not factor in any proper subfield of F.

Furthermore, by minimality, $F = k(\alpha_1, ..., \alpha_n)$ (i.e. F is generated by the roots of $f(x)$ in F)

The fact that a splitting field exists is immediate from the fact that $\overline{k}$ exists, namely if $f(x)$ is a polynomial, then $F \subseteq \overline{k}$ generated by the roots of $f(x)$ in $\overline{k}$. Furthermore, the splitting field is in fact unique (up to isomorphism), justifying replacing any "a splitting field" in the definition to "the splitting field"

Theorem 17.5.10: Uniqueness Of Splitting Field

Let k be a field and let $f(x) \in k[x]$. Then the splitting field F for $f(x)$ over k is unique up to isomorphism, and $[F : k] \leq (\deg(f))!$ (i.e. the factorial of the degree of f). Furthermore, if $i : k \to k'$ is an isomorphism, $f(x) \in k[x]$ and $g(x) = i(f(x))$, then i extends to an isomorphism from the splitting field of $g(x)$ over k' to the splitting field of $f(x)$ over k:

$$\begin{tikzcd} F \arrow[r, dashed, "\sim"] & F' \\ k \arrow[u, hook] \arrow[r, "\sim"] & k' \arrow[u, hook] \end{tikzcd}$$

***Proof* :**
We will first construct a splitting field explicitely with $[F : k] \leq \deg(f)$, and then use that field to show uniqueness. For existence, this is an application of induction on the degree of the polynomial. Degree 1 is trivial, covering the base-case, so let's say the splitting field was constructed for all fields and all polynomials of degree $\deg(f) - 1$. Let $f(x)$ be some polynomial that is not yet fully factored into linear terms; let $q(x)$ be some irreducible non-linear factor over k. Then

$$k \subseteq F' := \frac{k[x]}{(q(x))}$$

is an extension of degree $\deg q \leq \deg f$ in which $q(x)$ has a root (and so $f(x)$ has a root), say α, and so $q(x) = q'(x)(x - \alpha)$. Then we can define the polynomial $h(x) = f(x)/(x - \alpha) \in F'[x]$ of degree $\deg(f) - 1$. Thus, by the induction hypothesis, there exists a splitting field F for $h(x)$ over F' and

$$[F : F'] \leq (\deg(f) - 1)!$$

Since the factors of $f(x)$ over F are $(x - \alpha)$ and the factors of $h(x)$, it follows that F is a splitting field for $f(x)$, and

$$[F : k] = [F : F'][F' : k] \leq (\deg(f) - 1)!(\deg(f)) = (\deg(f))!$$

To prove that an isomorphism $i : k \to k'$ extends to isomorphisms of splitting fields as stated, let F be the splitting field for $f(x)$, and consider $k \to k' \subseteq \overline{k'}$. Since the extension $k \subseteq F$ is algebraic, by lemma 17.5.7 there exists a k-homomorphism $i : F \to \overline{k'}$. Since F is generated over k by the roots of $f(x)$ in F, we have

$$i(F) = k'(\alpha_1, ..., \alpha_d) \subseteq \overline{k'}$$

where the α_i's are the roots of $g(x) = i(f(x))$ in $\overline{k'}$. Therefore, $L := i(F)$ is independent of the given isomorphism i and the splitting field F. Applying this to i and to the identity $k' \to k'$ completes the proof (since both F and G are isomorphic to L)

Example 17.5.1: Splitting Fields

1. The splitting field of $x^2 - 2$ over $\mathbb{Q}$ is $\mathbb{Q}(\sqrt{2})$, since the two roots are $\pm\sqrt{2}$ and $-\sqrt{2} \in \mathbb{Q}(\sqrt{2})$. The splitting field of $(x-2)^n$ is also $\mathbb{Q}(\sqrt{2})$ for all $n \in \mathbb{N}_{>0}$, since we only need to add the root $\sqrt{2}$ (resp. $-\sqrt{2}$) for this polynomial to split.

2. Splitting field for $(x^2 - 2)(x^2 - 3)$ over $\mathbb{Q}$ is $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. Note that $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is equal to the composite field $\mathbb{Q}(\sqrt{2})\mathbb{Q}(\sqrt{3})$, i.e. $\mathbb{Q}(\sqrt{2})\mathbb{Q}(\sqrt{3}) = \mathbb{Q}(\sqrt{2}, \sqrt{3})$.

3. More generally, If $f(x)$ is a polynomial over k such that

$$f(x) = \prod g_i(x)$$

   where each $g_i(x)$ is irreduible over k, then the splitting field of $f(x)$ is the composite field of the splitting fields for each $g_i(x)$ since each splitting field is the smallest field containing all the roots of g_i, and the composite field is the smallest field containing all the splitting fields.

4. splitting field of $x^3 - 2$ over $\mathbb{Q}$ is not $\mathbb{Q}(\sqrt[3]{2})$, since we are missing two roots, namely:

$$\sqrt[3]{2}\left(\frac{-1 + i\sqrt{3}}{2}\right) \qquad \sqrt[3]{2}\left(\frac{-1 - i\sqrt{3}}{2}\right)$$

   Thus, the splitting field is $\mathbb{Q}(\sqrt[3]{2}, i\sqrt{3}) = \mathbb{Q}(\sqrt[3]{2}, \sqrt{-3})$.

5. The splitting field of $x^2 + 1$ over $\mathbb{Q}$ is $\mathbb{Q}(i)$, and the splitting field of the same polynomial $x^2 + 1$ over $\mathbb{R}$ is $\mathbb{C}$, showing that the splitting field is dependent on the base field.

6. Let F be the splitting field of $x^8 - 1$ over $\mathbb{Q}$. With some knowledge of complex analysis, we see that all the roots are 8th root of unit, that is,

$$\zeta = e^{\frac{2\pi i}{8}}$$

   is a generator for the roots (i.e. the primite elements), and so $F = \mathbb{Q}(\zeta)$. Visually, these roots look like so:

In fact, ζ is also the root of x^4+1 (since $\zeta^4 = -1$), and so is the root of the smaller irreducible polynomial $x^4 + 1$ over $\mathbb{Q}$. There is no smaller minimal polynomial for which ζ is a root (since smaller powers of ζ is not in $\mathbb{R}$) and so:

$$[\mathbb{Q}(\zeta) : \mathbb{Q}] = 4$$

showing that the degree of the splitting field is much less than $8! = 40'320$. To better undersatnd this field, notice that $\zeta^2 = i$, so $\sqrt{2} = \zeta + \zeta^7$. Thus:

$$\mathbb{Q}(i, \sqrt{2}) \subseteq \mathbb{Q}(\zeta)$$

Conversely, $\zeta = \frac{\sqrt{2}}{2}(1 + i) \in \mathbb{Q}(i, \sqrt{2})$, showing that:

$$\mathbb{Q}(i, \sqrt{2}) \supseteq \mathbb{Q}(\zeta)$$

showing that they are in fact equal.

As an exercise, show that $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(i, \sqrt{2})) = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$

7. Let $p(x) = x^4 - 1$. this polynomial in fact factors over $\mathbb{Q}$:

$$x^4 - 1 = (x - 1)(x + 1)(x^2 + 1)$$

and so the splitting field of $x^4 - 1$ is the splitting field of $x^2 + 1$, which is $\mathbb{Q}(i)$

8. Let $p(x) = x^4 + 2$. This polynomial is irreducible by Eisenstein. Given $\zeta = e^{\frac{2\pi i}{8}}$, then the roots of $p(x)$ are:

$$\sqrt[4]{2}\zeta,\ \sqrt[4]{2}\zeta^3,\ \sqrt[4]{2}\zeta^5,\ \sqrt[4]{2}\zeta^7$$

and so the splitting field is $K = \mathbb{Q}(\sqrt[4]{2}\zeta,\ \sqrt[4]{2}\zeta^3,\ \sqrt[4]{2}\zeta^5,\ \sqrt[4]{2}\zeta^7)$. We can simply the adjoint elements with the following observations:

$$K \subseteq \mathbb{Q}(\zeta, \sqrt[4]{2}) \subseteq \mathbb{Q}(i, \sqrt{2}, \sqrt[4]{2}) = \mathbb{Q}(i, \sqrt[4]{2})$$

so $K \subseteq \mathbb{Q}(i, \sqrt[4]{2})$. For the $\supseteq$ direaction, notice that:

$$\sqrt{2} = \frac{(\sqrt[4]{2}\zeta)^3}{\sqrt[4]{2}\zeta^3} \in K \qquad i = \frac{(\sqrt[4]{2})^2}{\sqrt{2}} \in K$$

So $\zeta = \frac{\sqrt{2}}{2}(1 + i) \in K$ and $\sqrt[4]{2} = \frac{\sqrt[4]{2}\zeta}{\zeta} \in K$, showing that $K \supseteq \mathbb{Q}(i, \sqrt[4]{2})$. With a little bit of computation, we see that:

$$[\mathbb{Q}(i, \sqrt[4]{2})] = 8$$

From Dummite and Foote. There are in fact a ton of good example that I can put here, check them out! Overall, small variation in polynomials can create quite large differences in splitting fields (as we've seen with x^4+2, x^4+1, x^4-1). This shows how precise the tools of field theory are in the study of polynomials.

Given a polynomial $f(x)$ over k, we can find a field F in which $f(x)$ factor completely over F. In fact, we have also gained a slightly stronger result for free: if $g(x)$ is irreducible in k has a root in F, then it in fact *also* splits. The idea is that the automorphism group acts transitively on all the roots of an irreducible polynomials (recall the discussion under corollary 17.2.15). If all the roots are inside the field, then the automorphism group will shuffle around all the roots within the field. Hence, if we are given that a single root is inside a field, and that the image of each automorphism in the automorphism group is inside the field, we would have that all the roots are within the field. As the action of the automorphism group is a conjugate action, another way of saying it is that the field is closed under conjugation of the automorphism group. To understand this phenomenon better, we give a name to fields with this property:

Definition 17.5.11: Normal Extension

let F/k be a field extension. Then F/k is called a *normal extension* if for every polynomial $f(x) \in k[x]$, $f(x)$ has a root if and only if $f(x)$ splits as a product of linear factors over F

Some authors like to say that F/k needs to also be *algebraic*. This so that we need not worry about any transcendental elements in the following proofs. It is no harm on making F/k possibly not algebraic, since we work with finite extensions in the majority of cases

A normal extension is *not necessarily* the algebraic closure since not necessarily all irreducible polynomials have a root, however the irreducible polynomials over the base field become the polynomials that split completely in the normal extension, constant polynomials, and irreducible polynomials that have no roots in the normal extension (or, they can be thought of as polynomials that have yet to have a root added to the extension).

The name normal is given to this type of extension is due to the fact that most of the nice properties of automorphism groups we have discussed in corollary 17.2.15 apply to normal extensions[19]

Theorem 17.5.12: Finite Normal Extension Equivalences

Let F/k be a field extension. Then the following are equivalent:

1. F is finite and normal
2. F is the splitting field of some polynomial $f(x) \in k[x]$
3. For all k-homomorphisms, $F \xrightarrow{\varphi} \overline{F}$ where $\varphi(k) = k$, $\varphi(F) = F$

***Proof* :**
We'll first show $2 \Leftrightarrow 3$ since it's simple. For $2 \Rightarrow 3$, let F be the splitting fifeld for some polynomial $f(x) \in k[x]$, then if $F \xrightarrow{\varphi} \overline{F}$ is a k-homomorpism, $F \xrightarrow{\varphi} \varphi(F)$ is a k-isomorphism. Hence, φ must map roots of $f(x)$ to roots of $f(x)$: If α is a root of $f(x)$, then we know that $m_{\alpha,k}(x) \big| f(x)$, and

[19] In fact, normal subgroups got their name from normal extensions! The connection will be made in the chapter 18 when discussing Galois Theory in theorem 18.1.16

by corrollary 17.2.15, $\varphi(\alpha)$ is another root of $m_{\alpha,k}$, and hence $f(x)$. Since F is a splitting field, it contains all the roots of $f(x)$, and so $\varphi(\alpha) \in F$. Since the k-homomorphism is determined by where it maps the roots of the generators of F (namely the roots of $f(x)$), $\varphi(F) \subseteq F$. Since φ must be injective, $F \subseteq \varphi(F)$, so $F = \varphi(F)$.

For $3 \Rightarrow 2$, since all k-homomorphisms $F \xrightarrow{\varphi} \overline{F}$ map $\varphi(F) = F$. If F/k is purely transcendental (i.e., no elements of $\alpha \in F - k$ is the the root of a polynomial), Then then clearly $\varphi(F) = F$ or else φ would not be injective. Thus, if F/k is not purely transcendental, choose an algebraic element $\alpha \in F$ over k, so $m_{\alpha,k}(x)$ is it's minimal polynomial. Since φ is defined by where it maps roots of minimal polynomials over $k[x]$, $\varphi(\alpha)$ is another root of $m_{\alpha,k}(x)$. Since $\varphi(F) = F$, $\varphi(\alpha) \in F$. Using corollary 17.2.15, we can see that there are as many automorphisms as distinct roots of $m_{\alpha,k}(x)$, and so every root of $m_{\alpha,k}(x)$ is in F. Since this is true for every algebraic element, we see that $\varphi(F) = F$.

Next, we show $1 \Rightarrow 2$. Assume F is finite and normal. Since it's finite, $F = k(\alpha_1, \alpha_2, ..., \alpha_n)$, where each α_i is algebraic over k. Next, let $m_i(x)$ be the minimal polynomial of α_i. Since F is normal, and each m_i has at least one root (namely α_i), $m_i(\alpha)$ splits. Then $f(x) = m_1(x) \cdots m_n(x)$ is a polynomial that splits over F. But then, F is the smallest field in which $f(x)$ splits (since $F = k(\alpha_1, ..., \alpha_n)$), and so F is the splitting field for $f(x)$.

Finally, we show $2 \Rightarrow 1$: assume F is the splitting field for some polynomial $f(x) \in k[x]$, and assume that $p(x) \in k[x]$ is an irreducible polynomial over k such that F contains a root α for $p(x)$. If $p(x)$ has degree less than or equal to 2, we're done, so let's say $\deg p(x) \geq 3$. Since $F \subseteq \overline{k}$, let $\beta \in \overline{k}$ be a different root of $p(x)$. We want to show that $\beta \in F$. This will show that F is normal, and by theorem 17.5.10, F must be finite.

Since $\mathrm{id} : k \to k$ is a field homomorphism, by theorem 17.2.13, there is an isomorphism $i : k(\alpha) \to k(\beta)$ sending $\alpha \mapsto \beta$. Note that $F(\beta) \subseteq \overline{k}$. Visually, we can imagine all of this in the commutative diagram:

$$\begin{tikzcd}
 & k(\alpha) \arrow[r, hook] \arrow[dd, hook, "\iota"] & F \arrow[r, hook] & \overline{k} \\
k \arrow[ur, hook] \arrow[dr, hook] & & & \\
 & k(\beta) \arrow[r, hook] & F(\beta) \arrow[r, hook] & \overline{k}
\end{tikzcd}$$

Note that we cannot draw the identity map from $\overline{k}$ to $\overline{k}$, since ι maps $\alpha \mapsto \beta$. Now, notice that F is the *splitting field* $f(x)$ over $k(\alpha)$. Indeed, F contains all the roots of $p(x)$ and is generated over k (and hence $k(\alpha)$) by the roots of $p(x)$. Similarly, $F(\beta)$ is the splitting field for $f(x)$ over $k(\beta)$. Then by the uniqueness of the splitting field (theorem 17.5.10), ι extends to an isomorphism $\iota' : F \to F(\beta)$. Now, since ι restricts to the identity on k, ι' is an isomorphism of k-vector spaces, in particular $\dim_k F = \dim_k F(\beta)$ (since splitting fields are always finite extensions). In particular, we get to take advantage of facts about liner algebra, in particular that if the dimensions of two vector-spaces are the same, then any injective linear transformation is in fact an isomrophism.

Now, we also have the natural inclusion map $\iota : F \to F(\beta)$ which is both a k-homomorphism and a k-vector spaces. Since F and $F(\beta)$ are k-vector spaces of the same [finite!] dimension, i must *also* be an isomorphism, i.e.

$$[F(\beta) : F] = 1$$

Thus, $\beta \in F$, completing the proof.

Note that it can be proven that F is normal (not necessarily finite) if and only if for any k-homomorphism $\varphi : F \to \overline{F}$, $\varphi(F) = F$. The proof naturally generlizes from the proof of $2 \Leftrightarrow 3$ since we can apply the results to each irreducible polynomial and their roots.

45: Remark

Notice how condition (3) of theorem 17.5.12 allows us to talk about "the" embedding of a Normal extension F/k into $\overline{k}$, since $\varphi(F) = F$ for any F. Recall that $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is not normal, and that we have shown that $\mathbb{Q}(\sqrt[3]{2})$ is isomorphic to 3 subfields of $\overline{\mathbb{Q}}$ which do not all overlap:

However, if the extensoin is normal, then there is a notion of "the" embedding since they all overlap:

In this way, having a [algebraic][a] normal extension F/k can be thought of as being more "precise" than a non-normal extension, since any other isomorphic copy of F is just an automorhpism of F.

[a] This is to ensure that $F \subseteq \overline{k}$. The algebraic elements of F is also what we care about when we look at k-homomorphisms $F \to \overline{k}$

(there is also a cool proof the professor gave in class, the Jan 20th lecture)

Example 17.5.2: Consequences
Let's say $p(x)$ is an irreducible polynomial over $\mathbb{Q}$ with a complex root that may be expressed as a polynomial in i and $\sqrt[4]{2}$ over $\mathbb{Q}$. Then *all* roots of $p(x)$ may be expressed as a polynomial in i and $\sqrt[4]{2}$. As we have shown in example 17.5.1, $\mathbb{Q}(\sqrt[4]{2}, i)$ is a splitting field over $\mathbb{Q}$, and hence a normal extension.

Unlike algebraic and finite extensions, normal extensions are more finicky when it comes to composing normal extensions:

Proposition 17.5.13: Tower Law On Normal Extensions

If $L/K/k$ and L/k is normal, then L/K is normal

Proof **:**
Let $f(x) \in K[x]$ be irreducible with a root $\alpha \in L$. Then $f(x) = m_{\alpha,K}(x)$ is the minimal polynomial over K. Note that

$$m_{\alpha,K}(x) | m_{\alpha,k}(x)$$

And since $m_{\alpha,k}(x)$ splits over L (since L/k is normal), $m_{\alpha,K}$ splits over L, as we sought to show.

Note that this result has little wiggle room:

Example 17.5.3: Low Tolerance for Normal Extension
Consider $L/K/k$ where:

$$\begin{array}{c} \mathbb{Q}(\sqrt[3]{2}, \zeta_3) \\ | \\ \mathbb{Q}(\sqrt[3]{2}) \\ | \\ \mathbb{Q} \end{array}$$

where σ_3 is one of the two other roots of $x^3 - 2$. Then while L/k is normal, K/k is *not* normal here (we have yet to add enough roots). Furthermore, if L/K and K/k is normal, this does not imply that L/k is normal. For example:

$$\begin{array}{c} \mathbb{Q}(\sqrt[4]{2}) \\ | \\ \mathbb{Q}(\sqrt{2}) \\ | \\ \mathbb{Q} \end{array}$$

To see this, notice that each extension is of degree 2, and hence is normal (exercise). However, $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}$ is not normal, since 2 of the roots of $x^4 - 2$ are complex, while $\mathbb{Q}(\sqrt[4]{2})$ is contained in the reals, and hence cannot have all the roots.

Notice how this behavior is mimicked by normal subgroups.

Definition 17.5.14: Normal Closure

Let K/k be an algebraic extenion. Then the *normal closure* of K given k is an algebraic extension N/K such that

1. N/k is normal
2. If $N \supseteq N' \supseteq k$, and N'/k is normal, then $N = N'$

Proposition 17.5.15: Existence Of Normal Closure

Let K/k be a finite extension. Then there exists a unique normal closure N up to isomorphism fixing k.

Note the finiteness condition given. It is true for algebraic ones too, but we did not yet prove that

Proof **:**
Since K is finite, let $K = k(\alpha_1, \alpha_2, ..., \alpha_n)$ where each $\alpha_i \in K$ is algebraic over k. Let

$$f(x) = \prod m_{\alpha_i,k}(x) \in k[x]$$

and let N be the splitting field of $f(x)$ over K. Then $[N : K] < \infty$. But then, N is the splitting field of $f(x)$ over k as $K = k(\alpha_1, \alpha_2, ..., \alpha_n)$. But this implies N/k is normal.

Finally, let's say $k \subseteq N' \subseteq N$ and N'/k is normal. Then by normality, N' has to split each $m_{\alpha_i,k}(x)$, and so splits $f(x)$, and so N/k *is* a normal closure.

Exercise 17.5.1

1. (Splitting Field proof using induction)
2. Show that If Let E/k and F/k be field extensions. Show that if E/k is normal, then so is EF/F. Show that if E/k and F/k are normal, so is EF/F.
3. Show that every polynomial of $\mathbb{R}[x]$ of degree ≥ 3 is reducible. Show that every irreducible polynomial is linear or is of degree two with negative discriminant, that is if $ax^2 + bx + c = 0$, then $b^2 - 4ac = 0$.
4. Let L/K be an algebraic field extension, and $\varphi_K : K \to \overline{F}$ be a homomorphism to an algebraically closed field. Then φ_K extends to a homomorphism $\varphi_L : L \to \overline{F}$ (use Zorn's lemma and a proof similar to lemma 17.5.7 with Zorn being applied to (E, φ_E) where E/K is a subextension of L/K and $\varphi_E : E \to \overline{F}$ extends φ_K)

17.6 Separable Extensions

We've now shown that there exists a field K over F such that a polynomial $f(x)$ splits (or a field $\overline{F}$ such that all polynomials in $F[x]$ split). A pattern has been occuring which you might have not noticed: every time we split an irreducible polynomial completely into linear factors, each factor was

unique, that is, we had that the multiplicity of every linear term was 1. Does that need to be the case for any polynomial over any field? The answer is no:

Example 17.6.1: Higher Multiplicity

Let p be a prime number, and take the field $\mathbb{F}_p(t)$. Consider the polynomial:

$$p(x) = x^p - t \in \mathbb{F}_p(t)[x]$$

Then $p(x)$ is irreducible: by Eisenstein's criterion, it is irreducible in $\mathbb{F}_p[t][x]$ (since (t) is prime in $\mathbb{F}_p[t]$), and hence it is prime in $\mathbb{F}_p(t)[x]$ by corrollary 11.2.2. Let u be a root of $p(x)$ in the extension $L/\mathbb{F}_p(t)$ (L can be the algebraic closure, or the splitting field). Then notice that:

$$x^p - t = (x - u)^p$$

Simply expand the right hand side of the equation. Thus, u has multiplicity p, i.e., the minimal polynomial over $\mathbb{F}_p(t)$ for $u \in L$ vanishes p times at u[a].

[a] $p(x)$ must be the irreducible polynomial, since if the degree was smaller then $p(x)$ would be reducible

Thus, in general, a polynomial $f(x) \in k[x]$ can factorize over a splitting field as:

$$f(x) = (x - \alpha_1)^{n_1}(x - \alpha_2)^{n_2} \cdots (x - \alpha_k)^{n_k}$$

where $\alpha_1, \alpha_2, ..., \alpha_k$ are distinct elements in the splitting fields, and $n_1, n_2, ..., n_k$ is the multiplicity of every root. However, this seems to be a rare occurence given the fields we have worked with. In fact, most field we will work with do not have this property. It will also become important in Galois theory that a polynomial would have to to linear polynomials with multiplicity 1[20]; for those who are curious, see what hapens to the automorphism group $\mathrm{Aut}_k(F)$ of the splitting field F if $(x - \alpha)^n$ is the linear factorization of $f(x)$ over k. Hence, we give a name to polynomials that factor with multiplicity 1:

Definition 17.6.1: Separable Polynomial

A polynomial over k is called *separable* if it factors into linear terms with multiplicity 1 in it's splitting field. Equivalently, it has no multiple roots (i.e., all its' roots are distinct). A polynomial which is not separable is called *inseparable*.

My intuition behind the name is that the roots cannot be completely "separated" from each other if the polynomial is *inseparable*. In Algebraic Geometry, polynomials that have roots with multiplicity higher than 1 (say multiplicity m) shall correspond to a collection of points that shall all be arbitrarily close to each other while not being equal, ostensibly being "inseparable"[21]. Naturally, the choice of splitting field does not affect whether a polynomial is separable since all splitting fields for a polynomial are isomorphic, and isomorphisms preserves roots (theorem 17.2.13). Even better, we can use *any* field in which the chosen polynomial $f(x)$ splits into linear factors (since there will be a homomorphism from the splitting field F to the field K in which $f(x)$ splits). For our purposes, that means that it is just as good to work over the algebraic closure $\overline{k}$.

[20] (the correspondence Galois group is trivial otherwise)

[21] See the plane with two origins for an example of such a geometric object

Definition 17.6.2: Separable Extension

Let F/k be a field extension and $\alpha \in F$. Then:

1. The element α is called *seprable* if α is algebraic and the minimal polynomial corresponding to α is seperable. Otherwise, the element α is called *inseparable*
2. F/K is called a *separable extension* if every element $\alpha \in F$ is separable, and an *inseparable extension* otherwise.

46: Remark

Note that it is not immediately obvious that if α is separable over k that $k(\alpha)/k$ is a separable extension (what if $1 + \alpha$ is inseparable?). We will answer this question after we develop more tools to work with separable elements.

Using the fact that separability is independent of the field in which it splits, we can use a clever trick in finding whether a polynomial is separable using a tool usually defined in calculus:

Definition 17.6.3: Derivative

Given a polynomial $f(x) \in k[x]$ where

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

then its *derivative* or *formal derivative* is the image of the function $D_x : k[x] \to k[x]$ where:

$$D_x f(x) = na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + 2a_2 x + a_1$$

or more succinctly

$$D_x \left(\sum_i^n a_i x^i \right) = \sum_i^n i a_i x^{i-1} \tag{17.2}$$

Notice that this derivative acts exactly like the derivative seen in a regular calculus course [22]. In a calculus course, we derive this formula using a limit construction. However, if we just think about the derivative (restricted to polynomials) as a function, then we get a purely algebraic definition of the derivative. Note that the i in equation (17.2) is well-defined since we always have the unique homomorphism $\iota : \mathbb{Z} \to k[x]$ since $\mathbb{Z}$ is initial in **Ring**, and so i is the element associated with $\iota(i)$. The *linearity* and *Leibniz property* of the derivative can both be proven with just this definition:

$$D_x(f(x) + g(x)) = D_x f(x) = D_x g(x)$$
$$D_x(f(x)g(x)) = (D_x f(x))g(x) + f(x)(D_x g(x))$$

We will often abbreviate notation and write $f'(x) := D_x f(x)$ when it is not ambiguous.

[22] A good resources to learn this would be Spivak's Analysis

Proposition 17.6.4: Separable Polynomial Test

A polynomial $f(x)$ has multiple roots α if and only if α is also a root of $D_x f(x)$, i.e., $f(x)$ and $D_x f(x)$ are both divisible by the minimal polynomial for α. In particular, $f(x)$ is separable if and only if it is relatively prime to its derivative

$$(f(x), f'(x)) = 1$$

Noet how $f(x)$ is *not necessarily irreducible*. This means that the derivative test is incredibly powerful, telling us that if $(f(x), f'(x)) = 1$, then any irreducible factor of $f(x)$ but all be distinct and have distinct roots.

***Proof* :**
First, let's say $f(x)$ is inseparable so that it has multiple roots in a splitting field F:

$$f(x) = (x - \alpha)^n g(x)$$

where $n \geq 2$ and $g(x) \in F[x]$. Then:

$$f'(x) = n(x - \alpha)^{n-1} g(x) + (x - \alpha)^n g'(x)$$

showing that $\gcd(f(x), f'(x)) \neq 1$ in $F[x]$. By corrollary 11.2.2 to Gauss's Lemma, $\gcd(f(x), f'(x)) \neq 1$ in $k[x]$

Conversely, Assume that $f(x)$ and $f'(x)$ share a root (say $\alpha \in \overline{k}$), so that $\gcd(f(x), f'(x)) \neq 1$ in $k[x]$. Since $f(x)$ has a root, we can write it as:

$$f(x) = (x - \alpha)h(x)$$

for $h(x) \in k[x]$ (or $\overline{k}[x]$?). Then the derivative is:

$$f'(x) = h(x) + (x - \alpha)h'(x)$$

Since $f(x)$ and $f'(x)$ share a root, it must be that $(x - \alpha) \big| h(x)$. But then:

$$(x - \alpha)^2 \big| f(x)$$

showing that $f(x)$ has multiple roots, hence inseparable, as we sought to show.

Using this theorem, we can see that $p(x) = x^p - t$ is inseparable since $p'(x) = px^{p-1} = 0$ (since the characteristic of $\mathbb{F}_p(t)$ is p). Thus, $\gcd(x^p - t, 0) = x^p - t \neq 1$. In fact, this captures a bigger result:

Proposition 17.6.5: Separability and irreducible polynomial

Let k be a field and $f(x) \in k[x]$ be an *inseparable and irreducible* polynomial. Then

$$f'(x) = 0$$

***Proof* :**
Since $f(x)$ is inseparable, by proposition 17.6.4, $f(x)$ and $f'(x)$ share a common irreducible factor $q(x)$, that is $f(x) = q(x)g(x)$ for non-constant $g(x) \in k[x]$. However, f is irreducible, and so $q(x)$ must be an associate of $f(x)$ (and hence be of equal degree). However, that implies that it has degree higher than $f'(x)$ if $f'(x) \neq 0$. Since $q(x) \big| f'(x)$, the only possibility left is that $f'(x) = 0$, as we sought to show.

Corollary 17.6.6: Separability Over Characteristic 0

let k be a field of characteristic 0 and $f(x) \in k[x]$ an irreducible polynomial. Then $f(x)$ is seperable.

***Proof* :**
Since f is irreducible, it must have degree at least 2, and so $f'(x) \neq 0$, showing it is separable by proposition 17.6.5

Thus, every algebraic extension of a characteristic 0 field is a separable extension, and fields in which an irreducible polynomial is inseperable must have characteristic p. Even better, given an inseperable polynomial:

$$f(x) = \sum_{i}^{n} a_i x^i$$

since $f'(x) = 0$:

$$f'(x) = \sum_{i}^{n} i a_i x^{i-1} = 0$$

thus, $ia_i = 0$ for all i, so ia_i is a multiple of p. If i itself is a multiple of p, then certainly $ia_i = 0$, and so it must be that $a_i = 0$ for all indicies that are *not* multiples of p. Therefore, we have reduced inseparable polynomials to polynomials of the form:

$$f(x) = a_0 + a_p x^p + a_{2p} + x^{2p} + \cdots + a_{np} x^{np} = \sum_{k=1}^{n} a_k (x^p)^k$$

Thus, an inseparable polynomial $f(x)$ can be thought of as a polynomial in x^p, $f(x) = g(x^p)$. Using this, we can define the following homomorphism which will be useful to characterize fields which have inseparable polynomials:

Definition 17.6.7: Frobenius Homomorphism

Let k be a field such that $p = \mathrm{ch}(k) > 0$. Then the *Frobenius homomorphism* is a map $k \to k$ defined by

$$x \mapsto x^p$$

It should be verified that this map is indeed a field homomorphism. the Frobenius must be injective, being a homomorphism between fields, however it need not be surjective. Using this homomorphism, we will find all fields in which being separable is identical to being algebraic:

Definition 17.6.8: Perfect Field

Let k be a field. Then k is called a *perfect field* if all irreducible polynomials in $k[x]$ are seperable, that is, every algebraic extension is a separable extension.

Proposition 17.6.9: Frobenius Homomorphism And Separability

Let k be a field. Then k is perfect if and only if either:

1. if $\mathrm{ch}(k) = 0$
2. if $\mathrm{ch}(k) > 0$ and the Frobenius homomorphism is surjective.

Proof **:**
If $\mathrm{ch}(k) = 0$, then by corrollary 17.6.6, any irreducible polynomial is already seperable, so let's say $\mathrm{ch}(k) = p$ for some prime p.

($\Leftarrow$) By our discussion from earlier, we know an inseparable irreducible polynomial must be of the form:

$$f(x) = \sum_{i=1}^{n} a_i (x^p)^i$$

Now, since the Frobenius homomorphism is surjective, for everry a_i, there exists a b_i such that $b_i \mapsto a_i$, that is $(b_i)^p = a_i$. Thus, using the fact that this map is indeed a ring homomorphism, we get:

$$f(x) = \sum_{i=1}^{n} b_i^p (x^p)^i = \sum_{i=1}^{n} (b_i x^i)^p = \left(\sum_{i=1}^{n} b_i x^i \right)^p = g(x)^p$$

where $g(x) = \sum_{i=1}^{n} b_i x^i$. But then $f(x)$ has become reducible! This contradicts our assumption of the irreducibility of $f(x)$, showing us that no such polynomial can exist.

($\Rightarrow$) Take the contra-positive, that $\mathrm{ch}(k) \neq 0$ (say $\mathrm{ch}(k) = p$) and that the Frobenius homomorphism is not surjective. Let's say that φ is the frobenius homomorhpism and that $t \in k \setminus \varphi(k)$. I claim that $x^p - t$ does not split into linear factors. Indeed, emulating what we have done in example 17.6.1, we get $x^p - t$ is irreducible, and that $(x - u)^p = x^p - t$

Here is another one: Suppose that T has characteristic $p > 0$, and that $x \mapsto x^p$ is not an automorphism. Since a ring homomorphism from a field into itself is injective or zero, then $x \mapsto x^p$ is not surjective.

Let $a \in T$ be such that $a \neq b^p$ for every $b \in T$. Let $f(x) = x^p - a$. We prove that f is irreducible and inseparable. If α is a root of $x^p - a$, then $x^p - a = (x - a)^p$, so α is a multiple root of f (with multiplicity p), hence f is inseparable. Now, let $g(x)$ be an irreducible factor of $f(x)$. Note that $\alpha \notin T$, otherwise $a = \alpha^p$, contrary to the assumption. Then $g(x) = (x - \alpha)^k$ for some $1 < k \leq p$. Using the binomial theorem, we have

$$g(x) = (x - \alpha)^k = x^k - k\alpha x^{-1} + \cdots + (-\alpha)^k$$

Therefore, $k\alpha \in T$. Since $\alpha \notin T$, then $k = p$, so $g = f$. Hence, f is irreducible.

Corollary 17.6.10: Finite Fields Are Perfect

let k be a finite field. Then k is perfect

Proof **:**
Since a finite field must have a $\mathrm{ch}(k) = p > 0$, the Frobenius homomorphism is well-defined. Furthermore, Since k is a field, the homomorphism is injective. But an injective map between two finite sets is surjective, hence the Frobenius homomorphism is surjective, and so by proposition 17.6.9 k is perfect.

Thus, we've shown that over most fields we have worked with so far, they are separable. Inseparable extensions do not behave so well (as we'll see, their galois group is [almost] trivial, and hence we get little information about them through galois theory), and so more focus is given in understanding the more common subclass of seperable extensions of a field.

17.6.1 Separability and Embeddings

It should be noted that there *are* separable extensions even if a field is not perfect. For example $x^6 - tx^3 + t \in \mathbb{F}_3(t)[x]$ is irreducible over $\mathbb{F}_3(t)$ and factors into unique linear terms in the splitting field, however the field is not perfect since $x^6 - x^3 + 1$ is irreducible and inseparable. Therefore, we need a better way of finding separable polynomial without relying on nice properties of the ground field. This will be part of the goal of this section.

To characterize separable extensions without relying on k being perfect, we will take advantage that a k-homomorphism is determined by where it maps the roots of minimal polynomials of elements of F. In this way, we get an idea of how many distinct roots a minmial polynomial contains. Recall that we have been able to characterize normal fields through k-homomorhpisms: if $\sigma : F \to \overline{k}$ is a k-homomorhpism and F/k is normal, then $\sigma(F) = F$, which as we talked about in remark 45 shows that normal extensions tells us that we can characterize all k-homomorhpisms as in fact being k-automorphisms (in the sense that all isomorphic copies of F are just symmetries of F). In general, separable extensions need not be normal, and hence they might have multiple embeddings into $\overline{k}$. However, understanding all possible k-homomorphisms still gives us a lot of information about our field as was just pointed out. Studying these embeddings will give us some more information about separable extensions, and will allow us to answer the question of whether an extension $k(\alpha)/k$ is separable if α is separable over k, and to see that any algebraic extension can further be broken down into a separable extension followed by a "purely inseparable" extension: F_{sep}/k and F/F_{sep}.

Definition 17.6.11: Separable Degree

Let F/k be a field extension. Then the number of k-homomorphism $F \to \overline{k}$ called the *separable degree* and is denoted

$$[F : k]_s = \mathrm{Hom}_k(F, \overline{k}) = \mathrm{Emb}_k(F)$$

where $\mathrm{Emb}_k(F)$ is shortand for all possible k-embeddings of F into the algebraic closure of F [a]

[a] Note that for any two algebraic closures $\overline{k}_1, \overline{k}_2$, $\mathrm{Hom}(F, \overline{k}_1) \cong \mathrm{Hom}(F, \overline{k}_2)$, making the definition independent of particular algebraic closure, and hence the notation $\mathrm{Emb}_k(F)$ unambiguous

For the curious reader who would want to see the connection between separability and the separable degree and wants a hint before looking at the solution, they may want to think about what happens to the root, since the following theorem is a generalization of theorem 17.2.13.

47

moment to ponder

Proposition 17.6.12: Separable Degree of Simple Extensions

Let $k \subseteq k(\alpha)$ be a simple algebraic extension. Then $[k(\alpha) : k]_s$ is equal to the number of distincts roots in $\overline{k}$ of the minimal polynomial of α. In particular,

$$[k(\alpha) : k]_s \leq [k(\alpha) : k]$$

with equality if and only if α is separable over k

Notice how this proposition gives us a way to check the separability of the element α without any appeal to k (i.e., without knowing whether k is perfect or not). In fact more is true: $[k(\alpha) : k]_s \mid [k(\alpha) : k]$. The proof of this fact is a bit more laborious than just showing the inequality since it requires the notion of *separable closure*, which is simply the intersection of all fields that are separable, and so we will leave it as an exercise (I think I'll actually incorporate this material).

Proof **:**
This proof will be very similar to that of corrollary 17.2.15. Consider the collection of k-homomorpihism $\{\iota_r : k(\alpha) \to \overline{k}\}$. By corrollary 17.2.15, we can associate each of these homomorphisms with where a root r of the minimal polynomial $m_{\alpha,k}(x)$ maps to, i.e. $\iota_r(\alpha) = r$. We will start by showing that this correspondence is bijective:

$$\{\iota_r : k(\alpha) \to \overline{k}\} \leftrightarrow \{r\}$$

Injectivity comes easily since ι_r fixes k, and so it is completely determined by where α maps to, which necessarily is a root of the minimal polynomial for α. To show surjectivity, consider $\beta \in \overline{k}$ which is some root of the minimal polynomial $m_{\alpha,k}(x)$. Then by lemma 17.5.7, $k(\beta) \subseteq \overline{k}$ exists, and by theorem 17.2.13, there exists a unique isomorphism $k(\alpha) \xrightarrow{\sim} k(\beta)$. Hence, combining with $k(\beta) \subseteq \overline{k}$, we get an extension $k(\alpha) \subseteq \overline{k}$ that maps α to the root β, showing surjectivity.

Thus, we have establisehd a bijection between the roots of the minimal polynomial and the number of embeddings of a simple algebraic extension. Since the degree of α is the degree of the minimal

polynomial, which is less than or equal to the number of roots the minimal polynomial contains, we have that:

$$[k(\alpha) : k]_s \leq [k(\alpha) : k]$$

If furthermore $k(\alpha)$ is separable, then the number of roots is the degree of the polynomial, and hence:

$$[k(\alpha) : k]_s = [k(\alpha) : k]$$

completing the proof.

Corollary 17.6.13: Tower Law For Finite Separable Extensions

Furthermore, if $F/E/k$ is a sequence of algebraic extensions, $[F : k]_s$ is finite if and only if $[F : E]_s$ and $[E : k]_s$ is finite, and furthermore:

$$[F : k]_s = [F : E]_s[E : k]_s$$

***Proof* :**
We'll show that if $[E : k]_s$ or $[F : E]_s$ are infinite, so must $[F : k]_s$, and then show that if they are both finite, so must $[F : k]_s$ be. Let $\varphi : E \to \overline{k}$ be a k-homomorhpism. Then since F/E exists and is algebraic over E (and hence k), by lemma 17.5.7, there exists an exension of φ to $\Phi : F \to \overline{k}$ fixing E (i.e. $\Phi|_E = \varphi$). Furthermore, an extension F into $\overline{E} = \overline{k}$ fixing E extends *a fortiori* the identity on k. Thus, if either $[E : k]_s$ or $[F : E]_s$ is infinite, so must $[F : k]_s$ be.

Conversely if $[F : E]_s$ and $[E : k]_s$ is finite, we'll show $[F : k]_s$ is finite and the multiplicity formula holds. This is almost immediate: for any embedding $F \to \overline{k}$ that extends k, we can can first find all extensions from k to E fixing k, of which there are $[E : k]_s$ possible ways of doing so, and then and then extend any chosen embedding from F to $\overline{E} = \overline{k}$ $(F \to \overline{E})$ fixing E in $[F : E]_s$ different ways, giving us a total of $[F : E]_s[E : k]_s$ total possible extensions, completing the proof.

Proposition 17.6.14: Finite Separable Extension Equivalence

Let F/k be a finite extension. Then $[F : k]_s \leq [F : k]$, and the following are equivalent:

1. $F = k(\alpha_1, ..., \alpha_n)$ where each α_i, is separable over k
2. F/k is a separable extension
3. $[F : k]_s = [F : k]$

***Proof* :**
We'll first generlize proposition 17.6.12 for finite extensions to show that $[F : k]_s \leq [F : k]$. Since F/k is finite, F is finitely generated over k, so let $F = k(\alpha_1, ..., \alpha_n)$. Using proposition 17.6.12,

corollary 17.6.13, and theorem 17.3.5:

$$\begin{aligned}[F:k]_s &= [k(\alpha_1,...,\alpha_{n-1})(\alpha_n):k(\alpha_1,...,\alpha_{n-1})]_s \cdots [k(\alpha_1):k]_s \\ &\leq [k(\alpha_1,...,\alpha_{n-1})(\alpha_n):k(\alpha_1,...,\alpha_{n-1})] \cdots [k(\alpha_1):k] \\ &= [F:k]\end{aligned}$$

showing the inequality

($1 \Rightarrow 3$) Since each α_i is separable over k, it is separable over $k(\alpha_1,...,\alpha_{i-1})$ (the minimal polynomial of α_i over the intermediate field divides the minimal polynomial over the ground field), and so the inequalites in the previous argument are in fact all equalities by corollary 17.6.13

($3 \Rightarrow 2$) Let $[F:k]_s = [F:k]$, take $\alpha \in F$, and consider $k \subseteq k(\alpha) \subseteq F$. Since F is a finite extension, by corollary 17.6.13,

$$[F:k(\alpha)]_s[k(\alpha):k]_s = [F:k]_s = [F:k] = [F:k(\alpha)][k(\alpha):k]$$

Thus, we have two numbers, say a, b such that $a \leq A$, $b \leq B$ and $ab = AB$. This forces $a = A$ and $b = B$, and so we can cancel $[F:k(\alpha)]_s$ and $[F:k(\alpha)]$ on both sides to get:

$$[k(\alpha):k]_s = [k(\alpha):k]$$

which, by proposition 17.6.12 tells us that α is separable. Since $\alpha \in F$ was arbitrary, this tells us that all elements of F are separable, and hence F/k is a separable extension.

($2 \Rightarrow 1$) Since F/k is separable and F is finite, $F = k(\alpha_1,...,\alpha_n)$, and each α_i must be separable by definition.

We can finally answer the quesion posed in remark ref:RemarkSeparableEx: if α is separable over k, then so is $1+\alpha$, α^2, and any other combination of α, since $k(\alpha)/k$ is separable.

Proposition 17.6.15: Tower Law For Separable Extensions

Let $F/E/k$ be a sequence of algebraic extensions. Then F/k is separable if and only if F/E and E/k is separable.

***Proof* :**
This is a generalization of the argument in $3 \Rightarrow 2$ of proposition 17.6.14. Let F/E and E/k be separable extensions (not necessarily finite), and take $\alpha \in F$. We want to show that α is separable over k, that is, $m_{k,\alpha}(x)$ is separable. Since α is separble over E, then $m_{\alpha,E}(x)$ is separable. Let $\alpha_1,...,\alpha_n$ be the coefficients of $m_{\alpha,E}(x)$. Since each of these elements are over E, they are each separable over k.

By proposition 17.6.14, $k(\alpha_1,...,\alpha_n)/k$ is a separable extension over k. We will show that $k(\alpha,\alpha_1,...,\alpha_n)/k$ is separable, completing the proof. To that end, Let $K = k(\alpha_1,...,\alpha_n)$, and

define:

$$L := K(\alpha) \cap E$$

Then since $L \subseteq E$ and E/k is separable, so if L/k. Since the minimal polynomial of α over L (note that α need not be in E) is the same as that of α over E, it is separable by assumption that F/E is separable. But then, $K(\alpha)/L$ is a finite extension... I'm not sure

I found another/this proof here: here

Note that along with the fact that an extension is normal (implying the minimal polynomial splits), we get a bijection between the roots of the minimal polynomial and the automorphisms fixing the base field. For this reason, this will be one of the definitions (out of many equivalent definitions) of a galois extension.

17.6.2 Separability and Simple Extensions

Finally, we show that all finite-degree separable fields (of infinite cardinality) are in fact simple. This means that when we were working with the finite extension F/k, we can always find an α where $F = k(\alpha)$. We will show the case for fields of finite cardinality in the next section.

Proposition 17.6.16: Algebraic And Simple Extensions

Let F/k be an algebraic extension. Then F is simple if and only if the number of distinct intermediate fields $k \subseteq E \subseteq F$ is finite

Proof **:**
Let F/k be simple, so that $F = k(\alpha)$. Since F is algebraic, let $m_{\alpha,k}(x)$ be the minimal polynomial over k. Embed $k(\alpha)$ into an algebraic closure $\overline{k}$. If E is an intermediary field $k \subseteq E \subseteq F$, then $k(\alpha) = F = E(\alpha)$ is also a simple algebraic extension. Denote $m_{E,\alpha}(x)$ to be the minimal polynomial for $E(\alpha)$ over E. Since $m_{k,\alpha}(x) \in E[x]$ for all E and $m_{k,\alpha}(\alpha) = 0$, we know that each $m_{E,\alpha}(x)$ is a factor of $m_{k,\alpha}(x)$.

I claim that E is determined by $m_{E,\alpha}(x)$. Since $m_{k,\alpha}(x)$ has finitely many factors in $\overline{k}$, this will prove there are only finitely many intermediate fields, completing one direction of the proof.

Proceeding with proving the claim, I will show that E is generated (over k) by the coefficients of $m_{E,\alpha}(x)$. To see this, let E' be the set generated by k and the coefficients of $m_{E,\alpha}(x)$. Then $m_{E,\alpha}(x) \in E'[x]$, and since $m_{E,\alpha}(x)$ is irreducible in E, it is irreducible in E', and in fact minimal. Note too that $E'(\alpha) = k(\alpha) = E(\alpha)$, so $[k(\alpha) : E'] = \deg(m_{E,\alpha}(x)) = [k(\alpha) : E]$

Now, we have $E' \subseteq E \subseteq k(\alpha)$, therefore by the tower law:

$$\deg(m_{E,\alpha}(x)) = [k(\alpha) : E'] = [k(\alpha) : E][E : E'] = \deg(m_{E,\alpha}(x))[E : E']$$

Thus, it must be that $[E : E'] = 1$, hence $E = E'$, showing that E is indeed completely determined by $m_{E,\alpha}(x)$, hence there can only be finiely many intermediate fields.

Conversely, let's assume there are finitely many intermediate field $k \subseteq E \subseteq F$. Then it must be that the extension F/k is finite, or else we could construct an infinite sequene of sub-extensions $k \subseteq k(\alpha_1) \subseteq k(\alpha_1, \alpha_2) \subseteq \cdots F$. Thus, F/k is algebraic. We will deal with the case of F being a finite field in corrollary 17.7.12. Thus, for now, assume k is a field of infinite cardinality. We have

to show that every $F = k(\alpha_1, ..., \alpha_n)$ is in fact simple.

Arguing induictively, it suffices to show that $k(\alpha, \beta)$ is simple. For every $c \in k$, we have:

$$k \subseteq k(c\alpha + \beta) \subseteq k(\alpha, \beta)$$

Since there are only finitely many intermediate fields, it must be that for some two $c, c' \in k$, $c \neq c$:

$$k(c\alpha + \beta) = k(c'\alpha + \beta)$$

But then:

$$\alpha = \frac{(c'\alpha + \beta) - (c\alpha + \beta)}{c' - c} \in k(c\alpha + \beta)$$
$$\beta = (c\alpha + \beta) - c\alpha \in k(c\alpha + \beta)$$

Thus, $k(\alpha, \beta) \subseteq k(c\alpha + \beta)$. Since $k(c\alpha + \beta) \subseteq k(\alpha, \beta)$, we have an equality, showing that it is indeed simple, completing the proof.

Theorem 17.6.17: Primitive Element Theorem

Let F/k be a finite separable extension. Then F is also a *simple* extension.

***Proof* :**

Arguing inductively, we may take $F = k(\alpha, \beta)$ with α and β separable over k (and hence algebraic). As before, the finite case will be dealt with in corrollary 17.7.12.

Let I be the number of k-homomorhpisms $\iota : F \to \overline{k}$ (so $|I| = [F : k]_s$). Choose two $\iota, \iota' \in I$ where $\iota \neq \iota$. If x is an indeterminate, then the polynomials

$$\iota(\alpha)x + \iota(\beta) \quad \iota'(\alpha)x + \iota'(\beta)$$

are different (or else $\iota'(\alpha) = \iota(\alpha)$ and $\iota'(\beta) = \iota(\beta)$ so ι, ι' would act in the same way on the whole of $k(\alpha, \beta)$ forcing $\iota = \iota'$)

Therefore, define:

$$f(x) = \prod_{\iota \neq \iota'} ((\iota(\alpha)x + \iota(\beta)) - (\iota'(\alpha)x + \iota'(\beta))) \in \overline{k}[x]$$

The polynomial $f(x)$ cannot be identically 0 since k is infinite. Thus, take $c \in k$ such that $f(c) \neq 0$. So, a distinct map $\iota \in I$ will map $\gamma = c\alpha + \beta$ to distinct elements $\iota(\alpha)c + \iota(\beta)$ (recall $\iota(c) = c$). Since $|I| = [F : k]_s$, and each $\iota(\gamma)$ is a root of the minimal polynomial of γ over k, we have:

$$[F : k]_s \leq [k(\gamma) : k] \leq [F : k]$$

By assumption, the extension is separable, so we have equalities above:

$$[F : k]_s = [k(\gamma) : k] = [F : k]$$

Thus:

$$F = k(\gamma)$$

as we sought to show.

Using this, we can extend corrollary 17.2.15 from simple to finite extensions:

Corollary 17.6.18: Order of $\mathrm{Aut}_k(F)$ Of Finite Extension

Let F/k be a finite seprable extension. Then:

$$|\operatorname{Aut}_k(F)| \leq [F : k]$$

with equlaity if and only if F/k is a finite separable normal extension.

***Proof* :**
By theorem 17.6.17, if K is finite and separable, it is simple, and so the inequality immediately holds by corrollary 17.2.15. Equality holds if the minimal polynomial splits, which it does over it's splitting field, which is a normal extension by theorem 17.5.12. Conversely, if K/F is normal, then $f(x)$ splits completly in F and has distinct roots since α is separable over k, establishing equalitly once again, completing the proof.

Note that the separability condition is necessary:

Example 17.6.2: Nonseparable Nonsimple Extension
Let $F = \mathbb{F}_p(u, v)$ be the field with two indeterminate variables over $\mathbb{F}_p$, and subfield $k = \mathbb{F}_p(s, t)$ such that $s = u^p$ and $t = v^p$. Then F/k is an algebraic extension; the minimal polynomial of u over k is $x^p - s$, and the minimal polynomial of v over $k(u)$ is $y^p - t$. Thus:

$$[F : k] = p^2$$

Arguing like before, As c ranges in k, we get the intermediate fields:

$$k \subseteq k(cu + v) \subseteq F$$

If any two chocies c, c' get give the same intermediate field, then we can deduce $k(u, v) = k(cu + v)$ as we've argued earlier. This would then give us:

$$[k(cu + v) : k] = p^2$$

However, the minimal polynomial of $cu + v$ has degree p:

$$(cu + v)^p = c^p + st \in k(s, t) = k$$

Thus, since $k = \mathbb{F}_p(s, t)$ is infinite, there are infinitely many intermediate fields, and it follows that F is *not* a simple extension of k by proposition 17.6.16

As a final result, we state the following that shall be handy for chapter 21, section 21.3:

Corollary 17.6.19: Lüroh's Theorem

Let $k(t)/k$ be a simple extension and $E \subseteq k(t)$ an intermediary field containing k. Then there exists a polynomial g such that

$$E = k(g(t))$$

Proof **:**
Exercise. Sketch of proof: use the primitive element theorem, use Gauss's lemma, and find the minimal polynomial of the primitive element.

Exercise 17.6.1

1. Show that If E/k and F/k be field extensions, E/k is separable, then so is EF/F. Show that if E/k and F/k are separable, so is EF/F.
2. (exercise on automatic differentiation? $k[x, \epsilon]/(\epsilon^2)$)
3. Let F/k is simple so that $F = k(\alpha)$. then if L is the splitting field for a [not necessarily separable] minimal polynomial $m_{\alpha,k}(x)$, $L = \prod_{\sigma \in \mathrm{Emb}_k(F)} \sigma(F)$. If F is normal, then $L = F$, and if F separable, we have an idea of the number of σ exists, namely the degree of $k(\alpha)$ over k)

17.7 Cyclotomic Polynomials and Finite Fields

In this section, we use the tools given to us to gain some important examples of field extensions and analyse the correpsonding $\mathrm{Aut}_k(F)$, and generalize theorem 17.6.17 to show finite seperable extensions of finite fields are in fact simple.

We start by using the tools that were developed to study the important class of polynomials of the form $x^n - 1$; such extensions will be called *cyclotomic extensions*. These polynomials will have particularily strong relations are are a great case-study to pratice everything that was developped. They are also important for the continuation of field theory to Kummer Theory, which is the study the case where $\mathrm{Aut}_k(F)$ is an abelian group, and they form an important first step towards Fermat's Last theorem through their application to Elliptic curves[23]

After having classified cyclotomic extensions, we shall finish the section with the study of extensions of finite fields. As it turns out, all finite extensions will be splitting fields of polynomials of the form $x(x^{p^d} - 1)$, meaning they will have a very simple structure and automorphism group.

17.7.1 Cyclotomic Polynomials and Fields

We start off with some reminders:

[23] if $\mathrm{ch}(F) \notin 2, 3$, then an elliptic curve is a polynomial of the form $y^2 = x^3 + ax + b$, $a, b \in F$. A more technical definition is given when $\mathrm{ch}(F) \in \{2, 3\}$ that would be presented in a course on algebraic number theory or algebraic geometry.

Definition 17.7.1: Roots Of Unity Reminder

We will denote $\zeta_n = e^{\frac{2\pi i}{n}}$, and the cyclic subgroup they generated to be μ_n. Note that:

$$\zeta_n^n = e^{\frac{2\pi i k}{k}} = e^0 = 1$$

If $\langle \zeta_n^k \rangle = \mu_n$, then this element is called the *primitve root of unity*.

The element ζ_n is always a root of $x^n - 1$, and each ζ_n^k, $1 \leq k \leq n$ where $\gcd(n, k) = 1$ consist of the rest of the roots of $x^n - 1$, for

$$(\zeta_n^k)^n = 1$$

and if $\gcd(n, k) = \ell$, then:

$$(\zeta_n^k)^\ell = 1$$

It is an easy mistake to think that $\zeta_2 = i$, however $\zeta_2 = -1$ and $\zeta_4 = i$. These elements always form a group under multiplication, which we labeled μ_n. This group always lies on the unit circle, for example:

If you know about the euler representation of complex numbers:

$$re^{i\theta} = r(\cos(\theta) + i\sin(\theta))$$

Then $e^{\frac{2\pi i}{n}} = \zeta_n = \cos(2\pi/n) + i\sin(2\pi/n)$. Other identities we get are:

1. $\zeta_3 = \frac{-1+i\sqrt{3}}{2}$
2. $\zeta_5 = \frac{\sqrt{5}-1}{4} + \frac{\sqrt{10+2\sqrt{5}}}{4} i$
3. $\zeta_6 = \frac{1+i\sqrt{3}}{2}$ (to not be confused with ζ_3)

4. $\zeta_8 = \frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}$

These can be deduced using the results from section 17.4. In paticular Gauss proved that primitive nth roots of unity ζ_n can be constructed using the square-roots, additions, subtraction, multiplication, and division of rational numbers, which we saw implies that n is a power of two, or in our case we would require that n is the product of two discint *fermat primes*.

Every root of unity of a root of $x^n - 1$ for some n, however this polynomial is not irreducible, namely $x^n - 1 = (x-1)p(x)$. The polynomial $p(x)$ will turn out to be irreducible, and hence is given a name:

Definition 17.7.2: Cyclotomic Polynomials

Let:

$$\Phi_n(x) = \prod_{\zeta \text{ primitive } n\text{th root of } 1} (x - \zeta) = \prod_{\substack{1 \leq m \leq n \\ \gcd(m,n)=1}} (x - \zeta_n^m)$$

Then the resulting polynomial is called the *nth cyclotomic polynomial*

As we have seen in the section on number theory (section 0.2), there are $\varphi(n)$ primitive roots of 1, where φ is Euler's totient function. The term cyclotomic comes from the greek *cyclo* (circle) and *tomos* (divide), owing to how the roots of unity divide the circle into regular sections. Though not obvious, this polynomial has rational coefficients and is irreducible over $\mathbb{Q}$ (which we'll show).

Example 17.7.1: Cyclotomic Polynomials

In the case of $n = p$ is a prime, it is simple to see it's irreducible. Every non-identity element of μ_p is a generator, i.e. every pth root of 1 is primitive except 1. In fact, 1 is a root of $x^p - 1$ ($1^p - 1 = 0$), and so we get:

$$\Phi_p(x) = \frac{x^p - 1}{x - 1} = x^{p-1} + x^{p-2} + \cdots + x + 1$$

which is irreducible over $\mathbb{Q}$ using Eistensteins trick along with shifting (noting that p divides $\binom{n}{k}$). Furthermore, $\mu_p = C_p = \mathbb{Z}/p\mathbb{Z}$.

Aside: for those who know Taylor series and the exponential function, notice how close $\Phi_p(x)$ resembles the pariatal sum for the taylor expansion of e^x:

$$1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots + \frac{x^p}{p!}$$

the main difference being the coefficient of the partial taylor expansion of e^x being $1/n!$. This reflects on the "rotation" nature of e^x, as well as it's link to roots of unity. However, note that e^x has no roots (even when expended to $\mathbb{C}$, since $e^z \cdot e^{-z} = 1$). Hence, though the partial expansion may have root, even roots that resemble the roots of unity, taylor expansion, the "infinite polynomial representation" need not have roots!

For ease of reference, we record what happens when n is not prime:

Lemma 17.7.3: Cyclotomic Polynomial divisor Representation

For all $n > 0$,

$$x^n - 1 = \prod_{1 \leq d \mid n} \Phi_d(x)$$

Proof **:**
For all ζ_n^k, $0 \leq k \leq n-1$, $(\zeta_n^k) - 1 = 0$. Hence, $x^n - 1$ splits into products

$$x^n - 1 = \prod_k^n (x - \zeta_k)$$

If $n = dk$ then $1 = \zeta_n^n = (\zeta_n^d)^k$, hence every dth root of unity is an nth root of unity. In particular, every primitive dth root of unity is an nth root of unity. Conversely, if $\zeta_n^d \in \mu_n$, then ζ_n^d forms cyclic subgroup, say μ_d where d is a divisor of n. But then ζ_n^d is a primitive dth root of unity for some divisor d. This establishes:

$$x^n - 1 = \prod_{1 \leq d \mid n} \left(\prod_{\substack{\zeta \text{ primitive} \\ d\text{th root of } 1}} (x - \zeta) \right) = \prod_{1 \leq d \mid n} \Phi_d(x)$$

as we sought to show.

Corollary 17.7.4: Integer Coefficients

The cyclotomic polynomials $\Phi_n(x)$ have integer coefficients

Proof **:**
This is a proof by [strong] induction. The case of $\Phi_1(x) = x - 1$ is immediate, so assume that for $m < n$, Φ_m is a monic polynomial with integer coefficients. Then:

$$f(x) = \prod_{\substack{1 \leq d \mid n \\ d < n}} \Phi_d(x)$$

is a monic polynomial with degree $< n$. Since $f(x), x^n - 1 \in \mathbb{Z}[x]$ are two monic polynomials with integer ceofficients, we may use the euclidean algorithm to find $q(x), r(x) \in \mathbb{Z}[x]$, $\deg(r(x)) < \deg(f(x))$ or $r(x) = 0$ where

$$x^n - 1 = f(x)q(x) + r(x)$$

Then by lemma 17.7.3 we have

$$x^n - 1 = f(x)\Phi_n(x)$$

Hence, we get

$$f(x)(q(x) - \Phi_n(x)) = r(x)$$

But that implies $\deg(r(x)) > \deg(f(x))$, a contradiction. Thus $r(x) = 0$, giving us $f(x)q(x) = f(x)\Phi_n(x)$, which implies $\Phi_n(x) = q(x) \in \mathbb{Z}[x]$, as we sought to show.

Example 17.7.2: Cyclotomic Polynomial

1. Let $p(x) = x^4 - 1 = \Phi_1(x)\Phi_2(x)\Phi_4(x)$. Then:

$$\Phi_4(x) = \frac{x^4 - 1}{x^2 - 1} = x^2 + 1$$

2. Let $p(x) = x^6 - 1 = \Phi_1(x)\Phi_2(x)\Phi_3(x)\Phi_6(x)$. Then:

$$\Phi_6(x) = \frac{x^6 - 1}{(x^2 - 1)(x^2 + x + 1)} = x^2 - x + 1$$

3. Let $p(x) = x^{12} - 1 = \Phi_1(x)\Phi_2(x)\Phi_3(x)\Phi_4(x)\Phi_6(x)\Phi_{12}(x)$. Then:

$$\Phi_{12}(x) = \frac{x^{12} - 1}{(x^6 - 1)(x^2 + 1)} = x^4 - x^2 + 1$$

It might seem that all the coefficients are 0 or ± 1. However, the first counter-example to this problem is when $n = 105$, and in general the coefficients can be as large as we might want

Proposition 17.7.5: Irreducibility Of Cyclotomic Polynomials

For all $n \in \mathbb{N}_{>0}$, $\Phi_n(x) \in \mathbb{Z}[x]$ is irreducible over $\mathbb{Q}$

Note that since $\Phi_n(x)$ is monic, by Gauss's lemma, irreducibility in $\mathbb{Q}[x]$ is equivalen to ireducibility in $\mathbb{Z}[x]$.

Proof **:**
p. 448 Aluffi

Definition 17.7.6: Cyclotomic Field

The splitting field $\mathbb{Q}(\zeta_n)$ for the polynomial $x^n - 1$ over $\mathbb{Q}$ is the nth *cyclotomic field*

Proposition 17.7.7: Cyclotomic Field Automorphism

$$\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta_n)) \cong (\mathbb{Z}/n\mathbb{Z})^\times$$

Proof **:**
First, the cardinality of both are the same since cyclotomic fields are a splitting field with degree $\varphi(n)$ (Euler's Totient) while $(\mathbb{Z}/n\mathbb{Z})^\times$ has order $\varphi(n)$ as well. Define

$$j : (\mathbb{Z}/n\mathbb{Z})^\times \to \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta_n)) \qquad [m]_n \mapsto (\zeta_n \mapsto \zeta_n^m)$$

This map is certainly independent of representative and sends roots of polynomials of $m_{\zeta_n,\mathbb{Q}}(x)$ to themselves making it an automorphism.

17.7.2 Finite Fields

Let F be a finite field of characteristic p. Then as we remarked under definition 17.1.7,

$$F/\mathbb{F}_p$$

where $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. Since F is finite, let $d = [F : \mathbb{F}_p]$. Since F is a d-dimensional vector-space over $\mathbb{F}_p$, $F \cong_{\mathbf{Vect}} \mathbb{F}_p^d$, hence the order of $|F|$ is a power of p: $|F| = p^d$.

With everything we've learned, we can now classify all finite fields. Fortunately for us, we get essentially the best possible result we might ask for when classifying objects: for every order p^d, we get a unique field up to isomorphism:

Theorem 17.7.8: Calssifying Finite Fields

Let $q = p^d$ be a power of the prime number p. Then the polynomial $x^q - x$ is separable over $\mathbb{F}_p$, and the splitting field of the polynomial $x^q - x$ over $\mathbb{F}_p$ is a field with exactly q elements. Conversely, let F be a field with exactly q elements. Then F is a splitting field for $x^q - x$ over $\mathbb{F}_p$

Hence, all finite extension of finite fields is abelian, and even better cyclic.

Proof **:**
Let F be the splitting field of $x^q - x$ over $\mathbb{F}_p$. Then necessarily F is finite (by our realier remark), and hence it's perfect, and hence it's a separable extension, meaning F contains q distinct roots of $x^q - x$. We need to show that every element of F is a root. Let R be the set of roots of $x^q - x$ from F. Then I claim that R is a field. Since F is the smallest field containing all the roots, it follows that $R = F$.

We simply need to show it's a subring, and since all nonzero elements are units, it is a field. For any $a, b \in R$, since $a^q - a = 0$, $b^q - b = 0$, then $a^q = a$ and $b^q = b$. Then using the binomial theorem and the characteristic of F, we get:

$$(a - b)^q = a^q + (-1)^q b^q = a - b$$

where $(-1)^q = -1$ if p is odd, and $(-1)^q = 1 = -1$ if $p = 2$. Furthermore, if $b \neq 0$, then:

$$(ab^{-1})q = a^q(b^q)^{-1} = ab^{-1}$$

showing that R is closed under both operations, and hence a subfield, and hence $R = F$

On the other hand, let F be a field with q elements. Then $|F^\times| = q - 1$, and so by lagrange theorem, for any $0 \neq a \in F^\times$, $|a| \big| q - 1$ Thus:

$$a \neq 0 \Rightarrow a^{q-1} = 1 \Rightarrow a^q - a = 0$$

showing that a *must* be a root of $x^q - x$. Furthermore, $0^q - 0 = 0$, showing that $x^q - x$ has q roots in F (i.e. all elements of F). But then, F is the splitting field of $x^q - x$ by the previous logic, as we sought to show.

Corollary 17.7.9: Unique Finite Field For Every Order

For every $q = p^d$, p a prime and $d \in \mathbb{N}_{>0}$, there exists a unique field of order q

Proof **:**
this follows from theroem 17.7.8 and the uniquness of the splitting field

Definition 17.7.10: Galois Field

Let F be a field of order $q = p^d$. Then F is called the *Galois Field* of order q, and we usually denote it as $\mathbb{F}_q$ or $\mathbb{F}_{p^d}$.

Example 17.7.3: Galois Fields
Good example by aluffi involving the reducibility of $x^4 + 1$ over *any* field $\mathbb{F}_p$

Corollary 17.7.11: Finite Field Extensions

Let p be a prime and let $d \leq e$ be positive integers. Then there is an extension $\mathbb{F}_{p^d} \subseteq \mathbb{F}_{p^e}$ if and only if $d \big| e$. If $d \big| e$, then there is exactly one such extension, in the sense that $\mathbb{F}_{p^e}$ conains a unique copy of $\mathbb{F}_{p^d}$.

Furthermore, all extensions $\mathbb{F}_{p^d} \subseteq \mathbb{F}_{p^e}$ are simple

Proof **:**
First, let's show that extensions are only dependent on the divisibility of the order. Let's say there is an extension $\mathbb{F}_P \subseteq \mathbb{F}_{p^d} \subseteq \mathbb{F}_{p^e}$. Then $[F_{p^d} : \mathbb{F}]$ divides $[\mathbb{F}_{p^e} : \mathbb{F}]$, meaning $d \big| e$.

Conversely, let's say $d \leq e$ such that $d \big| e$. Since:

$$p^e - 1 = (p^d - 1)((p^d)^{\frac{e}{d}-1} + (p^d)^{\frac{e}{d}-\frac{e}{d}})$$

we see that $p^d - 1$ divides $p^e - 1$, and so $x^{p^d-1} - 1$ divides $x^{p^e-1} - 1$ (Aluffi gave this as exercsie V.2.13). Therefore:

$$(x^{p^d} - x) \big| (x^{p^e} - x)$$

Since $\mathbb{F}_{p^e}$ is a splitting field for the second polynomial by theorem 17.7.8, it must contain a unique copy of the splitting field for the first plynomial (exercsie 4.5 in Aluffi, p. 442).

Finally, to show the extension is simple, recall that any finite field is necessarily cyclic, as we recalled in theorem 17.1.8 ($F^\times$ is locally cyclic, and since F is finite, $F^\times$ is cyclic). Let's say $\alpha \in F_{p^e}$ generated the multiplicative group of $\mathbb{F}_{p^e}$. Then clearly, α will generated $\mathbb{F}_{p^e}$ over any subfield, so if $d \big| e$, $\mathbb{F}_{p^e} = \mathbb{F}_{p^d}(\alpha)$, showing that $\mathbb{F}_{p^e}/\mathbb{F}_{p^e}$ is simple, completing the proof.

Corollary 17.7.12: Separable Finite Extensions Of Finite Fields

Let k be a finite field and F/k a finite separable extension. Then F is a simple extension.

Proof **:**
Most of the proof of this was already done by proposition 17.6.16. If k is a finite field, then so is F, and the extensions are automatically simple by corollary 17.7.11.

Using these results, we can be more gain a lot of information about polynomials over finite fields:

Corollary 17.7.13: Finite Fields And Irreducible Polynomials

Let F be a finite field. Then for all integers $n \geq 1$, there exists irreducible polynomials of degree n in $F[x]$

Proof **:**
Let $n \geq 1$ and $F = \mathbb{F}_{p^d}$. Then

$$\mathbb{F}_{p^d} \subseteq \mathbb{F}_{p^{dn}}$$

is a finite simple extension, say $\mathbb{F}_{p^{dn}} = \mathbb{F}_{p^d}(\alpha)$. Then sine

$$[\mathbb{F}_{p^{dn}} : \mathbb{F}_{p^d}] = n$$

and α has a minimal polynomial over $\mathbb{F}_{p^d} = F$, we see that we can find an irreducible polynomial of degree n in $F[x]$.

Note how this proof doesn't necessarily generalize for infinite fields, since we don't have so much control over the degree of the extension (worth pondering over).

Corollary 17.7.14: Factorizing $x^{q^n} - x$

here p. 443 aluffi

Proof **:**

Example 17.7.4: Factorizing Polynomials over Finite Fields
p. 443 Aluffi

Another result that finiteness brings is that the structure of $\mathrm{Aut}_{\mathbb{F}_p}(\mathbb{F}_{p^d})$ is very simple. We already know that any finite field is perfect, hence all algebraic extensions are separable, and so

$$|\mathrm{Aut}_{\mathbb{F}_p}(\mathbb{F}_{p^d})| = d$$

However, we can say even more about the structure of the group

Proposition 17.7.15: Structure of Aut Of Finite Fields

Let $\mathbb{F}_{p^d}/\mathbb{F}_p$ be a field extension of degree d. Then $\operatorname{Aut}_{\mathbb{F}_p}(\mathbb{F}_{p^d})$ is cyclic, and generated by the Frobenius isomorphism.

Proof **:**
Let $\varphi : \mathbb{F}_{p^d} \to \mathbb{F}_{p^d}$ be the map such that $x \mapsto x^p$. Since finite fields are perfect, φ is surjective, and hence an isomorphism. For any $x \in \mathbb{F}_p$, by Fermat's theorem (theorem 10.3.3)

$$x^p \bmod x \bmod p$$

and so φ is in fact a $\mathbb{F}_p$-isomorphism, meaning $\varphi \in \operatorname{Aut}_{\mathbb{F}_p}(\mathbb{F}_{p^d})$. Since $|\operatorname{Aut}_{\mathbb{F}_p}(\mathbb{F}_{p^d})| = d$, we need to show that $|\varphi| = d$.

Since φ is in a finite group, there exists an e such that $\varphi^e = \mathrm{id}$, so $x^{p^e} = x$ for all $x \in \mathbb{F}_{p^d}$. Then $x^{p^e} - x$ has p^d roots in $\mathbb{F}_{p^d}$ (since φ makes all elements of $\mathbb{F}_{p^d}$ satisfy this relation). Then since the number of roots of a polynomial cannot exeed the highes degree, we have that $p^d \leq p^e$, which implies $d \leq e$. Thus, in terms of order of groups, we get that:

$$|\operatorname{Aut}_{\mathbb{F}_p}(\mathbb{F}_{p^d})| \leq |\varphi|$$

but by Lagrange's theorem, $|\varphi| \big| |\operatorname{Aut}_{\mathbb{F}_p}(\mathbb{F}_{p^d})|$,

(A word on a way to construct an algebraic closure of finite fields, and on how the category of finite fields seems to be close to another known category)

To close off this section, we saw that all finite extensions of finite fields is a cyclic extension. This has not been shown for finite extension for $\mathbb{Q}$. However, it turns out that this is *almost* the case, namely all finite extensions of $\mathbb{Q}$ embed into a cyclic extensions. This requires a lot more build-up, and is easier to prove in a course on Class Field Theory[24]. These notes shall indeed prove this fact not using class field theory as the "reward" for the material covered in chapter 22 (see theorem 22.5.3).

Exercise 17.7.1

1. Let $f(x)$ be an irreducible polynomial of degree n over $\mathbb{F}_p$. Show that any field extension of $f(x)$ splits and is isomorphic to $x^{p^n} - x$
2. Let $k = \mathbb{F}_p$. Show that $f(x^p) = f(x)^p$.

[24]See CITE:HERE for my book on it

18

Galois Theory

For eons, mathematicians where looking for relations between coefficients and roots of a polynomial. Some of these were found with success, for example if $x^2 + bx + c$ has roots α and β, then $\alpha + \beta = -b$ and $\alpha\beta = c$[1]. More generally, a lot of focus was put into find the roots of such polynomials, which lead to the discovery of the quadratic[2], and eventually the cubic and quartic. However, the effort to understand these brought limited results to pressing questions like the solvability of the quintic, that is whether its roots can written in terms of it's coefficients and the arithmetic operations of $+, -, \times, \div$ and $\sqrt{-}$. It wasn't until the turn of the 18th century when serious progress has been made on this front. The key insight was to consider breaking down a polynomials as though it already had roots into the product of monomials and analyzing the relationship. The first immediate realization that could be made linking roots to coefficients is that given we can completely split a polynomial:

$$(x - a_1)(x - a_2) \cdots (x - a_n)$$

each coefficient of the original polynomial can be represented as a *symmetric function* (see 18.2.3). These form a natural relationship between the roots and the coefficients. However, as any roots of polynomial induce such equations, these don't give enough information to "distinguish" relationships between roots for a given polynomial. It shall be shown that most polynomials have no further relations besides the symmetric functions. However, there are many functions that shall have more relationships (think of $x^2 - a$ where squaring both roots gives the same coefficient a). This chapter shall make rigorous this realization. This realization almost managed to show the general quintic is unsolvable:

> "The quintic was almost proven to have no general solutions by radicals by Paolo Ruffini in 1799, whose key insight was to use permutation groups, not just a single permutation. His

[1] See section 18.2.3 for much details about these relations

[2] This is a little anachronistic, for when the quadratic formula was first discovered in Babylonia and more succinctly by Brahmagupta, the "modern" notion of polynomial and variables where not used, but rather it was shown by relating areas of circles and rectangles by the Babylonians

> solution contained a gap, which Cauchy considered minor, though this was not patched until the work of the Norwegian mathematician Niels Henrik Abel, who published a proof in 1824, thus establishing the Abel–Ruffini theorem."[3]

This theory established that there were unsolvable quintic's, but not how to find whether a given quintic is solvable. This is where the work of Évariste Galois comes in. He noticed that the key insight was to work with all possible relationships of the roots of the polynomial, assuming all the roots exist.

Example of Galois Theory before Formal Coverage

Take the polynomial:

$$x^4 - 5x^2 + 6 = (x^2 - 2)(x^2 - 3)$$

which reduces to 2 irreducible polynomials over $\mathbb{Q}$. Adjoining $\sqrt{2}$ and $\sqrt{3}$ to $\mathbb{Q}$, we can factor further and get:

$$= (x - \sqrt{2})(x + \sqrt{2})(x - \sqrt{3})(x + \sqrt{3}) \tag{18.1}$$

In this case, the roots are: $\sqrt{2}$, $-\sqrt{2}$, $\sqrt{3}$, and $-\sqrt{3}$. In equation 18.1, we may permute these for roots however we want and shall get that the answer does not change, namely we have that S_n acts on these roots.

However, these roots share relationships among themselves. For example:

$$\sqrt{2} + (-\sqrt{2}) = 0 \qquad \sqrt{3} + (-\sqrt{3}) = 0$$

There is also the equation $\sqrt{2}(\sqrt{-2}) - 2 = 0$ (equiv. $\sqrt{3}(\sqrt{-3}) - 3 = 0$) which says that we can multiply the roots to a coefficient of one of the quadratic irreducibles. This limit our permutations down significantly, leading us to consider only the trivial permutation and:

$$\begin{cases} \sqrt{2} \mapsto -\sqrt{2} \\ -\sqrt{2} \mapsto \sqrt{2} \\ \sqrt{3} \mapsto -\sqrt{3} \\ -\sqrt{3} \mapsto \sqrt{3} \end{cases}$$

Notice that if the following permutation is taken:

$$\begin{cases} \sqrt{2} \mapsto \sqrt{3} \\ -\sqrt{2} \mapsto -\sqrt{3} \\ \sqrt{3} \mapsto \sqrt{2} \\ -\sqrt{3} \mapsto -\sqrt{2} \end{cases}$$

Then at least one of the equation is no longer equals 0 ("has a solution"):

$$\sqrt{2} + (-\sqrt{3}) \neq 0$$

Thus, we only choose the permutations that keep the relationships between the roots.

[3] taken from Wikipedia for Galois Theory

Another example is the following: take $x^4 - 2$. This is irreducible by Eisenstein's. Adding the root $\alpha = \zeta_4 \sqrt[4]{2}$, we see that we can use it to form the other roots of this equation, namely

$$\zeta_4^2 \sqrt[4]{2} \qquad \zeta_4^3 \sqrt[4]{2} \qquad \sqrt[4]{2}$$

Thus there are polynomial relationships between the roots α and the other roots of this equation. As it turns out, this shall limit our permutations greatly, where from the original S_4 possible permutations, it must be that there is only the permutations of the form $\zeta_4 \sqrt[4]{2} \mapsto \zeta_4^k \sqrt[4]{2}$ for $1 \leq k \leq 4$, namely it is a cyclic group of order 4.

Re-framing this in the language of field extensions, if f is a polynomial over $\mathbb{Q}$, α is a root, then the algebraic extension $\mathbb{Q}(\alpha)/\mathbb{Q}$ contains all possible polynomials that can be generated with elements from $\mathbb{Q}$ and α. If any of these polynomials happen to be another root of f, that is $p(\alpha) = \beta$ where $f(\beta) = 0$, then $\mathbb{Q}(\alpha)/\mathbb{Q}$ contains a relationship between the roots. In fact, $\mathbb{Q}(\alpha)/\mathbb{Q}$ contains all possible [polynomial] relationships between α and the rest of the roots. Let's say one of these relationships is represented as :

$$P_{\mathbb{Q}}(\alpha, \beta) = 0$$

Then a $\mathbb{Q}$-automorphism (in this case nothing more than a field automorphism) would have to preserve this relationship. Hence, the set of $\mathbb{Q}$-automorphism captures the relationships between α and all possible roots. If $K/\mathbb{Q}$ is the splitting field of K, then the $\mathbb{Q}$-automorphisms of K would be restricted to automorphism that must satisfy all possible relationships between roots. If there were no relationships (you would need to add n roots to split f), then the automorphism group would be isomorphic to S_n. If there is any other type of group, that means there are some relationships between the coefficients.

The group described above is called the *galois group* and it is the aim of this chapter to be able to find some galois groups, to find how frequent certain galois groups are, to show the correspondence between field extension an galois groups, and answer some long-pending mathematical questions such as the solvability of [general] quintic using the newly developed theory.

18.1 Galois Extension and Galois Group

First, let's the recall the definition of an automorphism and define a new term:

Definition 18.1.1: Automorphism and fixed Element

1. An isomorphism σ of F with itself is called an *automorphism* of F. The collection of automorphisms of F is denoted $\mathrm{Aut}(F)$. If $\alpha \in F$, we write $\sigma\alpha$ for $\sigma(\alpha)$

2. An auotmorphism $\sigma \in \mathrm{Aut}(F)$ is said to *fix* an element $\alpha \in F$ if $\sigma\alpha = \alpha$. If k is a subset of F, then an automorphism σ is said to *fix* F if it fixes all the elements of F

$$\sigma a = a \qquad \forall a \in k$$

3. Let F/k be an extension of fields. Then $\mathrm{Aut}(F)$ is the collection of automorphissm of F, and $\mathrm{Aut}_k(F)$ is collection of automorphsism of F which fix F.

Example 18.1.1: Automorphism and Fixed F
The identity map id is always an automorphism.

Furthermore, the prime field of any field K is always fixed. If P is the prime-field of F, then for any $\sigma \in \mathrm{Aut}(K)$, $\sigma(1) = 1$, since $\sigma(0) = 0$. Thus, $\sigma(a) = a$ for every $a \in P$. This means that $\mathbb{Q}$ and $\mathbb{F}_p$ have only trivial automorphisms

$$\mathrm{Aut}(\mathbb{Q}) = \{\mathrm{id}\}, \mathrm{Aut}(\mathbb{F}_p) = \{\mathrm{id}\}$$

and if P is a prime field of F, then

$$\mathrm{Aut}_P(F) = \mathrm{Aut}(F)$$

Next, we recall a very useful criterion for finding automorphisms of algebraic extensions K/F

Proposition 18.1.2: Fixed Roots Criterion

Let F/k be a field extension and let $\alpha \in F$ be algebraic over k. Then for any $\sigma \in \mathrm{Aut}(F/k)$, $\sigma\alpha$ is a root of the minimal polynomial for α over k, i.e., $\mathrm{Aut}(F/k)$ permutes the roots of irreducible polynomials. Equivalently, any polynomial whith coefficeints in k having α as a root also has $\sigma\alpha$ as a root.
Furthermore, $\mathrm{Aut}_k(k(\alpha))$ acts *transitively* on the set of roots of $m_{k,\alpha}(x)$ (i.e. if β_1, β_2 are roots of $m_{k,\alpha}(x)$, then there exists a $\sigma \in \mathrm{Aut}_k(k(\alpha))$ such that $\sigma\beta_1 = \beta_2$)

***Proof* :**
Let

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$$

be the minimal polynomial with coefficients from F so that

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0$$

then, applying any automorphism σ to this polynomial will get us

$$\begin{aligned} \sigma(\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0) &= \sigma(0) \\ \sigma(\alpha)^n + a_{n-1}\sigma(\alpha)^{n-1} + \cdots + a_1\sigma(\alpha) + \sigma(a_0) &= 0 \qquad \sigma \text{ fixes } a_i \text{ for each } i \end{aligned}$$

since $a_i \in F$ for each i, showing that $\sigma(\alpha)$ is also a root of the minimal polynomial

Transitivty is left as an exercise (recall the appropriate theorem)

Example 18.1.2: Automorphism Groups

1. Take $K = \mathbb{Q}(\sqrt{2})$. Then $\mathrm{Aut}(\mathbb{Q}(\sqrt{2})) = \mathrm{Aut}(\mathbb{Q}(\sqrt{2})/\mathbb{Q})$. If $\sigma \in \mathrm{Aut}(\mathbb{Q}(\sqrt{2}))$, then by the fixed root criterion, it must be that

$$\sigma(\sqrt{2}) = \pm\sqrt{2}$$

and since σ must fix everything else in place, this means that we get

$$\sigma(a + b\sqrt{2}) = a \pm b\sqrt{2}$$

in other words, we've determined were every other element must map to! Thus, we have the maps

$$\sqrt{2} \mapsto \sqrt{2} \qquad \sqrt{2} \mapsto \sqrt{-2}$$

If we let the first map be represetned by 1 and the second by σ, then we get

$$\mathrm{Aut}(\mathbb{Q}(\sqrt{2})) = \{1, \sigma\} = C_2$$

that is, $\mathrm{Aut}(\mathbb{Q}(\sqrt{2}))$ is a cyclic group of order 2.

2. Let $K = \mathbb{Q}(\sqrt[3]{2})$. By theorem 17.2.2, every element of $\mathbb{Q}$ can be written as

$$a + b\sqrt[3]{2} + c(\sqrt[3]{2})^2$$

If we consider any $\sigma \in \mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2}))$, then we get

$$a + b\sigma(\sqrt[3]{2}) + c\sigma(\sqrt[3]{2}) \qquad a, b, c \in \mathbb{Q}$$

and since $\mathbb{Q}(\sqrt[3]{2})$ has non of it's other roots (recall example 17.3.4). Then that means that the only automorphism is the identity, meaning

$$\mathrm{Aut}(\mathbb{Q}(\sqrt[3]{2})) = \{1\}$$

3. Now let $K = \mathbb{Q}(\sqrt[3]{2}, \sqrt{-3})$, be the splitting field of $x^3 - 2$ in example (17.5.1), which unlike the previous example will insure we have interesting automorphisms since K contains all the roots of $x^3 - 2$. Let's label these roots $\sqrt[3]{2}, \omega, \omega^2$. Then:

$$[\mathbb{Q}(\sqrt[3]{2}, \sqrt{-3}) : \mathbb{Q}] = [\mathbb{Q}(\sqrt[3]{2}, \sqrt{-3}) : \mathbb{Q}(\sqrt[3]{2})] : [\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 2 \cdot 3 = 6$$

We know that $\mathbb{Q}(\sqrt[3]{2})$ contains one root of $x^3 - 2$ over $\mathbb{Q}$ ($\sqrt[3]{2} \in \mathbb{Q}(\sqrt[3]{2})$), and that $\mathbb{Q}(\sqrt[3]{2}, \omega)$ also contains the other two roots over $\mathbb{Q}$, and in particular over $\mathbb{Q}(\sqrt[3]{2})$ ($\omega, \omega^2 \in \mathbb{Q}(\sqrt[3]{2}, \omega)$.

Next, Since the degree of K is 6, we know it has a basis consisting of 6 elements, namely:

$$1,\ \sqrt[3]{2},\ \sqrt[3]{4},\ \omega,\ \omega\sqrt[3]{2},\ \omega\sqrt[3]{4}$$

Since any $\mathbb{Q}$-automorphism of K is a $\mathbb{Q}$-linear automorphism on K, it suffices to characterise these automorphisms by where by map these basis elements. By the fixed roots criterion, a $\mathbb{Q}$-automorphism σ on K can map $\sqrt[3]{2}$ to:

$$\sqrt[3]{2},\ \omega\sqrt[3]{2},\ \omega^2\sqrt[3]{2}$$

and map ω to:

$$\omega,\ \omega^2$$

Using this information, we can carry through computation's to see that $\mathrm{Aut}_{\mathbb{Q}}(K)$ is S_3. (I think we will encounter theorems that will make this easier later)

4. More on computing automorphisms groups here: here

Since any field homomorhpism must be injective, $\mathrm{Aut}_k(-)$ can in fact respect composition, making it a [contravariant] funtor[4]

Lemma 18.1.3: Field Automorphism Is Functorial

let $\mathbf{Field}_k$ be all fields such that any $F \in \mathbf{Field}_k$ has k as a subfield: $k \subseteq F$. Then $\mathrm{Aut}_k(-) : \mathbf{Field}_k \to \mathbf{Grp}$ sending F to $\mathrm{Aut}_k(F)$ and $F \to L$ to $\mathrm{Aut}_k(L) \to \mathrm{Aut}_k(F)$ via $\varphi \mapsto \varphi|_F$ is a contraviant functor

Proof **:**
simple definition pushing.

Though $\mathrm{Aut}_k(-)$ being functorial is quite powerful and a fact that shall be often used, it actually perserves even more structure, namely the latice of groups and interemetiate fields is [contravariantly] preserved. The following propositions builds up to this result. Since $\mathrm{Aut}_k(F)$ has a relatively easy to find structure, any connections between the structure of $\mathrm{Aut}_k(F)$ (which is a finite group if $[F : k] < \infty$) and F/k would allow us to use the full power of group theory (and better still finite group theory) to analise field extensions. We endevaour to deepen this connection in this chapter. One of the most important connection is that intermediate fields $k \subseteq E \subseteq F$ correspond to subgroups of $\mathrm{Aut}(F/k)$! This is called the *Galois correspondence*, and is what we will be building up for the rest of this section. We start by showing how subgroups of $\mathrm{Aut}_k(F)$ correspond to an intermediate field $k \subseteq E \subseteq F$, and conversely intermediate extension $k \subseteq E \subseteq F$ correspond a subgroup fo $\mathrm{Aut}_k(F)$

Proposition 18.1.4: Fixing F with automorphisms

Let $H \leqslant \mathrm{Aut}_k(F)$ be a subgroup of automorphisms of F (or $H \subseteq \mathrm{Aut}_k(F)$ more generally). Then the collection F^H of elements of F fixed by all the elements of H is a subfield of F extending k, $k \subseteq F^H \subseteq F$. Conversely, Let E be an intermediate field $k \subseteq E \subseteq F$. Then $\mathrm{Aut}_E(F) \leqslant \mathrm{Aut}_k(F)$.

Proof **:**
Let F^H be the collection of fixed elements by element $h \in H$. Then by definition, every element of k is fixed by H, so F^H is nonempty and $k \subseteq F^H$. Next, for any $a, b \in F^H$

$$h(a \pm b) = h(a) \pm h(b) = a \pm b \qquad h(ab) = h(a)h(b) = ab \qquad h(a^{-1}) = h(a)^{-1} = a^{-1}$$

that is, F^H is closed under the operations of F, and so $F^H \subseteq F$, showing that $k \subseteq F^H \subseteq F$.

For the other direction, let $\mathrm{Aut}_E(F)$ be the set of automorphisms that fix E. This subset is nonempty since $\mathrm{id}_F \in \mathrm{Aut}_E(F)$. If $\sigma \in \mathrm{Aut}_E(F)$, clearly $\sigma^{-1} \in \mathrm{Aut}_E(F)$. Finally, for any $\sigma, \tau \in \mathrm{Aut}_E(F)$, given $e \in E$,

$$\sigma \circ \tau^{-1}(e_1) = \sigma(\tau^{-1}(e)) = \sigma(e) = e$$

showing that $\sigma \circ \tau^{-1} \in \mathrm{Aut}_E(F)$, an so $\mathrm{Aut}_E(F) \leqslant \mathrm{Aut}_k(F)$, completing the proof.

[4]Note that $\mathrm{Aut}_{\mathbf{Grp}}(-)$ is *not* a functor; there are counter-examples, see exercise ref:HERE

Definition 18.1.5: Fixed Field

Let $H \subseteq \mathrm{Aut}_k(F)$. Then the subfield of F fixed by all the elements of H is called the *fixed field* of H, and is denoted F^H.

Example 18.1.3: Fixed Field

1. The fixed field of $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\sqrt{2}))$ is $\mathbb{Q}$. As we have seen in example 18.1.2.
2. The fixed field of $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\sqrt[3]{2}))$ is $\mathbb{Q}(\sqrt[3]{2})$, since $\mathrm{Aut}_{\mathbb{Q}}(\sqrt[3]{2}) = \{\mathrm{id}\}$. If we add the other roots of $x^3 - 2$, then the fixed field of $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\sqrt[3]{2}, \omega))$ is $\mathbb{Q}$.

Using this vocabulary, we can make the following correspondence betweeen F/k and its intermediate fields with $\mathrm{Aut}_k(F)$ and it's subgroups:

Proposition 18.1.6: Galois Correspondence

Let $k \subseteq F$ be a fields extension. Then there exists a correspondence:

$$\left\{\begin{matrix}\text{Intermediate fields } E \\ k \subseteq E \subseteq F\end{matrix}\right\} \underset{F^H}{\overset{\mathrm{Aut}_E(F)}{\rightleftarrows}} \left\{\begin{matrix}\text{subgroups } H \\ H \subseteq \mathrm{Aut}_k(F)\end{matrix}\right\}$$

sending an intermediate field E to the subgroup $\mathrm{Aut}_E(F)$, and a subgroup H of $\mathrm{Aut}_k(F)$ to the fixed field F^H. Furthermore, we have the following properies:

1. The correspondence is inclusion reversing, that is:
   - if $k \subseteq E_1 \subseteq E_2 \subseteq F$ are two subfield of F extending k , then $\mathrm{Aut}_{E_2}(F) \leqslant \mathrm{Aut}_{E_1}(F)$
   - if $H_1 \leqslant H_2 \leqslant \mathrm{Aut}_k(F)$ are two subgroups of automorphsism with associated fixed fields F^{H_1} and F^{H_2} respectively, then $F^{H_2} \subseteq F^{H_1}$.
2. For all subgroups $H \leqslant \mathrm{Aut}_k(F)$ and intermediate fields $k \subseteq E \subseteq F$
   - $E \subseteq F^{\mathrm{Aut}_E(F)}$
   - $H \subseteq \mathrm{Aut}_{F^H}(F)$
3. If E_1E_2 is the smallest subfield of F containing the two intermediate the E_1 and E_2, and $\langle G_1, G_2\rangle$ denotes the smallest subgroup of $\mathrm{Aut}_k(F)$ containing G_1 and G_2, then:

$$\begin{aligned}\mathrm{Aut}_{E_1E_2}(F) &= \mathrm{Aut}_{E_1}(F) \cap \mathrm{Aut}_{E_2}(F) \\ F^{\langle G_1, G_2\rangle} &= F^{G_1} \cap F^{G_2}\end{aligned}$$

Proof **:**
The Galois correspondence is an immediate consequence of proposition 18.1.4.

For the inclusion-reversal, consider the extension, $k \subseteq E_1 \subseteq E_2 \subseteq F$. Then $\mathrm{Aut}_{E_2}(F)$ *a fortiori* fixes all the elements of E_1, since $E_1 \subseteq E_2$, and so can only be *smaller* as the number of elements in Aut_{E_2} depend on the number of roots that can be premuted (as shown in the root permutation

Criterion). A similar argument can be made for $H_1 \leqslant H_2 \leqslant \mathrm{Aut}_k(F)$.

For the subsets, Take the intermediate field E, $k \subseteq E \subseteq F$. Then $F^{\mathrm{Aut}_E(F)}$ is the set of elements That are fixed by the automorphisms that fix E. Then it is immediate that $E \subseteq F^{\mathrm{Aut}_E(F)}$. Similarly for $H \subseteq \mathrm{Aut}_{F^H}(F)$.

For the minimality statements, If E_1E_2 is the smallest of F containing both E_1 and E_2, then the $\subseteq$ inclusion is immediate (since $\mathrm{Aut}_{E_1}(F) \cap \mathrm{Aut}_{E_2}(F)$ contains only elements that fix E_1 and E_2), while the $\supseteq$ comes from the fact that $\mathrm{Aut}_{E_1}(F)$ and $\mathrm{Aut}_{E_2}(F)$ are in themselves the minimal subgroups of $\mathrm{Aut}_k(F)$ that fix all elements of E_1 and E_2. A similar argument can be made for the second assertion.

It is natural to ask whether the galois correspondence is bijective. Before showing when it is, we will show that it is possible that it is not:

Example 18.1.4: non-bijective Galois Correspondence

Consider $Q(\sqrt[3]{2})/\mathbb{Q}$. Then we know that $[\mathbb{Q}(\sqrt{3})/\mathbb{Q}] = 3$, which is prime, and so the only intermediate fields are $\mathbb{Q}$ and $\mathbb{Q}(\sqrt[3]{2})$. For $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\sqrt[3]{2}))$, recall that $\mathbb{Q}(\sqrt[3]{2}) \subseteq \mathbb{R}$, and that $\sqrt[3]{2}$ is the only root of the minimal polynomial within $\mathbb{R}$. Thus, since automorphisms of fields permute roots, $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\sqrt[3]{2}))$ can only have one element:

$$\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\sqrt[3]{2})) = \{e\}$$

Hence, we have:

$$\{\mathbb{Q}, \mathbb{Q}(\sqrt[3]{2})\} \rightleftarrows \{\,\{e\}\,\} = \{\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\sqrt[3]{2}))\}$$

showing that the function from intermediate fields to subgroups of the automorphism group is not injective with respect to intermediate fields in general[a] (for every intermediate field, we may map it to the same subgroup of $\mathrm{Aut}_k(F)$). In this case, $\mathbb{Q} \subseteq F^{\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\sqrt[3]{2}))}$ and $\mathbb{Q}(\sqrt[3]{2}) \subseteq F^{\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\sqrt[3]{2}))}$

[a] Or similarly, the correspondence is not surjective with respect to subgroups of $\mathrm{Aut}_k(F)$

This examples shows that it possible that $E \subsetneq F^{\mathrm{Aut}_E(F)}$. Fortunately, in the finite case, we always have that for any $H \leqslant \mathrm{Aut}_k(F)$, there exists an intermediate field (namely F^H) such that $H = \mathrm{Aut}_{F^H}(F)$ (and hence, the correspondence is at least always surjective with respect to intermediate fields). To prove this result we need the following lemma which shows that the field F^H is essentially as nice as a field as we want. The high-level idea is that automorphisms in $H = \mathrm{Aut}_E(F)$ fix E, and F^H is *defined* to be the field that is fixed by H:

Lemma 18.1.7: Properties Of Fixed Field

Let F/k be a finite extension, and let $H \leqslant \mathrm{Aut}_k(F)$. Then F/F^H is normal and separable. It is furthermore finite and simple.

Note that we are not given that F^H or k is a perfect field, and so separability is a powerful result. Similarly, we do not know beforehand that F is a normal extension of k, yet F is a normal extension of F/F^H. Note too that we can take $H = \{e\}$ so that $F/F^{\{e\}}$ has all the aforementioned properies. As we've seen in the previous example, it is not necessarily the case that $F^{\{e\}} = k$. (in the previous example, we had $F^{\{e\}} = \mathbb{Q}(\sqrt[3]{2})$)

***Proof* :**
First, F/F^H is finite since F/k is finite, so by the tower law:

$$[F:k] = [F:F^H][F^H:k]$$

Next, to show F/F^H is a separable extension, we'll show any $\alpha \in F$ is separable over F^H. To do so, we use the following clever trick: Let $\alpha \in F$ be any element of F (possibly an element of F^H). We'll construct the minimal polynomial of α, and show that it is separable. Take $\sigma \in H$. Since F is finite over k, α is the root of some minimal polynomial, $m_{\alpha,k}(x)$. By the fixed roots criterion, $h\alpha$ must be a root of the minimal polynomial $m_{\alpha,k}(x)$ of α over k (which might also imply $h\alpha = \alpha$ if no other root are in F). Since $m_{\alpha,k}(x)$ has only finitely many roots, the H-orbit of α (i.e. all possibles distinct images of α under all possible $\sigma \in H$) consists of finitely many elements $\alpha_1, ..., \alpha_n$. Now, the group H acts on the H-orbit by premuting its elements. Therefore, every element of H keeps the polynomial:

$$q_\alpha(x) = (x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_n)$$

fixed (where $\alpha_1 = \alpha$). In other words, the coefficients of $q_\alpha(x)$ *must* be in the fixed field F^H: $q_\alpha(x) \in F^H[x]$! Next, we show that $q_\alpha(x) = m_{\alpha,F^H}(x)$, which will show that the minimal polynomial corresponding to α is separable, showing α is separable. To see this, let $g(x)$ be any other polynomial having α as a root. Then:

$$g(\alpha) = b_m\alpha^m + \cdots + b_1\alpha + b_0 = 0$$

and by definition, any $\sigma \in H$ will map α to another root α_i, so:

$$g(\alpha_i) = b_m\alpha_i^m + \cdots + b_1\alpha_i + b_0 = 0$$

meaning $g(x)$ is divisible by $x - \alpha_i$ for every i, and hence $q_\alpha(x) \big| g(x)$. Since the minimal polynomial of α over F^H also has α as a root, we have that $q_\alpha(x) \big| m_{\alpha,f^H}(x)$, and as we've shown before $m_{\alpha,F^H}(x) \big| q_\alpha(x)$. Thus, $q_\alpha(x) = m_{\alpha,F^H}(x)$ up to a constant. Thus, α is separable and since α was arbitrary, F/F^H is a separable extension (Note that it's possible that $\alpha \in F^H$, in which case $q_\alpha(x) = x - \alpha$, which is certainly the minimal polynomial of α over F^H and is separable).

Next, since F/F^H is finite and separable, then by theorem 17.6.17, it simple. Let's say $F = F^H(\beta)$, so $F^H(\beta)/F^H$ is simple. Let $q_\beta(x)$ be define as we've done above (note that $q_\beta(x) = m_{\beta,F^H}(x)$ and that it is not possible that $q_\beta(x) = x - \beta$ unless $F = F^H$)

Finally, since $q_\beta(x) \in F^H[x]$ splits in F, and F is generated over F^H (namely, F is generated by F and β), F is the splitting field of $q_\alpha(x)$ over F^H, and so $F^H \subseteq F$ is a normal extension by theorem 17.5.12, completing the proof (notice again how finiteness was important for this part of the proof)

(checkout other proofs of this that do not require using the fact it's a simple extension, since the primitive roots theorem is sometimes proven using Galois Theory)

This proposition provides us with a couple more results: The polynomial $q_\alpha(x)$ (say of degree n) can be expanded to:

$$q_\alpha(x) = x^n - s_1x^{n-1} + s_2x^{n-1} - \cdots + (-1)^{n-1}s_{n-1}x + (-1)^n s_n$$

where each s_i are functions $s_i(\alpha_1, ..., \alpha_n)$ and represent:

$$s_1(\alpha_1, ..., \alpha_n) = \alpha_1 + \alpha_2 + \cdots + \alpha_n = \sum_i \alpha_i$$

$$s_2(\alpha_1, ..., \alpha_n) = \sum_{i<j} \alpha_i \alpha_j$$

$$s_3(\alpha_1, ..., \alpha_n) = \sum_{i<j<k} \alpha_i \alpha_j \alpha_k$$

$$\vdots$$

$$s_n(\alpha_1, ..., \alpha_n) = \alpha_1 \alpha_2 \cdots \alpha_n$$

Since $q_\alpha(x) \in F^H$, then each of these coefficients are in F^H. Furthermore, $\deg(q_\alpha(x)) \leq |H|$

Proposition 18.1.8: $H = \operatorname{Aut}_{F^H}(F)$

Let F/k be a *finite* extension, and let H be a subgroup of $\operatorname{Aut}_k(F)$. Then $|H| = [F : F^H]$ and

$$H = \operatorname{Aut}_{F^H}(F)$$

In particular, the Galois correspondence is surejctive for a finite field (for any subgroup $H \leqslant \operatorname{Aut}_k(F)$, there exists an intermediate field (namely F^H) such that $H = \operatorname{Aut}_{F^H}(F)$)

Intuitively, we want to show that $\operatorname{Aut}_{F^H}(F)$ *only* fixes F^H, and so $H = \operatorname{Aut}_{F^H}(F)$.

Proof **:**
Let $H \leqslant \operatorname{Aut}_k(F)$. By proposition 18.1.6, we have that $H \leqslant \operatorname{Aut}_{F^H}(F)$, so $|H| \leq |\operatorname{Aut}_{F^H}(F)|$. Since $\operatorname{Aut}_{F^H}(F)$ is finite, verifying that $\geq$ holds will complete the proof.

To see this, recall that we've shown in lemma 18.1.7 that $F = F^H(\alpha)$ for some $\alpha \in F$. Thus, by corollary 17.2.15, $|\operatorname{Aut}_{F^H}(F)|$ equals the number of distinct roots of $m_{\alpha,F^H}(\alpha)$. Though we do not know the minimal polynomial $m_{\alpha,F^H}(x)$, we do know that α is a root of the polynomial $q_\alpha(x)$ defined in lemma 18.1.7. By construction, the number of roots of $q_\alpha(x)$ must be less than or equal $|H|$, giving us:

$$|\operatorname{Aut}_{F^H}(F)| \leq |H|$$

And since $|\operatorname{Aut}_{F^H}(F)| \geq |H|$, we get $|\operatorname{Aut}_{F^H}(F)| = |H|$ and since $H \subseteq \operatorname{Aut}_{F^H}(F)$, then $H = \operatorname{Aut}_{F^H}(F)$, giving the desired equality.

Finally, to show that $[F : F^H] = H$, since F/F^H is finite, separable, and normal, by corollary 17.6.18, $[F : F^H] = |\operatorname{Aut}_{F^H}(F)|$, completing the proof.

A particular consequence of this is that $\operatorname{Aut}_{F^{\operatorname{Aut}_E(F)}}(F) = \operatorname{Aut}_E(F)$, since the intermediate field that maps to $\operatorname{Aut}_{F^{\operatorname{Aut}_E(F)}}(F)$ is $F^{\operatorname{Aut}_E(F)}$, and so the group that it maps to is $\operatorname{Aut}_E(F)$.

Note that F/k being a *finite* extension was necesary for the proof. The Galois correspondence is in fact not necessarily surjective in the infinite case. This correspondence can be recovered given a suitable topology known as the *Krull topology* on $\operatorname{Aut}_k(F)$, which is used to limit the galois correspondence to *closed* subgroups of the automorphism group, which will regain surjectivity. More on this in appendix HERE (or another chapter?)

18.1.1 Galois Correspondence

We finally have reached a point where we can define a Galois extension. the following theorem is sometime also called the *Galois Extension Equivalences*, but it will be called by it's more famous name:

Theorem 18.1.9: Fundamental Theorem of Galois Theory I

Let F/k be a finite field extension. Then the following are equivalent:

1. F is the splitting field of a separable polynomial $f(x) \in k[x]$ over k
2. F/k is separable and normal
3. $|\operatorname{Aut}_k(F)| = [F : k]$
4. $k = F^{\operatorname{Aut}_k(F)}$ is the fixed field of $\operatorname{Aut}_k(F)$
5. F/k is separable, and if K/F is an algebraic extension and $\sigma \in \operatorname{Aut}_k(K)$, then $\sigma(F) = F$
6. The Galois correspondence for F/k is a bijection

***Proof* :**
Many of these have already been shown to be equivalent. We have shown:

- $1 \Leftrightarrow 2$ in theorem 17.5.12
- $2 \Leftrightarrow 3$ in corrollary 17.6.18
- $3 \Leftrightarrow 4$ follows from proposition 18.1.8, since $[F : F^{\operatorname{Aut}_k(F)}] = |\operatorname{Aut}_k(F)|$, and since $k \subseteq F^{\operatorname{Aut}_k(F)}$, we have that

$$k = F^{\operatorname{Aut}_k(F)} \text{ if and only if } |\operatorname{Aut}_k(F)| = [F : k]$$

- $2 \Leftrightarrow 5$ follows from the 3rd equivalence of theorem 17.5.12 along with the fact that finite separable extensions are simple (theorem 17.6.17)

Thus, we'll show $6 \Rightarrow 4$ and $1 \Rightarrow 6$, which will complete the chain of implications.

Starting with $6 \Rightarrow 4$, assume the galois correspondence is bijective, and consider $E = F^{\operatorname{Aut}_k(F)}$. By proposition 18.1.8, $\operatorname{Aut}_{F^{\operatorname{Aut}_k(F)}}(F) = \operatorname{Aut}_k(F)$, and so E and $F^{\operatorname{Aut}_k(F)}$ map to the same subgroup: $\operatorname{Aut}_k(F) = \operatorname{Aut}_E(F)$. Since the correlation is bijective, it must be that $k = E$, implying 4.

Next, we'll show $1 \Rightarrow 6$. Since F/k is a finite extension, by proposition 18.1.8, we already know the Galois correspondence is surjective (i.e., has a right-inverse), so it suffice to show it's injective, or equivalently that it has a left-inverse. To that end, we need to show that every intermediate field E equals the fixed field of the corresponding subgroup $\operatorname{Aut}_E(F)$, that is $E = F^{\operatorname{Aut}_E(F)}$. So, pick some E such that $k \subseteq E \subseteq F$. Since F is the splitting field for some separable polynomial $f(x) \in k[x] \subseteq E[x]$, we see that F is the splitting field of the separable polynomial $f(x)$ over E (i.e., condition (1) holds for F/E too). Since $1 \Leftrightarrow 4$, then $E = F^{\operatorname{Aut}_E}(F)$, thus $1 \Rightarrow 6$ completing the proof.

It can be helpful to have the following visual to keep track of where intermediate fields are with respect to their galois group:

$$\begin{aligned} \mathrm{Aut}_E(F) &= G \\ E &= F^G \end{aligned}$$

Definition 18.1.10: Galois Extension

Let F/k be a finite extension. Then F is called *Galois* or F/k is called a *Galois extension* if it satisfies any of the conditions in theorem 18.1.9.

Note that we have limited the definition to *finite galois extensions*, due to the requirement to bring in more tools to study the infinite equivalence. Nothing is lost in doing so, since infinite galois extensions will naturally generalize and improve upon the results we'll present. From now on, we will take galois extension to mean finite galois extension.

Corollary 18.1.11: Galois Extension Of Char 0

Let F/k be an extension of characteristic 0. Then F is galois over k if and only if F is finite and normal over k

Note that the finite condition was only added since we are focusing on finite Galois extensions.

Proof **:**
If F is galois then it is *a priori* normal. Conversely, if F is finite and normal, then by corollary 17.6.6, any algebraic extension of F is separable, and hence by theorem 18.1.9, it is Galois.

Definition 18.1.12: Galois Group

Let F/k be a Galois extension. Then the corresponding automrophsim group $\mathrm{Aut}_k(F)$ is called the *galois group* of the extension. We will denote it by $\mathrm{Gal}_k(F)$ or $\mathrm{Gal}(F/k)$

The symbol $\mathrm{Gal}_k(F)$ is simply to emphasize we are working over a Galois extension, and doesn't reflect any additional structrue or data given to $\mathrm{Aut}_k(F)$. Furthermore, (word on $\mathrm{Gal}_k(F)$ acting on $k[x]$ or $F[x]$?)

Definition 18.1.13: Galois Group Of A Polynomial

Let $f(x)$ be a separable polynomial over $k[x]$. Then the *galois group* $\mathrm{Gal}_k(f(x))$ is the galois group of the splitting field of $f(x)$).

Definition 18.1.14: Abelian and Cyclic Extensions

Let F/k be a galois extension. Then F is called a *cyclic* (resp. *abelian*) extension if it's corresponding Galois group is cyclic (resp. abelian)
More generally, if the Galois group G has some property (ex. nilpotent, solvable, free), then the respetive extension is said to have that same property (ex. nilpotent extension, solvable extension, free extension)

Note that any vocabulary already used (like simple extension) will be prioritized over the group theory vocabulary (a simple extension is not related to simple group)

The equivalent conditions for Galois extensions gives us many details on results we have encountered earlier:

48: Remark

1. For any $H \leqslant \operatorname{Aut}_k(F)$, F/F^H is a Galois extension, and every intermediate field $k \subseteq E \subseteq F$ appears as a fixed field for a unique $H \leqslant \operatorname{Aut}_k(F)$ (If F/k is not Galois, this is not the case, like for $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$). This also means that by the 4th condition of Galois extension equivlences:
$$F^H = F^{\operatorname{Aut}_{F^H}(F)}$$
and since every intermediate field appears as a fixed field, we get that:
$$E = F^{\operatorname{Aut}_E(F)}$$

2. Let F/k be any finite extension. Then $|\operatorname{Aut}_k(F)|$ divides $[F:k]$. Taking any $G = \operatorname{Aut}_k(F)$, then
$$[F:k] = [F:F^G][F^G:k] \overset{!}{=} |\operatorname{Aut}_{F^{\operatorname{Aut}_k(F)}}(F)|[F^G:k] = |\operatorname{Aut}_k(F)|[F^G:k]$$
where $\overset{!}{=}$ comes from theorem 18.1.9(3). If F/k is Galois, then $F^G = k$, meaninig $|\operatorname{Aut}_k(F)|[F^G:k] = |\operatorname{Aut}_k(F)|[k:k] = |\operatorname{Aut}_k(F)|$.

3. Let F/k be a Galois extension. Then $|\operatorname{Aut}_k(F)|$ divides —$\operatorname{Aut}_{F^H}(F)$—. Taking any $H \leqslant \operatorname{Aut}_k(F)$ and $G = \operatorname{Aut}_k(F)$, then
$$|\operatorname{Aut}_k(F)| = [F:k] = [F:F^H][F^H:k] = |\operatorname{Aut}_{F^H}(F)|[F^H:k]$$
Switching up the notation, this means that $|G| = |H|[F^H:k]$, meaning $[F^H:k]$ is the index of H in G: $[F^H:k] = [G:H]$. The fact that it equals to the index means that it corresponds bijectively with the cosets $g\operatorname{Aut}_k(F^H)$ where $g \in \operatorname{Aut}_k(F)$, however it is yet to be shown how these two sets can correspond, only that they are the same size. In theroem 18.1.16, we'll see when this set of cosets forms a group.

4. It may be really difficult to figure out if we have found all intermediate fields of an infinite in general, namely since in general, there are infinitely many possible choices. However, it is relatively easy to find the subgroups of a finite group! Hence, using finite group theory allows us to bring in the powerful tools having that finiteness brings (including Sylow's theorems).

5. Note how the 5th condition is almost the condition for an extension being normal, where just also assumed that the extension is separable. Note that the 5th condition still works if the extension is not separable. This can motivate the idea that we can somhow make separability a superfluous condition given some appropriate generalization, and indeed we can. If we replace Galois Group with an appropriate *algebraic group*, then we can recover the bijective correspondence[a]

6. Given the base field is perfect, for any extension F/k, we can always extend it to its normal closure $N/F/k$, which is normal over k, and so is Galois. Hence, many field theory problems over $\mathbb{Q}$ can be translated "wlog" to Galois theory

[a] algebraic groups are schemes (with with a group structure), for which one of the motivations for their invention was to distinguish roots higher multiplicities at a point

If $E = F^H$, the last observation can be seen visually as:

$$\begin{array}{ccc} F & & \\ \Big| & \} & |H|=[F:E] \\ E & & \\ \Big| & \} & |G:H|=[E:k] \\ k & & \end{array}$$

Example 18.1.5: Galois Groups

1. Let $K = \mathbb{Q}(\sqrt{2})/\mathbb{Q}$. Then since $[K : \mathbb{Q}] = 2$. and $|\operatorname{Aut}(K/\mathbb{Q})| = 2$ (as we've seen in example 18.1.2), then K is galois with galois group

$$\operatorname{Gal}(\mathbb{Q}(\sqrt{2})/\mathbb{Q}) = \{1, \sigma\} \cong \mathbb{Z}/2\mathbb{Z}$$

With the identity automorphism and

$$a + b\sqrt{2} \mapsto a - b\sqrt{2}$$

2. More generally, let K be any quadratic extension of F ($K = F(\sqrt{D})$). If $\operatorname{ch}(K) \neq 2$, then K is Galois.

3. The extension $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is *not* Galois, since $|\operatorname{Aut}(\sqrt[3]{2})| = 1 < 3 = [\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}]$. Notice that $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt[3]{2})$ is not a fixed field, since the only fixed field is $\mathbb{Q}(\sqrt[3]{2})$.

4. Let $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$ be an extension of $\mathbb{Q}$. Then since K is the splitting field of $(x^2-2)(x^2-3)$, it Galois. To determine its Galois group, we know from example (17.3.4) that $[K : \mathbb{Q}] = 4$, giving us the amount of automorphisms to look for. Since $\sqrt{2} \notin \mathbb{Q}(\sqrt{3})$, by the fixed roots criterion, we see that any automorphism $\sigma \in \operatorname{Aut}_{\mathbb{Q}}(K)$ must map $\sqrt{2}$ to $\pm\sqrt{2}$, and since $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$, we see that any automorphism $\tau \in \operatorname{Aut}_{\mathbb{Q}}(K)$ must map $\sqrt{3}$ map $\pm\sqrt{3}$.

   Going through defining a σ and a τ by permuting $\sqrt{2}$ and $\sqrt{3}$ we see that:

$$\operatorname{Gal}(K/\mathbb{Q}) = \{1, \sigma, \tau, \sigma\tau\} \cong V_4 \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$$

Next, we can use proposition 18.1.6 to find the associated fixed field given the subgroups of $\mathrm{Gal}(K/\mathbb{Q})$. Using some arithmetic, we see that:

$$\begin{array}{cc} \{1\} & \mathbb{Q}(\sqrt{2}, \sqrt{3}) \\ \\ \{1, \sigma\} & \mathbb{Q}(\sqrt{3}) \\ \\ \{1, \sigma\tau\} & \mathbb{Q}(\sqrt{6}) \\ \\ \{1, \tau\} & \mathbb{Q}(\sqrt{2}) \\ \\ \{1, \sigma, \tau, \sigma\tau\} & \mathbb{Q} \end{array}$$

which exhausts all possible intermediate fields of $\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q}$

5. Take the splitting field of $x^4 - 2$ (say we choose adjoints $\mathbb{Q}(i, \sqrt[4]{2})$). Since $\mathrm{ch}(\mathbb{Q}) = 0$, the polynomial must be separable, and hence $\mathbb{Q}(i, \sqrt[4]{2})$ must be Galois. Then we know that:

$$[\mathbb{Q}(i, \sqrt[4]{2}) : \mathbb{Q}] = [\mathbb{Q}(i, \sqrt[4]{2}) : \mathbb{Q}(\sqrt[4]{2})][\mathbb{Q}(\sqrt[4]{2}) : \mathbb{Q}] = 4 \cdot 2 = 8$$

Telling us the Galois group has 8 elements. Since $i^n \sqrt[4]{2}$ are the 4 roots of $x^4 - 2$, we see that

$$\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(i, \sqrt[4]{2})) \leqslant S_4$$

Since $|\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(i, \sqrt[4]{2}))| = 2 \cdot 4$, $\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(i, \sqrt[4]{2}))$ has a Sylow 2-subgroup of S_4. By Sylow's 2nd theorem (theorem 4.4.13) this group then *must* be D_8.

This makes finding the automorphisms quite easily. Notice that the automorphism σ:

$$\begin{cases} \sqrt[4]{2} \longrightarrow i\sqrt[4]{2} \\ \\ i \longrightarrow i \end{cases}$$

is cyclic of order 4, and the automorphism τ:

$$\begin{cases} \sqrt[4]{2} \longrightarrow \sqrt[4]{2} \\ \\ i \longrightarrow -i \end{cases}$$

is cyclic of order 2, and that:

$$\sigma\tau = \tau^{-1}\sigma$$

Which gives us that $\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(i, \sqrt[4]{2})) \cong D_8$! We correspond the subgroups with the intermediate fields in example 18.1.6.

6. As we've seen with cyclotomic fields, by proposition 17.7.7,
$$\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta_n)) \cong (\mathbb{Z}/n\mathbb{Z})^\times$$
giving us a Galois group for every $n \in \mathbb{N}_{>0}$.
7. By theorem 17.7.8, all finite field are Galois, since they are the splitting field over the separable polynomial $x^{p^d} - x$. As we've seen in proposition 17.7.15, the galois group is cyclic of order d:
$$\mathrm{Aut}_{\mathbb{F}_p}(\mathbb{F}_{p^d}) \cong \mathbb{Z}/d\mathbb{Z}$$
and is generated by the frobenius k-automorphism $x \mapsto x^p$
8. (An example of a normal, but not separable extension, is $F_2(x^{1/2})/F_2(x)$ (since $x^2 - t$ is inseparable over $\mathbb{F}_2(x)$), and so it's not Galois)

The correspondence by theorem 18.1.9 is in fact even stronger: we can correspond the intermediate lattice to the subgroup lattice, and we can (Something about how the III gal fund thm feels like it's defininga group-action). The first extension of the theorem is to give the correspondence between the lattices of subgroups of the galois group and of the field extensions. To show two lattices corresond, we need to show that if A_1 and A_2 are two objects, then A_1A_2 (i.e. the smallest object containing both A_1 and A_2) corresponding to B_1B_2, or if the relation is contravariant, A_1A_2 corresponds to $B_1 \cap B_1$. This is what we show with subgroups and intermediate fields:

Theorem 18.1.15: Fundamental Theorem Of Galois Theory II

Let F/k be a Galois extension. Then the Galois correspondence is an inclusion-reversing isomorphism between the lattice of intermediate subfield of F/k with the lattice of subgroups of $\mathrm{Aut}_k(F)$.

In particular, If E_1, E_2 are intermediate fields and G_1, G_2 are the correspondong subgroups of $\mathrm{Aut}_k(F)$, then $E_1 \cap E_2$ coresponds to $\langle G_1, G_2\rangle$ and E_1E_2 corresponds to $G_1 \cap G_2$

***Proof* :**
From theorem 18.1.9, we get the bijection between the intermediate fields and the subgroups of $\mathrm{Gal}_k(F)$, and from the 3rd property of the Galois Correspondence (proposition 18.1.6), we get that:
$$\mathrm{Aut}_{E_1E_2}(F) = G_1 \cap G_2 \qquad F^{\langle G_1, G_2\rangle} = E_1 \cap E_2$$
completing the proof.

We usually write the lattice of fields with k on the bottom and F on top, while the lattice of groups usualy has $\{e\}$ on top and $\mathrm{Aut}_k(F)$ on bottom to more easily associate the two lattices:

$$\begin{tikzcd}
F \arrow[d, dash] & \{e\} \arrow[d, dash] \\
E \arrow[d, dash] & \mathrm{Aut}_E(F) \arrow[d, dash] \\
k & \mathrm{Aut}_k(F)
\end{tikzcd}$$

Some authors mark the Galois group though their own edges:

$$\begin{tikzcd}
F \arrow[d, dash, "\mathrm{Aut}_E(F)"'] \arrow[dd, bend left, dash, "\mathrm{Aut}_k(F)"] \\
E \arrow[d, dash] \\
k
\end{tikzcd}$$

Example 18.1.6: Lattice Correspondence

1. As we've shown in example (18.1.5)(3), where we corresponded the subgroups of $\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(\sqrt{2}, \sqrt{3}))$ with the intermediate fields $\mathbb{Q} \subseteq E \subseteq \mathbb{Q}(\sqrt{2}, \sqrt{3})$, we found that $\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(\sqrt{2}, \sqrt{3})) \cong \mathbb{Z}_2 \times \mathbb{Z}_2$ and intermediate fields for each subgroup. We can make this result more precise. If the subgroup lattice is:

$$\begin{tikzcd}
 & \langle(0,0)\rangle \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
\langle(1,0)\rangle \arrow[dr, dash] & \langle(1,1)\rangle \arrow[d, dash] & \langle(0,1)\rangle \arrow[dl, dash] \\
 & \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} &
\end{tikzcd}$$

then the corresponding intermediate field lattice is:

$$\begin{tikzcd}
 & \mathbb{Q}(\sqrt{2}, \sqrt{3}) \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
\mathbb{Q}(\sqrt{2}) \arrow[dr, dash] & \mathbb{Q}(\sqrt{6}) \arrow[d, dash] & \mathbb{Q}(\sqrt{3}) \arrow[dl, dash] \\
 & \mathbb{Q} &
\end{tikzcd}$$

We can futhermore deduce that there are *no* further intermediate extensions $k \subseteq E \subseteq F$.

2. As we've shown in example (18.1.5)(4), $x^4 - 2$ has splitting field $\mathbb{Q}(i, \sqrt[4]{2})$ with Galois group

$\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(i, \sqrt[4]{2})) = D_8$. Then:

$$\begin{tikzcd}
 & & D_8 = \langle a, b \rangle \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & & \\
 & \langle a^2, b \rangle \arrow[dl, dash] \arrow[d] \arrow[dr, dash] & \langle a \rangle \arrow[d, dash] & \langle a^2, ab \rangle \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
\langle b \rangle \arrow[drr, dash] & \langle a^2 b \rangle \arrow[dr, dash] & \langle a^2 \rangle \arrow[d, dash] & \langle ab \rangle \arrow[dl, dash] & \langle a^3 b \rangle \arrow[dll, dash] \\
 & & \{e\} & &
\end{tikzcd}$$

with corresponding intermediate field lattice:

$$\begin{tikzcd}
 & & \mathbb{Q} \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & & \\
 & \mathbb{Q}(\sqrt[4]{4}) \arrow[dl, dash] \arrow[d] \arrow[dr, dash] & \mathbb{Q}(i) \arrow[d, dash] & \mathbb{Q}(\sqrt[4]{4} i) \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
\mathbb{Q}(\sqrt[4]{2}) \arrow[drr, dash] & \mathbb{Q}(\sqrt[4]{2} i) \arrow[dr, dash] & \mathbb{Q}(i, \sqrt[4]{4}) \arrow[d, dash] & \mathbb{Q}(\sqrt[4]{2}(1+i)) \arrow[dl, dash] & \mathbb{Q}(\sqrt[4]{2}(1-i)) \arrow[dll, dash] \\
 & & \mathbb{Q}(i, \sqrt[4]{2}) & &
\end{tikzcd}$$

3. Take $\mathbb{F}_{64}/\mathbb{F}_2$, which is a degree 6 extension with Galois group $\mathrm{Aut}_{\mathbb{F}_2}(\mathbb{F}_{64}) \cong \mathbb{Z}/6\mathbb{Z}$. the subgroup lattice is:

$$\begin{tikzcd}
 & \{0\} \arrow[dl, dash] \arrow[dr, dash] & \\
\{0, 3\} \arrow[dr, dash] & & \{0, 2, 4\} \arrow[dl, dash] \\
 & \mathbb{Z}/6\mathbb{Z} &
\end{tikzcd}$$

which corresponds to:

$$\begin{tikzcd}[column sep=small]
 & \mathbb{F}_{64} \arrow[dl, dash] \arrow[dr, dash] & \\
\mathbb{F}_8 \arrow[dr, dash] & & \mathbb{F}_4 \arrow[dl, dash] \\
 & \mathbb{F}_2 &
\end{tikzcd}
\qquad
\begin{tikzcd}[column sep=small]
 & \{e\} \arrow[dl, dash] \arrow[dr, dash] & \\
\{e, \varphi^3\} \arrow[dr, dash] & & \{e, \varphi^2, \varphi^4\} \arrow[dl, dash] \\
 & \{e, \varphi, \varphi^2, \varphi^3, \varphi^4, \varphi^5\} &
\end{tikzcd}$$

The next result comes from recalling that if $k \subseteq E \subseteq F$ is a series of field extensions where F/k is normal, this does not imply E/k is normal (as we have discussed in proposition 17.5.13 and example 17.5.3). Intuitively, this can be thought of as E having some but not all the required roots so that F is normal (ex. $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt[3]{2}) \subseteq \mathbb{Q}(\sqrt[3]{2}, \varphi)$). However, it is sometimes the case that the intermediate fields are normal (ex. $Q \subseteq \mathbb{Q}(\sqrt{2}) \subseteq \mathbb{Q}(\sqrt{2}, \sqrt{3})$). We may ask when more generally is an extension E/k also normal, or more generally Galois? The following theorem answers that question:

Theorem 18.1.16: Fundamental Theorem Of Galois Theory III

Let F/k be a galois extension, and let E be an intermediate field. Then E/k is Galois if and only if $\operatorname{Aut}_E(F)$ is *normal* in $\operatorname{Aut}_k(F)$. If this is the case, then

$$\operatorname{Aut}_k(E) = \operatorname{Aut}_k(F) / \operatorname{Aut}_E(F)$$

Hence, *normal* subgroups arise from *normal* extensions. In fact, the term "normal" originate in field theory, and the term normal subgroup was coined after it's field theory counterpart.

The proof essentially comes down to a couple of key observations about the cosets of $\operatorname{Aut}_k(F)/\operatorname{Aut}_E(F)$ and how we can surmize information about them through group actions! We have already observed in remark Ref:HERE (under definition 18.1.14) that E/k has the same cardinality as $\{g \operatorname{Aut}_E(F) \mid g \in \operatorname{Aut}_k(F)\}$. We will make this correspondence more precise.

(We have seen in theorem 17.5.12 that E is normal if and only if $\iota(E) = E$ for all ι. We are building up the consequences of this I think)

Let F/K be a Galois extension. Identify $F \subseteq \overline{k}$ within some algebraic closure of $\overline{k}$. Let $k \subseteq E \subseteq F$ be some intermediate field, and consider:

$$I = \left\{ \iota : E \to \overline{k} \;\middle|\; \iota \text{ is a } k\text{-homomorphism} \right\} = \operatorname{Hom}_k(E, \overline{k})$$

We can equivalently see E/k as an arbitrary extension of k where for some $\iota \in I$, $\iota(E) \subseteq F$.

Claim 1: for all $\iota \in I$, $\iota(E) \subseteq F$

Proof. Let ι be a $\iota \in I$ such that $\iota(E) \subseteq F$. Since E is separable (since $\iota(E) \subseteq F$ and F is separable) and finite, is is simple by theorem 17.6.17, Thus $E = k(\alpha)$ for some $\alpha \in \overline{k}$ and α is the root some minimal polynomial over k, $m_{\alpha,k}(x)$. Since $\iota(\alpha)$ determines ι, looking at where it maps is all that

matters. Since $\iota(\alpha)$ is *some* root of the minimum polynomial $m_{\alpha,k}(x)$, and F is normal, F contains *all* roots of the minimal polynomial. But then, for any ι, $\iota(E) \subseteq F$, as we sought to show. □

Notice that

Claim 2: $\mathrm{Aut}_k(F)$ acts on I: for $g \in \mathrm{Aut}_k(F)$ and $\iota \in I$, $g \circ \iota \in I$

Proof. Since $\iota(E) \subseteq F$, and $g : F \to F$, we can intermret ι as $\iota : E \to F$ and compose the functions □

claim 3: $\mathrm{Aut}_k(F)$ is transitive on I (i.e. on $\mathrm{Hom}_k(E, \overline{k})$)

Proof. We need to show that for any $\iota_1, \iota_2 \in I$, $g \circ \iota_1 = \iota_2$. This follows from uniqueness of the splitting field. Since $\iota_1(E)$ and $\iota_2(E)$ each contain k and are each contained in F, and F is a splitting field over k for some polynomial, say $f(x)$ (since F is Galois, it follows from theorem 18.1.9). Hence, *a fortiori*, F must be the splitting field of $f(x)$ over $\iota_1(E)$ and over $\iota_2(E)$. Note that:

$$\iota_2 \circ \iota_1^{-1} : \iota_1(E) \to \iota_2(E)$$

is an k-isomorphism, and hence by the uniqueness of the splitting field, this isomorphism is extended to a k-isomorphism:

$$g : F \to F$$

By construction, $g|_{\iota_1(E)} = \iota_2 \circ \iota_1^{-1}$, or equivalently:

$$g \circ \iota_1 = \iota_2$$

as we sought to show. □

Thus, the galois group $\mathrm{Gal}_k(F)$ always acts transitively on $\hom_k(E, \overline{k})$ (do we know how it linkes corresponding embeddings more precisely? Is that the III thm?)

Now, choose an $\iota \in I$ and identify E with the intermediate field $k \subseteq \iota(E) \subseteq F$. Then $\mathrm{Aut}_E(F)[= \mathrm{Aut}_{\iota(E)}(F)]$ is a subgroup of $\mathrm{Aut}_k(F)$ of elements that restrict to the identity on $E[= \iota(E)]$. More particularly, $\mathrm{Aut}_E(F)$ consists of the *stabalizers of* I (i.e. for all $g \in \mathrm{Aut}_E(F)$, $g \circ \iota = \iota$). Thus, by the Orbit-Stabalizer Thoerem, we can deduce the following result:

Let $G = \mathrm{Aut}_k(F)$. Then consdering I as a G-set, I is isomorphic to the set of left-cosets of $\mathrm{Aut}_E(F)$ in $\mathrm{Aut}_k(F)$

49: key Isomorphism

$$\hom_k(E, \overline{k}) \quad \cong_G \quad \{g \,\mathrm{Aut}_E(F) \mid g \in \mathrm{Aut}_k(F)\}$$

(Aluffi says he views this statement as a key statement in Galois Theory, so it's worth pondering over it more. I believe it's becasue the bijection is not just a bijection, but also a G-isomorphism, and we see that through this result.)

***Proof* :**
of theorem 18.1.16

Let $G = \mathrm{Aut}_k(F)$ and $I = \left\{\iota : E \to \overline{k} \;\middle|\; \iota \text{ is a } k\text{-homomorphism}\right\}$. Then as we've just seen, I is a

G-set, and it's stabalizer is $\mathrm{Aut}_E(F)$. By proposition ref:HERE (I don't think I've proven this, it's proposition II.9.12 in Aluffi), if ι_1, ι_2 are both in the stabalizer, they are conjugates: there exists a $g \in \mathrm{Aut}_k(F)$ such that $g \circ \iota_1 \circ g^{-1} = \iota_2$, in particular, ι and id are conjugates for any $\iota \in I$. Thus, if $\mathrm{Aut}_E(F)$ is normal:

$$\mathrm{Aut}_{\iota(E)}(F) = \mathrm{Aut}_E(F)$$

which, by the bijectivity of the Galois correspondence for F/k, we have that $\iota(E) = E$. But then E/k satisfies the 5th equivalence of Galois Extensions in theorem 18.1.9, meaning E/k is Galois

Conversely, if E/k is galois, then $\iota(E) = E$ for all ι (again by the 5th equivalence of Galois extensions in theorem 18.1.9). Define the homomorphism ρ which takes $g \in \mathrm{Aut}_k(F)$ and restricts it to $E[= \iota(E)]$:

$$\rho : \mathrm{Aut}_k(F) \to \mathrm{Aut}_k(E)$$

This homomorphism is surjective (since the G-action acts transitively on I), and $\ker(\rho)$ is all automorphism that restrict to the identity on E, that is, $\mathrm{Aut}_E(F)$. But then, we know that the kernel of a group homomorphism is normal, and by the 1st isomorphism theorem for groups:

$$\mathrm{Aut}_k(F)/\mathrm{Aut}_E(F) \cong \mathrm{Aut}_k(E)$$

completing the proof.

("It sends subgroup to subgroups by conjugation", i.e. if E is associatd to $H = \mathrm{Aut}_E(F)$, then for any $\sigma \in \mathrm{Aut}_k(F)$, $\sigma(E)$ is associated to $\mathrm{Aut}_{\sigma(E)}(F) = \sigma H \sigma^{-1}$)

(There is another really powerful result. Let's say $F/E/k$ is a sequence of extensions, F/k and E/k both Galois (and hence F/E is Galois. Then the map

$$\mathrm{Gal}_E(F) \times \mathrm{Gal}_k(E) \to \mathrm{Gal}_k(F) \qquad (\sigma, \tau) \mapsto \sigma \circ \tau$$

is is well defined and bijective (namely, since the identity must map to the identity, and since E/k is normal $\mathrm{Gal}_k(F)$ is determined by what happens to E, and then fixing E, what happens to F. Thus, constructing Galois groups is much easier in this case!!! Note, however, that this is a homomorhpism if and only if $E \cap F = k$. Once I read through Dummit's and foote on Composite fields, this will be incorporated))

Corollary 18.1.17: Isomorphism Of Categories

Let $\mathbf{Fld}_k^E$ be the collection of fields F such that $k \subseteq F \subseteq E$ whose morphisms are k-homomorphisms and $\mathcal{O}_G$ where $G = \mathrm{Aut}_k(E)$ be the category whose objects are left G-sets G/H for every $H \subseteq G$ and the morphisms are G-equivariant maps $G/H \to G/K$. Then as categories, $\mathcal{O}_G^{\mathrm{op}} \cong \mathbf{Flb}_k^E$

Proof **:**
see p.21 of Category theroy in Context if you need some help

(more in dummit and foote on Galois Composite fields! Really cool!)

Example 18.1.7: Application To Lattices
(Draw two latties, and show the normal extensions)

50: Foreshadowing

Can give the connection between Galois group/extensions to Algebraic geometry with regular covers and deck transformations

Exercise 18.1.1

1. Let $L = \mathbb{Q}(\zeta_p)$ where p is an odd prime. Show there exists a subfield $K \subseteq L$ such that $K = \mathbb{Q}(\sqrt{p'})$ where
$$p' = \begin{cases} p & p \equiv 1 \bmod 4 \\ -p & p \equiv 2, 3 \bmod 4 \end{cases}$$

2. Let $K/\mathbb{Q}$ be an odd Galois Extension, Show that L must be totally real.

18.1.2 Composite Fields

We now take what we proved earlier and do two things with it: we extend the results for composite fields and we classify all cyclic Galois Extensions (given an appropriate restriction on the base field) In the previous chapters, there has consistenly been a question working with composite fields (exercise ref:HERE), in particular, it was asked to be shown that:

> If E/k and F/k are field extensions, and E/k is (finite/algebraic/normal/separable resp.), then EF/F is (finite/algebraic/normal/separable resp). If E/k and F/k are (finite/algebraic/normal/separable resp.) then EF/k is (finite/algebraic/normal/separable resp.)

Since [finite] Galois extensions are finite, normal, and separable, we see immediately that if E/k is Galois, then EF/F is Galois, and if E/k and F/k are Galois, so is EF/k. The natural question to ask is what's the galois group of a composite field and of EF/F:

Proposition 18.1.18: Galois Composite Fields I

Suppose E/k is Galois and F/k is any field extension. Then EF/F is Galois, and $\mathrm{Aut}_F(EF) \cong \mathrm{Aut}_{E\cap F}(E)$. Visually, we see that:

$$\begin{tikzcd}
 & KF \arrow[dl, dash] \arrow[dr, dash, "G"] & \\
F \arrow[dr, dash, "G"'] & & K \arrow[dl, dash] \\
 & F \cap K \arrow[d] & \\
 & k &
\end{tikzcd}$$

The reader should ponder why does F/k need to be finite for this theorem?

***Proof* :**
Using the Galois extension equivalences, we can quickly prove that EF/k is Galois, namely since E/k is Galois, it is the splitting field of a separable polynomial $p(x) \in k[x] \in F[x]$. The roots of $p(x)$ generate E over k ,and so they generate EF over F, i.e., EF is the splitting field of $p(x)$ over F, thus EF/F is a Galois Extension

To show that $\mathrm{Aut}_F(EF) \cong \mathrm{Aut}_{E\cap F}(E)$, we'll first show we can embedd $\mathrm{Aut}_F(EF)$ into $\mathrm{Aut}_k(E)$, and then find that the subgroup of $\mathrm{Aut}_k(E)$ associated to this embedding is $\mathrm{Aut}_{E\cap F}(E)$. For the embedding, consider first any F-automorphism $\sigma : EF \to EF$ (and hence k-automorphism). Then we can restrict σ to E, giving us $\sigma : E \to \sigma(E)$. Since E is Galois, $\sigma(E) = E$. Thus, any element of $\mathrm{Aut}_F(EF)$ get's mapped to an element of $\mathrm{Aut}_k(E)$, that is, we have a natural group homomorphism:

$$\rho : \mathrm{Aut}_F(EF) \to \mathrm{Aut}_k(E)$$

This homomorphism is injective: If the only element that maps to the identity is the identity, then we're done. Indeed, if $\sigma \in \mathrm{Aut}_k(E)$ is the identity on E, since any element $\mathrm{Aut}_F(EF)$ is the identity on F, the coresponding σ in $\mathrm{Aut}_F(EF)$ must be the identity on EF, showing injectivity.

Thus, $\mathrm{Aut}_F(EF)$ embeds into $\mathrm{Aut}_k(E)$, and so it is associated with some $H \leqslant \mathrm{Aut}_k(E)$. By proposition 18.1.8, $H = \mathrm{Aut}_{E^H}(E)$. It suffices to show that $E^H = E \cap F$. For the $\supseteq$ direction, since every $\sigma \in \mathrm{Aut}_F(EF)$ fixes F, the restriction $\rho(\sigma)$ fixes $E \cap F$, hence $E^H \supseteq E \cap F$. For the $\subseteq$ direction, take $\alpha \in E^H$. Then $\alpha \in E$. By definition, $F(\alpha)$ is in the fixed field of $\mathrm{Aut}_F(EF)$, which is F sinc F is Galois over EF/F (1st Galois Theorem (18.1.9)(4)). Thus, $\alpha \in F$, and so $E^H \subseteq E \cap F$. Therefore, $H = E \cap F$, and:

$$\mathrm{Aut}_F(EF) \cong \mathrm{Aut}_{E\cap F}(E)$$

as we sought to show

Example 18.1.8: Composite Fields Determining Abelian Extensions

1. As we have seen in section 17.7.1, if ζ_n is the nth root of unity, $\mathbb{Q}(\zeta_n)$ is the splitting field of $x^n - 1$ over $\mathbb{Q}$, and hence is Galois since $\mathrm{ch}(\mathbb{Q}) = 0$. We have shown in proposition ref:HERE that $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta_n)) \cong (\mathbb{Z}/n\mathbb{Z})^\times$.

   Now, let k be any field of characteristic 0. Then the splitting field K/k for $x^n - 1$ is the composite of k and $\mathbb{Q}(\zeta)$, namely $K = k(\zeta)$. Since $\mathbb{Q}/(\zeta)/\mathbb{Q}$ is Galois and $k/\mathbb{Q}$ is a field extension, by proposition 18.1.18, we get that:

   $$\mathrm{Aut}_k(k(\zeta)) \cong \mathrm{Aut}_{\mathbb{Q}(\zeta)\cap k}(\mathbb{Q}(\zeta)) \leqslant \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta)) \cong (\mathbb{Z}/n\mathbb{Z})^\times$$

   showing that $\mathrm{Aut}_k(k(\zeta))$ is some subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$. in particular, *any* such extension *must* be abelian.

   The fact that every extension of the form $k(\zeta)$ is abelian is quite powerful. In fact, they are relatively easy to classify. The study of Galois extensions that are abelian[a] is called *Kummer Theory*. We touch a bit of Kummer theory in the following, since we need it for showing no quintic is unsolvable, and will later show that all finite abelian group show up as a field extension of $\mathbb{Q}$ in section 22.5.1[b].

2. Generalizing the results from the exercsies, we know that if E/k and F/k are Galois, so is

EF/k. Can you find a way to link $\mathrm{Gal}_k(EF)$ with $\mathrm{Gal}_k(E)$ and $\mathrm{Gal}_k(F)$? (Hint: Consider $\mathrm{Gal}_k(EF) \to \mathrm{Gal}_k(E) \times \mathrm{Gal}_k(F)$, $\sigma \mapsto (\sigma|_E, \sigma|_F)$)

[a] which also must contain n distinct roots of unity for some $n \in \mathbb{N}$
[b] Perhaps surprisingly, this requires the development of algebraic number theory which is done in chapter 22

Corollary 18.1.19: Order Of Galois Composite

Let E/k be a galois extension and F/k be any finite extension. Then:

$$[EF : k] = \frac{[E : k][F : k]}{[E \cap F : k]}$$

Proof **:**
exercsie

Note that this formula does not hold in general if niether of the extensions is Galois.

Example 18.1.9: Failure of Formula with no Galois Extension
Let $k = \mathbb{Q}$, $E = \mathbb{Q}(\sqrt[3]{2})$, and $F = \mathbb{Q}(\zeta_3 \sqrt[3]{2}))$, since $EF = \mathbb{Q}(\zeta_3 \sqrt[3]{2}), \sqrt{3}$ (with minimal polynomial $x^3 - 2$, and hence degree 6 extension) and:

$$6 = [\mathbb{Q}(\zeta_3 \sqrt[3]{2})] \neq \frac{[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}][\mathbb{Q}(\zeta_3 \sqrt[3]{2}) : \mathbb{Q}]}{[\mathbb{Q} : \mathbb{Q}]} = 9$$

Using this, we may answer the arguably more intesesting questions of the composite and intersection of fields being galois over the same base field:

Proposition 18.1.20: Galois Composite Fields II

Let $E/k, F/k$ be galois extensions of F. Then:

1. $E \cap F$ is galois over k
2. EF is galois over k. It's galois group is isomorphic to:
$$G = \{(\sigma, \tau) \mid \sigma|_{E \cap F} = \tau|_{E \cap F}\} \leqslant \mathrm{Gal}_k(E) \times \mathrm{Gal}_k(F)$$
3. If $E \cap F = k$, then:
$$\mathrm{Gal}_k(EF) \cong \mathrm{Gal}_k(E) \times \mathrm{Gal}_k(F)$$

Proof **:**
D&F p.593

Corollary 18.1.21: Galois Closure

Let F/k be a finite separable extension. Then there exists a minimal finite galois extension E/k such that $E/F/k$.

Proof **:**
Simply take the composite of the splitting fields of the minimal polynomials for the basis of F over k (which is separable since F/k is separable), nad intersect all the galois extensions of k containing F.

Definition 18.1.22: galoisClosure

Let F/K be a finite separable extension. Then the smallest extension $E/F/k$ such that E/k is galois is called the *galois closure* of F over k.

Exercise 18.1.2

1. Show there exists fields E, F such that $[EF : F] \neq [F : E \cap F]$

18.1.3 A bit of Character Theory

(a bit of character theory, more of it in chapter ref:HERE)

(as far as I can tell, this is needed to prove an essential part for the solvability of the quintic. In particular, it is used to say that F/k is galois and $\mathrm{Aut}_k(F)$ is cyclic if and only if $F = k(\delta)$, $\delta^m \in k$. I think it is used as a clean way to write "algebraically independent" since the values of the coefficients can be anything)

Definition 18.1.23: Group Character

Let G be a group and L a field. Then a *[linear] character* (denoted χ) is a homomorphism from G to the multiplicative group of L:

$$\chi : G \to L^\times$$

Example 18.1.10: Group Character
For our purposes, any field homomorphism $\sigma : F \to K$ induces a character $\chi_\sigma : F^\times \to K^\times$ (since σ maps the multiplicative group of F into the multiplicative group of K). Note that χ_σ captures essentially all the mapping information of σ, since χ only ommits mapping 0 from F to 0 from K, but we can simply deduce that. Thus, if we are working with χ_σ, we are showing that we do not want to remeber that the additive structure is preserved.

Definition 18.1.24: Character Linearly Independent

The characters $\chi_1, ..., \chi_n$ of G are *linearly independent* over L if they are linearly indepenent as functions on G, i.e. for any $a_i \in L$

$$a_1\chi_1 + a_2\chi_2 + \cdots + a_n\chi_n = 0 \quad \text{if and only if} \quad a_i = 0, 1 \leq i \leq n$$

Note that the expression on the left hand side of the above equality is a large function, and hence the it means that characters are linearly indepenndent if for all $g \in g$:

$$a_1\chi_1(g) + a_2\chi_2(g) + \cdots + a_n\chi_n(g) = 0 \quad \text{if and only if} \quad a_i = 0, 1 \leq i \leq n$$

Theorem 18.1.25: Linear Independence Of Characters

Let $\chi_1, ..., \chi_n$ be distinct characters of G with values in L. Then they are linearly independent over L

Proof **:**
Suppose that the characers are linearly dependent, that is, there exists (possibly many) sets of coefficeints coefficients $a_1, ..., a_n \in L$ where not all are zero and

$$a_1\chi_1 + a_2\chi_2 + \cdots + a_n\chi_n = 0$$

To construct a minimality argument, choose a set of not-all-zero coefficients of minimal cardinaliy that satisfies the above equation (such coefficeints exists since these character are linealy dependent). Thus, for any $g \in G$:

$$a_1\chi_1(g) + a_2\chi_2(g) + \cdots + a_n\chi(g) = 0 \tag{18.2}$$

Let g_0 be an element of G such that $\chi_1(g_0) \neq \chi_n(g_0)$ (which exists since $\chi_1 \neq \chi_n$). Since equation (18.2) holds for every element of G, we have that:

$$a_1\chi_1(g_0g) + a_2\chi_2(g_0g) + \cdots + a_n\chi_n(g_0g) = 0 \tag{18.3}$$

i.e.

$$a_1\chi_1(g_0)\chi_1(g) + a_2\chi_2(g_0)\chi_2(g) + \cdots + a_n\chi_n(g_0)\chi_n(g) = 0 \tag{18.4}$$

Now, multiplying equation (18.2) by $\chi_n(g_0)$ and subtracting by equation (18.4), we get:

$$[\chi_n(g_0) - \chi_1(g_0)]a_1\chi_1(g) + \cdots + [\chi_n(g_0) - \chi_{n-1}(g_0)]a_{n-1}\chi_{n-1}(g) = 0$$

which holds for all $g \in G$. Howver, the first coefficeint is nonzero, and this is a relatoin with fewer nonzero coefficeints than the minimal choice we have made earlier – a contradiction.

Corollary 18.1.26: linear independence of Automorphisms (Dedekind's Lemma)

Let $\sigma_1, \sigma_2, ..., \sigma_n$ be distinct embeddings of K into L. Then they are linearly indepednent as functions on K. In particular, distinct automorphism of a field K are linearly independet as functions on K.

Proof **:**
This is an immediate consequence of theorem 18.1.25

Using some character theory, we can

Proposition 18.1.27: Galois And Cyclic Extensions

Let F/k be an extension of degree m. Assume that k contains a primitive m-th root of 1 and $\mathrm{ch}(k)$ does not divide m. Then F/k is Galois and cyclic if and only if $F = k(\delta)$ with $\delta^m \in k$.

This proposition is part of the more general study of *Kummer theory* which studies abelian extensions.

Proof **:**
Let $\zeta \in k$ be a primitive m-th root of 1.

($\Leftarrow$) Let's say $F = k(\delta)$ with $\delta^m = c \in k$. Then all m roots of the polynomial $x^m - c$ are in F, in particular,

$$c, \zeta\delta, \zeta^2\delta, ..., \zeta^{m-1}\delta \in F$$

Furthermore, F must be generated by them. Hence F is the splitting field of the separable polynomial $x^m - c$ in k (separable since $\mathrm{ch}(k)$ does not divide m). Thus, F is Galois

Next, to see that $\mathrm{Aut}_k(F)$ is cyclic, note that for any $\varphi \in \mathrm{Aut}_k(F)$, it is determined by $\varphi(\delta)$, which his necessarily a root of $x^m - c$, and so $\varphi(\delta) = \zeta^i\delta$ for some i determined up to a multiple of m. Then it is easy to see that this can be used to define an isomorphism of $\mathrm{Aut}_k(F)$ with $\mathbb{Z}/m\mathbb{Z}$.

($\Rightarrow$) Conversely, assume that F/k is Galois and $\mathrm{Aut}_k(F)$ is cyclic (say, generated by φ). Since the automorphisms φ^i $(1 \leq i \leq m-1)$ are pairwise distinct, by corollary 18.1.26, they are linearly independent. Therefore, there exists an $\alpha \in F$ such that

$$\delta := \alpha + \zeta^{-1}\varphi(\alpha) + \cdots + \zeta^{-(m-2)}\varphi^{m-2}(\alpha) + \zeta^{-(m-1)}\varphi^{m-1}(\alpha) \neq 0$$

Then note that :

$$\varphi(\delta) := \varphi(\alpha) + \zeta^{-1}\varphi^2(\alpha) + \cdots + \zeta^{-(m-2)}\varphi^{m-1}(\alpha) + \zeta^{-(m-1)}\alpha = \zeta\delta$$

since $\zeta \in k$ so $\varphi(\zeta) = \zeta$. Then δ is *not* fixed by any nonidentity element of $\mathrm{Aut}_k(F)$, that is $\mathrm{Aut}_{k(\delta)}(F) = \{e\}$, so $k(\delta) = F$. Furthermore:

$$\varphi(\delta^m) = \varphi(\delta)^m = \zeta^m\delta^m = \delta^m$$

so, δ^m is fixed by φ, and hence by every element of $\mathrm{Aut}_k(F)$. Since F/k is Galois, it must be that $\delta^m \in k$, completing the proof.

Exercise 18.1.3

1. Let $k(t)/k$ be an infinite extension. Determine $\mathrm{Aut}_k(k(t))$

2. Show that for all n, there is an extension F of $\mathbb{Q}$ such that $F/\mathbb{Q}$ is not Galois. Similarly, show that for all n, there exists an extension L such that $\mathrm{Gal}_{\mathbb{Q}}(L) \cong \mathbb{Z}/n\mathbb{Z}$ (Hint: show that for all n, there exists a p such that $p \equiv 1 \bmod n$)

18.2 Application of Galois Theory

18.2.1 Fundamental Theorem of Algebra

(Some build-up with properties of $\mathbb{C}$ that are given a exercsie 5.23)

Theorem 18.2.1: Fundamental Theorem of Algebra

$\mathbb{C}$ is Algebraically closed

Proof **:**
Take $f(x) \in \mathbb{C}[x]$ to be a nonconstant polynomial. If $f(x)$ is irreducible, then since conjugation is a field-homomorphism, $f(x) \to \overline{f(x)}$, then $f(x)\overline{f(x)}$ is also irreducible. Since $f(x)\overline{f(x)} \in \mathbb{R}[x]$, we can assume without loss of generality that all coefficents are real. Let F be the splitting field of $f(x)$ over $\mathbb{R}$, and embed $F \subseteq \overline{\mathbb{R}}$ (we want to show that $\overline{\mathbb{R}} = \mathbb{C}$).

Now, consider the extension:

$$\mathbb{R} \subseteq F(i)$$

this extension is Galois: it is the splitting field of the square-free part of $f(x)(x^2+1)$ (any irreducible polynomial over $\mathbb{R}$ of degree ≥ 2 will have such a square free part, as should be verified TBD). Let $G = \mathrm{Aut}_{\mathbb{R}}(F(i))$ By Sylow's Theorem, G has a 2-sylow subgroup H (even if $2 \nmid |G|$, since then it would be $\{e\}$). Then the index $[G : H]$ is oddand so by the Galois correspondence, it corresponds to a finite separable extension $\mathbb{R} \subseteq E$ with $[E : \mathbb{R}]$ odd. Since every polynomial of odd degree has a root by elementary calculus [a], it must be that $\mathbb{R} \subseteq E$ is *trivial* (since TBD, was exercsie VII.5.23 in Aluffi which stated that if $k[x]$ had no irreducible polynomials of degre $n > 0$, then it has no separable extensions). Thus, $[G : H] = 1$, proving that G is a 2-group. Since $\mathbb{C} = \mathbb{R}(i) \subseteq F(i)$, the Galois group ofthe Galois extension $|C \subseteq F(i)$ is a subgrou pof G, and therefore it is a 2-group: $|\mathrm{Aut}_{\mathbb{C}}(F(i))| = 2^n$ for some $n \geq 0$.

To finish the proof, we must show that $n = 0$. By proposition ref:HERE, for any $1 \leq m \leq n$, a group of order p^n contains a subgroup of order p^m. For this case, we have that $\mathrm{Aut}_{\mathbb{C}}(F(i))$ contains a subgroup of order 2^{n-1}. However, by exercise 5.23 again, $\mathbb{C}$ contains *no* quadratic exensions of $\mathbb{C}$, since there are no irreducible polynomials in $\mathbb{C}[x]$.

Thus, $n = 0$, so $\mathrm{Aut}_{\mathbb{C}}(F(i))$ is trivial, and so $F(i) = \mathbb{C}$, in particular $\mathbb{C}$ contains the roots of $f(x)$.

[a] Namely, the intermediate value theorem

This proof can also be accomplished without Sylow's theorem's by HERE. However, since the tools were developped, it is good practice to use them.

Corollary 18.2.2: No Field Structure On $\mathbb{R}^n$, $n \geq 3$

There does not exist a field structure on $\mathbb{R}^n$ ($n \geq 3$) such that $\mathbb{R}$ includes into $\mathbb{R}^n$ (i.e. $\mathbb{R}^n$ is a field extension of $\mathbb{R}$)

Proof **:**
For the sake of contradiction, assume $\mathbb{R} \subseteq \mathbb{R}^n$ is a field extension for $n \geq 3$. Since $\mathbb{R}^n$ is an n-dimensional vector-space, it is a degree n algebraic extension. Hence, by lemma 17.5.7, there is an (injective) homomorhpism $\mathbb{R}^n \to \mathbb{C}$ fixing $\mathbb{R}$. However, the dimension of the right hand side is greater than 2, and hence no such homomorhpism exists, a contradiction.

Note that we had to be careful in how we state the above corollary. It is not true that there is no field structure on $\mathbb{R}^n$ ($n \geq 3$). Since $\mathbb{R}^n$ and $\mathbb{R}$ have the same cardinality, we may (by the axiom of choice) find a bijection from $\mathbb{R}^n$ to $\mathbb{R}$. Then we may put a field structure on $\mathbb{R}$ through the bijection. However, the field structure on $\mathbb{R}^n$ would certainly not have $\mathbb{R}$ as a subfield, and it impossible to determine the product of any two elements[5].

Theorem 18.2.3: Frobenius Theorem

There are only 3 finite-dimensional associative division algebras, namely

1. $\mathbb{R}$, the real numbers
2. $\mathbb{C}$, the complex numbers
3. $\mathbb{H}$, the Quaternions

Proof **:**
It's on Wikipedia, I thought it really cool

As an exercise, show that $\mathbb{H}$ is not a $\mathbb{C}$-algebra, and that D is the *Clifford algebra* of $\mathbb{R}^n$ (see ref:HERE (will put in section 15.4))

We leave this section with a mention of *Hurwitz's Theorem* which classifies unital real *not necessarily associative* algebras endowed with a positive definite quadratic form (recall quadratic forms are in correspondence with bilinear forms). There are 4 such structures, which are the 3 in Frobenius Theorem as well as the Octonions $\mathbb{O}$. These algebras had found some interesting applications in algebraic topology in the classification of vector fields on spheres, as well as classifying the homotopy groups (see part VII, section ref:HERE) of many classical groups

[5] loosely speaking, if we would be able to explicitly find a bijection from $\mathbb{R}^n$ to $\mathbb{R}$, but that would be a "countable process", but $\mathbb{R}^n$ and $\mathbb{R}$ is uncountable

18.2.2 Construction of Regular n-gon

18.2.3 Symmetric Functions

If $t_1, ..., t_n$ are indeterminates, then the polynomial:

$$P_n(x) := (x - t_1)(x - t_2) \cdots (x - t_n) \in \mathbb{Z}[t_1, ..., t_n][x]$$

is universal in the sense thati every polynomial of degree n with coefficeints in (say) an intergral domain may be obtained from $P_n(x)$ given a suitable choice for $t_1, ..., t_n$, choices form a possibly larger ring. Expanding this polynomial, we get that:

$$P_n(x) = x^n - s_1(t_1, ..., t_n)x^{n-1} + \cdots + (-1)^n s_n(t_1, ..., t_n)$$

where each s_i are the *elementary functions*. If $n = 3$, then:

$$\begin{aligned} s_1(t_1, t_2, t_3) &= t_1 + t_2 + t_3 \\ s_2(t_1, t_2, t_2) &= t_1t_2 + t_1t_3 + t_2t_3 \\ s_3(t_1, t_2, t_3) &= t_1t_2t_3 \end{aligned}$$

The function s_n are called symetric since they are invariant under permutation (essentially, since $P_n(x)$ is invariant under permutation). The name "elementary" comes because of the following theorem

Theorem 18.2.4: Fundamental Theorem On Symmetric Functions

Let K be a field and let $\varphi \in K(t_1, ..., t_n)$. Then φ is invariant under permutation (i.e. symmetric) if and only if it is a raioanl functoin (with ceofficeint in K) of the elementary symmetric functions $s_1, ..., s_n$.

The functions s_i are viewed as element sof $K[t_1, ..., t_n]$ which we can do unmabiguously since $\mathbb{Z}$ is initial in **Ring**.

Proof **:**
Let $F = K(t_1, ..., t_n)$ and $k = K(s_1, ..., s_n)$ be the subfield generated by the elemntary symmetric functions over K. Then F is a splitting field of the separable polynomial $P_n(x)$ over k, i.e. F is Galois and:

$$|\operatorname{Aut}_k(F)| = [F : k] \leq n!$$

Notice that the symmetric group S_n acts (faithfully) on F by permutting $t_1, ..., t_n$ and the action is the identity on $k = K(s_1, ..., s_n)$ since each s_i is symmetric. Thus, it is clear that S_n can be identitfied with some subgroup of $\operatorname{Aut}_k(F)$. However,

$$n! = |S_n| \leq |\operatorname{Aut}_k(F)| \leq n!$$

showing that $\operatorname{Aut}_k(F) \cong S_n$

Now, since F/k is Galois, we get $k = F^{S_n}$. This in fact completes the rpoof: if $\varphi \in K(t_1, .., t_n)$, then φ is invariant under the S_n action on $t_1, ..., t_n$ if and only if $\varphi \in k \in K(s_1, ..., s_n)$, as we sought to show.

One important result that was proven in the previous theorem is noted for future reference:

Lemma 18.2.5: S_n Extension

Let K be a field, $t_1, ..., t_n$ indeterminates, and let $s_1, ..., s_n$ be the elementary symmetric functions on $t_1, ..., t_n$. Then the extension:

$$K(s_1, ..., s_n) \subseteq K(t_1, ..., t_n)$$

is Galois, with Galois group S_n

From this, we can prove that every finite group is represented as a galois group:

Theorem 18.2.6: Finite Group Represented by Galois Group

Let G be a finite group. Then there exists a Galois extension F/k such that $\mathrm{Aut}_k(F) \cong G$

Proof **:**
exercsie

Noe that it is not certain if every finite group G appears as an extension of F over $\mathbb{Q}$. This is called the *inverse Galois Problem*.

(aluffi develops words on discriminants and the inverse galois problem. p. 473)

Corollary 18.2.7

With the abaove notation, the extension

$$K(s_1, ..., s_n)(\sqrt{D}) \subseteq K(t_1, ..., t_n)$$

is Galois, with Galois group A_n.

18.2.4 Solvability By Radicals

We are finally in a position to prove that an arbitrary degree 5 polynomial cannot be written in terms of it's coefficients using the oprations $+, -, \times, \div, \sqrt{--}$. Is is famously known, any quadratic polynomial over k ($\mathrm{ch}(k) \neq 2$) can be written in terms of:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

and similar equations exists for degree 3 and 4 polynomials, though they get horrendously messy.

51: Goal

irreducible polynoials of degree $n \geq 5$ are unsolvable

To avoid complications, we will assume in this section that $\mathrm{ch}(k) = 0$.

Definition 18.2.8: Radical Extension

Let F/k be a field extension. Then if $F = k(\alpha_1, \alpha_2, ..., \alpha_n)$ where

$$k \subseteq k(\alpha_1) \subseteq k(\alpha_1, \alpha_2), \subseteq \cdots \subseteq k(\alpha_1, ..., \alpha_n)$$

where for each $i \in \{1, ..., n\}$, there exists an n_i such that $\alpha^{n_i} \in k(\alpha_1, ..., \alpha_{i-1})$, then F/k is called a *radical extension*

Elements of radical extensions are easy to picture: if $k = \mathbb{Q}$, then simply nest as many nth-roots as you please, for example:

$$\sqrt[7]{3 + \sqrt[2]{19}^5 \sqrt[11]{3+2}}$$

it is also clear by definition that a radical extension is the composition of cyclic exensions provided that the appropraite rotos of 1 are in k.

The next result insures us that any [separable] radical extension is contained in a radical extensions that is *Galois*:

Lemma 18.2.9: Separable And Galois Radical Extensions

Let F/k be a separable radical extension. Then F is conatiend in a Galois radical extension

Proof **:**
Let F/K be a separable radical extension. Then F is a separable finite extension of k, and so by the primite root theorem (theorem 17.6.17), there exists an $\alpha \in F$ such that $F = k(\alpha)$. Let L be the splitting field of $k(\alpha)$. Then we can see that L is equal to the composite of all the field $\sigma(F)$ for every σ that embeds F into $\overline{k}$:

$$L = \sigma_1(F) \cdots \sigma_n(F)$$

Then it is easy to see that the composition of radical extensions is radical, showing that L is radical. Thus, L is the splitting field of a separable polynomial, and so it is Galois, completing the proof.

Lemma 18.2.10: Solvability In Char 0

Let k be a field such that $\mathrm{ch}(k) = 0$, and let $f(x) \in k[x]$ be an irreducible polynomial. Then the roots of $f(x)$ can be written in terms of field operations and radicals if and only if the splitting field fo $f(x)$ is contined in a Galois Radical extension

Proof **:**

($\Leftarrow$) If the splitting field $f(x)$ is contained in a radical extension (not necessarily Galois), then *a fortiori* the roots can be written using field operations and radicals.

($\Rightarrow$) Conversely, assume the roots of the irreducible polynomial $f(x)$ can be written in terms of field operations and radicals . Then any formula for a root t_1 of $f(x)$ can be turned into

a formula for t_i by applying an element σ_i of $\mathrm{Aut}_k(\overline{k})$ sending t_1 to t_i (this automorphism exists since $f(x)$ is irreducible). Thus, if the formula exists for one root, a formula exists for all roots. Thus, $k(t_1), ..., k(t_n)$ are each radical extensions and

$$L = k(t_1)k(t_2)\cdots k(t_n)$$

is a radical extension that contains all the roots of $f(x)$. Thus, by minimality of the splitting field, it is contained in a radical extension of k. By lemma 18.2.9, it must also be contained in a galois radical extension, as we sought to show.

Thus, we can characterize when a polynomial is solvable by radicals by when it is contained in a Galois radical extension. Since the extension is Galois, we have a corresponding Galois group to analyze. We will soon see that the appropriate type of group that is needed for $f(x)$ to be solvable is one we have encountered a long time ago:

Definition 18.2.11: Solvable Extension

Let F/k be a Galois extension. Then F/k is a *solvable extension* if $\mathrm{Aut}_k(F)$ is a solvable group

Recall that if G is a solvable group and finite, then G can be decomposed into a subnormal series with cyclic factors:

$$G = G_0 \trianglerighteq G_1 \trianglerighteq G_2 \trianglerighteq \cdots G_n = \{1\}$$

notice how this parrallels the fact that any radical extension is a sequence of cyclic extensions.

Lemma 18.2.12: Relating Radical And Solvable Extensions

Let F/k be a galois extension with $\mathrm{ch}(k) = 0$. If F/k has enough roots of 1, then F/k is radical if and only if it is solvable

What is meant by "enough roots of 1" is that k contains primitive M-th roots of 1 for a common multiple M of the order of the corresponding cyclic factors.

***Proof* :**
This is a generalization of proposition 18.1.27.

Let F/k be a radical extension, so that:

$$k \subseteq k(\delta_1) \subseteq \cdots \subseteq k(\delta_1, ..., \delta_r) = F$$

where $\delta_i^{m_i} \in k(\delta_1, ..., \delta_{i-1})$. By assumption, k contains primitive m_ith roots of 1 for each i. Next, consider the corresponding subgroups of $G = \mathrm{Aut}_k(F)$:

$$G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_r = \{e\}$$

where $G_i = \mathrm{Aut}_{k(\delta_1,...,\delta_i)}(F)$. Next, note That F is Galois over each intermediate field (by proposition 17.5.13), and each extension:

$$k(\delta_1, ..., \delta_{i-1}) \subseteq k(\delta_1, ..., \delta_i)$$

satisfies proposition 18.1.27, and hence each G_i is cyclic. By theorem 18.1.16, each G_i is normal (I think since the automorphism group is abelian) in the previous G_{i-1} with a cyclic quotient, and hence G is solvable.

Conversely, the same argument works in reverse using the converse implication in proposition 18.1.27 to show that an extension with enough roots of 1 is radical.

We were dependent on each extension containing enough roots of unity. In the following, we show that we can drop them:

Proposition 18.2.13: Solvable In Char 0

Let k be a field set $\mathrm{ch}(k) = 0$, and let F/k be a Galois extension. Then F/k is solvable if and only if it is containd in a Galois Radical extension

Technically, we only need that $\mathrm{ch}(k)$ not divide the relevant degrees of the extensions, however for our purposes it safe to assume $\mathrm{ch}(k) = 0$.

Proof **:**

($\Rightarrow$) Assume $\mathrm{Aut}_k(F)$ is solvable. Let M be the common multiple of the cyclic quotients and let ζ be a M-th root of 1 in an algebraic closure $\overline{k}$ of k. The splitting field $k(\zeta)$ of $x^M - 1$ is Galois over k since $\mathrm{ch}(k) = 0$. By proposition 18.1.18, the extension $k(\zeta) \subseteq F(\zeta)$ is also Galois, with Galois group isomorphic to $\mathrm{Aut}_{k(\zeta)\cap F}(F)$. Since $k(\zeta) \cap F$ is Galois over k (Aluffi says exercise 6.13), this subgroup is normal, and it follows (corrollary IV.3.13) that $\mathrm{Aut}_{k(\zeta)}(F(\zeta))$ is solvable.

Since $k(\zeta) \subseteq F(\zeta)$ has enough roots of 1, by lemma 18.2.12, it is radical. Thus, so is $k \subseteq F(\zeta)$ since $k \subseteq k(\zeta)$ is by trivially radical. Thus, F/k is contained in a radical extension.

($\Leftarrow$) Conversely, assume that F/k is Galois and contained in a radical Galois extension L/k. Let M be a common multiple of the order of the cyclic factors in this extensions,and let ζ be a primitive m-th root of 1. The compsoite $L(\zeta)$ is Galois and radical over $k(\zeta)$ (show that if E/k is radical and K/k is any extension, then EK/K is a radical extension), and it has enough roots of 1.

By lemma 18.2.12, $L(\zeta)/k(\zeta)$ is solvable. Then we actually have that $L(\zeta)/k$ is solvable. Indeed, this extension is Galois since it's a composition of galois (exercise 6.13). Applying the fundaemtnal theorem to:

$$k \subseteq k(\zeta) \subseteq L(\zeta)$$

to estalibhed that $\mathrm{Aut}_k(k(\zeta))$ is isomorphic to the quoitnet of $\mathrm{Aut}_k(L(\zeta))$ by its normal subgroup $\mathrm{Aut}_{k(\zeta)}(L(\zeta))$. Both $\mathrm{Aut}_{k\zeta)}(L(\zeta))$ and $\mathrm{Aut}_k(k(\zeta))$ are solvable (being abelian), and it follows that $\mathrm{Aut}_k(L(\zeta))$ is solvale.

Since F is an intermediate field and Galois over k, by the fundaemtnal theorem, $\mathrm{Aut}_k(F)$ is a homomorphic of $\mathrm{Aut}_k(L(\zeta))$. This shows that $\mathrm{Aut}_k(F)$ is a homomorphic image of a solvable group, hence it is itself solvable, and so we're done.

Corollary 18.2.14: Galois Criterion

Let k be a field such that $\mathrm{ch}(k) = 0$ and let $f(x) \in k[x]$ be an irreducible polynomial. Then f(x) is solvable by radicals if and only if its galois group is solvable

Proof **:**
This is an immediate consequence of lemma 18.2.10 and proposition 18.2.13.

Theorem 18.2.15: Abel's Theorem

Let $f(x)$ be an polynomial of degree 5 or higher. Then it admits no solutions by radicals

Proof **:**
By lemma 18.2.5, the Galois group of the general polynomial of degree n is S_n. The group S_n is not solvable for $n \geq 5$ (see ref:HERE), and hence the statement follows from the Galois Criterion.

Furthermore, we know that S_3 and S_4 *are* solvable, with subnormal series:

$$\{e\} \trianglelefteq \mathbb{Z}/2\mathbb{Z} \trianglelefteq D_6 = S_3$$

$$\{e\} \trianglelefteq V_4 \trianglelefteq A_4 \trianglelefteq S_4$$

hence all degree less than or equal to 4 polynomial have a root using radicals.

Exercise 18.2.1

1. Let $\alpha \in F$ for an algebraic extension $F/\mathbb{Q}$. In particular, we may think of F as being embedded in $\mathbb{C}$ so that $\alpha = a + bi$. Does there always exist an N such that $\alpha^N \in \mathbb{R}$?

18.3 Galois Groups of Polynomials

In this section, we shall focus on techniques to compute the galois group of polynomial. So far, much of the focus was on showing what properties can be deduced given properties of the galois group, there have not been many technics to find galois groups (for example, can you come up with a non-solvable degree 5 polynomial?). Hence, this section is a compilation of interesting facts and techniques useful for working with galois groups of poylnomials.

Lemma 18.3.1: Transitivity Of Galois Group

Let $f(x) \in k[x]$ be a separable irreucible polynomial of degree n. Then $\mathrm{Gal}_k(f(x))$ acts transitively on the set of roots of $f(x)$ in $\overline{k}$. In particular, $\mathrm{Gal}_k(f(x))$ may be identified with a *transitive* subgroup of the symmetric group S_n.

Proof **:**
exercise

By corrollary 17.2.15, given α a root of $m_{\alpha,k}(x)$, then the elements of $\mathrm{Gal}_k(m_{\alpha,k}(x))$ are determined by where this root maps. This motivates the following definition:

(In general)

$$\mathrm{Gal}_k(fh) \leqslant \mathrm{Gal}_k(f) \times \mathrm{Gal}_k(h) \leqslant S_{n_1} \times S_{n_2} \leqslant S_n$$

(but more relations can be added: here)

This gives us the limits of the possible subgroups of galois groups:

1. The only possible group for degree 2 irreducible polynomials is: $\mathbb{Z}/2\mathbb{Z}$
2. The only possible groups for degree 3 irreducible polynomials is:
$$\mathbb{Z}/3\mathbb{Z} \cong A_3, S_3$$
3. The only possible groups for degree 4 irreducible polynomials is:
$$S_4, A_4, D_8, \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}, \mathbb{Z}/4\mathbb{Z}$$

Definition 18.3.2: Galois Conjugate

Let $\alpha \in L$ be an algebraic element of L over k. Then the set of roots of the minimal polynomial $m_{\alpha,k}(x)$ is called the *galois conjugates*, *algebraic conjugates*, or *conjugate elements* of α over k.

The motivation comes from the case of $\alpha \in \mathbb{C}$ which are algebraic over $\mathbb{R}$. In this case, the minimal polynomial is either $x - \alpha$ (in which case it has one root), or it is of degree 2 with galois conjugate $\overline{\alpha}$. We thus see that the galois conjugate is the generalization of the complex conjugate.

Definition 18.3.3: Discriminant

Let $f(x) \in k[x]$ be a separable polynomial with roots $\alpha_1, ..., \alpha_n$ in its splitting field. Then the *discriminant* of $f(x)$ is the element $D = \Delta^2$ where
$$\Delta = \prod_{1 \leq i < j \leq n} (\alpha_i - \alpha_j)$$

If you have done some multilinear algebra, this discrimiant is the special case of the generlize the notion of discriminant for general multilinear algebra. We shall come back to the more general notion in alebraic number theory.

Every permutation of the roots fixes D, so D is fixed by the whole Galois group, therefore $D \in k$. Odd permutations move Δ (if $\mathrm{ch}(k) \neq 2$) and even permutations fix it. Tus, Δ is fixed by the Galois group G (i.e. $\Delta \in k$) if and only if $G \subseteq A_n$

Lemma 18.3.4: Discriminant And Galois Group

Let k be a field where $\mathrm{ch}(k) \neq 2$ and let $f(x) \in k[x]$ be a separable polynomial with discriminant D. Then the Galois group o f $f(x)$ is contaiend in the alternating group A_n if and only if D is a quare in k

Example 18.3.1: Discriminant

1. Let $f(x) = x^3 + ax^2 + bx + c$ be a general monic cubic. By shifting by $a/3$, we get
$$f(x - a/3) = x^3 + px + q$$

for appropriate p, q. Then it can be computed that:

$$D = -4p^3 - 27q^2$$

Then by our above lemma, if D is a square then the galois group is S_3, and A_3 otherwise. For example, the galois group of $x^3 - 2$ is S_3 since $D = -108$ is not a square in $\mathbb{Q}$, while $x^3 - 3x + 1$ has galois group $A_3 \cong \mathbb{Z}/3\mathbb{Z}$ since $D = 81 = 9^2$.

2. Let $f(x) \in \mathbb{Q}[x]$ be of degree p for prime p with $p - 2$ real roots and 2 non-real complex roots. Then $\mathrm{Gal}_{\mathbb{Q}}(f(x)) = S_p$. Prove this, then show that the galois group of $x^5 - 5x - 1$ over $\mathbb{Q}$ is S_5.

Other useful discriminants are:

$$\begin{aligned} \mathrm{D}(x^3 + ax + b) &= -4a^3 - 27b^2 \\ \mathrm{D}(x^4 + ax + b) &= -27a^4 + 256b^3 \\ \mathrm{D}(x^4 + ax^2 + b) &= 16b(a^2 - 4b)^2 \end{aligned}$$

(a word on the inverse galois problem? Sovled for every finite abelian group?)

Exercise 18.3.1

1. Let $K/\mathbb{Q}$ be Galois. Show that if $\mathrm{disc}(K)$ is square-free, then K has no intermediate fields

18.4 *Infinite Galois Extensions

In this final section, we shall explore the case of infinite separable normal extensions L/K. This section shall require the notion of topological groups and topological completion, which is covered in appendix C. Most of this material will be important for a course in *Class Field Theory* (see cite:HERE for a reference where this material is used).

In this section, a Galois extension is a normal separable extension, where the finiteness condition is dropped. We may want to keep the galois correspondence between the fixed fields and the subgroups, however there are simply way too many subgroups. Consider

$$\overline{\mathbb{Q}}/\mathbb{Q} \qquad \text{and} \qquad \mathrm{Gal}_{\mathbb{Q}}(\overline{\mathbb{Q}})$$

Then the reader should check that there are only countably many intermediary extension of degree 2 corresponding to the extensions of the form $\mathbb{Q}(\sqrt{\alpha})/\mathbb{Q}$ where α is either a prime p, ip, or a finite compositum of such elements. However, there are uncountably many index two subgroups. To see this, consider $\varphi_S : \mathrm{Gal}_{\mathbb{Q}}(\overline{\mathbb{Q}}) \to \mathbb{Z}/2\mathbb{Z}$ where S is a finite collection of primes where if σ fixes each $\sqrt{p}$. Then $\varphi_S(s) = 0$, and it is 1 otherwise. This is a group-homomorphism, as the composition of two maps that fix each $\sqrt{p}$ for $p \in S$ shall still fix those primes. Furthermore, There are certainly always morphisms τ such that $\varphi_S(\tau) = 1$ for each finite subset S. Thus, each φ_S is a surjective homomorphism, whose kernel is index 2. But then, there are uncountably distinct $\ker(\varphi_S)$. In fact, we don't even need S to be finite (exercise). In this case, we shall even have index two subgroups whose fixed field corresponds to $\mathbb{Q}$!

The reader may feel that there is some simple solutions that can be found, as it seems there are many galois subgroups corresponding to an intermediary. Indeed, we may still recover the galois correspondence if we appropriately limit which subgroups of $\mathrm{Gal}_{\mathbb{Q}}(\overline{\mathbb{Q}})$ we consider (and more generally which subgroups of $\mathrm{Gal}_K(L)$ we consider). This is done by taking the *Krull topology* on the galois group, which we shall naturally derive by first taking the *profinite topology*. The profinite topology shall make us limit our attention to normal subgroups of finite index (namely, such subgroups shall be open in the profinite topology), while in the Krull topology we shall limit our attention to normal subgroups of finite index which also have a [finite] galois group (namely, such subgroups shall be open in the Krull topology). Both of these topologies are achieved by showing that the infinite galois group, as a topological group, is a profinite group. The main idea is that all of the information of infinite galois extension is in fact fully captured by the finite normal extension, due to the fact galois groups represent the relationship of polynomials (where an infinite galois groups stores the information of infinitely many polynomials).

18.4.1 Profinite Groups

In this section, a brief introduction to profinite groups is given. This is not an in-depth coverage, but rather it gives enough information to describe infinite galois theory.

Definition 18.4.1: Profinite Group

Take an inverse system of finite groups with the discrete topology. Then a *profinite group* is a topological group that is the limit of the system. Given any topological group G, we can take its *profinite completion* by taking:

$$\widehat{G} = \varprojlim_{N} G/N$$

where N ranges over all finite index open normal subgroups ordered by containment.

Any group G can be given a topology by given it the profinite topology given by taking the cosets of finite-index normal subgroups as a basis. The reader should check that if G is a topological group and $\widehat{G}$ is its profinite completion, then the continuous homomorphism $\varphi : G \to \widehat{G}$ can be shown that G is dense in $\widehat{G}$.

Example 18.4.1: Profinite Groups

1. Let G be finite. Then all normal subgroups have finite index and there are finitely many. It follows that $G \cong \widehat{G}$.
2. Let $G = \mathbb{Z}$. Then
$$\widehat{Z} = \varprojlim_{n} \mathbb{Z}/n\mathbb{Z} \cong \prod_{p} \mathbb{Z}_p$$
where $\mathbb{Z}_p = \varprojlim_{n} \mathbb{Z}/p^n\mathbb{Z}$ is the p-adic completion of $\mathbb{Z}$.
3. Take $G = \mathbb{Q}$. Then $\mathbb{Q}$ has only one finite-index normal subgroup, namely itself. Hence $\widehat{\mathbb{Q}} = \{1\}$.

The above examples show that the map $\varphi : G \to \widehat{G}$ could be injective, surjective, both. The map will be injective if the intersections of all finite-index open normal subgroups of G is trivial (where

we would then say G is residually finite). The completion comes equipped with a natural universal property where a continuous homomorphism $\varphi : G \to H$ with H a profinite group has a unique continuous homomorphism such that the following diagram commutes:

$$\begin{tikzcd} G \arrow[r, "\varphi"] \arrow[dr, "\varphi"'] & \widehat{G} \arrow[d, "\exists!"] \\ & H \end{tikzcd}$$

A word on strongly complete? MIT notes p.275 Remark 26.15

Theorem 18.4.2: Characterizing Profinite Groups

A topological group is profintie if and only if it is a totally disconnected compact group

Proof **:**
exercise: Compactness comes from Tychnoff's Theorem, while totally disconnected shall come from the discreteness in each finite quotient

Corollary 18.4.3: Distinguishing Profinite Completion

Let G be a profinite group. Then G is naturally isomorphic to its profinite completion, or even stronger:

$$G \cong \varprojlim G/U$$

where U ranges over all open normal subgroups ordered by contain need.
If G is isomorphic to its profinite completion *as a group* (i.e. is strongly complete) if and only if every finite index subgroup is open

Proof **:**
MIT notes p.276

18.4.2 Infinite Galois Theory

Let us now find which topology to put on galois groups. Let us first make sure that the notion of there being a correspondence between subfields and fixed fixed fields of subgroups of galois group is still well-defined

Lemma 18.4.4: Fixed Field Infinite Extension

Let L/K be a galois extension with galois group $G = \mathrm{Gal}_K(L)$. Then if F/K is a normal subextension of L/K, then $\mathrm{Gal}_F(L) = H \trianglelefteq G$ is a normal subgroup of G wit fixed field F. As the subgroup is normal,we have the short exact sequence:

$$1 \to \mathrm{Gal}_F(L) \to \mathrm{Gal}_K(L) \to \mathrm{Gal}_K(F) \to 1$$

which implies:

$$G/H \cong \mathrm{Gal}_K(F)$$

***Proof* :**
As F/K is normal, the map $\sigma \mapsto \sigma|_F$ is a well-defined surjective map (as every $\tau \in \mathrm{Gal}_K(F)$ can be extended to an element $\sigma \in \mathrm{Gal}_K(L)$, see exercise ref:HERE) and has kernel $\mathrm{Gal}_F(L)$ by definition. Then by definition $F \subseteq L^H$, and $L^H \subseteq F$ since if $\alpha \in L^H \setminus F$, then we can construct an element of H that sends α to a distinct root $\alpha' \neq \alpha$ of its minimal polynomials f over F, completing the proof.

Hence, every normal subextension has (at least one) corresponding normal subgroup. However, there are many normal subgroups whose fixed field F is *not* equal to $\mathrm{Gal}_F(L)$, but is contained in $\mathrm{Gal}_F(L)$ The index two example of $\mathrm{Gal}_{\mathbb{Q}}(\overline{\mathbb{Q}})$ is the prototypical example, where there are uncountably many index 2 subgroups $H \leqslant \mathrm{Gal}_{\mathbb{Q}}(\overline{\mathbb{Q}})$ whose fixed field is $\mathbb{Q}$ but naturally H is not the galois subgroup of $\overline{\mathbb{Q}}/\mathbb{Q}$, nor is it the galois group of $\overline{\mathbb{Q}}/K$ for any subextension $K/\mathbb{Q}$. Hence, the following topology limits our attention to the right groups:

Definition 18.4.5: Krull Topology On Galois Group

Let L/K be a Galois (i.e. normal separable) extension with $G = \mathrm{Gal}_K(L)$. Then the *Krull Topology* on G is the basis given by the basis consisting of cosets of the subgroups

$$H_F = \mathrm{Gal}_F(L)$$

where F ranges over finite normal extensions of K in L.

In this topology, every open normal subgroups has finite index, but not every normal subgroup of finite index is open (think $\mathrm{Gal}_{\mathbb{Q}}(\overline{\mathbb{Q}})$ again). Hence, this topology is (strictly) coarser than the porfinite topology, however it is still a profinite group:

Proposition 18.4.6: Galois Group is Profinite Group

Let L/K be a galois extension. Then under the Krull topology, the restriction maps induce a natural isomorphism of topological groups:

$$\mathrm{Gal}_K(L) \cong \varprojlim_F \mathrm{Gal}_K(F)$$

where F ranges over finite galois extension of K in L. Hence, $\mathrm{Gal}_K(L)$ is a profinite group whose open normal subgroups are those of the form $\mathrm{Gal}_F(L)$ for some finite normal extension F/K.

***Proof* :**
Let's first show the map is bijective. For injectivity, for every $\alpha \in L$ it is algebraic over K, and hence liens ins some finite normal subextensions F/K, and so every automorphism in $\mathrm{Gal}_K(L)($ is uniquely determined by all its restrictions to finite normal F/K, showing φ is injective. For surjectivity, given any $(\sigma_F) \in \varprojlim_F \mathrm{Gal}_K(F)$, we may define a corresponding $\sigma \in \mathrm{Gal}_K(L)$ that maps to it by point-wise defining:

$$\sigma(\alpha) := \sigma_F(\vec{\alpha})$$

where F is the normal closure of $K(\alpha)$. This map is well-defined by construction of the inverse limit, giving surjectivity.

For continuity, Letting $G = \mathrm{Gal}_K(L)$ and $H_F = \mathrm{Gal}_F(L)$, we can view φ as the natural continuous map:

$$\varphi : G \to \varprojlim G/H_F$$

Finally, to show the inverse is continuous, we'll show the equivalent of φ being an open map, which translates to showing that φ maps open subgroups $H \subseteq G$ (i.e. the basis) to open subsets in $\varprojlim G/H_F$. Taking $H = \mathrm{Gal}_F(L)$, we get:

$$\varphi(H) = \{(\sigma_E) \mid \sigma_{E_{E\cap F}} = \mathrm{id}_{E\cap F}\} = \pi_F^{-1}(\mathrm{id}\,|_F)$$

as E/K ranges over all finite normal subextension of L/K and π_F is the projection from the limit to $\mathrm{Gal}_K(F)$. Then the singleton set $\{\mathrm{id}\,|_F\}$ is open in the discrete group $\mathrm{Gal}_K(F)$, and so its inverse is open, completing the proof.

Theorem 18.4.7: Fundamental Theorem of Galois Theory, Infinite Extensions

Let L/K be a galois extension with galois group $G = \mathrm{Gal}_K(L)$ endowed with the Krull topology. Then the maps

$$F \mapsto \mathrm{Gal}_F(L) \qquad H \mapsto L^H$$

define inclusion reversing bijections between subextensions F/K of L/K and closed subgroups H of G. Thus, subextensions of finite degree n correspond to subgroups of finite index n, and normal subextensions F/K correspond to normal subgroups $H \leqslant G$ such that as topological groups:

$$\mathrm{Gal}_K(F) \cong G/H$$

Proof **:**
First off, recall that every open subgroup is clopen, as the compliment is the union of non-trivial cosets, and so closed subgroups of finite index are open. Next, the correspondence between finite galois subextensions F/K and finite index closed normal subgroups H is given by proposition 18.4.6, and furthermore we have $[F : K] = [G : H]$ as $G/H \cong \mathrm{Gal}_K(F)$ by lemma 18.4.4.

For the correspondence maps, if F/K is any finite subextension with normal closure E, then $H = \mathrm{Gal}(L/F)$ contains the normal subgroup $N = \mathrm{Gal}(L/E)$ with finite index. The subgroup N is open and therefore closed, thus H is closed since it is a finite union of cosets of N. The fixed field of H is F (by the same argument as in the proof of lemma 18.4.4), thus finite subextensions correspond to closed subgroups of finite index. Conversely, every closed subgroup H of finite index has a fixed field F of finite degree, since the intersection of its conjugates is a normal closed subgroup $N = \mathrm{Gal}(L/E)$ of finite index whose fixed field E contains F and has finite degree. The degrees and indices match because $[G : N] = [G : H][H : N]$ and $[E : K] = [F : K][E : F]$; by the previous argument for finite normal subextensions, $[E : K] = [G : N]$ and $[E : F] = [H : N]$ (for the second equality, replace L/K with L/F and G with H).]

Next, for the closed part, any subextension F/K is the union of its finite subextensions E/K. The intersection of the corresponding closed finite index subgroups $\mathrm{Gal}(L/E)$ is equal to $\mathrm{Gal}(L/F)$, which is therefore closed. Conversely, every closed subgroup H of G is an intersection of basic closed subgroups, all of which have the form $\mathrm{Gal}(L/E)$ for some finite subextension E/K, thus $H = \mathrm{Gal}(L/F)$, where F is the union of the E, completing the proof.

Finally, we see that the isomorphism $\mathrm{Gal}_K(F) \cong G/H$ of topological groups follows from lemma 18.4.4, completing the proof.

Corollary 18.4.8: Closure in Krull Topology

Let L/K be a galois extension and $H \leqslant \mathrm{Gal}_K(L)$ with fixed field F. Then the closure $\overline{H}$ in the Krull Topology is $\mathrm{Gal}_F(L)$:

$$\overline{H} = \mathrm{Gal}_F(H)$$

Proof **:**
First, $H \subseteq \mathrm{Gal}_F(L)$ since by definition $\mathrm{Gal}_F(L)$ contains every $\sigma \in \mathrm{Gal}_K(L)$ that fixes F . Next, $\mathrm{Gal}_F(L)$ is closed and is the group characterized by $L^{\mathrm{Gal}_F(L)} = F$ by theorem 18.4.7. Thus

$$H \subseteq \overline{H} \subseteq \mathrm{Gal}_F(L)$$

Then in the bijection presented in theorem theorem 18.4.7, we have that they are equal, completing the proof.

Let us conclude with a result proven by Waterhouse in 1973 that generalizes theorem 18.2.6:

Theorem 18.4.9: Profinite Group Representable By Galois Group

Let G be a profinite group. Then G is isomorphic to the galois group of some galois extension L/K

***Proof* :**
MIT notes p.279

Just like for finite groups where $K = \mathbb{Q}$ is a difficult (as of writing this book unsolved) problem called the *inverse Galois Problem*, there is little control on what base field K we can choose.

Exercise 18.4.1

1. Find $G = \mathrm{Gal}_{\mathbb{F}_p}(\overline{\mathbb{F}}_p)$ recalling that for each degree n, there is only one extension of $\mathbb{F}_p$. Show that there are too many subgroups to correspond to the fixed fields. Define the Krull topology on G.

19

*More Structure on Fields

The most common fields the reader will encounter are $\mathbb{Q}, \mathbb{R}$, and $\mathbb{C}$. Each of these fields are natural in some way; $\mathbb{Q}$ is initial in $\mathbf{Fld}_0$, $\mathbb{R}$ is the smallest complete ordered field[1], and $\mathbb{C}$ is the smallest complete and algebraically closed field. A key property all three of these fields share is that there is a notion of *size*. For $\mathbb{Q}$ and $\mathbb{R}$, there is even the stronger notion of *ordering* for all the elements. In this chapter, we explore these properties

Recall that an order can be defined on a set (definition 0.1.7). In this section, we shall explore fields with orders, or fields for which these is a function that maps to an ordered field.

This chapter shall also incorporate some topology from chapter C as it is beneficial for a deeper understanding of the subjects presented in this chapter.

19.1 Ordered and Real Fields

Definition 19.1.1: Ordered Field

Let k be a field. Then k is a *totally ordered field* (or *ordered field* for short) if there exists a (total) order $\leq$ on k such that for all $x, y, z \in k$:

1. $x < y$, then $x + z < y + z$
2. If $z > 0$, then if $x < y$, $zx < zy$

If a field k admits a total ordering, then k is said to be a *formally real field*.

Famously, $\mathbb{Q}$ and $\mathbb{R}$ are ordered fields. The reader should prove that $\mathbb{C}$ is *not* an ordered field.

[1] More technically, it is the unique Dedekind-complete ordered field up to isomorphism

Proposition 19.1.2: Properties Of Ordered Fields

Let k be an ordered field and $x \in k$. Then

1. $x > 0$ if and only if $-x < 0$ (similarly, $x < 0$ if and only if $-x > 0$)
2. $x > y$ if and only if $x - y > 0$
3. $x < y$ if and only if $-x > -y$
4. $x > y > 0$, then $y^{-1} > x^{-1} > 0$
5. If $x < y$ and $z < 0$, then $xz = (-x)(-z) > (-y)(-z) = yz$
6. $x^2 > 0$ (hence $1 > 0$ always since $1 = 1^2 > 0$)

Furthermore, ordered fields have characteristic zero

Proof **:**
All of these should be done as an exercise. For the last one, note that

$$0 < 1 < 1 + 1 < 1 + 1 + 1 < \cdots$$

Note how this proposition immediately shows that $\mathbb{C}$ is not ordered since $-1 = i^2 \not> 0$, and that it can be shown that all finite fields are not ordered. In fact, this is a characterizing property of non-ordered fields

Proposition 19.1.3: Artin-Schreier Conditions

Let k be a field. Then k can be an ordered field. Then the following are equivalent:

1. k is an ordered field
2. -1 is *not* the sum of squares
3. 0 is *not* a (non-empty) sum of square

Proof **:**

($1 \Rightarrow 2$) see previous proposition.

($2 \Leftrightarrow 3$) If $-1 = \sum_i^n x_i^2$, then $0 = 1 + \sum_i^n x_i^2 = 1^2 + \sum_i^n x_i^2$. Conversely, if $0 = \sum_i^n x_i^2$ for nonzero x_i, then

$$-x_n^2 = \sum_i^{n-1} x_i^2 \qquad \Leftrightarrow \qquad -1 = \sum_i^{n-1} (\frac{x_i}{x_n})^2$$

Then by proposition 19.1.2, squares of elements cannot be negative, however $-1 < 0$.

($2 \Leftrightarrow 1$) Assume -1 is not a sum of squares. Let S be the set of all elements that are (non-empty) sums of squares. Notably, $0, -1 \notin S$, S is closed under addition, and

since k is commutative S is also closed under multiplication. Furthermore, it is a multiplicative subgroup of $k \setminus \{0\}$).

Such sets are ordered by inclusion and are certainly closed under chains, hence by Zorn's lemma there exists a maximal set $M \subseteq F$ that is closed under addition is a multiplicative subgroup of $k \setminus \{0\}$, which also cannot contain $-1, 0 \in M$. Then $M, \{0\}$, and $-M$ must be disjoint. I claim that

$$k = -M \cup \{0\} \cup M$$

Given this is proven, it is easy to see that k then must be ordered, where $x < y$ iff $y - x \in M$. Suppose $0 \neq a \in k$ such that $-a \notin M$. Define:

$$M' = \{x + ay \mid x, y \in M \cup \{0\}, (x \neq 0) \vee (y \neq y)\}$$

Certainly $M \subseteq M'$ and M' is closed under addition, and is closed under multiplication for if $x, y, z, t \in M \cup \{0\}$, Then:

$$(x + ay)(z + at) = (xz + a^2yt) + a(yz + xt)$$

where $(xz + a^2yt), yz + xt \in M \cup \{0\}$ (note that $a^2 \in S \subseteq M'$). Furthermore, it should be clear that $1 \in M'$ $1 \in S \subseteq M \subseteq M'$) and $0 \notin M'$ (or else $a \in -M$). Finally, if $t \in M'$, then:

$$t^{-1} = \frac{t}{t^2} = \frac{x}{t^2} + a\frac{y}{t^2} \in M'$$

since $t^2 \in M$. Thus, M is a multiplicative subgroup of $k \setminus \{0\}$, and hence by maximality $M = M'$, giving $a \in M$.

;w

Since any ordered field is characteristic zero, it must contain a subfield

$$Q = \left\{ \frac{m1}{n1} \;\middle|\; m, n \in \mathbb{Z} \right\}$$

field-isomorphic to $\mathbb{Q}$. In fact, Q is order-field isomorphic, for $n1 > 0$ in Q when $n > 0$ in $\mathbb{Z}$, and $m1/n1 > 0$ in Q when $m/n > 0$ in $\mathbb{Q}$, and so $a1/b1 > c1 > d$ in Q iff $a/b > c/d$ in $\mathbb{Q}$. Thus, we usually identify Q with $\mathbb{Q}$ as an ordered subfield. A property that we may take for granted that an ordered field may have is the following:

Definition 19.1.4: Archimedean Field

Let k be an ordered field. Then k is an *archimedean field* if every positive element of k is less than a positive integer

Though it may seem like all fields have this property, there do exists field without it, most notably the *hyperreals numbers* which can be thought of as the "order completion" of the reals (think of adding an element ϵ st $0 < \epsilon < x$ for all $x > 0$ and a ω st $0 < x < \omega$ for all $x > 0$). Fortunately, all the ordered fields we shall work with will be Archimedean due to the following classification result:

Theorem 19.1.5: Classification Of Archimedean Fields

Let k be an Archimedean field. Then k is isomorphic to a field:

$$\mathbb{Q} \subseteq k' \subseteq \mathbb{R}$$

Proof **:**
Grillet p.233

Let us focus on formally real fields. If k is formally real, then so is $k(x)$. For an algebraic extension

Proposition 19.1.6: Algebraic Extension Of Formally Real Field

Let k be a formally real field. Then if $\alpha^2 \in k$ such that $\alpha^2 > 0$, $k(\alpha)$ is a formally real field

Proof **:**
exercise

If an algebraic extension of a formally real field is formally real, then it is called a *real extension*.

Proposition 19.1.7: Odd Extension Of Formally Real Field

Let k be a formally real field. Then every finite extension of odd degree is a formally real field.

Proof **:**
Grillet p.234

Using this, we can find the real closure of a formally real field:

Definition 19.1.8: Closed Real Field

Let k be a formally real field. Then k is said to be *closed* if there is no real algebraic extension L/k.

Theorem 19.1.9: Detecting Real Closure

Let k be a formally real field. Then k is real closed if and only if:

1. every positive element of k is a square in k
2. every polynomial of odd degree has a root in k

From these condition, we get that $\overline{k} = k(i)$ where $i^2 = -1$.

Proof **:**
Grillet p.235

Due to this, a real closed field can be made into an ordered field in a unique way, namely $x \leq y$ if and only if $y - x = a^2$ for some $a \in k$. It should also be clear that the field of all algebraic real numbers is certainly a real closed field.

Corollary 19.1.10: Irreducible Elements In Real Closed Field

Let k be a real closed field. Then $f \in k[x]$ is irreducible if and only if f has degree 1 or f has degree 2 with no roots

Proof **:**
exercise

Finally, we have:

Theorem 19.1.11: Real Closure

Every formally real field has a real closure

Proof **:**
Grillet p.236

The reader should verify that any two real closures of k are order-field isomorphic.

Finally, we state without proof the following characterization of real-closed fields that show a naturalness to the choice of $\mathbb{R}$

Theorem 19.1.12: Artin-Schreir Theorem

Let $k \neq \overline{k}$ be a field. Then the following are equivalent:

($\Rightarrow$) k is real closed

($\Rightarrow$) $[\overline{k} : k]$ is finite

($\Rightarrow$) there is an upper bound for the degrees of irreducible polynomials in $k[x]$

Proof **:**
Grillet p.236-7

19.2 Absolute Values on General Fields

The reader [ought to have] showed that $\mathbb{C}$ is not an ordered field. The next best thing is some notion of "size". check this! Intuitively, size is a totally ordered concept, and hence the absolute value shall

be a function that will capture the intuition of size with codomain $\mathbb{R}$:

Definition 19.2.1: Absolute Value

Let k be a field. Then an *absolute value real valuation* on k is a function $v : k \to \mathbb{R}$ whose image is denoted $|x|_v$ such that

1. (positivity) $|x|_v \geq 0$ for all $x \in k$ and $|x|_v = 0$ if and only if $x = 0$
2. (multiplicative) $|xy|_v = |x|_v |y|_v$
3. (triangle inequality) $|x + y|_v \leq |x|_v + |y|_v$

If the absolute value is unambiguous, $|x|_v$ will be denoted $|x|$. If the image of v is a finitely generated abelian group, then $|-|_v$ is called a *discrete absolute value*

Note: The reader should be careful with the definition *real valuation* as there is a different concept called *valuation* that is not a generalization o specialization of the above (see definition 20.6.1).

If the reader is familiar with norms, then this can be thought of as the field-equivalent as norms are defined on vector-spaces. In fact, taking a field to be a vector-space over itself, the existence of a absolute value induces a norm. If the reader is familiar with some Lebesgue space theory, they may recall Hölder's inequality stating that:

$$\|fg\|_1 \leq \|f\|_p \|g\|_q \qquad \frac{1}{p} + \frac{1}{q} = 1$$

which contrasts with the strict equality result given above. Recall that equality happens if and only if $\alpha|f|^p = \beta|g|^q$ for $\alpha, \beta \neq 0$, that is they are "power multiples" of each other. This is certainly the case for fields. Furthermore, there is only one absolute value instead of the 3 norms present in Hölder's inequality. A norm that satisfies the 2nd condition (on an algebra!) is called a *multiplicative norm*, and implies certain geometric invariances in the space.

Example 19.2.1: Absolute Values

1. The usual absolute values on $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ all satisfy the above definition
2. For any field k, there always exists the trivial valuation

$$|x|_t = 1 \qquad x \neq 0$$

   This real valuation shall be an example of a *nonarchimedean* real valuation.

3. Let k be any field and take $k(x)$, the simple transcendental extension of k. Then there are two interesting valuations that may be defined. The first is called the infinity valuation:

$$|f/g|_\infty = \begin{cases} 2^{\deg f - \deg g} & f \neq 0 \\ 0 & f = 0 \end{cases}$$

and the 0 valuation:

$$|f/g|_0 = \begin{cases} 2^{\operatorname{ord} f - \operatorname{ord} g} & f \neq 0 \\ 0 & f = 0 \end{cases}$$

where $\operatorname{ord}(f)$ for a polynomial $f(x) = a_0 + a_1 x_1 + \cdots + a_n x^n$ to be the smallest $k \geq 0$ such that $a_k \neq 0$. The choice of 2 was arbitrary in both and can be replaced with any positive constant.

From the 0-valuation, we may define:

$$\left|\frac{f(x)}{g(x)}\right|_a = \left|\frac{f(x-a)}{g(x-a)}\right|_0$$

These valuations when restricted to k are all trivial.

With positivity and the triangle inequality, the absolute value can define a metric (see chapter C) and hence a metric topology on a field:

$$d(x, y) = |x - y|_v$$

For the p-adic case, the idea behind this metric is that $|x - y|$ is small is $x \equiv y \bmod p^n$ for a large n, that is a large power of p divides the difference.

Proposition 19.2.2: Equivalent Topoloies

Let v, w be two absolutes values on a field k. Then the following are equivalent:

($\Rightarrow$) v, w induce the same topology on k

($\Rightarrow$) Given $x \in k$, $|x|_v < 1$ if and only if $|x|_w < 1$

($\Rightarrow$) There exists a $n \in k$ such that $|x|_w = |x|_v^n$ for all $x \in k$

Proof **:**
exercise

The following can be htougt of as the CRT for real valutions

Proposition 19.2.3: Approximation Theorem

Let $| - |_1, ..., | - |_n$ be pairwise inequivalent valuations for a field K, and let $a_1, ..., a_n \in K$ be some choice elements. Then for every $\epsilon > 0$,there exists a $x \in K$ such that

$$|x - a_i|_i < \epsilon \qquad \forall 1 \leq i \leq n$$

Proof **:**
exercise, use the negation of the above proposition (or Neukrich p.118)

As mentioned earlier, the trivial valuation is non-Archimedean. The following show that non-Archimedean valuations can be thought of as absolute values induced by *ultra-metrics*

Definition 19.2.4: Archimedean Valuation

Let v be a real valuation on a field k. Then if for all $n \in \mathbb{N}$, there exists a $x \in k$ such that

$$|n|_v \leq |x|_v$$

the valuation is called an *Archimedean valuation*. On the other hand, if there exists an x such that for all $n \in \mathbb{N}$:

$$|n|_v \leq |x|_v \qquad \forall n \in \mathbb{N}$$

Then v is called a *non-Archimedean valuation*.

Proposition 19.2.5: Characterizing Non-Archimedean Valuation

Let v be a real valuation on a field k. Then the following are equivalent:

($\Rightarrow$) v is non-Archimedean

($\Rightarrow$) $|n|_v \leq 1$ for all $n \in \mathbb{N}$

($\Rightarrow$) $|n|_v \leq 1$ for all $n \in \mathbb{Z}$

($\Rightarrow$) for all $x, y \in k$,

$$|x+y|_v \leq \max(|x|_v, |y|_v)$$

The final inequality upgrading the triangle inequality is called the *strong triangle inequality* or *ultra-metric inequality*.

Note that if $|x| \neq |y|$ in an ultra-metric , then $|x+y| = \max(|x|, |y|)$!

Proof **:**

(1) $\Rightarrow$ (2) Let v be a non-Archimedean valuation. For the sake of contradiction, suppose that for some $n \in \mathbb{N}$, $|n|_v > 1$. Then by multiplicity we may have an $|n|_v$ with an arbitrarily large valuation:

$$|n^k| = |n|_v^k \xrightarrow{k \to \infty} \infty$$

Which would imply there is no $x \in k$ such that $|n|_v \leq |x|_v$ for all $n \in \mathbb{N}$.

(2) $\Rightarrow$ (3) show that $|-1| = |1|$ and use this to conclude.

(3) $\Rightarrow$ (4) Take:

$$\begin{aligned}
|x+y|^n &= \left| \sum_k^n \binom{n}{k} x^k y^{n-k} \right| \\
&\leq \sum_k^n \left| \binom{n}{k} \right| |x|^k |y|^{n-k} \\
&\leq \sum_k^n |x|^k |y|^{n-k} \\
&\leq (n+1) \max(|x|^n, |y|^n)
\end{aligned}$$

Hence

$$|x+y| \leq (n+1)^{1/n} \max(|x|, |y|)$$

As this is true for all n, and $(n+1)^{1/n} \xrightarrow{n\to\infty} 1$, we get the desired result

(4) $\Rightarrow$ (1) by induction we get:

$$|n| = |1+1+\cdots+1| \leq |1| = 1$$

giving the desired result.

Corollary 19.2.6: Values On different characteristic

Let k be a finite with characteristic $p \neq 0$. Then there does not exist an Archimedean valuation on k.

If k is a finite field, the only valuation is the trivial valuation.

Proof **:**
exercise.

In the case of characteristic 0, given $k = \mathbb{Q}$, all the absolute values may be classified. Define p-valuation to be:

$$v_p(m/n) = \begin{cases} p^{-k} & \frac{m}{n} = p^k \frac{m'}{n'} \neq 0,\ p \nmid m', n' \\ 0 & \frac{m}{n} = 0 \end{cases}$$

Then

Theorem 19.2.7: Classification Of Absolute Values On Rationals

Let $k = \mathbb{Q}$. Then every non-trivial valuation on k is either a p-valuation or the usual absolute value.

This result is due to Ostrowski.

Proof **:**
Let v be a nonarchemidian valuation. Then $|n|_v \leq 1$ for all $n \in \mathbb{Z}$. If $|n|_v = 1$ for all $0 \neq n \in \mathbb{Z}$ then

$$|m/n|_v |n|_v = |m|_v = 1$$

which can be sued to show that $|x|_v = 1$ for all $x \in \mathbb{Q}$. Thus, the following set is non-empty:

$$P = \{x \in \mathbb{Z} \mid |n|_v < 1\}$$

This can be shown to be an prime ideal of $\mathbb{Z}$ (exercise), and since $\mathbb{Z}$ is a PID we have that $P = (p)$ for appropriate prime $p \in \mathbb{Z}$. Let

$$c = |p|_v < 1$$

Now, if

$$\frac{m}{n} = p^k \frac{m'}{n'} \neq 0$$

where $p \nmid m', n'$, then by definition of the prime ideal:

$$|m'|_v = |n'|_v = 1 \qquad \Rightarrow \qquad n\left|\frac{m}{n}\right| = c^k$$

This valuation is not equivalent to v_p. Note that v_p and v_q are not equivalent as that would imply $|p|_v = 1$ (exercise).

The Archimedean is left as a real-analysis exercise, or see Grillet p.242,

In chapters 20, section 20.6, many of these concepts shall be generalized to more general valuations, namely we have been working with *real* valuations, however there are more generally G-valuations for totally ordered groups G.

In an analysis class, the absolute value is used to define the completion of a ring. If k is a field with an absolute value, then the completion $\widehat{k}$ of k is the collection of Cauchy-sequences under the appropriate quotient where k is dense in $\widehat{k}$, and the absolute value n extends naturally to $\widehat{v}$ in $\widehat{k}$.

19.2.1 Extending Valuation

Eventually here

19.2.2 Ostrowski Theorem

Show that if $(K, |-|)$ is a field with an Archimedean absolute value, then

$$\widehat{K} \underset{\sim}{\in} \{\mathbb{R}, \mathbb{C}\}$$

Neukirch p.124

Part V

Commutative Algebra

In this part of the book, we are no longer studying a new object as the central focus. Rather, it is the basic language that deals with *algebraic number theory* and *algebraic geometry*; this is the language of commutative algebra. Commutative algebra is an extension of ring theory where it is always assumed that the ring R is commutative. As shall be shown soon, this theory will prove to be very useful in developing two major studies of mathematics:

1. *Algebraic Geometry*: At the heart of Algebraic Geometry is the realization that all commutative rings have a corresponding *geometric object* (known as an *affine scheme*). We have already encountered the simplest of schemes in section 13.2.4 with the classification of hyper-planes of a finitely dimensional vector-space. Recall that each hyper-plane is associated the solution (or kernel) of linear equations. Schemes generalize this notion a great deal by working with polynomials over general commutative rings instead of restriting to linear equations (which are polynomials of degree 1), and by allowing the resulting geometric object to be "glued together" from kernels of polynomials instead of being the solution to polynomials. In chapter 21, we shall explore how to associate a scheme to every commutative ring and build-up the interesting theory behind this correspondence.

2. *Algebraic Number Theory*: Studying fields allowed us to study the solutions to polynomials over fields. However, fields are very restrictive (dually very versatile) algebraic structures that allowed for some very powerful results. When trying to replicate the same strategies for solving polynomial equations over more general rings, many of these strategies fall apart. To properly generlize this, we shall need to invent a new theory, namely *algebraic number theory* is exactly this theory. In chapter 22, we shall explore some results such as quadratic reciprocity and classifying all Abelian extensions (i.e. Kronnecker-Weber's Theorem). Most of the time, we focus on solutions to equation over $\mathbb{Z}$, as $\mathbb{Z}$ is the initial object of ring, or more concretely $\mathbb{Z}$ is an object that appears very naturally in the world.

In this chapter, some basic topological notions will be used. The geometric intuition presented in this chapter will be lost on those who have not taken or learned some basic topology (up to the separation axioms, in particular T_1 and T_2, and compactness/connectedness). Appendix C serves as a review of what topological concepts will be used in this chapter.

Chapter 21 and 22 can be thought of as covering the more "classical" versions of heir respective field. For example, modern algebraic geometry shall start with schemes, while algebraic number theory is developed over local fields and uses techniques from complex analysis. Most of the motivation for these developments come form more concrete questions that naturally arose when studying polynomials and their solutions. The reader shall certainly gain a lot by understanding the core ideas behind these two chapters before tackling their more abstract modernization's. The reader wishing to follow up these chapters with a more in-depth covering can checkout cite:ALG-GEO for the continuation of this book towards Algebraic Geometry and cite:CFT for a continuation of this book towards Class Field Theory.

Note that algebraic geometry is *not* algebraic topology. In algebraic topology, we usually start with a topological space (or manifold) and associate to it an algebraic object to find properties of the topological/geometrical object. In Algebraic Geometry, we start with a commutative ring (or a collection of commutative rings), and create a geometric object that can be used to study the ring.

20

More on Ring Theory, Ring Extensions, and Category Theory

We take a moment to revisit some ring theory and build up more ring theoretical tools. We will focus mainly on the extension of commutative rings, and the relation between ideals between two rings connected by a ring homomorphism.

As we progress in the rest of the book, more complicated constructions shall be coming to light. Thus, a few more categorical notions shall be introduced to organize these new ideas cohesively.

20.1 Categorical (co)Limits, Natural Transformations, and Adjoints

In this section, it will be taken that chapter 8 has been read, the rigorous definition of category, functor, and commutative diagram are all well understood, and simple examples of all of these concepts come to mind. Recall that a functor could be covariant or contravariant. To demonstrate, if we have a functor $F : C \to D$ between two categories and a commutative diagram in the category C as follows:

$$\begin{tikzcd}
b \arrow[r] & c \arrow[r] & f \\
a \arrow[u] \arrow[r] & d \arrow[u] \arrow[ur] \arrow[r] & e \arrow[u] \arrow[ul]
\end{tikzcd}$$

Then if F is covariant, the above commutave diagram gets sent to:

$$\begin{tikzcd}
F(b) \arrow[r] & F(c) \arrow[r] & F(f) \\
F(a) \arrow[u] \arrow[r] \arrow[ur] & F(d) \arrow[u] \arrow[r] \arrow[ur] & F(e) \arrow[u]
\end{tikzcd}$$

and if F is contravariant, the comutative diagram gets sent to:

$$\begin{tikzcd}
F(b) \arrow[d] & F(c) \arrow[l] \arrow[d] & F(f) \arrow[l] \arrow[d] \\
F(a) & F(d) \arrow[l] \arrow[ul] & F(e) \arrow[l] \arrow[ul]
\end{tikzcd}$$

With the futher material covered, a couple more interesting functors may be introduced which may enlighten the reader with some further intuitions on functors:

Example 20.1.1: Further Examples Of Functors

1. Ring and Modules both have natural forgetful functors to either **Set** or **Ab**. The adjoint of the forgetful functor **Ring** $\to$ **Ab** maps abelian groups to group-rings.

2. Let R be a ring and R^* be its group of units. Then any ring homomorhpism $R \to S$ can be induces a group homomorphism $R^* \to S^*$ (namely since ring homomorphisms sends units to units).

3. Let S be any set. Take $\hat{S}$ to be the category whose objects are subsets of S and morphisms are inclusions $\subseteq$. Then a contravriant functor $\hat{S} \to$ **Set** is called a *presheaf*. A typical example would assotiate to each $U \subseteq S$ the set of all function $\mathrm{Hom}_C(S, X)$ for some X (popular examples are $\mathbb{R}, \mathbb{C}, \mathbb{Z}, S^1$, though there are many others) and locally small category C^a, and each $U \subseteq V$ is associated to a restriction map $\mathrm{Hom}_C(V, X) \to \mathrm{Hom}_C(U, X)$, $\varphi \mapsto \varphi|_U$. These types of functors are very important and shall be further explore in algebraic geometry.

4. As a particular example of the above, let C be any locally small category and $X \in C$. Then:

$$A \mapsto \mathrm{Hom}_C(X, A) \qquad A \mapsto \mathrm{Hom}_C(A, X)$$

define a covariant and contravariant functor $C \to$ **Set** respectively; these functors are usually dentoed $\mathrm{Hom}_C(-, A)$ and $\mathrm{Hom}_C(A, -)$. To see where the morphisms of C get mapped to, it's best to see the following diagrams:

$$\begin{tikzcd}
A \arrow[r, "\alpha"] \arrow[dr, "f_a"'] & B \arrow[r, "\beta"] \arrow[d, dashed, "f_b"] & C \arrow[dl, "f_c"] \\
& X &
\end{tikzcd}
\qquad
\begin{tikzcd}
& X \arrow[dl, "f_a"'] \arrow[d, dashed, "f_b"] \arrow[dr, "f_c"] & \\
A \arrow[r, "\alpha"] & B \arrow[r, "\beta"] & C
\end{tikzcd}$$

Taking $\mathrm{Hom}_C(-, X)$, any morhpism $\alpha : A \to B$ gets maped to $\mathrm{Hom}_C(B, X) \to \mathrm{Hom}_C(A, X)$ where $f_b \mapsto f_a = (f_b \circ \alpha)$. The functors $\mathrm{Hom}_C(-, X)$ and $\mathrm{Hom}_C(X, -)$ are often abbreviated to h_X and h^X respectively.

5. Recall that if $C = R\text{-}\mathbf{Mod}$, then $\mathrm{Hom}_R(A, B)$ is an abelian group. If a functor F preserves this structure, then it is called *additive*. In this book, F would usually be a funtor between cateogires of modules. For example, $F : R\text{-}\mathbf{Mod} \to S - \mathbf{Mod}$ is an additive functor then:

$$\mathrm{Hom}_R(A, B) \to \mathrm{Hom}_S(F(A), F(B))$$

is a group homomorphism. for all $A, B \in \mathrm{Obj}(R\text{-}\mathbf{Mod})$. As an example: $S \otimes_R (-) : R\text{-}\mathbf{Mod} \to S - \mathbf{Mod}$ is an additive functor mapping $M \mapsto S \otimes_R M$ and $f : M \to N$ to $\iota \otimes f : S \otimes_R M \to S \otimes_R N$.

[a] recall that a locally small category is one where $\mathrm{Hom}_C(A, B)$ has at most a set amount of elements

Natural Transformations and Equivalence of Categories

Continuinig from chapter 8, it is important to establish when two categories are equivalent, that is a notion of "isomorphism of categories". If $F : C \to D$ were an isomorphism, Then there would exists a G such that $F \circ G = G \circ F = \mathrm{id}$. However, in general this definition is too restrictive to be of much use. This comes down to the fact that most objects in category theory are defined *up to isomorphism*[1]. In most cases, what we really need is that $F \circ G$ and $G \circ F$ are both "isomorphic" to id. In order to talk about functors being isomorphic, we must talk about morphisms between functors. This might seem like rather abstract concept. However, they have a very practical purpose: they represent when two "interpretations" provided by a functor are comparible *without* any supplementary information. This was briefly commented upon this notion in section 13.3 when discussing how when V is finite dimensional, $V \cong V^*$ and $V \cong V^{**}$. The isomorphism $V \cong V^*$ required a choice of basis, while $V \cong V^{**}$ can be done without the need of any such choice, hence making the later more *natural*. What this translates to is the function $(-)^* : \mathbf{Vect}_{\mathrm{fin}} \to \mathbf{Vect}_{\mathrm{fin}}$ is *not* naturally isomorphic to $\mathrm{id} : \mathbf{Vect}_{\mathrm{fin}} \to \mathbf{Vect}_{\mathrm{fin}}$, while $(-)^{**} : \mathbf{Vect}_{\mathrm{fin}} \to \mathbf{Vect}_{\mathrm{fin}}$ *is* naturally isomorphic to id. The following definition captures this idea:

Definition 20.1.1: Natural Transformation

Let $F : C \to D$ and $G : C \to D$ be two funtors between the same categories. Then a *natural transformation* from F to G, denoted $\alpha : F \Rightarrow D$[a] consists of a morphism $\alpha_c \in \mathrm{Mor}(\mathbf{D})$ for each $c \in \mathrm{Obj}(C)$, where $\alpha_c : F(c) \to G(c)$ and that for any morphism $f : c \to c' \in \mathrm{Mor}(\mathbf{C})$, the following diagram commutes.

$$\begin{tikzcd} F(c) \arrow[r, "\alpha_c"] \arrow[d, "F(f)"'] & G(c) \arrow[d, "G(f)"] \\ F(c') \arrow[r, "\alpha_{c'}"'] & G(c') \end{tikzcd}$$

A *natural isomorphism* is a natural transformation α such that each α_c is an isomorphism, in which case we often write $\alpha : F \cong G$.

[a] the notation $\Rightarrow$ instead of $\to$ is due to historical precedence; it has no relation to the logical notion of implication

[1] There will be examples of isomorphism of categories, famously the Galois Correspondence can be made into a category isomorphism between the right categories

(couple of interesting examples that are currently in the category chapter of the book.)

Natural transformation shall be covered in more depth in part VIII. At the moment, they are useful to define the following weaker definition of equivalence of categories

Definition 20.1.2: Equivalence Of Categories

Let C and D be two categories. Then C and D are said to be *equivalent* if there exists functors $F : C \to D$ and $G : D \to C$ such that $F \circ G \cong 1_C$ and $G \circ F \cong 1_D$, that is $F \circ G$ and $G \circ F$ are both naturally isomorphism to the identity functor. If C and D are naturally equivalent, then we shall write $\mathbb{C} \cong D$

It should be verified that equivalence of categories defines an equivalence relation. This definition is nice theoretically, but for practical purposes, the following equivalent definition is what will be used instead

Proposition 20.1.3: Equivalence Of Categories

Let C and D be categories. Then $C \cong D$ if and only if there exists a functor F such that

1. F is a *full* functors, that is for each $x, y \in C$, $\mathrm{Hom}_C(x, y) \to \mathrm{Hom}_D(F(x), F(y))$ is *surjective*
2. F is a *faithful* functors, that is for each $x, y \in C$, $\mathrm{Hom}_C(x, y) \to \mathrm{Hom}_D(F(x), F(y))$ is *injective*
3. F is an *essentialy surjective* functors, that is for each $y \in D$, there exists a $c \in C$ such that $F(c) \cong d$

Proof **:**
see ref:HERE in part VIII.

Example 20.1.2: Equivalence Of Categories
Let $\mathbf{Mat}_k$ be the categories whose objects are the natural numbers $\mathbb{N}$ an whose morphisms between any two numbers, $\mathrm{Hom}(m, n)$, is the set of $n \times m$ matrices with entries in the field k, where composition is defined as the product of matrices. Then $\mathbf{Mat}_k \cong \mathbf{Vect}_k^{\mathrm{fin}}$, namely each n gets mapped to k^n with the standard basis for k^n, and each matrix gets mapped to the linear transformation corresponding to the matrix given the standard basis. It is easy to see that this functor is fully faithful and essentially surjective, making these categories equivalent.

Limits and Colimits

The next important concepts is that of *limits and colimits*. If the reader knows some analysis, they might suspect there to be some similarities between these concepts, and abstractely there are, as shall be demonstrated. On a high-level, a limit and colimit is always taken given a commutative diagram, say for simplicity we have the diagram:

$$a \to b \to c$$

Focusing for a moment on limits, we may imagine that there are multiple possible morphism from an object, say x, to objects a, b, c such that the following diagram commutes:

$$\begin{tikzcd} & x \arrow[dl] \arrow[d] \arrow[dr] & \\ a \arrow[r] & b \arrow[r] & c \end{tikzcd}$$

For example, if a, b, c where natural numbers and the morphism $\alpha \to \beta$ exists if and only if $\alpha \big| \beta$, then x would have to divide a, b, c, then we may have the following commutative diagrams:

$$\begin{tikzcd} & 10 \arrow[dl] \arrow[d] \arrow[dr] & \\ 30 \arrow[r] & 90 \arrow[r] & 600 \end{tikzcd} \qquad \begin{tikzcd} & 2 \arrow[dl] \arrow[d] \arrow[dr] & \\ 30 \arrow[r] & 90 \arrow[r] & 600 \end{tikzcd} \qquad \begin{tikzcd} & 5 \arrow[dl] \arrow[d] \arrow[dr] & \\ 30 \arrow[r] & 90 \arrow[r] & 600 \end{tikzcd}$$

Then the limit of this diagram, say t is the "terminal" object of the above, that is for any x such that the diagram commutes:

$$\begin{tikzcd} & x \arrow[d] \arrow[ddl, bend right] \arrow[ddr, bend left] \arrow[ddrr, bend left] & \\ & t \arrow[dl] \arrow[d] \arrow[dr] & \\ a \arrow[r] & b \arrow[r] & c \end{tikzcd}$$

For example, in the above diagram, 30 is the terminal object, since: for any other number x such that $x \to 30$, $x \to 90$, and $x \to 600$ (that is $x \big| 30$, $x \big| 90$, and $x \big| 600$), then necessarily $x \to 30$, that is $x \big| 30$,

$$\begin{tikzcd} & x \arrow[d] \arrow[ddl, bend right] \arrow[ddr, bend left] \arrow[ddrr, bend left] & \\ & 30 \arrow[dl] \arrow[d] \arrow[dr] & \\ 30 \arrow[r] & 90 \arrow[r] & 600 \end{tikzcd}$$

Similar to how a limit in calculus and analysis is in some way the "end" of a sequence, the limit of a diagram is the "end" of a diagram. Dually, a *colimit* reverses the process. Just like limits, any colimit start with a given commutative diagram, say the same commmutative diagram that was used above:

$$a \to b \to c$$

Then given the above commutative diagram, there may be many morphisms from a, b, c to an object, say x, the following diagram always commutes:

$$\begin{tikzcd} a \arrow[r] \arrow[dr] & b \arrow[r] \arrow[d] & c \arrow[dl] \\ & x & \end{tikzcd}$$

For example, if a, b, c where sets and a morphism $\alpha \to \beta$ exists if and only if $\alpha \subseteq \beta$, then we may have the following commutative diagrams:

$$\begin{tikzcd}
\mathbb{N} \arrow[r] \arrow[dr] & \mathbb{N} \cup \{1/2\} \arrow[r] \arrow[d] & \mathbb{Q} \arrow[dl] \\
& \mathbb{Q} \cup \{\pi\} &
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbb{N} \arrow[r] \arrow[dr] & \mathbb{N} \cup \{1/2\} \arrow[r] \arrow[d] & \mathbb{Q} \arrow[dl] \\
& \mathbb{Q} \cup \{e\} &
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbb{N} \arrow[r] \arrow[dr] & \mathbb{N} \cup \{1/2\} \arrow[r] \arrow[d] & \mathbb{Q} \arrow[dl] \\
& \mathbb{R} &
\end{tikzcd}$$

then the colimit of this diagram, say n, is the "initial" object of the above, that is for any x, the following diagram always commutes:

$$\begin{tikzcd}
a \arrow[r] \arrow[dr] \arrow[ddr, bend right] & b \arrow[r] \arrow[d] \arrow[ddl, bend left] & c \arrow[dl] \arrow[ddl, bend right] \\
& n \arrow[d] & \\
& x &
\end{tikzcd}$$

For example, in the above diagram the colimit is $\mathbb{Q}$. To generlize this notion, the notion of commutative diagram must be abstracted:

Definition 20.1.4: Commutative Diagram

Let C be a category. Then a *diagram* in C is a function $F : I \to C$ whose domain is an *indexing category* which is a small category.

An indexing category is an arbitrary small category which consists of arbitrary objects and morphisms. The category is usually finite or countable and has very few morphisms. Omiting the identity morphisms, an indexing category may look like any of the following when all the objects and morphisms are drawn out as a directed graph:

1. $$\bullet \longrightarrow \bullet \longrightarrow \bullet$$

2. $$\bullet \qquad \bullet$$

3. $$\begin{tikzcd}
\bullet \arrow[d] & \bullet \arrow[l] \arrow[dl] \\
\bullet \arrow[r] \arrow[ur] & \bullet \arrow[u]
\end{tikzcd}$$

4. $$\bullet \longrightarrow \bullet \longrightarrow \bullet \longrightarrow \cdots$$

Definition 20.1.5: [Categorical] Limits

Let $\mathbf{C}$ be a category, $\mathbf{I}$ be an indexing category, and $F : \mathbf{I} \to \mathbf{C}$ a covariant functor. Then the *limit* of F (if it exists) is an object $L \in \mathbf{C}$ endowed with morhpisms $\varphi_I : L \to F(I)$ for all object $I \in \mathbf{I}$ such that

1. If $\alpha : I \to J$ is a morphism in $\mathbf{I}$, hen $\varphi_J = F(\alpha) \circ \varphi_I$, that is:

$$\begin{tikzcd} L \arrow[d, "\varphi_I"] \arrow[dr, "\varphi_J"] & \\ F(I) \arrow[r, "F(\alpha)"] & F(J) \end{tikzcd}$$

2. L is final with respect to this property that is, if M is anoher object enodwed with morphsism ψ_I that also satisfies the previous requirement, then there exists a unique morphism $M \to L$ making the relavnat diagrams commute

the limit is usually denoted $\varprojlim F$. A limit is sometimes also called an *inverse limit* or *projective limit*.

dually:

Definition 20.1.6: [Categorical] Colimit

Let $\mathbf{C}$ be a category, $\mathbf{I}$ be an indexing category, and $F : \mathbf{I} \to \mathbf{C}$ a covariant functor. Then the *limit* of F (if it exists) is an object $L \in \mathbf{C}$ endowed with morhpisms $\varphi_I : L \to F(I)$ for all object $I \in \mathbf{I}$ such that

1. If $\alpha : J \to I$ is a morphism in $\mathbf{I}$, hen $\varphi_I = \varphi_J \circ F(\alpha)$, that is:

$$\begin{tikzcd} L \arrow[d, "\varphi_I"] \arrow[dr, "\varphi_J"] & \\ F(I) \arrow[r, "F(\alpha)"] & F(J) \end{tikzcd}$$

2. L is initial with respect to this property that is, if M is anoher object enodwed with morphsism ψ_I that also satisfies the previous requirement, then there exists a unique morphism $L \to M$ making the relavnat diagrams commute

the limit is usually denoted $\varinjlim F$. A colimit is sometimes also called a *direct limit* or an *injective limit*.

Example 20.1.3: Limits and Colimits

It should be stated immediately that all universal properties are examples of limits and colimits, and the reader should take these for their basic understanding of (co)limits. We shall take most of our examples in **CRing** and R-**Mod** for commutative ring R, as it is the category most relevant to the following chapters and in algebraic geometry.

1. The simplest diagram to take the limit and colimit of is the diagram with a single dot:

$$\bullet$$

In **CRing**, the Limit and Colimit will just return the ring itself

$$\bullet \qquad \bullet$$

The limit of this diagram is called the *product*, and the colimit the *coproduct*. In **CRing**, the product and coproduct match, that is:

$$\begin{tikzcd} & R \times S \arrow[dl, "\pi_R"'] \arrow[dr, "\pi_S"] & \\ R & & S \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} R \arrow[dr, "\iota_R"'] & & S \arrow[dl, "\iota_S"] \\ & R \times S & \end{tikzcd}$$

Are the limit and colimits in **CRing**. In **Ring**, the colimit of a diagram would be the free product.

2. Let us now take a diagram with two arrows:

$$\bullet \rightrightarrows \bullet$$

The two arrows usually represent that we want certain values from the domain of the arrow to be equal in the codomain. This diagram is called the *equalizer*. We would be a bit more careful with which category to take the limit, as we usually want the limit to be a single arrow $K \to X \rightrightarrows Y$ to be the universal object. The quintessential example of a limit of such a diagram is with one of the arrows being the 0 morphism (when it exists in the category):

$$R \underset{f}{\overset{0}{\rightrightarrows}} S$$

In **CRing**, the object that would satisfy this arrow would have to commute in the diagram, namely:

$$f(R) = 0(R) = 0$$

Namely, the *kernel*, and $\ker(f) \xrightarrow{\iota} R \underset{0}{\overset{f}{\rightrightarrows}} S$.

3. We may wonder which limits may exist in a category. One famous result called the *Existence Theorem for Limits* is that all limits exist in a category (it is *complete*) iff all equalizers and (small) products exists (similarly for all colimits existing with equalizers and coproducts, this makes the category *cocomplete*). A small product means that it is closed under a "set-size" of indexing, something that is not an issue for our current investigation as we shall always have a set-size for our indexing, usually finite countable, or $|\mathbb{R}|$. The category **CRing** and R-**Mod** both have all [small] (co)products and (co)euqalizers, and hence they are both (co)complete.

   A category that is not complete but not cocomplete is the category of compact Hausdorff spaces (or by Gelfand Duality[a], unital commutative C*-algebras) which is closed under countable product (by Tychnoff's theorem and some more results from functional analysis[b]) and closed under equlizers, but is closed under coproducts as this can result in a discrete topology on $\mathbb{Z}$. The category of finite groups is neither complete nor cocomplete. f

4. Consider the indexing category I which may be represented as the following diagram:

$$\cdots \to 4 \to 3 \to 2 \to 1$$

Then $F : \mathbf{I} \to \mathbf{C}$ representes a coiceof objects A_i an morphism $\varphi_{ji} : A_i \to A_j$ such that $\varphi_{ii} = 1_{A_i}$ and $\varphi_{kj} \circ \varphi ji = \varphi_{ki}$ for all $i \geq j \geq k$ giving us the diagrm:

$$\cdots \xrightarrow{\varphi_{45}} A_4 \xrightarrow{\varphi_{34}} A_3 \xrightarrow{\varphi_{23}} A_2 \xrightarrow{\varphi_{12}} A_1 \tag{20.1}$$

Then $\varprojlim F$ is an object Aendowed with morphisms $\varphi_i : A \to A_i$ such that the following diagram commutes:

$$\begin{tikzcd}
 & & & A \arrow[dll, dashed, "\cdots"'] \arrow[dl, "\varphi_4"] \arrow[d, "\varphi_3"] \arrow[dr, "\varphi_2"] \arrow[drr, "\varphi_1"] & & \\
\cdots \arrow[r, "\varphi_{45}"] & A_4 \arrow[r, "\varphi_{34}"] & A_3 \arrow[r, "\varphi_{23}"] & A_2 \arrow[r, "\varphi_{12}"] & A_1
\end{tikzcd}$$

and A is terminal, that is any object object that satisfies the above diagram factors through A. Often, such a limit is dentoed $\varprojlim_i A_i$. Such a limit need not always exist, however in $R\mathbf{Mod}$, we have that they always exist. Indeed, say we are given a diagram as in equation (20.1). Take $\prod_i A_i$, which we may interpret as a set of sequences. Take $A \subseteq \prod_i A_i$ to be all sequence $(a_i)_{i>0}$ which are *coherent*, that is for all $i > 0$, take $a_i = \varphi_{i,i+1}(a_{i+1})$. The reader may verify that A is indeed a R-submodule of $\prod_i A_i$, and that the cannonical projection $\pi_i : A \to A_i$ will make the necessary diagram commute. Then it is a relatively simple check to see that $A = \lim_i A_i$. a very similar construction occurs for colimits, in particular with the indexing category:

$$1 \to 2 \to 3 \to \cdots$$

5. An important generalization of the product and coproduct is the *fibered product* (or pullback) and *fibered coproduct* (or pushout). The fibered product is given by the diagram:

$$\begin{tikzcd}
\bullet \arrow[dr] & & \bullet \arrow[dl] \\
 & \bullet &
\end{tikzcd}$$

The intuition is that the left and right dot are somehow glued at the same place for the bottom product. In $\mathbf{CRing}$ and R-$\mathbf{Mod}$, the fibered product of:

$$\begin{tikzcd}
R \arrow[dr, "f"'] & & S \arrow[dl, "g"] \\
 & Z &
\end{tikzcd}$$

is:

$$R \times_Z S = \{(r, s) \in R \times S \mid f(r) = g(s)\}$$

This is also called the pullback as it fits nicely into the diagram:

$$\begin{tikzcd}
 & R \times_Z S \arrow[dl, dashed] \arrow[dr, dashed] & \\
R \arrow[dr, "f"'] & & S \arrow[dl, "g"] \\
 & Z &
\end{tikzcd}$$

Notice that the intersection can be written as a fibered product, namely if $S, S' \subseteq R$, then:

$$\begin{tikzcd}
 & S \cap S' \arrow[dl, dashed] \arrow[dr, dashed] & \\
S \arrow[dr, hook, "\iota"'] & & S' \arrow[dl, hook', "\iota"] \\
 & R &
\end{tikzcd}$$

The fibered coproduct is the colimit of the codiagram:

$$\begin{tikzcd}
 & \bullet \arrow[dl] \arrow[dr] & \\
\bullet & & \bullet
\end{tikzcd}$$

Examples of coproducts come from tensor products over a [commutative] ring R:

$$\begin{tikzcd}
 & R \arrow[dl] \arrow[dr] & \\
M \arrow[dr, dashed] & & N \arrow[dl, dashed] \\
 & M \otimes_R N &
\end{tikzcd}$$

In **CRing**, we can consider R as $\mathbb{Z}$-modules.

[a] see eyntka cite:funcAnalysis
[b] see cite:HERE

Adjoint Functors

The final categorical that will be introduced is a special type of of equivalence of functors that is a little moer general than natural transformations:

Definition 20.1.7: Adjoint Functors

Let $F : \mathbf{C} \to \mathbf{D}$ and $G : \mathbf{D} \to \mathbf{C}$ be two functors. Then these functors are *adjoint*, in particular F is *left-adjoint* to G and G is *right-adjoint* to F) if there is a natural isomorphism:

$$\mathrm{Hom}_C(X, G(Y)) \xrightarrow{\sim} \mathrm{Hom}_D(F(X), Y)$$

for all objects $X \in C$ and $Y \in D$.

If you are more proficient with category theory, we would say there is a natural isomorphism of *bifunctors* $C \times D \to \mathbf{Set} : \mathrm{Hom}_C(-, G(-)) \overset{\sim}{\Rightarrow} \mathrm{Hom}_D(F(-), -)$. A functor F is called a *left-adjoint* functor (resp. *right-adjoint* functor) if there exists a corresponding right-adjoint functor (resp. left-adjoint functor).

Example 20.1.4: Adjoints

1. (The free-forgetful adjoint)
2. Any equivalence of categories naturally comes with a functor with an adjoint pair.
3. (adjoining an identity to a rng. Similarly wth adjoining identity to semi-groups, or building up from monoids to groups to rings)
4. (scalar extension is adjoint with respect to $\varphi : R \to S$ is left adjoint to the forgetful functor $G : S\text{-}\mathbf{Mod} \to R\text{-}\mathbf{Mod}$)
5. (Field of fractions has an adjoint. In particular the category $\mathbf{Dom}_m$ whose objects are integral domains and morphism are injective ring homomorphisms. Then the forgetful functor $\mathbf{Fld} \to \mathbf{Dom}_m$ is right-adjoint to the functor $\mathbf{Dom}_m \to \mathbf{Fld}$ that assigns each element of the domain its field of fraction)
6. (The inclusion map $\mathbf{Ab} \to \mathbf{Grp}$ has a left-adjoint given by the abelianization: $G^{\text{ab}} = G/[G, G]$)
7. If you know topology, let $F : \mathbf{Top} \to \mathbf{Set}$ be the usual forgetful functor. This functor has a left and right adjoint: the left-adjoints gives the discrete topology to each set, while the right adjoint give the triival toplogy to each set.
8. Another famous topological examples is the Stone-Cech compactification theorem, where the inclusion functo $F : \mathbf{CHaus} \to \mathbf{Top}$ of compact hausdorff spaces into topological spaces has a left adjoint known mapping topological spaces to their Stone-Cech compactification.
9. An important adjoints that will appear in two major theorems later in the book[a] comes when conidering posets as categories (the objects are the elements and the morphisms are the relations, x has a morphism to y if and only if $x \leq y$). Then a pair of adjoint functors between two posets (interpreted as categories) is called a *galois connection*. If the functor is bijective and contravariant, then this is usually called a *galois correspondence*.

More adjoints shall apear later in each following chapter[b]

[a] Galois Correspondence and Nullstellensatz
[b] The hom-tensor adjoint and the Frobenius Reciprocity being two famous examples

Lemma 20.1.8: Continuous And Cocontinuous Functor

Let $F : C \to D$ be a right-adjoint functor. and $: I \to C$ represent a diagram. Then:

$$F(\varprojlim A) \xrightarrow{\sim} \varprojlim(F \circ A)$$

if the limit exists. Similarly, if $G : C \to D$ is a right-adjoint functor

$$G(\varinjlim A) \xrightarrow{\sim} \varinjlim(G \circ A)$$

Analogously to calculus, when a function commutes with a limit, it is called *continuous*. Hence right-adjoint functors are continuous functors. Similarly, left-adjoint functors are *cocontinuous*.

Proof **:**
see ref:HERE

As a consequence, rightad-joint functors preserve producects and "preserve kernsl" when kernels make sense.

20.2 Noetherian Rings and Hilbert's Basis Theorem (HBT)

As with groups an modules, all R-algebras are quotient of a polynomial ring over many variables. If the polynomial ring is over finitely many variables. The fact that all algebras are quotient of polynomial rings will let the study of geometry and algebra.

Next, we recall some important definition and results that shall be central to the following section:

Definition 20.2.1: Noetherian Ring

Let R be a commutative ring. Then if R satisfies the ascending chain condition on Ideals, then R is Noetherian. In particular, if $\mathcal{I}$ is a collection of ideals ordered by inclusion $(I_1 \subseteq I_2 \subseteq \cdots I_k \cdots)$, then there exists an m such that for all $n \geq m$, $I_m = I_{m+1} = \cdots$.

Proposition 20.2.2: Extension of Noetherian Rings

Let R be a Noetherian ring. Then R/I is Noetherian for any ideal I. Conversely, if I and R/I is Noetherian, then R is Noetherian.

Proof **:**
This is proposition ref:HERE

As a consequence the image of any Noetherian ring is Noetherian.

Proposition 20.2.3: Equivalence of Noetherian Ring

Let R be a Noetherian ring. Then the folloing are equivalent

1. R is Noetherian
2. every nonempty collection of ideals of R has a maximal element under inclusion
3. every ideal of R is finitely generated

Proof **:**
this is proposition ref:HERE

Example 20.2.1: Exercise
Show that if R is a Noetherian ring and M is a finitely generated R-module, then M is Noetherian (Hint: remember the universal property of free-modules)

We will soon encounter rings that satisfy the descending chain condition (commonly abbreviated to D.C.C). Such rings are called *Artinian rings*. For the following result, let us upgrade definition 3.7.3 for modules:

Definition 20.2.4: Chain of Modules and Composition Series

Let M be a module. Then a *chain* of submodule of M is a sequence (M_i) of submodules st

$$M = M_0 \supsetneq M_1 \supsetneq \cdots \supsetneq M_n = 0$$

where the length of the chain is the number of $\supsetneq$. A maximal chain is called a *composition series* of M (equivalently, each M_{i-1}/M_i is *simple*. IF the chain is of infinite length, we shall say that M does not have a composition series.

Proposition 20.2.5: Finite Length, then Length is Invriant

Let M have a composition series of length n. Then every composition series of M has length N, and every chain in M can be extended to a composition series.

Proof **:**
Let $\ell(M)$ denote the least length of composition series of a module M To get the desired result, let us prove a few results:

1. $N \subseteq M \Rightarrow \ell(N) \leq \ell(M)$. Let (M_i) be a compositum series of M of minimum length, and take $N_i = N \cap M_i$. Then $N_{i-1}/N_i \subseteq M_{i-1}/M_i$, which as M_{i-1}/M_i is simple conclude the result

2. each $M_i \subsetneq M$ have length $\leq \ell(M)$. To see this, we may make a new chain with M_0 as the starting point:
$$M = M_0 \supsetneq M_1 \supsetneq \cdots \supsetneq M_k = 0$$
which by part 1 give us:
$$\ell(M) = \ell(M_0) \geq \ell(M_1) \geq \cdots \geq \ell(M_k) = 0$$

3. Take any composition series of M. Then if it has length k, it must be that $k \leq \ell(M)$ by part (2). Then by definition, $\ell(M)$, $k = \ell(M)$. Hence, all composition series have the same length.

Finally, consider any chain for M. If its length is $\ell(M)$, by (2) it must be a composition series, if its length is less than $\ell(M)$¡ it cannot be a composition series, and hence cannot be maximal. There does exists a lengthening by a new term, which as the number of possible terms that can be added is of finite length we may continue until we have a series of length $\ell(M)$, completing the proof.

Hilbert Basis Theorem

From theorem 11.1.1 that if F is a field, then $F[x]$ is a Euclidean Domain. By corollary 11.2.4, we showed that $F[x_1, x_2, ...]$ is a Unique Factorization Domain. However, it is not necessarily the fact that they are anything stronger than Unique Factorization Domains (with the counter-example being example 11.2.1 where $\mathbb{Z}[x, y]$ is not a PID, and hence not a UFD), and so it will certainly not be a Principal Ideal Domain. However, if we loosen the the condition from the ideals be generated by a single element to being generated by *finitely many* elements (i.e. a Noetherian ring), then we in fact get that extending by an indeterminate x does not affect the finiteness condition on ideals.

Theorem 20.2.6: Hilbert's Basis Theorem (HBT)

Let R be a Noetherian ring. Then $R[x]$ is a Noetherian ring

If you know some linear algebra, the term "basis" is not like the term basis in linear algerba. It is because of an archaic use of the word basis when Hilbert formulated the theorem near the end of the 19th century, where it had a closer to "generation" than basis.

Proof **:**
Recall that if $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, then a_n is the leading coefficient of $p(x)$. This terminology will often be used in this proof.

Let $I \subseteq R[x]$ be an ideal. The trick is to somehow make I linked to an ideal of R (which is finitely generated since R is Noetherian). Let L be the set of all leading coefficients of the ideal I. Then L is an ideal. First, $0 \in I$, so $0 \in L$. Next, take $f, g \in I$ be two polynomials in I so that $f = ax^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ and $g = bx^m + b_{n-1}x^{m-1} + \cdots + b_1 x + b_0$ with leading coefficents a and b and degrees n and m respectively, so that $a, b \in L$. Then since I is an ideal, $rf - g \in I$ which has leading coefficient $ra - b$, and so $ra - b \in L$. Now, since R is a noetherian ring and L is an ideal of R, L is finitely generated, say $L = \langle a_1, a_2, ..., a_n \rangle$. Then for each a_i, there is a polynomial f_i that has a_i as it's coefficient. Let's say e_i is the degree of f_i, so that the set $e_1, e_2, ..., e_n$ represents the degree's of the polynomials representing the generators of L. Let $N = \max\{e_1, e_2, ..., e_n\}$.

Next, we will construct a couple of smaller ideals. Let $d \in \{1, 2, 3, ..., N-1\}$ (i.e. d is strictly smaller tha nthe largest degree polynomial of L) and define $L_d \subseteq L$ to be the set of all leading coefficients of polynomials of I of degree d, together with 0. Then L_d is also an ideal of R as can be checked by a similar metho as last proof. Since R is Noetherian, there exists a set of benerators for L_d, say $L_d = \{b_{d,1}, b_{d,2}, ..., b_{d,k_d}\}$. For each geneartor of L_d, there is a polynomial $f_{d,i} \in L_d$ that has the generator $b_{d,i}$ as it's leading coefficient.

Now, we will show that

$$I = \langle f_1, f_2, ..., f_n, f_{1,1}, f_{1,2}, ...f_{1,k_1}, f_{2,1}, ..., f_{N-1,k_{N-1}} \rangle$$

that is, I is generated by the polynomial f_i and the polynomials $f_{d,i}$. Since there are finitely many of them, I would be finitely generting completing the proof. To show this, let I' represent the right hand side. Then it is clear that $I' \subseteq I$, since every generator of I' is in I, and I' is the smallest set that contains those generators. So, let's prove $I \subseteq I'$, and take a $f \in I$.

If $\deg(f) \geq N$, then since $f \in I$, it's leading coefficient (say a) is in L: $a \in L$. Then since L is finitely generated, it is an R-linear combination of the generators of L: $r_1 a_1 + r_2 a_2 + \cdots + r_n a_n$.

From this, we can define a new polynomial $g_1 = r_1 x^{d-e_1} f_1 + r_2 x^{d-e_2} f_2 + \cdots + r_n x^{d-e_n} f_n$ with the same leading coefficient a of f that is in I'. Then

$$f - g_1$$

is a polynomial in I of degree less than $\deg(f)$. If $\deg(f - g_1)$ is still greater than or equal to f, then repeat the same process, making the polynomial $(f - g_1) - g_2$. Continue on this way till we get a polynomial of degree less than N:

$$f - g_1 - g_2 - \cdots - g_\ell = f - (g_1 + g_2 + \cdots + g_\ell)$$

Let's label this above polynomial f'. Note that the sum of the g_i's can be re-written as an R-linear combination of elements of I' since they all belong to I':

$$f - (\alpha_1 f_1 + \alpha_2 f_2 + \cdots + \alpha_n f_n)$$

Now, $\deg(f') < N$. Since $f' \in I'$, it's leading coefficient (say a) is in L_d: $a \in L_d$. Then since L_d is finitely generated, it is an R-linear combination of the generators of L_d: $a = r_1 b_{d,1} + r_2 b_{d,2} + \cdots + r_{k_d} b_{d,k_d}$. From this, we can define a new polynomial $h_1 = r_1 f_{d,1} + r_2 f_{d,2} + \cdots + r_{k_d} f_{d,k_d}$ with the same leading coefficient as f' that is in I'. Then

$$f' - h_1$$

Is of degree less than f'. Just as before, we can continue this procedure, which will terminate once the degree reaches 0:

$$f' - h_1 - h_2 - \cdots - h_p = f' - (h_1 + h_2 + \cdots + h_p)$$

Note that $\sum_i h_i$ can be decomposed into a sum of polynomials with leading coefficients from L_d:

$$f' - (\beta_{1,1} f_{1,1} + \beta_{1,2} f_{1,2} + \cdots + \beta k, k_d f_{k,k_d})$$

Since our algorithm terminated after we have subtacted the constant (that is, the leading term of L_0), we are left with:

$$f - (g_1 + g_2 + \cdots + g_\ell + h_1 + h_2 + \cdots + h_p) = 0$$

which implies $f = g_1 + g_2 + \cdots + g_\ell + h_1 + h_2 + \cdots + h_p$. Re-writing slightly, we get that

$$f = \alpha_1 f_1 + \alpha_2 f_2 + \cdots + \alpha_n f_n + \beta_{1,1} f_{1,1} + \beta_{1,2} f_{1,2} + \cdots + \beta k, k_d f_{k,k_d}$$

which shows that f is a combination of elements of I' and so $f \in I'$. Since f was arbitrary, we cna conclude that $I \subseteq I'$.

Along with the quick assertion earlier that $I' \subseteq I$, that shows us that $I = I'$, showing that I was finitely generated. Since I was an arbitrary ideal, we are left with saying that every ideal of $R[x]$ is finitely generated, showing that $R[x]$ is also Noetherian, as we sought to show.

In the exercises there will be the outlining of an alternative proof which will give a different insight (see exercise ref:HERE)

Corollary 20.2.7: HBT on mult-variable Polynomials

Let R be a Noetherian ring. Then $R[x_1, x_2, ..., x_n]$ is a Noetherian ring

Proof **:**
Since R is a Noetherian Domain, by the Hilbert Basis Theroem, $R[x_1]$ is Noetherian. Since $R[x_1]$ is a Noetherian Domain, $(R[x_1])[x_2] = R[x_1, x_2]$ is Noetherian. Continuing Inductively, we see that $R[x_1, x_2, ..., x_n]$ is Noetherian, as we sought to show

You can now return to example 10.1.5 to fully appreciate how there exist Atomic Domains which are not Noetherian Domains.

This theorem is very powerful. Since $R[x]$ is an ideal of itself, it is also finitely generated and the polynomials which generate it finitely can be characterized by their leading coefficients. In section 20.2.1, we will expand this theory further find a computational way of getting these polynomials.

Since a field is Noeherian (only ideal is (1) and (0), both clearly finite), Hilbert's Basis Theorem and induction immediately give:

Corollary 20.2.8: HBT on Polynomial Rings of Several Variables

Every ideal in the polynomial ring $F[x_1, x_2, ..., x_n]$ with coefficients from a field F is finitely generated

Proof **:**
exercise

Recall that finitely generated R-algebras are quotient of polynomial rings R. By HBT, we have the following

Corollary 20.2.9: Finitely Generated Algebras Are Noetherian

Let A be a finitely generated R-alebra. Then if R is Noetherian, A is finitely generated if and only if it is Noetherian.

Proof **:**
exercise.

Polynomial rings being over a Noetherian ring being Noetherian is central to to the study of algebras thanks to the universal property of Algebras

(A word on how Artinian rings are the opposite with D.C.C. We won't study them till section 20.7, since they sort of seem like a "oh wow, Noetherian rings can do so much, what about the opposite?" and then we realise that artinian ring implies noetherian ring, and it has so much structure that we can easily classify them all with no further fought, so we'll just leave that for later. All in all, the Noetherian condition turns out to be "more interesting")

Exercise 20.2.1

1. Alternative HBT: (the one uses Diksons lemma)
2. Consider the chain $k[[x]]] \subseteq k[[x^{1/2}]] \subseteq k[[x^{1/6}]] \subseteq \cdots \subseteq k[[x^{1/n!}]]$. Taking the union of all of these forms an object called the *Puisseaux series*. Is the Puisseaux seris a ring, and if so is the Puisseaux series Noetherian?
3. Let S be a finitely generated R-algebra, where R is Noetherian. Prove that S is Noetherian.
4. We know that ideals of $k[x]$ are generated by 1 element if k is a field. By Hilbert's basis theorem, we know that every ideal of $k[x, y]$ is finitely generated. Is it true that ideals of $k[x, y]$ is generated by only 2 elements if k is a field?
5. Which of the following are Noetherian Rings:
    1. $\mathbb{Z}[x, 1/x]$ as a subring of $\mathbb{Q}$
    2. $R := \mathbb{R} + x\mathbb{C}[x]$
    3. $R := k + xL[x]$ where $[L : k] < \infty$ is a finite field extension
    4. $R := k + xk[x, y]$ as a subring of $k[x, y]$, k a field.
6. Let M be an R-module over a Noetherian ring R. Show that there exists a $x \in M$ such that $\mathrm{Ann}_R(x)$ is a prime ideal (large hint: consider the set $\mathcal{S} = \{\mathrm{Ann}_A(x) \mid x \in M\}$ of ideals, and use the fact that R is Noetherian to conclude $\mathcal{S}$ has a maximal element. Is this element prime?)

20.2.1 Gröbner Basis

here?

20.3 Localization

Localization shall be the generalization of section 9.5 to commutative rings and modules in general (including ones with zero-divisors). With the introduction of the tensor product, localization shall be shown to have the important property of producing *flat module*, which shall have important geometric consequences in algebraic geometry. From this, we shall define the notion of local properties and the famous Nakayama lemma which shall allow many proofs to be done "locally". Concluding this section, we shall wrap up the correspondence between primes of a ring R and primes in its integral extension.

The term "localization" comes from algebraic geometry, where this process of localizing mirrors the process of "localizing" for manifolds. If you have done some differential geometry, recall that the tangent space at a point p can be defined as the set of all germs at a point p whose value is 0, and then taking its quotient with its square: $\mathfrak{m}_p/(\mathfrak{m}_p)^2$. Localizing shall correspond to this process, where "inverting" elements around a point is the same as a function on a manifold being invertible on a small enough local neighborhood where $f(p) \neq 0$. The reader may question this correspondence, for in differential geometry we are working with smooth functions over a [smooth] manifold, while in the

algebraic case we simply have a [commutative] ring. However, the correspondence is more equivalent than it may at first seem. It shall be particularly clear by the end of chapter 21, where it shall be shown that there is a way of thinking of all commutative rings as a set of "functions" on a special geometric space (the geometric space is called an [affine] *scheme*), hence there is a direct parallel between smooth functions on the geometric object of a manifold and commutative rings where each element is interpreted as a function on a scheme.

20.3.1 Localization of Rings

This is a generalization of the constructing of ring of fraction from theorem 9.5.1 by allowing the set that's closed under multiplication D to have zero divisors. This will imply that R need not be embedded into $D^{-1}R$, but the universal property is still satisfied. Furthermore, we will have to solve the problem of elements that "should" be zero being equal to nonzero elements. If $a \in R$ and $d \in D$ where $as = 0$, then:

$$a = \frac{a}{1} = \frac{as}{s} = \frac{0}{s} = 0$$

this issue will be addressed in the proof:

Theorem 20.3.1: Universal Property of Localization

Let R be a commutative ring and D be a multiplicatively closed subset containing 1. Then there exists a ring $D^{-1}R$ such that the following diagram commutes

$$\begin{tikzcd} D^{-1}R \arrow[dr, "\Psi"] & \\ R \arrow[u, "\iota"] \arrow[r, "\psi"'] & S \end{tikzcd}$$

with the property that $\psi(D)$ is a subset of the units of S

Proof **:**
A lot of this proof is similar to theorem 9.5.1 (Ring of Fractions) and so it will be quickly presented pointing out where the generalization of D having zero divisors makes a difference.

Let $R \times D$ have the equivalence relation

$$(r, b) \sim (s, d) \;\Leftrightarrow\; x(rd - bs) = 0 \text{ for some } x \in D$$

Which is essentially the same as in theorem 9.5.1. The difference between the proof in the case where D does not have zero divisors' is with the fact that the terms are scaled by some $x \in D$. This is to make sure elements that are zero-divisors will belong to the [0] (as we'll show in the proceeding corollary). Usually, the equivalence class $[(r, d)]$ is denoted $\frac{r}{s}$ or r/s.

Define

$$\frac{r}{d} + \frac{s}{e} = \frac{re + sd}{de} \qquad \frac{r}{d}\frac{s}{e} = \frac{rs}{de}$$

Then these two operations form a well-defined commutative ring $(R \times D)/\sim$; this ring is denoted $D^{-1}R$. Note that $\frac{d}{1}$ is always a unit in $D^{-1}R$ (even when it is a zero-divisor, since it will be shown the ring will be a zero-ring)

Define $\iota : R \to D^{-1}R$ by $\iota(r) = \frac{r}{1}$. This map defines a natural ring homomorphism. Using this map, we can lift a map $\varphi : R \to S$ to $\Psi : D^{-1}R \to S$ where D is mapped to the units of S by defining:

$$\Psi(\frac{r}{d}) = \psi(r)\psi(d)^{-1}$$

This map is well-defined: Notice if $\frac{r}{d} = \frac{s}{e}$ then by definition of the equivalence relation, for some $x \in D$, $x(re - sd) = 0$ Thus $\varphi(x(re - sd)) = \varphi(x)(\varphi(re) - \varphi(sd)) = 0$. Since x is a unit in S and S is a commutative ring, s is not a zero-divisor, and so $\varphi(re) - \varphi(sd) = 0$ Moving values around, we get

$$\varphi(r)\varphi(d)^{-1} = \varphi(s)\varphi(e)^{-1}$$

Thus, Ψ is a well-defined ring homomorphism. Furthermore, by construction, $\Psi \circ \iota = \varphi$

Finally, Ψ is unique. If there was another Ψ' such that $\Psi' \circ \iota = \varphi$, then:

$$\varphi(r) = \Psi'(\iota(r)) = \Psi'\left(\frac{r}{1}\right)$$

Thus Ψ' maps $\frac{r}{1}$ in the same way as Ψ. For the $\frac{1}{d}$ terms, notice that

$$\frac{1}{d} = \left(\frac{d}{1}\right)^{-1}$$

and so $\varphi(d) = \Psi'(\frac{d}{1})$, and so, by properties of ring homomorphisms

$$\Psi'\left(\frac{d}{1}\right)^{-1} = \Psi'\left(\frac{1}{d}\right)$$

showing that $\Psi' = \Psi$, fulfilling the last condition needed to prove for the universal property

As before, $D^{-1}R$ is still called the ring of fractions. However, there is a more common name with a geometric intuition:

Definition 20.3.2: Localization of R at D

Let R be a commutative ring and D be a multiplicatively closed subset of R. Then the ring $D^{-1}R$ produced in the previous theorem is the *ring of fractions of R with respect to D*, or the *localization of R at D*. This ring is also sometimes denoted $R[D^{-1}]$.
If $D = \{f^n \mid n \in \mathbb{Z}_{\geq 0}\}$ for some $f \in R$, we usually denote $D^{-1}R = R_f$.

If D is not a multiplicatively closed set, we define $R[D^{-1}] := R[\overline{D}^{-1}]$ where $\overline{D^{-1}}$ is the multiplicative closure of D.

The fact that D can contain zero divisors means that ι can now have non-trivial kernel:

Corollary 20.3.3: Kernel of ι

Let ι be the natural inclusion of R into $D^{-1}R$ for some appropriate D. Then

1. $\ker(\iota) = \{r \in R \mid xr = 0 \text{ for some } x \in D\}$. In particular, ι is an injection if and only if D doesn't contain 0 or zero-divisors
2. $D^{-1}R = 0$ if and only if $0 \in D$, or more generally, if D contains no nilpotent elements (since D is multiplicatively closed)

Proof **:**

1. If $\iota(r) = \frac{r}{1} = 0$, then we have that $(r, 1) \sim (0, 1)$ and so $x(1r - 0) = xr = 0$ for some $x \in D$. Thus, ι is an injection if and only if D does not contain 0 or zero-divisors
2. $D^{-1}R = 0$ if and only if $(1, 1) \sim (0, 1)$ if and only if $x1 = 0$ if and only if $x = 0$. Thus, $0 \in D$

This let's us get a sense of how adding zero-divisors to D changes the resulting localization of R at D.

Example 20.3.1: Localization

All examples in section 9.5 are valid examples, including the trivial one of any field $\mathbb{F}$, $\mathbb{Z}$ to $\mathbb{Q}$ with $D = \mathbb{Z} - \{0\}$, and $\mathbb{F}[x]$ to $\mathbb{F}(x)$ with $D = F[x] - \{0\}$. We shall give some more interesting examples here (all rings are commutative):

1. If R is an integral domain, then we can always take $D = R - \{0\}$. The ring $D^{-1}R$ then always a field, and is called the *field of fractions of* R, and is sometimes denoted Q. Then for any smaller multiplicatively closed subset E, $E^{-1}R$ is a subring of Q with elements $r/e \in Q$ with $r \in R$ and $e \in E$
2. Let R be a ring and $f \in R$. Define $D = \{f^n \mid n \in \mathbb{Z}_{\geq 0}\}$. Then we'll denote $D^{-1}R = R_f$. Then R_f is a ring in which the element f has become a unit. In a sense, it is now possible to solve the question
$$fx - 1 = 0$$
In fact, this intuition can be made more rigorous showing that
$$R_f \cong \frac{R[x]}{fx - 1}$$
This is similar to how we worked with field extensions, and in fact is an important link to connect the intuition of localization to a geometric intuition. Keep this example in the back of your mind for now.
3. This example is called *localizing at a prime*. Let R be a ring and P a prime ideal of R. Then the set $D = R - P$ is multiplicatively closed. To see that, let $x, y \in D$. If it were the case that $xy \in P$, then either $x \in P$ or $y \in P$. But then one of them wouldn't be in D – a contradiction. Thus, D is indeed multiplicatively closed. The localization $D^{-1}R$ is called the *localization of* R *at* P and is commonly denoted as R_P. In this ring, all elements not in P become units. As a simple example, take $R = \mathbb{Z}$ and $P = (p)$ for some prime p. Then by

the nice divisibility properties of integers, the set $\mathbb{Z}_{(p)}$ can with a little work be shown to be:

$$\mathbb{Z}_{(p)} = \left\{ \frac{a}{b} \in \mathbb{Q} \;\middle|\; p \not| b \right\}$$

that is, it is all elements of $\mathbb{Q}$ for which the denominator is not divisible by p. Can you prove that $\mathbb{Z}_{(p)}$ has only one maximal ideal (and hence, is a local ring)? Is this true of all localization at primes?

It is tempting to say that if $R_{(p)}$ is the localization of a ring R at the prime p, then $a/b \in R_{(p)}$ if $b \notin (p)$, however this is not necessarily the case as we are dealing with an equivalence class! The reader may then ask if we can take a better representative (like the representative a/b in $\mathbb{Z}$ such that $\gcd(a, b) = 1$), however R is not necessarily a UFD and so this may not be well-defined.

4. The following examples is a way to "localise" functions. Let k be a field and V be any set. Define R to be functions from V to k containing the constant functions ($F \subseteq \mathrm{Hom}_{\mathbf{Set}}^{+\text{ constants}}(V, k)$) with $+$ and $\cdot$ defined by component-wise addition and multiplication on the image: $R = (F, +, \cdot)$. For example, take all continuous functions from $[0, 1]$ to $\mathbb{R}$. For any $a \in V$, let $M_a \subseteq R$ be the set of all functions that vanish at a: $f(a) = 0$. Then M_a is the kernel of the evaluation map from R to k:

$$\mathrm{ev}_a : R \to k \qquad f \mapsto f(a) \qquad \ker(\mathrm{ev}_a) = M_a$$

Since R contains constant functions, ev_a is surjective and so $k \cong R/M_a$. By proposition 9.4.11, M_a is a maximal idea, and hence a prime ideal. Thus, we can define the localization of R at M_a to be:

$$R_{M_a} = \left\{ \frac{f}{g} \;\middle|\; f, g \in R,\ g(a) \neq 0 \right\}$$

where ev_a is extended as:

$$(f/g)(a) = f(a)/g(a)$$

This then makes R_{M_a} a way of defining rational functions.

5. Let $R = \mathbb{C}[x]$, and take $p = (x - a)$. Since $\mathbb{C}[x]$ is a Principal Ideal Domain and $(x - a)$ is irreducible, $(x - a)$ is prime, and so we can define

$$\mathbb{C}[x]_{(x-a)} = \left\{ p(x)/q(x) \;\middle|\; x - a \not| q(x) \right\}$$

which can be thought of all rational functions which do not vanish at a.

Localization has many nice geometric interpretations. As we will see (T)BD, all algebraic varieties (i.e. irreducible algebraic sets) are related to the prime ideals. Hence, understanding how ideals of R correspond to ideals of $D^{-1}R$ will bring us to understanding a geometric connection.

For notational simplicity, let $D^{-1}R\iota(I) = {}^{\iota}I$ for some ideal I (note that this is *not* $\iota(I)$) . Let $\pi : D^{-1}R \to R$ be the function mapping $r/d \mapsto r$ so that for some ideal $I \subseteq D^{-1}R$, $R\pi(I) = \iota^{-1}(I) = {}^{\pi}I$. The ideal ${}^{\iota}I$ is sometimes called the *extension* of I and ${}^{\pi}I$ the *contraction* of I. Since every element of ${}^{\iota}I$ can be written of the form a/d where $a \in I$ and $d \in D$, the notation $D^{-1}I$ is also frequently used.

Proposition 20.3.4: Ideals of Localization

Let R be a ring and $D^{-1}R$ be the localization of R at D for appropriate D. Then:

1. For every ideal $J \subseteq D^{-1}R$,
$$J = J^{ce}$$
As a consequence, every ideal of $D^{-1}R$ is an extension of an ideal in R, and every ideal of D^{-1} contracts to a distinct ideal in R

2. If $I \subseteq R$ is an ideal of R, then:
$$(\lim I)^e = \{r \in R \mid dr \in I \text{ for some } d \in D\}$$
As a consequence, $I^e = D^{-1}R$ if and only if $I \cap D \neq \emptyset$

3. There is a bijection between the set of prime ideals with $P \cap D = \emptyset$ and the set of prime ideals of $D^{-1}R$:
$$\left\{\begin{matrix}\text{prime ideals of } R \text{ with} E \\ P \cap D = \emptyset\end{matrix}\right\} \underset{(-)^c}{\overset{(-)^e}{\rightleftarrows}} \left\{\text{prime ideals of } D^{-1}R\right\}$$

4. If R is Noetherian, then $D^{-1}R$ is Noetherian. Similarly if R is Artinian

Proof **:**

1. Notice that by construction, ${}^{\iota}({}^{\pi}J) \subseteq J$. For the other direction, notice that for any $a/d \in J$ it is always the case that $a/1 = d(a/d) \in J$, that is there exists a a/d such that $d(a/d) = a/1$. Since $\iota(\pi(a/d)) = a/1$. and so by definition, of ${}^{\iota}({}^{\pi}J)$, $(1/d)(a/1) = a/d \in {}^{\iota}({}^{\pi}J)$ (since it closed by multiplication of elements from $D^{-1}R$.

2. Let $I' = \{r \in R \mid dr \in I \text{ for some } d \in D\}$ be the set in the question. We'll show $I' = {}^{\pi}(\lim I)$. For the $\subseteq$ direction, take $r \in I'$. Then by definition there exists a $d \in D$ such that $dr = a \in I$. Then $r/1 = a/d \in \lim I$, so contracting we get $r \in {}^{\pi}(\lim I)$. For the $\supseteq$ direction, take $r \in {}^{\pi}(\lim I)$. Then $r/1 = a/d \in \lim I$ for some $a \in I$ and $d \in D$. Then by the definition of the equivalence class, we have $x(dr - a) = 0$ for some $x \in D$, and hence $xdr = xa$ where $xa \in I$. Since $xd \in D$, $r \in I'$. The rest of the assertions of (2) follow quickly.

3. To map prime ideals from $D^{-1}R$ to R, recall that pre-images of prime ideals is prime (see exercise ref:HERE in section ref:HERE). Thus, by construction, the contraction must map prime ideals of $D^{-1}R$ to prime ideals that are disjoint from D.

   For the converse, let I be an ideal in R that is disjoint from D, and consider $Q = {}^{\iota}I$. We'll show Q is prime. Suppose $(a/d_1)(b/d_2) = (ab/d_1d_2) \in Q$. Then $ab/d_1d_2 = c/d$ for some $c \in P$ and $d \in D$. Then by definition of the equivalent relation, $x(dab - d_1d_2c) = 0$ for some $x \in D$. Since $xdab - xd_1d_2c = 0 \in P$ and P is closed under subtraction and left-multiplication of elements of R, it must be that $xdab = (xd)(ab) \in P$. Since P is disjoint from D, it must be that $ab \in P$. Since P is prime, either a or b is an element of P. Thus, either a/d_1 or b/d_2 is an element of Q, showing that Q is a prime ideal. Thus, extending prime ideals that are disjoint from D map to prime ideals in $D^{-1}R$.

Finally, by (2), since P is disjoint from D we have $P = {}^{\pi}({}^{\iota}P)$, showing that extending and contracting are inverses of each other, finishing the proof for (3).

4. Given some ascending chain of ideals in $D^{-1}R$, they contract to ascending chain of ideals in R. Since R is Noetherian, they stabilize. Thus, the ideals that get contracted must also stabilize. Similarly for Artinian Rings.

The ideal in part (2) of proposition 20.3.4 is given a name:

Definition 20.3.5: Saturated Ideal

Let R be a commutative ring, $I \subseteq R$ is an ideal and $D^{-1}R$ is the localization of R at D for appropriate D. Then *saturation* of the ideal I with respect to D is ${}^{\pi}(\lim I)$. If $I = {}^{\pi}(\lim I)$, then I is said to be *saturated* or *D-saturated* with respect to D

(Re-write the previosu results in terms of $\operatorname{Spec} R$)

and show that for localization at a prime, we have the following image:

If we instead have R_f to be the localization at f (so $D = \{f^n \mid n \in \mathbb{Z}\}$, then in fact there is a bijection between $\operatorname{Spec} R \cong_{\mathbf{Set}} \operatorname{Spec} R_f \oplus \operatorname{Spec} R/(f)$. Later on, we will see that we can upgrade this bijection into a homeomorphism

Proposition 20.3.6: Localization Property Preservation

Let R be a ring and let $D^{-1}R$ be the localization of R. Then the following properties of R are preserved in $D^{-1}R$:

1. no zero divisors
2. Noetherian
3. Integrally closed

Proof **:**
exercise

Exercise 20.3.1

1. Let R be a ring, f be an element of R ,and R_f be the localization of R at $\{f^n \mid n \in \mathbb{Z}_{\geq 0}\}$. Then

   1. $$R_f \cong \frac{R[x]}{fx - 1}$$

   2. $R_f \cong R_{f^n}$ for $n \geq 1$

   3. If f is nilpotent, then $R_f = 0$

2. Let R be a normal domain and S a multiplicatively closed set. Then $S^{-1}R$ is a normal domain

20.3.2 Localization of Modules

Since every ring is a module over itself, we can see how localization works for this special case of modules. We extend the concept of localization now to more general modules. The construction is essentially the same thing: if M is an R-module, and $D \subseteq R$ is multiplicatively closed, define the same equivalence class on $M \times D$ and the operation's on $(M \times D)/\sim$ as

$$\frac{m}{d} + \frac{m'}{d'} = \frac{d'm + dm'}{dd'} \quad \text{and} \quad r \cdot \frac{m}{d} = \frac{r \cdot m}{d}$$

Then this is indeed a module with the proof proceeeding essentially the same as theorem 20.3.1, and so is left as a sanity check. Just like $R[D^{-1}]$ is the localization of R at D, we will denote $M[D^{-1}]$ to be the localization of M at D. Thus, an R-module M is extended to an $D^{-1}R$-module $M[D^{-1}]$.

Proposition 20.3.7: Zeros of $M[D^{-1}]$

Let R be a ring, $D \subseteq R$ be a multiplicatively closed subset, and M be a module over R. Then m goes to 0 in $M[U^{-1}]$ if and only if m is annihilated by an element $d \in D$. If M is finitely generated, then $M[U^{-1}] = 0$ if and only if M is annihilated by an element in D.

If R lies in a field K and M is an R-module that is contained in a k-vector space, then $\iota : M \to D^{-1}M$ is injective iff $M \xrightarrow{\cdot s} M$ is injective for all $s \in D$.

***Proof* :**
Exercise

Localization modules works well with homomorphisms: let $\varphi : M \to N$ be an R-module homomorphism. Then by the universal property, we see that this is extended to an $D^{-1}R$-module homomorphism $D^{-1}\varphi : M[D^{-1}] \to N[D^{-1}]$ by mapping:

$$D^{-1}\varphi\left(\frac{m}{d}\right) = \frac{\varphi(m)}{d}$$

Furthermore, $D^{-1}(\varphi \circ \psi) = (D^{-1}\varphi) \circ (D^{-1}\psi)$, i.e. $D^{-1}(-)$ is a functor from Mod_R to $\mathrm{Mod}_{D^{-1}R}$. We may ask for certain properties of this functor, most basic perhaps would be if $D^{-1}(-)$ is an exact functor, that it preserves kernels and cokernels. This is the case:

Proposition 20.3.8: Localization Is Exact

Let M_1, M_2, M_3 be R-modules such that $0 \to M_1 \xrightarrow{f} M_2 \xrightarrow{g} M_3 \to 0$ is a short exact sequence. Then

$$0 \to M_1[D^{-1}] \xrightarrow{D^{-1}f} M_2[D^{-1}] \xrightarrow{D^{-1}g} M_3[D^{-1}] \to 0$$

is a short exact sequence, that is $D^{-1}(__)$ is an exact functor.

Proof :
since $g \circ f =$, $D^{-1}g \circ D^{-1}f = D^{-1}(0) = 0$, and so $\text{im}(D^{-1}f) \subseteq \ker(D^{-1}g)$. For the reverse inclusion, let $m/d \in \ker(D^{-1}g)$ so that $g\left(\frac{m}{d}\right) = \frac{g(m)}{d} = 0$ in $M_3[D^{-1}]$. Thus, there exists an s such that $sg(m) = 0$ in M_3. Since g is an R-module homomorphism, $sg(m) = g(sm)$, and so $sm \in \ker(g) = \text{im}(f)$, implying for some $m' \in M_3$, $f(m') = tm$. Putting it all together, we get that in $M_2[D^{-1}]$:

$$\frac{m}{d} = \frac{f(m')}{ds} = (D^{-1}f)\left(\frac{m'}{ds}\right) \in \text{im}(D^{-1}f)$$

showing that $\ker(D^{-1}g) \subseteq \text{im}(D^{-1}f)$, which along with the fact that $\text{im}(D^{-1}f) \subseteq \ker(D^{-1}g)$ show's that

$$\ker(D^{-1}g) = \text{im}(D^{-1}f)$$

completing the proof.

Notice that as we are defining a $D^{-1}R$-module, we are extending scalars. If we use section 16.4, then we can use the following observation to arrive at this result more quickly:

Proposition 20.3.9: Characterization of Localization

Let M be an R-module and $M[D^{-1}]$ a localiation. Then:

$$M[D^{-1}] \cong_R D^{-1}R \otimes_R M$$

Proof :
Define the map:

$$(r/d) \times m \mapsto \frac{r \cdot m}{d}$$

then it is easy to that this map is R-bilinear, and hence by the universal property of the tensor-product, there exists an R-module homomorphism:

$$f : D^{-1}R \otimes M \to M[D^{-1}]$$

It is clear that f is surjective and is uniquely defined with respect to the universal property. It remains to check that f is injective. First, show that for any element in $D^{-1}R \otimes M$, it can be written as:

$$\frac{1}{d} \otimes m$$

for some $d \in D$ and $m \in M$. Let's say $f(1/d \otimes m) = m/d = 0$. Then $sm = 0$ for some $s \in D$. Then:

$$\frac{1}{d} \otimes m = \frac{s}{ds} \otimes m = \frac{1}{ds} \otimes sm = \frac{1}{ds} \otimes 0 = 0$$

showing that f is injective, and hence an R-module isomorphism, completing the proof.

As a direct conssequence, $D^{-1}R$ is a *flat R-module*.

Corollary 20.3.10: Localization Commuting With Tensor

Let M and N be R-modules, $D \subseteq R$ a semi-group. Then there exists a unique isomorphism:

$$\varphi : M[D^{-1}] \otimes_{D^{-1}R} N[D^{-1}] \to (M \otimes_A N)[D^{-1}] \qquad (m/d) \otimes (n/s) \mapsto (m \otimes n)/ds$$

In particular, if p is any prime ideal:

$$M_p \otimes_{A_p} N_p \cong (M \otimes_A N)_p$$

Proof **:**
Use proposition 20.3.9 and the appropriate isomorphism theorem for modules.

(Some more cool results from Micheal's p. 43)

20.3.3 Local Properties

Localization allows us to diminish our problem to a ring much closer to a field, in particular if we want to study a family of prime ideals contained in some prime ideal p. In fact, there are many results about rings that can be deduced by studying rings *locally*, namely by studying rings at each prime ideal[2]. This motivates the following concept:

Definition 20.3.11: Local Properties

Let P be a property of a ring (ex. commutative, free, etc.). Then a property P of a ring R (resp. of an R-module M) is called a *local property* if it respects the following:

$$R \text{ (resp. } M\text{) has property } P \Leftrightarrow R_p \text{ (resp. } M_p\text{) has property } P \text{ for each prime ideal } p \subseteq R$$

In otherwords, a property is local if we can move back and forth between studying any ring (resp. module) localized at a prime. Here are some examples of local properties:

Proposition 20.3.12: Zero Module Is A Local Property

Let M be an R-module. Then the following are equivalent:

1. $M = 0$
2. $M_p = 0$ or all prime ideals $p \subseteq R$
3. $M_m = 0$ for all maximal ideal $m \subseteq R$

Proof **:**
Certainy (1) $\Rightarrow$ (2) $\Rightarrow$ (3). For (3) $\Rightarrow$ (1), let $0 \neq x \in M$ be a nonzero element of an R-module M with $0 \neq 1$. Let $I = \text{Ann}(x)$. Then I is contained in some maximal ideal m: $I \subseteq m$. Consider M_m. Since $M_m = 0$, $x/1 = 0$, implying $rx = 0$ for some $r \in R - m$. Howver $\text{Ann}(x) \subseteq m$, meaning so

[2]This observation is at the basis of interpreting rings as geometric objects

such element can exist, a contradiction, showing that $M = 0$ as we sought to show.

Another property that only needs to be verified locally is the injectivity/surjectivity of a function:

Proposition 20.3.13: Injectivity/Surjectivity/Isomorphism Is A Local Property

Let $\varphi : M \to N$ be an R-module homomorphism. Then the following are equivalent:

1. φ is an injection/surjection/isomorphism
2. $\varphi_p : M_p \to N_p$ is injection/surjection/isomorphism for all prime ideals $p \subseteq R$
3. $\varphi_m : M_m \to N_m$ is injection/surjection/isomorphism for all maximal ideals $m \subseteq R$

***Proof* :**
The proof shall be doen for injectivity, surjectivitiy follows easly and isomorphism is combining the two.

(1) $\Rightarrow$ (2) Since $0 \to M \xrightarrow{\varphi} N$ is exact, by proposition 20.3.8, $0 \to M_p \xrightarrow{\varphi_p} N_p$ is exact, hence φ_p is injective for any prime ideal $p \subseteq R$.

(2) $\Rightarrow$ (3) Every maximal idael is prime

(3) $\Rightarrow$ (1) Let $K = \ker(\varphi)$. Then the sequence $0 \to K \to M \to N$ is exact, and hence proposition 20.3.8 $0 \to K_m \to M_m \to N_m$ is exact, therefore $K'_m \cong \ker(\varphi_m) = 0$ since φ_m is injective. Thus, by proposition 20.3.12 φ is injective.

to show surjectivity, simply reverse the arrows.

Corollary 20.3.14: Exactness Is A Local Property

Let A, B, C be R-modules and let $1 \to A_{\mathfrak{m}} \to B_{\mathfrak{m}} \to C_{\mathfrak{m}}$ for each maximal ideal $\mathfrak{m} \subseteq R$. Then there exists maps such that

$$0 \to A \to B \to C \to 0$$

is exact

***Proof* :**
take advantage of the fact that injectivity/surjectivity are local properties

A final example that will be useful is that faltness is a local property:

Proposition 20.3.15: Flatness Is A Local Property

Let M be an R-module. Then the following are equivalent:

1. M is a flat R-module
2. M_p is a flat R_p-module for each prime ideal $p \subseteq R$
3. M_m is a flat R_m-module for each maximal ideal $m \subseteq R$

Proof **:**
(1) $\Rightarrow$ (2) comes from pervious lemmas and (2) $\Rightarrow$ (3) is immediate. For (3) $\Rightarrow$ (1), let $N \to P$ be any R-module homomorphism and $m \subseteq R$ any maximal idea. Then:

$$\begin{aligned} N \to P \text{ is injective } &\Rightarrow N_m \to P_m \text{ is injective} && \text{proposition 20.3.13} \\ &\Rightarrow N_m \otimes_{R_m} M_m \to P_m \otimes_{R_m} M_m \text{ is injective} && \text{flatness property} \\ &\Rightarrow (N \otimes_R M)_m \to (P \otimes_R M)_m \text{ is injective} && \text{corollary 20.3.10} \\ &\Rightarrow N \otimes_R M \to P \otimes_R M \text{ is injective} && \text{proposition 20.3.13} \end{aligned}$$

Hence, M is flat.

Proposition 20.3.16: Integral Closure Is A Local Property

Let R be an integral domain. Then the following are equivalent:

1. R is integrally closed
2. R_p is integrally closed for each prime ideal $p \subseteq R$
3. R_m is integrally closed for each maximal ideal $m \subseteq R$

Proof **:**
exercise.

The following is a way of proving results locally:

Proposition 20.3.17: Local Properties On Modules

Let R be a subring of a field k, and let M be an R-module contained in a k-vector space V (equiv. the map $M \to M \otimes_R k$ is injective). Then:

$$M = \bigcap_{\mathfrak{m}} M_{\mathfrak{m}} = \bigcap_{\mathfrak{p}} M_{\mathfrak{p}}$$

Proof **:**
Certainly $M \subseteq \cap_{\mathfrak{m}} M_{\mathfrak{m}}$. For the other direciton, take $x \in \cap_{\mathfrak{m}} M - \mathfrak{m}$. For each maximal ideal $\mathfrak{m}$, $x = m/s$ for some $m \in M$ and $s \in R - \mathfrak{m}$. Hence $sx \in M$. Define the ideal

$$I_a = \{a \in R \mid ax \in M\}$$

Then $as \in I_a$. But by construction $s \notin \mathfrak{m}$,hence $I_a \not\subseteq \mathfrak{m}$. But then it must be that $I_a = R$, hence $x = 1 \cdot x \in M$, the giving the result

Extending to prime ideals is left as an exercise.

Corollary 20.3.18: Local Properties Of Ideals

Let R be an integral domain. Then every ideal $I \subseteq R$ is equal to the intersection of its localizations at the maximal ideals of R, and also to the intersections of its localization at the prime ideals of R

Hence, to show a property of an ideal I in an integral domain R, it suffices to show that the properly holds for all its localization $I_{\mathfrak{p}}$ and it is preserved under intersections.

Exercise 20.3.2

1. Show that M is projective if and only if each M_p is free.

20.3.4 Local Rings

Recall that the localization of $\mathbb{Z}$ at any prime ideal (p), $\mathbb{Z}_{(p)}$, has a unique maximal ideal. It was left as an exercise to see whether any localization at a prime, R_p left R with only a single maximal ideal. This is indeed the case, meaning there is a unique field associated to R_p, letting us study R in terms of the collection of field induced by R_p. Due to this, it is worth-while studying rings with a unique maximal ideal:

Definition 20.3.19: Local Ring

Let R be a ring (not necessarily commutative). Then if R has a unique two-sided maximal ideal, then R is said to be a *local ring*.

As a consequence, all simple rings (including fields and $M_k(n)$ for $n \geq 2$) are trivially local rings.

Proposition 20.3.20: Local Ring Relation to left/right Maximal Ideal

Let R be a ring (not necessarily commutative). Then the following are equivalent:

1. R is a local ring
2. R has a unique maximal left-ideal
3. R has a unique maximal right-ideal

Proof **:**
exercise

In the commutative case, this translates to R simply having a unique maximl ideal. This is particularly enlightening: if x is not in M, then x must be a unit, or else $(x) \subseteq M$.

Example 20.3.2: Local Rings

1. Every field is trivially a local ring.
2. $\mathbb{Z}/p^n\mathbb{Z}$ is a local ring with $n > 1$ with maximal ideal $\overline{p}$
3. if $f \in k[x]$ is irreducible then $k[x]/(f(x)^n)$ is a local ring with maximal ideal $\overline{(f(x))}$
4. (geometric example) Let M be a smooth manifold (real or complex). Then define $\mathcal{O}_{M,x}$ to be the germ of smooth functions at x, that is, two smooth functions $f, g \in C^\infty(M)$ are in the same equivalence relation $f \sim g$ (i.e. $[f] = [g]$ are the same germ) if there exists some open set $U \subseteq M$ containing x such that $f|_U = g|_U$. Clearly, $\mathcal{O}_{M,x}$ is a ring (using the ring operation of the codomain $\mathbb{R}$ or $\mathbb{C}$). The maximal ideal of $\mathcal{O}_{M,x}$ are those functions that are 0 at x

Proposition 20.3.21: Local Ring Equivalences

Let R be a local ring. Then the following are equivalent:

1. R is local ring.
2. $1 \neq 0$, and the sum of any two non-units is a non-unit
3. $1 \neq 0$, and if $x \in R$ then either x is a unit, or $1 - x$ is a unit
4. If a finite sum is a unit, then one term has to be a unit

Notice too that the final condition can include the "vacuously empty" sum, which sums to 0, and since 0 is not a unit, $0 \neq 1$ (TBD).

Proof **:**
here

Using localization and quotienting, we have a large amount of control over the prime ideals we want to work with: given a (commutative) ring R, if we take R_p, we eliminates all ideals except those contained in p, while R/p eliminates except those contained not contained in p (i.e. containing p). This can be visualized like so:

Hence, R_p/q^e (resp. $(R/p)_{\overline{q}}$) will only keep all the ideals between p and q, and if $p = q$, then the result is a field, known as the *residue field*:

Definition 20.3.22: Residue Field

Let R be a commutative local ring with maximal ideal m. Then R/m is called the *residue field*.

The quotient $R/\mathfrak{p}$ and the ring $R_\mathfrak{p}$ are canonically related: the ring $R_\mathfrak{p}$ can be thought of as the localization of $R/\mathfrak{p}$:

Proposition 20.3.23: Localization and Quotient Commute

Let $\mathfrak{p} \subseteq R$ be prime ideal of a ring R, and let $\mathfrak{m}_\mathfrak{p}$ be the unique maximal ideal of $R_\mathfrak{p}$. Then localizing at $\mathfrak{p}$ and then taking the field of fractions is isomorphic to taking quotient at $\mathfrak{p}$ and then localizing (at $\overline{fp} = \overline{0}$) Then there is a canonical embedding:

$$R/\mathfrak{p} \hookrightarrow R_\mathfrak{p}/\mathfrak{m}_\mathfrak{p}$$

that maps $R_\mathfrak{p}/\mathfrak{m}_\mathfrak{p}$ into the field of fractions of $R/\mathfrak{p}$

This may be visualized as the following diagram:

$$\begin{tikzcd}
 & R/\mathfrak{p} \arrow[dr, "\text{localize}"] & \\
R \arrow[ur, "\text{quotient}"] \arrow[dr, "\text{localize}"'] & & K(R/\mathfrak{p}) \cong R_\mathfrak{p}/\mathfrak{p}R_\mathfrak{p} \\
 & R_\mathfrak{p} \arrow[ur, "\text{quotient}"'] &
\end{tikzcd}$$

Proof **:**
exercise

20.3.5 Nakayama's Lemma

(Atiyah Michael in Comm alg has some interesting results that feel are worth putting in here: p.21 on fin. gen modules)

As mentioned at the beginning, localization let's us get geometric information on a local level. Nakayama's lemma will be one way we translate that information to a global level. In this section, we will restrict to finitely generated modules (in exercise ref:HERE, we will work with the infinite case).

Let's say M is a finitely generated module over R. If we were to localize M at a prime, we would get a local ring $R[D^{-1}]$. For simplicity, let's denote this ring G. As we've shown, G has only one maximal ideal m. Modding G by m, we get the field G/m. M over G/m is a vector-space, and so it has a basis. The heart of Nakayama's lemma is that this can be reversed: Using the basis of M over G/m, we can find a basis for M over G!

We will guild this up in three steps, which are collectively called Nakayama's Lemma:

Definition 20.3.24: Jacobson Radical

Let R be a ring. Then the *jacobson radical* of R, denoted $\mathrm{Jac}(R)$ is the intersection fo all maximal ideals of R:

$$\mathrm{Jac}(R) = \bigcap_{\substack{M \subsetneq R \\ \text{maximal}}} M$$

(analogous to Fratini subgroup)

Proposition 20.3.25: Properties Of Jacobson Radical

Let $J \subseteq R$ be the Jacobson radical of a commutative ring R. Then:

1. If $I \subsetneq R$ is a proper ideal, then the ideal generated by I and J is also proper: $(I, J) \subsetneq R$
2. The jacobson radical is nilpoent (i.e. if $\mathrm{rad}(J)^n = 0$ for some n)
3. $x \in J$ if and only if $1 - rx$ is a unit for every $r \in R$

Proof **:**

1. If I is maximal, then $J \subseteq I$ and we're done, and if not then $I \subseteq M$ for some maximal ideal so $(I, J) \subseteq M \subsetneq R$
2. Let J be the jacobson radical. By the D.C.C. the following chain stabalizes

$$J \supset J^2 \subset \cdots J^n = J^{n+1} = \cdots$$

Consider the set of all annihilators of J^n:

$$I = \{x \in R \mid xJ^n = 0\}$$

I claim that $I = (1)$, which will show that $1J^n = J^n = 0$ (i.e. J is nilpotent). For the sake of contradiction, let's say it is not I. Then pick $I' \supsetneq I$ that is minimal (i.e. there is no ideal

between I' and I). Such an ideal exists by the D.C.C. condition. Then by minimality, it must be that I' is generated by adding just one more element, that is

$$I' = I = Rx \qquad x \in R,\ x \notin I$$

Thus, I'/I is a simple R-module by minimality (i.e. has no other ideals in it). However, it is easy to see that any simple R-module is of the form R/m for some maximal ideal m. Thus,

$$I'/I \cong A/m$$

multiplying both sidses by m, we "kill" the right hand side and get

$$mI' \subseteq I$$

Thus $JI' \subseteq I.$, so $xJ \in I$, and so by definition of I,

$$xJ^{n+1} = 0$$

but by assumption $xJ^{n+1} = J^n$, but that implies $x \in I$. However, $x \notin I$, a contradiction.

3. Doing the contra-positive, let's say $x \notin \text{Jac}(R)$. Then $x \notin M'$ for some maximal ideal M'. By maximality, $M + (x) = (1)$, so $y + rx = 1$, implying $1 - rx = y \in M'$, showing that $1 - rx$ cannot be a unit (this does not show "forall $r \in R$"). Conversley, if $1 - rx$ is not a unit, $1 - rx \in M$ for some M. However, $1, rx \notin M$, since if they were then $1 \in M$. Thus, $rx \notin \text{Jac}(R)$ for all r, but then $x \notin \text{Jac}(R)$.

Lemma 20.3.26: Nakayama's Lemma I

Let M be a finitely generated R-module, and suppose $\alpha \subseteq R$ is an ideal that is contained in all maximal ideals of R (i.e. $\alpha \subseteq \text{Jac}(R)$). Then:

$$\text{if } \alpha M = M \text{ then } M = \{0\}$$

***Proof* :**
We will proceed by induction on the number of generators of M. If $n = 1$, then $\langle x_1 \rangle = M$. If $\alpha M = M$, we have that:

$$x_1 = a_1 x_1 \text{ for some } a_1 \in \alpha$$

we get that $(1 - a_1)x_1 = 0$, meaning $1 - a_1$ annihilates x_1.

However, $(1 - a_1)$ is in fact a unit (as will be shown in a moment). Since it's a unit, there exists a b such that

$$0 = b \cdot 0 = b(1 - a_1)x_1 = x_1$$

Thus it must be that $x_1 = 0$, showing that $M = \{0\}$. So it's left to show that $1 - a_1$ is a unit. Let's say $(1 - a_1)$ was a proper ideal. Then it must lie in some maximal ideal. Since $a_1 \in \alpha$ and α is a subset of all maximal ideals, it must be that $1 - \alpha_1 + \alpha_1 = 1$ is in a maximal ideal, however that would make the ideal a non-proper ideal – a contradiction.

For the case of $n = k$, let $x_1, x_2, ..., x_k$ generate M, and suppose $\alpha M = M$. Then

$$x_k = a_1 x_1 + a_2 x_2 + \cdots + a_k x_k$$

so $(1-a_k)x_k$ is generated by $k-1$ elements. Thus by the induction hypothesis, those elements must in fact generate the zero-module, and so

$$(1-a_k)x_k = 0$$

and by the same rasoning as before, it must be that $x_k = 0$, and so $M = \{0\}$, as we sought to show

If we restrict to $R = A$ being a local ring and $\alpha = m$ the maximal ideal of A, we will have the first step in constructing a basis for M over A:

Lemma 20.3.27: Nakayama's Lemma II

Let M be an A-module over a local ring A and let m be the maximal ideal, and let F be a submodule of M. Then:

$$\text{if } M = F + mM \text{ then } M = F$$

Proof **:**
Take $M/F = (F + mM)/F = mM/F$ so that $M/F = mM/F$. Then by lemma 20.3.26, $M/F = \{0\}$, so that means that M moded by F makes everything 0, and so

$$M = F$$

Finally, we put these two together to create create a basis for a module over a local ring using a basis for a smaller module!

Lemma 20.3.28: Nakayama's Lemma III

Let M be a A-module over a local ring A and m the maximal ideal of A. If $x_1, x_2, ..., x_n$ generated M/mM (M mod mM), then they also generators for M

Proof **:**
Let $F = \langle x_1, x_2, ..., x_n \rangle = M/mM$. Then

$$M = F + mM$$

Then by lemma 20.3.27, $M = F$, and so M has $\langle x_1, x_2, ..., x_n \rangle$ as generators, as we sought to show

Nakayama's lemma let's us reverse some constructions. As we saw in chapter 14, a [finitely generated?] projective module over a PID is free. Using Nakayama's lemma, we can show that a finitely generated module over a local ring is also free!

Theorem 20.3.29: Projective Module over Local Ring

Let M be a finitely generated projective module over a local ring A. Then M is free. In particular, If if $\{x_1, ..., x_n\}$ are elements of M who's residue class $\{\overline{x_1}, ..., \overline{x_1}\}$ form a basis for M/mM over A/m, then $x_1, ..., x_n$ is a *basis* for M over A.

Furthermore, if $\overline{x_1}, ..., \overline{x_n}$ are linearly independent over A/m, then this collection can be completed to a basis for M over A

***Proof* :**
First, Note that $M/\mathfrak{m}M$ over $A/\mathfrak{m}$ is indeed a vector space with action $(a + \mathfrak{m})(m + \mathfrak{m}M) = (am + \mathfrak{m}M)$ being well-defined.

Since M is finitely generated, let $\{x_1, x_2, ..., x_n\}$ be the generating set. Let F be a free module with basis $e_1, e_2, ..., e_n$, and consider the homomorphism $f : F \to M$, $f(e_i) = x_i$. We'll show that f is in fact an isomorphism. Since $x_1, ..., x_n$ generate M mod $\mathfrak{m}M$, by Nakayama's lemma III, they generate M¡ and so f is surjective. Next, since M is projective, f splits, i.e., $F = P_0 \oplus P_1$ where $P_0 = \ker(f)$ and P_1 and $P_1 \cong M$ through f. We'll show $P_0 = 0$.

Take some $\sum a_i x_i \in \ker(f)$ so that $\sum a_i x_i = 0$. Then reducing to the mod, we have $\sum a_i x_i + \mathfrak{m}M = 0 + \mathfrak{m}M = \sum (a_i \mathfrak{m}) \cdot (x_i + \mathfrak{m}M) = 0 + \mathfrak{m}M$. Since the $\overline{x}_i$ form a basis for $M/\mathfrak{m}M$ over $A/\mathfrak{m}A$, it must be that $a_i \in \mathfrak{m}$. Thus, we have $\sum a_i e_i \in \mathfrak{m}F$. Therefore:

$$\ker(f) \subseteq \mathfrak{m}F = \mathfrak{m}\ker(f) \oplus \mathfrak{m}M$$

Thus, $\mathfrak{m}\ker(f) = \ker(f)$. Since $\ker(f)$ is the direct summand of finitely generated modules, it is itself finitely generated. Thus, by Nakayama's lemma, $\ker(f) = 0$, and so f is an isomorphism! Therefore, we have $F \cong M$, and so M is free!

Notice that this immediately implies that the $x_1, ..., x_n$ are linealy independent since they form a basis for M over A (since f is an isomorphism and $f(e_i) = x_i$). Finally, if we extend $\overline{x}_1, ..\overline{x}_r$ to $\overline{x}_1, ..., \overline{x}_r, ...\overline{x}_n$ to be a basis for $M/\mathfrak{m}M$, then we just showed they will indeed form a basis for M.

A cool result of this theorem is a way to show some properties f based on f on the respective local modules. Let M be module over a local ring A. Then there is an associated module $M(\mathfrak{m}) = M/\mathfrak{m}M$ Then $__(\mathfrak{m})$ induces a functor mapping M to $M(\mathfrak{m})$ and

$$f : M \to N \quad \mapsto \quad f_{(\mathfrak{m})} : M(\mathfrak{m}) \to N(\mathfrak{m})$$

Notice that if f is surjective, then $f_{(\mathfrak{m})}$ is surjective. The magic of Nakayama's lemma is that the converse is true too!

Proposition 20.3.30: Local Homomorphisms

Let M, N be finitely generated modules over a local ring A and $f : M \to N$ be an A-module homomorphism. Then:

1. If f is surjective if and only if $f_{(\mathfrak{m})}$ is surjective
2. Assume f is injective. Then if $f_{(\mathfrak{m})}$ is surjective, f is an isomorphism
3. If M and N are free, f is injective (resp. an isomorphism) if and only if $f_{(\mathfrak{m})}$ is injective (resp. an isomorphism)

Found this page that proves the results quite nicely: http://homepage.divms.uiowa.edu/ acwood/-files/notes/nakayamaslemma.pdf

Proof **:**

1. Let $f : M \to N$ be an R-module homomorphism and let the inducd $f_{(m)}$ be surjective so that $f(M \bmod \mathfrak{m}M) = N \bmod \mathfrak{m}N$. Since N is finitely generated, so is any quotient of it, and so $N \bmod \mathfrak{m}N$ is finitely generated by $\{\overline{x}_1, \overline{x}_2, ..., \overline{x}_n\}$ By Nakayama's lemma, $\{x_1, x_2, ..., x_n\}$ generated N. Since $x_1, ..., x_n \in f(M)$, f is also surjective

   More consisely, since $f_{(\mathfrak{m})}$ is surjective, $N = \operatorname{im}(f) + \mathfrak{m}N$. Thus, by Nakayama's lemma, $N = \operatorname{im}(f)$.

2. here

3. Assume M and N are free and $f_{(\mathfrak{m})} M/\mathfrak{m}M \to N/\mathfrak{m}N$ is injective. Since M is free, let $\{x_1, ..., x_n\} \subseteq M$ so that $\{\overline{x}_1, ..., \overline{x}_n\}$ is a basis for $M/\mathfrak{m}M$. Then by theorem 20.3.29, $x_1, ..., x_n$ forms a basis for M. Suppose now that $f(a_1x_1 + \cdots + a_nx_n) = 0$. By definition of $f_{(\mathfrak{m})}$, $f_{(\mathfrak{m})}(\overline{a}_1\overline{x}_1 + \cdots + \overline{a}_n\overline{x}_n) = \overline{0}$. Since $f_{(\mathfrak{m})}$ is injective, we have $\overline{a}_1\overline{x}_1 + \cdots + \overline{a}_n\overline{x}_n = 0$, and since $\overline{x}_1, ..., \overline{x}_n$ is a basis, we have that the coefficients are all $\overline{0}$, that is $a_i \in \mathfrak{m}$.

   To show that $\ker(f) = 0$, we'll use Nakayama's lemma, so it suffices to show that $\ker(f) \subseteq \mathfrak{m}\ker(f)$. By what we've just shown, we have $\ker(f) \subseteq \mathfrak{m}M$. If $f_{(\mathfrak{m})}$ were surjective, then this proof is rather quick. By part (1), we have that f is surjective. Thus the following short exact sequence splits:
   $$0 \to \ker(f) \to M \to N \to 0$$
   since N is free, So $M \cong \ker(f) \oplus N$, and $\ker(f) \subseteq \mathfrak{m}M \cong \ker(f) \oplus N$, and so $\ker(f) \subseteq \mathfrak{m}\ker(f)$, and so by Nakayama's lemma $\ker(f) = 0$.

   If $f_{(\mathfrak{m})}$ is not surjective, then we need to do a little more work. We'll take advantage of the basis lifting property we proved in the previous theorem. Let $\{\overline{x}_1, ..., \overline{x}_n\}$ be a basis for $M/\mathfrak{m}M$ over $A/\mathfrak{m}A$ which induces a basis $\{x_1, .., x_n\}$ for M over A. $\langle f_{(\mathfrak{m})}(\overline{x}_1), ..., f_{(\mathfrak{m})}(\overline{x}_n)\rangle = \operatorname{im}(f_{(\mathfrak{m})})$ is a submodule of $N/\mathfrak{m}N$ which have that generating set as a basis (since $f_{(\mathfrak{m})}$ is injective). Thus, if:
   $$\overline{a}_1 f_{(\mathfrak{m})}(\overline{x}_1) + \cdots + \overline{a}_n f_{(\mathfrak{m})}(\overline{x}_n) = 0$$
   then since f is a homomorphism, we have
   $$f_{(\mathfrak{m})}(\overline{a}_1\overline{x}_1 + \cdots + \overline{a}_1\overline{x}_1) = 0$$

and since these elements form a basis and f is injective, $\overline{a}_i = 0$ for all i. Now, extend this to a basis for $N/\mathfrak{m}N$ over $A/\mathfrak{m}A$:

$$\{f_{(\mathfrak{m})}(\overline{x}_1), ..., f_{(\mathfrak{m})}(\overline{x}_n), \overline{y}_1, ..., \overline{y}_m\}$$

Since N is free, this basis lifts to a basis for N over A, in particular:

$$\{f(x_1), ..., f(x_n), y_1, ..., y_m\}$$

since $f_{(\mathfrak{m})}(\overline{x}_i) = f_{(\mathfrak{m})}(x_i + \mathfrak{m}M) = f(x_i) = \mathfrak{m}N$. With this, we can now get our conclusion: since the $f(x_i)$ are linearly independent, we have that:

$$a_1 f(x_1) + \cdots + a_n f(x_n) = f(a_1 x_1 + \cdots + a_n x_n) = 0$$

then it must be that the $a_i = 0$ for all the i, but then this is the condition for a function to be injective, and so f is injective, as we sought to show

commented region here:

Proposition 20.3.31

Let P and P' be A-module over a local ring A. Then $P \cong P'$ if and only if $P/\mathfrak{m}P \cong P'/\mathfrak{m}P'$.

Proof **:**
Any finitely generated projective module over a local ring is free, and so apply the previous proposition

There is a re-formulation of Nakayama's lemma using the tensor product, however I ommited it for now.

(Another application of Nakayama's Lemma)

Recall that if R is a finite integral domain, then it is necessarily a field (corollary 9.1.15). Naturally, this is not necessarily the case for infinite rings ($\mathbb{Z}$ being the prototypical example). We may thus take $Q = \text{Frac}(R)$, and consdier the R-module Q

Corollary 20.3.32: Infinite Generation Of Ring Of Fractions

Let $Q = \text{Frac}(R)$ be the ring of fractions of Q for an infinite ring R. Then considered as an R-module, Q is infinitely generated.

Proof **:**
Let's assume that Q was a finitely generated R module. Then $Q_{\mathfrak{m}}$ is a finitely generated $Q_{\mathfrak{m}}$. But then $\mathfrak{m}_{\mathfrak{m}} Q = Q$, and so Nakayama's lemma implies $F = 0$, a contradiction.

Recall that when dealing with infinite functions, we may have some strange results, like a surjection $f : M \to M$ that is not an injection. Nakayama's lemma insures we get some sanity here:

Corollary 20.3.33: Surjective Implies Injective

Let M be a finitely generated R-module. Then if $f : M \to M$ is a surjective R-module homomorphism, it is injecive.

Note how no assumption about R or M being Noetherian!!

Proof **:**
Need a slightly more general version of Nakayama's lemma, see coment in document, really cool!

20.4 Completion

The next type of construction we'll do is called a *completion*. These will be in the appropriate sense "stronger" then localizations, which will result in them being more restrictive, though also more powerful. A completion of a ring in many ways parallel's the completion of the rationals numbers $\mathbb{Q}$ to the real numbers $\mathbb{R}$. In fact, defining the appropriate metric, this process is a special case of the cauchy-sequence completion of the rational numbers to the reals. Namely, the completion of a ring is the equivalence class of all Cauchy-sequences which converge to the same point. Due to the topological background needed for this perspective, we shall relegate the equivalence of what shall be explicated in this section and Cauchy-sequences to section C.5.

There were rings that we have observed earlier that are completions of simpler rings; we will use completions to construct rings like the ring of *formal power series* $R[[x]]$ from $R[x]$, the ring $R[x_1, x_2, ...]$, the ring with *infinitely many monomial* from $R[x_1, ..., x_n]$, and the *p-adic integers* $\mathbb{Z}_p$ from $\mathbb{Z}$..

Completions will be handy for us in many ways. One important use that we will see when looking into algebraic geometry is looking at the "singularities" of schemes (or more particularly algebraic curves). Singularities will be points that don't be had as nicely as most other points, for example in the following image (which is the zero set of the equation $y^2 - x^2(x+1) = 0$) does not behave well at $(0, 0)$:

Figure 20.1: zero set of $y^2 - x^2(x+1) = 0$

This may seem similar to the geometric notion of localization, and this is indeed the case. Indeed, in the case of $\mathbb{Z}_{(p)}$ the p-adic integers are the completion of this ring, namely $\mathbb{Z}_p$ is the completion of $\mathbb{Z}_{(p)}$.

Completions shall also play a very important role in algebraic number theory. In this book, a more "classical" approach to algebraic number theory was taken, which shall give the reader great insights on the fundamental of algebraic number theory. However, a large part of *class number theory* shall rely on special completions of rings.

In this section, topological notion's will be introduced, and so an acquaintance of the material in appendix C is recommended. It should be noted that the topologies that will be discussed will be very different from that of Euclidean topologies, a new notion of size that is rather different from Euclidean distance shall be developed. The reader that is familiar with the construction of the real numbers using Cauchy sequences shall find it much easier to follow this section.

20.4.1 Constructing Completions

We will first define what is a "sequence". Let's say we have a collection of countable rings $R_1, R_2, ..., R_n$ and homomorphisms:

$$\cdots \xrightarrow{f_3} R_3 \xrightarrow{f_2} R_2 \xrightarrow{f_1} R_1$$

Then a sequence, or a *coherent sequence* is an element $(x_1, x_2, ...) \in \prod_{i=1}^{\infty} R_i$ where:

$$f_i(x_{i+1}) = x_i$$

If we take the special case of $I \subseteq R$ being an ideal and

$$\cdots \xrightarrow{f_3} R/(I^3) \xrightarrow{f_2} R/(I^2) \xrightarrow{f_1} R/I$$

then we will define a form of completion from these maps. For ease of notation, the sequence is usually written in the reverse order, as the reader shall understand once examples will be presented

Definition 20.4.1: Completion Of A Ring

Let R be a ring and $I \subseteq R$ be an ideal, and:

$$R/I \xleftarrow{f_1} R/(I^2) \xleftarrow{f_2} R/(I^3) \leftarrow \cdots$$

Then:

$$\hat{R} := \left\{ (a_1, a_2, ...,) \in \prod_{i=1}^{\infty} R_i \;\middle|\; f_i(a_{i+1}) = a_i \right\}$$

which is often denoted as:

$$\varprojlim R/I^n := \hat{R}$$

and is called the *limit* (or in older textbook the *inverse limit* or *projective limit*). The collection $\{R/I^n\}$ is called the *inverse system*, with the function usually taken as implicit

A couple of remarks are in order:

1. There is a natural topology and a metric in which these elements can actually be thought of as Cauchy sequences and $\hat{R}$ is indeed the completion of R. If $\{R/\mathfrak{a}^n R\}$ is an inverse system with

ideal $\mathfrak{a} \subseteq R$, and any $x \in R$, define a neighborhood basis (or local basis) $\{x + \mathfrak{a}^n R\}_{n=1}^{\infty}$. Then this forms a basis for a topology. This topology is called the $\mathfrak{a}$-*adic topology*, U is open in this topology if and only if for each $x \in U$, there exists a positive integer n such that

$$x + \mathfrak{a}^n R \subseteq U$$

Then R is Hausdorff if and only if $\cap \mathfrak{a}^n R = 0$. If this is the case, define the metric $d(x, y) = 2^{-n}$ whenever $x \neq y$ where n is an integer such that $x - y \in \mathfrak{a}^n R - \mathfrak{a}^{n-1} R$. The constant 2 does not matter, it can be replaced with any $C > 1$ and it will give the same topology.

This exact same process can be repeated on an arbitrary R-module, where the inverse system would be $\{M/a^n M\}$, and for a given ideal $I \subseteq R$, we would define the I-adic topology on M.

2. This is a special case of a more general construction of a *limit* in the category of **Ring**. Namely, it is the limit of a special form of the diagram:

$$\cdots \to \bullet \to \bullet \to \bullet$$

where each arrow has the coherence condition stated above.

3. By how we've defined it, any inverse system will be surjective. Such systems are sometimes *surjective systems*.

4. If R is Noetherian and integral, then $\bigcap_n I^n = 0$, and hence $R \subseteq \hat{R}$. For those who shall read appendix C.4, this shows that inegral Noetherian domains are Hausdorff in the I-adic topology.

Example 20.4.1: Completion of Rings

1. Let $R = k[x]$ and $I = (x)$. Then:

$$\varprojlim (k[x]/(x)^n) = k[[x]]$$

To see this, consider:

$$\cdots k[x]/(x^3) \to k[x]/(x^2) \to k[x]/(x) = k$$

Notice that in each $k[x]/(x^k)$, can be thought of the collection of all polynomial of degree up to $k - 1$. Hence, a coherent sequence $(a_1, a_2, ...)$ can be thought of as an "infinitely long polynomial" and would usually be denoted

$$\sum_k^{\infty} a_k x^k$$

Similarly, the reader may verify that $\widehat{k[x, y]} = k[[x, y]]$ with $I = (x, y)$

2. Let $R = \mathbb{Z}$ and $I = (10)$. Then $\varprojlim \mathbb{Z}/(10)^n$ is represented by all infinitely long numbers. Just like before, each $\mathbb{Z}/(10)^k$ can be thought of as as holding information up to the $k - 1$th position. So if we take:

$$Z/(10) \leftarrow Z/(100) \leftarrow \mathbb{Z}/(1000) \leftarrow$$

and construct a coherent chain, we may chose elements like 5, 25, 425, ... and so on. The reader may see that this would represent an infinitely long number. The resulting ring is

often denoted $\mathbb{Z}_{10}$ and is called the 10-adic integers. Any element of $\mathbb{Z}_{10}$ could formally be represented as

$$\sum_{k}^{\infty} a_k 10^k$$

3. Let $R_n = k[x_1, ..., x_n]$ be the polynomial in n variables. Then the completion is $k[x_1, ..., x_n, ...]$

Completions are a way to construct rings. It would be nice if exactness interact nicely with it. If $\{A_n\}$, $\{B_n\}$, $\{C_n\}$ are three inverse systems where:

$$\begin{tikzcd}
0 \arrow[r] & A_{n+1} \arrow[r, "\psi"] \arrow[d, "\alpha"] & B_{n+1} \arrow[r, "\varphi"] \arrow[d, "\beta"] & C_{n+1} \arrow[r] \arrow[d, "\gamma"] & 0 \\
0 \arrow[r] & A'_n \arrow[r, "\psi'"] & B'_n \arrow[r, "\varphi'"] & C'_n \arrow[r] & 0
\end{tikzcd}$$

Then we'll denote it as $0 \to \{A_n\} \to \{B_n\} \to \{C_n\} \to 0$. Then we naturally get the induced homomorphisms:

$$0 \to \varprojlim A_n \to \varprojlim B_n \to \varprojlim C_n \to 0$$

however this is not always exact. The following proposition shows us when it is:

Proposition 20.4.2: Limits and Exactness for inverse Systems

Let $0 \to \{A_n\} \to \{B_n\} \to \{C_n\} \to 0$ be an exact sequence of inverse systems. Then:

$$0 \to \varprojlim A_n \to \varprojlim B_n \to \varprojlim C_n$$

is always exact. If $\{A_n\}$ is a surjective system, then:

$$0 \to \varprojlim A_n \to \varprojlim B_n \to \varprojlim C_n \to 0$$

is always exact

Hence, when working with completions induced by surjective systems, the completion functor $\widehat{(-)}$ preserves injective and surjective functions[3].

Proof **:**
This is an application of Snakes lemma. Let $A = \prod_{n=1}^{\infty} An$, and

$$d^A : A \to A \qquad d^A(a_n) = a_n - f_{n+1}(a_{n+1})$$

Then by construction

$$\ker d^A = \varprojlim A_n$$

We may define B, C via d^B and d^C respectively. Now, as we have an exact sequence of inverse

[3]In part VII, this shall imply that the completion functor has trivial cohomology, while general inverse systems shall have non-trivial cohomology

systems, we get the following diagram:

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r] \arrow[d, "d^A"] & B \arrow[r] \arrow[d, "d^B"] & C \arrow[r] \arrow[d, "d^C"] & 0 \\ 0 \arrow[r] & A \arrow[r] & B \arrow[r] & C \arrow[r] & 0 \end{tikzcd}$$

Then by proposition 16.1.7 we get:

$$0- > \ker d^A \to \ker d^B \to \ker d^c \to \operatorname{coker} d^A \to \operatorname{coker} d^B \to \operatorname{coker} d^C \to 0$$

Finally, if the system is a surjective system, we can solve inuctively the equation:

$$x_n - f_{n+1}(x_{n+1}) = a_n$$

to show that d^A is surjective, completing the proof.

The following result shall be stated as it gives a very fascinating result, however for the full background the reader should read section C.5:

Corollary 20.4.3: Completion of Group Exact Sequences

Let $0 \to G' \to G \xrightarrow{p} G'' \to 0$ be an exact sequence of groups. Let G have a topology given by the subgroups $\{G_n\}$, and induce the topology on G', G'' via the subgroups $\{G'_n \cap G_n\}$ and $\{pG_n\}$. Then:

$$0 \to \hat{G}' \to \hat{G} \to \hat{G}'' \to 0$$

is exact

Proof **:**
Apply proposition 20.4.2 to

$$0 \to \frac{G'}{G' \cap G_n} \to \frac{G}{G_n} \to \frac{G''}{pG_n} \to 0$$

Corollary 20.4.4: Quotient in Completion

Let $\hat{G}_n \leqslant \hat{G}$. Then:

$$\hat{G}/\hat{G}_n \cong G/G_n$$

Proof **:**
exercise

As a consequence, in particular interest is the case of $\mathbb{Z}_{10}$ where

$$\mathbb{Z}_{10} \cong \mathbb{Z}_2 \oplus \mathbb{Z}_5$$

Hence $\mathbb{Z}_{10}$ is decomposeable to smaller rings. These rings are also in many ways nicer: notice that as a consequence of the above isomorhpism that $\mathbb{Z}_{10}$ has zero divisors (any product of two non-trivial

rings has zero-divisors)[4], showing that in general the completion may introduce zero-divisors, while $\mathbb{Z}_2$ and $\mathbb{Z}_5$ do not introduce zero-divisors. Furthermore, $\mathbb{Z}_p$ introduces more units (this was more evident with $k[[x]]$ where the units were explicitly found in proposition 9.2.8). For this reason, we would concentrate on the p-adic numbers $\mathbb{Z}_p$. Any element of $\mathbb{Z}_p$ could be represented as

$$\sum_k^\infty a_k p^k$$

We may generalize the result for finding units in completions:

Lemma 20.4.5: Units In Completion

Let R be a ring and $\hat{R}$ a completion of R via p. Then if $u \in R/p^n$ is a unit, $u \in R/p^{2n}$ is a unit.

Proof **:**
Let u be a unit so $1 + us \in p^n$. Squaring, we get:

$$1 + u(2s + us^2) \in p^{2n}$$

Hence, it is a unit in R/p^{2n}.

The reader should use this to show that u is a unit in $\hat{R}$. This also generalizes proposition 9.2.8, for this directly shows that any power series with nonzero constant is a unit. In the case where $p = \mathfrak{m}$ is a maximal ideal (as in the case for p-adic integers and $k[x]$), we get the following result:

Lemma 20.4.6: Completion At Maximal Ideals

Let R be a ring and $\mathfrak{m}$ a maximal ideal. Then $\varprojlim R/\mathfrak{m}^n$ is a local ring, that is it contains a single maximal field.

Proof **:**
Notice that all elements in R that are not in $\mathfrak{m}$ are units in $R/\mathfrak{m}$. Thus, by lemma 20.4.5 they are units in $\varprojlim R/\mathfrak{m}^n$.

Thus, there is a natural map:

$$R_\mathfrak{m} \to \varprojlim R/\mathfrak{m}^n$$

where $R_\mathfrak{m}$ is the localization. Recall that in $R_\mathfrak{m}$, the elements are constructed by taking formal inverses, hence completion at a maximal ideal shows a "dual" way of creating units by finding them in some way "at infinity". Note that it is important that $\mathfrak{m}$ be maximal, for this will fail if an ideal is only prime (however, one may localize the ring at the prime ideal to make it maximal and then take the completion).

When working with localizaiton, we saw that a localization of an integral domain is an integral domain. This is no longer true for completion:

[4] Can you find two explicit elements?

Example 20.4.2: Completion Looses Integral Domain
Let

$$R = \frac{k[x,y]}{(y^2 - x^3 - x^2)}$$

with k a field of characteristic 0. Then completing this ring, since completion is exact, we get:

$$\hat{R} = \frac{k[[x,y]]}{(y^2 - x^3 - x^2)} \cong \frac{k[[x,y]]}{(y - x\sqrt{1+x})(y + x\sqrt{1+x})}$$

where $\sqrt{1+x}$ is the short-hand for the power-series associated to $\sqrt{1+x}$. Then it is easy to see that this ring has zero-divisors. This shall eventually be made rather clear when a geometric perspective is taking. The ring R shall in a concrete sense have the curve $y^2 = x^3 + x^2$ associated to it:

Figure 20.2: $y^2 = x^3 + x^2$

This curve shall be shown to be *irreducible* in a concrete way. The localization shall represent this curve minus some finite amount of points. The completion at (x, y) shall represent "zooming in" at the point (x, y), at which point geometric object shall roughly correspond to:

Figure 20.3: Approximate Geometric picture corresponding to $\varprojlim R/(x,y)^n$

Then this shall be shown to be decomposable in a rigorous sense into two components (the two strokes of the "x" in the above picture).

Definition 20.4.7: Complete Ring and Module

Let R be a ring and M an R-mdoule. Then R is said to be *complete* if

$$R \cong \hat{R}$$

And hence $R \cong \varprojlim R_n$. M is said to be *complete* if

$$M \cong \hat{M}$$

where $\hat{M}$ is given by taking the limit of $G = M$ and $G_n = R_n \cdot M$. In this case, the ring is complete and we can say it is an $\hat{R}$-module.

Show that if we take $R_n = \mathfrak{a}^n$ for an ideal, then $f(\mathfrak{a}^n M) = \mathfrak{a}^n f(M) \subseteq \mathfrak{a}^n N$, and so any R-module homomorphisms lifts to an $\hat{R}$-module homomorphism, and similarity in the more general case.

20.4.2 Filtrations

The following construction is a more general version of an inverse system that drops many of the algebraic properties and is a more "pure topological" definition. It is being presented here for its use in the study of graded algebras, as well as it's use when defining *spectral sequences* in chapter 29

Definition 20.4.8: Filtration

Let M be an R-module. Then a *filtration* of M is an infinite chain:

$$M = M_0 \supseteq M_1 \supseteq \cdots \supseteq M_n \supseteq \cdots$$

where each M_n is a submodule of M. The chain is usually denoted (M_n). If furthermore $\mathfrak{a} \subseteq R$ is an ideal, then an $\mathfrak{a}$-filtration of M is a filtration st

$$\mathfrak{a}M_n \subseteq M_{n+1} \qquad \forall n \in \mathbb{N}$$

A filtration is called stable if for some $N \in \mathbb{N}$, for all $N \geq n$, $M_N = M_{N+1}$. A filtration is called $\mathfrak{a}$-stable if for some $N \geq n$, $\mathfrak{a}M_N = M_{n+1}$.

An $\mathfrak{a}$-stable filtration can be thought of as the generalization of a graded module which is in some way finite. The following lemma will be important to compare two $\mathfrak{a}$-stable filtrations:

Lemma 20.4.9: Comparing Stable $\mathfrak{a}$-Filtartions

Let $(M_n), (M'_n)$ be stable $\mathfrak{a}$-filtrations of M. Then they have bounded difference, that is, there exists an integer n_0 such that $M_{n+n_0} \subseteq M'_n$ and $M'_{n+n_0} \subseteq M_n$ for all $n \geq 0$.

This will imply that all stable $\mathfrak{a}$-filtrations determine the same topology on M (The $\mathfrak{a}$-topology)

Proof **:**
Let $M'_n = \mathfrak{a}M$. Then as $\mathfrak{a}M_N \subseteq M_{n+1}$, $\mathfrak{a}^n M \subseteq M_n$. Furthermore, $\mathfrak{a}M_n = M_{n+1}$ for all $n \neq n_0$ for appropriate n_0, and so $\mathfrak{a}^n M_{n_0} \subseteq \mathfrak{a}M$, completing the proof.

The key use of this lemma comes when defining the α-adic topology given in appendix C.4, namely all stable $\mathfrak{a}$-filtrations produce the same topology.

Exercise 20.4.1

1. Show that $\mathbb{Z}_p \cap \mathbb{Q} \cong \mathbb{Z}_{(p)}$
2. Show that $\hat{\hat{R}} \cong \hat{R}$

20.4.3 More on Graded Algebras

Graded algebras were initially explored in section 15.4. In this section, we expand further on graded-algebras, explore their relation to the Neotherian property, and look at some important associated graded algebras given an ideal I or module M. This exploration shall bear fruitful results that will be important in the development of *projective algebraic geometry* and *homological algebra*.

First, recall that a ring R is a graded ring if $R = \bigoplus_i R_i$ where $R_i R_j \subseteq R_{i+j}$, which makes R into an R_0-algebra, and each R_i is an R_0-module. A graded R-module is a module $M = \bigoplus_0^\infty M_i$ such that $R_m M_n \subseteq M_{m+n}$. If M is also a graded ring, then M is an R-graded algebra. Elements of M_n are called *homogeneous*, and an ideal $I \subseteq R$ is called homogeneous if $I = \bigoplus_i^\infty (R_i \cap I)$. A ring or

R-module homeomorphism is called graded if it respects the graded structure ($f(M_n) \subseteq N_n$). Each graded ring has an ideal $R_+ \bigoplus_{n>0} R_n$. The elements of R_n for each n are called the *homogeneous elements*.

Proposition 20.4.10: Noetherian Graded Rings

Let R be a graded ring. Then the following are equivalent:

($\Rightarrow$) R is Noetherian

($\Rightarrow$) R_0 is Noetherian and R is a finitely generated R_0 algebra

Proof **:**

(1) $\Rightarrow$ (2) As $R_0 \cong R/R_+$, R_0 is Noetherian. As $R_+ \subseteq R$ is an ideal of R, it is finitely generated, say $x_1, ..., x_s$. Then these elements can be used to construct homogeneous s element, let's use the same label and say they are of order $k_1, ..., k_s$ (all necessarily order greater than 0). Let $R' \subseteq R$ be the ring generated as an R_0 algebra over the homogeneous elements $x_1, ..., x_s$. Then I claim that $R = R'$, which will the proof. First, let' show $A_n \subseteq A'$ using induction. It is clear if $n = 0$. For $n > 0$, taking $y \in A_n$, as $y \in A_+$, y is a linear combination of the x_i

$$y = \sum_i^s a_i x_i$$

with $a_i \in A_{n-k_i}$ (and $A_m = 0$ of $m < 0$). Then by the inductive hypothesis, each a_i is a polynomial in the x's with coefficients in A_0. Hence, by the principle of induction, the same holds for y, we have that $A_n \subseteq A'$. As $A' \subseteq A$ by definition, we have $A = A'$, as we sought to show

(2) $\Rightarrow$ (1) Follow's from HBT

Any ring R induces a graded ring by taking $\mathfrak{a} \subseteq R$ and taking $R^* = \bigoplus_n^\infty \mathfrak{a}^n$. Similarly, any R-module M induces a graded R^*-module $M^* = \bigoplus_n M_n$ where M_n is the $\mathfrak{a}$-filtration. The "canonical" grading comes from $S = R[x_1, ..., x_n]$ with $\mathfrak{a} = (x_1, ..., x_n)$ which will produce the homogeneous grading of S.

In the case where R is Noetherian and $\mathfrak{a}$ is finitely generated (say by $x_1, ..., x_s$), then $R^* = R[x_1, ..., x_s]$ and is Noetherian by proposition 20.4.10.

Proposition 20.4.11: Topology of Noetherian Graded Module

Let R be a Noetherian ring, M a finitely generated R-module, (M_n) an $\mathfrak{a}$-filtration of M. Then the following are equivalent:

($\Rightarrow$) M^* is a finitely generated R^*-module

($\Rightarrow$) The filtration (M_n) is stable

Proof **:**
This is about creating the right conditions to see the filter is stable. Take $Q_n = \bigoplus_{r=0}^{n} M_r$. As each M_r is finitely generated, Q_n is finitely generated. Then Q_n forms a subgroup of M^* but not in general a R^*-submodule. We may however use it to generate one:

$$M_n^* = M_0 \oplus \cdots M_n \oplus \mathfrak{a}M_n \oplus \mathfrak{a}^2 M_n \oplus \cdots \oplus \mathfrak{a}^r M_n \oplus \cdots$$

Now, as Q_n is finitely generated as an R-module, M_n^* is finitely generated as an R^*-module, and the M_n^* form an ascending chain whose unions is M^*.

Finally, we now have that as R^* is Noetherian ,M^* is finitely generated as an R^*-module if and only if the chain stops ($M^* = M_{n_0}^*$) if and only if the filtration is table.

Corollary 20.4.12: Graded Noetherian then Component Noetherian

Let R be graded and Noetherian, and suppose M is a finitely generated graded R-module. Then M_n is a finitely-generated R_0-module for all $n \in \mathbb{N}$

Proof **:**
Let

$$d_j = \deg m_j$$

Then every element of $m \in M_n$ can be written as:

$$m = \sum_j f_j(x) m_j \qquad f_j(x) \in R$$

where f_j is homogeneous of degree $n - d_j$ (and is zero if $n < d_j$). Hence, M_n is finitely generated as an R_0-module generated by all $g_j(x)m_j$ where $g_j(x)$ is a monomial in x_i of total degree $n - d_j$, as we sought to show.

20.4.4 Artin-Rees Lemma

This is an excellent video: here

The Artin-Rees lemma is a central result that will allow us to compare the induced and $\mathfrak{a}$-adic topology on a submodule, giving us insight on their topologies.

Theorem 20.4.13: Artin-Rees Lemma

Let R be a Noetherian ring, $\mathfrak{a} \subseteq R$ an ideal, M a finitely generated R-module, and (M_n) a stable $\mathfrak{a}$-filtration of M. Then if $M' \subseteq M$ is a submodule of M, $(M' \cap M_n)$ is a stable $\mathfrak{a}$-filtration of M'

Proof **:**
As $\mathfrak{a}(M' \cap M_n) \subseteq \mathfrak{a}M' \cap \mathfrak{a}M_n \subseteq M' \cap M_{n+1}$, $(M' \cap M_n)$ is an $\mathfrak{a}$-filtration, and so it defies a graded R^*-module which is a submodule of M^*, and so it is finitely generated, letting us apply

by proposition 20.4.11 to completing the proof.

Corollary 20.4.14: Graded Sub-module, Computational Shortcut

et R be a Noetherian ring, $\mathfrak{a} \subseteq R$ an ideal, M a finitely generated R-module, and (M_n) a stable $\mathfrak{a}$-filtration of M. Then there exists a k such that

$$(\mathfrak{a}^n M) \cap M' = \mathfrak{a}^{n-k} \left((\mathfrak{a}^k M) \cap M' \right)$$

for all $n \geq k$

Proof **:**
exercise.

Theorem 20.4.15: Submodule Adic Topology Equivalence

Let R be a Noetherian ring, $\mathfrak{a} \subseteq R$ an ideal, M a finitely generated R-module, $M' \subseteq M$ a submodule. Then the filtrations

$$\mathfrak{a}^n M' \qquad \text{and} \qquad (\mathfrak{a}^n M) \cap M'$$

have bounded difference, hence the particular $\mathfrak{a}$-topology of m" coincides with the induced topology given by the $\mathfrak{a}$-topology of M

Proof **:**
Direct result of corollary 20.4.14 and theorem 20.4.13.

Corollary 20.4.16: Module Completion Preserves Exactness

Let

$$0 \to M' \to M \to M'' \to 0$$

be a short exact sequence of finitely generated R-modules over a Noetherian ring R. Then if $\mathfrak{a} \subseteq R$ is an ideal, the $\mathfrak{a}$-adic completion functor preserves exactness:

$$0 \to \hat{M}' \to \hat{M} \to \hat{M}'' \to 0$$

Proof **:**
Theorem 20.4.15 provides the condition needed to use corollary 20.4.3.

Next, as we have a natural homomorphism $R \to \hat{R}$, we can think of $\hat{R}$ as an R-algebra, and hence we have an extension of scalars of an R-module M given to an $\hat{R}$-module $\hat{R} \otimes_R M$. This gives a natural sequence of maps that we may want to understand:

$$\hat{R} \otimes_R M \to \hat{R} \otimes_A \hat{M} \to \hat{R} \otimes_{\hat{R}} \hat{M} = \hat{M}$$

In general, these maps are neither injective nor surjective. Under nicer conditions, we have a better control:

Proposition 20.4.17: Module Completion as Tensor Product

Let R be any ring. Then if$M$ is finitely generated,

$$\hat{R} \otimes_R M \twoheadrightarrow \hat{M}$$

is surjective. If furthermore R is Neotherian, then:

$$\hat{R} \otimes_R M \cong \hat{M}$$

Proof **:**
Let $F \cong R^n$ so that $M \cong F/N$ for appropriate N. The conclusion can be drawn from the following commutative diagram:

$$\begin{tikzcd}
 & \hat{R} \otimes_R N \arrow[r] \arrow[d, "\gamma"] & \hat{R} \otimes_R F \arrow[r] \arrow[d, "\beta"] & \hat{R} \otimes_R M \arrow[r] \arrow[d, "\alpha"] & 0 \\
0 \arrow[r] & \hat{N} \arrow[r] & \hat{F} \arrow[r, "\delta"] & \hat{M} \arrow[r] & 0
\end{tikzcd}$$

where the top line is given by the right-exactness of the tensor product, and the bellow line is given by corollary 20.4.3. Note that as completion commutes with finite direct sums (use corollary 20.4.3), we have that $\hat{R} \otimes_R F \cong \hat{F}$, and so α is surjective, giving us the first part. If R is Noetherian, hen N is finitely generated and so γ is surjective, and so we get exactness of the bottom line. Then by chasing the diagram we will get hat α is injective, and hence an isomorphism, completing the proof.

From this, we get that when R is Noetherian, the extension of scalar functor $\widehat{(-)}$ is equivalent to the functor $\hat{R} \otimes_R (-)$ from the category of R-modules to the category of $\hat{R}$-modules is exact. As a consequence:

Corollary 20.4.18: Flatness of Completion

Let R be a Neotherian ring, $\mathfrak{a} \subseteq R$ an ideal, $\hat{R}$ the $\mathfrak{a}$-adic completing of R. Then $\hat{R}$ is a flat R-algebra.

Proof **:**
direct result of above discussion.

Proposition 20.4.19: Properties of Adic Completion

Let R be Noetherian and $\hat{R}$ its $\mathfrak{a}$-adic completion. Then:

1. $\hat{\mathfrak{a}} = \hat{R}\mathfrak{a} = \hat{R} \otimes_R \mathfrak{a}$
2. $\widehat{\mathfrak{a}^n} = \hat{\mathfrak{a}}^n$
3. $\mathfrak{a}^n/\mathfrak{a}^{n+1} \cong \hat{\mathfrak{a}}^n/\hat{\mathfrak{a}}^{n+1}$
4. $\hat{\mathfrak{a}}$ is contained in the Jacobson radical of $\hat{R}$

***Proof* :**
exercise (or see Atiyah p.109)

We shall show the last one. By (2) and the fact that $\hat{\hat{R}} \cong \hat{R}$, we see that $\hat{R}$ is complete in the $\mathfrak{a}$-adic topology. Hence for any $x \in \hat{\mathfrak{a}}$,

$$(1-x)^{-1} = 1 + x + x^2 + \cdots$$

converges in $\hat{R}$, and so $1 - x$ i a unique. Then this implies that $\hat{\mathfrak{a}}$ is contained in the Jacobson radical of $\hat{R}$, completing the proof.

Let us next characterize the kernel of the map $M \to \hat{M}$

Theorem 20.4.20: Krull's Theorem

LEt R be a Noetherian ring, $\mathfrak{a} \subseteq R$ an ideal, M a finitely generated R-module, and $\hat{M}$ the $\mathfrak{a}$-completion of M, Then the kernel

$$E = \bigcap_{n=1}^{\infty} \mathfrak{a}^n M$$

given by the map $M \to \hat{M}$ consist of all $x \in M$ that are annihilated by some element $1 + \mathfrak{a}$.

***Proof* :**
It suffices to show that $(1-\alpha)E = 0$. IT is clear that if $(1-\alpha)x = 0$

$$x = \alpha x = \alpha^2 x = \cdots \in \bigcap_{n=1}^{\infty} \mathfrak{a}^n M = E$$

Hence we shall show the converse. As $E \bigcap \mathfrak{a}^n M$, it is the intersection of all neighborhoods of $0 \in M$. Thus, in the induced topology, E is the only neighborhood of $0 \in E$, and hence the only neighborhood in the $\mathfrak{a}$-adic topology. As $\mathfrak{a}E$ is a neighborhood in the $\mathfrak{a}$-topology, we have that $\mathfrak{a}E = E$. As M is finitely generated and R is Noetherian, E finitely generated, and hence we have that:

$$\mathfrak{a}E = E \Rightarrow (1-\alpha)E = 0$$

for some $\alpha \in \mathfrak{a}$, completing the proof.

Proposition 20.4.21: Localization is A Subring of Completion

Let R be Noetherian and let $S = 1 + \mathfrak{a}$. Then

$$S^{-1}R \hookrightarrow \hat{R}$$

Proof **:**
The map $R \to \hat{R}$ and $R \to S^{-1}R$ have the same kernel, and for any $\alpha \in \hat{\mathfrak{a}}$ we have

$$(1-\alpha)^{-1} = 1 = \alpha + \alpha^2 + \cdots$$

converges in $\hat{A}$, and hence all S become a unit in $\hat{A}$. By the universal property of $S^{-1}A$, there is a natural homomorphism $S^{-1}A \to \hat{A}$ which is injective by theorem 20.4.20.

The Noetherian condition is necessary, an easy example lies within the realm of analysis:

Example 20.4.3: Localization and Completion Without Noetherian Property
Take R to be the collection of smooth functions on $\mathbb{R}$, $C^\infty(\mathbb{R})$. Let $\mathfrak{a}$ be the ideal of all f which vanish at 0. This ideal is maximal as $R/\mathfrak{a} \cong \mathbb{R}$. Then $\mathfrak{a}$ is generated by x (if you know some multi-variable calculus prove that any f that vanishes at 0 can be written as $f(x) = xg(x)$ for g smooth, it is a variation of Hadamard's lemma), and $E = \bigcap_{n=1}^\infty \mathfrak{a}^n$ is the set of all $f \in R$ such that all derivatives vanish at the origin. On the other hand, f is annihilated by some element $1 + \alpha$ for $\alpha \in \mathfrak{a}$ iff f vanishes identically in some neighborhood of 0. Now, the function e^{-1/x^2} has vanishing derivative at 0, but is not identically zero near 0. Hence the kernels of the $R \to \hat{R}$ and $\mathbb{R} \to S^{-1}R$ with $S = 1 + \mathfrak{a}$ do not coincide.

This can also be a way of showing that R cannot be Noetherian.

Corollary 20.4.22: Krull Theorem, Noetherian Domain

Let R be a Noetherian domain and $(1) \neq \mathfrak{a} \subseteq R$ an ideal. Then

$$\bigcap \mathfrak{a}^n = 0$$

Proof **:**
Exercise; use the fact that $1 + \mathfrak{a}$ does not contain any zero-divisors.

Corollary 20.4.23: Krull Theorem, Hausdorff Topology

Let R be a Noetherian ring, $\mathfrak{a} \subseteq R$ an ideal contained in the Jacobson radical, and M a finitely generated R-module. Then the $\mathfrak{a}$-topology of M is Hausdroff, i.e.

$$\bigcap \mathfrak{a}^n M = 0$$

Proof **:**
By the Artin-Rees lemma, every element of $1 + \mathfrak{a}$ is a unit.

Corollary 20.4.24: Krull Theorem, Kernel Into Localization

Let R be a Noetherian ring, $\mathfrak{p} \subseteq R$ a prime ideal. Then the intersection of all $\mathfrak{p}$-primary ideals of A is the kernel of $A \to A_{\mathfrak{p}}$

Proof **:**
The $\mathfrak{p}$-primary ideals of A are the ideals contained in between $\mathfrak{p}$ and $\mathfrak{m}^n$. Corollary 20.4.23 implies that the intersection of all $\mathfrak{p}$-primary ideals is 0. Then we use the correspondence between completions and localization by first lifting R from it's localization and applying the map.

20.4.5 Associated Graded Rings: Noetherian Closed under Completion

The following will let us show another way of characterizing complete rings, and let us prove that the completion of a Noetherian ring is Noetherian.

Definition 20.4.25: Associated Graded Ring

Let R be a ring and $\mathfrak{a} \subseteq R$ an ideal. Then the *associated graded ring* is

$$G_{\mathfrak{a}}(R) = \bigoplus_{n=1}^{\infty} \mathfrak{a}^n/\mathfrak{a}^{n+1}$$

where $\mathfrak{a}^0 = A$. If unambiguous, $G(R)$ will be written instead of $G_{\mathfrak{a}}(R)$. For an R-module M with an $\mathfrak{a}$-filtration (M_n), the graded $G(R)$-module is given by:

$$G(M) = \bigoplus_{n=0}^{\infty} M_n/M_{n+1} = \bigoplus_{n=0}^{\infty} G_n(M)$$

Multiplication is certainly well-defined: for each $x_n \in \mathfrak{a}^n$, $\overline{x}_n$ is the image of x_n in $\mathfrak{a}^n/\mathfrak{a}^{n+1}$. Then $\overline{x}_m\overline{x}_m$ is $\overline{x_m x_n}$ which is the image of $x_m x_n$ in $\mathfrak{a}^{m+n}/\mathfrak{a}^{m+n+1}$ (if the reader is not convinced, check to see if this is well-defined).

Proposition 20.4.26: Properties of Associated Graded Rings

Let R be a Noetherian ring and $\mathfrak{a} \subseteq R$ an ideal. Then:

1. $G_{\mathfrak{a}}(R)$ is Noetherian
2. $G_{\mathfrak{a}}(R)$ is isomorphic to $G_{\hat{\mathfrak{a}}}(\hat{R})$ as graded rings
3. If M is a finitely generated R-module and (M_n) is a stable $\mathfrak{a}$-filtration of M, then $G(M)$ is a finitely generated graded $G_{\mathfrak{a}}(R)$-module

Proof **:**

1. Use HBT in the right way
2. Use proposition 20.4.19(3)
3. exercise

Lemma 20.4.27: Associated Graded Ideal and Completion Map Relation

Let $\varphi : A \to B$ be a homomorphism of filtered groups ($\varphi(A_n) \subseteq B_n$), and let $G(\varphi) : G(A) \to G(B)$, $\hat{\varphi}\hat{A} \to \hat{B}$ be the induced homomorphisms of the associated graded group and complete group respectively. Then:

1. $G(\varphi)$ is injective $\Rightarrow \hat{\varphi}$ is injective
2. $G(\varphi)$ is surjective $\Rightarrow \hat{\varphi}$ is surjective

Proof **:**
The key is to look at the following commutative diagram of exact sequences:

$$\begin{tikzcd}
0 \arrow[r] & A_n/A_{n+1} \arrow[r] \arrow[d, "G_n(\varphi)"] & A/A_{n+1} \arrow[r] \arrow[d, "\alpha_{n+1}"] & A/A_n \arrow[r] \arrow[d, "\alpha_n"] & 0 \\
0 \arrow[r] & B_n/B_{n+1} \arrow[r] & B/B_{n+1} \arrow[r] & B/B_n \arrow[r] & 0
\end{tikzcd}$$

which by the snake lemma gives us the usual:

$$0 \to \ker G_n(\varphi) \to \ker \alpha_{n+1} \to \ker \alpha_n \to \operatorname{coker} G_n(\varphi) \to \operatorname{coker} \alpha_{n+1} \to \operatorname{coker} \alpha_n \to 0$$

Now, by induction on n, we see that:

1. $\ker \alpha_n = 0$
2. $\operatorname{coker} \alpha_n = 0$

taking the limit of the homomorphisms α_n and applying proposition 20.4.2 completes the proof.

Proposition 20.4.28: Partial Converse of lemma 20.4.27(3)

Let R b a ring, $\mathfrak{a} \subseteq R$ an ideal of R, M an R-module, and (M_n) an $\mathfrak{a}$-filtration of M. Suppose that R is complete in the $\mathfrak{a}$-topology, that M is Hausdroff in its filtration topology so that $\bigcap_n M)n = 0$, and that $G(M)$ is a finitely generated $G(R)$-module.

Then M is a finitely generated R-module.

Proof **:**
As $G(M)$ is finitely generated, pick some finite set of generators. Split them up into the homogeneous components, say ζ_i (with $1 \leq i \leq \nu$) where ζ_i has degree $n(i)$, and is the image of some $x_i \in M_{n(i)}$. Let F^i be the module with the stable $\mathfrak{a}$-filtration given by $F^i_K = \mathfrak{a}^{k+n(i)}$ and let $F = \bigoplus_{i=1}^r F^i$. Then the mapping that takes the generator 1 of each F^i to x_i gives a homomorphism of filtered groups

$$\varphi : F \to M$$

Hence it induces a homomorphism of $G(A)$-modules. By construction this map is surjective. Hence by lemma 20.4.27 $\hat{\varphi}$ is surjective. Now consider the diagram:

$$\begin{tikzcd} F \arrow[r, "\varphi"] \arrow[d, "\alpha"'] & M \arrow[d, "\beta"] \\ \hat{F} \arrow[r, "\hat{\varphi}"] & \hat{M} \end{tikzcd}$$

As F is free and $A = \hat{A}$, α is an isomorphism. Next, as M is Hausdorff, β is injective. . As $\hat{\varphi}$ is surjective, φ is surjective, and so $x_1, ..., x_r$ generate M as an A-module completing the proof.

Corollary 20.4.29: Partial Converse in Noetherian Case

Let R b a ring, $\mathfrak{a} \subseteq R$ an ideal of R, M an R-module, and (M_n) an $\mathfrak{a}$-filtration of M. Suppose that R is complete in the $\mathfrak{a}$-topology, that M is Hausdroff in its filtration topology so that $\bigcap_n M)n = 0$, and that $G(M)$ is a finitely generated $G(R)$-module.

Then if $G(M)$ is a Noetherian $G(R)$-module, then M is a Noetherian R-module.

Proof **:**
If we show that every submodule $M' \subseteq M$ is finitely generated, we're done. Let $M'_n = M" \cap M_n$. Then (M'_n) is an $\mathfrak{a}$-filtration of M', and the embedding $M'_n \to M_N$ induces an injective homomorphism:

$$M'_n/M'_{n+1} \to M_n/M_{n+1}$$

Hence we gt a embedding of $G(M')$ in $G(M)$, $G(M)$ is Noetherian, $G(M')$ is finitely generated, and so M' is Hausdroff. As $\cap M'_n \subseteq \cap M_N = 0$, we get that M' is finitely generated, as we sought to show.

Theorem 20.4.30: Noetherian Closed Under Completion

Let R be a Noetherian ring and $\mathfrak{a} \subseteq R$ an ideal. Then the $\mathfrak{a}$-completion $\hat{R}$ of R is Noetherian

Proof **:**
As:

$$G_{\mathfrak{a}}(R) = G_{\hat{\mathfrak{a}}}(\hat{R})$$

is Noetherian, apply corollary 20.4.29 to the complete ring $\hat{R}$ and taking $M = \hat{R}$.

Corollary 20.4.31: Power Series Ring is Noetherian

Let R be a Noetherian ring, and let $S = R[[x_1, ..., x_n]$ be the power series ring in n variable.s then S is Noetherian.

Proof **:**
The ring $R[x_1, ..., x_n]$ is Noetherian by Hilbert's Basis Theorem, and S is the completion for the $(x_1, ..., x_n)$-adic topology.

20.5 Ring Extension and Integral Extension

In this section, we generalize the notion of extension from fields to commutative rings. Just like how $k \subseteq F$ can be seen as a field extension, if we let k and R be rings instead, we will et the accompying notion of ring extension:

Definition 20.5.1: Ring Extension

Let S be a commutative ring and $R \subseteq S$ a subring. Then we say that S is a *ring extension* of R. Equivalently, if there exists an injective ring homomorphism $\varphi : R \to S$, then S is a ring extension of R.

Unlike fields, we must specify the homomorphism to be injective, since ring homomorphisms need not be injective. The notation S/R is also used just as in field, though more scarsely. Just like how any field extension F/k makes F into a k-vector-space and we have special k-homomorphisms letting k-commute between the homomorphism, a ring extension S/R have an analogue: in fact, a k-homomorphism is *precisely* a k-algebra homomorphism!

Recall that any *R-algebra* can be defined as a ring homomorpihsm $\varphi : R \to S$, *provided* that $\varphi(R) \subseteq Z(S)$. Since we are working with commutative rings, this is *always* the case, and so we can always thing of a ring extension S over R is an R-algebra! This is why we focus on *commutative rings*. There is a similar notion of ring extension for non-commutative which is closer to the notion of group extension (i.e. a short exact sequence extending the object), however as we have pointed out in remark ringExtRemark, ring extensions don't break down into nice building blocks since kernels of rings are not rings, which means we would have to extend a ring R by an abelian group (i.e. two-sided ideal) I. This very much changes how we would study rings. However, we can always study the structure of any commutaive ring S by how behaves as an R-algebra similarly to how we studied fields, which is what we'll do in these chapters.

Perhaps the first notion we might want to genelize is the equivalent of an *algebraic element*, since they capture the relations within a ring extension with respect to the base ring. The following is the exactly this generlization:

Definition 20.5.2: Integral Elements

Let $R \subseteq S$.

1. An element $s \in S$ is called an *integral over* R if s is the root of a monic polynomial over $R[x]$:
$$s^n + r_{n-1}s^{n-1} + \cdots + r_1 s + r_0 = 0 \qquad r_i \in R$$
2. S is called *integral over* R or an *integral extension of* R if every element of S is integral over R. The integral closure of an integral domain is also called the *normalization of* R. Sometimes, the integral extension will be written as S/R
3. The set $\overline{R}^S$ containing all the elements of S that are integral over R is called the *integral closure of* R
4. If $R = \overline{R}^S$, then R is said to be *integrally closed over* S.
5. If R is an integral domain that is integrally closed in it's field of fractions F, $\overline{R}^F$, then R is called *normal domain*.

The fact that the polynomial it satisfies has to be monic can be thought of as not wanting to add relations that produce "units" or unlit-like elements (ex. $3x - 1 = 0$ implies there is a new element that behaves like $1/3$). For a more specific motivation, we can think of the following: If $f(x) \in R[x]$ is a polynomial of minimal degree with root r, and $h(x) \in R[x]$ is another monic polynomials with minimal degree with root r, then we would want that $h(x)|f(x)$, just like for minimal polynomials over fields. In the process of doing so, we would require the coefficient of the leading terms to divide. Since they can be arbitrary, we would require the coefficients of these polynomial to be units, or just 1, in order for the division to follow through. It follows that it is a natural condition that all polynomials be monic for an integral extension.

Another motivation comes from Gauss's lemma, which states that a polynomial $R[x]$ is irreducible iff it is irreducible in $k[x]$, granted that gcd of the coefficients of $p(x)$ is 1, or more simply that the leading coefficient is 1. The parallel between factorization in $k[x]$ and $R[x]$ will become very important in chapter 22.

Integral elements share many of the same properties as algebraic elements:

Proposition 20.5.3: Integral Element Equivalences

Let $R \subseteq S$ be a subring of S. Then the following are equivalent:

1. s is integral over R
2. $R[s]$ is a finitely generated R-module
3. The element s is in T for some subring T where $R \subseteq T \subseteq S$, that is a finitely generated R-module

***Proof* :**

$1 \Rightarrow 2$ Recall that $R[s]$ can be expressed as all R-linear combinations of powers of s. Since sd is integral over R:

$$s^n + r_n s^{n-1} + \cdots + r_1 s + r_0 = 0$$

in particular:

$$s^n = -(r_n s^{n-1} + \cdots + r_1 s + r_0)$$

It follows that for all $m \geq n$, s^m can be expressed in terms of $1, s_1, ..., s_{n-1}$ (for the same reasons as we've shown for algebraic elements). Hence:

$$R[s] = R + Rs + \cdots + Rs^{n-1}$$

Notice here how if the polynomial was not monic, we might not have been able to isolae s^n, causing complications.

$2 \Rightarrow 3$ Let $T = R[s]$. Then the reuslt immediately follows

$3 \Rightarrow 1$ We take advantage of rational cannonical forms and Cramer's rule (see ref:HERE) (remember that if R is an integarl domain, we can take it's field of fractions, and the resulting matri multiplication remains the same). Since T is finitely generated, let $v_1, v_2, ..., v_n$ be a finite generating set. Since $s \in T$ and T is a ring, $sv_i \in T$ for all $1 \leq i \leq n$. Thus:

$$sv_i = \sum_j a_{ij} v_j$$

Giving us the equation n:

$$\sum_j (\delta_{ij} s - a_{ij}) v_j = 0 \qquad 1 \leq i \leq n$$

where δ_{ij} is one when $i = j$ and 0 otherwise (i.e. the Kronecker delta). Letting B be the matrix such that the entries are $B_{ij} = \delta_{ij} s - a_{ij}$ and letting v be the $n \times 1$ matrix whose entries are $v_1, ..., v_n$, we get:

$$Bv = 0$$

Thus, by Cramer's rule (ref:HERE):

$$(\det B) v_i = 0 \qquad 1 \leq i \leq n$$

Now, since 1 is an R-linear combination of $v_1, v_2, ..., v_n$,

$$\det B = (\det B) 1 = (\det B) \sum_i a_i v_i = \sum a_i ((\det B) v_i) = 0$$

i.e. $\det B = 0$. Now, B is by definition $B = sI - A$ (where $A_{ij} = a_{ij}$). Thus, s is the root of a monic polynomial $\det(xI - A) \in R[x]$, namely the characteristic polynomial of A, and so is the root of a monic polynomial with coefficients in R, giving (1)

Corollary 20.5.4: Integral Elements Form Rings

Let $R \subseteq S$ be a subring of S, and $s, t \in S$. Then:

1. If s and t are integral over R, so is $s \pm t$ and st
2. $\overline{R}^S$ is a subring of S containing R, and is integrally closed[a]

[a] If we were working over rng's, $\overline{R}^S$ need not be integrally closed

Proof **:**
Let s, t be integral over R, sy that by proposition 20.5.3 $R[s]$ and $R[t]$ are finitely generated R-modules, say:

$$R[s] = Rs_1 + \cdots + Rs_n$$
$$R[t] = Rt_1 + \cdots + Rt_m$$

Then:

$$R[s,t] = Rs_1t_1 + \cdots + Rs_it_j + \cdots + Rs_nt_m$$

is a ring containing $s \pm t$ and st that is also a finitely generated R-module. It follows that The $\overline{R}^S$ is a subring which contains R, and is trivially integrally closed.

(note sure if I want the comment about "the process of taking integral closure stops after one step because integraity is transitive" that dummit puts on p. 693)

Note that integral closure is dependent on the ring over which we are closing; we do not have the equivalent notion of "integral closure of R" without specifying a ring over which it is integrally closed (Why?)

Corollary 20.5.5: Relating Finite Type And Finite Degree

Let S/R be an integral extension. Then S is a finitely generated R-module if and only if S is finite type over R and integral

Notice how if S is not integral, then it is possible that S is of finite type over R but not finitely generated over R, the most cannonical example being $R[x]$ over R (since $R[x]$ is finitely generated as an R-algebra by 1 and x, but $1, x, x^2, ...$ is an infinite set that is a basis for $R[x]$)

Proof **:**
If S is finitely generarted as an R-module, it is certainly finite type over R, so we'll show the converse (which is a generalization of theorem 17.3.8). Take Since S is of finite type, let's say $s_1, s_2, ..., s_n$ generates S as a ring. Take

$$R \subseteq R[s_1] \subseteq R[s_1, s_2] \cdots \subseteq R[s_1, s_2, ..., s_n]$$

Then each extension is a finitely generated field extension. Then (show as an exercise), $R[s_1, ..., s_n]$ is finitely generated over R.

Before giving examples, we'll show that we have been studying special types of integral closures when we worked with field exensions in part IV:

Proposition 20.5.6: Integral Extension over Fields

Let S/R be an integral extension where S is an integral domain. Then R is a field if and only if S is a field

Note that if S was not an integral domain or commutative, this theorem does not work: Take $M_2(k)$ over any field. Then $M_2(k)$ is a 4-dimensional k-algebar, but certianly not a field (it contains zero-divisors, and is not commutative).

Proof **:**

($\Rightarrow$) Let R be a field and take $0 \neq s \in S$. Since s is integral over R, we have:

$$s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0 = 0 \qquad a_i \in R$$

Without loss of generality, let $a_0 \neq 0$; otherwise, factor an s, and using the fact that S is an integral domain, we see that either $s = 0$ or $s^{n-1} + \cdots + a_1 = 0$, and since $s \neq 0$, then we simply choose $s^{n-1} + \cdots + a_1 = 0$ to be our equation. Thus, moving a_0 and factoring s we get:

$$s(s^{n-1} + \cdots + a_1) = -a_0$$

Since R is a field, $\frac{-1}{a_0} \in R$, so:

$$s \cdot \big((-1/a_0)(s^{n-1} + \cdots + a_1)\big) = 1$$

showing that s has a multiplictative inverse. Since s was arbitrary, we have that S is a field.

($\Leftarrow$) Let S be a field and take $0 \neq r \in R$ (automatically implying that R is an integral domain). Since $r \in R$, $r \in S$, and so $r^{-1} \in S$. Since S is an integral extension of R:

$$r^{-n} + a_{n-1}r^{-(n-1)} + \cdots + a_1 r^{-1} + a_0 = 0$$

Multiply the entire expression by $r^{-(n+1)}$, and move all the terms except the r^{-1} terms to the right hand side to get:

$$r^{-1} = -(a_{n-1} + \cdots + a_1 r^{n-2} + a_0 r^{n-1})$$

which is a linear combination of elements in R, showing that $r^{-1} \in R$, and so R is a field.

Thus, any algebraic field extension is an integral extension.

Example 20.5.1: Integral Extensions

1. As we've just shown, all algebraic extensions of fields are integral extensions.
2. Let R be a Unique Factorization Domain and let $F = \text{Frac}(R)$. Then $\overline{R}^F = R$. The $\supseteq$ direction is obvious, so consider the $\subseteq$ direciton: suppose that $a/b \in \overline{R}^F$ with $\gcd(a, b) = 0$.

Then by definition

$$(\frac{a}{b})^n + r_{n_1} + (\frac{a}{b})^{n-1} + \cdots + r_1\frac{a}{b} + r_0 = 0$$

Then:

$$a^n = b(-r_{n-1}a^{n-1} - \cdots - r_1ab^{n-2} - r_0b^{n-1})$$

so $b|a^n$, and hence $b|a$. Since $\gcd(a,b) = 1$, it must thus be that b is a unit, and hence we have that $a/b \in R$, showing that $\overline{R}^F = R$. Hence, all UFD's are normal domains. This should make intuitive sense, since any element $a/b \in \text{Frac}(R)$ would satisfy the equation $bx - a = 0$, which is not-monic.

More generally, any polynomial ring over a UFD R, $R[x_1, ..., x_n]$ is integrally closed over it's field of frac $R(x_1, ..., x_n)$ (by theorem 11.2.3), showing every polynomial ring over an Unique Factorization Domain is a normal domain.

3. Let S be integral over R and $I \subseteq S$ be an ideal. Then S/I is integral over $R/(R \cap I)$. To see this, notice that a monic polynomial over R where $s \in S$ is a root, when reduced modulo I gives a monic polynomial over $R/(R \cap I)$ where $\overline{s} \in S/I$ is a root.

4. Let k be any field and take $k[x, y]$ (which is integrally closd since it's a UFD). Take the prime ideal $(x^2 - y^3) \subseteq k[x, y]$ (see exercise ref:HERE, 14 in section 9.1 in Dummit and Foote). Consider the integral domain

$$R = k[x, y]/(x^2 - y^3) = k[\overline{x}, \overline{y}]$$

Notice that R is *not* normal: the element $\overline{x}/\overline{y}$ is the solution to the equation

$$\left(\frac{\overline{x}}{\overline{y}}\right)^3 - \overline{x} = 0$$

but evidentily $\overline{x}/\overline{y} \notin R$. Thus, R cannot be a Unique Factorization Domain.

If you wish to practice, another example is $\mathbb{Z}[\sqrt{-3}]$. It's integral closure over it's field of fraction's is:

$$\mathbb{Z}\left[\frac{1+\sqrt{3}}{2}\right]$$

20.5.1 Extension and Contraction of Ideals

Probably the biggest difference between fields and rings is that ideals are nontrivial. Thus, it is meaningful to ask what happens to ideals in (general) ring extensions. Recall that if $f : R \to S$ is a ring homomorphism, and $J \subseteq S$ is an ideal, $f^{-1}(S)$ is an ideal (re-prove this if you're not convinced). Howver, if $I \subseteq R$ is an ideal, $f(I)$ need not be an ideal (take $\mathbb{Z} \hookrightarrow \mathbb{Q}$ with any non-trivial ideal of $\mathbb{Z}$). The next best thing would be to define the smallest ideal of B containing $f(I)$:

Definition 20.5.7: Extension and Contraction

Let $\varphi : R \to S$ be a homomorphism between commutative rings and let $I \subseteq R$, $J \subseteq S$.

1. The *extension* of I to S is the ideal $\varphi(I)S$, commonly denoted I^e
2. The *contraction* of J in R is the ideal $\varphi^{-1}(J)$, commonly denoted I^c

In the case where φ is the natural inclusion $\iota : R \to S$, then the extension of I is IS, and the contraction of J is $J \cap R$. If it is clear that $I \subseteq R$, there are multiple extension of R, say S/R and T/R, then the extension of I is sometiems denoted I_S or I_T to emphasize to what ring I is being extended too S or T respectively.

Proposition 20.5.8: Properties of Extensions and Contractions

1. $I \subseteq IS \cap R$, or more generally that I is contained in the contraction of its extension to S ($I \subseteq I^{ec}$)
2. $(J \cap R)S \subseteq J$, or more genrally that J is contained in the extension of its contraction in R ($J \supseteq J^{ce}$)
3. $I^e = I^{ece}$
4. $J^c = J^{cec}$
5. If J is prime, J^c is prime. If I is prime, I^e need not be prime (consider $\mathbb{Z} \hookrightarrow \mathbb{Q}$). If $\varphi : R \to S$ is surjective, then I^e is prime.
6. If J is maximal, J^c need not be maximal, and similarly if I is maximal then I^e need not be maximal.

Proof **:**
exercise

Lemma 20.5.9: Extension And Contraction With Common Operations

Let $\mathfrak{a}_1, \mathfrak{a}_2 \subseteq A$ be ideals of A and $\mathfrak{b}_1, \mathfrak{b}_2 \subseteq B$ be ideals of B. Then:

$$\begin{aligned}
(\mathfrak{a}_1 + \mathfrak{a}_2)^e &= \mathfrak{a}_1^e + \mathfrak{a}_2^e & (\mathfrak{b}_1 + \mathfrak{b}_2)^c &\supseteq \mathfrak{b}_1^c + \mathfrak{b}_2^c \\
(\mathfrak{a}_1 \cap \mathfrak{a}_2)^e &\subseteq \mathfrak{a}_1^e \cap \mathfrak{a}_2^e & (\mathfrak{b}_1 \cap \mathfrak{b}_2)^c &= \mathfrak{b}_1^c \cap \mathfrak{b}_2^c \\
(\mathfrak{a}_1\mathfrak{a}_2)^e &= \mathfrak{a}_1^e\mathfrak{a}_2^e & (\mathfrak{b}_1\mathfrak{b}_2)^c &\supseteq \mathfrak{b}_1^c\mathfrak{b}_2^c \\
(\mathfrak{a}_1 : \mathfrak{a}_2)^e &= (\mathfrak{a}_1^e : \mathfrak{a}_2^e) & (\mathfrak{b}_1 : \mathfrak{b}_2)^c &\subseteq (\mathfrak{b}_1^c : \mathfrak{b}_2^c) \\
\sqrt{\mathfrak{a}}^e &\subseteq \sqrt{\mathfrak{a}^e} & \sqrt{\mathfrak{b}}^c &= \sqrt{\mathfrak{b}^c}
\end{aligned}$$

Proof **:**
exercise.

If $\varphi : R \to S$ is a ring homomorphism, we may decompose φ into the following:

$$R \xrightarrow{p} \varphi(R) \overset{j}{\hookrightarrow} S$$

where p maps R surjectively onto $\varphi(R)$, and j injects into S. Relating the ideals of R to the ideals of $\varphi(R)$ is very easy:

Proposition 20.5.10: Correspondence Between Ideals And Image Of Ideals

Let $\varphi : R \to S$ be a ring homomorhpism. Then there exists a bijective correspondence:

$$\{I \subseteq \varphi(R)\} \underset{J^c}{\overset{I^e}{\rightleftarrows}} \{\ker(\varphi) \subseteq J \subseteq R\}$$

Furthermore, prime ideals of $\varphi(R)$ correspond to prime ideals of R.

Proof **:**
exercise

On the otherhand, the correpsondence of $\varphi(R)$ and S is much harder, and shall be covered in secction 20.5.2 when S/R is an integral extension. As shown in proposition 20.5.8, every prime ideal "lies over" another prime ideal. This motivates the following vocuabilary:

Definition 20.5.11: Lying Over

Let $I \subseteq S$ be an ideal and S/R an integral extension. If $J = I \cap R$, I is said to be *lying over* J.

Example 20.5.2: Extending ideals
Take $\varphi : \mathbb{Z} \to \mathbb{Z}[i]$ to be the inclusion. Then a prime ideal $(p) \subseteq \mathbb{Z}$ may or may not remain prime when extended to $\mathbb{Z}[i]$. Using Fermat's theorem of sums of squares (theorem 10.3.3), it can be deduced that:

1. $(2)^e = (1+i)^2$
2. If $p \equiv 1 \bmod 4$, then p^e is the product of two distinct prime ideals (ex. $(5)^e = (2+i)(2-i)$)
3. if $p \equiv 3 \bmod 4$, then p^e is prime in $\mathbb{Z}[i]$

In general, it is rather hard to determine when $\varphi(p)$ is prime when p is prime. If $\varphi : R \to S$ is an integarl extension, then the situation becomes much better, however it will require the notion of *localization* which shall be covered in the next section. One special case is note mentioning:

Proposition 20.5.12: Maximal Ideals Lying Over

Let S/R and $\mathfrak{m} \subseteq S$ be an ideal. Then $\mathfrak{m} \cap R$ is maximal if and only if $\mathfrak{m}$ is maximal

Proof **:**
Consider the integral domain $S/\mathfrak{m}$. Then $S/\mathfrak{m}$ is an integral extension of $R/(\mathfrak{m} \cap R)$. Then by proposition 20.5.6, $S/\mathfrak{m}$ is a field if and only if $R/(\mathfrak{m} \cap R)$ is a field. But then $\mathfrak{m}$ is maximal if and only if $\mathfrak{m} \cap R$ is maximal.

20.5.2 Integral Extensions and Prime Ideals

We return to the question of the extension of ideals, in particular prime ideals, given S/R:

Theorem 20.5.13: Integral Extension Prime Correspondence (Cohen–Seidenberg Theorem)

Let S be an integral extension of the integral domain R

1. (**Lying-over Property**) Let $\mathfrak{p} \subseteq R$ be a prime ideal. Then there is a prime ideal $P \subseteq S$ such that $\mathfrak{p} = P \cap R$. Furthermore, $\mathfrak{p}$ is maximal if and only if P is maximal.
2. (**Going-up Property**) Let $p_1 \subseteq p_2 \subseteq \cdots \subseteq p_n$ be a chain of prime ideals of R, and suppose there exists a chain $q_1 \subseteq q_2 \subseteq \cdots \subseteq q_m$, $(m < n)$ such that $p_i = q_i \cap R$. Then there exists prime ideals $q_{m+1}, ..., q_n$ that complete the chain: $q_m \subseteq q_{m+1} \subseteq \cdots \subseteq q_n$, $p_i = q_i \cap R$ for all i.
3. (**Going-down Property**) If R is a normal domain, $p_1 \supseteq p_2 \supseteq \cdots \supseteq p_n$ be a chain of prime ideals of R, and suppose there exists a chain $q_1 \supseteq q_2 \supseteq \cdots \supseteq q_m$, $(m < n)$ such that $p_i = q_i \cap R$. Then there exists prime ideals $q_{m+1}, ..., q_n$ that complete the chain: $q_m \supseteq q_{m+1} \supseteq \cdots \supseteq q_n$, $p_i = q_i \cap R$ for all i.
4. (**Incomparability property**) Let q_1 and q_2 lie over p. Then either $q_1 = q_2$, or they do not contain each other: $q_1 \nsubseteq q_2$ and $q_2 \nsubseteq q_1$

Proof **:**

1. We'll take advantage of the corresponence theorem for rings and their localization. Let $D = R - P$. Then as we saw, D is multiplicatively closed, and so $R[D^{-1}]$ is well-defined. Since D is also a multiplicatvely closed set of S, $S[D^{-1}]$ is also well-defined. Furthermore, $S[D^{-1}]/R[D^{-1}]$ is an integral extensions of $R[D^{-1}]$: for any $s \in S[D^{-1}]$, if $s \in S$, clearly S is still integral. If $t = \frac{s}{d}$ for some $s \in S$ and $d \in D$, then letting

$$s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0 = 0$$

This equation is still zero if we multiply by d^{-n}:

$$\frac{s^n}{d^n} + \frac{a_{n-1}s^{n-1}}{d^n} + \cdots + \frac{a_1 s}{d^n} + \frac{a_0}{d^n} = 0$$

and shuffling the coefficients we get:

$$\frac{s^n}{d^n} + \frac{da_{n-1}s^{n-1}}{d^{n-1}} + \cdots + \frac{d^{n-1}a_1 s}{d} + \frac{a_0}{d^n} = 0$$

and replacing each $d^{n-i}a_i = b_i$, and $s/d = t$,

$$t^n + b_{n-1}t^{n-1} + \cdots + b_1 t + b_0 = 0$$

showing that t is integral over $R[D^{-1}]$, and so $S[D^{-1}]$ is an integral extension of $R[D^{-1}]$. Let $m \subseteq S[D^{-1}]$ be a maximal ideal. Then by the corresponence of ideals from a ring and it's localization, we have that $m \cap R_p$ is a maximal ideal in R_p, and $m \cap R_p$ is the extension of P to the local ring R_p, so it contracts to p. It might be easier to see if we have a visual:

$$\begin{tikzcd} S \arrow[r, "\pi"] & S[D^{-1}] \\ R \arrow[u, "\iota"] \arrow[r, "\pi"] & R[D^{-1}] = R_p \arrow[u, "\iota"'] \end{tikzcd}$$

The maximal ideal m is maximal in $S[D^{-1}]$, it's pre-image is a maximal ideal of $R[D^{-1}]$, and the pre-image of that map is p, showing that $p = m \cap R$, as we sought to show.

2. This can be reducd to showing that if $p_1 \subseteq p_2$ and q_1 lies above p_1 such that $p_1 = q_1 \cap R$, then there exists a q_2 that lies above p_2 ($q_2 = p_2 \cap R$) and $q_1 \subseteq q_2$. First, let $\overline{S} = S/q_1$ and $\overline{R} = R/p_1$. Then $\overline{S}/\overline{R}$ is an integral extension by similar reasoning to what we've shown in part 1. By the first question in this homeowrk, we have that p_2/p_1 is a prime ideal of $\overline{R}$. Then by part (1), we have that there exists an ideal $q_2/q_1 \subseteq \overline{S}$ that lies abouve p_2/p_1, in particular:

$$q_2/q_1 \cap \overline{R} = p_2/p_1$$

At this point we are essentially done: taking the pre-image of q_2/q_1, we get q_2, and by the correspondence theorem we get

$$q_2 \cap R = p_2$$

3. here

4. For the sake of contradiction, let $q_1 \subseteq q_2$, and let both lie over p: $q_1 \cap R = q_2 \cap R = p$. Consider $\overline{S} = S/q_1$ and $\overline{R} = R/p$. Then as we've shown earlier, $\overline{S}/\overline{R}$ is an integral extension, and $q_2/q_1 \subseteq \overline{S}$ is a prime ideal. Consider:

$$(q_2/q_1) \cap \overline{R} = (q_2 \cap R)/(R \cap q_1) = (R \cap q_1)/(R \cap q_1) = \{\overline{0}\}$$

Thus, $q_2/q_1 = (\{0\}) \subseteq \overline{S}$, and so $q_2 \subseteq q_1$, so $q_2 \subseteq q_1 \subseteq q_2$, showing that they are in fact equal.

Exercise 20.5.1

1. Let $I \subseteq R$ be an ideal and R a ring. Then if we consider integral extension of *rngs*, show that I^{n+1}/I^n is an integral extension.

2. Let $I \subseteq R$, $f : R \to S$ an integral extension and $I^e \subseteq S$ the extension of S. Then:

$$S/I^e \cong S \otimes_R R/I$$

20.6 Valuation Ring and Local Rings

Recall from chapter 19, section 19.2 that the absolute value defines a notion of size on any field. If the field is finite, the absolute is trivial and hence doesn't define any notion of size satisfying the axioms of absolute value besides $|x| = 1$ if $x \neq 0$. By theorem 19.2.7, up to isomorphism the only absolute values on $\mathbb{Q}$ are $|-|_p$ and $|-|$. The these absolute values where defined using a constant c^{-m}, however the choice of this constant did not affect the topology, and hence what really matters was the exponent. In this section, we define a function that captures the exponent information. It shall be a useful concept as a link between geometric and algebraic intuition's on certain fields.

Definition 20.6.1: $\mathbb{R}$-Valuation

Let K be a field Append ∞ to $\mathbb{R}$ and preserve the total ordering by adding the properties that $g \leq \infty$ for all $g \in \mathbb{R}$ and $g + \infty = \infty + g = \infty + \infty = \infty$. Then a function $\nu : K \to \mathbb{R} \cup \{\infty\}$ is called a $\mathbb{R}$-*valuation* if

1. $\nu(K^\times) \subseteq \mathbb{R}$ and $\nu|_{K^\times} : K^\times \to \mathbb{R}$ is a group-homomorphism, that is $\nu(x \cdot y) = \nu(x) + \nu(y)$.
2. $\nu(x + y) \geq \min\{\nu(x), \nu(y)\}$
3. $\nu(x) = \infty$ if and only if $x = 0$

It may be more clear that $\mathbb{R}$-valuations come as a generalization of non-Archimedean absolute values from the original definition given by Emil Artin (and hence why these are called valuations). To make sense of this definition, we shall introduce the slightly more general notion of a totally ordered abelian group.

Definition 20.6.2: Totally Ordered [Abelian] Group

Let G be an abelian group. Then G is said to be a *totally ordered group*, or simply an *ordered group* if $\leq$ is a total ordering on G that is translation invariant, that is for all $x, y, z \in G$ where $x \leq y$:

$$zx \leq zy$$

It was shown by ref:HERE[5] that an abelian group is totally order iff it is torsion-free. Not all torsion-free abelian groups can be extended to a totally ordered field k, and hence they cannot fit into $\mathbb{Q} \subseteq k \subseteq \mathbb{R}$ (see theorem 19.1.5). Hence, this is a proper generalization, but not one that will not be needed beyond connecting $\mathbb{R}$-valuations to absolute values defined in chapter 19:

Example 20.6.1: Valuation Ring

1. Puiseau series has valuation group $\mathbb{Q}$. See Richard Borchard Schemes 22: Valuation rings

Now, in definition 20.6.1, the $(\mathbb{R}, +)$ can be replaced with $(G, *)$. The following shows how this links to absolute values:

[5] got it from this which lead to the comment here:

Definition 20.6.3: G-Valuation Equivalent Definition

Let K be a field. Append 0 to G and preserve the total ordering by adding the properties that $0 \leq g$ for all $g \in G$ and $0 * g = g * 0 = 0 * 0 = 0$. Then a function $\nu : K \to G \cup \{0\}$ is called a *G-valuation* if

1. $\nu(K^\times) \subseteq G$ and $\nu|_{K^\times} : K^\times \to G$ is a group-homomorphism, that is $\nu(x \cdot y) = \nu(x) * \nu(y)$.
2. $\nu(x + y) \leq \max\{\nu(x), \nu(y)\}$
3. $\nu(x) = 0$ if and only if $x = 0$

The image of $\nu(K^\times)$ is called the *value group*.

If $G = \mathbb{Z}$, or equivalently $\mathbb{R}$-evaluation definition when $v(K^*) \subseteq \mathbb{Z}$, a special name is given:

Definition 20.6.4: Discrete Valuation

Let K be a field. Then a function $\nu : K \to \mathbb{Z} \cup \{\infty\}$ is called a *discrete valuation* if

1. $\nu(x \cdot y) = \nu(x) + \nu(y)$
2. $\nu(x + y) \geq \min\{\nu(x), \nu(y)\}$
3. $\nu(x) = \infty$ if and only if $x = 0$

Example 20.6.2: Valuations

1. Given any non-Archimedean absolute value v_p with values in $(\mathbb{R}^+, \cdot)$, we get a $\mathbb{R}^+$ valuation. Hence, each v_p is a valuation on $\mathbb{Q}$ in the equivalent definition, and v_∞, v_0 are valuations on $k(x)$ for any field k (what has to be modified for the modern definition?).

Given a field k, it may be asked what valuations may be put on it. This is in correspondence with certain sub-rings of k:

Definition 20.6.5: Valuation Ring

Let ν be a valuation on K. Then

$$\mathcal{O}_K = \{x \in K \mid \nu(x) \geq 0\} \qquad (\text{equiv. } \{x \in K \mid \nu(x) \leq 1\})$$

forms a ring, called the *Valuation Ring*. If ν is a discrete valuation, the ring is called a *discrete valuation ring*.

The reader should verify this is indeed a ring.

Proposition 20.6.6: Valuation Ring Properties

Let k be a field with valuation ν and let $\mathcal{O}_k$ be the valuation ring associated to it. Then:

1. Given $x \in k \setminus \{0\}$, either $x \in \mathcal{O}_k$ or $x^{-1} \in \mathcal{O}_k$ (where both are possible at once). As a consequence, k is a field of fractions of $\mathcal{O}_k$. [a]
2. The group of units is $\mathcal{O}_K^* = \{x \in k \mid \nu(x) = 0\}$ (equiv. $\nu(x) = 1$)
3. $\mathcal{O}_k$ has exactly one maximal ideal $\mathfrak{m}_v = \{x \in k \mid v(x) > 0\} = \mathcal{O}_k \setminus \mathcal{O}_k^*$ (equiv $v(x) < 1$)
4. The ideal of $\mathcal{O}_k$ form a total ordering under $\subseteq$
5. The ring $\mathcal{O}_k$ is integrally closed (i.e. a normal domain)

[a] Note that the either-or implies a general Γ-valuation!

Proof **:**
Points 1-3 are left as an exercise.

For 4, first show it on principal ideals, namely if $(a), (b) \subseteq \mathcal{O}_K$, then $(a) \subseteq (b)$ or $(b) \subseteq (a)$. Do this by noticing that either $ab^{-1} \in \mathcal{O}_K$ or $a^{-1}b \in \mathcal{O}_K$. For general ideals, if $\mathfrak{a}$ and $\mathfrak{b}$ are ideals that are not ordered, then there exists $a \in \mathfrak{a} \setminus \mathfrak{b}$ and $b \in \mathfrak{b} \setminus \mathfrak{a}$. Consider $(a), (b)$ which must be ordered namely $(a) \subseteq (b)$ or $(b) \subseteq (a)$. Say $(a) \subseteq (b)$. Then $(a) \cap (b) = (a)$. But then $a \in (b)$ meaning $a = bx$ for some $x \in \mathcal{O}_K$. But as $bx \in \mathfrak{b}$, this would imply $a \notin \mathfrak{a} \setminus \mathfrak{b}$, a contradiction.

We shall show that $\mathcal{O}_K$ is integrally close. Say that $x \in K$ is integral over $\mathcal{O}_K$, for the sake of contradiction say $x \notin \mathcal{O}_K$. As it is integral:

$$x^n + a_1 x^{n-1} + \cdots + a_n = 0$$

with $a_i \in \mathcal{O}_K$. As $x \notin \mathcal{O}_K$, $x^{-1} \in \mathcal{O}_K$. But then:

$$x = -a_1 - a_2 x^{-1} - \cdots - a_n = 0$$

But then $x \in \mathcal{O}_K$, a contradiction, completing the proof.

These properties in fact characterize valuation rings:

Proposition 20.6.7: Characterizing Valuation Ring

Let k be a field, $R \subseteq k$ a sub-ring, and R^* the group of units of R. Then the following are equivalent:

($\Rightarrow$) R is a valuation ring of k

($\Rightarrow$) $k = \text{Frac}(R)$ and the ideals form a chain:

($\Rightarrow$) for $x \in k \setminus \{0\}$, either $x \in R$ or $x^{-1} \in R$

Then $G = (k \setminus \{0\})/R^*$ is a totally ordered abelian group and

$$\nu_R : x \mapsto xR^*$$

is a valuation on k. The induced valuation ring is R. As a consequence, two valuations on the same field if and only if they have the same valuation ring

Proof **:**
exercise.

20.6.1 Discrete Valuation Ring

Limiting to discrete valuation rings, then there is a smallest positive value s such that:

$$\nu(K^*) = s\mathbb{Z}$$

We can normalize this result and get an equivalent ν where $\nu(K^*) = \mathbb{Z}$ and $\mathcal{O}_K$, $\mathcal{O}_K^*$, and $\mathfrak{p}$ are the same. Then there must be an element such that

$$\nu(\pi) = 1$$

Labelling this element as π, we can use it to show that any $x \in K \setminus \{0\}$ can be written as :

$$x = u\pi^m$$

for a unit $u \in K^\times$ ($\nu(u) = 0$). To see this, let

$$\nu(x) = m \Rightarrow \nu(xp^{-m}) = 0 \Rightarrow u = x\pi^{-m} \in \mathcal{O}_K^*$$

which gives the result. This works as $m \in \mathbb{N}$. (discrete) valuation ring, this in fact characterizes them:

Proposition 20.6.8: Characterizing DVR

Let R be an integral domain and $k = \text{Frac}(R)$. Then R is the valuation ring of a discrete valuation ν on k if and only if R is a PID with a unique nonzero prime ideal.

Proof **:**
exercise

Example 20.6.3: Discrete Valuation Ring

1. The ring of p-adic integers $\mathbb{Z}_p$ is a discrete valuation ring.
2. For any prime p, $\mathbb{Z}_{(p)}$ the localization of $\mathbb{Z}$ at the prime (p) is a discrete valuation ring.
3. The ring $k[[x]]$ of formal power series is a DVR with unique non-zero prime ideal (x).
4. (If you know some analysis) Take the ring of Real or complex power series that converge to 0. Show this is a DVR.
5. (If you know some complex analysis) Let X be a Riemann surface. and let $M(X)$ be the collection of meromorphic functions $X \to \mathbb{C} \cup \{\infty\}$. Then $M(X)$ is a field. For each point $pinX$, define a discrete valuation on $M(X)$ where
$$\nu_p(f) = i$$
where i is the largest number st
$$\frac{f(z)}{(z-p)^i}$$
can be extended to a holomorphic function (i.e. the order of the root/pole).

It shall in algebraic geometry that DVR's appear naturally in the setting of "smooth projective algebraic curves".

Proposition 20.6.9: Representation Of Elements Of DVR

Let k be a field with DVR $\mathcal{O}_k$. Let $\mathcal{T} \subseteq \mathcal{O}_k$ where $0 \in \mathcal{T}$ and one element of every coset of $\mathfrak{m}$ is in $\mathcal{T}$. Then every element $x \in \mathcal{O}_k$ is the sum of a unique power series:
$$\sum_{i \geq 0} r_i p^i$$
with coefficients $r_i \in \mathcal{T}$. Every nonzero $x \in k$ is the sum of unique laurent series $\sum_{i \geq m} r_i p^i$ with coefficients $r_i \in \mathcal{T}$ for all $i \geq m$ and $r_m \neq 0$

Proof **:**
exercise.

Exercise 20.6.1

1. Show the following properties (or examples) of ordered groups (ex. torsion-free)
2. Let $\mathcal{O}$ be a DVR with unique prime $\mathfrak{p}$. Show that $\mathfrak{p}^n/\mathfrak{p}^{n+1} \cong \mathcal{O}/\mathfrak{p}$ for $n \geq 0$

3. Let $x \in R$ for a valuating ring R. Then $1 - x$ is a unit. To make this intuitive, consider the fact that $1/(1-x)$ always has a power-series expansion (see lemma 9.2.7)

20.7 Artinian Rings

As we've seen, being Noetherian has many powerful consequences by allowing to bring in a consistent notion of finiteness into rings, lending itself useful in being able to define concepts like dimension. It is thus a natural question to ask if the dual condition can give us any interesting results, namely the *descending chain condition*. In this section, we explore the properties of rings and commutative rings satisfying the descending chain condition. It turns out that the D.C.C. is much more restrictive,

(The place they seem to be important is that a simple ring that is also Artinian is, by Artin-Wederburn, isomorphic to a product of matrix ring over divisor rings)

Definition 20.7.1: Artinian module and ring

Let M be a R-module. Then if M satisfies the *descending chain condition* on submodules, it is an *artinian module*. More precisely, for any collection of submodules such that

$$M_1 \supseteq M_2 \supseteq \cdots \supseteq M_k \supseteq$$

then for some n, we get $M_n = M_{n+1} = \cdots$.
If we let $M = R$ be a module over itself, then we say that R is an *Artinian ring*.

Naturally, an Artinian ring has the D.C.C. condition on its ideals.

Proposition 20.7.2: Artinian Module Equivalent

M is Artinian if and only if every non-empty set of submodules of M contains a minimal element under inclusion

Proof **:**
Same proof as for the Noetherian case, just reversed

Just like Noetherian modules, if $N \subseteq M$ and M/N are artinian, so is M.

Lemma 20.7.3: Artinian Module Closed Under Quotient

Let M be an Artinian module and $N \subseteq M$. Then M/N is also Artinian

Proof **:**
exercise

Example 20.7.1: Artinian, But Not Noetherian Module
Take $\mathbb{Z}[1/2]/\mathbb{Z}$ over $\mathbb{Z}$. This ring is in the appropriate sense isomorphic to:

$$\mathbb{Z}/\mathbb{Z} \subseteq \mathbb{Z}/2\mathbb{Z} \subseteq \mathbb{Z}/4\mathbb{Z} \subseteq \cdots$$

which we see is an infinite increasing sequence. With a bit of work, we can show these are the only submodules, and hence we have an Artinian, but not Noetherian ring.

However, unlike Noetherian rings, Artinian rings need to satisfy stronger properties on their ideals:

Proposition 20.7.4: Artinian, Prime $\Rightarrow$ Maximal

Let R be an Artinian ring and $P \subseteq R$ be a prime ideal. Then P is maximal

Notice how this differs from Noetherian rings: $\mathbb{Z}[x]$ is Noetherian, (x) is prime, but not maximal since $(x) \subsetneq (2, x)$. Thus, being Artinian immediately imposes stronger conditions on its ideals.

***Proof* :**
Consider R/P, which is also artinian by lemma 20.7.3. We'll show $R/P = F$ is a field. COnsider $x \in F \setminus \{0\}$. Then by the descendign chain condition:

$$(x) \supseteq (x^2) \supseteq \cdots$$

must staxalize, that is for some n:

$$(x^n) = (x^{n+1})$$

which implies $x^n = x^{n+1}c$. By the Cancellion law, we get $1 = xc$, showing that x is in fact a unit. Since every element is a unit in F, we have that F is in fact a field, and so P is maximal, as we sought to show.

(This seems out of place; will move into comm alg somewhere)

Definition 20.7.5: Krull Dimension

Let R be a commutative ring. Then the *krull dimension* of R (or simply the *dimension* of R) is the longest possible length of distinct prime ideals (i.e. the supremum of the length of chains of [distinct] prime ideals):

$$P_1 \subsetneq P_2 \subsetneq \cdots \subsetneq P_n \subsetneq \cdots$$

where we count the number of inclusions $\subsetneq$. If R admits arbitarily long chains of distinct prime ideals, then R is said to be infinite dimensional. The Krull dimension of a ring is often denoted $\operatorname{kdim}(R)$ or $\dim(R)$.

(mension how krull dimension matches the transcendence degree if R is a finitely generated k-algebra. The advantage of this definition is that it does not reffer to anything about finitely generated k-algebra, just arbitrary commutative rings)

Example 20.7.2: Krull Dimension

1. If $R = F$ is a field, then $\dim(F) = 0$
2. if R is a principal ideal domain, then since all prime ideals are maximal, $\dim(R) = 1$.
3. We will prove later that if k is a field, then $k[x_1, ..., x_n]$ has Krull dimension n.
4. We will prove later that if R is a noetherian ring with dimension n, then $\dim(R[x]) = n + 1$
5. We will soon show that any artinian ring has dimension 0

Theorem 20.7.6: Hopkin's Theorem

Let R be an Artinian ring. Then:

1. There are only finitely many maximal ideals
2. The quotient $R/(\mathrm{Jac}(R))$ is the direct product of a finite number of fields, where if there are n maximal ideals, we have a direct product of n fields R/M_i.
3. Every prime ideal of R is maximal. Thus $\mathrm{kdim}(R) = 0$. As a consequence, $\mathrm{Jac}(R) = \mathrm{Nil}(R)$ and is a nilpotent ideal, $(\mathrm{Jac}(R))^n = 0$ for some $n \geq 1$
4. R is isomorphic to a direct product of a finite number of local artinian rings
5. Every Artinian ring is Noetherian

Note that the last conditions is not true for Artinian Modules.

Proof **:**

1. Consider the chain:
$$m_1 \supseteq m_1 \cap m_2 \supseteq \cdots$$
I first claim that each inclusion of this chain is proper. To see this, recall that if m_i and m_j are non-equal maximal ideal, it must be that $m_i + m_j = 1$. Thus, any collection of maximal ideals has this pair-wise condition. By the Chinese Remainder Theorem:
$$R/(m_1 \cap \cdots \cap m_n) \cong R/m_1 \times R/m_2 \times \cdots \times R/m_n$$
Now simply see that the element $(\overline{0}, ..., \overline{0}, \overline{1})$ is non-zero, and hence not contain in $m_1 \cap \cdots \cap m_n$, showing that each inclusion *must* be proper! Thus, by the D.C.C., the above chain must stabalize, meaning there can only be finitely many maximal ideals.
2. The proof follows immediately from the previous observation of the CRT
3. We've shown $\mathrm{Jac}(R)$ is nilpotent, hence $\mathrm{Jac}(R) \subseteq \sqrt{(0)}$. Since every prime ideal contains the nilradical (recall if $0 = x^n \in P$, then either $x \in P$ or $x^{n-1} \in P$, which inductively shows that $x \in P$), hence contain $\mathrm{Jac}(R)$. Thus, considering the natural projection $\pi : R \to R/\mathrm{Jac}(R)$, it is clear that $\pi(P)$ is also prime. Since
$$R/\mathrm{Jac}(R) \cong k_1 \times \cdots \times k_n$$

and since ideals in product rings $R_1 \times R_2$ are of the form $I_1 \times I_2$, we see that the image being prime implies it must be a subset of the codomain where either the component is all of k_i, or 0. It follow's that the image if a maximal ideal, and so by the correspondence theorem P must be a maximal ideal in R, showing that all prime ideals are maximal

4. Let $m_1, ..., m_n$ be the maximal ideals of R and let n be the number such that $(\text{Jac}(R))^k = 0$. Then
$$\prod_i^n M_i^k \subseteq \left(\prod_i^n M_i\right)^k \subseteq (\text{Jac}(R))^k = 0$$
Thus, by the CRT
$$R \cong (R/M_1^k) \times \cdots \times (R/M_n^k)$$
Notice that each R/M_i^k is an artinian with unique maximal ideal M_i/M_i^k.

5. By part (4), it suffices to show that local Artinian rings are Noetherian. If m is the maximal ideal of the local artinian ring R, then $m = \text{Jac}(R)$, so $M^k = (\text{Jac}(R))^k = 0$ for some $k \in \mathbb{N}_{>0}$. Hence $R \cong R/M^k$. It follows (as an exercise) that R is Noetherian if and only if R is artinian

Corollary 20.7.7: Artinian Ring Equivalent Condition I

Let R be a ring. Then R is Artinian if and only if R is Noetherian and $\dim(R) = 0$

Proof **:**
We have shown that R is Artinian implies R is Noetherian, and all prime ideals are maximal implies $\dim(R) = 0$. For the converse, by corollary 21.2.22, there exists prime ideals $P_1, P_2, ..., P_n$ such that
$$(0) = P_1 P_2 \cdots P_n$$
These prime ideals are all maximal since $\dim(R) = 0$. By the chinese remainder theorem, R is isomorphic to the direct product of a finite number of Noetherian rings of the form R/m^k where m is a maximal ideal in R. By theorem 20.7.6, R/m^k is artinian, and so it follows that R is artinian.

Corollary 20.7.8: Artinian Ring Equivalent Condition II

Let k be a field and R a finitely generated k-algebra. Then R is artinian if and only if R is finitely generated as a k-module.

For artinian modules, we have a nice characterization if $R = k$ is a field

Lemma 20.7.9: Artin+Noetherian Module

Let M be a vector space over k. Then the following are equivalent:

1. M satisfies the A.C.C.
2. M satisfies the D.C.C.
3. M is finite-dimensional

***Proof* :**
Naturally, (3) implies both (1) and (2). Conversely, assume $\dim_k(M) = \infty$, so pick some infinite collection of linearly independent vectors. Then we have an infinite strict chain:

$$\langle e_1 \rangle \subsetneq \langle e_1, e_2 \rangle \cdots$$

and an infinite descending chain:

$$\langle e_1, e_2, ... \rangle \supsetneq \langle e_2, e_3, ... \rangle \supseteq \cdots$$

completing the proof

Example 20.7.3: Artinian Rings

1. Let $n > 1$ be an integer, and let $R = \mathbb{Z}/n\mathbb{Z}$. Since R is finite, it must be Artinian. If $n = p_1^{\alpha_1} \cdots p_n^{\alpha_n}$, for some distinct primes p_n and $\alpha_i \in \mathbb{N}_{>0}$, then by the ftofgag:

$$\mathbb{Z}/n\mathbb{Z} \cong (\mathbb{Z}/p_1^{\alpha_1}\mathbb{Z}) \times (\mathbb{Z}/p_2^{\alpha_2}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_n^{\alpha_n}\mathbb{Z})$$

   where each $\mathbb{Z}/p_i^{\alpha_i}\mathbb{Z}$ is a local artinian ring with maximal idea $p_i/p_i^{\alpha_i}$. The jacobson radical of this ring is $(p_1 p_2 p_n)$.

2. Let R be a finite dimensional k-algebra over some field k (it need not be commutative!). Then R is an Artinian ring, since any descending chain of R is a descending chain of k-subspces of R, which is bounded by $\dim_k(R)$. Note that if R is not commutative, R is not the finite direct product of local artinian rings, but the finite product of indecomposable artinian rings (i.e. artinian rings which cannot be split into direct sum of submodules).

3. Let $k[x]$ be the polynomial ring over a field k and $f \in k[x]$ be a nonzero polynomial. Then if $f(x) = f_1^{n_1}(x) \cdots f_k^{n_k}(x)$, by the CRT we have:

$$k[x]/(f(x)) \cong k[x]/(f_1^{n_1}(x)) \times \cdots \times k[x]/(f_k^{n_k}(x))$$

   which is artinian, since $k[x]/(f(x))$ is a finite dimensional k-module (each $k[x]/(f_i^{n_i}(x)$ is a local artinian ring, with maximla ideal $(f_i(x))/(f_i^{n_i}(x))$). If $n_i = 1$ for all i, $k[x]/(f(x))$ is a product of fields, and if furtheremore there was only 1 n_i (i.e. $i \in \{1\}$), then $k[x]/(f(x))$ would be a field extension

We shall give names to modules that satisfy both the Artinian and Noetherian condition:

Definition 20.7.10: Modules of Finite Length

Let M be an R-module that is Noetherian and Artinian. Then M is an R-module of *finite length*.

By proposition 20.2.5, each such module has a unique $\ell(M) \in \mathbb{N}$. It can be shown That the Jordan-Hölder Theorem applies to this class of objects. The Jordan-Hölder progrma will be dependend on the ring-action:

Example 20.7.4: Module of Finite Length

The simplest example would be when $R = k$ is a field, which gives vector-spaces. In this case, for any V

$$\ell(V) = \dim_k(V)$$

Proposition 20.7.11: Finite Length Additive Under Extensions

Let

$$0 \to M' \to M \to M'' \to 0$$

Be a short exact sequence of finite length modules. Then:

$$\ell(M) = \ell(M') + \ell(M''')$$

Proof **:**

exercise

Exercise 20.7.1

1. Let M be a Noetherian and Artinian A-module, $\varphi : M \to M$ an A-module homomorphism. Show that there exists an $n \geq 0$ such that $M = \ker(\varphi^n) \oplus \operatorname{im}(\varphi^n)$.
2. Let M be an R-module. Then M is of finite length iff it has a composition series.

21

Relating Algebra to Geometry

For at least 2 millennia, a central domain of study in mathematics is geometry, which studies shapes and spaces as well as their geometric properties. Let us use the term "geometrical object" as a catch-all term for what would usually be considered a shape or a space. Examples of geometrical objects studies in history are:

1. Circles, or more generally ellipsoids and cones.
2. Curves on a plane
3. The Platonic solids
4. The Cartesian plane
5. Spheres and tori
6. Möbius bands
7. Graphs of functions

and so forth, the list stretching out endlessly. There have been many efforts in studying geometrical objects. Many efforts have historically concentrated in what would be called the *Euclidean approach* or *ruler and compass approach*, which takes most of its inspiration from the continuation of Euclid's book *Elements* written around 300BC. This paradigm has the 5 axioms of Euclidean Geometry stated on the first page of Euclid's *Elements* from which using a ruler and compass Euclid manages to deduce a large portion of what would now be considered classical Analytic geometry. However, there are some fundamental limitations to the Euclidean method, not the least that one of the axioms eliminates a lot of fruitful geometric objects from consideration that a modern mathematician or a physicist would work with (namely, the infamous 5th axiom, also known as the *parallel postulate*). For example, geometry on a sphere does not follow all the same rules as the geometry on a flat

surface. It wasn't until the 18th and 19th century when mathematics was going through a renaissance that these realizations where brought to the forefront and new geometrical paradigms where being invented and discovered.

Nowadays, we understand that the space Euclid was working with was *finite dimensional real vector spaces* where there is no notion of an origin (the idea of an origin was introduced by *René Descartes* in trying to quantify certain geometric problems). Conversely, given a finite-dimensional real vector-space where we drop the property of there being a 0 is equivalent to euclidean space (the geometrical results of these two spaces are the same). Hence, the geometrical space in Euclidean Geometry becomes the familiar finite dimensional real vector space. Most of the work following euclidean geometry would be presented in a class in Analytic Geometry[1].

It is now understood that there are many different paradigms in which we may pursue the study of geometry, each studying a different area of interest. Most of these paradigms fall under the umbrella terms of *Differential Geometry*, *Topology*, *Discrete Geometry*, *Algebraic/Complex Geometry*, or *Classical Geometry*[2]. Most of these paradigms are *not* mutually exclusive and heavily overlap, as they often study many similar objects. The study of *Algebraic Geometry* is the study of the shapes that are given by *algebraic functions*, which mostly constitute *polynomials* or ratio's of polynomials.

In this chapter, we shall show that shapes given by polynomials (or their ratios) can be described and characterized by all the functions on the shape, which shall often be the quotient of a polynomial ring. What is meant by "all the function on the shape" shall be made precise in this chapter. When we say the shapes given by polynomials, we shall mean the *zero sets* of the polynomials, or the *vanishing* points of the polynomial. Famously, the polynomial $x^2 + y^2 - 1 \in k[x, y]$ is interpreted as a unit circle by considering all points where $x^2 - y^2 - 1 = 0$ or $x^2 + y^2 = 1$:

Figure 21.1: Circle represented by the equation $x^2 + y^2 - 1 \in \mathbb{R}[x, y]$

We shall see that the properties of the quotient polynomial rings gives a lot of information about the zero-sets of a collection polynomials. Conversely, we shall see that we often can get algebraic results

[1] A course that in the 21st century has become rather scarce

[2] which includes Euclidean Geometry, Affine Geometry, Conformal Geometry, Analytic Geometry, among other fields

by looking at the geometry of the zero-sets of polynomials. In other words, there is a symbiotic relationship between algebra and geometry where the two influence each others developments[3].

The goal of this chapter is to build up the dictionary between geometry and algebra so that the reader can easily convert back and forth between the two fields. This chapter ends by briefly outlining modern algebraic geometry which generalizes the dictionary from quotients of polynomials rings and zero's of polynomials to schemes and commutative rings.

52: Quick Side-Note

A common misconception given by a strong reliance on graphs in most first-year undergraduate calculus/analysis class is to associate the zero set of a function with the graph of a function. For example, $f(x) = x$ has a zero-set of only 0, while $f(x, y) = x - y$ has the zero set to be $\Gamma(f)$. This example may seem trivial, however it is surprising easy to switch these two, as we shall be looking at the shape of zero-sets, and most students are comfortable looking at the shape of a function as it graph.

21.1 Affine Algebraic Sets

In this whole chapter, R is a commutative ring and k be a field.

So far when polynomials were encountered, they were treated as a commutative ring with an indeterminate extension (or equivalently, they can be thought of as a ring over a free monoid). There is a natural way of thinking about any such polynomial as a function (which is probably how most first encounter polynomials): if $f \in k[x_1, x_2, ..., x_n]$, then if $(a_1, a_2, ..., a_n) \in k^n$, let $f(a_1, a_2, ..., a_n) \in k$ be the element which you get when replacing each x_i with a_i. Then we can make f into a function through the following map[4]

$$k[x_1, ..., x_n] \to \mathrm{Hom}_{\mathbf{Set}}(k^n, k) \qquad f \mapsto ((a_1, ...a_n) \mapsto f(a_1, ..., a_n)) \tag{21.1}$$

where we shall consider the constants $c \in k$ to map to the functions that shall return the value of c:

$$c \mapsto ((a_1, ..., a_n) \mapsto c)$$

Note that the codomain of the function in equation 21.1 must be the hom-set over the category of set, and not any other category we've extensively worked over so far (**Grp**, **Ring**, **Alg**, etc.), in particular since we will have some very non-linear functions (a simple example would be $x^2 \in k[x]$). On the other hand, the image of the above function is not just a collection of all set functions as they must all satisfy polynomials relations (can you find a set-function that is not given by a polynomial?[5]). The subset given by the function in equation (21.1) shall be called the set of *regular functions*: $\mathrm{Hom}_{\mathbf{Reg}}(k^n, k)$. Before properly boxing this definition, we shall take some more time pondering what the domain of the hom-set should be, namely we shall want the domain to represent Cartesian space without the extra algebraic structure that is usually implied by the notation k^n. For example, k^n usually denotes the n-dimensional vector-space over k with an origin, and more generally

[3]There are some mathematicians that may even go so far as to say that algebra and geometry feel like two sides of the same coin

[4]For those familiar with algebraic geometry, the following map can be interpreted as taking a subset of the compatible germs over the global section

[5]Over finite fields (and only over finite fields) all set functions are polynomials, see exercise ref:HERE

we think of k^n as having addition and scalar multiplication satisfying the axioms of a vector-space. Vector-spaces are "defined" for linear polynomials, with subspaces that are naturally "linear" (recall that subspaces are given by solutions of linear polynomials, see theorem 13.2.23). We shall want a new type of space for which subspaces are solutions to general polynomial equations.

21.1.1 Affine Sets

Definition 21.1.1: Affine Space

Let k be any field. Then the collection of the n-tuples of k^n as a set shall be called *affine space*. It's elements shall be called *points* and the space shall be denoted $\mathbb{A}_k^n$.

This space is called *affine* space rather than just "space" as there shall be an equally important (or arguably more important) space called *projective space* that shall be introduced in section 21.4.

With affine space established, let us now define the image of the function in equation (21.1):

Definition 21.1.2: Regular Function [on Affine Space]

Let F represent the function in equation (21.1). Then the collection $\mathrm{im}(F)$ is called the *regular functions*, and is usually denoted $k[\mathbb{A}_k^n]$ or $\mathrm{Hom}_{\mathbf{Reg}}(\mathbb{A}_k^n, \mathbb{A}_k^1)$.

The set $\mathrm{Hom}_{\mathbf{Reg}}(\mathbb{A}_k^n, \mathbb{A}_k^1)$ forms a natural k-algebra structure given by point-wise operations:

$$f + g \equiv f(a) + g(a) \qquad fg \equiv f(a)g(a) \qquad c \cdot f = cf(a) \qquad \forall a \in k \tag{21.2}$$

With this k-algebra structure, it can be shown that $k[x_1, ..., x_n] \cong_k \mathrm{Hom}_{\mathbf{Reg}}(\mathbb{A}_k^n, \mathbb{A}_k^1)$, and so the two sets are often used interchangeably. If the distinction needs to be made clear, we shall write

$$k[X_1, ..., X_n] = \mathrm{Hom}_{\mathbf{Reg}}(\mathbb{A}_k^n, \mathbb{A}_k^1) = k[\mathbb{A}_n^k]$$

where each X_i is a regular function that is the projection onto the ith coordinate, and call an element of $k[X_1, ..., X_n]$ a *polynomial function* or *polynomial map*. This distinction shall become more important once there can be multiple polynomials that will be associated to the same function, which shall be the case in section 21.2.1.

Due to this isomorphism, we can define a function from $k[x_1, ..., x_n]$ to k directly by directly substituting the values for the indeterminates with a point in the affine space. Namely, for each $a \in \mathbb{A}_k^n$, we take:

$$\mathrm{ev}_a^k : k[x_1, ..., x_n] \to k \qquad f(x) \mapsto f(a)$$

where a and x are understood to be n-tuples of values and indeterminates respectively. These are all k-algebra homomorphisms, for

$$\mathrm{ev}_a^k(f(x) + cg(x)) = \mathrm{ev}_a^k(f(x)) + c\,\mathrm{ev}_a^k(g(x)) \qquad \mathrm{ev}_a^k(F(x)g(x)) = \mathrm{ev}_a^k(f(x)\,\mathrm{ev}_a^k(g(x))$$

The subscript of the evaluation function represents the point of evaluation, while the superscript represents the values we choose to plug for every indeterminate, where the values are usually the same as the base field, however they may differ, for example it is common to take $\mathbb{Z}$ or $\mathbb{Q}$[6]. Thus, a regular function can be thought of as the collection of evaluation functions on a polynomial.

[6]These shall be called $\mathbb{Z}$-valued and $\mathbb{Q}$-valued points of the polynomial, or more precisely the variety (as we shall soon see)

53: Foreshadowing

If you know elementary analysis or calculus, then you are probably aware that the types of functions we most commonly use are *continuous*. Continuous functions are a very important tool in analysing the propertie of the space we're studying (ex. $\mathbb{R}$, $\mathbb{R}^n$, $L^2(\mathbb{R})$, etc.), they can be thought of as the natural types of functions that preserve some local properties of the space. In Algebra, we can think of polynomial functions as the natural types of functions. This comparison between continuous functions mapping to $\mathbb{R}$ and polynomial functions mapping to k will in fact be incredibly precise, as we will see in the last section of this chapter.

From such polynomials functions, we can create many interesting shapes depending on the zeros', or locus, of polynomial:

Example 21.1.1: Zeros' of Polynomials

1. As we saw, $x^2+y^2-1 \in \mathbb{R}[x,y]$ gives a circle, see figure (21.1). More precisely, the polynomial shall be associated to an element of $k[\mathbb{A}^n_k]$, where plugin in any point on from $\mathbb{A}^2_{\mathbb{R}}$ that is on the circle (ex. $(1,0)$ or $(\sqrt{2}/2, \sqrt{2}/2)$ shall give 0. From here on out, we shall think of the polynomial $p(x)$ as representing this regular function, and assume it is clear from context when we are treating it $p(x)$ like element of the polynomial ring and when we are treating it like a regular function.

2. Take $y - x^2 \in \mathbb{R}[x,y]$. Then this shall give the usual parabola $y = x^2$. If we take $x - y^2 \in \mathbb{R}[x,y]$, we shall get a parabola rotated 90 degree's (or $\pi/2$ radians) clockwise. If we take the polynomial $(y-x^2)(x-y^2) \in \mathbb{R}[x,y]$, then the zero set of this new polynomial shall be the zero-set of the two multiplied polynomials:

3. Take $(x^2+y^2)^2 - (x^2-y^2) \in \mathbb{R}[x,y]$. This shape is called the *Bernoulli Lemniscate*:

4. Take $p(x) \in k[x]$ over any field k. Then the shape associated to $p(x)$ are the *roots* of the polynomials. For example, taking $x^2 - 2 \in \mathbb{Q}(\sqrt{2})[x]$ shall give $\{\sqrt{2}, -\sqrt{2}\} \subseteq \mathbb{A}^2_{\mathbb{Q}(\sqrt{2})}$ as these are the solution to $x^2 - 2 = 0$. More generally, the shape given by a 1-dimensional polynomial shall always the collection of roots that is present in the underlying field. This means that $x^2 - 2 \in \mathbb{Q}[x]$ would give the empty set, as there are no rational roots. By the fundamental theorem of algebra, any polynomial $p(x) \in \mathbb{C}[x]$ of degree d shall have d roots, and hence the shape associated to $p(x)$ shall have d points.

5. Take $x^2 + y^2 + z^2 - 1 \in \mathbb{R}[x, y, z]$. Then this is the classical unit sphere centered at the origin:

$$x^2 + y^2 + z^2 = 1$$

What polynomial would you choose to center the unit sphere at different points of $\mathbb{A}^3_{\mathbb{R}}$? How would you change the radius of the sphere?

6. Let us now consider taking two shapes: the sphere given by $x^2 + y^2 + y^2 - 4$ and the hyperboloid given by $x^2 + y^2 - z^2 - 1$. The hyperboloid looks like so:

Then the intersection of these two shapes shall produce two circles, visualized like so:

In other words, $(x^2 + y^2 + z^2 - 4)(x^2 + y^2 - z^2 - 1)$ shall have the same zero set as two circle at height $\pm\sqrt{3/2}$ with radius $\sqrt{5/2}$.

7. Consider the following shape given by $t \mapsto (t, t^2, t^3)$:

It can be proven that this shape is *not* the zero-set of a single polynomial $p(x, y, z) \in k[x, y, z]$. On the other hand, this shape, is the intersection of two zero-sets of the surfaces $y = x^2$ and $z = xy$.

More generally, given any polynomial in $k[x_1, x_2, ..., x_n]$, we can associate with it a geometric shape based on it's zero set!

Definition 21.1.3: [Affine] Algebraic Set

Let $\mathbb{A}^n$ be an affine n-space over a field k, and let $S \subseteq k[\mathbb{A}^n]$. Then:

$$\mathcal{Z}(S) \subseteq \mathbb{A}^n \quad \mathcal{Z}(S) := \{(a_1, a_2, ..., a_n) \mid f(a_1, a_2, ..., a_n) = 0, \text{ for all } f \in S\}$$

If for $V \subseteq \mathbb{A}^n$, there exists an $S \subseteq k[\mathbb{A}^n]$ such that $V = \mathcal{Z}(S)$, then V is called an *affine algebraic set*, *locus*, or zero set. If it is unambiguous, the set may be called an *algebraic set*[a].

[a] In some textbooks, this set is called a *variety*. The term variety shall be reserved for definition definition 21.2.13

From now on, when taking a subset of $\mathbb{A}^n$, $A \subseteq \mathbb{A}^n$, it will be assumed that A is an algebraic set unless specified otherwise (i.e. it would be stated "for some subset $A \subseteq \mathbb{A}^n$)

Example 21.1.2: Affine Algebraic Sets

1. If $n = 1$, then the zero's of a polynomial will always be a points, that is, the roots of the polynomial. The only polynomial with infinitely many points (if working over a field of characteristic 0) is 0 (more precisely, the regular function associated to 0), which has all of k as its vanishing locus. Furthermore, the polynomial 1 vanishes nowhere, and hence the corresponding algebraic set is $\emptyset$. Therefore, in $\mathbb{A}^1$, the affine sets are

$$\emptyset \qquad \text{finite sets} \qquad k$$

   This also shows that not all sets are algebraic sets (for example, if $k = \mathbb{R}$, then $[0, 1]$ is not an algebraic set in $\mathbb{A}^1_{\mathbb{R}}$)

2. If $n = 2$, Then take $f \in k[\mathbb{A}^2]$ such that $f(x, y) = x^2 - y^2 - 1$. Then

$$\mathbb{Z}(f) = \left\{(x, y) \in \mathbb{A}^2 \;\middle|\; x^2 + y^2 = 1\right\}$$

If $k = \mathbb{R}$, then notice that this the common expression for a circle! Notice too that $Z(f)$ is not finite, and so cannot be a product of two algebraic sets from $K[\mathbb{A}^1]$.

Furthermore, this examples shows that there are lot's more zero sets in $\mathbb{A}^2$ than there are in $\mathbb{A}^1 \times \mathbb{A}^1$. This is in stark contrast to how every element in the vector space $\mathbb{R}^2$ can be written as an element in $\mathbb{R}$ and an element in $\mathbb{R}$, i.e. $\mathbb{R}^2 \cong \mathbb{R} \times \mathbb{R}$ as vector spaces.

3. For any $\mathbb{A}^n$, can you always find a set of functions $F \subseteq k[\mathbb{A}^n]$ such that $|V(F)| = 1$ (i.e. is every singleton an algebraic set for every $\mathbb{A}^n$ over any field?) This shall be answered in section 21.2.5.

4. What is the vanishing set of linear polynomials?

5. In the first example, we produced the entire space using the zero polynomial. Is it necessary that we use the zero-polynomial to make the zero set the entire space? The answer is no: simply take $k = \mathbb{Z}/2\mathbb{Z}$ and $p(x) = x^2 + x$. However, this is a problem exclusive to finite fields. Can you prove that if k is an infinite field, $p \in k[x_1, x_2, ..., x_n]$ and $p(a_1, a_2, ..., a_n) = 0$ for all $a_i \in k$ then $p \equiv 0$? [exercise ref:HERE (freely use any polynomial has at most finitely many roots, and then induction on the number of terms)]

6. Let k be a finite field and take any subset $S \subseteq \mathbb{A}^n_k$. Show that S is always an algebraic set.

7. Let $f \in k[x_1, x_2, ..., x_n]$. Then $\mathcal{Z}(f)$ is called a *hyper-surface*. For example $S^1 = \mathcal{Z}(x^2 + y^2 - r^2)$ represents a circle of radius r. Another example would be $\mathcal{Z}(xy - 1)$ which represents the hyperbola $y = 1/x$. More generally, any conic section is a hyper-surface. In higher dimensions, $\mathcal{Z}(x^2 + y^2/4 + z^2/9 - 1)$ represents an ellipsoid in $\mathbb{R}^3$.

8. For any polynoimal function $f : k \to k$, we can define $F : k^2 \to k$ where $F(x, y) = f(x) - y$. Then $\mathcal{Z}(F)$ consists of all points where $f(x) - y = 0$, i.e. $f(x) = y$, which is the pair of points $\{(a, b) \mid f(a) = b\}$, i.e. the graph of f! More generally, if $f : k^n \to k$ is a polynomial, the polynomial $F : k^{2n} \to k$ gives rise to the polynomial $F : (\overrightarrow{x}, \overrightarrow{y}) = f(\overrightarrow{x}) - \overrightarrow{y}$ whose zero set is the graph of f.

Proposition 21.1.4: Properties of Affine Algebraic Sets

Let $S, T \subseteq k[\mathbb{A}^n]$. Then

1. If $S \subseteq T$ then $\mathcal{Z}(T) \subseteq \mathcal{Z}(S)$ (i.e. $\mathcal{Z}$ is *contravariant* in respect to inclusion)
2. If $(S) = I \subseteq k[\mathbb{A}^n]$ is the ideal of S, then $\mathcal{Z}(I) = \mathcal{Z}(S)$.
3. For any arbitrary collection of algebraic sets $\{S_i\}_{i \in I}$,
$$\bigcap_{i \in I} \mathcal{Z}(S_i) = \mathcal{Z}\left(\bigcup_{i \in I} S_i\right)$$
that is, the arbitrary intersection of algebraic sets is an algebraic sets, and is equal to the algebraic set of the intersections of the $\{S_i\}_{i \in I}$
4. For any finite collection of algebraic sets $\{S_i\}_{i=1}^n$,
$$\bigcup_{i=1}^n \mathcal{Z}(S_i) = \mathcal{Z}\left(\prod_{i=1}^n S_i\right)$$
that is, the finite union of algebraic sets is an algebraic set, and is equal to the product of the S_i as ideals (from part 2, we can assume $(S_i) = S_i$ without loss of generality).
5. If $0, 1 \in k[\mathbb{A}^n]$ denote the zero and constant function, then
$$\mathcal{Z}(0) = \mathbb{A}^n \qquad \mathbb{Z}(1) = \emptyset$$
6. If I is an ideal of $k[\mathbb{A}^n]$, then
$$\mathcal{Z}(I) = \mathcal{Z}(I^m)\ \forall m \in \mathbb{N}_{>0}$$
that is, $\mathcal{Z}$ is invariant to taking higher powers of the ideals

These are all staright-forwards proofs, and so should all be done before reading these proofs for the best learning experience.

Proof **:**

1. Let $S \subseteq T$, and consider $x \in \mathcal{Z}(T)$. Then for all $f \in T$, $f(x) = 0$. Since $S \subseteq T$, for all $g \in T$, $g(x) = 0$. Thus, $x \in \mathcal{Z}(S)$, establishing $\mathcal{Z}(T) \subseteq \mathcal{Z}(S)$.
2. Let $S \subseteq k[\mathbb{A}^n]$ be any set and let $I = (S)$. We want to show that $\mathcal{Z}(I) = \mathcal{Z}(S)$. Since $S \subseteq I$, the $\mathcal{Z}(I) \subseteq \mathcal{Z}(S)$ conclusion comes from the first part. For the reverse, consider $x\mathcal{Z}(S)$. Then for all $t \in S$, $t(x) = 0$. If we pick any two $f, g \in S$, then $(f + g)(x) = f(x) + g(x) = 0$. If we pick any $c(x) \in k[\mathbb{A}^n]$ and any $f \in S$, then $(cf)(x) = c(x)f(x) = c(x)0 = 0$. Thus, if we enlrage S to be closed under addition and multiplication by elements of $k^[\mathbb{A}^n]$, those polynomial will also be zero on x, and so $\mathcal{Z}(S) \subseteq \mathcal{Z}(I)$, establishing $\mathcal{Z}(S) = \mathcal{Z}(I)$
3. We want to show that $\bigcap_{i \in I} \mathcal{Z}(S_i) = \mathcal{Z}\left(\bigcup_{i \in I} S_i\right)$ for an arbitrary indexing set I. This is just

some set theory: by definition:

$$x \in \mathcal{Z}(\cup_{i\in I} S_i) = \{y \in \mathbb{A}^n \mid f(y) = 0,\ \forall f \in \cup_{i\in I} S_i\} = \mathcal{Z}\left(\sum_\alpha I_\alpha\right)$$

which is true if and only if x is the solution for polynomials in every set S_i. The colleciton of such elements is:

$$\bigcap_{i\in I} \mathcal{Z}(S_i)$$

4. We want to show that $\bigcup_{i=1}^n \mathcal{Z}(S_i) = \mathcal{Z}\left(\prod_{i=1}^n S_i\right)$. By definition:

$$x \in \mathcal{Z}\left(\prod_{i=1}^n S_i\right) = \left\{ y \in \mathbb{A}^n \;\middle|\; f(y) = 0\ \forall f \in \prod_{i=1}^n S_i \right\}$$

where by definition $f(x) = \sum_{j=1}^m (f_{1_j} f_{2_j} \cdots f_{n_j})(x) = \sum_{j=1}^m f_{1_j}(x) f_{2_j}(x) \cdots f_{n_j}(x) = 0$. Notice that for a particular term in the sum to be zero, we only need on of the f_{i_j} to be zero. Therefore, the sum is zero if and only if one polynomial in every term of the sum is zero. Since this also applies to sums with only one term in it, then if x is the solution to one polynomial in S_i, then it will be part of the zero's of $\prod_{i=1}^n S_i$. But this is just another way of saying:

$$x \in \bigcup_{i=1}^n \mathcal{Z}(S_i)$$

5. Notice that $0(x) = 0$ for all $x \in \mathbb{A}^n$ and $1(x) = 1 \neq 0$ for all $x \in \mathbb{A}^n$, and so

$$\mathcal{Z}(0) = \mathbb{A}^n \qquad \mathcal{Z}(1) = \emptyset$$

6. Since $I^m \subseteq I$ (recall that $1 \notin I \subsetneq R$), then by the first part $\mathcal{Z}(I) \subseteq \mathcal{Z}(I^m)$. For the $\supseteq$ direction, if $x \in \mathcal{Z}(I^m)$, then for every $f \in I^m$, $f(x) = 0$. By definition of I^m, $f = \sum_{i=1}^n f_i^m$ for some $n \in \mathbb{N}$. In order for $\sum_{i=1}^n f_i^m(x) = 0$, it must be that every $f_i(x) = 0$. But since every $f_i \in I$, and the summand can express any linear combination of elements of I raised to the power of m, it must be that $g(x) = 0$ for all $g \in I$, and so $\mathcal{Z}(I^m) \subseteq \mathcal{Z}(I)$. Combining with the previous direction we get $\mathcal{Z}(I) = \mathcal{Z}(I^m)$.

A couple of remarks are in order. From (2), we can without loss of generality restrict $\mathcal{Z}$ be a function from the ideals of $k[x_1, ..., x_n]$ to the algebraic sets in $\mathbb{A}^n$:

$$\mathcal{Z} : \{\text{ideals in } k[x_1, ..., x_n]\} \to \{\text{affine algebraic sets in } \mathbb{A}^n\}$$

Furthermore, $\mathcal{Z}$ is a surjection from the ideals of $k[x_1, x_2, ..., x_n]$ and the algebraic sets (the ideals I^m for each $m \geq 1$ map to the same algebraic set). Next, the finiteness condition on the product of ideals is a necessary condition for closure of algebraic sets:

Example 21.1.3: Not Closed Under Arbitrary Unions

Take $\mathbb{A}^1_{\mathbb{R}}$ and the sets $V(x - a)$ for $a \in \mathbb{Z}$. Then prove that the union of these sets is *not* an algebraic set (you may freely use the fact that polynomials have at most finite roots, and recall that $\mathbb{R}[x]$ is Noetherian).

Next, we took the product of ideal $\prod_i^n I_i$ instead of the intersection. Since it always the case that $IJ \subseteq I \cap J$ then in the cases where $IJ \subsetneq I \cap J$, $\mathcal{Z}(I \cap J) \subsetneq \mathcal{Z}(I) \cup \mathcal{Z}(J)$[7]. However, since $\mathbb{Z}(I^2) = \mathbb{Z}(I)$, there is a trick we can use in which $\mathcal{Z}(IJ) = \mathcal{Z}(I \cap J)$, making the distinction obsolete (see exercise ref:HERE)

Perhaps most importantly for this chapter, notice that these properties are the properties of a topology (see appendix C):

Definition 21.1.5: Zariski Topology

Let $\mathbb{A}^n$ be an n-affine space over k. Then the set

$$\mathcal{T}_{\mathcal{Z}} = \{\mathcal{Z}(I) \mid \text{I is an ideal of } k[\mathbb{A}^n]\}$$

from the set of closed sets in the *Zariski topology* $(\mathbb{A}^n, \mathcal{T}_{\mathcal{Z}})$

We can therefore try and study this topological space! This is essentially the foundation for studying geometrical concepts in algebra, and vice-versa, it is the foundation for studying algebraic concepts in geometry, and so we will spend a lot of time establishing the link between the ideals of $k[\mathbb{A}^n]$ and the Zariski topology. We will explore the Zariski topology soon, we first see if we can find some function that is the inverse of $\mathcal{Z}$, that is we start from an algebraic set in $\mathbb{A}^n$ (or more generally any subset of $\mathbb{A}^n$) and associate it to an ideal in $k[\mathbb{A}^n]$, letting us strive towards some way of corresponding the two in a manner similar to Galois Theory[8]

As the topology was defined using closed sets (sets for which a collection of polynomials is zero on that set), a moment should be given to give a geometric interpretation to open sets. Open sets are those for which a collection of polynomials are *not all zero* at once. If S is a singleton, then it will be precisely the set of points where the function $f \in S$ does not vanish. If there is more than one function in S, then if one is not a scalar multiple or a power of the other, it is provable that there is a polynomial $g \in S$ and a point $p \in \mathbb{A}_k^n \setminus \mathcal{Z}(S)$ where $g(p) = 0$. This shows that closed sets have more global properties, and that if working with open sets it may be preferable to work with singleton sets, sets that *don't vanish* anywhere[9].

Towards the goal of finding an inverse for $\mathcal{Z}$, notice that we already have that $\mathcal{Z}$ is a surjection of ideals of $k[\mathbb{A}^n]$ to the algebraic sets. Thus, it is meaningful to ask about having a function from the set of algebraic sets to the set of ideals of $k[x_1, x_2, ..., x_n]$. To understand this question, we start off we a slightly more general question (just like we started with subsets of $k[x_1, ..., x_n]$ instead of ideals):

Definition 21.1.6: Vanishing Ideal

Let $V \subseteq \mathbb{A}^n$ be a subset of the affine n-space. Then

$$\mathcal{I}(V) := \{f \in k[\mathbb{A}^n] \mid f(v) = 0, \ \forall v \in V\}$$

is the *vanishing ideal* of V.

[7]The condition that $IJ = I \cap J$ is surprisingly technical, requiring that $\mathrm{Tor}^1(R/I, R/J) = 0$, see ref:HERE

[8]In fact,the order reversing nature of this correspondence will end up making this correspondence a *Galois correspondence*, which name is inspired by how the Galois correspondence between intermediate fields and subgroups was one of the first well-known order-reversing correspondences

[9]see definition 21.1.13 for a further discussion on open sets generated by singleton

Notice that $\mathcal{I}(V)$ was not defined as an ideal, however it is indeed an ideal; it is important for the sake of intuition that the reader should prove this at least once. Even better, $\mathcal{I}(V)$ is the *largest* ideal of $k[\mathbb{A}^n]$ that corresponds to the set V. Just like $\mathcal{Z}$, $\mathcal{I}$ can be interpreted as a function

$$\mathcal{I} : \mathbb{C}\mathrm{P}(\mathbb{A}_k^n) = \{V \subseteq \mathbb{A}_k^n\} \to \{\text{ideals of } k[\mathbb{A}^n]\}$$

Example 21.1.4: Vanishing Ideal

1. Let $\mathbb{A}^2$ be an affine space over the field $\mathbb{R}$ and let $X \subseteq \mathbb{A}^2$ be the x-axis. Then $\mathcal{I}(X) = (y)$ in $\mathbb{R}[\mathbb{A}^2]$ (or equivalently $(y) \subseteq \mathbb{R}[x, y]$).
2. Let k be infinite. Then $\mathcal{I}(\mathbb{A}^n) = 0$.
3. Find the vanishing ideals if of subsets $\mathbb{A}_k^n$ if k is finite.

Just like $\mathcal{Z}$, $\mathcal{I}$ shares similar properties:

Proposition 21.1.7: Properties of Vanishing Ideals

let $A, B \subseteq \mathbb{A}^n$ be two algebraic sets. Then

1. If $A \subseteq B$, then $\mathcal{I}(B) \subseteq \mathcal{I}(A)$ (i.e. $\mathcal{I}$ is *contravariant* with respect to inclusion)
2. $\mathcal{I}(A \cup B) = \mathcal{I}(A) \cap \mathcal{I}(B)$
3. $\mathcal{I}(\emptyset) = k[\mathbb{A}^n]$. If k is infinite, $\mathcal{I}(\mathbb{A}^n) = 0$

***Proof* :**
These are similar to proposition 21.1.4.

Notice that we have omitted showing what $\mathcal{I}(A \cap B)$ is. One problem is that $\mathcal{I}(A) \cup \mathcal{I}(B)$ is not necessarily an ideal (since the union of two ideals is not necessarily an ideal). It might be tempting to say that since

$$(\mathcal{I}(A) \cup \mathcal{I}(B)) = \mathcal{I}(A) + \mathcal{I}(B)$$

we can conclude that:

$$\mathcal{I}(A \cap B) = \mathcal{I}(A) + \mathcal{I}(B)$$

However, this is not generally the case (consider the algebraic sets given by the ideals $(x^2 - y)$ and (x)) . We will require some more machinery to prove the appropriate equality (i.e. the Nullstellensatz), but it can already be shown with the current definitions that $\mathcal{I}(A \cap B) \supseteq \mathcal{I}(A) + \mathcal{I}(B)$ (see exercise ref:HERE). With the properties of $\mathcal{I}$ established, we can start getting an idea of what ideals of $k[\mathbb{A}^n]$ are necessary to restrict $\mathcal{Z}$ to make it a bijection:

Proposition 21.1.8: Zero Sets and Coordinate Ideals

Let $k[\mathbb{A}^n]$ be the set of polynomial fructions from $\mathbb{A}^n$ to k. Then

1. If $V \subseteq \mathbb{A}^n$, then $V \subseteq \mathcal{Z}(\mathcal{I}(V))$. If $I \subseteq k[\mathbb{A}^n]$ is an ideal of $k[\mathbb{A}^n]$, then $I \subseteq \mathcal{I}(\mathcal{Z}(I))$.
2. If $V = \mathcal{Z}(I)$ is an algebraic set, then $V = \mathcal{Z}(\mathcal{I}(V))$. Similarly, if $I = \mathcal{I}(V)$, then $I = \mathcal{I}(\mathcal{Z}(I))$. In other words,

$$\mathcal{I}(\mathcal{Z}(\mathcal{I}(V))) = \mathcal{I}(V) \qquad \mathcal{Z}(\mathcal{I}(\mathcal{Z}(V))) = \mathcal{Z}(V)$$

Proof **:**

1. For $V \subseteq \mathcal{Z}(\mathcal{I}(V))$, since $\mathcal{I}(V)$ is the set of all functions such that $f(V) \equiv 0$ (including at least the zero polynomial), then V is certainly part of it's zero set.

   Conversely, for $I \subseteq \mathcal{I}(\mathcal{Z}(I))$, the coordinate ideal of $\mathcal{Z}(I)$ certainly contains I since for any $f \in I$ and $x \in \mathcal{Z}(I)$, $f(x) = 0$. Thus, $I \subseteq \mathcal{I}(\mathcal{Z}(I))$
2. Let $V = \mathcal{Z}(I)$. Using part 1, it suffices to prove $V \supseteq \mathcal{Z}(\mathcal{I}(V))$. If $x \in \mathcal{Z}(\mathcal{I}(V))$, then for every $f \in \mathcal{I}(V)$, $f(x) = 0$. Since $V = \mathcal{Z}(I)$, f is defined as the function such that

What the 2nd part of proposition 21.1.8 tells us is that if we restrict to just the algebraic sets and vanishing ideals of $\mathbb{A}^n$ and $k[\mathbb{A}^n]$ respectively, then $\mathcal{Z}$ and $\mathcal{I}$ are indeed inverses of each other. The goal then would be to be able to characterize vanishing ideals, that is ideals that are of the form $I = \mathcal{I}(V)$ for some algebraic set V. It is in fact a rather involved process to find the nature of the ideal $\mathcal{I}(V)$; a lot of section 21.2.2 will be building up to get to the theorem which will tell us the answer: *Hilbert's Nullstellensatz*. Before we get to answering this question, there are a few other loose ends that we should tie up that are important. For example, we introduced the Zariski topology. What properties must the Zariski topology have, and what does that imply for the corresponding ring $k[\mathbb{A}^n]$? Another important realization is that as $k[x_1, ..., x_n]$ is Noetherian, every ideal is finitely generated: $I = (f_1, ..., f_m) = (f_1) \cup (f_2) \cup \cdots \cup (f_m)$. BY the properties of zero sets:

$$\mathcal{Z}(I) = \mathcal{Z}(f_1) \cap \mathcal{Z}(f_2) \cap \cdots \cap (f_m)$$

In other words, there seem to be a natural theory of decomposition of algebraic sets. Furthermore, we have been shown theoretically what properties algebraic sets have, but how would we go about actually computing one? These questions will receive answers in the following section.

21.1.2 Zariski Topology

In this section, we will be assuming basic topological knowledge from the reader (see appendix C)

A couple of remarks can already be done about the Zariski topology: notice that $\mathcal{T}_{\mathcal{Z}}$ is rarely Hausdorff (also known as T_2): if $\mathbb{A}^1 = k$ is any infinite field, as we've establsihed that the only algebraic sets (and hence closed sets) are the empty set, finite sets, and $\mathbb{A}^1_k$ and so the only non-trivial open sets in $\mathcal{T}_{\mathcal{Z}}$ would be co-finite sets (i.e. $\mathcal{T}_{\mathcal{Z}}$ is the co-finite topology). Then as we can see, if we pick any two points $a, b \in \mathbb{A}^1$, we cannot separate them by two open disjoint sets. However, over

any $\mathbb{A}^1 = k$, the singleton's are always algebraic sets (giving a hint to those who were working on example 21.1.2), and so it is always a Fréchet space (or T_1)[10].

This result generalizes for any k^n

Proposition 21.1.9: Zariski Topology Always T_1

Let k be an infinite field and $\mathbb{A}_k^n$ have the Zariski topology. Then $\mathbb{A}_k^n$ is T_1, i.e. every singleton is always closed, and $\mathbb{A}_k^n$ is not Hausdorff.

***Proof* :**
For any $a = (a_1, a_2, ..., a_n) \in k^n$, Let $I = (x_1 - a_1, ..., x_n - a_n)$. Then $\mathcal{Z}(I) = \{a\}$, showing that singleton's are closed.

To show's it not Hausdorff, we already showed it in the case of $n = 1$ in example 21.1.2, so we'll do k^n with $n > 1$. Take $\mathbb{A}_k^1 \times \{0\}^{n-1}$ with the sub-space topology. Then any closet subset must be finite. Since this is true for any arbitrary axis we choose, the closed set that intersects each of these subspaces can have at most finitely many points, and so the compliment has co-finitely many points. It is now easy to verify that the space is not Hausdorff.

Two topological concepts which should be immediately introduced to be explored as the chapter develops are the following:

Definition 21.1.10: Zariski Dense And Closure

Let $A \subseteq \mathbb{A}^n$ be any subset of the affine space $\mathbb{A}^n$. Then the *zariski closure* of A is the smallest algebraic set containing A, denoted $\overline{A}$. Geometrically, it is the smallest set $\mathcal{A}$ for which every point in A is the zero set of a collection of polynomials.

If $\overline{A} = \mathbb{A}^n$, then A is said to be *zariski dense*. More generally if $A \subseteq V \subseteq \mathbb{A}^n$ such that $\overline{A} = V$, then A is said to be zariski dense in V.

As an exmple, if we give $\mathbb{R}$ the Zariski topology, the closure of any finite set is itself, while the closure of any infinite set is all of $\mathbb{R}$. The Zariski closure of a set is in fact quite easy to characterize using $\mathcal{Z}$ and $\mathcal{I}$ and follows quickly using proposition 21.1.8:

Corollary 21.1.11: Zariski Closure

Let $A \subseteq \mathbb{A}^n$. Then:

$$\overline{A} = \mathcal{Z}(\mathcal{I}(A))$$

***Proof* :**
Hint: $\mathcal{I}(A)$ is the largest ideal on which every element of A is in the zero set, and so $\mathcal{Z}(\mathcal{I}(A))$ is the smallest zero set containing A.

[10] For those familiar with algebraic topology, schemes are almost never T_1, as prime ideals that are not maximal are not closed points

Sticking to $k = \mathbb{R}$ or $k = \mathbb{C}$ for a moment, we can compare the euclidean topology on $\mathbb{R}^n$ with the Zariski topology on $\mathbb{A}^n_{\mathbb{R}}$. (Coarser, and I think is the corsest topology on which all polynomial functions are continuous)

Another important topological consideration would be to find a topological basis for the Zariski topology. Recall a subset of the topology on a space X is a basis B if any open set of X is the union of a family of sets from B.

Proposition 21.1.12: Basis For Zariski Topology

Let $A \subseteq k[\mathbb{A}^n]$. Define

$$D(A) = \mathbb{A}^n_k \setminus \mathcal{Z}(A) = \{a \in \mathbb{A}^n_k \mid f(a) \neq 0,\ \forall f \in A\}$$

Then the collection $\{D(f)\}_{f \in k[A^n]}$ forms a [topological] basis for the Zariski topology on $\mathbb{A}^n$.

The letter D can be thought of as a mnemonic for *doesn't* vanish.

Proof **:**
It suffice to show that for any $a \in U$ for an open set U, there exists an f such that $a \in D(f) \subseteq U$, since then U is the union of $D(f)$'s. Since U is open, there exists an ideal $I \subseteq k[\mathbb{A}^n]$ such that $U = \mathbb{A}^n_k \setminus \mathcal{Z}(I)$. Since $a \in U$, for any $f \in I$, $f(a) \neq 0$, so $a \in D(f)$. Since $(f) \subseteq I$,

$$\mathcal{Z}(I) \subseteq \mathcal{Z}(f)$$

which, after taking the compliment of these sets, we get:

$$a \in D(f) \subseteq U$$

and since a was arbitrary, we see that U is the union of sets $D(f)$, as we sought to show.

The collection of $D(f)$ for each $f \in k[x_1, ..., x_n]$ forms the coarsest basis:

Definition 21.1.13: Distinguished Open Sets

Let $f \in k[x_1, ..., x_n]$. Then the set

$$D(f) = \mathbb{A}^n_k \setminus \mathcal{Z}(f) = \{a \in \mathbb{A}^n_k \mid f(a) \neq 0\}$$

is called he distinguished open set.

The next property we get about the Zariski topology is that it's always Noetherian, that is, if there is an ascending chain of open sets, it must stabilize:

Proposition 21.1.14: Zariski Topology Is Noetherian

Let $\mathbb{A}^n_k$ have the Zariski topology. Then $\mathbb{A}^n_k$ is Noetherian.

Proof **:**
Let $U_1 \subsetneq U_2 \subsetneq \cdots$ be an ascending chain of open sets, so there is a corresponding

$$X_1 \supsetneq X_2 \supsetneq \cdots$$

descending chain of closed sets. Then there is a coresponding

$$I(X_1) \subsetneq I(X_2) \subsetneq \cdots$$

ascending chain of ideals. Then we know that $k[\mathbb{A}^n]$ is *noetherian*, and so this chain must stabalize at some n:

$$I(X_n) = I(X_{n+1}) = \cdots$$

meaning the respective closed chain, and hence open chain, must stabalize, as we sought to show.

We will soon show that the Zariski topology on $\mathbb{A}_k^n$ is in fact always *compact*, but we require the Nullstellensatz to prove this result [11].

The next thing we should consider the continuous functions. As the topology is the weakest in which the zero-sets is closed, it may not surprise the reader that it is the coarsest topology in which polynomials are continuous

Definition 21.1.15: Regular Map on Affine Space

A map $\varphi : \mathbb{A}_k^n \to \mathbb{A}_k^m$ is called a *regular map* if there exists polynomials $\varphi_1, \varphi_2, ..., \varphi_m \subseteq k[x_1, x_2, ..., x_m]$ such that

$$\varphi(a_1, a_2, ..., a_n) = (\varphi_1(a_1, a_2, ..., a_n), \varphi_2(a_1, a_2, ..., a_n)..., \varphi_m(a_1, a_2, ...a_n))$$

Lemma 21.1.16: Coordinate Ring and Regular Map

The map $\varphi : \mathbb{A}_k^n \to \mathbb{A}_k^m$ is regular if and only if the induced map $\varphi^\# : k[\mathbb{A}_k^m] \to k[\mathbb{A}_k^n]$ is a ring homomorphism.

Proof **:**
Let $\varphi : \mathbb{A}_k^n \to \mathbb{A}_k^m$ be a regular map. Then $\varphi^\#$ maps a function $f \in k[\mathbb{A}_k^m]$ to $f \circ \varphi k[\mathbb{A}_k^n]$, namely $f \circ \varphi$ is a polynomial map. As φ has polynomials in each component, $f \circ \varphi \in k[\mathbb{A}_k^n]$. It is then easy to check that it is a ring homomorphism.

Conversely, if $\varphi^\# : k[\mathbb{A}_k^m] \to k[\mathbb{A}_k^n]$ is a k-algebra homomorphism, it map $f \to \varphi^\#(f) = f \circ \varphi \in k[\mathbb{A}_k^n]$ is as a polynomial map. If one of the components of φ where not a polynomial, then take $f = x_i$, so that the composition $f \circ \varphi_i$ is not a polynomial, giving a contradiction. Hence φ must be a regular map, completing the proof.

[11] For those familiar with algebraic geometry, almost all schemes are compact, making the notion much less useful. Instead, the term the notion of *properness* shall be the appropriate replacement

Lemma 21.1.17: Regular Map is Continuous

A regular map $\varphi : \mathbb{A}_k^n \to \mathbb{A}_k^m$ is continuous.

Proof **:**
It suffices to check that the pre-image of every closed set is closed. If $V = \mathcal{Z}(I) \subseteq \mathbb{A}_k^m$ is an algebraic set, we need to show there exists an ideal J such that

$$\mathcal{Z}(J) = \varphi^{-1}(V)$$

To that end:

$$\begin{aligned}\varphi^{-1}(V) &= \{a \in \mathbb{A}_k^n \mid \varphi(a) \in V\} \\ &= \{a \in \mathbb{A}_k^n \mid f(\varphi(a)) = 0 \ \forall f \in I\} \\ &= \left\{a \in \mathbb{A}_k^n \;\middle|\; \varphi^{\#}(f)(a) = 0 \ \forall f \in I\right\}\end{aligned}$$

Consider the set $\varphi^{\#}(I) = \{g \in k[\mathbb{A}_k^n] \mid g = f \circ \varphi\}$. This set would be the desired pre-image ideal, except it is in general not an ideal as it need not be closed under multiplication (recall the image of an ideal need not be an ideal). Thus, take $J = \langle \varphi^{\#}(I) \rangle$. Then it is clear that

$$\mathcal{Z}(J) = \varphi^{-1}(V)$$

and as V was an arbitrary closed set, φ is continuous, as we sought to show.

Note that *not all* continuous maps arise from a ring homomorphism:

Example 21.1.5: Continuous, No Associated Ring Homomorphism
Take $f : \mathbb{A}_{\mathbb{R}}^1 \to \mathbb{A}_{\mathbb{R}}^1$ given by:

$$f(a) = \begin{cases} 2 & a = 1 \\ 1 & a = 2 \\ a & \text{otherwise} \end{cases}$$

Then note that the pre-image of every closed set is closed, and hence it is continuous. However, there are no real polynomial functions that only permute two elements whose only root is 0.

There needs to be more information for the full and faithful correspondence between regular maps and ring homomorphisms. This extra information is a *sheaf* and a *sheaf morphism*. We shall not define these in this book, instead the reader may refer to cite:HERE.

Exercise 21.1.1

1. (if you are interested in cateogires and topology) Show that the Zariski topology is the coarsest topology where singletons are closed.
2. Prove that $(I \cap J)^2 \subseteq IJ \subseteq (I \cap J)$, and conclude that $\mathcal{Z}(I \cap J) = \mathcal{Z}(IJ)$.
3. Let $k = \mathbb{F}_q$ be a finite field. Show that each set-function $\mathrm{Hom}_{\mathbf{Set}}(k^n, k)$ is represented by a polynomial (hint: find a polynomial that is a characteristic function $\chi_{\overline{a}}(\overline{a}) = 1$ and 0 otherwise

and show these generate all the set functions). Conclude the Zariski topology is discrete over finite fields. Thus, the Zariski topology over finite fields reduces to set-theoretic counting arguments.

4. Show that the Zariski topology on $\mathbb{A}_k^n$ is Hausdorff if and only if k is finite.
5. Let $V \subseteq \mathbb{A}_k^n$ and $W \subseteq \mathbb{A}_k^m$ be algebraic sets. Show that $V \times W$ is an algebraic set in $\mathbb{A}_k^{n+m}$. Conclude that $\mathbb{A}_k^n \times \mathbb{A}_k^m \cong \mathbb{A}_k^{n+m}$[12].

21.2 Properties of Affine Sets and Duality

In this section, we explore more properties of affine sets. As affine sets correspond to ideals, we shall see that there is a correspondence between the two objects. In particular it shall will show that properties of algebraic sets will often have corresponding properties in their coordinate rings, and vice-versa. This theme shall play out often in this chapter, where a property proven in a geometric or algebraic setting shall have a corresponding result in the respective dual.

21.2.1 Regular Functions and Duality

The collection of regular function $\mathrm{Hom}_{\mathbf{Reg}}(\mathbb{A}_k^n, \mathbb{A}_k^1)$ can be thought of as the generalization of the dual vector-space, where polynomial functions are taken instead of linear functions. We saw that for each vector subspace $W \subseteq V$, there is a dual space $W^\vee = \mathrm{Hom}_k(W, k)$ which it is isomorphic to. It is thus natural to ask what is the appropriate dual space for an algebraic set $V \subseteq \mathbb{A}_k^n$.

This dual space correspond to some topology on the space V. It is rather easy to describe this topology. In particular, if we have any subset $A \subseteq \mathbb{A}_k^n$, we can define the Zariski topology on A by taking the intersection of all open sets of A. Note that all of the properties of the Zariski topology apply to A since all singleton's are closed (proposition 21.1.9). We will always limit ourselves to taking A to be an algebraic set, since algebraic set are the object of interest[13]. the dual space we wish to establish is a corresponding "polynomial ring" which will generate all of the closed sets of A, just like how $k[x_1, x_2, ..., x_n]$ is needed to generate the closed sets of $\mathbb{A}^n$. Without looking at the next definition, can you see what the definition of such a ring may be?

54

taking a moment

To elucidate the coming definition, let us come back to the algebraic set X defined given by the vanishing set of $x^2 + y^2 - 1$. Then on X, the polynomial $(x^2 + y^2 - 1)$ and $(x^2 + y^2 - 1)y$ map to the same regular function, namely if we take

$$F : k[x, y] \to \mathrm{Hom}_{\mathbf{Reg}}(X, k) \qquad f \mapsto ((a, b) \to f(a, b)) \tag{21.3}$$

we see that both of these shall map to the zero function, namely the function where $f(x, y) = 0$ for all $(x, y) \in X$. Next, notice that $\mathrm{Hom}_{\mathbf{Reg}}(X, \mathbb{A}_k^1)$ is naturally a commutative ring given by point-wise

[12]This seemingly simple result shall be shown to be greatly complicated over projective algebraic sets, see section 21.4.4

[13]Once we show that $\mathbb{A}_k^n$ is compact in the Zariski topology $A \subseteq \mathbb{A}_k^n$ will also be compact, being a closed set of a compact space

addition and multiplication of functions (see equation 21.2). Hence, the map in equation (21.3) has a kernel. As the map is surjective, we get a bijection between

$$\frac{k[x,y]}{\ker(F)} \cong_k \mathrm{Hom}_{\mathbf{Reg}}(X,k)$$

What is the ideal $\ker(F)$? The reader should clue in that this is by definition $\mathcal{I}(X)$ as defined in definition 21.1.6! This leads to the following definition:

Definition 21.2.1: [Affine] Coordinate Ring

Let $k[x_1,...,x_n]$ be a polynomial ring over n variables, X an algebraic set, and $\mathcal{I}(X)$ the corresponding vanishing ideal. Then:

$$k[X] = \frac{k[x_1,...,x_n]}{\mathcal{I}(X)}$$

is called the *coordinate ring* of X. The corresponding regular functions shall be denoted $k[X]$.

Notice how this notation is the same as for regular functions: $k[\mathbb{A}^n]$. Notice that the coordinate ring of $X = \mathbb{A}^n_k$ is $k[X_1,...,X_n]$ since $I(\mathbb{A}^n) = 0$. We can naturally bring the notion of regular function to algebraic sets:

Definition 21.2.2: Regular Maps on Algebraic Sets

Let $X \subseteq \mathbb{A}^n_k$, $Y \subseteq \mathbb{A}^m_k$ be algebraic sets. Then a map $\varphi : X \to Y$ is called a *regular function*, *regular map*, or *morphism*, if there exists polynomials $\varphi_1, \varphi_2, ..., \varphi_m \subseteq k[x_1, x_2, ..., x_m]$ such that

$$\varphi(a_1, a_2, ..., a_n) = (\varphi_1(a_1, a_2, ..., a_n), \varphi_2(a_1, a_2, ..., a_n)..., \varphi_m(a_1, a_2, ...a_n))$$

This definition also allows us to upgrade definition 21.1.2 and definition 21.1.15. To check that $\varphi : X \to Y$ is well-defined, it suffices to check that $\varphi_1, ..., \varphi_m$ as elements of $k[X]$ satisfy the equations of Y.

Proposition 21.2.3: Regular Maps Are Continuous

Let $\varphi : X \to Y$ be a regular function where X, Y are equipped with the Zariski topology. Then φ is continuous.

Proof **:**
Let $V \subseteq Y$ be a closet subset. It suffices to show that $\varphi^{-1}(V)$ is closed. As V is closed, there exists polynomials $f_1, ..., f_n$ such that for each $x \in V$, $f_1(x) = \cdots = f_n(x) = 0$. If $a \in \varphi^{-1}(V)$, then $\varphi(a) \in V$, so

$$f_1(\varphi(a)) = \cdots = f_n(\varphi(a)) = 0 \tag{21.4}$$

As $\varphi = (\varphi_1, ..., \varphi_n)$ as all polynomials, notice that equation (21.4) is a collection of polynomials in $k[X]$ where a vanishes. As this is true for all points in $\varphi^{-1}(V)$, we see that:

$$\varphi^{-1}(V) = \mathcal{Z}(f_1 \circ \varphi, ..., f_n \circ \varphi)$$

showing it is closed, as we sought to show.

Example 21.2.1: Coordinate Ring and Regular Mapping

1. Let $V = \mathcal{Z}(xy-1)$, which represents the hyperbola $y = 1/x$ in $\mathbb{R}^2$. Then the coordinate ring $\mathbb{R}[V] = \mathbb{R}[x,y]/(xy-1)$. In $\mathbb{R}[V]$, $f(x,y) = x$ and $g(x,y) = x + xy - 1$ are the same function since $f - g \in \mathcal{I}(V)$. In $\mathbb{R}[V]$, $\overline{xy} = \overline{1}$, giving us that $\mathbb{R}[V] \cong \mathbb{R}[x, 1/x]$. Thus, for any $f \in \mathbb{R}[x,y]$ and $(a,b) \in V$, $\overline{f}(a,b) = f(a, 1/a)$ for any f that maps to $\overline{f}$.

   Take the regular map $\varphi : V \to \mathbb{A}^1_k$ mapping $(x,y) \mapsto x$. The image of this map is $\mathbb{A}_k \setminus \{0\}$, which is certainly *not* a closed set (can the reader see why). It is an open set, being the compliment of the closed set $\{0\}$. However, if we take the map to be $\varphi' : V \to \mathbb{A}^2_k$ mapping $(x,y) \mapsto (x,0)$, then the image is no loner an open set, and hence it is possible for the images of regular maps to be neither open or closed. Further exploration the properties of the image of regular maps shall be done section section 21.4.3.

2. Take $k = \mathbb{F}_p$ and consider $\mathbb{A}^n_{\mathbb{F}_p}$. Then the map:

$$\varphi(a_1, ..., a_n) = (a_1^p, ..., a_n^p)$$

   is a regular function. If $X \subseteq \mathbb{A}^n_{\mathbb{F}_p}$ is an algebraic set, notice that φ will always take X to itself, since if f is one of the polynomials that define X so that $f(a) \equiv 0 \bmod p$, then:

$$f(a_1^p, ..., a_n^p) = (f(a_1, ..., a_n))^p \equiv (0)^p \equiv 0 \bmod p$$

   The regular function φ is called the *Frobenius mapping*

3. Take $t \mapsto (t^3, t^2)$. This is a regular function and maps $\mathbb{A}^1_{\mathbb{R}}$ to the curve given by $x^3 = y^2$ in $\mathbb{A}^2_{\mathbb{R}}$.

(give examples, and then show in the examples that the representation of a variety is "coordinate dependent", and so we will define coordinate rings in a "coordinate-free" way)

To sum up the correspondence between an algebraic set and its dual:

Theorem 21.2.4: Regular Function Algebraic Correspondence

Let $V \subseteq \mathbb{A}^n$ and $W \subseteq \mathbb{A}^m$. then there is a bijective correpspondence between the morphisms between V and W and the associated k-algebra homomorphisms, more precisely

$$\{\text{regular functions from } V \text{ to } W\} \leftrightarrow \{k\text{-algebra homomorphisms from } k[W] \text{ to } k[V]\}$$

$$\left\{\begin{matrix}\text{regular functions} \\ \varphi : V \to W\end{matrix}\right\} \underset{\psi^a}{\overset{f \circ \varphi}{\rightleftarrows}} \left\{\begin{matrix}k\text{-algebra homomorphism} \\ \varphi^\# : k[W] \to k[V]\end{matrix}\right\}$$

More generally, if k-**Aff** is the category whose objects are algebraic sets and morhpism are regular functions and k-$\mathbf{Alg}_0$ is the category whose objecst are coordinate rings (reduced, commutative, finite-type k-algebras) and morphisms are k-homomorphisms, then k-$\mathbf{Aff}^{\text{op}} \cong k$-**Alg**.

Proof **:**
it as an exercise

Thus, if $f : X \to Y$, then $(f^{\#})^a = f$ and if $\varphi : k[Y] \to k[X]$, then $(\varphi^a)^{\#} = \varphi$. Similarly to the dual theory provided in section 13.3, the dual map gives information about our original map:

Proposition 21.2.5: Regular Map Properties

Let $\varphi : X \to Y$ be a regular map with $\varphi^{\#} : k[Y] \to k[X]$ the associated k-algebra homomorphism. Then:

1. The kernel of $\tilde{\varphi}$ is $\mathcal{I}(\varphi(X))$: $\ker \tilde{\varphi} = \mathcal{I}(\varphi(X))$. Furthermore, the Zariski closure of $\varphi(X)$ is $\mathcal{Z}(\ker(\tilde{\varphi}))$: $\overline{\varphi(X)} = \mathcal{Z}(\ker(\tilde{\varphi}))$.
2. As a consequence, the homomorphism $\tilde{\varphi}$ is injective if and only if $\varphi(X)$ is Zariski dense in Y, or equivalently if $\overline{f(X)} = Y$
3. $\varphi(v) = w$ if and only if $\varphi^{\#}(\mathcal{I}(v)) = \mathcal{I}(w)$

Proof **:**
exercise.

Let us now take a moment to look at isomorphism of algebraic sets:

Definition 21.2.6: Isomorphism of Varieties

let X, Y be two algebraic sets. Then if there exists a regular function $f : X \to Y$ with regular inverse, then the two varieties are said to be *isomorphic*

Note that this was not called a homeomorphism between two varieties. All isomorphisms are homeomorphisms, but not all homeomorphisms are isomorphisms

Example 21.2.2: Isomorphism of Varieties

1. Consider $X = \mathcal{Z}(y - x^n) \subseteq \mathbb{A}^2_k$. Then $X \cong \mathbb{A}^1_k$ with $\varphi : (x, y) = x$ that has inverse $\psi(t) = (t, t^n)$
2. Take $X = \mathcal{Z}(x^3 - y^2) \subseteq \mathbb{A}^2_k$. Then X can be paramaterized by the regular map $f : \mathbb{A}^1_k \to X$ by $f(t) = (t^2, t^3)$. However, this is not an isomorphism: the inverse map is:
$$g(x, y) = \frac{y}{x}$$
however, there is no polynomial associated to g (can the reader see/prove why?), and hence it is not an isomorphism.
3. The reader may have recalled that $\mathcal{Z}(x^3 - y^2)$ has a cusp at 0, and believe this is the obstruction to the inverse. It is in fact a bit more subtle. Consider the map $\varphi : \mathbb{A}^1_k \to \mathbb{A}^1_k$ given by $\varphi(t) = t^3$. This is a bijective regular map, however it certainly has no regular inverse, namely $\psi(t) = t^{1/3}$ is not a regular function.

On the other hand, the map given in the previous example had a rational polynomial inverse. Rational functions appear as localization of the coordinate ring, and so they offer a different class of functions and a different classification problem that turns out to be much easier, as we shall demonstrate in section 21.3.

It is in general difficult to ascertain isomorphic classes of algebraic sets or varieties. Some isomorphism classes that we shall state here but not prove are:

1. Smooth genus 0 curves are isomorphic to conic curves in $\mathbb{P}^2_k$
2. classes of elliptic curves can be classified by their j-invariant
3. Algebraic surfaces are birationally equivalent (defined in section 21.3), but have different isomorphism classes given their Kodaira dimension.

Exercise 21.2.1

1. Show that $k[X \times Y] \cong k[X] \otimes_k k[Y]$ for X, Y algebraic sets. Show this is another way of showing that $\mathbb{A}^n_k \times \mathbb{A}^m_k \cong_{\mathbf{Reg}} \mathbb{A}^{n+m}_k$
2. Show that $X \times Y$ for algebraic sets X, Y is unique up to isomorphism, that is if $\varphi : X \to X'$ and $\psi : Y \to Y'$ are isomorphism, there is a unique isomorphism from $X \times Y$ to $X' \times Y'$ commuting with φ and ψ.

21.2.2 Closed Sets and Radical Ideals

As we've seen in proposition 21.1.4, given some polynomial f or collection of polynomials $S \subseteq k[x_1, ..., x_n]$, the set of zero's of these polynomials does not change if we take higher powers of the polynomials (ex. $f, f^2, f^3, ...$ all produce the same zero set). Thus, studying the set $\mathcal{Z}(I^n)$ is equivalent to studying $\mathcal{Z}(I)$. This property inspires the following definition:

Definition 21.2.7: Radical Ideal

Let $I \subseteq R$ be an ideal of a commutative ring R. Then:

1. The *radical of* I in R is the collection elements for which some power of them lie in I, that is:
$$\sqrt{I} = \mathrm{rad}(I) := \left\{a \in R \mid a^k \in I \text{ for some } k \in \mathbb{N}\right\}$$
here $\mathrm{rad}(I)$ stands for the radical of I. It is often written as $\sqrt{I}$.
2. The radical of the zero ideal ($\mathrm{rad}(0) = \sqrt{(0)}$) is called the *nilradical* of R
3. An ideal I is called a *radical ideal* if $I = \mathrm{rad}(I)$ ($I = \sqrt{I}$)

Note that the nilradical by definition contains all the nilpotent elements of R, and that $\mathrm{rad}(\mathrm{rad}(I)) = \mathrm{rad}(I)$.

Proposition 21.2.8: $\mathrm{rad}(I)$ Is An Ideal

Let $I \subseteq R$ be an ideal of a commutative ring. Then $\mathrm{rad}(I)$ is indeed an ideal (containing I). Furthermore, $\mathrm{rad}(I)/I$ is the nilradical of R/I.

The last statement shows us that R/I has no nilpotent elements if and only if $I = \mathrm{rad}(I)$.

Proof **:**
Exercise

A type of ideal in which we've seen a similar property is the prime ideal, in which $ab \in I$ if and only if $a \in I$ or $b \in I$. One important way of looking at radical is in how they correspond to prime ideals. The prime ideals in a sense "generate" the radical ideals:

Proposition 21.2.9: Radical Through Prime Ideals

Let $I \subsetneq R$ be a proper ideal of a commutative ring and $\mathrm{rad}(I)$ the radical ideal of I. Then $\mathrm{rad}(I)$ is the intersection of all prime ideals containing I. In particular, the nilradical is the intersection of all prime ideals of R.

Proof **:**
Since I is the intersection of all prime ideals containing I if and only if (0) in R/I is the intersection of all prime ideals of R/I (by proposition 21.2.8), it suffices to prove it in the case of $I = (0)$ (i.e. that the nilradical is the intersection of all prime ideals of R). Let

$$N' = \bigcap_{\substack{P \subseteq R \\ P \text{ prime}}} P$$

Let N represent the nilradical.

($N \subseteq N'$) Let $x \in N$ so that there exists some n such that $x^n = 0$. If P is any prime, then $0 = x^n \in P$. Since P is prime, then eiher $x \in P$ or $x^{n-1} \in P$. If we assume $x^{n-1} \in P$, we can continue this process until we get $x^2 \in P$, so it must be that $x \in P$. Since P and x were arbitary, we get atht $x \in N'$

($N \supseteq N'$) We'll use Zorn's lemma and show that if $a \notin N$, then $a \notin N'$. Given a is not an element of the nilradical ideal, define the family of ideals $\mathcal{S}$ that does not contain any power of a:

$$\mathcal{S} = \left\{ I \subseteq R \;\middle|\; \forall k \in \mathbb{N}_{>0},\ a^k \notin I \right\}$$

This collection is nonempty since $(0) \in \mathcal{S}$ (since $a \notin N$). Furthermore, if a^k (for any k) is not contained in any ideal in a chain $I_1 \subseteq I_2 \subseteq \cdots$, then it is not contained in the union of those elements. Thus, chains in $\mathcal{S}$ have upper bounds. By Zorn's lemma, $\mathcal{S}$ must contain a maximal element, say P. I claim that P is in fact a prime ideal. Suppose for contradiction that $xy \in P$ but $x, y \notin P$. By maximality of P, it must be that the larger ideals $(x) + P$ and $(y) + P$ contains a power of a, say $a^n \in (x) + P$ and $a^m \in (y) + P$. But then:

$$a^{n+m} \in (xy) + P = P$$

contardicting the fact that $P \in \mathcal{S}$. Thus, P is a prime ideal that does not contain a, and so $a \notin N'$, as we sought to show

Corollary 21.2.10: Prime Or Maximal Then Radical

Let $I \subseteq R$ be a prime ideal. Then I is a radical ideal

Proof **:**
Take

$$P = \sqrt{P}$$

Example 21.2.3: Radical Ideals

1. Let R be a UFD. Given an ideal, finds its radical.
2. The ideal $(x^3 - y^2)$ is prime in $k[x, y]$, and hence radical
3. If $l_1, l_2, ..., l_m$ is a collection of linear polynomials in $k[x_1, ..., x_n]$, then $I = (l_1, l_2, ..., l_m)$ is either $k[x_1, x_2, ..., x_n]$ or a prime ideal, and hence radical.

In the context of this chapter, since we are working over polynomials rings, we are working over Noetherian rings. Due to this, the elements of a radical ideal $\text{rad}(I)$ can be related back to elements of I quite naturally:

Proposition 21.2.11: Radical Ideals In Noetherian Rings

Let R be a noetherian ring and $I \subseteq R$ an ideal of R. Then there exists a $k \in \mathbb{N}$ such that $I^k \subseteq \text{rad}(I)$. In particular, if N is the nilradical ideal of R, then $N^k = 0$ for some $k \in \mathbb{N}$ (i.e. N is a nilpotent ideal)

Proof **:**
Let $I \subseteq R$ be any ideal and $\text{rad}(I)$ its corresponding radical ideal. Since R is Notherian, $\text{rad}(I)$ is finitely generated, say $\text{rad}(I) = (a_1, a_2, ..., a_n)$. By definition of $\text{rad}(I)$, for each a_i there exists a k_i such that $a^{k_i} \in I$. Take $k = \max(k_i)$. Then each element of $(\text{rad}(I))^{kn}$ is generated by elements of the form $a_1^{d_1} \cdots a_n^{d_n}$ where $d_1, d_2, ..., d_n = kn$ and at least one fact must have $d_i \geq k$. Then by construction, for that factor, $a_i^{d_i} \in I$, and hence each generator of $(\text{rad}(I))^{kn}$ is in I, showing that

$$(\text{rad}(I))^{kn} \subseteq I$$

as we sought to show.

Now, returning to the fact that $\mathcal{Z}(I) = \mathcal{Z}(I^k)$, we now have that $\mathcal{Z}(I) = \mathcal{Z}(\sqrt{I})$. Remember how we were trying to characterize the ideals for which $\mathcal{Z}$ and $\mathcal{I}$ are inverses, i.e. what are the ideals $\mathcal{I}(\mathcal{Z}(I))$? We can now state one direction of this:

Proposition 21.2.12: Radical and Algebraic Sets

Let $I \subseteq R$ be an ideal where R is a polynomial ring. Then

$$\sqrt{I} \subseteq \mathcal{I}(\mathcal{Z}(I))$$

Proof **:**
exercise

It would be great if we could get $\sqrt{I} = \mathcal{I}(\mathcal{Z}(I))$, however this is not true in general: consider $I = (x^2 + 1) \subseteq \mathbb{R}[x]$. Since $x^2 = 1$ is irreducible, it is a maximal ideal, and so $I = \sqrt{I}$. Then as we are over the reals, $\mathcal{Z}(I) = \emptyset$, so $\mathcal{I}(\mathcal{Z}(I)) = \mathbb{R}[x]$ But $I \not\supseteq \mathbb{R}[x]$, showing equality doesn't hold. The blocker here was that $x^2 + 1$ doesn't factor in $\mathbb{R}[x]$: if we were working over $\mathbb{C}$, then we would get that $(x^2 + 1) = ((x + i)(x - i))$, telling us that $\mathcal{Z}(I) = \{i, -i\}$, which by the basic properties of $\mathcal{I}$ we have that $\mathcal{I}(\mathcal{Z}(I)) = ((x - i)(x + i)) = (x^2 + 1) = I$, showing us in this case that $I = \sqrt{I}$. In fact, the underlying field k being algebraically closed is the *only* other condition we need for equality to hold; this will be theorem 21.2.26 (Hilbert's Nullstellensatz). It is tricky to show this, since it is in a sense a very simple condition: All we need to characterize the zero sets of the polynomials (over a closed field) $\mathcal{Z}(I)$ in terms of an ideal is to take the radical of I.

21.2.3 Affine Variety and Prime Ideals

We now return to the earlier fact that $\mathcal{Z}(I) = \mathcal{Z}(f_1, ..., f_n) = \mathcal{Z}(f_1) \cap \cdots \cap \mathcal{Z}(f_n)$. This tells us that we can decompose algebraic sets into smaller components. We will therefore take a moment to see if we can find "decomposition" properties of algebraic sets, and see if there is a corresponding property the coordinate ideals will poses

Definition 21.2.13: [Affine] Variety

let $V \subseteq \mathbb{A}^n$ be an algebraic set. The set V is reducible if for $V_1, V_2 \subseteq V$ where $V_1, V_2 \neq V$

$$V = V_1 \cup V_2$$

If there does not exist such sets, we say that V is *irreducible*, or more commonly an *algebraic variety*

55: Remark

Note that this notion of irreducibility is in fact a topological notion, we simply choose the Zariski topology. In particular, X is irreducible if $X = X_1 \cup X_2$ for some closed subsets X_1 and X_2, then either $X = X_1$ or $X = X_2$. For those who are curious, it may be shown that irreducibility implies connected. Furthermore, the condition transaltes to open sets to say that if

$$\emptyset = U_1 \cap U_2$$

where U_1, U_2 open, then either $U_1 = \emptyset$ or $U_2 = \emptyset$. It follows that it is equivalent to say that every open set is *dense* in an irreducible space. This contrasts heavily with any Hausdorff space, since by definition a Hausdorff space *must* have disjoint open subsets if the space is not empty or a singleton.

The name *algebraic variety* comes from HERE. Note that this condition automatically implies that V_1 and V_2 are non-empty (do you see why?). If V is an irreducible algebraic set, then the corresponding ideal has a very nice property:

Proposition 21.2.14: Coordinate Ideal of Algebraic Variety

A set $V \subseteq \mathbb{A}^n$ is an algebraic variety if and only if $I(V)$ is a prime ideal

Proof **:**

($\Rightarrow$) Taking the contrapositive, let's assume $I(V)$ is not prime. Then there exists $a, b \in k[x_1, .., x_n]$ such that $ab \in I(V)$ but $a \notin I(V)$ and $b \notin I(V)$. But then:

$$V = (V \cap \mathcal{Z}(a)) \cup (V \cap \mathcal{Z}(b))$$

with $V \cap V(a) \subsetneq V$, $V \cap V(b) \subsetneq V$, showing that V is in fact reducible

($\Leftarrow$) Doing the contrapositive again, let's assume V is reducible, so $V = V_1 \cup V_2$, where both $V_i \subsetneq V$. Then by the contra-positive correspondence of inclusion between algebraic sets and ideals, $I(V) \subsetneq I(V_i)$, $i \in \{1, 2\}$. Since the inclusion is proper, choose $a \in I(V_1)$ and $b \in I(V_b)$ such that $a, b \notin I(V)$. Then since a and b are polynomials which are zero on their respective algebraic sets and $V = V_1 \cup V_2$, $ab \in I(V)$. However, $a, b \notin I(V)$, showing that $I(V)$ is not prime

Corollary 21.2.15: Equivalent characterization

Let $V \subseteq \mathbb{A}^n$ be an algebraic set. Then V is an algebraic variety if and only if $k[V]$ is an integegral domain

This means there is a certain correspondence between being a prime and being irreducible as a variety. To deepen the conection, let's prove that every algebraic set can be decomposed into algebraic varieties:

Theorem 21.2.16: Irreducible Decomposition

Let $V \subseteq \mathbb{A}^n$ be an algebraic set. Then there exists a unique set of algebraic varieties $V_1, ..., V_n$ such that

$$V = \bigcup_i V_i \qquad V_i \not\subseteq V_j \quad i \neq j$$

Proof **:**

First, recall that since $k[x_1, ..., x_n]$ is Noetherian, any collection of ideals contains a maximal ideal. By the correspondence between ideals and algebraic sets, there exists a corresponding minimal algebraic sets who is not contained in any other algebraic set (or else it's corresponding ideal would contain the maximal ideal, a contradiction).

Now, Let $S = \{V \subseteq \mathbb{A}^n \mid V \text{ is not the union of finitely many algebraic varieties}\}$. If S were empty, we're done, so For the sake of contradiction let's assume it's not. Then the corresponding set of ideals must contain a maximal ideal, so S contains a corresponding minimal algebraic set. Since $V \in S$, V cannot be a variety, and so there exists a decomposition $V = V_a \cup V_a$. Since V_a and V_b are not in S (or else that would contradict minimality), they both admit a finite decomposition into irreducibles, say:

$$V_a = V_{1,1} \cup \cdots \cup V_{1,n_1} \qquad V_b = V_{2,1} \cup \cdots \cup V_{2,n_2}$$

But then

$$V = \bigcup_{i,j} V_{i,j}$$

showing that V is in fact decomposable into irreducible, a contradiction to the fact that $V \in S$. Thus, S must be empty.

If it happens that $V_{i,j} \subseteq V_{i',j'}$ for appropriate indices, simply discard the smaller one, so that the remaining $V_1, ..., V_n$ will have the property of $V_i \not\subseteq V_j$ for $i \neq j$, fulfilling the 2nd condition.

For uniqueness, suppose that there was a second decomposition $V = W_1 \cup \cdots \cup W_m$. Then by some set theory manipulation, we see that:

$$V_i = \bigcup_j (W_j \cap V_i)$$

so $V_i \subseteq W_{j(i)}$ for some appropriate $j(i)$. Symmetrically, we have that $W_{j(i)} \subseteq V_k$, implying $V_i \subseteq V_k$. But since no proper-subsets exists, it must be that $i = k$ and $V_i = V_k$, so $V_i = W_{i(j)}$, meaning each V_i equals to some $W_{i(j)}$. Symmetrically, each W_j is equal to some $V_{i(j)}$, completing the proof.

Definition 21.2.17: Irreducible Decomposition

Let V be an algebraic set. Then the *irreducible decomposition of* V is

$$V = V_1 \cup \cdots V_n \quad V_i \not\subseteq V_j i \neq j$$

for appropriate algebraic varieties V_i, $1 \leq i \leq n$

56: Remark

In a more general topological setting, it can be shown that if X is a Noetherian space (i.e. satisfies A.C.C. on open sets, so D.C.C. on closed sets), X can be decomposed into irreducible sets just liked was done in theorem 21.2.16. It then follows easily that any closed subset of X has this property too.

Example 21.2.4: Affine Varieties

1. All linear polynomials produce affine varieties. Thus, the study of linear algebra can be thought as the study of varieties with multiplicity 1.
2. $\mathcal{Z}(xy-1) \subseteq \mathbb{A}^2_{\mathbb{R}}$ is an affine variety, as $\mathbb{R}[x,y]/(xy-1) \cong \mathbb{R}[x,1/x]$. Note that $xy-1$ has two branches, however the variety is still irreducible and connected in the Zariski topology. Show that there is no variety that is only one branch of $xy-1$ (show that if two affine varieties agree on an open subset, they agree everywhere).

Primary Decomposition

let us now parallel the discussion of irreducible decomposition using rings.

(Note how this would imply the inersection of two primary ideals is not primary, and that is indeed the case: $(x),(y) \subseteq R[x,y]$ are prime ideals if R is integral, howeve $(x) \cap (y) = (xy)$, but (xy) is claerly not prime (x and y is not in (xy)). However, this does imply the intersection of radical ideals is radical)

In the integers any $n \in \mathbb{Z}$ can be factored into $n = p_1^{n_1} \cdots p_k^{n_k}$. for primes p_i. Unique factorization does not hold in general integral domains, however unique factorization of ideals holds in Noetherian rings! The generalization of prime power is a *primary ideal*.

Definition 21.2.18: Primary Ideal

Let R be a ring and $I \subseteq R$ an ideal. Then I is said to be a *primary ideal* if

$$xy \in I \Rightarrow x \in I \qquad \text{or} \qquad y^n \in I \text{ for some } n > 0$$

Translated into quotients, this is saying

$$I \text{ is primary} \qquad \Leftrightarrow \qquad R/I \neq 0 \text{ and every zero-divisor in } R/I \text{ is nilpotent}$$

Certainly every prime ideal is primary, and the contradction of a primary ideal is primary since if $f : A \to B$ is a homomorphism and I is a primary ideal of B ,then $A/I^c \cong B/I$.

Lemma 21.2.19: Radical Of Primary Ideal

let $I \subseteq R$ be a primary ideal. Then $\sqrt{I}$ is the smallest prime ideal containing I

Proof **:**
It suffices to show that $\sqrt{I}$ is prime. Let $xy \in \sqrt{I}$ so that $(xy)^m \in I$. Then for some $m > 0$, either $x^m \in I$ or $y^{nm} \in I$ for some $n > 0$. But that's the same as etiehr $x \in \sqrt{I}$ or $y \in \sqrt{I}$, completing the proof.

Example 21.2.5: Primary Ideals

1. In $\mathbb{Z}$, the primary ideals are (0) and (p^n) for prime p.
2. In $A = k[x, y]$ take $q = (x, y^2)$. Then $A/q \cong k[y]/(y^2)$, showing q is primary. The radical of q is (x, y). Notice that $\mathrm{rad}(I)^2 \subseteq q \subseteq \mathrm{rad}(I)$, showing that primary ideals are not necessarily a prime-power.
3. Note that if P is a prime ideal, P^n is *not* necessarily a primary ideal, even though its radical is the prime ideal p. For example, let $A = k[x, y, z]/(xy - z^2)$ and $\overline{x}, \overline{y}, \overline{z}$ are the images of x, y, z. Then $p = (\overline{x}, \overline{y})$ is prime ($A/p \cong k[y]$ is an integral domain) and $\overline{xy} = \overline{z}^2 \in p^2$, but $\overline{x} \notin p^2$ and $\overline{y} \notin \sqrt{p^2} = p$. Thus, p^2 is not primary.

We do get a partial converse if our ideal is maximal

Proposition 21.2.20: Radial Maximal, Then Primary

Let $I \subseteq R$ be an idealst $\sqrt{I}$ is maximal. Then I is primary. Hence, the power of a maximal ideal is primary

Proof **:**
Let $\sqrt{I} = m$. Ten the image of m in A/I is the nilradical of A/I, hence A/I has only one prime ideal. Then every element of A/I is either a unit or nilpotent, so every zero-divisor of A/I is nilpotent, completing the proof.

Theorem 21.2.21: Lasker-Noether Theorem

Let A be a Noetherian domain of dimension 1. Then every non-zero ideal $\mathfrak{a} \subseteq A$ can be uniquely expressed as a product of primary ideals whose radicals are all distinct

Proof **:**
exercise

Corollary 21.2.22: Prime Ideals In Noetherian Rings

Let R be a Noetherian ring, $I \subseteq R$ be an ideal, $P \subseteq R$ a prime ideal. Then:

1. $I \subseteq P$ if and only if P contains one of the associated primes of I. In other words, $I \subseteq P$ if and only if P contains one of the isolated prime of I, i.e., the isolated primes of I that are precisely the minimal elements in the set of all prime ideals containing I. In particular, there are only finitely many minimal elements among the prime ideals containing I
2. The radical of I is the intersection of the associated prime ideal of I, and so also the intersection of the isolated prime ideals of I
3. There exists prime ideals $P_1, ..., P_n$ (not necessarily distinct) containing I such that

$$P_1 P_2 \cdots P_n \subseteq I$$

Proof **:**

1. here
2. here
3. For the sake of contradiction, let's say there exist ideal(s) which do not contain a product of prime ideals. Then such a set of ideals, say $\mathcal{S}$ is non-empty and so it has a maximal memeber M. Then M cannot be prime since then it's trivially be a product of prime ideals, and so there must exist $r, s \in R \setminus M$ such that $rs \in M$ but $r, s \notin M$. Then the ideal $M + (r)$ and $M + (s)$ are strictly bigger than M and so cotnain a product of prime ideals say $P = p_1, p_2, ..., p_n$. Hence, so does the product of these ideals. However:

$$P \subseteq (M + (r))(M + (s)) \subseteq M$$

a contradiction, hence every ideal contains a product of prime ideals.

Exercise 21.2.2

1. Show that the product of two irreducible sets is irreducible.
2. Show that the intersection of two irreducibles sets need not be irreducible (hint: consider the varieties given by $x^2 - yw = 0$ and $xy - zw = 0$; show their intersection gives a twisted cubic curve and a line).

21.2.4 Computation of Algebraic Sets and k-algebras

That title is straight up from Dummit and Foote. It basically uses Gröbner bases to find such values. This feels like something to eventually understand.

21.2.5 Hilbert's Nullstellensatz

In this section, we shall make precise the correspondence between which ideals of the coordinate ring of an affine set correspond to closed subsets of the affine set. In this exploration, we shall also produce a new tower condition on ring extensions via Noether's Normalization theorem which builds upon the theme of separating an extension to "property" and "not property" (ex. separable and inseparable, algebraic and transcendental[14]) Recall if $a_1, a_2, ..., a_n \in S$ for a k-algebra S, these elements are called *algebraically independent* if

$$k[x_1, x_2, ..., x_n] \to k[a_1, a_2, ..., a_n] \qquad x_i \mapsto a_i$$

is an isomorphism. Noether's Normalization theorem is the generization of the idea that any field extension can be split into a transcendental extension, followed by an algebraic extension:

Theorem 21.2.23: Noether's Normalization Theorem

Let k be a field and suppose that $A = k[r_1, ..., r_n]$ is a finitely generated k-algebra. Then for some q, $0 \leq q \leq n$, there are algebraically independent elements $y_1, y_2, ..., y_q \in A$ such that A is integral over $k[y_1, y_2, ..., y_q]$

***Proof* :**
If $r_1, r_2, ..., r_n$ are algebraically independent, then A is an integral extension of itself. Otherwise, if $r_1, ..., r_n$ are algebraically *dependent*, then there exists some non-trivial polynomial $f(x_1, ..., x_n) \in k[x_1, x_2, ..., x_n]$ such that $f(r_1, r_2, ..., r_n) = 0$. f is a multinomial, monomials of the form $ax_1^{m_1}x_2^{m_2}\cdots x_n^{m_n}$ which have degree $m_1 + \cdots + m_n$. The degree of f is taken to be the maximum of the degree of all the monomials. Renumbering the polynomial if necessary, we assume that f is a non-constant polynomial in x_n (non-constant since $r_1, r_2, ..., r_n$ are algebraically dependent) with coefficeints in $k[x_1, ..., x_{m-1}]$

Now, we will perform a sort of change of varaibles to f that "normalizes" f into a *monic polynomial in* x_n (giving the namesake to this theorem). The coefficents of this f as a polynomial in x_n will form a subring A generated over k by $m-1$ elements, at which point we shall imply induction (the base case will be trivial).

First, define the integers $\alpha_i = (1+d)^i$. Using these, define the variables $X_i = x_i - x_n^{\alpha_i}$, $1 \leq i \leq m-1$. Define a new polynomial:

$$g(X_1, ..., X_{m-1}, x_n) = f(X_1 + x_n^{\alpha_1}, ..., X_n + x_n^{\alpha_{m-1}}, x_n)$$

so that $g \in k[X_1, ..., X_{m-1}, x_n]$. The key property of this polynomial is the following: for each monomial term in f, it contributes a signle term of the form a constant (in $k[x_1, ..., x_{m-1}]$) times x_n^e to g. Importantly, the choice of α_I insured that distinct monomials in f give different values of e; we can view the degrees of the monomials in the new variables as integers expressed in base $d+1$. Let's say N is the highest power of x_n. Then we get that:

$$g = cx_n^N + \sum_{i=1}^{N-1} h_i(X_1, ..., X_{m-1})x_n^i$$

[14]Another such example will appear in the next chapter, and shall be ramified and unramified

for some nonzero $c \in k$. Returning to the original fact that $f(r_1, ..., r_n) = 0$, letting $s_i = r_i - r_n^{\alpha_i}$, we get:

$$\frac{1}{c} g(s_1, ..., s_{n-1}, r_n) = \frac{1}{c} f(r_1, ..., r_{n-1}, r_n) = 0$$

showing that r_n is integral over $B = k[s_1, ..., s_{m-1}]$. Furthermore, each r_i is integral over $B[r_n]$ since r_i is the root of the monic polynomial $x - (s_i + r_n^{\alpha_i})$, thus A integral over $B[r_n]$. By transitivity of integrality, A is integral over B. Since B is a k-algebra generated by $n-1$ elements, by the principle of induction (HERE) (Dummit simply concludes here)

Noether's Normalizaiton can be thought of as a very geometric result: we shall see that any finitely generated k-algebra shall be an n-dimensional affine set along with some extra important information given by the integral extension, more on this in section 21.5. One very important is that if we have an extension F/k that is a finite-type k-algebra, is finitely generated if we allow for both addition and multiplication, it is finitely gendered by *just* allowing addition!!

Theorem 21.2.24: Zariski's Lemma

Let F/k be a field extension, and assume F is a finite-type k-algebra. Then F/k is a finite (and hence algebraic) extension.

***Proof* :**
Since F is a finite-type k-algebra, $F = k[r_1, r_2, ..., r_n]$. By Noether's Normalization Theorem, there exists a $q \in \mathbb{N}$ (possibly 0) enumerating variables $y_1, ..., y_q \in F$ such that F is integral over $k[y_1, y_2, ..., y_q]$, i.e. F is a finitely generated $k[y_1, ..., y_q]$-module. We can re-phrase this by saying we have a finite integral extension $F/k[y_1, y_2, ..., y_q]$. By proposition 20.5.6, since F is an integral domain, F is a field if and only if $k[y_1, y_2, ..., y_q]$ is a field. But then, $k[y_1, y_2, ..., y_q]$ is a field if and only if there are no indeterminate variables, i.e. $q = 0$. Thus, we get the finite integral extension F/k, which is equivalent to saying we get a finite field extension, as we sought to show.

The last two theorem's are true no matter what field we are working over. The following theorem requires k to be algebraically closed.

Theorem 21.2.25: Hilbert's Nullstellensatz - Weak Version

Let k be an algebraically closed field. Then M is a maximal ideal of the polynomial ring $k[x_1, ..., x_n]$ if and only if

$$M = (x_1 - a_1, x_2 - a_2, ..., x_n - a_n)$$

for any $a_1, a_2, ..., a_n \in k$. As a result, the maps:

$$\left\{\text{Points in } \mathbb{A}^n\right\} \underset{\mathcal{Z}}{\overset{\mathcal{I}}{\rightleftarrows}} \left\{\text{Maximal ideals in } k[\mathbb{A}^n]\right\}$$

are in bijective correspondence, and if $I \subseteq k[\mathbb{A}^n]$ is a proper ideal, then $\mathcal{Z}(I) \neq \emptyset$

Proof **:**
Certainly, $(x_1 - a_1, ..., x_n - a_n)$ is a maximal ideal in $k[x_1, x_2, ..., x_n]$ since it's quotient is k. Conversely, suppose M is a maximal ideal in $k[x_1, x_2, ..., x_n]$. Consider $F = k[x_1, x_2, ..., x_n]/M$. Then since $k \subseteq F$, F/k is a field extension that is also a finitely generated k-algebra (generated by $\overline{x_1}, ...\overline{x_n}$). By Zariski's lemma, F/k is a finite field extension, and hence is algebraic over k.

Now, we take advantage of the fact that k is algebraically closed, meaning there does not exist a non-trivial algebraic extension of k (lemma 17.5.2), implying that $F \cong_k k$, in particular we $\overline{x_i} \mapsto a_i$ for some $a_i \in k$. Thus, this is a k-algebra homomorhpism, and $\overline{a_i} \in F$, $\overline{x_i - a_i} \mapsto a_i - a_i = 0$, hich since the to fiels are isomorphic, they are in bijective corresponence, iplying $\overline{x_i - a_i} = \overline{0}$, i.e. $x_i - a_i \in M$. Thus, we must have $x_i - a_i \in M$ for every x_i, implying that $(x_1 - a_1, ..., x_n - a_n) \subseteq M$, which by maximality tells us that $(x_1 - a_1, ..., x_n - a_n) = M$.

As a consequence, if I is any ideal of $k[x_1, x_2, ..., x_n]$, then $I \subseteq M$ for some maximal ideal M. By what we've just shown, for some $a_1, ..., a_n$, $M = (x_1 - a_1, ..., x_n - a_n)$. Thus, $(a_1, ..., a_n) \in \mathcal{Z}(I)$, showing that $\mathcal{Z}(I)$ is non-empty.

Theorem 21.2.26: Hilbert's Nullstellensatz

Let k be an algebraically closed field. Then for any ideal $I \subseteq k[x_1, x_2, ..., x_n]$,

$$\mathcal{I}(\mathcal{Z}(I)) = \sqrt{I} = \text{rad}(I)$$

Furthermore, the maps $\mathcal{I}$ and $\mathcal{Z}$ are order-reversing bijections:

$$\{\text{Affine Sets}\} = \{\text{Closed Sets}\} \underset{\mathcal{Z}}{\overset{\mathcal{I}}{\rightleftarrows}} \{\text{Radical Ideals}\}$$

Proof **:**
First, we've already shown that $\text{rad}(I) \subseteq \mathcal{I}(\mathcal{Z}(I))$ (proposition 21.2.12), so it remains to show that $\text{rad}(I) \supseteq \mathcal{I}(\mathcal{Z}(I))$. By Hilbert's basis theorem, we have that $I = (f_1, f_2, ..., f_n)$. Let $g \in \mathcal{I}(\mathcal{Z}(I))$. If we show that $g^N \in I$ for some $N \in \mathbb{N}_{>0}$, this would imply $g \in \text{rad}(I)$. To that end, we employ the following trick known as the Rabinowitsch trick: introduce a new variable x_{n+1} and consider the ideal:

$$(f_1, ..., f_n, x_{n+1}g - 1) = I' \subseteq k[x_1, ..., x_n, x_{n+1}]$$

By definition of g, at any point where each $f_1, f_2, ..., f_n$ vanish, g also vanishes (since $g \in \mathcal{I}(\mathcal{Z}(I))$). But then, $x_{n+1}g - 1$ is nonzero! Thus, $\mathcal{Z}(I') = \emptyset$ in $\mathbb{A}^{n+1}$, and by Hilbert's weak Nullstellensatz, I' cannot be a proper ideal, and so $1 \in I'$. Since $1 \in I'$, we can write it as:

$$1 = a_1 f_1 + \cdots + a_n f_n + a_{n+1}(x_{n+1}g - 1) \qquad a_i \in k[x_1, ..., x_{n+1}]$$

For the following steps, it is insightful to expand the equation:

$$1 = \sum_i^n a_i(x_1, ..., x_{n+1}) f_1(x_1, ..., x_n) + a_{n+1}(x_1, ..., x_{n+1})(x_{n+1}g(x_1, ..., x_n) - 1)$$

Now, since x_{n+1} is a free variable, this equation still holds true if we substitue x_{n+1} for another expression (it will simply limit the equation to solution which also satisfy the substitued expression).

We wil choose $x_{n+1} = 1/g(x_1, ..., x_n)$ so we get the following expression in $K(x_1, ..., x_n)$:

$$1 = \sum_i^n a_i \left(x_1, ..., x_n, \frac{1}{g(x_1, ...x_n)} \right) f_i(x_1, ..., x_n)$$

Importantly for us, the only terms in the denominator would be some powers of $g(x_1, ..., x_n)$. Thus, for suitably large $N \in \mathbb{N}$, we may re-write the right hand side as:

$$1 = \sum_i^n \frac{b_i\,(x_1, ..., x_n)\, f_i(x_1, ..., x_n)}{g(x_1, ..., x_n)^N}$$

given appropriately substitued $h_1, ..., h_n \in k[x_1, ..., x_n]$. Multiplying both sides by the denominator, we get that:

$$g(x_1, ..., x_n)^N = \sum_i^n b_i(x_1, ..., x_n) f_i(x_1, ..., x_n)$$

Notice the right hand side is an element in the ideal I, and os $g^N \in I$! Thus, by definition, $g \in \mathrm{rad}(I)$. Since g was arbitrary, we see that $\mathcal{I}(\mathcal{Z}(I)) \subseteq \mathrm{rad}(I)$, and so we get $\mathcal{I}(\mathcal{Z}(I)) = \mathrm{rad}(I)$.

Furthermore, we want to show the given inverse bijections: If I is a radical ideal, then we've just shown that

$$I = \mathcal{I}(\mathcal{Z}(I))$$

conversely, if $A \subseteq k^n$ is an algebraic set, consider $\mathcal{I}(A)$. We want to check that $\mathcal{I}(A)$ is radical, so let's say $f^n \in \mathcal{I}(A)$, so $f^n(a)$ vanishes on all points of A. But then $f(a)$ vanishes at all points (since the field k has no nilpotent elements). Thus $f \in \mathcal{I}(A)$, showing that all algebraic sets map to radical ideals, as we sought to show.

Corollary 21.2.27: Nullstellensatz in Algebraic Closure

Let k be any field with algebraic closure $\overline{k}$, and $I \subseteq k[x_1, ..., x_n]$ an ideal. Then:

$$\mathcal{I}_k(\mathcal{Z}_{\overline{k}}(I)) = \sqrt{I}$$

As a particular result, $I = (1)$ if and only if there is no common zeros in $\overline{k}^n$ of the polynomials in I

Proof **:**
exercise.

More on Algebra-Geometry Dictionary

With these results established, let us recap all the algebraic/geometric correspondences and prove some more. There are some immediate consequence of these result. An immediate easy result is that it can finally be concluded that:

$$\mathcal{I}(A \cap B) = \sqrt{\mathcal{I}(A) + \mathcal{I}(B)}$$

Next, we now have made precise the relation between radical ideals and prime ideals, as we have fully geometrically classified radical and prime ideals. Next, we get the following dictionary between the language of geometry and the language of algebra:

Geometry	Algebra
Affine sets V	coordinate rings $k[V]$
points of V	maximal ideals of $k[V]$
closed subsets of V	radical ideals of $k[V]$
sub-varieties of V	prime ideals of $k[V]$
regular function: $\varphi : V \to W$	k-algebra homomorphism $\tilde{\varphi} : k[W] \to k[V]$

We shall keep expanding this dictionary throughout this part of the book (i.e. part V). As a teaser, we shall show that all PID's correspond to smooth curves (see section section 22.7.

Now that we have established this definition between algebraic sets and it's algebraic counter-part, we can expand some more on the Zariski Topology:

Proposition 21.2.28: Zariski Topology is Compact

Let $\mathbb{A}_k^n$ have the Zariski topology. Then $\mathbb{A}_k^n$ is compact

***Proof* :**
Let $\mathcal{O}$ index an open cover for $\mathbb{A}_k^n$

$$\mathbb{A}_k^n = \bigcup_{\alpha \in \mathcal{O}} U_\alpha$$

By proposition 21.1.12, any open set is the union of sets of the form $D(f)$, so we may re-write this as:

$$\mathbb{A}_k^n = \bigcup_{\alpha \in \mathcal{O}'} D(f_\alpha)$$

appropriately re-indexing. Then taking the compliment, we get that it's equivalent that:

$$\emptyset = \bigcap_{\alpha \in \mathcal{O}'} \mathcal{Z}(f_\alpha) = \mathcal{Z}\left(\sum_{\alpha \in \mathcal{O}'} (f_\alpha)\right)$$

By the Nullstellensatz, we get that it's equivalent that

$$1 \in \sum_{\alpha \in \mathcal{O}'} (f_\alpha)$$

Thus, for appropriate coefficients $g_i \in k$, we get that it's equivalent that:

$$1 = g_1 f_1 + \cdots + g_n f_n \in (f_1, ..., f_n) = \sum_{i=1}^{n} (f_i)$$

But then, we get that

$$\emptyset = \bigcap_{i=1}^{n} \mathcal{Z}(f_i) \quad \Leftrightarrow \quad \mathbb{A}_k^n = \bigcup_{i=1}^{n} D(f_i)$$

Recall that for any algebraic subset $V \subseteq \mathbb{A}_k^n$, the corresponding coordinate ring $k[V]$ is used to define the subspace topology. Using the Nullstellensatz, we may deduce that it cannot have nilpotent elements:

Corollary 21.2.29: Coordinate Ring Always Reduced

Let $V \subseteq \mathbb{A}_k^n$. be an algebraic subset and $k[V]$ the assocaited coordinate ring. Then $k[V]$ is has no nilpotent elements

Proof **:**
Since $k[V] = k[x_1, x_2, ..., x_n]/I(V)$, and $I(V)$ is a radical ideal by the Nullstellensatz, the quotient cannot contain any nilpotent elements, as we sought to show.

Corollary 21.2.30: Nullstellensatz For Subspaces

Let $V \subseteq \mathbb{A}_k^n$ be an algebraic subset (i.e. a closed subset) with cooresponding coordinate ring $k[V]$. Then we can define $\mathcal{Z}$ and $\mathcal{I}$ as:

$$\left\{\begin{matrix}\text{affine algebraic subsets} \\ \text{of } V\end{matrix}\right\} \underset{\mathcal{Z}}{\overset{\mathcal{I}}{\rightleftarrows}} \left\{\begin{matrix}\text{Radical Ideals} \\ \text{of } k[V]\end{matrix}\right\}$$

Proof **:**
This follows immediately from the correspondence theorem and the Nullstellensatz

Next, sticking to the case where $k = \overline{k}$ is algebraically closed, from this section we can deduce that any coordinate ring is finitely generated, nilpotent-free (i.e. reduced) and is a k-algebra over an algebraically closed field. We can show that all such rings appear as coordinate rings for an affine set.

We first have the following:

Definition 21.2.31: Reduced Ring

Let R be a ring. Then R is said to be *reduced* if $x^n = 0$ implies $x = 0$

(finish here)

Proposition 21.2.32: Coordinate Ring Correspondence

Let R be a ring. Then if R is:

1. Finitely generated
2. Reduced
3. over an algebraically closed field

then R is isomorphic to a coordinate ring for an appropriate dimension n of $\mathbb{A}_{\overline{k}}^n$ and $\mathcal{Z}(S) \subseteq \mathbb{A}_{\overline{k}}^n$.

Proof **:**
exercise. I think I should prob. write it down actually

Notice that through this result, we can define an affine set without the reference of an ambient space, similar to how we may define a manifold without the need an ambient embedding space. This idea shall be greatly generalized in section 21.7 and algebraic geometry more generally (see cite:HERE).

Exercise 21.2.3

1. Perhaps some on the Nullstellensatz over non-algebraically closedfields (In particular, I think if we take $\mathcal{Z}$ over the algebraic closure of k, then we recover the Nullstellensatz)

21.3 Rational Maps: Classifying Varieties

In this section, we finally take advantage of the fact that we may localize $k[X]$ to $k(X)$. If $k[X]$ is an integral domain, then $k(X) \cong k(X_1, ..., X_n)$. An important consequence of working with varieties is that any open subset must be dense (the closure has to be the entire variety). Hence, it suffices to define rational morphisms on open subsets, which we shall see the consequences of in this section. Note how this different from manifold theory, but similar to complex analysis.

Definition 21.3.1: Rational Functions

Let X be an irreducible affine variety. Then the fraction field of $k[X]$ is called the field of *rational function on* X, and is denoted $k(X)$. A function in $k(X)$ is called a *rational function*.

If $\varphi \in k(X)$, then $\varphi = \frac{f}{g}$ where $g \notin I(X)$ (or else the denominator would be the zero function). When dealing with polynomials, we shall say that two rational polynomial f/g and f'/g' represent the same rational function if and only if $fg' - f'g \in I(X)$.

Another way of constructing $k(X)$ is the following: take the collection

$$\mathcal{O}_X = \{\varphi = f/g \in k(x_1, ..., x_n) \mid g \notin I(X)\}$$

then take the subset $\mathfrak{m}_X = \{\varphi = f/g \in \mathcal{O}_X \mid f \in I(X)\}$. Show that:

$$k(X) \cong \mathcal{O}_X/\mathfrak{m}_X$$

This shall become the dominant way of interpreting the set of rational functions when working with quasi-projective varieties (see 21.4.3).

A key difference between regular and rational functions is that a rational function need not be defined at every point. In fact, it *has* to be the case that a rational function is not defined on a finite set of points for it to not be equivalent to a regular function. To see this, consider the following definition:

Definition 21.3.2: Regular Point

Let $\varphi \in k(X)$ be a rational function. Then a *regular point* at $x \in X$ is a point where we can represent φ as f/g, $\varphi(x) = f(x)/g(x)$, and $g(x) \neq 0$. In this case, the element $f(x)/g(x)$ in k is called the *value* of φ at x and is denoted $\varphi(x)$

For example, take $\varphi = 1/x$ on $\mathbb{A}^1_k$. Then all points except 0 are regular points. If $\varphi = f/g$ is regular at a point x, then it is regular in an open neighborhood of X, namely the neighborhood $D(g) \cap X$. Some authors put that a rational function is regular at a point if it is regular in an open neighborhood.

Proposition 21.3.3: Regular at All Points, then Regular

Let $\varphi \in k(X)$ be a rational function that is regular at all points. Then φ is a regular function

***Proof* :**
Let $\varphi \in k(X)$ be a rational function that is regular at every point. Then for each $x \in X$, we can represent φ as f_x/g_x. Take $\mathfrak{a} = (g_x)_{x \in X}$. Then as $k[X]$ is Noetherian, this is a finitely generated ideal:

$$\mathfrak{a} = (g_{x_1}, ..., g_{x_n})$$

Note that the g_{x_i}'s cannot have a common zero $x \in X$, or else all the functions in the ideal would vanish at x, however we know that $g_x(x) \neq 0$. Then this gives an equivalent result to the Nullstellensatz (see exercise ref:HERE), namely that:

$$\mathfrak{a} = (1)$$

In particular, there exists $u_i \in k[X]$ such that

$$\sum_i^n u_i g_i = 1$$

Now, by what we've covered $\varphi = f_i/g_i$ on open neighborhoods U_i that cover X. Multiply both sides of the equation by φ, we get:

$$\varphi = \sum_i^n u_i f_i$$

showing it is a regular function, completing the proof.

The Nullstellensatz is essential for the proof, i.e. that k is an algebraically closed field, or else this result is false!

Example 21.3.1: Rational Function That is Regular
the function

$$\frac{x}{x+100}$$

is regular on $\mathcal{Z}(x^2 + y^2 - 1) \subseteq \mathbb{A}^2_{\mathbb{R}}$ even though it is not a polynomial: there does not exist a point (x, y) satisfying $x^2 + y^2 - 1$ such that $x + 100 = 0$ namely since all points x where $x^2 + y^2 - 1 = 0$ must be within $[-1, 1]$[a]

[a] This failure is *not* present when working with the scheme $\operatorname{Spec} \mathbb{R}[x, y]/(x^2 + y^2 - 1)$ as $x + 100$ has a zero at the prime ideal $(x - 100, y^2 + 9999)$, and so this "anomaly" is present if we don't work in the right context

Lemma 21.3.4: Domain of Definition

Let $\varphi \in k(X)$ be a rational function. Then set of regular points is non-empty, open and dense.

Proof **:**
exercise.

Definition 21.3.5: Domain of Definition

Let $\varphi \in k(X)$ a rational function. Then the set of regular point is called the *domain of definition.*

Lemma 21.3.6: Collection of Rational Points

Let $\varphi_1, ..., \varphi_n \in K(X)$ be rational functions. Then their common domain of definition is open and non-empty

Proof **:**
It is certainly open as it is the finite union open sets, and it is non-empty for X is irreducible (see exercise ref:HERE)

Lemma 21.3.7: Suffice To Define On Open Set

Let $\varphi \in K(X)$ be a rational function. Then if φ is specified on some non-empty open subset $U \subseteq X$, then there is a unique extension of φ to all of X.

Proof **:**
Take $\varphi \in K(X)$. If $\varphi = 0$, there is nothing to prove, so assume $\varphi \neq 0$. Say we specify what φ is on an open non-empty subset $U \subseteq X$, call this function $\varphi|_U$. Then let's say that φ' is a rational function where φ' on U is equal to $\varphi|_U$ (where a-priori it is unknown where there is a unique extension, and hence it may be another function). Consider $\varphi - \varphi' = \varphi''$. On U, this function is identically 0. Take the representative $\varphi'' = f/g$ on U so that $f = 0$. Then take $X_1 \subseteq X$ given by $f = 0$. This is not all of X as $\varphi \neq 0$, however we then have $X_1 \cup (X \setminus U) = X$ where both are proper subsets, contradicting irreducibility of X, completing the proof.

Definition 21.3.8: Rational Mapping

Let X, Y be varieties. Then a *rational mapping* $\varphi : X \dashrightarrow Y$ is a map given by

$$\varphi = (\varphi_1, ..., \varphi_n)$$

where each φ is is a (regular or rational) function and for each $x \in X$, $(\varphi_1(x), ..., \varphi_n(x)) \in Y$ where all of the functions φ_i are regular.

The image of the rational map shall be denoted $\varphi(X)$, even though not every point may be regular. Note that a rational map shall necessarily map an open subset $U \subseteq X$ into Y, and hence it is well-defined by lemma 21.3.7. by lemma 21.3.7, an equivalent definition of a rational map is that it is the collection of equivalence relations $[(\varphi, U)]$ with $U \subseteq X$ open and $\varphi : U \to Y$ where $[(\varphi, U)] \sim [(\psi \sim V)]$ if and only if

$$\varphi|_{U \cap V} = \psi|_{U \cap V}$$

This equivalent to definition 21.3.8, and is more prominent in the context of algebraic geometry. Its upside is that each representative is a well-defined map and hence can be seen as the appropriate quotient of some hom-sets, while its downside is that a rational mapping is now an equivalence relationship.

Definition 21.3.9: Birational Map

Let $\varphi : X \dashrightarrow Y$ be a rational map. Then φ is said to be *biraitonal* if the image of φ is dense and there exists a rational map $\psi : Y \dashrightarrow X$ with dense image such that $\varphi \circ \varphi = \psi \circ \varphi = 1$.

Proposition 21.3.10: Correspondence Rational Mapping and Rational Polynomials

Let $\varphi : X \dashrightarrow Y$ be a rational mapping. Then $\varphi^* : k(Y) \to k(X)$ is a ring isomorphism where φ^* determines an isomorphic embedding of $k(Y)$ into $k(X)$. Furthermore, if $\varphi : X \dashrightarrow Y$ and $\psi : Y \dashrightarrow Z$ are rational mapping with dense images, then $\psi \circ \varphi$ has dense image and $(\psi \circ \varphi)^* = \psi^* \circ \varphi^*$.

Furthermore, φ is birational if and only if φ^* is an isomorphism.

Proof **:**
exercise

Definition 21.3.11: Rational Closed Set

Let V be a variety. Then if V is birational to $\mathbb{A}^n_k$ or $\mathbb{P}^n_k$, then V is called *affine rational* or *projective rational* respectively. If V has dimension $1, 2$ or codimension 1, then it is a *rational curve*, *rational surface*, or *rational hypersurface* respectively.

Example 21.3.2: Rational Mapping and Birational Maps

1. Naturally, all curves of order 1 (i.e., straight lines) are tautologically rational. More generally any linear subspace is rational.
2. Consider the curve given by
$$y^2 = x^2 + x^3$$
which looks like:

Label this curve C. Then this is a rational curve. To see this, consider that the line given by $y = tx$ passing through the origin intersects the curve at a single point (not counting the origin). Then substituting $y = tx$ into our equation, we get:

$$x^2(t^2 - x - 1) = 0$$

The root $x = 0$ corresponds to the origin, and the other roots $x = t^2 - 1$ corresponds to non-trivial solution. Hence, we get $x = t^2 - 1, y = t(t^2 - 1)$, and the rational map:

$$\varphi(t) = (t^2 - 1, t(t^2 - 1))$$

For the inverse, start with any line for example the line $x = 1$. For any $(x_0, y_0) = p \in C$, make a line passing through the origin O and the point O, this is the line PO. Then this line intersects Q. Then with an argument similar to the above one, we can see that there is a rational map $\psi : C \to \mathbb{A}^1_k$, which furthermore is the inverse of the above map, giving birationality. Note that this map is certainly not an isomorphism or a homeomorphism.

If we take $k = \mathbb{Q}$, then this gives all rational solutions to the equation $y^2 = x^3 + x^2$. The case is a little more subtle if we were looking for integer solutions to the equation, which is explored in chapter 22.

3. Let us show that all irreducible curves C of order 2 are rational (note that the above example is an order 3 curve, the maximal degree being 3). Let C be given by the equation $f(x, y)$. Take any point $(x_0, y_0) \in C$, and equation of the line through (x_0, y_0) with slope t

$$y - y_0 = t(x - x_0)$$

Label this line L_t. Let us find the points of intersection of the curve C with line L_t. To do so, we substitute:

$$f(x, y_0 + t(x - x_0)) = 0$$

which is equivalent to the mapping $\mathbb{A}^1_k$ to the curve. This mapping is well-defined (notice it is still of degree 2). Then one of the roots is certainly x_0. Then let a denote the coefficient

of x after dividing the equation $f(x, y_0 + t(x - x_0))$ by the coefficient of x^2. Then we can see that the remaining roots must be:

$$x + x_0 = -a \qquad \Rightarrow \qquad x = -x_0 - a$$

Then x is a rational function of t, and substituting the expression for x into L_t, we often an expression for y as a rational function of t.

Note that if the coordinates x_0, y_0 of $(x_0, y_0) \in C$ and the coefficients of $f(x, y)$ belong to some subfield $k_0 \subseteq k$, then so do the coefficients of the function that gives the parameterization. In this case, we can find the general form of the solution in rational numbers of a Diophantine equation of degree 2 by knowing *at least* one solution! However, it turns out that knowing there is at least one solution can sometimes be tricky[a]

4. More generally, an irreducible hypersurface $X \subseteq \mathbb{A}^n_k$ given by a polynomial of degree 2 $f(x_1, ..., x_n) = 0$ is rational. The proof generalizes the above and is left as an exercise. The geometric picture is projection the hypersurface to a hyper-plane via a projection that does not pass through some point x. The point x can be any point that is not a "vertex" on X, namely that for at least one variable:

$$\frac{\partial f}{\partial x_i}(x) \neq 0$$

5. The curve order 3 given by $x^3 + y^3 = 1$ is *not* rational. More generally, $x^n + y^n = 1$ is not rational for $n \geq 3$ (if $n \nmid \operatorname{char} k$). To see this, let C_n be the curve given by $x^n + y^n = 1$ and suppose it is rational so that there exists rational functions φ, ψ such that $x = \varphi(t), y = \psi(t)$. Then we may find polynomials p, q, r relatively prime such that

$$\varphi(t) = \frac{p(t)}{r(t)} \qquad \psi(t) = \frac{q(t)}{r(t)}$$

where the denominator is the same for simplicity (the same argument follows with more steps by taking p/r and q/s). Then plugging into our equation we get:

$$p(t)^n + q(t)^n - r(t)^n = 0$$

for all $t \in k$. Then differentiating, we get a non-zero solution (as $n \nmid \operatorname{char} k$) and so we have:

$$p(t)^{n-1}p'(t) + q(t)^{n-1}q'(t) - r(t)^{n-1}r'(t) = 0$$

We now have two systems equations. If we assume that $p^{n-1}, q^{n-1}, -r^{n-1}$ are variable, we can consider the above two equations as being a system of linear equations which has has the matrix:

$$\begin{pmatrix} p & q & r \\ p' & q' & r' \end{pmatrix}$$

Solving this system of linear equations, we get that $p^{n-1}, q^{n-1}, -r^{n-1}$ are proportional too

$$qr' - rq' \qquad rp' - pr' \qquad pq' - qp'$$

As p, q, r are relatively prime, we get that:

$$p^{n-1} \big| (qr' - rq') \qquad q^{n-1} \big| (rp' - pr') \qquad r^{n-1} \big| (pq' - qp')$$

Finally, if a, b, c are the degrees of p, q, r . If $a \geq b \geq c$, then

$$(n-1)a \leq b + c - 1$$

which can be show to be a as $n \geq 3$. A similar result occurs for all other cases, leading to a contradiction.

[a] see Legendre theorem in cite:HERE

Birational mappings are weaker than isomorphisms (of varieties). This problem can be translated algebraically to the difference between having $k[X]$ and $k[Y]$ be isomorphic and $k(X)$ and $k(Y)$ be isomorphic. The reader may recall from section 21.2.5 that $k[X]$ and $k[Y]$ must be finitely generated reduced k-algebras, $k(X)$ and $k(Y)$ finite field extensions. As the structure of field extensions is much simpler, we expect birational mappings to be much simpler. In fact, we may even classify all affine variety up to birational isomorphism:

Theorem 21.3.12: Classification of Varieties Up To Birational Map

Every variety X is birationally isomorphic to a hypersurface of some affine space $\mathbb{A}^n_k$.

Thus, birationality of varieties reduces to the classification of hypersurfaces. Note that being birationally isomorphic to a hypersurface does not imply a curve or plane is rational. For example, the curve given by $x^3 + y^3 = 1$ is already a hypersurface in $\mathbb{A}^2$ as it is given by $\mathcal{Z}(x^3 + y^3 - 1)$, however it is not a rational curve as it is not birational to $\mathbb{A}^1_k$.

Proof **:**
Take the field $k(X) \cong k[x_1, ..., x_n]/I(X)$ which is a finitely generated k-algebra, namely it can be generated by the functions $t_1, ..., t_n$. Take the collection of all algebraically independent subsets of $t_1, ..., t_n$ and (as $k(X)$ is Noetherian) choose a maximal subset, index it as $t_1, ..., t_d$. Then every $y \in k(X)$ must have the be algebraically dependent on $t_1, ..., t_d$ for the set $t_1, ..., t_d, y$ algebraically dependent, that is is that there exists a function f such that

$$f(t_1, ..., t_d, y) = 0$$

and furthermore, the polynomial representative $f(x_1, ..., x_d, x_{d+1})$ can be chosen to be irreducible. So choose a polynomial $f(x_1, ..., x_d, x_{d+1})$ such that $f(t_1, ..., t_d, t_{d+1}) = 0$. Then we shall show that $k(X)$ is isomorphic to $k(Y)$ where $k(Y)$ is given by the hypersurface defined by f.

Let us start by showing that adjoining t_{d+1} to $k(t_1, ..., t_d)$ gives a separable extension. To do this, let us first show that $\frac{\partial f}{\partial x_i}(x_1, ..., x_{d+1}) \neq 0$ for at least on index i. If k has characteristic 0, it is clear as f is irreducible with degree higher than 0. If k has characteristic p, then for the sake of contradiction assume that it were not the case. Then each x_i would occur in f with degrees that must be a multiple of the characteristic p of k. Hence we can write f as:

$$f = \sum a_{i_1,...,i_{d+1}} x_1^{pi_1} x_2^{pi_2} \cdots x_{d+1}^{pi_{d+1}}$$

Setting $b_{i_1,...,i_{d+1}}$ so that $b^p_{i_1,...,i_{d+1}} = b_{i_1,...,i_{d+1}}$, we can define a g such that $f = g^p$, contradicting the irreducibility of f.

Now, as $\frac{\partial f}{\partial x_i} \neq 0$, the polynomial f must have monomials with x_i occurring. Then it must be that $t_1, ..., t_{i-1}, t_{i+1}, ..., d_{d+1}$ are algebraically independent over k. First, t_i is algebraic over

$k(t_1, ..., t_{i-1}, t_{i+1}, ..., t_{d+1})$ as f has nonzero terms with x_i (and hence correspondingly t_i). Thus if the elements $t_1, ..., t_{i-1}, t_{i+1}, ..., t_{d+1}$ were dependent, the transcendence degree of $k(t_1, ..., t_{d+1})$ would be less than d, contradicting the algebraic independence of $t_1, ...t_d$.

Thus, renumber $x_1, ..., x_n$ so that $t_1, ..., t_n$ are independent over k and $\frac{\partial f}{\partial x_{d+1}} \neq 0$. Then t_{d+1} is separable over $k(t_1, ..., t_d)$. This gives the right circumstances for the primitive root theorem: as t_{d+2} is algebraic over $k(t_1, ..., t_d)$, we have:

$$k(t_1, ..., t_d, t_{d+1}, t_{d+2}) = k(t_1, ..., t_d)(y)$$

Now, repeating this process, we may adjoin the elements $t_{d+1}, ..., t_n$ and replacing y with the appropriate element, so that we can represent $k(X)$ as $k(z_1, ..., z_{d+1})$ where $z_1, ..., z_d$ are algebraically independent over k and:

$$f(z_1, ..., z_d, z_{d+1}) = 0$$

The polynomial f is irreducible over k ,and $\frac{\partial f}{\partial x_{d+1}} \neq 0$. Let Y be the closed set defined by the above f. Then $k(Y)$ is isomorphic to $k(X)$ by construction, which by **??** means that X and Y are birationally isomorphic, as we sought to show.

Note that $k(X)/k(z_1, ..., z_d)$ is a finite separable extension, and that the primitive elements $z_1, ..., z_{d+1}$ can be chosen as linear combinations of the $x_1, ..., x_n$ for appropriate c_{ij}:

$$z_i = \sum_j^n c_{ij} x_j$$

Then the mapping $(x_1, ..., x_n) \to (z_1, ..., z_{d+1})$ is a projection from $\mathbb{A}_k^n$ that is parallel to the linear subspace given by:

$$\sum_j^n c_{ij} x_j = 0 \qquad 1 \leq i \leq d+1$$

Thus, the birational mapping theorem says that there is an appropriate hypersurface (not necessarily a hyperplane) on which we may project in a rational manner the variety! From this perspective, linear algebra results which are birationally invariant can be translated over to varieties, which we shall soon see gives us some interesting results!

Exercise 21.3.1

1. Complete the proof of proposition 21.3.3
2. Let X be irreducible. Show that any finite collection of rational functions on X has a non-empty common domain of definition

21.4 Projective Variety

So far, all of the algebraic sets were called *affine* algebraic sets, as they were all a subset of $\mathbb{A}_{\mathbb{C}}^n$. These offer great geometrical intuition, however they are lacking in certain algebraic qualities, namely they don't have "enough" functions that are defined on them. To see this, it is simplest to resort to an analogy from calculus. The reader may recall that the function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = 1/x$ is

famously undefined at $x = 0$, as we cannot divide by 0. Graphing this function, we that this function is certainly not well-defined at 0, and can qualitatively be thought of as being far away from being continuous:

However, we can enlarge $\mathbb{R}$ to make this function well-defined by adding a *point at infinity*. Namely, we may take $\mathbb{R}_\infty = \mathbb{R} \cup \{\infty\}$ where we extend the arithmetics of $\mathbb{R}$ to $\mathbb{R}_\infty$ by doing:

$$a + \infty = \infty + a = \infty \qquad b \cdot \infty = \infty \cdot b = \infty$$

where $a \in \mathbb{R}$ and $0 \neq b \in \mathbb{R}_\infty$ We then say that if (x_n) diverges, then

$$\lim_{n \to \infty} x_n = \infty$$

In this case, we know have that the function $1/x$ has a well-defined value of ∞ at 0, and even better under the right frame-work (provided in differential geometry), the function $f(x) = 1/x$ can be shown to be smooth on $\mathbb{R}_\infty$. In this way, we have insured that we cannot have an undefined rational polynomials on $\mathbb{R}_\infty$! The space $\mathbb{R}_\infty$ is called the *projective line*. This "closure" on the set of definable rational polynomials makes the projective line an algebraically nicer space to work with, in a similar way that $\mathbb{C}$ is the better space to work in as all polynomials have roots.

Why is the space called projective? This is understood from the following construction of projective space that extends to higher dimensions. Let us start with one dimensions. We may define $\mathbb{R}_\infty$ as the

$$\mathbb{R}\mathrm{P}^1 := \frac{\mathbb{R}^2 \setminus \{(0,0)\}}{\{(x,y) \sim (\lambda x, \lambda y), \forall \lambda \in \mathbb{R}_{\lambda \neq 0}\}}$$

by taking the representative of each equivalence class to have norm 1, notice that we may associate all the points of $\mathbb{R}\mathrm{P}^1$ to S^1, and we can associate points of $\mathbb{R}$ with the equivalence class where $y \neq 0$.

The equivalence class $[(\lambda x, 0)]$ can be thought of as the "point at infinity" that we've added, hence we may think of $\mathbb{R}\mathrm{P}^1$ as

$$\mathbb{R}\mathrm{P}^1 = \mathbb{R}^1 \sqcup \mathbb{R}^0 \tag{21.5}$$

The term "projective" come from the idea that parallel lines will meet at the "point at infinity". In terms of "closure" that we talked about earlier, the reader may recall that in classical geometry, two non-parallel lines will always have a solution. In $\mathbb{R}\mathrm{P}^1$, we have added the extra property that there is one point at which parallel lines meet; they also have a solution. In terms of linear algebra, every system of linear equations has a solution.

We may extend this further, and take $\mathbb{R}^2$ and add "points at infinity". In this case, there are many possible pairs of parallel lines. We may consider that up to scaling, we have many new points that we may add. Formally, let:

$$\mathbb{R}\mathrm{P}^2 := \frac{\mathbb{R}^3 \setminus \{(0,0,0)\}}{\{(x,y,z) \sim (\lambda x, \lambda y, \lambda z), \forall \lambda \in \mathbb{R}_{\lambda \neq 0}\}}$$

We may imagine that this insures that any rational functions in two variables always has solution. Adding multiple points at infinity insures that the function $1/y$, $1/x$, $1/xy$ and so forth are all considered distinct when evaluating ($1/y$ is zero when evaluated at one infinity but not another). In terms of parallel lines, we may think that parallel lines in any angle will have a unique solution (note that the distance between lines not matter, they still "meet" at the same point at infinity). The shape $\mathbb{R}\mathrm{P}^2$ can be thought of as $\mathbb{R}^2$ with a circle of points at infinity added where opposites are glued. Can the reader use this intuition to generalize equation (21.5) to:

$$\mathbb{R}\mathrm{P}^2 = \mathbb{R}^2 \sqcup \mathbb{R}^1 \sqcup \mathbb{R}^0$$

The reader familiar with differential geometry will know that $\mathbb{R}\mathrm{P}^2$ is a two dimensional smooth manifold, and is one of the simplest example of a shape that is 2-dimensional but cannot be embedded in 3-dimensions without self-intersection. The following image from Wikipedia may help in trying to visualize (note that this is an *immersion*, meaning it is an approximation of the shape by embedding it in a space in which there is not enough dimensions to visualize it without intersection[15])

Figure 21.2: Wikipedia: Immersion of $\mathbb{R}\mathrm{P}^2$

We may continue this process and define $\mathbb{R}\mathrm{P}^n$, or more generally $\mathbb{P}^n_k$ for some field k. These shall have a larger class of functions that we may define on them (we shall eventually say that there are an "ample" amount of function[16]). Thus, we shall treat $\mathbb{P}^n_k$ as an extension of $\mathbb{A}^n_k$, and consider the equivalence classes of $\mathbb{P}^n_k$ as *points*.

[15] See CITE:DIFF-GEO

[16] More technically speaking, we shall say that $\mathbb{P}^n_k$ has *ample line bundles*, a technicality that shall be elucidated in CITE:ALG-GEO

An important technicality that we may encounter is that we must convert a function that is defined on $\mathbb{A}_k^n$ to a function defined on $\mathbb{P}_k^n$, namely the points of $\mathbb{P}_k^n$ are now equivalence relations and hence a function whose domain is $\mathbb{P}_k^n$ must be independent of representative. This shall form an important class of rings that are "'good representatives" to extend polynomials to $\mathbb{P}_k^n$. As we are extending results from $\mathbb{A}_k^n$, this class of rings shall have a similar dictionary as that provided between affine varieties and radical ideals of coordinate rings, namely we shall show that polynomials solutions over $\mathbb{P}_k^n$ shall correspond to *graded ideals* of *graded rings*.

21.4.1 Definitions and Properties

Let us first formally define projective space

Definition 21.4.1: Projective Space

Let k be a field. Then *projective space* over k of dimension n is:

$$\mathbb{P}_k^n = \frac{k^{n+1} \setminus \{0\}}{(x \sim \lambda x,\ \forall x \in k^{n+1})}$$

The points of $\mathbb{P}_k^n$ are often denoted $[x_0 : x_1 : \cdots : x_n]$.

Regular polynomials function need not be well-defined on projective space, for example $f(x, y) = x^2+y$ does not map two elements in the same equivalence class to the same point. We must limit to polynomials that maps the representatives of an equivalence class to the same point. These are precisely the homogeneous polynomials:

Definition 21.4.2: Homogeneous Polynomials (Algebraic Form)

Let $f \in k[x_1, ..., x_n]$ be a polynomial. Then f is said to be a *homogeneous polynomial* of degree d or n *algebraic d-form* if

$$f(\lambda x_1, ..., \lambda x_n) = \lambda^d f(x_1, ..., x_n)$$

The term *algebraic form* is a bit out-dated, but is often still used for the degree two case, where such algebraic forms are called *quadratic forms*. This definition can be defined on more general class of functions. The reader comfortable with some analysis will see in exercise 21.4.1 that any (at least) d-times differential degree d homogeneous function is in fact a homogeneous polynomial[17].

Example 21.4.1: Homogeneous Polynomials

All homogeneous polynomials $f \in k[x_1, ..., x_n]$ are polynomials for which all monomials have the same degree (exercise). For example:

1. $f(x, y, z) = x^3 + xy^2 - 2y^3$ is homogeneous
2. $f(x, y) = x^2 + y^3$ is not homogeneous

If $(a_1, ..., a_n) \in k^n$ is a solution to a homogeneous polynomials, then all points on the line passing

[17] homogeneity does not imply differentiability or continuity, can you think of a counter-example?

through this point and the origin is also a solution, hence homogeneous polynomials are the "natural" for projective space, as projective space is an equivalence relation on lines.

It is tempting conclude that if f is homogeneous, then $f : \mathbb{P}^n_k \to \mathbb{P}^1_k$ is well-defined, however these function are not well-defined as $\mathbb{P}^1_k$ does not contain a point $[0 : \cdots : 0]$ while all homogeneous polynomials (over an algebraically closed field) will map an element to 0 (by the Nullstellensatz). If we ignore the zero's of f, we still get some odd behaviors, namely if $f(a_1, ..., a_n) \neq 0$, then:

$$[f(a_1, ..., a_n)] = \left[\frac{f(a_1, ..., a_n)}{f(a_1, ..., a_n)}\right] = [1]$$

This may unsettle the reader, as a homogeneous polynomials seems to act like a constant function rather than a proper polynomial on projective space. As it turns out, we can't extend (non-constant) polynomials from affine space to projective in a well-defined manner (see section 21.4.2) unless it is a constant function, which the above example is a first demonstration on why this is the case. Instead, we shall see that we can have non-trivial *rational* functions (omitting a finite number of points from the domain), and non-trivial function on open/closed subsets of $\mathbb{P}^n_k$.

As the zero-set of homogeneous polynomials in $\mathbb{P}^n_k$ must be a collection of lines in $\mathbb{A}^{n+1}_k$ before taking the quotient, we shall give a name to the zero-set in affine space:

Definition 21.4.3: Affine Cone

Let f be a homogeneous polynomial. Then it's zero-set in $\mathbb{A}^{n+1}$ is called an *affine cone*.

For example, the polynomial $x^2 + y^2 - z^2$ is the zero-set of in $\mathbb{A}^{2+1}_k$ looks like:

Figure 21.3: zero set of $x^2 + y^2 - z^2$

This equation in $\mathbb{P}^2$ would resemble a circle. This is no coincidence: as mentioned in the introduction of the section, projective space $\mathbb{P}^n$ is usually thought of as an extension of $\mathbb{A}^n$ by adding points at infinity in every straight direction and gluing all "opposite points by straight lines at infinity". The polynomial $x^2 + y^2 - z^2$ is very similar to $x^2 + y^2 - 1$. We can do this to any polynomial, any polynomial can be homogenized, and hence can be thought of as extending from $\mathbb{A}^n_k$ to $\mathbb{P}^n_k$:

Definition 21.4.4: Homogenization

Let $f \in k[x_1, ..., x_n]$ be a polynomial. Then the *homogenization* of f over degree d is the polynomial ${}^h f \in k[x_0, x_1, ..., x_n]$ satisfying:

$$ {}^h f(x_0, x_1, ..., x_n) = x_0^d f\left(\frac{x_1}{x_0}, ..., \frac{x_n}{x_0}\right) $$

Conversely, for a homogeneous polynomial $f(x_0, ..., x_n)$. We may *dehomogenize* it with respect to x_i by setting $x_i = 1$

$$ {}^d f(x_0, ..., x_n) = f(x_0, ..., x_{i-1}, 1, x_{i+1}, ..., x_n) $$

Homogenization should be thought of as the way we extend zero-sets from $\mathbb{A}^n$ into $\mathbb{P}^n$. Sometimes, these extensions of zero-sets add not extra points, while at other times they may add multiple:

Example 21.4.2: Homogenizing Polynomials

1. Take $x^2 + y^2 - 1$. Homogenizing (with degree 2) and labelling $x_0 = z$, we get $x^2 + y^2 - z^2$. If we homogenizing in degree n, we would get:

   $$ z^{n-2}x^2 + z^{n-2}y^2 = z^n $$

   Note that homogenizing at different degrees will not effect the shape. Homogenization in a degree greater than the "base degree" shall always have an extra solution, namely all points of the form $(x, y, 0)$, however these are "artificially added" points which appear when de-homogenizing in the non-z variables as an extra line of points, and so we shall only homogenize to insure the parity of all degree's of the monomials with the highest degree.

   In projective coordinates, the "classical" solutions of the form $x_0^2 + y_0^2 = 1$ become:

   $$ [x_0^2, y_0^2, 1] $$

   To see if there are new solutions, we may set the last coordinate to 0 (symbolizing going "off to infinity"), in which case we could ask for x_0, y_0 such that

   $$ [x_0^2 : y_0^2 : 0] $$

   is a solution. This is only possible if $x_0 = y_0 = 0$, but $[0 : 0 : 0] \notin \mathbb{P}^2$, hence there is no added points when homogenizing.

2. Take $xy - 1$. It's shape looks like:

Figure 21.4: zero-set of $xy = 1$

Homogenizing, we get $xy - z^2$. A solution $x_0y_0 - 1 = 0$ in projective coordinates is of the form:

$$[x_0 : y_0 : 1]$$

setting $z = 0$, we can ask if there are any points where:

$$[x_0 : y_0 : 0]$$

is a solution. And indeed, if $x_0 = 0$ or $y_0 = 0$, we get a solution, since

$$0y_0 - 0 = 0 \qquad x_0 0 - 0 = 0$$

We can see by visualizing the zero set that there seems to be two extra point added at infinite, one for the y-axis and one for the x-axis (remember that opposite points are considered the same).

If we de-homogenize at x, we get $y - z^2 = 0$, equivalently $y = z^2$, and de-homogenizing at y we get $x - z^2 = 0$, equivalently $x = z^2$. In both of these cases, we see that $x = z = 0$ and $y = z = 0$ is a solution. Hence, by de-homogenizing at each variable, we can find the vanishing set of a homogeneous polynomial over projective space. This can be seen more clearly when looking at the affine cone of the homogenized polynomial:

Figure 21.5: affine cone: $xy - z^2$

3. Take $x^3 + xy^2 - 2y^3 - 1$. It's shape looks like:

Figure 21.6: $x^3 + xy^2 - 2y^3 = 1$

When homogenizing, the affine cone gives:

Figure 21.7: affine cone of $x^3 = xy^2 - 2y^3 = z^3$

are there any added points, and if so what are they?

4. Take $f(x) = x^2$. Its vanishing set is the single point 0. Then ${}^h f(x_0, x) = x^2$ is the degree 2 homogenization. Then this has a single zero, namely $[1 : 0]$, which correspond to the single root of x^2 in $\mathbb{A}^1_k$

5. Take $f(x, y) = x^2 - y$ whose zero set is the graph of the usual quadratic. Then ${}^h f(x_0, x, y) = x^2 - x_0 y$. Then it has the roots:
$$[1 : y_0^2, y_0]$$
As well as:
$$[0 : 1 : 0]$$
in particular, we added a single point at infinity. This is the point where the parabola "meets". Notice that it wasn't that there was two points at infinite but just 1. The idea is that the "angle" with which the function approaches infinity tends towards $\pi/2$, and hence the edges of the graph shall meet at infinity. This shows that parabola's can be thought of as infinitely large ellipsoids[a].

   Can the reader see different sorts of connections between circles, parabolas, ellipses, and hyperbolas through projective space?

[a] for those comfortable with some Riemann geometry, can the reader see why projective space can be thought of as having positive curvature?

The usual polynomials ring $k[x_0, ..., x_n]$ has non homogeneous polynomials, however it has a natural grading which allows us to only focus on the homogeneous elements:

Lemma 21.4.5: Homogeneous Decomposition

Let $f \in k[x_1, ..., x_n]$ be a polynomial with maximal degree d. Then f can be decomposed into a unique sum of distinct degree homogeneous polynomials.

Furthermore, the collection $R_d \subseteq k[x_1, ..., x_n]$ of degree d polynomials forms a vector-space over k.

Proof **:**
exercise

Hence, $R = k[x_1, ..., x_n]$ can be written as a direct sum:

$$R = \sum_{d \geq 0} R_d$$

where each R_d is the collection of homogeneous elements of degree d. Furthermore, notice that $R_d R_e = R_{d+e}$, hence these always from $\mathbb{N}$-graded algebra over R_0 (recall section 20.4.3). By proposition 15.4.6 and lemma 21.4.5, we may work with $\mathbb{N}^{\geq 0}$-graded algebra's and their homogeneous ideals. This sets us up nicely for the following:

Definition 21.4.6: Projective Algebraic Set

Let $R = k[x_0, x_1 ..., x_n]$ with the usual homogeneous graded structure. Let $I \subseteq R$ be a homogeneous ideal. Then a *projective algebraic set* is:

$$Z_{\mathbb{P}}(I) = \{p \in \mathbb{P}^n_k \mid f(p) = 0, \ \forall f \in I\}$$

If it is clear from context, the $\mathbb{P}$ subscript will be omitted and it shall be written as $Z(I)$.

By lemma 21.4.5, we can find the decomposition $f(\lambda x_1, ..., \lambda x_n) = f_0(x_1, ..., x_n) + \lambda f_1(x_1, ..., x_n) + \cdots + \lambda^r f_r(x_1, ..., x_n)$. Then if k is an infinite field, if $f(\lambda x_1, ..., \lambda x_n) = 0$ for all $\lambda \neq 0$ ($\lambda \in k$), then all the homogeneous components are equal to 0, giving us a way of characterizing polynomials which shall give projective algebraic sets. This can also let us characterize homogeneous ideals: an ideal is homogeneous if and only if for every $f \in I$, all of it's homogeneous components are in the ideal.

Note that we could have taken an arbitrary homogeneous set $S \subseteq R$, and then taken the homogeneous closure ${}^h\langle S \rangle = I$. We can further define $I(Y)$ in a similar way as we did for affine algebraic sets, namely

$$\mathcal{I}_{\mathbb{P}}(Y) = \{f \in R \mid f \text{ is homogeneous and } f(p) = 0, \ \forall p \in Y\}$$

We may thus define the *homogeneous coordinate ring* of Y to be:

$$k[Y] = \frac{R}{(\mathcal{I}(Y))}$$

Example 21.4.3: Zero Sets of Projective Space

Example 21.4.2 contained multiple examples of zero sets. All of those examples were from a single homogeneous polynomial.

1. Let us show that every point $[a_0 : \cdots : a_n] \in \mathbb{P}^n_k$ is the zero set of homogeneous polynomials. First, we need that
$$f(x_0, ..., x_n) = \lambda(a_0, ...a_n)$$
for some $\lambda \neq 0$. This is the case when each 2×2 minor of the matrix
$$\begin{pmatrix} x_0 & x_1 & \cdots & x_n \\ a_0 & a_1 & \cdots & a_n \end{pmatrix}$$
vanish, namely that
$$\begin{vmatrix} x_i & x_j \\ a_i & a_j \end{vmatrix} = a_j x_i - a_i x_j = 0$$
This creates $\binom{n}{2}$ equations, however n of these equations suffices (one for each variable). In other words, we get a point via linear homogeneous polynomials with the appropriate coefficients. This show's how "linear" projective space is. These equations can be thought of as the intersection of all the hyper-planes passing through $[a_0 : \cdots : a_n]$. Equivalently, if you know some differential geometry, it shows that we can view $\mathbb{P}^n_k$ as the Grassmanian $G(1, n+1)$.

   The coordinate ring associated to a point would be:
$$\frac{k[x_0, ..., x_n]}{\langle a_i x_i - a_j x_j \big| 0 \leq i, j \leq n \rangle} \cong k[x_0]$$
That is, the coordinate ring is *not* just k at a single point! The reader may think of this as the grading condition for the coordinate rings.

2. In $\mathbb{P}^1_k$, the only closed sets are points, the empty set, and the entire space. We shall see why when we show that any closed subset of projective space can be restricted to a closed subset of affine space, in which these are the only possibilities

3. In $\mathbb{P}^2_k$ has many more closed sets just like $\mathbb{A}^2_k$. Any line is given by $a_0x_0 + a_1x_1 + a_2x_2$ and any conic by $ax_0^2 + bx_0x_1 + cx_1^2 + dx_0x_2 + ex_1x_2 + fx_2^2$. A famous set of curves is the Fermat curves given by $x_0^n + x_1^n + x_2^n$. There are many more equations.

   To give one example of a coordinate ring, the coordinate ring for a line is isomorphic to $k[s, t]$ with $\deg(s) = \deg(t) = 1$.

4. In $\mathbb{P}^3_k$, we have the first non-rational smooth curves given by the set of three equations
$$x_0x_2 - x_1^2,\ x_0x_3 - x_1x_2,\ x_1x_3 - x_2^2$$
Another example in $\mathbb{P}^3_k$ are the quadratic surfaces given by $\sum_{i,j} a_{ij}x_ix_j$. The coordinate ring of the twisted cubic is:
$$\frac{k[x_0, x_1, x_2, x_3]}{(x_0x_2 - x_1^2, x_0x_3 - x_1x_2, x_1x_3 - x_2^2)} \cong k[s^3, s^2t, st^2, t^3] \subseteq k[s, t]$$
Notice this is the binary cubic form (i.e. 2-element degree 3 form).

By proposition 15.4.7, we get that homogeneous ideals are closed under ideal addition, multiplication, intersection, and taking the radical:

$$I + J \qquad IJ \qquad I \cap J \qquad \sqrt{I}$$

Furthermore, to check that a homogeneous ideal I is prime, it suffices to check the condition $fg \in I$ iff $f \in I$ or $g \in I$ with homogeneous polynomials. The addition can naturally be extended to arbitrary sums, allowing us to define the Zariski topology on $\mathbb{P}^n_k$ when focusing on the homogeneous ideals of a $k[x_0, ..., x_n]$:

Definition 21.4.7: Zariski Topology on Projective Space

Letting the projective algebraic sets to be the closed subsets of $\mathbb{P}^n_k$, we get the Zariski topology.

With a topology, we may define all of the previous notion such as compactness, connectedness, dimension, irreducibility, and so forth. The irreducibility condition in projective space becomes a condition on homogeneous prime ideals, form which we define

Definition 21.4.8: Projective Variety

A *projective variety* is an irreducible projective algebraic set.

If k is algebraically closed, then we may naturally want the Nullstellensatz to apply between projective space and homogeneous polynomials. In particular, if $I \subseteq k[x_1, ..., x_n]$ then $Z(I) = \emptyset$ if and only if $I = (1)$. However, this is not the case:

Example 21.4.4: Nullstellensatz False in Projective Space

Take $I_1 = (x_1, ..., x_n) \subseteq k[x_1, ..., x_n]$. Then notice on projective space these ideals have no common zeros (as $0 \notin \mathbb{P}^n_k$), hence $Z(I_s) = \emptyset$. More generally, let $I_s \subseteq k[x_1, ..., x_n]$ denote the ideal consisting of all polynomials which only have terms of degree at least s. Then as each of these ideals contain x_i^s, $1 \leq i \leq n$, it can only vanish at the origin, which isn't present in $\mathbb{P}^n_k$.

Fortunately, the above ideals are the only ideals for which this could happen, namely because these ideals correspond to ideals that contain the origin in affine space:

Lemma 21.4.9: Characterizing Irrelevant Ideals

Let k be algebraically closed and $I \subseteq k[x_1, ..., x_n]$ be a homogeneous ideal. Then $Z(I) = \emptyset$ if and only if $I_s \subseteq I$ for some $s > 0$ where $I_s = (x_1^s, ..., x_n^s)$.

Proof **:**

Certainly any ideal containing I_s give the empty zero set, so conversely assume $Z(I) = \emptyset$. As $k[x_1, ..., x_n]$ is Noetherian, let

$$I = (f_1, ..., f_n)$$

where each f_i is homogeneous. First, as $Z(I) = \emptyset$, I claim the polynomials

$$f_i(1, y_1, ..., y_n) \qquad \left\{ y_i = \frac{x_i}{x_0} \right.$$

have no common roots. For if $(\alpha_1, ..., \alpha_n)$ was a common root of the homogeneous polynomials, $(1, \alpha_1, ..., \alpha_n)$ would be a common root of $f_1, ..., f_n$ and hence a non-empty zero set.

Now, by Hilbert's theorem, there must exist polynomials g_I such that

$$\sum f_i(1, y_1, ..., y_n) g_i(y_1, ..., y_n) = 1$$

Substituting $y_j = x_j/x_0$ and multiplying by the common denominator (which is $x_0^{m_0}$, we get that:

$$x_0^{m_0} \in I$$

By de-homogenizing at different points, we can similarly get $x_i^{m_i} \in I$. Finally $m = \max(m_0, ...m_n)$, and $d = \max(\deg(f_i))$, letting $s = (m-1)(d+1)+1$, then in any monomial $x_0^{a_0} \cdots x_n a_n$ where $\sum a_i \geq s$, at least one of the s_i must occur with an exponent $a_i \geq m$, and as $s_i^{m_i} \in I$, we have that this term is contained in I, and so $I_s \subseteq I$, as we sought to show.

Definition 21.4.10: Irrelevant Ideal

Let $I_s \subseteq k[x_1, ..., x_n]$ be the collection of all homogeneous polynomials for which each term has degree at least s. Then I_s is called the *irrelevant ideal* of degree s

21.4.2 Properties of Projective space: Closure, Covering, Coordinates

A common characterization of $\mathbb{P}_k^n$ is by showing that it can be constructed by gluing together $n+1$ copies of $\mathbb{A}_k^n$. Let $R = k[x_0, x_1, ..., x_n]$, and let H_i be the zero set of the (homogeneous) polynomial x_i for each i, giving us $H_0, H_1, ..., H_n$. These sets certainly cover $\mathbb{P}_k^n$ (their union is $\mathbb{P}_k^n$). If we consider the map $\mathbb{A}_k^n \to \mathbb{P}_k^n$ mapping $(a_1, ..., a_n) \mapsto [1 : a_1 : \cdots : a_n]$, then the hyperplane H_0 is sometimes denotes H_0 and is called the *hyperplane at infinity*. For any hyperplane H_i, we have a map

$$H_i \to \mathbb{P}_k^{n-1} \qquad [a_1, ..., a_{i-1}, 0, a_{i+1}, ..., a_n] \to [a_1, ..., a_{i-1}, a_{i+1}, ..., a_n]$$

Once we define regular maps, the reader can verify that $H_i \cong \mathbb{P}_k^{n-1}$. Let

$$D(x_i) = U_i = \mathbb{P}_k^n - H_i$$

Then these sets are open and cover $\mathbb{P}_k^n$, namely as $0 \notin \mathbb{P}_k^n$, any point $[a_0 : ...a_n]$ must have at least one nonzero coefficient $a_\ell \neq 0$. Then this points lies in U_ℓ. Then we may define a map:

$$\varphi_i : U_i \to \mathbb{A}_k^n \qquad [a_0 : ... : a_n] \mapsto \left(\frac{a_0}{a_i}, ..., \frac{a_n}{a_i}\right)$$

where we omit $\frac{a_i}{a_i}$. Notice how this is similar to de-homogenization. Verify that the map is well-defined, namely that it is independent of representative of points $U_i \subseteq \mathbb{P}_k^n$. Then:

Proposition 21.4.11: Projective Space Affine Cover

Let φ_i be defined as above. Then φ_i is a homeomorphism:

$$U_i \cong \mathbb{A}_k^n$$

Proof **:**
Each φ_i is clearly a bijection. Verify that the image and pre-image of closed sets is closed, which completes the proof.

This implies that we have a decomposition:

$$\mathbb{P}_k^n = \mathbb{A}_k^n \sqcup \mathbb{P}_k^{n-1}$$

For example, if $n = 2$, then:

$$\mathbb{P}_k^2 = \mathbb{A}_k^2 \sqcup \mathbb{P}_k^1$$

The reader should think about this example as part of their intuitions on projective space. This decomposition can be extended further, breaking down $\mathbb{P}_k^{n-1}$ into an affine and projective space. Repeating inductively, we get:

$$\mathbb{P}_k^n = \mathbb{A}_k^n \sqcup \mathbb{A}_k^{n-1} \sqcup \cdots \sqcup \mathbb{A}_k^1 \sqcup \mathbb{A}_k^0$$

Another "immediate" result from our above discussion is that projective space $\mathbb{P}_k^n$ can be covered with $n+1$ copies of $\mathbb{A}_k^n$. This shall become the dominant way of viewing projective space in modern algebraic geometry (similar to how smooth manifolds are viewed as patches of euclidean space glued together), making this result important enough to be a theorem:

Theorem 21.4.12: Projective Variety Affine Cover

Let Y be a projective variety. Then Y is covered by open set $Y \cap U_i$ where each U_i is homeomorphic to $\mathbb{A}_k^n$

Proof **:**
simple, but important exercise.

For the reader's awareness, we box the following[18]:

Corollary 21.4.13: Transition Between Affine Cover

For each H_i, H_j there exists a transition map $\varphi_{ij} : H_i \cap H_j \to H_i \cap H_j$ given by:

$$\begin{aligned}(x_0, ..., \hat{x}_i, ..., x_n) &\mapsto [x_0, ..., x_{i-1}, 1, x_{i+1}, ..., x_n] \\ &\mapsto \left[\frac{x_0}{x_j}, ..., \frac{x_{i-1}}{x_j}, \frac{1}{x_j}, \frac{x_{i+1}}{x_j}, ..., \frac{x_j}{x_j} ..., \frac{x_n}{x_j}\right] \\ &\mapsto \left(\frac{x_0}{x_j}, ..., \frac{x_{i-1}}{x_j}, \frac{1}{x_j}, \frac{x_{i+1}}{x_j}, ..., \hat{1}_j ..., \frac{x_n}{x_j}\right)\end{aligned}$$

where the hats represents omission of the variable, and it was assumed $i < j$ without loss of generality.

[18] this shall be the transition maps to verify the gluing is well-defined

***Proof* :**
immediate.

Let us now use these results to gain a further intuition on which regular function can extend to $\mathbb{P}^n_k$, and what is the closure of certain affine closed sets in $\mathbb{P}^n_k$. Let us start off by trying to extend $f \in k[\mathbb{A}^1_k]$:

$$f : \mathbb{A}^1_k \to k, \quad f(x) = x \qquad \text{to} \qquad f : \mathbb{P}^1_k \to k$$

which would be in $k[\mathbb{P}^1_k]$. Let's say this is defined on U_0. Then as U_0 is dense in $\mathbb{P}^1_k$, there is a unique extension to $\mathbb{P}^1_k$ (see exercise ref:HERE). Then embedding U_0 into $\mathbb{P}^1_k$, we get that:

$$f([1 : x]) = x$$

Then extending to the rest of $\mathbb{P}^1_k$ means finding where the points $[x_0 : x] = [x_0/x : 1]$ map to. To be consistent, we need that $f([1/x : 1]) = x$. Re-labeling $y = 1/x$ we must have that $f([y : 1]) = 1/y$ for all $y \in U_1$. This shall agree with $f([x_0 : x]) = x/x_0$ whenever $x_0 \neq 0$, thus by uniqueness this is the only possible extension. However:

$$f([0 : 1]) = \frac{1}{0} \notin k$$

As we see, f is *not* a polynomial (it is undefined at $y = 0$), and so it cannot be extended to all of $\mathbb{P}^1_k$. This reasoning extends to any (non-constant) regular function over $\mathbb{A}^n_k$, however we must delay the proof till corollary 21.4.27 as it is more technical to deal with the general case.

If we replace the codomain with $\mathbb{P}^1_k$, then we can fix these problems, namely $1/0$ is well-defined: it is the point at infinity (it is $[1 : 0]$). Hence, $f : \mathbb{P}^1_k \to \mathbb{P}^1_k$ where $f([x_0 : x]) = [x_0 : x]$ is a well-defined *regular map* (recall that a regular map between algebraic sets, as opposed to a regular functions, allowed for a codomain different from $\mathbb{A}^1_k$). We shall cover regular and rational maps in section 21.4.3.

Instead of extending regular functions from affine space to projective space, we can extend closed sets of $\mathbb{A}^n_k$ to $\mathbb{P}^n_k$, and conversely restrict closed sets of $\mathbb{P}^n_k$ to any U_i. Let us first take a closed subset $V \subseteq \mathbb{P}^n_k$, and consider its restriction to $U_i \cong \mathbb{A}^n_k$. As $U_i = D(x_i)$ is open, $V \cap U_i = W_i$ is open in V (with the subset topology and Zariski topology of U_i[19]). Let us show that W_i is closed in $U_i \cong \mathbb{A}^n_k$, that is it is the zero set of the appropriate polynomials. If $f_1, ...f_n$ are the homogeneous polynomials defining V[20] so that

$$f_1 = \cdots = f_n = 0$$

with $\deg f_i = n_i$, then on U_i each f_j is converted to:

$$x_i^{n_j} f_j \left(\frac{x_0}{x_j}, ..., 1_j, ..., \frac{x_n}{x_j} \right)$$

For example $f(x_0, x, y) = xy - x_0^2$ would be converted to $xy - 1$ in U_0. Then we can see that $V = \cup W_i$, that is each V can be covered by closed subsets from each U_i.

Conversely, if $W \subseteq \mathbb{A}^n_k$ is a closed subset, we may take the closure in $\mathbb{P}^n_k$, $\overline{W}$, which is the intersection of all closed sets containing W where we interpret W as being set in one of the U_i's. Some closed W are also closed in $\mathbb{P}^n_k$, for example $x^2 + y^2 - 1$ over $\mathbb{C}[x, y]$ when embedded becomes the zero set of

[19] see exercise ref:HERE to show they are equivalent
[20] Recall that we are working over Noetherian rings, and hence each ideal is finitely generated

$x^2 + y^2 - z^2$ whose zero's are all of the form $[\cos(a) : \sin(a) : 1]$ for $a \in \mathbb{C}^*$. On the other hand, if we take $xy - 1$, then then consider $xy - z^2$, we get that the traditional zero's become:

$$\left[a : \frac{1}{a} : 1\right]$$

and we now get a new 0 when $z = 0$ giving two more points:

$$[1 : 0 : 0] \qquad [0 : 1 : 0]$$

which we can think of as adding the two points at infinity of the graph of $y = 1/x$. Another example would be $y - x^2$. When extending to $\mathbb{P}^2_k$, we get $yz - x^2$. By setting $x = 0$, we get that $y = 0$ is a solution (which already exists in $\mathbb{A}^2_k$) as well as the new solution:

$$[0 : 0 : 1]$$

which is a new point at infinity that makes the parabola shape akin to an infinite ellipse.

Finally, notice there is more than one natural set of coordinates on projective space. It is natural to consider the coordinates of $\mathbb{P}^n_k$ to be given by degree 1 homogeneous functions, namely each point $x \in \mathbb{P}^n_k$ is of the form $[x_0 : \cdots : x_n]$ where $[\lambda x_0 : \cdots : \lambda x_n] = \lambda[x_0 : \cdots : x_n]$. However, we may also just as naturally choose to represent each point in any degree of homogeneity, namely:

$$[\lambda w_0 : \cdots \lambda w_n] = \lambda^k [w_0 : \cdots \lambda w_n]$$

where we can have $w_i = (x_i)^k$.

Quasi-Projective Variety

As we saw in the above, affine and projective varieties are very natural linked to each other. Many results that are proven in classical algebraic geometry apply both to projective and affine varieties. Furthermore, the open subset of projective variety (for example $\mathbb{P}^1_k$ minus the point at infinity and zero) have natural algebraic structure but are not the vanishing set of polynomials (they are the opposite, they are the non-vanishing set of polynomials). The properties of such sets often are strongly tied to varieties, or are even varieties themselves (for example $\mathbb{P}^1_k$ without the point at infinity is the closed set $\mathbb{A}^1_K$ in $\mathbb{A}^1_k$, but is an open set in $\mathbb{P}^1_k$). The definition of quasi-projective varieties aims to unify all of these.

If $X \subseteq \mathbb{P}^n_k$ is closed and $U_i = X \cap H_i$ is open in X, they are in fact *closed* in $H_i \cong \mathbb{A}^n_k$. This can be seen by de-homogenizing: if X is given by equations $f_0, ..., f_n$ where $\deg f_i = n_i$, then U_i is given by:

$$x_i^{-n_j} f_j = f(x_0, .., x_{i-1}, 1, x_{i+1}, ..., x_n) = 0$$

If we have $U \subseteq H_i$ which is closed, then we may consider the closure $\overline{U} \subseteq \mathbb{P}^n_k$ by taking the polynomials that generate U and homogenizing them, namely if f is one such polynomial where $\deg f = d$, then the equations describing $\overline{U}$ are of the form

$$x_i^l f(x_1/x_i, ..., x_n/x_i) = 0$$

Which implies that:

$$U = \overline{U} \cap \mathbb{H}_i$$

Hence, if $X \subseteq \mathbb{P}^n_k$, then an open subset of X will correspond to a closed subset of $\mathbb{A}^n_k$. As X is an open subset of itself, it corresponds to an open subset of a closed projective algebraic set (intuitively, this can be thought of as considering the points at infinity). Thus, open subsets of closed projective sets encapsulate affine and projective varieties:

Definition 21.4.14: Quasi-projective Variety

A *quasi-projective variety* is an open subset of a closed projective variety.

A closed subset $C \subseteq U$ of a quasi-projective variety is the intersection of C with a closed subset $\mathbb{P}^n_k$. It should be checked that the irreducible composition given in theorem 21.2.16 applies to quasi-projective varieties (see exercise ref:HERE). If Y, $Y \subseteq X \subseteq \mathbb{P}^n_k$ is a quasi-projective variety in $\mathbb{P}^n_k$, then Y is said to be a subvariety of X.

It must be noted that the collection of quasi-projective varieties is larger than the collection affine and projective varieties. For example $\mathbb{A}^2 \setminus \{(0,0)\}$ is a quasi-projective variety, but is not affine or projective.

21.4.3 Regular Maps

Let us now define regular mappings on quasi-projective varieties. This definition should extend that of affine varieties where we had a polynomials function for each coordinate. As mentioned before, the global regular functions on $\mathbb{P}^n_k$ must be constant, however even removing one point (the point at infinity) will allow us to define polynomials. If we further limit our attention open subsets (subsets on which polynomials don't vanish), we may define even more maps, namely rational maps for which the denominator is never zero.

Recall that a rational map on $\mathbb{P}^n_k$ is the quotient of two homogeneous polynomials, with degree $r - s$ where r, s is the degree of the numerator and denominator respectively. A regular point x of a rational function f would be a point where f can be represented by p/q with $q(x) \neq 0$.

Definition 21.4.15: Regular Point [on Quasi-projective Variety]

Let X be a quasi-projective variety, and $f = p/q$ a regular homogeneous rational function or map of degree 0. If $q(x) \neq 0$ (with codomain k) for some point x, then it is nonzero in some open neighborhood of x, and we say that f is *regular* in a neighborhood of x. If f is regular at all points, then f is said to be *regular*.

When we defined a regular function before, it was an element of the coordinate ring. We have now defined it to be locally regular. When k is algebraically closed, these two are in fact the same. If X is irreducible, then this is the result of proposition 21.3.3. For the general case, say $f = p_x/q_x$ where $q_x \neq 0$ on U_x so that on U_x:

$$q_x f = p_x$$

Then this equation holds on all of X. to see this, choose a regular functions that vanishes on $X \setminus U_x$ but not at x (recall that U_x is dense in X), and multiply it be $p_x q_x$. Then we see that 2 must hold outside of U_x as both sides of the equality will vanish. Then like the irreducible case, we can now

find points $x_1, ..., x_n$ and regular functions $h_1, ..., h_n$ such that

$$\sum_i^n q_x h_i = 1$$

Then for each x_i, multiplying by h_i and adding up we get:

$$f = \sum_i^n p_x h_i$$

and hence f is regular in the same sense given in definition 21.2.2.

Definition 21.4.16: Regular Map [On Quasi-projective Variety]

Let $f : X \to Y$ be a mapping of quasi-projective varieties and $Y \subseteq \mathbb{P}_k^m$. Then the mapping is called *regular* if for every $x \in X$ and every open affine set $U \subseteq H_i^m$ containing the point $f(x)$, there exists a neighborhood U of x such that $f(U) \subseteq H_i$ and the mapping $f : U \to H_i^m$ is regular (see definition 21.3.2)

The collection of regular function is denoted $k[X]$ or $\mathcal{O}(X)$[a], and the collection of regular mapping between two quasi-projective varieties is denoted $\mathrm{Hom}_{\mathbf{Reg}}(X, Y)$.

[a]This notation shall become more prominent in modern algebraic geometry

This definition of regular function is well-defined and is independent of choice of H_i: if $f(x) = (y_0, ..., \hat{1}_i, ..., y_m) \in H_i$ (where the hat indicates omission), it is also contained in H_j when $y_j \neq 0$, and the coordinates of this point in H_i are of the form:

$$(y_0/y_j, ..., 1/y_j, ..., \hat{1}_j, ..., y_m/y_j)$$

Hence, given the mapping $f : U \to H_i$ is given by

$$(f_0, .., \hat{1}_i, .., f_m)$$

then $f : U \to H_j^m$ is given by $(f_0/f_j, ..., 1/f_j, ..., \hat{1}_j, ..., f_m/f_j)$. As $f_j(x) \neq 0$ and the set U' of points of U where $f_j = 0$ is open. Then on U', the functions $f_0/f_j, ..., 1/f_j, ..., \hat{1}_j, ..., f_m/f_j$ are regular, and so $f : U' \to H_j^m$ is regular.

If we have an affine closed set, then the regular mapping $f : X \to Y$ is determined by the mapping $f^* : k[Y] \to k[X]$. If X is isomorphic to a closed subset of an affine space, it shall be called an *affine variety*, and if it isomorphic to a closet subset of projective space, it shall be called *projective variety*. Note all quasi-projective varieties are affine or projective (for example $\mathbb{A}_k^2 \setminus \{(0, 0)\}$).

Let us now characterize regular maps form quasi-projective varieties to projective space[21]. To understand such a map, start with a regular mapping f on X and a point $x \in X$ where $f(x) \in \mathbb{A}_k^m = H_0^m$ so that there is an open neighborhood where $f = (f_1, ..., f_m) : U \to H_i^m$. Then as it is regular we can write $f_i = p_i/q_i$ where p_i, q_i are o the same degree in homogeneous coordinates and $q_i(x) \neq 0$. Taking the l.c.m. of such fractions we get that $f_i = F_i/F_0$ where $F_0, ..., F_m$ are forms of the same degree and $F_0(x) \neq 0$; overall we can think of the point $f(x) \in H_i^m$ as the point:

$$f(x) = [F_0(x) : \cdots : F_m(x)] \in \mathbb{P}_k^m$$

[21]This is in fact a larger exploration that shall culminating in understanding line bundles on projective scheme, see cite:ALG-GEO

in projective space. Note that there is more than one representative for such a regular function, namely if

$$F(x) = [F_0(x) : \cdots : F_m(x)] \qquad G(x) = [G_0(x) : \cdots : G_m(x)]$$

then they are the same point if and only if $F_i G_j = F_j G_i$ on X for all $0 \leq i, j \leq m$. This gives us an explicit condition for when $Y = \mathbb{P}_k^m$:

Proposition 21.4.17: Equivalent Regular Function

Let $f : X \to \mathbb{P}_m^k$ be a set function from an irreducible quasi-projective variety. Then it is a regular mapping if and only if f can be characterized as a collection of maps:

$$[F_0 : \cdots : F_m]$$

all of the same degree in homogeneous coordinates where two mappings are identical if

$$f_i g_j = f_j g_i \text{ on } X \text{ for all } 0 \leq i, j \leq m.$$

Next, let's see that checking that a quasi-projective variety is closed is a open-local condition:

Lemma 21.4.18: Closedness of Quasi-projective Variety is Open-local

Let $Y \subseteq X$ be a quasi-projective variety. Then Y is closed if and only if $X = \cup_\alpha U_\alpha$ where U_α is open and $Y \cap U_\alpha$ is closed for every α. Namely, closedness is open-local for quasi-projective varieties

***Proof* :**
Let $X = \bigcup_\alpha U_\alpha$ with U_α open and $Y \cap U_\alpha$ is closed. As U_α is open,

$$U_\alpha = X - Z_\alpha$$

where Z_α is closed, and by definition:

$$U_\alpha Y = U_\alpha \cap T_\alpha$$

If $Y = \bigcap_\alpha (Z_\alpha \cup T_\alpha)$, then Y is closed. But this is a routine check, completing the proof.

The following condition shall shows that any type of variety is fundamentally locally affine:

Lemma 21.4.19: Varieties Are Locally Affine

Let X be a quasi-projective variety. Then every $x \in X$ has an open neighborhood that is isomorphic to an affine variety

***Proof* :**
Take $X \subseteq \mathbb{P}_k^n$. If $x \in H_0^n$, then $x \in X \cap H_0^n$ and by definition of a quasi-projective variety:

$$X \cap H_0^n = Y - Y_1$$

where Y_1, Y are closed subsets of H_0^n. As $x \in Y$, there exits a polynomial f with coordinates in

H_0^n where $f = 0$ on Y_1 and $f(x) \neq 0$. Let take the ideal (f) which corresponds to all points on Y where f is 0. By treating the ideal like set of point, $D(f) = Y - (f)$ is a neighborhood of x.

I claim this neighborhood is isomorphic to an affine variety. If $(f_1, ..., f_m)$ is the ideal associated to $Y \subseteq H_0^n$, then define the variety in $\mathbb{A}_k^{n+1}$ via:

$$f_1(x_1, ..., x_n) = \cdots = f_m(x_1, ..., x_n) = 0 \qquad f(x_1, ..., x_n)x_{n+1} = 1$$

Then the mapping

$$\varphi(x_1, ..., x_{n+1}) \to (x_1, ..., x_n)$$

give a regular mapping from Z into $D(f)$, and

$$\psi(x_1, ..., x_n) \to (x_1, ..., x_n, f(x_1, ..., x_n)^{-1})$$

is a regular mapping of $D(f)$ into Z where $\varphi \circ \psi = \psi \circ \varphi = \mathrm{id}$, as we sought to show.

This result also showed that distinguished (definition 21.1.13) are naturally associated to an affine algebraic set, namely:

$$k[D(f)] = k[X][1/f]$$

In particular,f is nonzero on $D(f)$, this is well-defined. Overall, when checking if a mapping between quasi-projective is regular, it suffices to check that any affine variety in the codomain has a close pre-image.

Example 21.4.5: Regular Mappings

1. *Projection*: Let $E \subseteq \mathbb{P}_k^n$ be a d-dimensional subspace given by $n - d$ linearly independent linear equations $(L_1, ..., L_{n-d})$. In homogeneous coordinates, the equations L_i are called *linear forms*. The mapping:

   $$\pi(x) = (L_1(x) : \cdots : L_{n-d}(x))$$

   is well-defined and is called the projection with center E. Notice this set is regular on $\mathbb{P}_k^n \setminus E$ as the L_i's do not vanish simultaneously. We for any $X \subseteq \mathbb{P}_k^n$ where $X \cap E = \emptyset$ and X is closed, we have a regular mapping

   $$\pi : X \to \mathbb{P}_k^{n-d-1}$$

   Geometrically, we can interpret the mapping as follows. Take any $(n - d - 1)$-dimensional subspace $V \subseteq \mathbb{P}_k^n$ such that $H \cap E = \emptyset$. Through any point $x \in \mathbb{P}_k^n \setminus E$ and the space E, there passes a unique $(d + 1)$-dimensional subspace E_x. Then this subspace intersects H at a unique point, namely $\pi(x)$. If X intersects E, but does not contain it, then this is a rational mapping.

   We have seen the case where $d = 0$ before (think of the hyperbola to $\mathbb{A}_k^1$)

2. *Veronese Mapping*: Consider the collection of all projective hypersurfaces on $\mathbb{P}_k^n$. return to this, Shafarevich p.40

Rational Functions on Quasi-projective Varieties

When working with affine varieties, we had the quotient of two regular functions. However, mentioned a few times, on $\mathbb{P}^n_k$ there are only constant regular functions, and a rational function, being the ratio of two regular functions, would still be constant.

The solution is to recall the definition of $\mathcal{O}_X$, namely:

$$\mathcal{O}_X = \{\varphi = f/g \in k[x_1, ..., x_n] \mid g \notin I(X)\}$$

where f and g are of the same degree. Then take the subset $\mathfrak{m}_X = \{\varphi = f/g \in \mathcal{O}_X \mid f \in I(X)\}$.

Definition 21.4.20: Rational Function On quasi-projective Variety

Let X be a quasi-projective variety. Then the set of rational functions on X is given by:

$$k(X) \cong \mathcal{O}_X/\mathfrak{m}_X$$

By a reasoning similar to lemma 21.3.7, a homogeneous function vanishes on X if and only if it vanishes on an open subset $U \subseteq X$. Hence $k(X) = k(U)$. As a consequence:

$$k(X) = k(\overline{X})$$

Due to this, when studying rational functions on quasi-projective spaces, we may reduce to studying rational functions to affine or projective varieties, which often leads most results to focus on the irreducible case. The set of regular points shall be called the *domain of definition*. If $f : X \to \mathbb{P}^m_K$, then it is determined by specifying $m+1$ form $[F_0 : \cdots : F_m]$ of $n+1$ homogeneous coordinates of projective space $\mathbb{P}^n_k$ containing X (where at least one form must not vanish on X). Furthermore, if f can be given by functions $[f_0 : \cdots f_m]$ where each f_i are regular at all $x \in X$ and not all vanish at this point, then the map is regular at x, and hence determines a regular mapping of some neighborhood x into $\mathbb{P}^m_k$.

We saw before that birational maps are weaker than isomorphisms. However, rational maps are regular on an open dense set (as the reader should verify with quasi-projective varieties). This leads to the following result:

Proposition 21.4.21: Reduction To Open Subset

Let X, Y be irreducible varieties. Then X and Y are birationally isomorphic if and only if the contains isomorphic open subsets $U \subseteq X$ and $V \subseteq Y$.

Proof **:**
Let $f : X \to Y$ be a birational isomorphism so that $g : Y \to X$ is it's inverse. Let $U_1 \subseteq X$ and $V_1 \subseteq Y$ be the domains of regularity for f, g respectively (domains in which f, g are regular). As $f(U_1)$ is dense in Y, $f^{-1}(V_1) \cap U_1$ is non-empty, and is open. Letting $U = f^{-1}(V_1) \cap U_1$, $V = g^{-1}(U_1) \cap V_1$, then we can see that $f(U) = V$, $g(V) = U$, $f \circ g = g \circ f = \mathrm{id}$, completing the proof.

21.4.4 Products and Mappings

In exercise ref:HERE, the reader showed that the (direct) product algebraic sets $X \subseteq \mathbb{A}_k^n$ and $Y \subseteq \mathbb{A}_k^m$ is algebraic, and lives in the space $\mathbb{A}_k^{n+m}$. If $X' \subseteq \mathbb{A}_k^p$ and $Y' \subseteq \mathbb{A}_k^q$ are algebraic sets st $X \cong X'$ and $Y \cong Y'$, then $X \times Y \cong X' \times Y' \subseteq \mathbb{A}_k^{p+q}$, showing that this definition is embedding-independent. Though not as immediate as affine space, there does exist an N such that $\mathbb{P}_k^n \times \mathbb{P}_k^m \cong \mathbb{P}_k^N$, though $N \neq n+m$. In this section, we shall start by looking into finding the right N. Defining the product of projective spaces shall also allow us to HERE.

To start, recall that $\mathbb{A}_k^n \times \mathbb{A}_k^m \cong \mathbb{A}_k^{n+m}$. A first attempt at understanding $\mathbb{P}_k^n \times \mathbb{P}_k^m$ would be to consider $\mathbb{P}_k^{n+m}$. However, this space is dependent on the representation of the homogeneous points in $\mathbb{P}_k^n$ and $\mathbb{P}_k^m$, for example if $[2:0] = [1:0] \in \mathbb{P}_k^1$, and $[3:0] = [1:0] \in \mathbb{P}_k^1$,then:

$$[1:1:0] \neq [2:3:0] \in \mathbb{P}_k^2$$

Furthermore, if we choose $[x_1 : x_2] \in \mathbb{P}_k^1$ and $[y_1 : y_2] \in \mathbb{P}_k^1$, which point in $\mathbb{P}_k^2$ ought to represent this point? To resolve this, we shall "twist-multiply"

$$[x_1y_1 : x_1y_2 : x_2y_1 : x_2y_2] \in \mathbb{P}_k^{(1+1)(1+1)-1}$$

Notice that for $[\lambda x_1 : \lambda x_2]$ and $[\mu y_1 : \mu y_2]$ then:

$$[\lambda\mu x_1y_1 : \lambda\mu x_1y_2 : \lambda\mu x_2y_1 : \lambda\mu x_2y_2] = [x_1y_1 : x_1y_2 : x_2y_1 : x_2y_2] \in \mathbb{P}_k^3$$

and hence is independent of representation. Generalizing, this, we may define for any homogeneous

$$x = [x_0 : \cdots : x_n], y = [y_0 : \cdots : y_m] \qquad w = [w_{ij}] \in \mathbb{P}_k^{(n+1)(m+1)-1},\ w_{ij} = x_iy_j$$

Consider now the embedding $\varphi : \mathbb{P}_k^n \times \mathbb{P}_k^m \to \mathbb{P}_k^{(n+1)(m+1)-1}$ given by the above. The image is indeed a closed subset, for it satisfies the equations:

$$w_{ij}w_{kl} - w_{ji}w_{lk} = 0 \tag{21.6}$$

Let us show that this construction is natural by showing that it is covered by hyperplanes H_{ij} satisfying the same property as projective space. Consider how the sets H_i get transformed under the embedding φ. To keep track of the dimensions better, let us assume we are working over a field k and write $\mathbb{A}_k^n$ as $\mathbb{A}^n$. Then let $\mathbb{A}_i^n = H_i$, and:

$$\varphi(\mathbb{A}_i^n \times \mathbb{A}_j^m) = W_{ij} = \mathbb{A}_{ij}^{(n+1)(m+1)-1} \cap \varphi(\mathbb{P}_k^n \times \mathbb{P}_k^m)$$

where $\mathbb{A}_{ij}^{(n+1)(m+1)-1}$ is the set of all points where $w_{ij} \neq 0$. Let $N = (n+1)(m+1) - 1$. Consider now the set $\mathbb{A}_{00}^N$. Take a point in projective space where $w_{00} \neq 0$ so that it corresponds to a point in $\mathbb{A}_{00}^N$:

$$[w_{ij}] = \varphi(x, y) \in \mathbb{P}_k^N$$

Take $z_{ij} = w_{ij}/w_{00}$, $x_i = u_i/u_0$, $y_j = v_j/v_0$ as the inhomogeneous coordinates. Then:

$$z_{i0} = x_i \qquad z_{0j} = y_j \qquad z_{ij} = x_iy_j = z_{i0}z_{0j}$$

But then $\mathbb{A}_0^N \cong \mathbb{A}^{n+m}$ with coordinates $(x_1, ..., x_n, y_1, ..., y_n)$ with the isomorphism given by φ! Hence φ is indeed the natural choice for the embedding of $\mathbb{P}_k^n \times \mathbb{P}_k^m$. This may feel more natural if we

recall that each point of projective space $[a_0 : \cdots a_n] \in \mathbb{P}^n_k$ is closed and given by a system of linear equations

$$a_i x_j - a_j x_i = 0$$

mirroring equation (21.6). Notice that we may interpret the point $[w_{ij}]$ as an $(n+1) \times (m+1)$ matrix given by multiplying the points x and y as a $(n+1, 1)$ matrix and a $(1, m+1)$ matrix respectively (note the resulting rank of the matrix will be 1). From this, we may define the product of two quasi-projective varieties

Definition 21.4.22: Serge Embedding: Product of Projective Space

Let $V \subseteq \mathbb{P}^n_k, W \subseteq \mathbb{P}^m_k$ be two quasi-projective varieties. Then:

$$V \times W \subseteq \varphi(\mathbb{P}^n_k \times \mathbb{P}^m_k) = \mathbb{P}^{(n+1)(m+1)-1}_k$$

is the product of the quasi-projective varieties V, W.

Example 21.4.6: Product of Projective Lines

Let $V = W = \mathbb{P}^1_k$ and consider $\varphi(\mathbb{P}^1_k \times \mathbb{P}^1_k) \subseteq \mathbb{P}^3_k$. This set satisfies the quation:

$$w_{11}w_{00} - w_{01}w_{10} = 0$$

which is a (non-degenerate) quadratic Q in $\mathbb{P}^3_k$. Let us consider the set

$$\varphi(\{\alpha\}, \mathbb{P}^1_k)$$

then given $\alpha = [\alpha_0 : \alpha_1]$, these are all points in $\mathbb{P}^3_k$ such that

$$\alpha_1 w_{00} - \alpha_0 w_{10} = 0 \qquad \alpha_1 w_{01} - \alpha_0 w_{11} = 0$$

These equations give lines in $\mathbb{P}^3_k$. Similarly for $\varphi(\mathbb{P}^1_k, \{\beta\})$. The following visual from Hartshorne is an excellent visual:

Figure 21.8: Serge Embedding of Product of Projective Lines

The closed subsets of $\operatorname{im}\varphi \subseteq \mathbb{P}_k^{(n+1)(m+1)-1}$ correspond well with the closed subsets of $\mathbb{P}_k^n \times \mathbb{P}_k^m$. Let us start with seeing which subset of $\mathbb{P}_k^n \times \mathbb{P}_k^m$ are algebraic subsets under the Serge Embedding. Let $X \subseteq \mathbb{P}_k^{(n+1)(m+1)-1}$ be a sub-variety so that it is defined by homogeneous equations $f_i(w_{00}, ..., w_{nm})$. Then these induce the homogeneous polynomials $g_i(u_0, ..., u_n, v_0, ..., v_m)$ on $\mathbb{P}_k^n \times \mathbb{P}_k^m$ where the u_i, v_j are the same homogeneous degree. Conversely, if there are two systems of homogeneous equations taking variables of the same homogeneous degree in u_i, v_j, then there exists homogeneous polynomials taking variables in $u_i v_j$. If the degree's of the variables u_i and v_j differ, say the former are of degree r and the latter of degree s, then we can create the system of homogeneous equations:

$$v_j^{r-s} f_i(u_0, ..., u_n, v_0, ..., v_m) = 0$$

assuming $r > s$. One important consequence of this is that we can directly check for a set $Z \subseteq \mathbb{P}_k^n \times \mathbb{P}_k^m$ being closed by looking for a single polynomial in $\mathbb{P}_k^{(n+1)(m+1)-1}$ instead of looking at for two polynomials in each separate projective space and verifying the set is the direct product of those two closed sets. Let us box this for future reference:

Proposition 21.4.23: Serge Embedding is Closed

Let $\varphi : \mathbb{P}_k^n \times \mathbb{P}_k^m \to \mathbb{P}_k^{(n+1)(m+1)-1}$ be the Serge embedding. Then $\operatorname{im}\varphi$ is closed. Furthermore, $X \subseteq \mathbb{P}_k^n \times \mathbb{P}_k^m$ is closed if and only if $\varphi(X) \subseteq \operatorname{im}\varphi$ is closed.

Proof **:**
formally put together the above argument.

This is actually rather special: as we saw in example ref:HERE (in particular, consider $X = V(xy-1)$, $Y = \mathbb{A}_k^1$, and $\pi : (a, b) \mapsto a$) the image of an regular map need not be closed, but we have there that we have a condition for which the image is closed. However, the image of a projective variety will always be projective! If the reader has done some analysis this should remind the reader of the fact that the image of compact sets are compact. Though all affine and projective algebraic sets are compact (they are closed subsets of the compact space $\mathbb{A}_k^n$ or $\mathbb{P}_k^n$), projective varieties behave more like compact sets. In Algebraic Geometry, cite:HERE, this distinction shall be made by showing that projective sets having the property of being *proper*[22]. To see this, we require a few lemmas:

Lemma 21.4.24: Graph of Regular Mapping

Let $\varphi : X \to Y$ be a regular mapping between quasi-projective varieties. Then the graph

$$\Gamma(\varphi) = \{(x, \varphi(x) \mid x \in X\}$$

is closed in $X \times Y$

If the Zariski topology was Hausdorff, then we could have concluded using classical topological results, however some more care has to be done as the topology is not Hausdorff.

[22] In analysis, a map is proper is the pre-image of a compact set is compact

Proof **:**
First, we may take Y to be a projective space; for an appropriate choice of m, $Y \subseteq \mathbb{P}_k^m$, so that $X \times Y \subseteq X \times \mathbb{P}_k^m$. Then by lemma 21.3.7, f determine a mapping $\overline{f} : X \to \mathbb{P}_k^m$ and $\Gamma_f = \Gamma_{\overline{f}} \cap (X \times Y)$. As $X \times Y$ is closed (as shown in exercises and proposition 21.4.23), if $\Gamma_{\overline{f}}$ is closed then Γ_f is closed.

Next, define $(f, i) : X \times \mathbb{P}_k^m \to \mathbb{P}_k^m \times \mathbb{P}_k^m$ where i is the identity map. Then Γ_f is the inverse image of Γ_i under (f, i). As this is a continuous map, the inverse image of a closed set is closed, and hence it suffices to show that Γ_i is closed.

Now, Γ_i are points $(x, y) \in \mathbb{P}_k^m \times \mathbb{P}_k^m$ of the form

$$[u_0 : \cdots u_m],\ [v_0 : \cdots v_m]$$

where $u_i v_j = u_j v_i$. By proposition 21.4.23, Γ_i is closed, and $\Gamma_{\overline{f}}$ is closed, and so so Γ_f is closed, completing the proof.

Proposition 21.4.25: Projection of Projective Variety

Let X be a projective variety and Y a quasi-projective variety. Then the projection map:

$$\pi_2 : X \times Y \to Y$$

is a closed map (the image of a closed set is closed)

Hence, the projection of a projective variety can always be written as a system of polynomial equations! This is not true for affine varieties, consider $X = V(xy - 1)$ and $Y = \mathbb{A}_k^1$

Proof **:**
Let us first reduce the problem: as $X \subseteq \mathbb{P}_k^n$ is closed, $X \times Y \subseteq \mathbb{P}_k^n \times Y$ is a closed subset, hence if $Z \subseteq X$ is closed, it is closed in $Z \times Y \subseteq \mathbb{P}_k^n \times Y$, and so it is enough to show it for the case of $X = \mathbb{P}_k^n$. Next, As being closed is affine local, it suffices to cover Y by affine open subsets U_i and prove the theorem on each of these. Thus, it suffices to assume that Y is an affine variety. Finally, as Y is closed in $\mathbb{A}_k^m$, $\mathbb{P}_k^n \times Y \subseteq \mathbb{P}_k^n \times \mathbb{A}_k^m$ is closed, and so we can reduce further to $Y = \mathbb{A}_k^m$.

Thus, let us characterize closed subsets $Z \subseteq \mathbb{P}_k^n \times \mathbb{A}_k^m$. By definition of the Zariski topology, and by a simple generalization of proposition 21.4.23, we see that Z is closed if and only if it is the zero set of the collection of polynomials

$$g_i([u_0 : \cdots : u_n], y_1, ..., y_m) \qquad 1 \leq i \leq t$$

In particular, we do not need to create two zero sets for each component and check that the their product produces Z: we can directly work with a set of polynomials. Let us write $g_i(u; y)$ to represent these polynomials with the two different inputs.

Now, take any $y_0 \in \mathbb{A}_k^m$. Then $\pi^{-1}(y_0)$ consists of all solutions (whether empty or non-empty) to the system of equations

$$g_i(u; y_0) = 0 \qquad 1 \leq i \leq t$$

Take the collection of y_0 such that the system of polynomials equations $g_i(u; y_0) = 0$ has a nonzero solution; call this collection $T \subseteq \mathbb{A}_k^m$. We want to show that T is closed, for then the projection π_2

is a closed map. By lemma 21.4.9, the system of polynomials equations $g_i(u; y_0) = 0$ has a nonzero solution if and only if:

$$I_s \not\subseteq (g_1(u, y_0), ..., g_t(u, y_0)) \qquad \forall s \in \mathbb{N}$$

where I_s is the irrelevant ideal of degree $s \geq 1$. Thus, we shall that the set T_s of points $y_0 \in \mathbb{A}_k^m$ for which $I_s \not\subseteq (g_1(u, y_0), ..., g_t(u, y_0))$ is closed, in particular if we find polynomial equations whose zero set is T_s, then as $T = \bigcap_s T_s$, T is closed which completing the proof.

To that end, let k_i be the homogeneous degree of $g_i(u; y)$ in the u_j variables. Let $(M^\alpha)_\alpha$ be the monomials of degree s in $u_0, ..., u_n$ with some total ordering (note this set is finite). Then $I_s \subseteq (g_1(u, y_0), ..., g_t(u, y_0))$ if and only if each monomial M^α can be expressed as:

$$M^\alpha = \sum_{i=1}^{t} g_i(u, y_0) F_{i,\alpha}(u) \tag{21.7}$$

As each g_i has homogeneous degree k_i, it must be that $F_{i,\alpha}$ has degree $s - k_i$ or $F_{i,\alpha} = 0$ if $k_i > s$. Then we may replace $F_{i,\alpha}$ with elements in $(N_i^\beta)_\beta$ which are monomial of degree $s - k_i$ written in some total ordering. Then equation (21.7) holds if and only if M^α is a linear combination of the polynomials $g_i(u, u_0)N_i^\beta$. As this is true for each M^α, this is equivalent to saying that the polynomials $g_i(u, y_0)N_i^\beta$ span the entire vector space S of homogeneous polynomials of degree s in $u_0, ..., u_n$. As $I_s \not\subseteq (g_1(u, y_0), ..., g_t(u, y_0))$, we have that $g_i(u, y_0)N_i\beta$ *do not* span S.

Now, take the matrix (a_α, β) that represents the coefficients of the M^α terms as a linear combination of the $g_i(u, u_0)N_i^\beta$. As not all M^α can be written as such terms, this matrix cannot be full rank. Hence all $\sigma \times \sigma$ minors (where $\sigma = \dim(S)$ are polynomials in the coefficients of the polynomials $g_i(u, y_0)$, in particular they are polynomials in the coordinates of the point y_0. But then we get a system of polynomials each point y_0, showing that T_s is closed!

Thus, since the arbitrary intersection of closed sets is closed $T = \bigcap_s T_s$ is closed, completing the proof.

For those interested in algebraic geometry, the above proof is in part in proving that projective varieties are *proper*.

Corollary 21.4.26: Image of Projective Variety is Closed

Let $f : X \to Y$ be a regular map where X is a projective variety. Then $\operatorname{im} f$ is closed

***Proof* :**
Let $\pi : X \times Y \to Y$ be the natural projection. Consider $\Gamma_f \subseteq Y$. Then certainly:

$$\operatorname{im} f = \pi(\Gamma_f)$$

and hence by proposition 21.4.25 $\operatorname{im} f$ is closed, completing the proof.

We can finally rigorously show that the collection of regular functions on projective space, and more generally projective varieties, are only constant functions:

Corollary 21.4.27: Regular Function on Projective Variety

Let φ be a regular function non an irreducible projective variety. Then $\varphi \in k$, i.e. φ is constant.

Proof **:**
Take $\varphi : X \to \mathbb{A}^1_k$. As $\mathbb{A}^1_k$ is dense in $\mathbb{P}^1_k$, there is a unique extension to a regular map $\overline{\varphi} : X \to \mathbb{P}^1_k$. As X is a projective variety, by corollary 21.4.26 $\overline{\varphi}(X) \subseteq \mathbb{P}^1_k$ is closed. As $\overline{\varphi}(X) \subseteq \mathbb{A}^1_k$, $x_\infty \notin \overline{\varphi}(X)$, and so by irreducibility of $\mathbb{P}^1_k$ it must be that $\varphi(X) = \mathbb{A}^1_k$ or some finite set $S \subseteq \mathbb{A}^1_k$. It cannot be the latter as $\overline{\varphi}(X) = \mathbb{A}^1_k$ needs to be closed in $\mathbb{P}^1_k$ but $\mathbb{A}^1_k$ is open. Hence $\varphi(X)$ is a collection of points. Then as these points are distinct and points are closed, their pre-image would make X the non-trivial union of closed sets, contradicting its irreducibility, and hence it must be that the image is a single point, in other words a constant, as we sought to show.

Corollary 21.4.28: Regular Map on Projective Variety

Let $f : X \to Y$ be a regular map from a projective variety to an affine variety. Then $\operatorname{im} f$ is a point

Proof **:**
Take $f(x) = (\varphi_1(x), ..., \varphi_m(x))$ where each φ_i is a regular function. Then by corollary 21.4.27 these are all constant, but then

$$f(X) = (\alpha_1, ..., \alpha_m)$$

completing the proof.

One final somewhat surprising application is that we can determine whether a set of homogeneous polynomials is irreducible by just looking at their coefficients:

Proposition 21.4.29: Parameterizing Reducible Sets

The collection of points $\zeta \in \mathbb{P}_k^{\nu_{n,m}}$ where $\mathbb{P}_k^{\nu_{n,m}}$ is the Veronese embeddings space, which correspond to reducible polynomials form a closed subset.

Proof **:**
Let Z the set of points $\zeta \in \mathbb{P}^{v_n,m}$ corresponding to reducible polynomials, and by $Z_j (j = 1, \ldots, m-1)$ the set of all points corresponding to polynomials Z that split into factors of degree j and $m-j$. Clearly $Z = \bigcup Z_j$, hence need only prove that each Z_j is closed.

Consider the projective spaces $\mathbb{P}^{\nu_{n,j}}$ and $\mathbb{P}^{\nu_{n,m-j}}$. The multiplication of two polynomials of degree j and $m-j$ determines a mapping

$$f : \mathbb{P}^{v_n,j} \times \mathbb{P}^{v_n,m-j} \to \mathbb{P}^{v_n,m}$$

which is regular. Naturally, $Z_j = f(\mathbb{P}^{v_n,j} \times \mathbb{P}^{v_n,m-j})$. Then as the product of projective spaces is a projective variety, it follows from that Z_j is closed, as we sought to show.

As an example, in the case where $m = n = 2$, then f is reducible iff the determinant of the 2×2 is zero: $\det(a_{ij})_{i,j} = 0$.

*Finite Maps

These maps shall be the mirror of integral extensions, and shall be one of the easiest examples of *proper morphisms* which is a somewhat complicated but fundamental type of morphism in scheme theory.

Definition 21.4.30: Finite Map [on Affine Varieties]

Let $f : X \to Y$ be a regular map between affine varieties where $f(X)$ is dense in Y and $f^* : k[Y] \hookrightarrow k[X]$ is the induced map. Then if $k[X]$ is integral over Y, then f is called a *finite map*

The composition of finite maps is finite. If $f : X \to Y$ is finite, then its fibers must be finite (see exercise ref:HERE). A reasonable assumption would be that a map is finite if and only if its fibers are finite, but that is not the case: an example of a non-finite map would be $\pi_1 : V(xy - 1) \to \mathbb{A}^1_k$ as $k \hookrightarrow k[x, 1/x]$ is not an integral extension. Instead, one way of thinking of a finite map $f : X \to Y$ is that given an element $y \in Y$, as y is moved around in Y and we look at $f^{-1}(y)$, none of the roots could tend to infinity, that is none can disappear. This intuition can captured by the following proposition:

Proposition 21.4.31: Finite Maps Are Surjective

Let $f : X \to Y$ be a finite map between affine varieties. Then f is surjective.

The key result that we shall employ is lying-over property of integral extensions (see theorem 20.5.13), namely that if B/A is an integral extensions, if $\mathfrak{a} \subseteq A$ then $\mathfrak{a}B \subseteq B$. In particular, if $\mathfrak{a} \not\subseteq A$, then $\mathfrak{a}B \not\subseteq B$.

Proof **:**
Let $f : X \to Y$ be a finite map between affine varieties. Take any $y \in Y$ and let $\mathfrak{m}_y \subseteq k[Y]$ be the functions that take the value 0 at y: if $f \in \mathfrak{m}_y$, then $f(y) = 0$. Let $x_1, ..., x_n$ be coordinate functions on Y and $y = (\alpha_1, ..., \alpha_n)$ so that

$$\mathfrak{m}_Y = (x_1 - \alpha_1, ..., x_n - \alpha_n)$$

Then the equations of the variety $f^{-1}(y)$ have the form:

$$f^*(x_1)\alpha_1, ..., f^*(x_n) = \alpha_n$$

By the Nullstellensatz, $f^{-1}(y) = 0$ iff the equations $f^*(x_i) - \alpha_i$ generate the trivial ieal, that is:

$$(f^*(x_1) - \alpha_1, ..., f^*(x_n) - \alpha_n) = k[X]$$

Viewing $k[Y]$ as a subring of $k[X]$, we see that the above is equivalent to:

$$(x_1, \alpha - 1, ..., x_n - \alpha_n) = k[X]$$

In other words, $\mathfrak{m}_y k[X] = k[X]$. Then since $k[X]$ is integral over $k[Y]$, we get the result by the going-up property, which completes the proof.

Corollary 21.4.32: Finite Maps Are Closed Maps

Let $f : X \to Y$ be a finite map between affine varieties. Then f is a closed map.

***Proof* :**
It suffices to check on irreducible closed subsets $Z \subseteq X$. By the above theorem the restriction of f to Z, $f|_Z : Z \to \overline{f(Z)}$, is a finite map between affine varieties, and so by surjectivity we get $f(Z) = \overline{f(Z)}$, but then $f(Z)$ is closed, completing the proof.

Let us now build-up to defining finite maps for quasi-projectile varieties. For that, we start by showing that finiteness is a local property:

Lemma 21.4.33: Finiteness is Local

Let $f : X \to Y$ be a regular map of affine varieties. Then if for every point $y \in Y$, there is an affine neighborhood U of y such that $V = f^{-1}(U)$ is affine and $f : V \to U$ is finite, then f is finite

***Proof* :**
Let $k[Y] = A$ and $k[X] = B$. Take the principal open set around each point of Y, and as Y is compact take a finite subcover, label these $D(g_\alpha)$ so that $Y = \bigcup_\alpha D(g_\alpha)$. Then:

$$V_\alpha = f^{-1}(D(g_\alpha)) = D(f^*(g_\alpha))$$

and

$$k[D(g_\alpha)] = A[1/g_\alpha] \qquad k[V_\alpha] = B[1/g_\alpha]$$

As the map is finite, $B[1/g_\alpha]$ has a finite $A[1/g_\alpha]$-basis, say $w_{i,\alpha}$. Then we can assume that $w_{i,\alpha} \in B$, since if the basis consisted of elements $w_{i,\alpha}/g_\alpha^{m_i}$ where $w_{i,\alpha} \in B$ then the elements $w_{i,\alpha}$ would also be a basis.

Now, take the union of all bases $w_{i,\alpha}$. I claim these form an A-basis for B. To see this, take any $b \in B$. Then for each α we can write it as :

$$b = \sum_i \frac{a_{i,\alpha}}{g_\alpha^{n_\alpha}} w_{i,\alpha}$$

As $g_\alpha^{n_\alpha}$ generate the entire A (as the union of the distinguished sets $D(g_\alpha)$ is Y), there exists $h_\alpha \in A$ such that

$$\sum_\alpha g_\alpha^{n_\alpha} h_\alpha = 1$$

Thus:

$$b = b \sum_\alpha g_\alpha^{n_\alpha} h_\alpha = \sum_i \sum_\alpha a_{i,\alpha} h_\alpha w_{i,\alpha}$$

as we sought to show.

Thus, we can define finite maps for quasi-projective varieties:

Definition 21.4.34: Finite Map [on Quasi-Projective Varieties]

Let $f : X \to Y$ be a regular map between quasi-projective varieties. Then f is said to be *finite* if any point $y \in Y$ has an affine local neighborhood V such that $U = f^{-1}(V)$ is affine and $f : U \to V$ is a finite map between affine varieties.

Naturally, $f^{-1}(y)$ is finite for every $y \in Y$, and it follows easily that any finite map is surjective. Note that the density condition is now a local condition. One consequence of these results that we shall not prove is that if $f : X \to Y$ between quasi-projective is a regular map (not necessarily finite) and $f(X)$ dense in Y, then $f(X)$ contains an open subset of Y. This is *not* true in analysis, the cannonical example being the wrapping of the real line around the torus with an irrational period which will be a dense map with empty interior. For a proof of the existence of an open subset in a dense mapping, see Shafarevich cite:HERE.

Normalization

An important culminating result is that every irreducible projective variety is in some way only a "finite extension" from projective space. These results shall allow us to diminish many arguments about irreducible varieties to arguments about irreducible varieties into arguments about affine/projective spaces that are closed under finite extension! These ideas shall come back more in modern algebraic geometry (see cite:HERE algGeo) We first need a technical lemma:

Lemma 21.4.35: Projection is Finite Mapping

let X is closed in $\mathbb{P}^n$ and $X \subset \mathbb{P}^n - E$, where E is a d-dimensional linear subspace. Then the projection $\pi : X \to \mathbb{P}^{n-d-1}$ with center at E determines a finite mapping $X \to \pi(X)$.

Proof **:**
Let $y_0, \dots, y_{n-d-1}$ be homogeneous coordinates in $\mathbb{P}^{n-d-1}$ and let π be given by the formulae $y_j = L_j(x)$, $x \in X$ $(j = 0, \dots, n-d-1)$. Naturally, $U_i = \pi^{-1}(\mathbb{A}^{n-d-1}) \cap X$ is given by the condition $L_i(x) \neq 0$ and is an affine open subset of X. We want to show that $\pi : U_i \to \mathbb{A}^{n-d-1} \cap \pi(X)$ is a finite mapping. To that end, first note that every function $g \in k[U_i]$ is of the form $g = G_i(x_0, \dots, x_n)/L_i^r$, where G_i is a form of degree m. Consider the mapping $\pi_1 : X \to \mathbb{P}^{n-d} : z_j = L_j^m(x)$ $(j = 0, \dots, n-d-1)$, $z_{n-d} = G_i(x)$, where $z_0, \dots, z_{n-d}$ are homogeneous coordinates in $\mathbb{P}^{n-d}$. It is a regular mapping and its image $\pi_1(X)$ is closed in $\mathbb{P}^{n-d}$. Let $F_1 = \cdots = F_s = 0$ be its equations. Since $X \subset \mathbb{P}^n - E$, the forms $L_i (i = 0, \dots, n-d-1)$ do not vanish simultaneously on X. Hence, the point $0 = [0 : \cdots : 1]$ is not contained in $\pi_1(X)$, in other words, that the equations $z_0 = \cdots = z_{n-d-1} = F_1 = \cdots = F_s = 0$ have no solution in $\mathbb{P}^{n-d}$. Hence, for some $l > 0$:

$$(z_0, \dots, z_{n-d-1}, F_1, \dots, F_s) \supset T_l$$

in particular, $z_{n-d}^l \in (z_0, \dots, z_{n-d-1}, F_1, \dots, F_s)$. This means that

$$z_{n-d}^l = \sum_{j=0}^{n-d-1} z_j H_j + \sum_{j=1}^{s} F_j P_j,$$

where H_j and P_j are polynomials. Denote by $H^{(q)}$ the homogeneous component of H of degree q. Then:

$$\Phi(z_0, \ldots, z_{n-d}) = z_{n-d}^l - \sum z_j H_j^{(l-1)} = 0 \quad \text{on } \pi_1(X) \tag{21.8}$$

The homogeneous polynomial Φ is of degree l and as a polynomial in z_{n-d} it has the leading coefficient 1:

$$\Phi = z_{n-d}^l - \sum_{j=0}^{l-1} A_{l-j}(z_0, \ldots, z_{n-d-1}) z_{n-d}^j. \tag{21.9}$$

Substituting the transformation formulae for π_1 in equation (21.8) we find that $\Phi(L_0^m, \ldots, L_{n-d-1}^m, G_i) = 0$ on X, with Φ of the form equation (21.9). Dividing this relation by L_i^{ml} we obtain the required relation

$$g^l + \sum_{j=0}^{l-1} A_{l-j}(x_0, \ldots, 1, \ldots, x_{n-d-1}) g^j = 0,$$

where $x_r = y_r/y_i$ are coordinates in $\mathbb{A}^{n-d-1}$. But then we have that the mapping is finite, completing the proof.

Theorem 21.4.36: Normalization of Projective Variety

Let X be an irreducible projective variety. Then there exists a finite mapping $\varphi : X \twoheadrightarrow \mathbb{P}_k^m$ for appropriate m.

***Proof* :**
Take $X \subseteq \mathbb{P}_k^n$. If $X = \mathbb{P}_k^n$ we're done so take $X \neq \mathbb{P}_k^n$. Take $x \in \mathbb{P}_k^n \setminus X$ and take the projection from x, $\varphi_x : X \to \mathbb{P}_k^{n-1}$. This map is regular by example 21.4.5, and the image is closed, hence projective. This map is furthermore finite by lemma 21.4.35. If $\varphi(X) \neq \mathbb{P}_k^{n-1}$. We can repeat this argument. All the same properties will remain, as the composition of closed/finite maps is closed/finite. This process shall terminate since if $n - \ell = 1$ then it must be that $\varphi(X) = \mathbb{P}_k^1$. Hence we get that for appropriate m a finite mapping

$$\varphi(X) = \mathbb{P}_k^m$$

completing the proof.

Theorem 21.4.37: Normalization of Affine Variety

Let X be an irreducible variety variety. Then there exists a finite mapping $\varphi : X \twoheadrightarrow \mathbb{A}_k^m$ for appropriate m.

Note that this is the analog of Noether's Normalization theorem for rings that are finitely generated k-algebras without zero-divisors, and in fact can be proven using it, see exercise 21.4.1. This proof is more generalizable to more general rings, as is important for modern algebraic geometry (see cite:HERE algGeo, or Hartshorne's chapter on Curves and Surfaces).

***Proof* :**
Take $X \subseteq \mathbb{A}_k^n$. If $X = \mathbb{A}_k^n$ we're done so take $X \neq \mathbb{A}_k^n$. Consider $\mathbb{A}_k^n \cong U_i$, X is open in $\mathbb{P}_k^n$ so take $\overline{X}$. Then choose a point $x \in \mathbb{P}_k^n \setminus \mathbb{A}_k^n$ where $x \notin \overline{X}$ (which exists as $X \neq \mathbb{A}_k^n$). Then take the same projection $\varphi\overline{X} \to \mathbb{P}_k^{n-1}$. Then it is not hard to see that X is projected onto $\mathbb{A}_k^{n-1}$. We may again repeat this process like in theorem 21.4.36 a surjective map:

$$\varphi(X) = \mathbb{A}_k^m$$

completing the proof.

Exercise 21.4.1

1. Let $R_d \subseteq k[x_1, ..., x_n]$ be the vector space of homogeneous polynomials of degree d. Show that $\dim_k R_d = \binom{d+n-1}{n=1}$
2. Show that any two lines in projective space $\mathbb{P}_k^n$ intersect, that is if L_1, L_2 are two equation of lines and we set $L_1 = 0, L_2 = 0$, then these two zero-sets must have a common point.
3. if $\frac{\partial f}{\partial x_i}$ is the formal partial derivative applied to a function f, then show that if f is a homogeneous polynomial of degree d that

$$df = \sum_{i=1}^{n} x_i \frac{\partial f}{\partial x_i}$$

4. Let R be a graded ring so that $1 = \sum_{k=0}^{\infty} r_k$. Show that r_0 is the multiplicative identity of R_0.
5. Check the decomposition theorem for quasi-projective varieties.
6. (Calculus exercise) Let f be a k-times differentiable function degree k homogeneous function. Show that f is a degree k homogeneous polynomial. Hence homogeneity characterizes homogeneous polynomials.
7. Good exercise!
8. Since Noether's Normalization theorem was already proven earlier in this chapter, I think a good exercise would be converting the result from Shafarevich into some guided exercises here (theorem 7-10 chapter 1 section 5) here
9. Prove the affine normalization theorem using Noether's Normalization Theorem.

21.4.5 Bézout's Theorem

Arguably one of the most important results on [classical] projective varieties is the fact that we can extend he fundamental theorem of algebra to is fullest extend. The fundamental Theorem of algebra states that a polynomial of degree n over an algebraically closed fields has has n roots. This can also be thought of as the polynomial f and the line $x = 0$ intersecting at n locations (counting multiplicity). Generalizing this, we shall see that we can have two polynomials f, g of degree n and

m that, when embedded into projective space, intersect at nm places (counting multiplicity). Let us first establish some vocabulary:

Definition 21.4.38: Projective Plane Curve

A projective plane curve is an equivalence class of nonconstant homogeneosu polynomial in $k[x, y, z]$, where $f \sim g$ if and only if $f = \alpha g$ for some $\alpha \in k^*$. The degree of a curve is defined to be the degree of the defining homogeneous polynomial.

Theorem 21.4.39: Bézout's Theorem

Let $C = Z_{\mathbb{P}}(f)$ and $D = Z_{\mathbb{P}}(g)$ be projective plane curves that do not have a common component. Then, if C has degree m and D has degree n, C and D intersect in mn points counting multiplicity.

Proof **:**
(will put down soon)

21.5 Dimension Theory

When looking at the variety associated to $k[x, y, z]$, namely $\mathbb{A}^3_k$, we may natural think of this space as being 3-dimensional, and in this special case match with our notion of a 3-dimensional vector-space. We may think that varieties also have a notion of dimension, for example the curve with the corresponding corresponding coordinate ring $\frac{k[x,y]}{(y^2-x)}$ Should be thought of as a 1-dimensional variety

This variety should even be thought of as being smooth, as compared to the variety $\frac{k[x,y]}{y^2-x^3}$ which has a cusp at $(x, y) = (0, 0)$.

We may hope that the notion of dimension is as simple as it was for vector spaces, however this turns out to be a more subtle problem. For example, continuous maps do not preserve dimension (there is a continuous map from $\mathbb{R}$ onto $\mathbb{R}^2$ giving by Peano space-fill curve). Homeomorphisms between open sets of $\mathbb{R}^n$ and $\mathbb{R}^m$ do force the condition that $n = m$, however this proof requires a good amount of homological build-up[23].

The most general definition of dimension can be defined in the topological level and is called the *Legesgue Covering Dimension* where a space X is dimension at most n if every open cover can be refined to an open cover where each point is covered by at most $n+1$ sets (see CITE:TOP). However, this definition will fail spectacularly in the Zariski topology as most open sets are huge (in fact, most Zariski topologies will have infinite Lebesgue covering dimension).

There are a few other notion of dimension (to name some notable example: the Brower's-Menger-Urysohn's Boundary Dimension, the Hausdorff Dimension, Deviation of Poset). One to point out that is meaningful that we have seen with the Nullstellensatz is by looking at the transcendental degree of a field extension. For example, if we take $k[V]$ to be some coordinate ring for a variety (so that $k[V]$ is an integral domain), we may take the field of fractions $k(V)$. Then taking the transcendental degree of $k(V)/k$, we may define this to be the *transcendental dimension* of $k[V]$. This works well on smooth shapes, and by exercise ref:HERE we see that all affine and projective varieties have finite maps to $\mathbb{A}^n_k$ or $\mathbb{P}^m_k$ respectively. Furthermore, it is not to difficult to prove that the dimension of these spaces is invariant (see exercise ref:HERE). However this definition fails when there the coordinate ring is not an integral domain[24]. Furthermore, it does not work so well once we try to generalize this notion to spectrum in section **??** (for example, we shall show that the geometric object associated to $\mathbb{Z}$ will have transcendence degree 0, while in the definition that we shall settle on it will have dimension 1).

One fundamental limiter is that these notion of dimension define it in a *global way*; the entire space has some dimension. This turns out to be a misnomer, as the example of the curve $y^2 = x^3$ shows (it *drops* a dimension at $(x, y) = (0, 0)$). To anyone who has done differential geometry, they may already be comfortable with the fact that the dimension of a smooth manifold is given by the vector space dimension of the tangent space at a point (which for a smooth-manifold should be invariant

[23] see CITE:ALGTOP

[24] It also fails as an adequate definition when localizing a ring or taking it's completion, which are two important tools that we shall often use on shapes

to the choice of point, and hence a well-defined global notion). Then the reader well-versed in the equivalent definition's of tangent spaces of a manifold may recall that the tangent space can be defined as the dual of $\mathfrak{m}/\mathfrak{m}^2$ where $\mathfrak{m}$ is the maximal ideal of C_a^∞ (the set of germs of C^∞ on a manifold at the point a). This definition is agnostic of the differential structure, and so can be used on varieties. However, it fails in some important cases; to reiterate the same example the tangent space at $(x, y) = (0, 0)$ of $y^2 = x^3$ will be dimension 2, even though we can see it ought to be 1. We shall explore this notion starting in section 21.6.

21.5.1 Krull dimension and Transcendental Degree

What has been settled on as the most useful notion of global dimension that translates well between algebra and geometry is the following:[25]

Definition 21.5.1: Topological Krull Dimension

Let X be a topological space. Then the *Krull dimension* of X is the supremum of the length of all chains:

$$Z_0 \subsetneq Z_1 \subsetneq \cdots \subsetneq Z_n$$

of distinct irreducible subsets of X The Krull dimension of a topological space will be denoted kdim

This definition is not interesting in euclidean topology, for example for $\mathbb{R}^n$ as only points are irreducible we have $\operatorname{kdim}(\mathbb{R}^n) = 0$. If the dimension is finite, then it is equal to the length of the longest chain of irreducible closed subsets. The Krull dimension of $\mathbb{A}^1_{\overline{k}}$ is always 1, $\operatorname{kdim}(\mathbb{A}^1_{\overline{k}}) = 1$, as the only irreducible closed subsets of $\mathbb{A}^1_{\overline{k}}$ are the single point and whole space. Another simple example are Principal Ideal Domains which all have Krull dimension 1. As irreducible closed subsets correspond to prime ideals, we have the equivalent notion for primes:

Definition 21.5.2: Height of Prime

Let $\mathfrak{p} \subseteq R$ be a prime ideal. Then the *height* of a prime ideal is the supremum of all integers n such there is a chain of distinct prime ideals:

$$\mathfrak{p}_0 \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_n = \mathfrak{p}$$

The *dimension* of R is the supremum of the heights of all prime ideals.

Before giving some example, it will be useful to establish some basic properties first:

Proposition 21.5.3: Dimension of Affine Algebraic Set

Let $Y \subseteq \mathbb{A}^n_k$ be an affine algebraic set. Then $\operatorname{kdim}(Y)$ is equal to the dimension of the affine coordinate ring $k[Y]$

[25] The Krull dimension turns out to be equivalent to the Brouwer-Menger-Uryson definition of dimension, which is useful for cohomological intuitions

Proof **:**
The closed irreducible subsets of Y correspond to the prime ideals of $k[x_1, ..., x_n]$ containing $I(Y)$, which correspond to the prime ideals of $k[Y]$. But then $\text{kdim}(Y)$ is the ieght of the longest prime ideals in $k[Y]$, as we sought to show.

When covering Noether's Normalization (theorem 21.2.23), it was observed that this can be used to define the dimension of a ring. The following definition captures this idea:

Definition 21.5.4: Transcendental Dimension

Let X be a quasi-projective irreducible variety. Then the *transcendental dimension* of X is the transcendental degree of the field $k(X)$. It shall be denoted $\dim X$.

Proposition 21.5.5: Transcendental and Krull Dimension

Let k be a field and R a finitely generated k-algebra. Then

$$\text{kdim}(R) = \text{trdeg}_k(R)$$

Furthermore, if $\mathfrak{p} \subseteq R$, then:

$$\text{height}(\mathfrak{p}) + \text{kdim}(R/\mathfrak{p}) = \text{kdim}(R)$$

Proof **:**
exercise

The conditions are necessary for the equality of both equations:

Example 21.5.1: Different Krull and Transcendental Dimension

1. Take $R = k(x)$. Then it has Krull dimension 0, but transcendental dimension 1
2. In general the Krull dimension equalling the transcendental dimension is well-defined because of the finite generation: Let R be a ring with transcendence dimension n, and let the following be a chain of prime ideals
$$\mathfrak{p}_0 \subsetneq \cdots \subsetneq \mathfrak{p}_m = \mathfrak{p}$$
so that the height of $\mathfrak{p}$ is m. Choose $p_i \in \mathfrak{p}_i \setminus \mathfrak{p}_{i-i}$, let let $B = k[p_1, ..., p_n]$ be the subalgebra $B \subseteq R$ generated by these elements. Define the primes $\mathfrak{q}_i = \mathfrak{p}_i \cap B$ which still form a chain as $p_i \in \mathfrak{q}_i \setminus \mathfrak{q}_{i-1}$. Then $\text{Frac}(B) \subseteq \text{Frac}(R)$, hence the transcendence dimension of $\text{Frac}(B)$ is at most n. As B is finitely generated, the Krull dimension must be at most n (or else we can create a longer chain). But then $m \leq n$.

   For those who know some scheme theory, we can think of this phenomenon as localizing at a non-closed point. The $k(x)$ is the localization at the generic point of $\mathbb{A}^1_k$, while the algebraic dimension is a field.

Corollary 21.5.6: Dimension of Affine Space

The Krull dimension of affine space $\mathbb{A}_k^n$ is n.

Proof **:**
apply proposition 21.5.5.

Codimension

Definition 21.5.7: Codimension

Let $Y \subseteq X$ be a closed subvariety of X. Then

$$\operatorname{codim} Y = \dim X - \dim Y$$

Note that dimension is thus determined by any open subset of a variety, as $k(U) = k(X)$ implies $\dim U = \dim X$. Verify that planar curves have dimension 1, finite sets have dimension 0, $\mathbb{A}_k^n$ and $\mathbb{P}_k^n$ have dimensions n, and that $\dim(X \times Y) = \dim(X) + \dim(Y)$.

Definition 21.5.8: Curves and Surfaces

Let X be an algebraic variety. Then:

1. If X is 1-dimensional, it is called a *curve*
2. If X is 2-dimensional, it is called a *surface*
3. If X is 1-codimensional, it is called a *hypersurface*.

Proposition 21.5.9: Dimension Between Irreducible Varieties

Let $Y \subseteq X$. Then $\dim Y \leq \dim X$. If Y is closed and $\dim(Y) = \dim(X)$, then $X = Y$

There are lots of great theorem from Shafarevich that I will put down when I have the time

Shafarevich p.54

Main idea's that I want to have down:

1. irred components of hypersurface have codim 1, and variety whose components have codim 1 is hypersurface (their ideals are also principal).
2. any subvariety $X \subseteq \prod \mathbb{P}_k^{n_i}$ whose components all have codim 1 is given by a singel equation, homogeneous in each groups of variables.
3. Look at intersecting hypersurfaces.
4. Non-vanishing form on irred proj var will add a single dimension of restriction: $\dim X_F = \dim X - 1$

5. results relating dim and codim and forms, incl under intersections

6. results on fibers!

7. Classifying lines on $\mathbb{P}^3_k$.

8. Chow coordinates of proj variety

Exercise 21.5.1

1. Let X, Y be irreducible varieties. Show that $\dim(X \times Y) = \dim(X) + \dim(Y)$ (hint: use the transcendental degree definition)

21.6 *Local Dimension: Hilbert Function

As mentioned earlier, there are flaws in the other ways of defining dimension. The proper way of defining dimension is by using *Hilbert Functions*. This notion shall become more important in cite:ALGGEO, however the key concepts are useful in a more nuanced interpretation of dimension.

Additive Functions

Definition 21.6.1: Additive Function on Collection

Let C be a class of R-modules and $\lambda : C \to \mathbb{Z}$ a function. Then λ is *additive* if for each short exact sequence

$$0 \to M' \to M \to M'' \to 0$$

with the R-modules belonging to C we have

$$\lambda(M') - \lambda(M) + \lambda(M'') = 0 \qquad \text{equiv.} \qquad \lambda(M') + \lambda(M'') = \lambda(M)$$

Example 21.6.1: Additive Function

1. Let $R = k$ and let C be all finite-dimensional k-vector spaces. Then $V \mapsto \dim V$ is an additive function on C.

2. Let C be the collection of all Artinian ring (recall they are also Noetherian). Then the length is well-defined, and in fact $\ell(-)$ is an additive function.

Proposition 21.6.2: Additive Function on Exact Sequence

Let
$$0 \to M_0 \to M_1 \to \cdots \to M_n \to 0$$
be an exact sequence of R-modules. Then for any additive function λ on C,
$$\sum_{i=0}^{n} (-1)^i \lambda(M_i) = 0$$

Proof **:**
Recall any exact sequence can be split into short exact sequences:
$$0 \to N_i \to M_i \to N_{i+1} \to 0$$
where $N_0 = N_{n+1} = 0$. Then $\lambda(M_i) = \lambda(N_i) + \lambda(N_{i+1})$, and taking the alternating sum of the $\lambda(M_i)$ will simply insure everything cancels out.

Hilbert Function

Let $R = \bigoplus_{n=0}^{\infty} R_n$ be a Noetherian graded ring. Then R_0 is Noetherian, and say that R is generated by $r_1, ..., r_s$ as an R_0-algebra, which without loss of generality we may take to be homogeneous elements of (strictly positive) degrees $k_1, ..., k_s$. The ring I keep in mind is any homogenized curve, for example:
$$\frac{k[x_0, x, y]}{(x_0 x^2 - y^3)}$$
Let M be a finitely generated graded R-module so that M can be generated by a finite number of homogeneous elements $m_1, ..., m_t$. This can be thought of as a fiber bundle on our curve (ex. a vector bundle, line bundle, or tangent bundle). By corollary 20.4.12, the following is well-defined:

Definition 21.6.3: Poincaré Series

The *Poincaré Series* of M with respect to λ is the generating function of $\lambda(M_n)$, that is the power series:
$$P(M, t) = \sum_{n=0}^{\infty} \lambda(M_n) t^n \in \mathbb{Z}[[x]]$$

Proposition 21.6.4: Poincaré Series is Rational Function

Let $P(M, t)$ be the Poincaré series for M with respect to λ, where M is a finitely generated R-module over a graded ring R that is a finitely generated R_0-algebra generated by s elements each with homogeneous degree k_i. Then $P(M, t)$ is a rational function in t of the form:
$$P(M, t) = \frac{f(t)}{\prod_{i=1}^{s} (1 - t^{k_i})} \qquad f(t) \in \mathbb{Z}[t]$$

The polynomial $f(t)$ can be thought of as holding information about the lack of freeness.

***Proof* :**
We shall proceed by induction on s. Starting with $s = 0$ so that $R = R_0$ and $R_n = 0$ for all $n > 0$. Then M is a finitely generated R_0-module, and so the generators have at most some degree n_0, and hence $M_n = 0$ for $n > n_0$. But then $P(M, t)$ is a polynomials, as $\lambda(0) = 0$ (which can be verified using short exact sequences). It's coefficients will be the dimensions of the M_n's.

Now assume the result holds for any R_0-algebra R generated by $s - 1$ elements and let R be generated by s elements, $r_1, ..., r_s$, each with degree k_i (which without loss of generality we may take are > 0). We shall find how to reduce to $s - 1$ elements to apply the inductive hypothesis. To that end, consider the R-module homomorphism $M \xrightarrow{\cdot r_s} M$ that will map $M_n \to M_{n+k_s}$. This induces an exact sequence:

$$0 \to r_s M_n \to M_n \xrightarrow{r_s} M_{n+k_s} \to M_{n+k_s}/r_s M_n \to 0$$

Let $r_s M_n = K_n$ and $M_{n+k_s}/r_s M_n = L_{n+k_s}$ for each $n \geq 0$. Now, applying the additive function λ to the above short exact sequence, we get by proposition 21.6.2 that for all n:

$$\lambda(K_n) - \lambda(M_n) + \lambda(M_{n+k_s}) - \lambda(L_{n+k_s}) = 0$$

which re-arranged we get:

$$\lambda(M_{n+k_s}) - \lambda(M_n) = \lambda(L_{n+k_s}) - \lambda(K_n)$$

Take $K = \bigoplus K_n$ and $L = \bigoplus L_n = M/r_s M$. Then these are both finitely generated, being a submodule and a quotient of a finitely generated R-module M. As they are both annihilated by r_s, we have that they are both $R_0[r_1, ..., r_{s-1}]$-modules. This shall let us apply the induce hypothesis to the following chain of equalities:

$$\begin{aligned}
(1 - t^{k_s})P(M, t) &= \sum_{0}^{\infty} \lambda(M_n)t^n - \sum_{n-0}^{\infty} \lambda(M_n)t^{n+k_s} \\
&= g(t) + \sum_{n=0}^{\infty} \left(\lambda(M_{n+k_s}) - \lambda(M_n)\right) t^{n+k_s} \\
&= g(t) + \sum_{n=0}^{\infty} \left(\lambda(L_{n+k_s}) - \lambda(K_n)\right) t^{n+k_s} \\
&= P(L, t) - t^{k_s} P(K, t) \\
&\overset{\text{I.H.}}{=} \frac{f_L(t) - t^{k_s} f_K(t)}{\prod_{i=1}^{s-1} 1 - t^{k_i}} \\
&= \frac{f(t)}{\prod_{i=1}^{s-1} 1 - t^{k_i}}
\end{aligned}$$

where

$$g(t) = \sum_{n=0}^{k_s - 1} \lambda(M_n)t^n \in \mathbb{Z}[x]$$

Rearranging, we get:

$$P(M, t) = \frac{f(t)}{\prod_{i=1}^{s}(1 - t^{k_i})} \qquad f(t) \in \mathbb{Z}[t]$$

as we sought to show.

Hence, $P(M,t)$ has a pole at $t=1$. Let $d(M)$ denote the order of the pole; this pole shall in many ways give a measure of the size of M relative to λ.

For the rest of this section, let the degree of the zero polynomial be -1 and $\binom{n}{-1}=0$ for all $n \geq 0$ and if $n=1$, then $\binom{-1}{-1}=-1$.

Corollary 21.6.5: Hilbert Polynomial of M

If $k_i = 1$ for each i, then for sufficiently large n, $\lambda(M_n)$ is a polynomial in n (with rational coefficients) of degree $d-1$

***Proof* :**
Assuming $k_i = 1$, from proposition 21.6.4, we get that each $\lambda(M_n)$ is the coefficient of t^n in the expansion of:

$$\frac{f(t)}{(1-t)^s}$$

If $f(t)$ has $(1-t)$ factors, we may cancel them out so that $s=d$, let $g(t)$ be the polynomial such that

$$\frac{f(t)}{(1-t)^s} = \frac{g(t)}{(1-t)^d}$$

where $g(1) \neq 0$. Let $g(t) = \sum_{k=0}^{N} a_k t^k$. Now, power-series of $(1-t)^{-d}$ is:

$$(1-t)^{-d} = \sum_{k=0}^{\infty} \binom{d+k-1}{d-1} t^k$$

Grouping terms we get for all terms larger than the highest degree of g:

$$\lambda(M_n) = \sum_{k=0}^{\infty} a_k \binom{d+n-k-1}{d-1} \qquad \forall n \geq N$$

Then distributing the right hand side, we get that it is a polynomial in n, its leading term being

$$\frac{\sum a_k}{(d-1)!} n^{d-1}$$

showing it is of degree $d-1$, completing the proof.

Note that $f(n)$ may be an integer for all integers n even though the coefficients of f are not all integers, for example $\frac{1}{2}x(x+1)$. In fact, all integer-valued polynomials will be an integer linear combination of the terms (see exercise ref:HERE):

$$\binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \ldots$$

Definition 21.6.6: Hilbert Function

Let R be a graded and finitely generated generated R_0 algebra where R_0 is Artinian. Let M be a finitely generated graded R-module. THen the *Hilbert function* of M is given by:

$$H_M(n) := \ell_{R_0}(M_n)$$

for all $n \in \mathbb{N}$

Proposition 21.6.7: Non-Zero Divisor and Order of Pole

Let $x \in R_k$ be an element that is *not* a zero-divisor in M. Then:

$$d(M/xM) = d(M) - 1$$

Proof **:**
As x is not a zero-divisor we have that $xm = 0$ implies $m = 0$. Then in the exact sequence:

$$0 \to K_n \to M_n \xrightarrow{x} M_{n+k_s} \to L_{n+k_s} \to 0$$

we have $K_n = 0$, and $L = M/xM$. Then by equation (**??**) we get:

$$d(L) = d(M) - 1$$

which and as $L = M/xM$, we have $d(M/xM) = d(M) - 1$, as we sought to show.

Example 21.6.2: Additive Function and Hilbert Polynomial
An example the reader may want to try is with A_0 being an Artin ring (or a field) and $\lambda(M)$ is the length of M; $\ell(M)$. By proposition 20.7.11 the length is additive. For a concrete example:

1. Let $R = R_0[r_1, ..., r_s]$, where R_0 is an Artinian ring (say a field k) and each r_i is an indeterminate (the simplest case scenario). Then R_n is a free R_0-module generated by monomials $r_1^{m_1}, ..., r_s^{m_s}$ with $\sum m_i = n$. There are

$$\binom{s+n-1}{s-1}$$

such monomials, and hence by proposition 21.6.7

$$P(R, t) = (1-t)^{-s}$$

For high enough degree, the Hilbert polynomial i:

$$\begin{aligned} h_R(t) &= \frac{1}{N!}(x+N)(x+N-1)\cdots(x+1) \\ &= \frac{1}{N!}x^N + \frac{N=1}{2(N-1)!}x^{N_1} + \cdots + 1 \end{aligned}$$

Proposition 21.6.8: Hilbert Function On Local Ring

Let R be a Noetherian local ring, $\mathfrak{m}$ its assocaited maximal ideal, $\mathfrak{q}$ an $\mathfrak{m}$-primary ideal, M a finitely generated R-module, and (M_n) a stable $\mathfrak{q}$-filtration of M. Then:

1. M/M_n is of finite length for each $n \geq 0$
2. for all sufficiently large n, this legnth is a polynomial $g(n)$ of degree $\leq s$ in n, where s is the lesat number of generators of $\mathfrak{q}$
3. The degree and leading coefficient of $g(n)$ depends only on M and $\mathfrak{q}$, not on the chosen filtration

***Proof* :**

1. Take $G(R) = \bigoplus_n \mathfrak{q}^n/\mathfrak{q}^{n+1}$ and $G(M) = \bigoplus_n M_n/M_{n+1}$. Then by corollary 20.7.7 $G_0(A) = A/\mathfrak{q}$ is an Artinian ring. By proposition 20.4.26, $G(R)$ is Noetherian and $G(M)$ is a finitely generated graded $G(R)$-module. As each $G_n(M) = M_N/M_{n+1}$ is a Noetherian R-module annihilated by $\mathfrak{q}$, it is a Noetherian $R/\mathfrak{q}$-module, and hence is of finite length (as $R/\mathfrak{q}$ is Artinian). It follows that M/M_n is of finite length and :

$$\ell_n = \ell(M/M_n) = \sum_{r=1}^{n} \ell(M_{r-1}/M_r)$$

2. Let $x_1, ..., x_s$ generate $\mathfrak{q}$, and take $\overline{x}_i \in \mathfrak{q}/\mathfrak{q}^2$, which generate $G(R)$ as an $R/\mathfrak{q}$-algebra. As each $\overline{x}_i$ has degree 1, we have by corollary 21.6.5

$$\ell(M_n/M_{n+1}) = f(n)$$

where $f(n)$ is a polynomial in n of degree $\leq s-1$ for large n. Then by (1) we have $\ell_{n+1} - \ell_n = f(n)$, we get that ℓ_n is a polynomial $g(n)$ of degree $\leq s$ for large n

3. Let $(\tilde{M}_n)$ be another stable $\mathfrak{q}$-filtration of M and $\tilde{g}(n) = \ell(M/\tilde{M}_n)$. Then by **??**, the two have bounded difference, meaning there exists an integer n_0 such that

$$M_{n+n_0} \subseteq \tilde{M}_n \qquad \text{and} \qquad \tilde{M}_{n+n_0} \subseteq M_N$$

for all $n \geq 0$. Hence:

$$g(n+n_0) \geq \tilde{g}(n) \qquad \text{and} \qquad \tilde{g}(n+n_0) \geq g(n)$$

as $g, \tilde{g}$ are polynomials for all large n, we get:

$$\lim_{n\to\infty} \frac{g(n)}{\tilde{g}(n)} = 1$$

and so g and $\tilde{g}$ have the same degree and leading coefficients, completing the proof.

Definition 21.6.9: Characteristic Polynomial of $\mathfrak{m}$-primary Ideal

Let R be a local ring with maximal ideal $\mathfrak{m}$, $\mathfrak{q}$ a $\mathfrak{m}$-primary ideal, and $g(n)$ be the polynomial corresponding to the filtration $(\mathfrak{q}^n M)$. Then it is denoted by $\chi_{\mathfrak{q}}^m(n)$ and is called the *characteristic polynomial* of the $\mathfrak{m}$-primary ideal $\mathfrak{q}$, and we have:

$$\chi_{\mathfrak{q}}^M(n) = \ell(M/\mathfrak{q}^n M)$$

If $M = R$, then we often abbreviate to $\chi_{\mathfrak{q}}(n) := \chi_{\mathfrak{q}}^R(n)$.

Corollary 21.6.10: Characteristic Polynomial For Local Ring

Let R be a local ring and $\mathfrak{q}$ a $\mathfrak{m}$-primary ideal. Then for all large n, $\chi_{\mathfrak{q}}(n) = \ell(R/\mathfrak{q}^n R)$ is a polynomial of degree $\leq s$, where s is the least number of generators of $\mathfrak{q}$

***Proof* :**
exercise (apply proposition proposition 21.6.8)

Proposition 21.6.11: Degree of Characteristic Polynomial of Local Ring

Let R be a local ring with maximal ideal $\mathfrak{m}$ and $\mathfrak{q}$ a $\mathfrak{m}$-primary ideal. Then:

$$\deg \chi_{\mathfrak{q}}(n) = \deg_{\mathfrak{m}} \chi(n)$$

***Proof* :**
By definition of $\mathfrak{m}$-primary ideals, $\mathfrak{m} \supseteq \mathfrak{q} \supseteq \mathfrak{m}^r$ for some r. Then $\mathfrak{m}^n \supseteq \mathfrak{q}^n \supseteq \mathfrak{m}^{rn}$ and hence:

$$\chi_{\mathfrak{m}}(n) \leq \chi_{\mathfrak{q}}(n) \leq \chi_{\mathfrak{m}}(rn)$$

for all large n. Letting $n \to \infty$ gives the desired result.

We hence have:

$$\deg \chi_{\mathfrak{q}}(n) = d(R) = d(G_{\mathfrak{m}}(R))$$

Let us apply this to projective varieties. Recall from exercise ref:HERE that if $k[X] = S/I(X)$ is the projective coordinate ring with grading:

$$k[X]_n = \frac{S_n + I(X)}{I(X)}$$

with $S_0 = k$ is a field, and hence an Artinian ring which allows us to use the above results on projective varieties. Note too that $\ell_k(k[X]_n) = \dim_k \left(\frac{S_n + I(X)}{I(X)}\right)$ as we can see that $k[X]_n$ is a vector space and hence a (and hence the) composition series can be given via the usual filtration of a vector-space (in this case, each S_n is generated by the homogeneous elements of degree n given from the s generators). This leads to the following:

Definition 21.6.12: Hilbert Function On Projective Variety

Let $X \subseteq \mathbb{P}^n_k$ be a projective variety with corresponding projective coordinate ring $k[X] = S/I(X)$ where $S = k[x_0, ..., x_n]$ with k algebraically closed and $I(X)$ the homogeneous radical ideal. Then the *Hilbert Function*for all $n \in \mathbb{N}$ of X is:

$$H_X(n) = H_{k[X]}(n) = \ell_k(k[X]_n) = \dim_k \left(\frac{S_n + I(X)}{I(X)} \right)$$

for all $n \in \mathbb{N}$. For sufficiently large n, this agrees wit hthe *Hilbert polynomial* of X given by:

$$h_X(t) := h_{k[X]}(t) \in \mathbb{Q}[x]$$

21.6.1 Dimension of Noetherian Local Ring

Dimension is a local property, that is to be extracted from a local ring. In this section let R be a Noetherian local ring, with $\mathfrak{m}$ its maximal ideal. Let

$$\delta(R) = \text{least number of generators of an } \mathfrak{m}\text{-primary ideal of } R$$

We shall show that:

$$\delta(R) = d(R) = \operatorname{kdim} R$$

by showing a sequence of inequalities starting and ending with $\delta(R)$. by corollary 21.6.10 and proposition 21.6.11, we already have:

$$\delta(R) \geq d(R)$$

Let us next prove an analogous version of proposition 21.6.7

Proposition 21.6.13: Degree Diminish by Quotient

Let R be a local ring with maximal ideal $\mathfrak{m}$ and $\mathfrak{q}$ a $\mathfrak{m}$-primary ideal. Let M be a finitely generated R-module, and $x \in R$ a non-zero-divisor in M. Take $M' = m/xM$. THen:

$$\deg \chi_{\mathfrak{q}}^{M'} \leq \deg \chi_{\mathfrak{q}}^{M} - 1$$

Proof **:**
Take $N = xM$ so that $N \cong_R M$ (as x is not a zero divisor). Takign $N_n = N \cap \mathfrak{q}^n M$ we have :

$$0 \to N/N_n \to M/\mathfrak{q}^n M \to M'/\mathfrak{q}^n M' \to 0$$

Hence, given $g(n) = \ell(N/N_n)$ we ahve:

$$g(n) - \chi_{\mathfrak{q}}^{M}(n) + \chi_{\mathfrak{q}}(M'(n) = 0$$

for large n. Then by Artin-Rees lemma, (N_n) is a stable $\mathfrak{q}$-filtration of N. As $N \cong_R M$, by proposition 21.6.8(3) we have that $g(n)$ and $\chi_{\mathfrak{q}}^{m}(n)$ have the same leading terms, completing the proof.

Corollary 21.6.14: Non-zero Divisor and Order of Pole for Local Ring

Let R be a Noetherian local ring, x a non-zero-divisor in R. Then:

$$d(R/xR) \leq d(R) - 1$$

Proof **:**
Take $M = R$ in proposition 21.6.7.

Proposition 21.6.15: Pole and Dimension of Local Ring

$$d(R) \geq \operatorname{kdim}(R)$$

Proof **:**
We proceed by induction. Let $d = d(R)$. IF $d = 0$, then $\ell(R/\mathfrak{m}^n)$ is constant for all large n, and so $\mathfrak{m}^N = \mathfrak{m}^{n+1}$,and so by Nakayama's lemma $\mathfrak{m}^n = 0$. THus, R is an Artinian ring, and hence $\operatorname{kdim}(R) = 0$

Now suppose $d > 0$ and that

$$\mathfrak{p}_0 \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_r$$

is any chain of prime ideals in R. Take $x \in \mathfrak{p}_1$ where $x \notin \mathfrak{p}_0$ and $R' = R/\mathfrak{p}_0$. Then $\overline{x'} \neq \overline{0} \in R'$. As R' is an integral domain, by corollary 21.6.14:

$$d(R'/x'R') \leq d(R') - 1$$

Furthermore, of $\mathfrak{m}'$ is the maximal ideal of R', then $R'/(\mathfrak{m}')^n$ is the image of $R/\mathfrak{m}^n$, and so

$$\ell(R/\mathfrak{m}^n) \geq \ell(R'/(\mathfrak{m}')^n) \qquad \Rightarrow \qquad d(R) \geq d(R') \qquad \Rightarrow \qquad d(R'/x'R') \leq d(R) - 1 = d - 1$$

Thus, by the inductive hypothesis ,the length of the chain of prime ideals in $R'/x'R'$ is less than or equal to $d - 1$ ($\leq d - 1$). As the aimge sof $\mathfrak{p}_1, ..., \mathfrak{p}_r$ in $R'/x'R'$ for m oa chain of length $r - 1$, we have

$$r - 1 \leq d - 1 \qquad \text{iff} \qquad r \leq d \qquad \text{iff} \qquad \operatorname{kdim}(R) \leq d$$

as we sought to show.

Corollary 21.6.16: Finite Dimension of Local Noetherian Ring

Let R be a local Noetherian ring. Then R is finite

Proof **:**
apply the above result.

Corollary 21.6.17: Height and Dimension

$$\operatorname{height}(\mathfrak{p}) = \operatorname{kdim} R_{\mathfrak{p}}$$

Proof **:**
Apply the above, noting that the height is the supremum of the length of the prime ideals;

$$\mathfrak{p}_0 \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_r = \mathfrak{p}$$

Corollary 21.6.18: D.C.C. on Prime Ideals of Noetherian Ring

Let R be a Noetherian ring. Then each prime ideal has finite height, and the set of prime ideals satisfies the descending chain condition.

Proof **:**
exercise.

Proposition 21.6.19: Generation of Local Noetherian Ring and Dimension

Let R be a Noetherian local ring of dimension d. Then there exists an $\mathfrak{m}$-primary ideal in R generated by d elements $r_1, ..., r_d$, hence:

$$\dim(R) \geq \delta(R)$$

Proof **:**
We shall cleverly construct $r_1, ..., r_d$ inductively in such a way so that every prime ideal containing $(r_1, ..., r_i)$ has height $\geq i$ for each i.

Suppose $i > 0$ and $r_1, ..., r_{i-1}$ has been constructed. Let $\mathfrak{p}_j$ with $1 \leq j \leq s$be the minimal prime ideals of $(r_1, ..., r_{i-1})$ (if any such minimal prime ideals exist) which has height exactly $i - 1$. As $i - 1 < d = \operatorname{kdim} R = \operatorname{height}(\mathfrak{m})$, we have

$$\mathfrak{m} \neq \mathfrak{p}_j \qquad (1 \leq j \leq s) \qquad \overset{\text{cor. 21.6.16}}{\Rightarrow} \qquad \mathfrak{m} \neq \bigcup_{j=1}^{s} \mathfrak{p}_j$$

Now, choose $r_i \in \mathfrak{m} \setminus \cup \mathfrak{p}_j$. Let $\mathfrak{q}$ be any prime containing $(r_1, ..., r_i)$. Then $\mathfrak{q}$ contains some minimal prime ideal $\mathfrak{p}$ of $(r_1, ..., r - i - 1)$. If $\mathfrak{p} = \mathfrak{p}_j$ for some j, we get $r_i \in \mathfrak{q}$ and $f_i \notin \mathfrak{p}$, and so

$$\mathfrak{q} \supsetneq \mathfrak{p}$$

implying that the height of $\mathfrak{q}$ is $\geq i$. If $\mathfrak{p} \neq \mathfrak{p}_j$ (for $1 \leq j \leq s$) then $\operatorname{height}(\mathfrak{p}) \geq i$,and so $\operatorname{height}(\mathfrak{q}) \geq i$. Thus, every prime ideal containing $(r_1, ..., r_i)$ has height $\geq i$.

Now, consider $(r_1, ..., r_d)$. If $\mathfrak{p}$ is a prime ideal of this ideal, then $\mathfrak{p}$ has height $\geq d$, and hence it must be that $\mathfrak{p} = \mathfrak{m}$ or else $\mathfrak{p} \subsetneq \mathfrak{m}$ implying $\operatorname{height}(\mathfrak{p}) < \operatorname{height}(\mathfrak{m}) = d$, which is impossible. Hence, the ideal $(r_1, ..., r_s)$ is $\mathfrak{m}$-primary, completing the proof.

Summing up all these results:

Theorem 21.6.20: Dimension Theorem

Le R be a local Noetherian ring. then the following are equivalent:

1. $\operatorname{kdim}(R)$
2. $\deg \chi_{\mathfrak{m}}(n) = \ell(R/\mathfrak{m}^n)$
3. the least number of generator of an $\mathfrak{m}$-primary ideal of R

Corollary 21.6.21: Dimension and Tangent Space

$$\operatorname{kdim}(R) \leq \dim_k(\mathfrak{m}/\mathfrak{m}^2)$$

Proof **:**
Take $r_i \in \mathfrak{m}$ $(1 \leq i \leq s)$ such that their image sin $\mathfrak{m}/\mathfrak{m}^2$ form basis. Then the r_i generate $\mathfrak{m}$, hence:

$$\dim_k(\mathfrak{m}/\mathfrak{m}^2) = s \overset{!}{\geq} \operatorname{kdim}(R)$$

Where the $\overset{!}{\geq}$ inequality comes from proposition 21.6.19.

Corollary 21.6.22: Height of Minimal Ideals belonging to Set

Let R be a Noetherian ring and let $r_1, ..., r_s \in R$. THen every minimal ideal $\mathfrak{p}$ belonging to $(r_1, ..., r_s 0$ has height $\leq s$

Proof **:**
Take $(r_1, ..., r_s 0$ in $R_{\mathfrak{p}}$, which becomes a $\mathfrak{p}^e$-primary ideal. Hence:

$$s \geq \dim R_{\mathfrak{p}} = \operatorname{height}(\mathfrak{p})$$

Corollary 21.6.23: Krull Principal Ideal Theorem

Let R be a Neotherian ring and let $x \in R$ be an elemtn wiht is neither a zero-divisor nor a unit. Then every minimal prime ideal $\mathfrak{p}$ of (x) has height 1

Proof **:**
By corollary 21.6.22, $\operatorname{height}(\mathfrak{p}) \leq 1$. If $\mathfrak{p} = 0$, then $\mathfrak{p}$ is a prime ideal belonging to 0, but then every element of $\mathfrak{p}$ is a zero-divisor, contradicting our assumption.

Corollary 21.6.24: Quotient of Local Noetherian Ring

LEt R be a local Noetherian ring, $x \in \mathfrak{m} \subseteq M$ an element of the maximal ideal which is not a zero-divisor. Then:

$$\operatorname{kdim} R/xR = \operatorname{kdim} R - 1$$

Proof **:**
Let $d = \operatorname{kdim}(R/xR)$. By corollary 21.6.14 and theorem 21.6.20 we get

$$d \leq \operatorname{kdim}(R) - 1$$

On the other hand, if r_i $(1 \leq i \leq d)$ are the element sof $\mathfrak{m}$ whose image in R/xR generage an $(\mathfrak{m}/xR)$-primary ideal, then the ideal $(x, r_1, ..., r_d)$ in R is $\mathfrak{m}$-primary, and so

$$d + 1 \geq \operatorname{kdim}(R)$$

Corollary 21.6.25: Dimension of Completion

Let $\hat{R}$ be the $\mathfrak{m}$-adic completion of the local Noetherian ring R. Then

$$\operatorname{kdim}(R) = \operatorname{kdim}(\hat{R})$$

Proof **:**
As $R/\mathfrak{m}^n \cong \hat{R}/\hat{\mathfrak{m}}^n$, $\chi_{\mathfrak{m}}(n)$ $\chi_{\hat{\mathfrak{m}}}(n)$, and so apply the dimension theorem.

The following will be important for an equivalent definition of regular local ring

Proposition 21.6.26: System of Parameters

Let $r_1, ..., r_d$ be a system of parameters for R and let $\mathfrak{q} = (r_1, ..., r_d)$ be the $\mathfrak{m}$-primary ideal generated by them. Let $f(x_1, ..., x_d)$ be a homogeneous polynomial of degree s with coefficient in R. Assume that :

$$f(r_1, ..., r_d) \in \mathfrak{q}^{s+1}$$

Then all the coefficients of f lie in $\mathfrak{m}$

Proof **:**
Take the surjective homomorphism:

$$\alpha : (R/\mathfrak{q})[x_1, ..., x_d] \to G_{\mathfrak{q}}(R)$$

where $x_i \mapsto \overline{r}_i$. Then by assumption, $f \in \ker(\alpha)$. Then as $\overline{f}$ can is not a zero-divisor, we have:

$$\begin{aligned} d(G_{\mathfrak{q}}(R)) &\leq d((R/\mathfrak{q})[x_1, ..., x_d]/(\overline{f})) \\ &= d((R/\mathfrak{q})[x_1, ..., x_d]) - 1 \\ &= d - 1 \end{aligned}$$

But we know that $d(G_{\mathfrak{q}}(R)) = d$ by the Dimension Theorem, a contradiction.

If R contains a field k mapping isomorphically onto the residue field $R/\mathfrak{m}$, we get the following

Corollary 21.6.27: Residue Field and Algebraic Independence

Let $k \subseteq R$ be a field mapping isomorphically onto $R/\mathfrak{m}$ and let $r_1, ..., r_d$ be a system of parameters. Then $r_1, ..., r_d$ are algebraically independent over k

***Proof* :**
For the sake of contradiction, assume $f(r_1, ..., r_d)0$ where f is a polynomial with coefficients in k. If $f \not\equiv 0$, then $f = f_s + g$ where g has some higher terms and f_s is homogeneous of degree s. Then $f_s \not\equiv 0$. By proposition 21.6.26, f_s has all coeficients in $\mathfrak{m}$. As f_s has coefficients in k we get

$$f_s \equiv 0$$

but we can't both have $f_s \equiv 0$ and $f_s \not\equiv 0$, a contradiction. Hence, no such f can exists, and these algebraically independent.

21.6.2 Smoothness: Regular Local Rings

Theorem 21.6.28: Regular Local Ring Equivalent Conditions

Let R be a local Noetherian ring of dimensin d with maximal ideal $\mathfrak{m}$ and residue field $k = R/\mathfrak{m}$. Then the following are equivalent:

1. $G_{\mathfrak{m}}(R) \cong k[t_1, ..., t_d]$ where the t_i are independent and indeterminate
2. $\dim_k(\mathfrak{m}/\mathfrak{m}^2) = d$
3. $\mathfrak{m}$ can be generated by d elements

***Proof* :**
These are all straight-forward. (1) $\Rightarrow$ (2) and (2) $\Rightarrow$ (3) are immediate (use the above corollaries). For (3) $\Rightarrow$ (1), let $\mathfrak{m} = (r_1, ..., r_d)$. Then by proposition 21.6.26 the map

$$\alpha : k[r_1, ..., r_d] \to G_{\mathfrak{m}}(R)$$

is an isomorphism of graded rings.

Definition 21.6.29: Regular Local Ring

Let R be a local Noetherian ring. THen if R satisfies any of the conditions in theorem 21.6.28, R is called a *regular local ring.*

Lemma 21.6.30: Regular Local Ring Always Integral Domain

Let R be a ring and $\mathfrak{a}$ an ideal of R such that

$$\cap_n \mathfrak{a}^n = 0$$

Suppose that $G_{\mathfrak{a}}(R)$ is an integral domain. Then R is an integral domain

Proof **:**
Let $x, y \in R$ be non-zero elements. As $\cap_n \mathfrak{a}^n = 0$, there exists $r, s \geq 0$ such that

$$\begin{aligned} x &\in \mathfrak{a}^r & x &\notin \mathfrak{a}^{r+1} \\ y &\in \mathfrak{a}^s & y &\notin \mathfrak{a}^{s+1} \end{aligned}$$

Let $\overline{x}, \overline{y}$ be the images of x, y in $G_r(R)$ and $G_s(R)$ respectively. Then $\overline{x} \neq 0$ and $\overline{y} \neq 0$. Hence $\overline{xy} \neq 0$, and so $xy \neq 0$, completing the proof.

Corollary 21.6.31: Regular Local Rings of Dimension 1

Let R be a local Noetherian ring of dimension 1. Then R is a regular local ring if and only if R is a DVR.

Proof **:**
exercise.

This gives a nice condition for 1-dimensional regular local rings. For higher-dimensions, it is unfortunately not so easy.

Corollary 21.6.32: Regular Local Ring, then Normal

Let R be a regular local Noetherian ring. Then R is normal.

Proof **:**
Let R be a local ring. If $G_{\mathfrak{m}}(R)$ is a normal domain, then we can show that R is a normal domain, and hence regular local ring is normal, completing the proof.

The converse does not hold in general in dimension > 1, namely a normal domain need not be regular (see CITE:ALG-GEO)

Proposition 21.6.33: Completion and Regularity

Let R be a Noetherian local ring. Then R is regular iff $\overline{R}$ is regular

Proof **:**
This is a matter of putting together the right results. Show that $\overline{R}$ is a Noetherian local ring of the same dimension as R with maximal ideal $\overline{\mathfrak{m}}$. Then assert that $G_{\mathfrak{m}}(R) \cong G_{\overline{\mathfrak{m}}}(\overline{R})$ to complete the proof.

In this case, it must be that $\overline{R}$ is also an integral domain. For the geometric interpretation, see CITE:ALG-GEO.

(Atiyah still have a section on transcendental dimension, if you so wish to add it)

Exercise 21.6.1

1. (Compare Gelfand-Kirillov Dimension to Krull dimension in Commutative case; the former can be defined more generally an can give fractional dimensions)
2. Define the *depth* of a prime $\mathfrak{p}$ to be the chains of ideals which start at $\mathfrak{p}$. Show that $\operatorname{depth}(\mathfrak{p}) = \dim R/\mathfrak{p}$ if R is Noetherian, and give an example where this is infinite. Show this is always finite if R is a local ring.
3. Let R be a Noetherian Ring. Show that $\operatorname{kdim} R[x] = \operatorname{kdim} R + 1$

21.7 *Prelude to Algebraic Geometry

We have now created a dictionary translating properties of algebraic sets to coordinate rings. In particular, we have:

$$\left\{\text{Affine Varieties}\right\} \underset{\varphi}{\overset{V(S)}{\rightleftarrows}} \left\{\begin{array}{c}\text{Finitely Generated,}\\ \text{Reduced rings}\\ \text{over algebraically closed fields}\end{array}\right\} \tag{21.10}$$

This allows for a certain subset of commutative rings to be interpreted as geometric objects, namely zero sets of polynomials. The question arises then if this duality can be extended to general commutative rings? In particular, does there exist some topological space such that:

$$\left\{???\right\} \underset{V(S)}{\overset{R[x]/V}{\rightleftarrows}} \left\{\text{Commutative Ring}\right\}$$

Note the necessity of commutativity, since many of the theorem on prime ideals such as ref:HERE, ref:HERE, and ref:HERE rely on commutativity[26].

In this section, we build-up the object that should should be put on the other side, known as an *affine scheme*. An affine scheme will be defined in such a way that if we restricted to finitely generated reduced rings over an algebraically closed field, the affine scheme be an affine variety, making the following diagram the correct generalization of equation (21.10)

[26]The theory of "non-commutative geometry" using non-commutative rings can be explored in *Spectral Theory* in a class on functional analysis (see EYTNKA functional analysis for more details)

$$\left\{\text{Affine Scheme}\right\} \underset{V(S)}{\overset{R[x]/V}{\rightleftarrows}} \left\{\text{Commutative Ring}\right\} \tag{21.11}$$

Why do this? This abstraction is not simply for abstractions-sake. It is because the dictionary between algebra and geometry that has been built up to is incredibly useful, and will fix some of the problems of algebraic varieties, for example:

1. There are many algebras over more general rings that have geometric interpretations. For a rather trivial but illustrative example, the coordinate ring of a line whose points are in some arbitrary ring k (the coordinate ring being $k[x]$) currently does not have a good representation since the Nullstellensatz limits to Algebraically closed fields. Another example that may at first seem un-geometric but turns out to have many strong geometric properties are rings such as $\mathbb{Z}$, $\mathbb{Z}[i]$ (the Gaussian integers), and more generally *algebraic number rings* (explored in the chapter 22). Mathematicians have found that these corresponds naturally to curves in a plane (as we shall see in chapter 22, section 22.7).

2. As we saw in section 21.5, localization gives us powerful tools to explore the behavior of a variety at a point. For example, if we take the coordinate ring of an affine line, $k[x]$, takes the closure $k(x)$, or study $k[x]_{(x)}$, we love the fact that we are finitely generated. However, as we saw these certainly still have interpretations as algebraic correspondences to geometric objects.

3. Finally, varieties loose multiplicity information. Saw we have the variety over $\mathbb{C}$ defined by the equations
$$y = x^2 \qquad y = 0$$
then if we take their intersection, we get a point $(0, 0)$. The corresponding affine variety is given by taking the union of quotient of the two coordinate rings, namely:
$$\frac{k[x, y]}{(y, x^2 - y)} = \frac{k[x]}{(x^2)}$$
Usually, we would say omit the x^2 term and just use $k[x]/(x)$ for the coordinate ring, however this eliminate the multiplicity information, namely that the variety $y = x^2$ "doubly intersects" the variety $y = 0$ the variety $y = 0$. The ring $k[x]/(x^2)$ keeps track that we have a "double-point", while $k[x]/(x)$ does not. This is a achieved by the fact that the ring has nilpotent elements.

In this section, we shall build up the definition of an affine scheme. We shall then for the first time properly venture into the realm of geometry and define a new object known as an *scheme* that shall will mirror the properties of a (smooth) manifold. (TBD)

After having presented schemes, we shall give a few notable examples, but the full exploration of scheme theory and algebraic geometry deserves a book on its own, and so they shall be relegated to future readings after that point. There is still a point in introducing schemes since the concept shall come a few times times later in this book (notably in number theory).

Topology of Affine Schemes

(Mention also how Schemes take away the need to have an ambient affine or projective space, that is where the)

A natural first place to start is by interpreting the set of maximal ideals of R as the "points" on which we can define a topology (analogous to the Zariski topology). If this were the case, we would want that the functor translating a ring homomorphism $\varphi : R \to S$ has an associated continuous function mapping the maximal ideal $M \subseteq S$ to a maximal ideal$\varphi^{-1}(M)$ (recall that regular function corresponds contra-variantly to k-algebra homomorhpisms). The problem is that the pre-image of maximal ideals need not be maximal (ex. take $\mathbb{Z} \hookrightarrow \mathbb{Q}$. Then (0) is maximal in $\mathbb{Q}$, but certainly not maximal in $\mathbb{Z}$). Fortunately, if we loosen our requirement a bit and consider prime ideals, then it is indeed the case that $\varphi^{-1}(\mathfrak{p})$ is prime. This inspires the following definition:

Definition 21.7.1: Spectrum Of A Ring

Let R be a commuttaive ring. Then the *spectrum* or *prime spectrum* of the ring R is the collection of all prime ideals of R, and is denoted $\mathrm{Spec}(R)$:

$$\mathrm{Spec}(R) := \{\mathfrak{p} \subseteq R \mid \mathfrak{p} \text{ is a prime ideal of } R\}$$

Similarly, the collection of all maximal ideals is called the *maximal spectrum* of R and is denoted $\mathrm{mSpec}(R)$.

Example 21.7.1: Spec Of Rings

1. For any field k, $\mathrm{Spec}(k) = \mathrm{mSpec}(k) = \{(0)\}$
2. Let $R = \mathbb{Z}$. Then $\mathrm{Spec}(\mathbb{Z}) = \{(p) \mid p \text{ is prime}\}$ and $\mathrm{mSpec}(\mathbb{Z}) = \mathrm{Spec}(\mathbb{Z}) \setminus \{(0)\}$
3. $\mathrm{Spec}(\mathbb{Z}[x])$ consists of (really cool explanation by Richard Borcherd at the end of his video on spectrums of a ring!!)

Proposition 21.7.2: $\mathrm{Spec}(-)$ Is Functorial

Let R, S be rings and $\varphi : R \to S$ a ring homomorphis. Then $\mathrm{Spec}(-)$ induces

$$\varphi^* : \mathrm{Spec}(S) \to \mathrm{Spec}(R) \qquad p \mapsto \varphi^{-1}(p)$$

Making it a contravariant functor $\mathrm{Spec}(-) : \mathbf{Ring} \to \mathbf{Set}$

Soon, we shall see we may induce a natural topology on $\mathrm{Spec}(R)$ for any ring R, making it a functor to **Top**

Proof **:**
The key is to recall (or re-prove) that if p is a prime ideal, then $\varphi^{-1}(p)$ is a prime ideal.

Next, when we defined the algebraic sets, they where the zero sets of some collection of functions. When the ring we worked over was $k[x]$, we already had a good notion of function that was readily available to us, letting us define a map from $k[x]$ to $\mathrm{Hom}_{\mathrm{Set}}(V, k)$ where:

$$f \mapsto (p \mapsto f(p))$$

Howveer, an arbitrary element of a ring no longer has notion of being "evaluatable". We thus must come up with an analogous idea of "evalutaing an element at a point" for an arbitrary ring. As a start, we already have that our domain of or function would be $\operatorname{Spec}(R)$ instead of $\mathbb{A}^n$: Any algebraic set (i.e. closed set of $\mathbb{A}^n$) is the union algebraic verities, and so we would want that any closed set of $\operatorname{Spec}(R)$ (once we define a suitable topology) also be the union of prime ideals. But how would an element $x \in R$ then be "evaluated" at a "point" $\mathfrak{p} \in \operatorname{Spec}(R)$?

We generlize this to the following: let's first consider when $f \in k[x]$. If $\operatorname{Spec}(k[x])$ is the spectrum of $k[x]$, then the *value* of f at $p \in \operatorname{Spec}(R)$ is the element of k representing f in the quotient $K[V]/\mathcal{I}(p) \cong k$. Thus, the "value of f" in $k[V]$ at $p \in V$ can be viewed as the element $\overline{f} \in k[V]/\mathcal{I}(p) \cong k$. Using this intuition, we can generlize the notion of evaluating a funciton:

$$R \to \operatorname{Hom}_{\mathbf{Set}}\left(\operatorname{Spec}(R), \bigsqcup_{\mathfrak{p} \in \operatorname{Spec}(R)} R/\mathfrak{p}\right) \qquad f \mapsto (\mathfrak{p} \to f \bmod \mathfrak{p})$$

Notice how this generalizes equation (21.1) seen at the beginning of section 21.1.

Definition 21.7.3: Value Of Element

Let $f \in R$. then the "value" of f at the "point" $\mathfrak{p} \in \operatorname{Spec}(R)$ is the element:

$$f(\mathfrak{p}) := \overline{f} \in R/\mathfrak{p}$$

Next, we see that we can define the maps $\mathcal{Z}$ and $\mathcal{I}$ as follows. For $\mathcal{Z}$, we get:

$$\mathcal{Z}(A) = \{\mathfrak{p} \in \operatorname{Spec}(R) \mid A \subseteq \mathfrak{p}\} \subseteq \operatorname{Spec}(R)$$

That is, $\mathcal{Z}(A)$ is the colection of all prime ideals containing A. Clearly, $\mathcal{Z}(A) = \mathcal{Z}(I)$ where $I = (A)$, so we only need to consider ideals. To see that this relates to the notion of a zero-locus as we've seen with polynomial rings, notice that if $p \in \mathcal{Z}(I)$, then $I \subseteq \mathfrak{p}$, so $f \in \mathfrak{p}$ for every $f \in I$. If we view $f \in R$ as a "function" on $\operatorname{Spec}(R)$, we see that $\mathfrak{p} \in \mathcal{Z}(I)$ if and only if

$$f(\mathfrak{p}) = f \bmod \mathfrak{p} = 0 \in R/\mathfrak{p}$$

for all $f \in I$, showing that, in a sense, $\mathcal{Z}(I)$ consists of points in $\operatorname{Spec}(R)$ at which all the "functions" of I are zero.

For the other direction, the analogous concept would be:

$$\mathcal{I}(Y) = \bigcap_{\mathfrak{p} \in Y} \mathfrak{p}$$

Proposition 21.7.4: Properties Of Correspondent Maps on Spectrum

Let R be a ring. Then:

1. For any ideal $I \subseteq R$, $\mathcal{Z}(I) = \mathcal{Z}(\operatorname{rad}(I)) = \mathcal{Z}(\mathcal{I}(\mathcal{Z}(I)))$ and $\mathcal{I}(\mathcal{Z}(I)) = \operatorname{rad}(I)$
2. For any ideal $I, J \subseteq R$ $\mathcal{Z}(I \cap J) = \mathcal{Z}(IJ) = \mathcal{Z}(I) \cup \mathcal{Z}(J)$
3. If $\{I_j\}$ is an arbitrary collectio of prime ideals of R, then $\mathcal{Z}(\cup_j I) = \cap_j \mathcal{Z}(I_j)$

Proof **:**
Dummit p. 732

1. If P is a prime ideal containing I, then $\mathfrak{p}$ contains $\text{rad}(I)$, hence $\mathcal{Z}(I) = \mathcal{Z}(\text{rad}(I))$. Since $\text{rad}(I)$ is the inersection of all prime ideals containing I, the definition of $\mathcal{I}(I)$ gives $\mathcal{Z}(\text{rad}(I)) = \mathcal{Z}(\mathcal{I}(I))$. Similarly:

$$\mathcal{I}(\mathcal{Z}(I)) = \mathcal{I} \bigcap_{\mathfrak{p} \in \mathcal{Z}(I)} \mathfrak{p} = \bigcap_{I \subseteq \mathfrak{p}} \mathfrak{p} = \text{rad}(I)$$

2. essentially immediate

3. essentially immediate

This shows that we are back in the condition's for the Nullstellensatz: every $\mathcal{Z}(I)$ in $\text{Spec}(R)$ occurs as a radical ideal, which is unique since $\mathcal{I}(\mathcal{Z}(I)) = \text{rad}(I)$. Furthermore, the 2 other statements shows that we have also kept the same topology on $\text{Spec}(R)$ as we had for polynomial rings, in particular, we can define a topology on

$$\mathcal{T} = \{\mathcal{Z}(I) \mid I \text{ is an ideal of } R\}$$

Definition 21.7.5: Zariski Topology

Let R be a ring and $\text{Spec}(R)$ the corresponding spectrum. Then the collection of $\mathcal{Z}(I)$ defines the closed sets of the *zariski topology* on $\text{Spec}(R)$.

Definition 21.7.6: Closed Points

Let $m \subseteq R$ be a maximal ideal of R and $\text{Spec}(R)$ be the specturm of R. Then the maximal ideals of R in $\text{Spec}(R)$ are called the *closed points* of R.

It is easy to see that maximal ideals are closed, since the zariski closure of them is the set of all prime ideals containing the maximal ideal, which since the ideal is maximal can only be itself, and hence is closed. Thus, the Zariski topology on $\text{Spec}(R)$ is still T_1, and again rarely Hausdroff.

Definition 21.7.7: Basic Open Set

Let R be a ring and $f \in R$. Then $D(f)$ or X_f denotes the collection of all prime ideals in $\text{Spec}(R)$ that do not contain f, i.e. $D(f)$ is the set of points in $\text{Spec}(R)$ for which the value of $f \in R$ is nonzero.

I would suggest you take a moment to go back and see that the properties of the Zariski topology we have shown for $k[x]$ generalize well for the general Zariski topology on $\text{Spec}(R)$ (including irreducibility). One notable one to point out is the generalization of the Nullstellensatz:

Proposition 21.7.8: Nullstellensatz On Spectrum

Let R be a ring and $\mathrm{Spec}(R)$ the spectrum with the Zariski topology. Then the maps $\mathcal{Z}$ and $\mathcal{I}$ define order-reversing bijections:

$$\left\{\begin{matrix}\text{Zariski closed subset of} \\ \mathrm{Spec}(R)\end{matrix}\right\} \underset{\mathcal{Z}}{\overset{\mathcal{I}}{\rightleftarrows}} \left\{\text{Radical Ideals of } R\right\}$$

Proof **:**
Use proposition 21.7.4.

Example 21.7.2: Zariski Topology

1. on $R = \mathbb{Z}$
2. on $R = \mathbb{Z}[x]$
3. Let $R = k[V] = k[x_1, ..., x_n]/I(V)$ be a coordinate ring. Then we have set up the generlization so that all the same properties work for $\mathrm{Spec}(R)$.

The 3rd example given can be generlized:

Definition 21.7.9: Affine k-Algebra

Let R be an algebra over a field k. Then if R is finitely-generated, k is algebarically closed, R has no nilpotent elements, then it is called an *affine k-algebra*

The lack of nilpotence is important for the last part of the Nullstellensatz to correspond exactly. (Words on showing that affine k-algebra are exactly the ring arising as the ring of functions on affine algebraic sets over algebraically cosed fields)

(the corersponding k-algebra homomorphism maps maximal ideals to maximal ideals. In genearl, this is not the case, which is why we have $\mathrm{Spec}(R)$ have prime ideals. The following generlizes the notion of regular function)

Proposition 21.7.10: Spectrum Map

Let $\varphi : R \to S$ be a ring homomorphism. Then there is an indcued map $\varphi^* : \mathrm{Spec}(S) \to \mathrm{Spec}(R)$ mapping

$$\varphi^*(P) = \varphi^{-1}(P)$$

This map is continuous with respect to to the Zariski Topology

Proof **:**
exercise

(many examples)

Proposition 21.7.11: Properties Of Zariski Topology

Let $f \in R$, $X = \text{Spec}(R)$, and X_f be the principal open set. Then

1. $X_f = X$ if and only if f is a unit, and $X_f = \emptyset$ if and only if f is nilpotent
2. $X_f \cap X_g = X_{fg}$
3. $X_f \subseteq X_{g_1} \cup \cdots \cup X_{g_n}$ if and only if $f \in \text{rad}(g_1, ..., g_n)$. In particular, $X_f = X_g$ if and only if $\text{rad}(f) = \text{rad}(g)$
4. The principal open sets from a basis for he Zariski topology on X
5. The natural map from R to R_f induces a homeomorphism from $\text{Spec}(R_f)$ to X_f where R_f is the localization of R at f
6. X is Compact
7. If $\varphi : R \to S$ is a ring homomorphism, with corresponding $\varphi^* : \text{Spec}(S) \to \text{Spec}(R)$, then letting $Y = \text{Spec}(S)$, the pre-image of X_f is $Y_{\varphi(f)}$ in Y

Proof **:**
should be done

(now define localization for zariski topology. Build up intuition)

Definition 21.7.12: Presheaf

Let $X = \text{Spec}(R)$ and $U \subseteq X$ be an open subset. If $U = \emptyset$, define $\mathcal{O}(U) = 0$. Otherwise, define $\mathcal{O}(U)$ to be the set of functions $s : U \to \bigsqcup_{Q \in U} R_Q$ from U to the disjoint union of the localizations R_Q for $Q \in U$ with the following two properties:

1. $s(Q) \in R_Q$ for every $Q \in U$
2. For every $P \in U$, there exists an open neighbourhood $X_f \subseteq U$ containing P and an element a/f^n in the localization R_f defining s on X_f, i.e. $s(Q) = a/f^n \in R_Q$ for every $Q \in X_f$.

notice that for $s, t \in \mathcal{O}(U)$, $s + t, st \in \mathcal{O}(U)$, showing that it is a ring.

Definition 21.7.13: Sheaf

Let R be a commutative ring, $X = \text{Spec}(R)$.

1. The collection of rings $\mathcal{O}(U)$ for the collection of Zariski open sets of X together with the restriction maps $\mathcal{O}(U) \to \mathcal{O}(U')$ for $U' \subseteq U$ is called the *[structure] sheaf* on X, and is denoted $\mathcal{O}$ (or $\mathcal{O}_X$)
2. The elements s of $\mathcal{O}(U)$ are called *sections* of $\mathcal{O}$ over U. The elements of $\mathcal{O}(X)$ are called *global sections* of $\mathcal{O}$

(a proposition on gobal section's of sheafs)

Definition 21.7.14: Stalk

Let $(\mathrm{Spec}(R), \mathcal{O})$ be a structure sheaf on $\mathrm{Spec}(R)$. Then a stalk at $p \in \mathrm{Spec}(R)$ is

$$\mathcal{O}_p = \varinjlim F(\mathcal{O})$$

Definition 21.7.15: Affine Scheme

Let R be a commutative ring. Then the pair $(\mathrm{Spec}(R), \mathcal{O}_{\mathrm{Spec}(R)})$ is called an *affine scheme*.

(define scheme morphism, and show every ring homomorphism induce scheme morphisms)

(a word on how a *scheme* is a locally ringed space in which each point is in a neighbourhood which is isomorphic to an affine scheme)

(Also, incorporate this: it will make the definition of a morphism seem more topological)

> map $\varphi^* : W^* \to V^*$ that pre-composes a linear functional $W \to \mathbb{k}$ with φ to obtain a linear functional $V \xrightarrow{\phi} W \xrightarrow{\omega} \mathbb{k}$.
>
> (iii) The functor $\mathcal{O}\colon \mathsf{Top}^{\mathrm{op}} \to \mathsf{Poset}$ that carries a space X to its poset $\mathcal{O}(X)$ of open subsets is contravariant on the category of spaces: a continuous map $f\colon X \to Y$ gives rise to a function $f^{-1}\colon \mathcal{O}(Y) \to \mathcal{O}(X)$ that carries an open subset $U \subset Y$ to its preimage $f^{-1}(U)$, which is open in X; this is the definition of **continuity**. A similar functor $\mathcal{C}\colon \mathsf{Top}^{\mathrm{op}} \to \mathsf{Poset}$ carries a space to its poset of closed subsets.
>
> (iv) There is a contravariant functor $\mathrm{Spec}\colon \mathsf{CRing}^{\mathrm{op}} \to \mathsf{Top}$ that sends a commutative ring R to its set $\mathrm{Spec}(R)$ of prime ideals given the **Zariski topology**. The closed subsets in the Zariski topology are those subsets $V_I \subset \mathrm{Spec}(R)$ of prime ideals containing a fixed ideal $I \subset R$. This construction is contravariantly functorial: for any ring homomorphism $\phi\colon R \to S$ and prime ideal $\mathfrak{p} \subset S$, the inverse image $\phi^{-1}(\mathfrak{p}) \subset R$ defines a prime ideal of R, and the inverse image function $\phi^{-1}\colon \mathrm{Spec}(S) \to \mathrm{Spec}(R)$ is continuous with respect to the Zariski topology.
>
> (v) For a generic small category C, a functor $\mathsf{C}^{\mathrm{op}} \to \mathsf{Set}$ is called a (set-valued) **presheaf** on C. A typical example is the functor $\mathcal{O}(X)^{\mathrm{op}} \to \mathsf{Set}$ whose domain is the poset $\mathcal{O}(X)$ of open subsets of a topological space X and whose value at $U \subset X$ is the set of continuous real-valued functions on U. The action on morphisms is by restriction. This presheaf is a *sheaf*, if it satisfies an axiom that is introduced in Definition 3.3.4.
>
> (vi) Presheaves on the category $\mathbb{\Delta}$, of finite non-empty ordinals and order-preserving

(Given a topological space X, we can define an algebra by taking the set of all continuous functions $C(X)$ with the operations being pointwise. Now, given just the information $C(X)$, does there exist a X such that $C(X)$ is isomorphic to A? In other words, can we recover X? The answer is yes if X is a locally compact topological space! This is the algebra/topology duality. The same is true if X is a Riemann manifold, though you need more information, something called the *spectral triple*. Later, Grothendick generlized this and we got to create form any algebra A a topological space $\mathrm{Spec}(A)$!!! Given the appropriate topological properties on $\mathrm{Spec}(A)$, we can recover A!! In other words, there is this way to go back and forth between algebra and geometry. This same exists in analysis too: the dual of $C_c(X)$ for a LCH X is intimately related to Radon measure, and the dual of $C_c(X)$ with the appropriate conditions let's us generlize the notion of derivatives to locally integrable functions!!!)

22

Relating Algebra to Number Theory

A central object of study in mathematics is, naturally, the natural numbers $\mathbb{N}$ and, if the negatives are included, the integers $\mathbb{Z}$. So far, it is arguable to say that these objects have played a relatively minor algebraic role in this book: we know $\mathbb{N}$ (resp. $\mathbb{Z}$) is the initial abelian monoid (resp. ring) in their respective categories, that they are free objects, and that from $\mathbb{N}$ we may construct $\mathbb{Z}$, then $\mathbb{Q}$, and from there is any other field of characteristic 0. One aspect of $\mathbb{N}$ and $\mathbb{Z}$ we have barely touched on are their *primes* (recall the Fundamental Theorem of Arithmetic's, corollary 10.1.24). Exercise ref:HERE showed that there are infinitely many primes, and since $\mathbb{Z}$ is a PID we have that being irreducible is equivalent to being prime. However, we have explored little the nature of primes, and more generally have explored little of functions with integer solutions (elements of $f \in \mathbb{Z}[x_1, ..., x_n]$ are called *Diophantine equations*).

Unlike fields, it is in general much harder to ascertain whether an algebraic equation has solutions over the integers (or even integral extensions of the integers). Perhaps the most famous number theory problem demonstrating this difficulty is the following:

Show that for $n \geq 3$, the equation $x^n + y^n = z^n$ has no integer solutions

This problem is known as *Fermat's Last Theorem.* This is trivial if $n = 1$ (ex. $1 + 2 = 3$). It is more interesting to see the case of $n = 2$, that is we want to find all $(a, b, c) \in \mathbb{Z}^3$ ($\gcd(a, b, c) = 1$) such that $x^2 + y^2 = z^2$. Such a triple is called a *Pythagorean triple* and there are infinitely many. However, for $n \geq 3$, it is incredibly more difficult to find whether solutions exist or not, partly due to the fact that many rings that naturally arise in solving this case are *not* unique factorization domains, and partly because there turn out to be some "bad number" that make many of the number theory strategies fail. This particular problem has spurred a lot of the development of modern algebraic and analytic number theory, and shall be touched on later due to some interesting general theory

that came from it. In this section we shall cover the mathematics that build built-up in solving Diophantine equations, including some special cases of Fermat's last theorem, and prepare the reader for Class Field theory which shall definitively answer the question of factoring integer-polynomials mod primes[1].

One final intuition that shall become more evident is that the theory of prime factorization in field extensions tie into solving irreducible polynomials mod some prime ideals. This theory extends results from the classical results of solutions or lack thereof for an irreducible polynomial over $\mathbb{Q}$ when considering it mod p for all primes $p \in \mathbb{Z}$. Recall that factorization and irreducibility of monic polynomials over $\mathbb{Q}$, more generally a finite extension $K/\mathbb{Q}$ is exactly reflected to the factorization and irreducibility of monic polynomials over $\mathbb{Z}$ (which will have a ring $\mathcal{O}_K$ that shall be the equivalent[2]) by Gauss's lemma (Proposition 11.2.1) . This was the historical origin of number theory, as it is much easier to work with polynomials finite fields (recall proposition 11.3.6 and example 11.3.3).

Material Overview

In this chapter, we shall start by exploring when do $\mathbb{Z}$-polynomials solutions over a field k. The space of solutions shall naturally be a subset of the field k. It shall be labeled as $\mathcal{O}_K$ and called *rings of integers*, and will have many important properties to determine. Similarly to how we extended fields to find solutions to polynomials over fields, rings of integers shall be extended (in parallel to a field extension) to find more solutions. Some key functionals such as the trace, norm, and discriminant shall become important tools to find properties of elements of $\mathcal{O}_K$ and the structure of $\mathcal{O}_K$ itself. As we classify $\mathcal{O}_K$, we shall see that we may consider $\mathcal{O}_K$ as a *lattice* when embedded in $\mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2}$ for appropriate $r_1, r_1 \in \mathbb{N}$. This geometric interpretation of $\mathcal{O}_K$ shall add some important new perspectives onto the elements of $\mathcal{O}_K$[3], and will allow us to define some important inequalities.

After having done a sweep over the structure of $\mathcal{O}_K$, we shall tackle some of it's algebraic structure. Namely, we shall see that they form a new type of ring called a *Dedekind domain*. Dedekind domains are very important types of rings; for those who are keen on algebraic geometry, Dedekind domains are the algebraic counter-part to 1-dimensional smooth curves (or schemes). In fact, we have seen Dedekind domains before in the previous chapter (for example, $\mathbb{Q}[x, y]/(x - y^2)$). Hence, Dedekind domains will be studied in their own right. They shall have a very rich structure: for example, many $\mathcal{O}_K$ will *not* be unique factorization domain (a fact that led to many wrong proofs of Fermat's last theorem), however we shall be able to define an abelian group structure on the ideals of $\mathcal{O}_K$ which *will* have a unique factorization (in the case where $\mathcal{O}_K$ is a PID, we shall see the ideal group I_K is isomorphic to $\mathbb{Q}^\times$, exactly mirroring the structure of $\mathrm{Frac}(\mathbb{N})$, however in general it may be much more complicated[4]). The discussion on Dedekind domains shall finish with a discussion of the unit's in Dedekind rings, which shall be important for finding the factorization of ideals.

After we realise we ought to be focusing on the *prime ideals* of $\mathcal{O}_K$ rather than the irreducible elements of $\mathcal{O}_K$ (for they are not necessarily prime), we shall start exploring how prime may further decompose as we extend to larger rings of integers. This helps in finding solutions to $\mathbb{Z}$-polynomials for the prime ideals in larger rings will have very rigid structure that can be taken advantage of. We have already seen this before: It shall turn out that $\mathcal{O}_{\mathbb{Q}(i)} = \mathbb{Z}[i]$ (the Gaussian integers). In the

[1] In particular, the *Artin Reciprocity Law* shall give the over-arching result showing how to factorize Diophantine equations

[2] See definition 22.1.2

[3] and of the norm

[4] which is measured by the *class group*, as we shall see

gaussian integers, we saw that 2 is not prime, for

$$2 = (1+i)(1-i) = 1^2 - i^2 = 1 + 1$$

or in terms of ideals, we shall have $(2) = (1+i)(1-i)$. Hence, 2 has further factored into new primes. In section 10.3, this was used to prove Fermat's Theorem for sums of squares. More generally, we shall see that any prime ideal $\mathfrak{p} \subseteq \mathcal{O}_K$ may factor:

$$\mathfrak{p}\mathcal{O}_L = \mathfrak{B}_1^{e_1} \cdots \mathfrak{B}_k^{e_k} \tag{22.1}$$

Though it may currently seem unrelated, the study of the factorization of primes *almost* mirrors that of the factorization of irreducible polynomials mod $\mathfrak{p}$ after extending fields. In fact, it shall turn out to be the case that *all but finitely many primes* shall factor exactly like how a polynomial factors, the "bad primes" essentially being primes for which an exponent in equation (22.1) is greater than 1, ≥ 2. This sort of exploration leads to a lot of very interesting theory that have applications beyond classical number theory, for example we shall be able to fully classify all abelian extensions of $\mathbb{Q}$.

For those who get motivation through algebraic geometry, algebraic number theory naturally ties into it; it may help to think of the main object of study of this chapter (rings of integers) as further expanding the dictionary between commutative algebra and geometry, namely they shall be important for *normal schemes* which links to the notion of a smooth curve[5]

At the end of this chapter, we shall briefly touch on some analytic number theory, for it's study is central to modern number theory and the author of this books considers that anyone that has reached this stage of the book would find it enlightening to see how the Riemann-zeta function play an important role in number theory.

57: Warning

Some elementary analysis will be assumed in this section, including some properties of the real numbers.

22.1 Ring of Integers

In this chapter, whenever the notation $L/K/\mathbb{Q}$ is used for some letter's L, K, it is implies that L/K are both finite extensions of $\mathbb{Q}$, and L is an extension of K. Unless otherwise specified, the characteristic of any field is implicitly assumed to be 0. We start by defining our terms.

Definition 22.1.1: Algebraic Number Fields

Let K be a field. If $K/\mathbb{Q}$ is a finite extension of $\mathbb{Q}$, then K is called an *algebraic number field* or *number field* if unambiguous.

Example 22.1.1: Number Fields

1. Let $\zeta_n = e^{2\pi/n}$ be a root of unity. Then $\mathbb{Q}(\zeta_n)$ is a number field, called the *cyclotomic field*.
2. Let $m \in \mathbb{Z}$ be not-a-square. Then Fields of the form $\mathbb{Q}(\sqrt{m})$ is called a *quadratic field*. Such

[5] See a textbook in algebraic geometry for further details

fields always have degree 2. notice that we can consider any $m \in \mathbb{Z}$, however we get many equality of fields, for example $\mathbb{Q}(\sqrt{12}) = \mathbb{Q}(\sqrt{3})$. The field $\mathbb{Q}(i)$ is both a quadratic field and a cyclotomic field since $i = \zeta_4$. As an exercise, show that $\mathbb{Q}(\sqrt{-3})$ is a cyclotomic field.

3. The field $\mathbb{Q}(\sqrt{n}, \sqrt{m})$ is called a *biquadratic* field where n, m are both not-a-square.

4. The field $\mathbb{R}$ and $\mathbb{C}$ are *not* algebraic number fields, are they are not alegbraic extensions of $\mathbb{Q}$ (they both contain transcendental elements).

5. Similarly, $k(x)$ for any field k is not an algebraic number field. However, in many ways this field is very similar to algebraic number fields, and shall be given some more attention in section 22.2.

Of key interest to this chapter the study of integer polynomials with solutions fields:

Definition 22.1.2: Ring Of Integers

Let k be a field. Then the *ring of integers* of k is the integral closure of $\mathbb{Z}$ over k:

$$\mathcal{O}_k := \overline{\mathbb{Z}}^k$$

Since we will often work over $\mathbb{Z}$, it is useful to know that instead of taking all roots of polynomials, it is equivalent to take all the roots of *irreducible polynomials*; by using Gauss's lemma, if $f = gh$, $f \in \mathbb{Z}[x]$ and $g, h \in \mathbb{Q}[x]$, then in fact $g, h \in \mathbb{Z}[x]$. Hence, the smallest degree polynomial for which α is a root will be irreducible. An important ring of integer that is worth isolating is the following:

Definition 22.1.3: Algebraic Integers

The collection of elements

$$\mathcal{O}_{\overline{\mathbb{Q}}} = \mathcal{O}_{\mathbb{C}} := \overline{\mathbb{Z}}^{\mathbb{C}}$$

are called the *algebraic integers*, and is sometimes denoted $\mathbb{A}$

Just like for fields, there may be multiple "integral closures", and they are all isomorphic which motivates saying they are *the* algebraic integers.

Lemma 22.1.4: Equivalence Definition

Let $K/\mathbb{Q}$ be a finite extension and let $\mathcal{O}_{\mathbb{C}}$ be the algebraic integers. Then:

$$\mathcal{O}_{\mathbb{C}} \cap K = \mathcal{O}_K$$

Proof **:**
exercise.

Example 22.1.2: Ring of Integers

1. $\mathcal{O}_{\mathbb{Q}} = \mathbb{Z}$. For this reason, the integers in algebraic number theory are sometimes called the rational integers

2. $\mathcal{O}_{\mathbb{Q}(i)} = \mathbb{Z}[i]$. To see this, let $\alpha = a + bi$, $a, b \in \mathbb{Q}$. Then for $\alpha \in \mathcal{O}_{\mathbb{Q}(i)}$, we need to show that the minimal polynomial of α:

$$\alpha^2 - 2a\alpha + (a^2 + b^2) = 0$$

has integer coefficients, that is $-2a$ and $a^2 + b^2$ are integers. Since $\mathcal{O}_{\mathbb{Q}(i)}$ is a ring, if $a + bi \in \mathcal{O}_{\mathbb{Q}(i)}$, then so is $a + bi - c - di = a' + b'i$ for $c, d \in \mathbb{Z}$ where $a', b' \leq 1$. But then from $(a')^2 + (a')^2 = k \in \mathbb{Z}$, we get $a', b' \in \{-1, 0, 1\}$. Then that implies $a, b \in \mathbb{Z}$, showing that $\mathcal{O}_{\mathbb{Q}(i)} = \mathbb{Z}[i]$.

3. In general, it is difficult to find the ring of integers. The difficulty of finding the ring of integers can even show up for quadratic extensions. Consider $F = \mathbb{Q}(\sqrt{-3})$. Then we might be tempted to say that $\mathcal{O}_F = \mathbb{Z}[\sqrt{-3}]$. However, this is not the case, take:

$$\omega = \frac{-1 + \sqrt{-3}}{2} \in \mathbb{Q}(\sqrt{-3})$$

Then $\omega^2 + \omega + 1 = 0$, however $\omega \notin \mathbb{Z}[\sqrt{-3}]$.

4. In general, we will need more tools to find $\mathcal{O}_k$, but in the quadratic case, since the polynomial are relatively simple to work with, we may find all rings of integers. Let $k = \mathbb{Q}(\sqrt{m})$. Then:

$$\mathcal{O}_k = \mathbb{Z}[\sqrt{m}] \qquad m \equiv 2, 3 \bmod 4$$
$$\mathcal{O}_k = \mathbb{Z}\left[\frac{1 + \sqrt{m}}{2}\right] \qquad m \equiv 1 \bmod 4$$

To see this, let $\alpha = r + s\sqrt{m}$, $r, s \in \mathbb{Q}$. If $s = 0$, we've done, hence if $s \neq 0$ we get the monic irreducible polynomial:

$$x^2 - 2rx + r^2 - ms^2$$

Hence, α is algebraic if and only if $2r$ and $r^2 - ms^2$ are integers. Since $2r \in \mathbb{Z}$, $r \in \frac{1}{2}\mathbb{Z}$. Considering now $r^2 - ms^2 \in \mathbb{Z}$, if $r = \alpha/2$ for $\alpha \in \mathbb{Z}$, then subtracting both sides by $(\alpha^2 - 1)/4$ we get that $1/4 - ms^2 \in \mathbb{Z}$ or $m^2 = \mathbb{Z} + 1/4$. Now, this equality holds true if $s = \beta/2$ and $m \equiv 1 \bmod 4$ or $s = \beta$ and $m \equiv 2, 3 \bmod 4$. This now gives our restrictions on our coefficients.

We should think of ring of integers as being a lattice given by inclusion with $(\mathbb{Q}, \mathbb{Z})$ the initial object and $(\mathbb{C}, \mathcal{O}_{\mathbb{C}})$ the final object. We may want to relate these two lattices with a pair of adjoint functors. To do this, if we have $\varphi : F \to L$ be a field homomorphism (not necessarily a k-homomorphism), then $\varphi|_{\mathcal{O}_F} : \mathcal{O}_F \to \mathcal{O}_L$ is certainly well defined since if $\alpha \in \mathcal{O}_F$, then:

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} + \alpha^n = 0 \qquad a_i \in \mathbb{Z}$$

Then $\varphi(a_i\alpha^i) = a_i\varphi(\alpha)^i$, since $a_i \in \mathbb{Z}$. Hence $\varphi(\alpha) \in \mathcal{O}_L$. From this, we may define $\mathcal{O}_- : \mathbf{Fld}^0 \to \mathbf{O}$ where $\mathbf{Fld}^0$ are fields of characteristic 0 and $\mathbf{O}$ is a full subcategory of $\mathbf{Ring}$ containing rings of integers as objects and ring homomorphisms

$$\begin{tikzcd} F \arrow[r, hook, "\varphi"] \arrow[d, "\mathcal{O}_-"] & L \arrow[d, "\mathcal{O}_-"] \\ \mathcal{O}_F \arrow[r, "\varphi|_{\mathcal{O}_F}"] & \mathcal{O}_L \end{tikzcd}$$

Now given a map $\psi : \mathcal{O}_F \to \mathcal{O}_K$, can we find a map $\Psi : F \to K$ which is furthermore we an adjoint to the above forgetful functor. To find this functor, we first find how to retrieve the field from the ring of integers:

Proposition 22.1.5: Field of Fraction of Integral Closures of Extensions

Let R be an integral domain, $K = K(R)$ its field of fraction, and $S = \overline{R}^L$ where L/K is an algebraic extension. Then $L = K(S)$. Equivalently, $L = \langle S \rangle_{K(R)}$.

Concretely, we would have $R = \mathbb{Z}$ and $K(R) = \mathbb{Q}$ so that $S = \overline{\mathbb{Z}}^L = \mathcal{O}_L$. Then $L = K(\mathcal{O}_L)$.

Proof **:**
Let R, K, S and L be defined as in the question. We shall first show that if $\alpha \in L$ is algebraic over K, then there exists a $d \in R$ such that $d\alpha$ is integral over R. Since α is algebraic over K,

$$\alpha^m + a_1\alpha^{m-1} + \cdots + a_m = 0 \qquad a_i \in K$$

Let d be the least common multiple of the denominators of a_i so that $da_i \in R$ for all i. Then multiplying the above equation by d^m, we get:

$$d^m\alpha^m + d^m a_1\alpha^{m-1} + \cdots + d^m a_m = 0 \qquad a_i \in K$$

re-writing:

$$(d\alpha)^m + da_1(d\alpha)^{m-1} + \cdots + d^m a_m = 0 \qquad a_i \in K,$$

where each $a_i d, ..., a_m d^m \in R$, hence $d\alpha$ is integral over R. Note that this immediately gives the equivalence $L = \langle S \rangle_{K(S)}$.

With that, to see that $L = K(S)$, notice that every $\alpha \in L$ can be written as $\alpha = b/d$ for $d \in R$, hence $L \subseteq K(S)$. Conversely, if $a/b \in K(S)$, where $a, b \in S$ and each satisfy an R-polynomial, then they are certainly satisfy a K-polynomial, and furthermore S is all elements of L that satisfy an R-polynomial, hence $a/b \in L$, giving equality.

Corollary 22.1.6: Basis For Number Field

Let $F/\mathbb{Q}$ be a finite extension of degree n and $(\alpha_1, \alpha_2, ..., \alpha_n)$ an $\mathbb{Q}$-basis for F. Then there exists an $m_i \in \mathbb{Z}$ such that $m_1\alpha_1, m_2\alpha_2, ..., m_n\alpha_n \in \mathcal{O}_F$

Proof **:**
let $R = \mathbb{Z}$ in proposition 22.1.5.

Proposition 22.1.7: Relating Rings Of Integers

Let $K/\mathbb{Q}$ be a number field and $\mathcal{O}_K$ the ring of integers of K. Then:

1. If L/K, then $\mathcal{O}_L \cap K = \mathcal{O}_K$. In particular $\mathcal{O}_K \cap \mathbb{Q} = \mathbb{Z}$.
2. $\mathcal{O}_K \otimes_{\mathbb{Z}} \mathbb{Q} \cong K$. Namely, $\mathcal{O}_- \otimes_{\mathbb{Z}} \mathbb{Q} : \mathbf{O} \to \mathbf{Fld}^0$ is a functor.

Giving us a correspondence that is covariant under inclusion:

$$\left\{\begin{matrix} \varphi : F \to K \\ \mathrm{char}(F) = \mathrm{char}(K) = 0 \end{matrix}\right\} \underset{(-)\otimes_{\mathbb{Z}}\mathbb{Q}}{\overset{\overline{\mathbb{Z}}^{(-)}}{\rightleftarrows}} \left\{\begin{matrix} \varphi : \mathcal{O}_F \to \mathcal{O}_K \\ K(\mathcal{O}_F) = F,\ K(\mathcal{O}_K) = K \end{matrix}\right\}$$

Or similarly:

$$\mathrm{Hom}_K(L_1, L_2) \xrightarrow{\sim} \mathrm{Hom}_A(B_1, B_2)$$

Proof **:**

1. This is essentially by definition: $\mathcal{O}_k = \overline{\mathbb{Z}}^k$ are all integral elements over $\mathbb{Z}$ that belonging k. Hence,$\mathcal{O}_k \cap \mathbb{Q}$ must be integral over k and also rational. But the only integral elements that are rational are the integers, hence we have equality. Similarly, $\mathcal{O}_L \cap k$ are all integral elements over $\mathbb{Z}$ that also belong to k, which by definition is $\mathcal{O}_k$
2. This comes from proposition 20.3.9 and its corollary.

This tells us how to relate $\mathcal{O}_k$ with k and a bit about how to move down the lattice, but we have not said much about the structure of $\mathcal{O}_k$. We know it is a ring, but is it a PID? a UFD? Every ring is a $\mathbb{Z}$-module; it is a finitely generated $\mathbb{Z}$-module? Noetherian $\mathbb{Z}$-module? Projective? The strategy in example 22.1.2 to find the structure of $\mathcal{O}_k$ gets much harder for more complicated roots of irreducible, even once we have degree 3, and especially beyond degree 4. Our first goal is to establish the structure of $\mathcal{O}_K$. To find the structure, we will require some tools for detecting elements of $\mathcal{O}_K$: we shall find numerical values for the elements of K to determine whether they are elements of $\mathcal{O}_K$, i.e. we shall find functionals. We shall see that $\mathcal{O}_K$ will be a free $\mathbb{Z}$-module of finite rank, in fact it shall be a particular $\mathbb{Z}$-module known as a *complete* lattice, as shall be demonstrated in section 22.1.4.

22.1.1 Functionals: Field Trace and Field Norms

(this link is useful: here)

Let M/k be a finite extension, and take $\alpha \in M$. We shall define two invariances on α that will detect whether α is algebraic over k or not, detect when it is an algebraic integer, and eventually help us find the structure of $\mathcal{O}_M$ as well as the factorization of elements of $\mathcal{O}_M$.

Define $L_\alpha : M \to M$ by $m \mapsto \alpha m$. This is a k-linear transformation, since:

$$L_\alpha(a + kb) = \alpha(a + kb) = \alpha a + k\alpha b = L_\alpha(a) + kL_\alpha(b)$$

Thus, we can represent this linear transformation in a matrix form (usually with basis given by α).

Since the trace and determinant are basis independent calculations, it is meaningful to ponder what the trace/determinant of this transformation and what information it may provide us:

Definition 22.1.8: Trace And Norm

Let M/k be a finite extension, $L_\alpha : M \to M$ be a linear transformation defined by multiplication Then we call the trace of this linear transformation the *trace* of α and the determinant of this linear transformation the *norm* of α.

Example 22.1.3: Trace And Norm

1. Let $\mathbb{C}/\mathbb{R}$ be the extension and $\alpha = x + iy$. Pick basis $\{1, i\}$. Then multiplication is:

$$\begin{pmatrix} x & y \\ -y & x \end{pmatrix}$$

and so the trace is $2x = 2\operatorname{Re}\alpha$ and the norm is $x^2 + y^2 = \|\alpha\|^2$. If you know some complex analysis, then we see that the norm does indeed match with our usual understanding of the norm, and the trace can be thought of as a generalization of taking the real part of the complex number (up to a factor).

2. Let $\alpha = a + b \cdot \left(\frac{1+\sqrt{-7}}{2}\right) \in \mathbb{Z}\left[\frac{1+\sqrt{-7}}{2}\right]$. Then:

$$\operatorname{Nm}_{\mathbb{Q}(\sqrt{-7})/\mathbb{Q}}(\alpha) = \alpha^2 + \alpha\beta + 2\beta^2$$

3. Let $M = k(\alpha)$ be a finite extension, and let

$$m_{\alpha,k}(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

be the minimal polynomial. Then this is the same as the characteristic polynomial (since plugging in L_α into the minimal polynomial produces the zero transformation by Caley Hamilton [a]). Then the trace of α is $-a_{n-1}$ and the norm is $\pm a_0$ depending on whether $m_{\alpha,k}(x)$ is even or odd.

[a] Alternatively, choose basis $1, \alpha, ..., \alpha^{n-1}$ and compute

Naturally, if F/k, $a_1, a_2 \in F$ and $b \in k$:

$$\operatorname{tr}_{F/k}(a_1 + a_2) = \operatorname{tr}_{F/k}(a_1) + \operatorname{tr}_{F/k}(a_2) \quad \operatorname{tr}_{F/k}(ba_1) = b\operatorname{tr}_{F/k}(a_1) \quad \operatorname{tr}_{F/k}(a) = na$$

and:

$$\operatorname{Nm}_{F/k}(a_1 a_2) = \operatorname{Nm}_{F/k}(a_1)\operatorname{Nm}_{F/k}(a_2) \quad \operatorname{Nm}_{F/k}(ba_1) = b^n \operatorname{Nm}_{F/k}(a_1) \quad \operatorname{Nm}_{F/k}(b) = b^n$$

Most of the properties of the trace and norm come from linear algebra, and furthermore since every finite separable field is a primitive element generating the entire field, the norm and trace always shows up in the minimal polynomial (as was pointed out in section ref:HERE). In section 18.2.3, we saw another way of characterizing the trace and norm, which will be emphasized here due to its importance:

Proposition 22.1.9: Trace and Norm via Galois Conjugates

Let M/k be a finite extension where $[M:k] = n$ and let $\alpha \in M$ be algebraic over k. Then:

$$\mathrm{tr}_{M/k}(\alpha) = [M:k(\alpha)] \cdot \sum_{\substack{r \\ \text{root of } m_{\alpha,k}}} r \qquad \mathrm{Nm}_{M/k}(\alpha) = \left(\prod_{\substack{r \\ \text{root of } m_{\alpha,k}}} r \right)^{[M:k(\alpha)]}$$

Proof **:**
Let $\sigma_i(\alpha) \in \overline{k}$ be the roots of $m_{\alpha,k}(x)$ where $\sigma_i : M \to \overline{k}$ are the embeddings of M into the algebraic closure of k. Suppose first that $M = k(\alpha)$ and is a separable extension. Pick a basis $1, \alpha, ..., \alpha^{n-1}$. Then concretely computing the trace, we get the $n-1$th coefficient of

$$\alpha^n - a_{n-1}\alpha^{n-1} - \cdots - a_1\alpha - a_0 = 0$$

which, if we factor out the above polynomial in a normal closure and collect the terms of the roots as we've done in section 18.2.3, we see that

$$\alpha_{n-1} = \sum_i \sigma_i(\alpha)$$

that is, it is the sum of the roots. Similarly, we get that $\mathrm{Nm}_{F/k}(\alpha) = a_0 = \prod_i \sigma_i(\alpha)$, giving us both results.

For the general finite extension case, we may use induction and transitivity. The inseparable case is left as an exercise.

Corollary 22.1.10: Trace and Norm Via Embeddings

Let L/k be a separable extension of degree n, an $\sigma_1, \sigma_2, ..., \sigma_n$ the embedding of L into it's Galois closure or $\overline{k}$. Then for $\alpha \in L$:

$$\mathrm{tr}_{L/k}(\alpha) = \sum_i^n \sigma_i(\alpha) \qquad \mathrm{Nm}_{L/k}(\alpha) = \prod_i^n \sigma_i(\alpha)$$

Proof **:**
this is just a re-phrasing of the previous result. Note that there may be $\sigma_i(\alpha)$ that repeat since σ_i may be fixed in $k(\alpha)$.

This immediately shows us that for algebraic elements, $\mathrm{tr}_{M/k}(\alpha), \mathrm{Nm}_{M/k}(\alpha) \in k$, since $\pm a_{n-1}, a_0 \in k$. An easy way to remember that that the trace gives a_{n-1} and Norm gives a_0 is that it is not possible that a minimal polynomial have 0 constant term since

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x$$

is reducible. Hence, $\mathrm{Nm}_{M/k}(\alpha) = a_1$ while $\mathrm{tr}_{M/k}(\alpha) = a_{n-1}$. To relate this back to our goal of understanding $\mathcal{O}_M$, we would like to identify when $\alpha \in M$ belongs in $\mathcal{O}_M$ using tr and Nm.

This is exactly when the coefficients are integers, that is $\mathrm{tr}_{M/k}(\alpha), \mathrm{Nm}_{M/k}(\alpha) \in \mathbb{Z}$. We've in fact proved something slightly more general that is important to point out for when we are working with intermediary extensions:

Corollary 22.1.11: Trace and Norm on Rings of Integers

Let R be an integrally closed integral domain ($\overline{R}^K = R$), and let $F/K(R)$ be a finite field extension of the field of fraction of R. Then if $\alpha \in F$ is integral over R,

$$\mathrm{tr}_{F/K(R)}(\alpha) \in R \qquad \mathrm{Nm}_{F/K(R)}(\alpha) \in R$$

A particular consequence of this is $\mathrm{tr}_{L/k}(\mathcal{O}_L) \subseteq \mathcal{O}_k$

Proof **:**
If α is integral, then certainly so are all its Galois conjugates. But then adding and multiplying all the Galois conjugates remains integral, meaning the coefficients of the corresponding polynomials remains in R, hence the trace and norm of α is in R.

In particular, if $\alpha \in \mathcal{O}_L = \overline{\mathbb{Z}}^L \subseteq \overline{k}^L$, $\mathrm{tr}_{L/k}(\alpha) \in \mathcal{O}_L \cap k = \mathcal{O}_k$ by proposition 22.1.7.

If $R = \mathbb{Z}$ and $F/\mathbb{Q}$ is a finite extension, we get can detect potential integral elements using the trace and norm! Since we may be working with intermediate fields $M/L/k$, it is useful to know how to compute the trace and norm given information about the trace and norm of the intermediate fields:

Proposition 22.1.12: Trace and Norm w.r.t. Towers

Let $M/L/K$ be finite extensions. Then:

$$\mathrm{tr}_{M/K}(\alpha) = \mathrm{tr}_{L/K}(\mathrm{tr}_{M/L}(\alpha)) \qquad \mathrm{Nm}_{M/K}(\alpha) = \mathrm{Nm}_{L/K}(\mathrm{Nm}_{M/L}(\alpha))$$

Proof **:**
Let $\sigma_1, \sigma_2, ..., \sigma_n$ be the embeddings of L in $\mathbb{C}$ which fix K pointwise and let $\tau_1, \tau_2, ..., \tau_m$ be the embeddings in $\mathbb{C}$ fixing L pointwise. We will want to compose σ_i with the τ_j, so we will extend our embeddings as follows: Let N/M be the normal closure of M so that all σ_i, τ_j can be extended to automorphisms of N; relabel our extensions σ_i, τ_j (i.e. keep the same labelling). Then:

$$\mathrm{tr}_{L/k}(\mathrm{tr}_{M/L}(\alpha) = \sum_i^n \sigma_i \left(\sum_j^m \tau_j(\alpha) \right) = \sum_{i,j} \sigma_i \tau_j(\alpha)$$

$$\mathrm{Nm}_{L/k}(\mathrm{Nm}_{M/L}(\alpha) = \prod_i^n \sigma_i \left(\prod_j^m \tau_j(\alpha) \right) = \prod_{i,j} \sigma_i \tau_j(\alpha)$$

It only remains to show that the mappings $\sigma_i \tau_j$ when restricted to M give the embeddings of M in $\mathbb{C}$ fixing k pointwise. Certainly they fix k pointwise and there is the right number of them, so this follows by showing they are all distinct (TBD).

Note that it is not necessarily sufficient that the trace and norm is integral:

Example 22.1.4: Trace And Norm Insufficient To Detect Integral Elements
Let $L = \mathbb{Q}(\alpha)$ where $\alpha^4 + 2\alpha^3 + 1/2\alpha^2 + 1/2\alpha + 2\alpha = 0$. Note that $x^4 + 2x^3 + 1/2x^2 + 1/2x + 3$ is irreducible, hence this is a degree 4 extension. Furthermore,

$$\mathrm{tr}_{L/k}(\alpha) = 2 \qquad \mathrm{tr}_{L/K}(\alpha) = 3$$

But certainly $\alpha \notin \mathcal{O}_L$.

A notable exception to this is for quadratic extensions, since any extension of degree 2 is quadratic, giving only 2 possible coefficients:

Example 22.1.5: Rings of Integers of Quadratic Extensions
Let $K = \mathbb{Q}(\sqrt{n})$ where n is square free, and consider $\alpha = a + b\sqrt{n} \in \mathcal{O}_K$ where a, b are square free. Then:

$$\mathrm{tr}_{K/\mathbb{Q}}(\alpha) = 2a \in \mathbb{Z} \qquad \mathrm{Nm}_{K/\mathbb{Q}}(\alpha) = (a + b\sqrt{n})(a - b\sqrt{n}) = a^2 - nb^2 \in \mathbb{Z}$$

We may ask when $2a \in \mathbb{Z}$, that is when is it possible that $a = 1/2$? Let's say it was, so that

$$\left(\frac{1}{2}\right)^2 - nb^2 \in \mathbb{Z} \qquad \Leftrightarrow \qquad \frac{b^2 n}{4} \in \mathbb{Z} + \frac{1}{4}$$

hence, $4b^2 n \in \mathbb{Z}$. Since n is square-free, $n \neq m^2$ implying that we cannot have $b^2 = p^2/m^2 \cdot m^2$, so $2b \in \mathbb{Z}$. Thus, if $a = 1/2$, it must be that $b = 1/2$. Similarly, it is easy to see that if b is an integer, a has to be an integer.

Now, is $\alpha = \frac{1}{2} + \frac{1}{2}\sqrt{n}$ an algebraic integer? For this to be the case, it would have to solve the equation:

$$\alpha^2 - \alpha + \frac{1-n}{4} = 0$$

or equivalently, $\mathrm{Nm}(\alpha) = \frac{1-n}{4}$. Hence, we need $n \equiv 1 \bmod 4$. Thus, we have

$$\mathcal{O}_{\mathbb{Q}(\sqrt{d})} = \begin{cases} \mathbb{Z}[\sqrt{d}] & d \not\equiv 1 \bmod 4 \\ \mathbb{Z}\left[\frac{1+\sqrt{n}}{2}\right] & d \equiv 1 \bmod 4 \end{cases}$$

A final result in this section is to find the group of units of $\mathcal{O}_K$. Note that in general, if $\alpha \in \mathcal{O}_K$, then $1/\alpha \notin \mathcal{O}_K$; for example $2 \in \mathcal{O}_\mathbb{Q}$, but certainly $1/2 \notin \mathcal{O}_\mathbb{Q}$. In fact, the only units in $\mathcal{O}_\mathbb{Q}$ are ± 1. If $\mathcal{O}_K$ contains roots of unity, then certainly their inverse is in $\mathcal{O}_K$ (for if $\alpha = \zeta_n$ $1/\zeta = \zeta_n^{n-1}$). There are in general more units in $\mathcal{O}_K$, and the group of units of a ring of integers shall be fully characterized in section 22.2.4. With the norm, we have a way of detecting units:

Proposition 22.1.13: Norm And Units

Let L/K be a finite field extension. Then

$$\alpha \in \mathcal{O}_L^* \qquad \Leftrightarrow \qquad \mathrm{Nm}_{L/K}(\alpha) \in \mathcal{O}_K^*$$

In the space case where $K = \mathbb{Q}$, then $\alpha \in \mathcal{O}_L^*$ if and only if $\mathrm{Nm}_{L/\mathbb{Q}}(\alpha) = \pm 1$.

Proof **:**
Let $\alpha \in \mathcal{O}_L^*$ so that $\mathrm{Nm}_{L/K}(\alpha) \in \mathcal{O}_K$. If α is a unit in $\mathcal{O}_L$, then $1/\alpha \in \mathcal{O}_L$ exists. Then:

$$1 = \mathrm{Nm}_{L/K}(1) = \mathrm{Nm}_{L/K}(\alpha \cdot 1/\alpha) = \mathrm{Nm}_{L/K}(\alpha)\,\mathrm{Nm}_{L/K}(1/\alpha)$$

hence, $\mathrm{Nm}_{L/K}(\alpha), \mathrm{Nm}_{L/K}(1/\alpha) \in \mathcal{O}_K^\times$. Conversely, if $\mathrm{Nm}_{L/K}(\alpha) \in \mathcal{O}_K$, then there exists a β such that

$$\beta\,\mathrm{Nm}_{L/K}(\alpha) = 1$$

Then by definition of the norm:

$$1 = \beta\,\mathrm{Nm}_{L/K}(\alpha) = \beta \prod_{\sigma} \sigma(\alpha) = \alpha \left(\beta \prod_{\sigma \neq \mathrm{id}} \sigma(\alpha) \right) = \alpha y$$

where $y \in \mathcal{O}_L$ since each $\sigma(\alpha) \in \mathcal{O}_L$, which gives the result.

Proposition 22.1.14: Norm And Irreducible

Let $L/\mathbb{Q}$ be a finite extension and $\alpha \in \mathcal{O}_L$. Then if p is prime:

$$\mathrm{Nm}_{L/\mathbb{Q}}(\alpha) = p \implies \alpha \text{ is irreducible}$$

Proof **:**
Suppose α where reducible so that $\alpha = \beta\gamma$ where β, γ or not units, hence have norm greater than 1. Then:

$$\begin{aligned} \mathrm{Nm}_{L/\mathbb{Q}}(\alpha) &= \mathrm{Nm}_{L/\mathbb{Q}}(\beta\gamma) \\ &= \mathrm{Nm}_{L/\mathbb{Q}}(\beta)\,\mathrm{Nm}_{L/\mathbb{Q}}(\gamma) \\ &= p \end{aligned}$$

a contradiction.

Note that the converse is not true

Example 22.1.6: Prime, But Norm Not Prime
Let $K = \mathbb{Q}(\sqrt{2})$ so that $\mathcal{O}_K = \mathbb{Z}[\sqrt{2}]$. Then $\mathrm{Nm}_{K/\mathbb{Q}}(5) = 5^2 = 25$, which is certainly not prime, however 5 is prime in $\mathbb{Z}[\sqrt{2}]$. This is at this stage a tedious calculation: assume that it factors $5 = \alpha\beta$ where $\alpha, \beta \in \mathbb{Z}[\sqrt{2}]$, $\alpha = a + bi$, $\beta = c + di$, and:

$$\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) = a - 2b = 5 = c - 2d = \mathrm{Nm}_{K/\mathbb{Q}}(\beta)$$

but no such a, b, c, d exist. Later, we shall be able to deduce this almost immediately using Dedekind-Kummer's Theorem (see theorem 22.4.6).

In lemma 22.2.23, we shall give a partial converse which shall as a corrollary imply:

$$\alpha \text{ prime} \implies \mathrm{Nm}_{K/\mathbb{Q}}(\alpha) = p^n$$

for appropriate n. Note how α must be prime, which is *stronger* than irreducible. However, we will often be working over rings where we have few prime elements but many irreducible elements. In section 22.2 we shall show how to "fix" this problem to make the norm more useful.

22.1.2 Functionals: Discriminant

As we saw, if $\alpha \in F/\mathbb{Q}$ is integral, then $\mathrm{tr}(\alpha), \mathrm{Nm}(\alpha) \in \mathbb{Z}$. However, the converse is not necessarily true as we saw in example 22.1.4. We will require a more powerful result, one that will verify that an element is an element of a ring of integers. For this, we shall bring in some multilinear algebra and define a new tool that shall become invaluable finding a *basis* for $\mathcal{O}_K$ (recall that it was mentioned that $\mathcal{O}_K$ shall be proven to be a finitely generated free $\mathbb{Z}$-module). Given a basis can be found, then an element can be determined to be integral using the basis. The tool that shall be used to detect a basis for $\mathcal{O}_K$ is called the *field discriminant*. The norm and the trace shall both be used in the definition of the field discriminant.

Some Multi-Linear Algebra

Recall in section 13.3.2 that multi-linear forms have matrix representations and so the determinant of multi-linear maps is well-defined:

Definition 22.1.15: [General] Discriminant

Let $\varphi : V \times V \to k$ be a symmetric k-bilinear map ($\varphi(x, y) = \varphi(y, x)$). Then given a basis $v_1, v_2, ..., v_n$, the *discriminant* of φ is:

$$\mathrm{disc}(\varphi) = \det \begin{pmatrix} \varphi(v_1, v_1) & \varphi(v_1, v_2) & \cdots & \varphi(v_1, v_n) \\ \varphi(v_2, v_1) & \varphi(v_2, v_2) & \cdots & \varphi(v_2, v_n) \\ \vdots & \ddots & \vdots & \\ \varphi(v_n, v_1) & \varphi(v_n, v_2) & \cdots & \varphi(v_n, v_n) \end{pmatrix}$$

and is denoted $\mathrm{disc}(\varphi)$.

Let $\{\beta_1, \beta_2, ..., \beta_n\}$ be a basis for V and $\{v_1, v_2, ..., v_n\}$ be some collection of elements so that

$$v_i = \sum_i a_i \beta_i$$

Then:

$$\varphi(v_k, v_l) = \sum_{i,j} \varphi(a_{ki}\beta_i, a_{lj}\beta_j) = a_{ki}\varphi(\beta_i, \beta_j)a_{kl}$$

hence, as matrices, we have the (component-wise) equality:

$$(\varphi(v_k, v_l)) = A \cdot (\varphi(e_i, e_j)) \cdot A^{\mathrm{tr}}$$

where $A = (a_{ij})$. Then if we have a change of basis, we have the determinant of the matrix is:

$$\det((\varphi(f_i, f_j)) = \det(A)^2 \cdot \det((\varphi(\beta_i, \beta_j))) \tag{22.2}$$

Since $A \in \mathrm{GL}(k)$, we see that the discriminant is defined up to the square of a unit. Hence, if $\mathrm{disc}(\varphi) \neq 0$, $\mathrm{disc}(\varphi)$ is well-defined in $k^\times/(k^\times)^2$ If $K = \mathbb{Q}$, then $(\mathbb{Q}^\times)^2 = 1$, hence the discriminant is well-defined in $\mathbb{Q}^\times$.

Invertible matrices allowed us to define an isomorphism between V and $V^\vee$ when V is finite dimensional. We shall require this same property for bilinear forms. For this, the notion of non-generate form needs to be upgraded for bilinear forms:

Definition 22.1.16: Non-degenerate Form

let φ be a bilinear form. Then φ is *non-generate* if φ has a non-zero discriminant

Proposition 22.1.17: Non-degenerate Form Equivalences

let φ be a bilinear form. Then φ is *non-generate* if any of the following equivalent conditions hold:

1. φ has a non-zero discriminant
2. the left kernel is zero: $\{v \in V \mid \varphi(v, x) = 0,\ \forall x \in V\}$
3. the right kernel is zero: $\{v \in V \mid \varphi(x, v) = 0,\ \forall x \in V\}$

Proof **:**
exercise

if φ is non-degenerate, it always induces an isomorphism $V \mapsto V^\vee$, $v \mapsto (x \mapsto \varphi(v, x))$. With this, build-up, we shall define a particularly important symmetric k-bilinear form that will be useful in finding the rank of the ring of integers:

Definition 22.1.18: Trace Pairing

Let F/k be a finite field extension. Consider:

$$\mathrm{tr}_{F/k}(\alpha\beta) : F \times F \to k \qquad (\alpha, \beta) \mapsto \mathrm{tr}_{F/k}(\alpha\beta)$$

Then $\mathrm{tr}_{F/k}(\alpha\beta)$ is called the *trace-pairing*.

Note that the trace pairing is non-degenerate. We see in definition 22.1.16 that $\mathrm{tr}_{F/k}$ is non-degenerate if the left kernel is trivial. If $\alpha \in F$, then

$$\mathrm{tr}_{F/k}(\alpha\alpha^{-1}) = \mathrm{tr}_{F/k}(1) = [F : k] \neq 0$$

giving the result. Hence, we have an isomorphism:

$$F \xrightarrow{\sim} \mathrm{Hom}_k(F, k) = F^* \qquad x \mapsto (y \mapsto \mathrm{tr}_{F/K}(xy))$$

We shall soon require a particular case of the above that is singled out here:

Definition 22.1.19: Integral Trace Dual

Let F/k be a finite extension. Then:

$$F^{\mathrm{tr}\,\vee} = F^{\vee} = \left\{ b \in F \;\middle|\; a \in F,\ \mathrm{tr}_{F/k}(ab) \in \mathbb{Z} \right\}$$

is called the *Integral trace dual* of F. If unambiguous, the trace dual is denoted $F^{\vee}$.

Field Discriminant

With a general overview of discriminants, we return to detecting algebraic integers more precisely using the discriminant. First the determinant, det, is an alternating n-form, meaning $(\det)^2$ is a symmetric n-form. Note that this is the same discriminant introduced in section 18.3; the equivalence shall soon be shown.

Definition 22.1.20: [Field] Discriminant

Let L/k be a finite extension of degree n, and let $\sigma_1, \sigma_2, ..., \sigma_n \in \mathrm{Emb}_k(L)$ be the k-embeddings of L in $\mathbb{C}$. Then for any n-tuple of elements $\alpha_1, \alpha_2, ..., \alpha_n \in L$, define the *discriminant* of $\alpha_1, \alpha_2, ..., \alpha_n$ to be:

$$\mathrm{disc}_{L/k}(\alpha_1, \alpha_2, ..., \alpha_n) = \det \begin{pmatrix} \sigma_1(\alpha_1) & \cdots & \sigma_1(\alpha_n) \\ \vdots & \ddots & \vdots \\ \sigma_n(\alpha_1) & \cdots & \sigma_n(\alpha_n) \end{pmatrix}^2 = \det((\sigma_i(\alpha_j))_{ij})^2$$

Notice that the square makes the discriminant independent of the ordering of the σ_i and α_j. If we chose a different set of elements $(\beta_1, \beta_2, ..., \beta_n)$ that differ from $(\alpha_1, \alpha_2, ..., \alpha_n)$ by a change of basis, that is $\vec{\alpha} = A\vec{\beta}$, then as we observed in equation 22.2, we get

$$\mathrm{disc}_{L/k}(\beta_1, \beta_2, ..., \beta_n) = \det(A)^2 \cdot \mathrm{disc}_{L/k}(\alpha_1, \alpha_2, ..., \alpha_n)$$

meaning the discriminant is not quite basis independent, meaning it is properly defined for operators in $L^{\times}/L^{\times 2}$ (where $L^{\times}$ are the units of L, since det maps invertible matrices to units, see proposition 13.4.8).

Example 22.1.7: Field Discriminant

1. Let $K = \mathbb{Q}(\sqrt{m})$ where m is square free. Then K has a basis $A = \{1, \sqrt{m}\}$ and $B = \left\{1, \frac{1+\sqrt{m}}{2}\right\}$. If $m \equiv 2, 3 \bmod 4$, then the former is a basis for $\mathcal{O}_K$ and if $m \equiv 1 \bmod 4$, then then the latter is a basis for $\mathcal{O}_K$. Computing both, we get:

$$\mathrm{disc}_{K/\mathbb{Q}}(A) = \det \begin{pmatrix} 1 & \sqrt{m} \\ 1 & -\sqrt{m} \end{pmatrix}^2 = (-\sqrt{m} - \sqrt{m})^2 = 4m$$

$$\mathrm{disc}_{K/\mathbb{Q}}(B) = \det \begin{pmatrix} 1 & \frac{1+\sqrt{m}}{2} \\ 1 & \frac{1-\sqrt{m}}{2} \end{pmatrix}^2 = \left(\frac{-2\sqrt{m}}{2}\right)^2 = m$$

Yet another basis for $\mathbb{Q}(\sqrt{m})$ is $C = \{1, \sqrt{m}/10\}$. In that case:

$$\operatorname{disc}_{K/\mathbb{Q}}(C) = \det \begin{pmatrix} 1 & \frac{\sqrt{m}}{10} \\ 1 & -\frac{\sqrt{m}}{10} \end{pmatrix} = \left(\frac{-\sqrt{m}}{5} \right)^2 = \frac{m}{25}$$

However, C is not a basis for $\mathcal{O}_K$, namely since the denominator is too big. Notice that the first two discriminant integers, while the third one is not.

2. Let $K = \mathbb{Q}(\sqrt[3]{2})$. Then $A = \{1, \sqrt[3]{2}, \sqrt[3]{3}\}$ is a basis for K, and:

$$\operatorname{disc}_{K/\mathbb{Q}}(A) = \begin{pmatrix} 1 & \sqrt[3]{2} & \sqrt[3]{4} \\ 1 & \sqrt[3]{2}\zeta_3 & \sqrt[3]{3}\zeta_3^2 \\ 1 & \sqrt[3]{2}\zeta_3^2 & \sqrt[3]{3}\zeta_3 \end{pmatrix}^2 = -108 = -2^2 \cdot 3^3$$

showing that the discriminant can be negative. We shall come back to seeing how this is useful for finding whether $\mathcal{O}_K = \mathbb{Z}[\sqrt[3]{2}]$.

The following alternative form is useful for an alternate way of computing the discriminant as well as important for deducing some theoretical properties of the elements inputted into the discriminant:

Proposition 22.1.21: Discriminant Via Trace-Pairing

$$\operatorname{disc}_{L/k}(\alpha_1, \alpha_2, ..., \alpha_n) = \det(\operatorname{tr}_{L/k}(\alpha_i \alpha_j))$$

Proof **:**
This follows immediately from the matrix equation. If $[\sigma_i(\alpha_j)]$ represents the matrix:

$$[\sigma_j(\alpha_i)][\sigma_i(\alpha_j)] = [\sigma_1(\alpha_i\alpha_j) + \cdots + \sigma_n(\alpha_i\alpha_j)] = [\operatorname{tr}_{L/k}(\alpha_i\alpha_j)]$$

Using the fact that the determinant is multiplicative and invariant under transposes, we get our result.

In particular, the above form allows for the detection of sets of algebraic integers:

Corollary 22.1.22: Detecting Algebraic Integers using Discriminant

Let K be a number field and $\alpha_1, \alpha_2, ..., \alpha_n$ a $\mathbb{Q}$-basis. Then $\operatorname{disc}_{K/\mathbb{Q}}(\alpha_1, ..., \alpha_n) \in \mathbb{Q}$, and if all $\alpha_1, \alpha_2, ..., \alpha_n$ are algebraic integers, then $\operatorname{disc}(\alpha_1, \alpha_2, ..., \alpha_n) \in \mathbb{Z}$.

Proof **:**
we know that $\operatorname{tr}_{K/\mathbb{Q}}(\alpha_i\alpha_j) \in \mathbb{Q}$ by proposition 22.1.9 and are integers via corollary 22.1.11, hence $\operatorname{disc}_{K/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n) \in \mathbb{Q}$, and if they are all algebraic integers then $\operatorname{disc}_{K/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n) \in \mathbb{Z}$.

Corollary 22.1.23: Discriminant And Linear Independence

$\operatorname{disc}_{K/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n) = 0$ if and only if $\alpha_1, \alpha_2, ..., \alpha_n$ are $\mathbb{Q}$-linearly dependent.

***Proof* :**
If the $\alpha_1, \alpha_2, ..., \alpha_n$ are $\mathbb{Q}$-linearly dependent, then two of the columns would be linearly dependent, hence $\operatorname{disc}_{K/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n) = 0$. Conversely, assuming $\operatorname{disc}_{K/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n) = 0$, then the rows of the matrix $[T(\alpha_i\alpha_j)]$ are linearly dependent; let's label them R_i so that for some $a_1, a_2, ..., a_n$ which are not all zero:

$$a_1R_1 + \cdots + a_nR_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

For the sake of contradiction, let's say that the $\alpha_1, \alpha_2, ..., \alpha_n$ are $\mathbb{Q}$-linearly independent. Then they form a $\mathbb{Q}$-basis for K^n. Consider:

$$\alpha = a_1\alpha_n 1 + \cdots + a_n\alpha_n$$

where $\alpha \neq 0$ since some $a_i \neq 0$ and $(\alpha_1, \alpha_2, ..., \alpha_n)$ is linearly independent. Then $(\alpha\alpha_1, ..., \alpha\alpha_n)$ also forms a $\mathbb{Q}$-basis for K. By the linear dependence of the row space, we have that for each j

$$\operatorname{tr}_{K/\mathbb{Q}}(\alpha\alpha_j) = \sum_i a_i \operatorname{tr}_{K/\mathbb{Q}}(\alpha_i\alpha_j) = 0$$

But not for any $t \in K$, $t = \sum b_i\alpha\alpha_j$, and by our above observation that implies that $\operatorname{tr}_{K/\mathbb{Q}}(t) = 0$, however that is a contradiction since we know that $\operatorname{tr}_{K/\mathbb{Q}}(1) = n$. Hence, it must be that $(\alpha_1, \alpha_2, ..., \alpha_n)$ are linearly independent, completing the proof.

The corollary can be thought of as motivation for the name discriminant: If $K/\mathbb{Q}$ is a number field of dimension n, the discriminant will "discriminate" sets that are bases for $K/\mathbb{Q}$.

Since number fields are finite separable extensions, we always find a primitive element for a separable algebraic extension, the following theorem gives another important interpretation of the discriminant:

Proposition 22.1.24: Discriminant via Primitive Element

Let L/K be a finite separable field extension, with α a primitive element and $m_{\alpha,k}(x) \in K[x]$ the minimal polynomial with roots $\alpha_1, ..., \alpha_n$ in the splitting field of L. Then the *[field] discriminant* of L[a] is:

$$\Delta_K^2 = \left(\prod_{1 \leq i < j \leq n} (\alpha_i - \alpha_j) \right)^2 = \mathrm{Nm}_{L/K}(m'_{\alpha,k}(\alpha))$$

where $m'_{\alpha,k}(x)$ is the derivative of the minimal polynomial. If unambiguous the discriminant is often denoted

$$\mathrm{disc}_{L/K}(\alpha)$$

[a] In section 18.3, we called this the discriminant of $m_{\alpha,k}(x)$

***Proof* :**

Take as basis $1, \alpha, \alpha^2, ..., \alpha^{n-1}$. This comes down to noticing that $(\mathrm{tr}_{B/A}(\alpha^i \alpha^j))$ form the matrix:

$$\begin{pmatrix} 1 & \mathrm{tr}(\alpha) & \mathrm{tr}(\alpha^2) & \cdots & \mathrm{tr}(\alpha^{n-1}) \\ \mathrm{tr}(\alpha) & \mathrm{tr}(\alpha^2) & \mathrm{tr}(\alpha^3) & \cdots & \mathrm{tr}(\alpha^n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}(\alpha^{n-1}) & \mathrm{tr}(\alpha^n) & \mathrm{tr}(\alpha^{n+1}) & \cdots & \mathrm{tr}(\alpha^{2n-2}) \end{pmatrix}$$

gives us:

$$\begin{pmatrix} \sum_i \alpha_i^0 & \sum_i \alpha_i^1 & \sum_i \alpha_i^2 & \cdots & \sum_i \alpha_i^{n-1} \\ \sum_i \alpha_i^1 \alpha & \sum_i \alpha_i^2 & \sum_i \alpha_i^3 & \cdots & \sum_i \alpha_i^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_i \alpha_i^{n-1} & \sum_i \alpha^n & \sum_i \alpha_i^2 & \cdots & \sum_i \alpha_i^{2n-2} \end{pmatrix}$$

This matrix can be split into:

$$\begin{pmatrix} 1 & 1 & \cdots \\ \alpha_1 & \alpha_2 & \cdots \\ \alpha_1^2 & \alpha_2^2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots \\ 1 & \alpha_2 & \alpha_2^2 & \cdots \\ \vdots & \vdots & \ddots & \end{pmatrix}$$

which are the Vandermonde matrices we saw when dealing with the symmetric group. Then by simple computation we get:

$$\det(\mathrm{tr}_{B/A}(\alpha_i \alpha_j)) = \prod_{1 \leq i < j \leq n} (x_i - x_j)$$

For the final equality, notice that:

$$\mathrm{Nm}_{L/K}(m_{\alpha,k}(\alpha)) = \prod_{r=1}^{n} \sigma_r(m'_{\alpha,k}(\alpha)) = \prod_{r=1}^{n} m'_{\alpha,k}(\alpha_r) \stackrel{!}{=} \prod_{s \neq r} (\alpha_r - \alpha_s)$$

where $\stackrel{!}{=}$ comes from differentiating $m_{\alpha,k}(x)$ and plugging in α_r, completing the proof.

If $L = K[\alpha]$, then we shall often write $\operatorname{disc}_{L/K}(\alpha)$ instead of $\operatorname{disc}_{L/K}(\alpha, \alpha_2, .., \alpha_n)$. Note that this theorem shows that the discriminant cannot detect a basis on a non-separable extension:

Corollary 22.1.25: Discriminant And Separability

Let F/k be a finite extension. Then $\operatorname{disc}_{F/k}$ is a nonzero function if and only if F is separable

If you recall that the resultant is zero if a two polynomials share a root, then this shows how the discriminant is also related to the resultant.

Proof **:**
By proposition 22.1.24, we have the discriminant is the product of the differences of roots. If all the roots are distinct, the discriminant is nonzero, and similarly if the extension is inseparable so that there are repeated roots, the discriminant is zero.

Example 22.1.8: Discriminant On Cyclotomic extension
Let ζ_n be an pth primitive root of unity and consider $K = \mathbb{Q}[\zeta_p]$. Then recall from section 17.7.1 that it's minimal polynomial is:

$$f(x) = 1 + x + x^2 + \cdots + x^{p-1}$$

Then to compute $f'(\zeta_n)$, recall that

$$x^p - 1 = (x-1)f(x) \implies px^{p-1} = f(x) + (x-1)f'(x) \implies f'(\zeta_n) = \frac{p}{\zeta_p(\zeta_p - 1)}$$

Taking the norm, we get:

$$\operatorname{Nm}_{K/\mathbb{Q}}(f'(\zeta_p)) = \frac{\operatorname{Nm}_{K/\mathbb{Q}}(p)}{\operatorname{Nm}_{K/\mathbb{Q}}(\zeta_p)\operatorname{Nm}_{K/\mathbb{Q}}(\zeta_p) - 1} = \frac{p^{p-1}}{1 \cdot p}$$

where $\operatorname{Nm}_{K/\mathbb{Q}}(\zeta_p - 1) = p$ is left as an exercise. Hence:

$$\operatorname{disc}_{K/\mathbb{Q}}(\zeta_p) = \operatorname{Nm}_{K/\mathbb{Q}}(f'(\zeta_p)) = p^{p-2}$$

The precise value of $\operatorname{disc}(\zeta_n)$ shall be covered in lemma 22.4.48, however it is relatively easy to see that for arbitrary n, $\operatorname{disc}(\zeta_n) \big| n^{\varphi(n)}$ where φ is the Euler totient. If f is the [monic] irreducible polynomial for ζ_n over $\mathbb{Q}$ for arbitrary $n \in \mathbb{N}$, we get that

$$x^n - 1 = f(x)g(x) \qquad g(x) \in \mathbb{Z}[x]$$

Differentiating and evaluating at ζ_n, we get $n = \zeta_n f'(\zeta_n)g(\zeta_n)$. Taking the norm, we get:

$$n^{\varphi(n)} = \pm \operatorname{disc}(\zeta_n)\operatorname{Nm}(\zeta_n g(\zeta_n))$$

which since $\operatorname{Nm}(\zeta_n g(\zeta_n)) \in \mathbb{Z}$, we get that $\operatorname{disc}(\zeta_n) \big| n^{\varphi(n)}$.

Exercise 22.1.1

1. Show that if F/K is finite and separable, then $\mathrm{tr}_{L/K}(a)$ cannot be 0 for all $a \in F$ (Hint: use Dedekind's lemma, corrollary 18.1.26). Show that this implies the bilienar form $(x, y) \mapsto \mathrm{tr}_{F/K}(xy)$ is non-degenerate.
2. (series of question for Stickelberger's theorem, that $d_K \equiv 0, 1 \bmod 4$)

22.1.3 Abelian Group Structure

We now prove that $\mathcal{O}_K$ is a free $\mathbb{Z}$-module of finite rank. From the above discussion, we can start seeing how this is useful in finding elements of $\mathcal{O}_K$: given it is finitely generated and we know the rank, we can look for generating linearly independent sets, i.e. a basis, for $\mathcal{O}_K$ (since $\mathcal{O}_K$ is a module over a PID).

We in fact can prove something slightly stronger that shall be useful later in this chapter:

Proposition 22.1.26: Generation of Ring of Integers

Let R be a normal integral domain, ($\overline{R}^{K(R)} = R$), $K = K(R)$ the field of fraction of R, and let $S = \overline{R}^F$ where F/K is a separable extension of degree n. Then there exists free A-submodules M' of F such that if $M = R^n$,

$$M \subseteq S \subseteq M'$$

Hence, S is a finitely generated R-module if and only if R is Noetherian, and it is free of rank n if R is a Principal Ideal Domain.

Hence, the integral closure over a larger field that is a finite extension shall only extend the integral closure by a relatively controlled manner; the more condition on R the better the control.

We are usually considering $\mathcal{O}_F$ for some number field $F/\mathbb{Q}$, the generality is to work over the cases that are less nice (as we shall sometimes do) and to have some important theoretical build-up for algebraic geometry. To see this result more concretely, it is worth stating the following immediate corollary:

Corollary 22.1.27: $\mathcal{O}_K$ is a Free $\mathbb{Z}$-Module

Let $F/\mathbb{Q}$ be a number field of degree n. Then $\mathcal{O}_F$ is a free $\mathbb{Z}$-module of degree n.

Proof **:**
Of proposition 22.1.26
Let $K = K(R)$ and let $B = \{\beta_1, \beta_2, ..., \beta_n\}$ be a basis for F over K. We shall use this basis to squish S between two finitely generated R-modules (note that S is an R-module since S/R is an integral extension)

To define the R-module M, as we saw in the proof of proposition 22.1.5, there exists a nonzero $d \in R$ such that $d\beta_i \in S$. Clearly $dB = \{d\beta_1, d\beta_2, ..., d\beta_n\}$ is still a K-basis for F. Now, consider

$$\langle d\beta_1, d\beta_2, ..., d\beta_n \rangle_R = M$$

Since S is naturally an R-module and $d\beta_i \in S$, we have that $M \subseteq S$. Without loss of generality, we may now assume that $\beta_i \in S$ for all $1 \leq i \leq n$, that is $B \subseteq S$ and is a basis for F.

For the second R-module, define: the *integral trace dual*

$$M^{\mathrm{tr}\,\vee} = \left\{ b \in F \;\middle|\; \forall \ell \in M, \mathrm{tr}_{F/K}(b, \ell) \in \mathbb{Z} \right\}$$

Since the trace-pairing is nondegenerate, there exists a dual K-basis $B' = \{\beta_1', ..., \beta_m'\}$ of F given by $\mathrm{tr}(\beta_i \beta_j) = \delta_{ij}$. Notably, $M \cong_{\mathbb{Z}} M^{\mathrm{tr}\,\vee}$

They key result we shall prove is:

$$M = R\beta_1 + R\beta_1 + \cdots + R\beta_n \subseteq S \subseteq R\beta_1' + R\beta_2' + \cdots + R\beta_m' = M^{\vee} \tag{22.3}$$

The first inclusion has already been proved, hence we show the second inclusion. Let $b \in S$ be arbitrary. Since $S \subseteq F$ and B' is a basis for F, we may write b as

$$b = \sum_i b_i \beta_i' \qquad b_i \in K$$

If we show that each $b_i \in R$, we are done. First, Since $\beta_i, b \in S$, their product is in S since S is a ring, $b\beta_i \in S$, hence $\mathrm{tr}_{F/K}(b\beta_i) \in R$ for each i since S is integral over R. Computing, we get:

$$\mathrm{tr}(b\beta_i) = \mathrm{tr}(\sum_j b_j \beta_j' \beta_i) = \sum_j b_j \,\mathrm{tr}(\beta_j' \beta_i) = \sum_j b_j \delta_{ij} = b_i$$

Hence, $b_i \in R$, showing equation (22.3). If R is Noetherian, then M' is a Noetherian R-module, and hence S would be finitely generated. If furthermore R is a Principal Ideal Domain, then any submodule of a free module (over a PID) is free, and S is free of rank less than or equal to n. Since it contains a free R-module of rank n, it has to be free rank greater than or equal to n, and so it has rank n as an R-module, completing the proof.

Note that this implies that if $\mathcal{O}_K$ is a PID and L/K a separable extension, then *every* nonzero finitely generated $\mathcal{O}_L$-module is a free A -module of rank $[L : K]$.

Corollary 22.1.28: Maximality of Ring Of Integers for Number Fields

The ring of integers of a number field L is the largest subring that is finitely generated as a $\mathbb{Z}$-module.

Proof **:**
By the above proposition, $\mathcal{O}_L$ is finitely generated as a $\mathbb{Z}$-module If $B \subseteq L$ is a subring that is also a finitely generated $\mathbb{Z}$-submodule, then every element of B is integral over $\mathbb{Z}$ (proposition 20.5.3(3)). hence, $B \subseteq \mathcal{O}_L$.

With proposition22.1.26, we now have an "upper bound" on how far away a ring of integers is form simply being $\mathbb{Z}[\alpha]$ where $K = \mathbb{Q}(\alpha)$.

Example 22.1.9: Finding Dual
Let $K = \mathbb{Q}(\sqrt{5})$. Then we know that $L = \langle 1, \sqrt{5} \rangle \subseteq \mathcal{O}_K$. Let's try to bound by above the elements.

Let $b = r + s\sqrt{5} \in \mathbb{Q}(\sqrt{5})$ be an arbitrary element. Then:

$$\mathrm{tr}(b \cdot 1) = 2r \qquad r(b\sqrt{5}) = 10s$$

Hence, if $2r \in \mathbb{Z}$ and $10s \in \mathbb{Z}$, we get:

$$L^\vee = \left\langle \frac{1}{2}, \frac{\sqrt{5}}{10} \right\rangle_{\mathbb{Z}}$$

Note that $L^\vee$ is bigger than $\mathcal{O}_K$, for example $1/2 \notin \mathcal{O}_K$. By similar reasoning, we can check that $\frac{1}{\sqrt{5}} \in \mathcal{O}_K^\vee$, but it can be verified that it is not in $\mathcal{O}_K$, and so $\mathcal{O}_K^\vee \neq \mathcal{O}_K$.

Note that separability is a necessary hypothesis to insure the non-degeneracy of the trace-pairing, meaning this result works at least over over number fields, function fields, and finite fields. Naturally, if R is not a Principal Ideal Domain, then S need not be a free R-module. For now, we shall always let $R = \mathbb{Z}$ so that $K = \mathbb{Q}$ and $F/\mathbb{Q}$ is a number field. Note that we are only working with the addition structure of $\mathcal{O}_K$, while $\mathcal{O}_K$ is a ring. Thus, when talking about generators as a $\mathbb{Z}$-module, it must not be confounded with a set of generators for the ring. For example $(\alpha) \subseteq \mathcal{O}_K$ is a principal ideal and hence generated by a single element, but it may be generated by many elements as a $\mathbb{Z}$-module.

Definition 22.1.29: Integral Basis

Let K be a number field. Then a $\mathbb{Z}$-basis for $\mathcal{O}_K$ is called an *integral basis* for K.

The previous proposition also gives more motivation for squaring the determinant in the definition of the discriminant. By corollary 22.1.22, if $(\alpha_1, \alpha_2, ..., \alpha_n)$ is a $\mathbb{Z}$-basis for $\mathcal{O}_F$, then $\mathrm{disc}_{F/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n) \in \mathbb{Z}$ and any change of basis A must map to a unit, hence $\det(A) \in \mathbb{Z}^\times = \{-1, 1\}$. Squaring thus guarantees that the discriminant is basis independent for rings of integers. Now that we have explicitly found that $\mathcal{O}_K$ is a finitely generated $\mathbb{Z}$-module in the case of K being a number field, the term *field discriminant*, or just *discriminant* for short, shall imply that the input is an integral basis for $\mathcal{O}_K$, unless explicitly stated otherwise:

Definition 22.1.30: Discriminant For Ring of Integers

Let K be a number field. Then if $A = \{\alpha_1, \alpha_2, ..., \alpha_n\}$ is an integral basis for $\mathcal{O}_K$, then:

$$d_K := \mathrm{disc}(\mathcal{O}_K) := \mathrm{disc}_{K/\mathbb{Q}}(A) = \mathrm{disc}(\mathcal{O}_K)$$

is the *field discriminant* of K. It is sometimes also denoted $\mathrm{disc}(K)$ if clear from context.

The discriminant can be thought of as capturing some information on how far $\mathcal{O}_K$ is from being

Thus, in example 22.1.7 where $K = \mathbb{Q}(\sqrt{m})$, we have:

$$d_K = \begin{cases} 4m & m \equiv 2, 3 \bmod 4 \\ m & m \equiv 1 \bmod 4 \end{cases}$$

As we saw, we may have many bases for K, not all of them necessarily basis for $\mathcal{O}_K$. Given $K/\mathbb{Q}$ of dimension n, we may have many $M \subseteq B$ of rank n. Since M is of rank n, then if $\mathcal{O}_K \subseteq M$, $\mathcal{O}_K/M$

is of rank 0, and if $M \subseteq \mathcal{O}_K$, then $M/\mathcal{O}_K$ is also of rank 0 (some elementary group theory can be used to prove this, see exercise ref:HERE (fund thm of fin gen ab grp)). This naturally leads to asking how the discriminants of such $\mathbb{Z}$-modules compare:

Proposition 22.1.31: Discriminant And Comparing Basis

Let $B/\mathbb{Z}$ be a free $\mathbb{Z}$-module of rank n, and let $A \subseteq B$ be a $\mathbb{Z}$-submodule of finite index (i.e. $[B : A] < \infty$). Then if $\operatorname{disc}_{B/\mathbb{Z}}(\gamma_1, \gamma_2, ..., \gamma_n) \neq 0$, then:

$$\operatorname{disc}(A) = [B : A]^2 \operatorname{disc}(B)$$

Proof **:**
Let $\beta_1, \beta_2, ..., \beta_n$ is a basis for B. By the previous corollary, if $N \subseteq B$ is of finite index, then it must be a free module of rank n and so the discriminant is nonzero, and similarly the other way around.

Then there exists a change of basis matrix A, and we get $\operatorname{disc}(N) = \det(A)^2 \operatorname{disc}(B)$. It is not hard to see that $\det(A) = [B : N]$ (it will be easier to visualize after finishing section 22.1.4), which complete the proof.

Corollary 22.1.32: Subset And Same Discriminant, Then Equal

Let $F/\mathbb{Q}$ be number field, Let $A, B \subseteq F$ be free $\mathbb{Z}$-modules, $B \subseteq A$ and $\operatorname{disc}(A) = \operatorname{disc}(B)$. Then $B = A$. Hence, we may use discriminants to identify free $\mathbb{Z}$-modules of finite rank.

Proof **:**
Since the discriminants match and

$$\operatorname{disc}(B) = [B : A]^2 \operatorname{disc}(N)$$

we have $[B : A] = 1$, meaning $B = N$.

Corollary 22.1.33: Isomorphism Check Of Number Fields

Let $\mathbb{Q}(\alpha), \mathbb{Q}(\beta)$ be two number fields. Then if the discriminant is different, the number fields are not isomorphic

Proof **:**
If the rings of integers were the same, their index would be 1, but since their discriminant is not equal this is not the case.

Note that it is possible that there are multiple fields with the same discriminant:

Example 22.1.10: Same Discriminant, Different Fields
Take $\mathbb{Q}(\alpha), \mathbb{Q}(\beta), \mathbb{Q}(\gamma)$ where

$$a^3 - 18\alpha - 6 = 0 \qquad \beta^3 - 36\beta - 78 = 0 \qquad \gamma^3 - 58\gamma - 150 = 0$$

Then each has discriminant 22356, but these fields are not isomorphic (as shall be shown in example 22.4.2)

It should be noted that there are only finitely many fields for each discriminant value, however we shall not prove this fact[6]

The next couple of theorems give some computational tools on finding the ring of integers:

Proposition 22.1.34: Computation Of Dual For Ring Of Integers

Let K be a number field with basis $(\alpha_1, \alpha_2, ..., \alpha_n)$ of algebraic integers and let $d = \text{disc}_{K/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n)$. Then every $\alpha \in \mathcal{O}_k$ can be expressed as:

$$\frac{m_1\alpha_1}{d} + \cdots + \frac{m_k\alpha_k}{d}$$

where $m_i \in \mathbb{Z}$ and each m_i^2 is divisible by d $(d|m_i^2)$

Proof **:**
First, $d \neq 0$ since $\alpha_1, \alpha_2, ..., \alpha_n$ forms a basis for K and $d \in \mathbb{Z}$ since the α_i are algebraic integers (which exists by corollary 22.1.6).

Let $\alpha = x_1\alpha_1 + \cdots + x_n\alpha_n$ for some $x_i \in \mathbb{Q}$. Letting $\sigma_1, \sigma_2, ..., \sigma_n$ denote the embeddings of K into $\mathbb{C}$, we get

$$\sigma_i(\alpha) = x_1\sigma(\alpha_1) + \cdots + x_n\sigma(\alpha_n)$$

for each $1 \leq i \leq n$. We can solve for the x_j's using Cramers rule, namely

$$x_j = y_j/\delta$$

where $\delta = \det(\sigma_i(\alpha_j))$ and y_j is obtained by replacing the jth column of the previous determinant by $\sigma_i(\alpha)$. Clearly y_j and δ are algebraic integers, and furthermore $\delta^2 = \text{disc}_{K/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n)$. Hence, letting $d = \text{disc}_{K/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n)$ we have

$$dx_j = \delta y_j$$

showing that the rational number dx_j is an algebraic integer, and so $dx_j \in \mathbb{Z}$. Let $m_i = dx_j$.

What's left to show is that $m_j^2/d \in \mathbb{Z}$, so it suffices to show that it is an algebraic integer. But $m_j^2/d = y_j^2$, which completes the proof.

Hence, for any $\mathcal{O}_K$, if $\alpha_1, \alpha_2, ..., \alpha_n$ is a basis for $K/\mathbb{Q}$

$$\bigoplus \alpha_i\mathbb{Z} \subseteq \mathcal{O}_K \subseteq \bigoplus \frac{\alpha_i}{d}\mathbb{Z}$$

[6]For those interested, it is the *Hermite-Minkowski's Theorem*

And if $K = \mathbb{Q}(\alpha)$ so that $1, \alpha, ..., \alpha^{n-1}$ is a basis for K, then by proposition 22.1.24 $d = \left(\prod_{i \neq j}(\alpha_i - \alpha_j)\right)^2$ and:

$$\mathbb{Z}[\alpha] \subseteq \mathcal{O}_K \subseteq \mathbb{Z}\left[\frac{\alpha}{d}\right]$$

Note that in general, $\mathbb{Z}[\alpha] \neq \mathcal{O}_K$; a counter-example shall be given in example 22.4.4. In the special case where $\mathbb{Z}[\alpha] = \mathcal{O}_K$, we give such a number ring a name:

Definition 22.1.35: Monogenic Number Ring

Let $K/\mathbb{Q}$ be a number field, $\mathcal{O}_K$ be the associated number ring, and α the primitive element so that $K = \mathbb{Q}(\alpha)$. Then if

$$\mathbb{Z}[\alpha] = \mathcal{O}_K$$

Then $\mathcal{O}_K$ is called *monogenic* and is said to be a *monogenic number ring.*

Since all number fields are monogenic, that is $K = \mathbb{Q}(\alpha)$ for some appropriate α, it is reasonable to ask if $\mathcal{O}_K$ can be expressed using α. The following theorem gives this structure:

Theorem 22.1.36: Basis For Ring Of Integers

Let $\alpha \in \mathcal{O}_k$ and suppose α has degree n over $\mathbb{Q}$. Then there exists $d_1, d_2, ..., d_{n-1} \in \mathbb{Z}$, and monic polynomial $f_i \in \mathbb{Z}[x]$ of degree i for $1 \leq i \leq n-1$ such that

$$\mathcal{O}_K = \mathbb{Z} \oplus \frac{f_1(\alpha)}{d_1}\mathbb{Z} \oplus \cdots \oplus \frac{f_{n-1}(\alpha)}{d_{n-1}}\mathbb{Z}$$

is an integral basis where

$$d_1 \big| d_2 \big| \cdots \big| d_{n-1}$$

and the d_i's are unique.

Proof **:**
Marcus p.26

Example 22.1.11: Ring Of Integers Of Cubic Extension

Let $K = \mathbb{Q}(\sqrt[3]{2})$ and $\alpha = \sqrt[3]{2}$. We want to find $\mathcal{O}_K$. First $\operatorname{rank} \mathcal{O}_K = 3$ since $[K : \mathbb{Q}] = 3$, and $L = \mathbb{Z}[\sqrt[3]{2}] \subseteq \mathcal{O}_K$. If A is the change of basis from L to $\mathcal{O}_K$, then $\det(A) = [\mathcal{O}_K : \mathbb{Z}[\alpha]]^a$. Thus, by doing a change of basis from $\mathbb{Z}[\alpha]$ to $\mathcal{O}_K$ by A, we get that:

$$\operatorname{disc}(\mathbb{Z}[\alpha]) = [\mathcal{O}_K : \mathbb{Z}[\alpha]]^2 \operatorname{disc}(\mathcal{O}_K)$$

Thus, computing the discriminant of the left hand side we get:

$$\operatorname{disc}(\mathbb{Z}[\alpha]) = \det \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 6 \\ 0 & 6 & 0 \end{pmatrix} = -3 \cdot 36 = -(2^2 \cdot 3^2) \cdot 3 = -108$$

giving us that $[\mathcal{O}_K : \mathbb{Z}[\alpha]]^2 \big| 108 \Rightarrow [\mathcal{O}_K : \mathbb{Z}[\alpha]] \big| 2 \cdot 3 = 6$. If we show the index is 1, the proof is done. We shall prove the contradiction for the case of $[\mathcal{O}_K : \mathbb{Z}[\alpha]] = 3$, the proof for $[\mathcal{O}_K : \mathbb{Z}[\alpha]] = 2$ is similar, and from that we can also conclude that $[\mathcal{O}_K : \mathbb{Z}[\alpha]] \nmid 6$.

Let's say $[\mathcal{O}_K : \mathbb{Z}[\alpha]] = 3$. Then $\mathcal{O}_K/\mathbb{Z}[\alpha] \cong \mathbb{Z}/3\mathbb{Z}$. Any element $\beta \in \mathcal{O}_K$ lands in one of the cosets of $\mathcal{O}_K/\mathbb{Z}[\alpha]$ which implies that $a, b, c \in \frac{1}{3}\mathbb{Z}$. We want to show that a, b, c must be integers (lands in the zero coset). Since the ring of integers is closed under subtraction, we may without loss of generality subtract the non-fractional part of any element β we choose so that $a, b, c \in \{-1, 0, 1\}$[b]. Hence we take the norm of $\beta = (a/3) + (b/3)\alpha + (c/3)\alpha^2$

$$N(\beta) = \det \begin{pmatrix} a/3 & 2c/3 & 2b/3 \\ b/3 & a/3 & 2c/3 \\ c/3 & b/3 & a/3 \end{pmatrix} = \frac{1}{27}(a^3 + 2b^2 + 4c^3 - 6abc)$$

Since β is algebraic, $N(\beta) \in \mathbb{Z}$ meaning $27 \big| (a^3 + 2b^2 + 4c^3 - 6abc)$. But, unless $a = b = c = 0$:

$$|a^3 + 2b^2 + 4c^3 - 6abc| \leq 15$$

showing that no such combination of a, b, c exist, meaning β is an algebraic integer if $a, b, c \in \mathbb{Z}$. A similar argument can be made in the case of $[\mathcal{O}_K : \mathbb{Z}[\alpha]] = 2$, hence we have that $\mathcal{O}_K = \mathbb{Z}[\alpha]$, completing the proof.

[a] since the index is finite, a basis for $\mathbb{Z}[\alpha]$ may be scaled to form a basis for $\mathcal{O}_K$, which translates to the determinant result

[b] Equivalently, any coset can be chosen to have such representatives

In some special cases, the ring of integers of more composite fields may be found given ring of integers of each respective field:

Theorem 22.1.37: Ring of Integers of Composites

Let $K_1/K, K_2/K$ be Galois with degree n and m and with coprime discriminants $d_{K_1} = d$ and $d_{K_2} = d'$ respectively. Let $L = K_1K_2$ and $K_1 \cap K_2 = K$. Let $\omega_1, \omega_2, ..., \omega_n$ and $\omega'_1, \omega'_2, ..., \omega'_m$ be an integral basis for K_1/K and K_2/K. Then $\omega_i\omega'_j$ is an integral basis for K_1K_2 with discriminant $(d_{K_1})^m(d_{K_2})^n = d^m(d')^n$, namely

$$\mathcal{O}_L = \mathcal{O}_{K_1}\mathcal{O}_{K_2}$$

Proof **:**
Since $K_1 \cap K_2 = K$, $[K_1K_2 : K] = nm$, so the nm products $\omega_i\omega'_j$ form a basis of K_1K_2/K. Now, let α be an integral element of K_1K_2 over K, and write:

$$\alpha = \sum_{i,j} a_{ij}\omega_i\omega'_j \qquad a_{ij} \in K$$

We must show that $a_{ij} \in \mathcal{O}_K$. Let

$$\beta_j = \sum_i a_{ij}\omega_i$$

$$\mathrm{Gal}_{K_2}(L) = \{\sigma_1, \sigma_2, ..., \sigma_n\}$$
$$\mathrm{Gal}_{K_1}(L) = \{\tau_1, \tau_2, ..., \tau_m\}$$

Then since $[L : K] = nm$, by proposition 18.1.20$\mathrm{Gal}_K(L) = \{\sigma_k\tau_\ell \mid 1 \leq k \leq n \; 1 \leq \ell \leq m\}$. Furthermore, let:

$$T = \tau_\ell(\omega'_j) \qquad a = (\tau_1(\alpha), ..., \tau_m(\alpha))^t \qquad b = (\beta_1, ..., \beta_m)^t$$

Then by definition we have $\det(T)^2 = d'$ and $a = Tb$. If T^* is the adjoint of T, then some linear algebra manipulation gives:

$$\det(T)b = T^*a$$

Now, since T^* and a have integral entries (in L), $d'b$ has integral entries in K_1¡ i.e.

$$d'\beta_j = \sum_k d/a_{ij}\omega_i$$

Hence, $d'a_{ij} \in \mathcal{O}_K$. Now, changing (ω_i) and $\omega'_j)$, we get that $da_{ij} \in A$, hence since the discriminants are relatively prime, we get for appropriate x, x':

$$a_{ij}xda_{ij} + x'd'a_{ij} \in \mathcal{O}_K$$

Hence, $\omega_i\omega'_j$ form an integral basis for L/K, completing the first part of the proof.

We next find the discriminant the ring of integers of L/K. Since we have the Galois group of L over K, take the $(nm \times m)$ matrix

$$M = (\sigma_k\tau_\ell(\omega_i\omega'_j)) = (\sigma_k(\omega_i)\tau_\ell(\omega'_j))$$

Note that M can be interpreted as an $(m \times m)$-matrix with entries being $(n \times n)$-matrices where each (ℓ, j)-entry is the matrix $Q\tau_\ell\omega'_j$ where $Q = (\sigma_k\omega_i)$:

$$M = \begin{pmatrix} Q & & 0 \\ & \ddots & \\ 0 & & Q \end{pmatrix} \begin{pmatrix} E\tau_1\omega'_1 & \cdots & E\tau_m(\omega'_1) \\ \vdots & \ddots & \vdots \\ E\tau_1(\omega'_m) & \cdots & E\tau_m(\omega'_m) \end{pmatrix}$$

where E is an $(n \times n)$-unit matrix. Now, changing indices to compare the two representations of M, we get:

$$\operatorname{disc}(\mathcal{O}_L) = \det(M)^2 = \det(Q)^{2m} = \det\left((\tau_\ell(\omega'_j))\right)^{2n} = d^m(d')^n$$

as we sought to show.

Note that for the general separable case, we need to replace our assumption with

$$K_1 \cap K_1 = K \qquad K_1,\ K_2 \text{ are linearly disjoint}$$

If the discriminants are not relatively prime, then there is a partial generalization in the case where $K = \mathbb{Q}$:

Corollary 22.1.38: Ring of Integers of Composite, Partial Generalization

Let $K_1/\mathbb{Q}$ and $K_2/\mathbb{Q}$ be finite separable extensions, $d_1 = \operatorname{disc}(\mathcal{O}_{K_1})$ and $d_2 = \operatorname{disc}(\mathcal{O}_{K_2})$. Then if $L = K_1K_2$:

$$\mathcal{O}_L \subseteq \frac{1}{\gcd(d_1, d_2)}\mathcal{O}_{K_1}\mathcal{O}_{K_2}$$

Proof **:**
Marcus p.25

Example: Ring of Integers of Cyclotomic Field

Let's use the above results to show that if $K = \mathbb{Q}(\zeta_n)$ then $\mathcal{O}_K = \mathbb{Z}[\zeta_n]$, that is the ring of integers of K is monogenic. Besides quadratic extensions, this is so far the only other ring of integers we have classified (in particular since they are rather difficult to classify). We first require some lemmas:

Lemma 22.1.39: Cyclotomic Field Lemmas

Let $n = p^r$ where $n \geq 3$. Then:

1. $\mathbb{Z}[\zeta_n] = \mathbb{Z}[1 - \zeta_n]$ and $\operatorname{disc}(\zeta_n) = \operatorname{disc}(1 - \zeta_n)$
2. $\prod_k (1 - \zeta_n^k) = p$

Proof **:**

1. Certainly $\mathbb{Z}[\zeta_n] = \mathbb{Z}[1 - \zeta_n]$ since $\zeta_n = 1 - (1 - \zeta_n)$, implying the discriminants are equal. Similarly, by proposition 22.1.24
$$\operatorname{disc}(\zeta_n) = \prod_{r<s} (\alpha_r - \alpha_s)^2 = \prod_{r<s} ((1 - \alpha_r) - (1 - \alpha_s))^2 = \operatorname{disc}(1 - \zeta_n)$$
2. Let
$$f(x) = \frac{x^{p^r - 1}}{x^{p^{r-1}} - 1} = 1 + x^{p^{r-1}} + x^{2p^{r-1}} + \cdots x^{(p-1)p^{r-1}}$$
Then for $p \nmid k$, ζ_n^k are the roofs of f, since they are roots of $x^{p^r} - 1$, but not of $x^{p^{r-1}} - 1$. Thus:
$$f(x) = \prod_{p \nmid k} (x - \zeta_n^k)$$
Since there is exactly $\varphi(p^r) = (p-1)p^{r-1}$ where φ is the Euler totient. Thus, setting $x = 1$ gives us the desired result.

Proposition 22.1.40: Ring Of Integers Of Prime-Power Cyclotomic Field

Let $n = p^r$ and $K = \mathbb{Q}(\zeta_n)$. Then:
$$\mathcal{O}_K = \mathbb{Z}[\zeta_n]$$

Proof **:**
We shall show that $\mathcal{O}_K = \mathbb{Z}[1 - \zeta_n]$. By theorem 22.1.36, any $\alpha \in \mathcal{O}_K$ can be written as:
$$\alpha = \frac{n_1 + n_2(1 - \zeta_n) + \cdots + n_m(1 - \zeta_n)^{m-1}}{\operatorname{disc}(1 - \zeta_n)}$$
where $m = \varphi(p^r)$, $n_i \in \mathbb{Z}$, and $\operatorname{disc}(1 - \zeta_n) = \operatorname{disc}(\zeta_n)$. As we saw in example 22.1.8, if $n = p$ is a prime, $\operatorname{disc}(\zeta_n) = p^{p-2}$. Since $n = p^r$, by example 22.1.8, d is a power of p.

Let's now assume for the sake of contradiction that $\mathcal{O}_K \neq \mathbb{Z}[1 - \zeta_n]$, so that there must exist some

$\alpha \in \mathcal{O}_K$ such that not all n_i are divisible by d, namely $\mathcal{O}_K$ contains an element of the form:

$$\beta = \frac{n_i(1-\zeta_n)^{i-1} + n_{i+1}(1-\zeta_n)^i + \cdots + n_m(1-\zeta_n)^{m-1}}{p}$$

where $i \leq n$, $n_i \in \mathbb{Z}$, not all n_i are divisible by p. Now, by lemma 22.1.39 since $(1-\zeta_n)\big|(1-\zeta_n)^k$ in $\mathbb{Z}[\zeta_n]$,

$$\frac{p}{(1-\zeta_n)^m} \in \mathbb{Z}[\zeta_n] \implies \frac{p}{(1-\zeta_n)^i} \in \mathbb{Z}[\zeta_n] \implies \frac{\beta p}{(1-\zeta_n)^i} \in \mathcal{O}_K$$

Subtracting the terms that are in $\mathcal{O}_K$ (by closure of multiplication) we get

$$\frac{n_i}{1-\zeta_n} \in \mathcal{O}_K \implies \mathrm{Nm}_{K/\mathbb{Q}}\left(\frac{n_i}{1-\zeta_n}\right) \in \mathbb{Z} \implies \mathrm{Nm}_{K/\mathbb{Q}}(1-\zeta_n)\big|\,\mathrm{Nm}_{K/\mathbb{Q}}(n_i)$$

However, this is a contradiction since $\mathrm{Nm}_{K/\mathbb{Q}}(n_i) = n_i^m$, but by lemma 22.1.39 $\mathrm{Nm}_{K/\mathbb{Q}}(1-\zeta_n) = p$, Hence

$$\mathcal{O}_K = \mathbb{Z}[\zeta_n]$$

as we sought to show.

We now look to generalize to any cyclotomic field. The key idea is that the ring of integers of cyclotomic fields will be the composite of simpler ring of integers

Proposition 22.1.41: Ring of Integers of Cyclotomic Field

Let $K = \mathbb{Q}(\zeta_n)$ for any $n \in \mathbb{N}$. Then:

$$\mathcal{O}_K \cong \mathbb{Z}[\zeta_n]$$

***Proof* :**
We shall do a form of induction. This was proven if n is a prime power, so let's assume it's not so that we may write $n = n_1 n_2$ for relatively prime $n_1, n_2 > 1$. We shall work by induction on n_1, n_2. The base case was proposition 22.1.40, hence set:

$$\begin{aligned} \omega_1 &= \zeta_{n_1} & \omega_2 &= \zeta_{n_2} \\ K_1 &= \mathbb{Q}(\omega_1) & K_2 &= \mathbb{Q}(\omega_2) \\ \mathcal{O}_{K_1} &= \mathbb{Z}[\omega_1] & \mathcal{O}_{K_2} &= \mathbb{Z}[\omega_2] \end{aligned}$$

First, it is clear that

$$\zeta_n^{n_1} = \omega_2 \qquad \zeta_n^{n_2} = \omega_1 \implies \zeta_n = \omega_1^r \omega_2^s,\ r, s \in \mathbb{Z}$$

Thus, $L = K_1 K_2$. Next, as we saw in example 22.1.8,

$$\mathrm{disc}(\omega_1)\big|n_1 \qquad \mathrm{disc}(\omega_2)\big|n_2$$

since $\gcd(n_1, n_2) = 1$, we see the discriminants are relatively prime. Since $\varphi(n) = \varphi(n_1)\varphi(n_2)$, we by theorem 22.1.37 that

$$\mathcal{O}_K = \mathcal{O}_{K_1}\mathcal{O}_{K_2} = \mathbb{Z}[\omega_1]\mathbb{Z}[\omega_2] = \mathbb{Z}[\zeta_n]$$

as we sought to show.

Exercise 22.1.2

1. Generalize proposition 22.1.31 for modules over $\mathcal{O}_K$.
2. Give a counter-example where $\mathcal{O}_{KL} = \mathcal{O}_K\mathcal{O}_L$?

22.1.4 Lattice Structure: Geometry of Number Rings

An important word of warning must be given about free $\mathbb{Z}$-modules due to the existence of non-invertible elements. It is possible that A, B are two free $\mathbb{Z}$-module of rank 2, one contains the basis of the other, however $A \subsetneq B$! This is in contract to the case of modules over fields where if A, B are two k-vector-spaces of the same dimension and $A \subseteq B$, then $A = B$[7]. Even more behavior can happen: consider $\mathbb{Q} \otimes_{\mathbb{Z}} A$. Then this is a vector-space with a finite dimension. However, the dimension need not match the rank of A:

Example 22.1.12: Behavior of $\mathbb{Z}$-modules

1. Let $A = \langle (1,0),(0,1)\rangle_{\mathbb{Z}}$ and $B = \langle (1/2,0),(0,1/2)\rangle_{\mathbb{Z}}$. Then by construction they are both isomorphic as abelian groups to $\mathbb{Z}^2$, however it is clear that $A \subsetneq B$.
2. Let A be a free $\mathbb{Z}$-module of rank 2. Then we may extend scalar to make A into a free $\mathbb{R}$-module; either take $\mathbb{R} \otimes_{\mathbb{Z}} A$ or consider $\langle a_1, a_1\rangle_{\mathbb{R}}$ for some free generating set $\{a_1, a_2\}$. Then $\mathbb{R} \otimes_{\mathbb{Z}} A$ *need not be dimension 2*. For example, take
$$A = \langle (1,1),(\sqrt{2},\sqrt{2})\rangle_{\mathbb{Z}}$$
Then $\mathbb{R} \otimes_{\mathbb{Z}} A$ is one dimensional
3. By construction of $\mathcal{O}_k$, if $\langle \alpha_1, \alpha_2, ..., \alpha_n\rangle_{\mathbb{Z}} = \mathcal{O}_k$, then $\langle \alpha_1, \alpha_2, ..., \alpha_n\rangle_{\mathbb{Q}} = k$, Thus $\mathcal{O}_k$ is a little more than just a $\mathbb{Z}$-module. We shall explore this in this section.

The third example shows that the rank of the ring of integers more coincides with the dimension of the natural vector-space when extending the actions to a field. As was explored in chapter 13, [finite dimensional] vector spaces have easy visualizations using orthonormal bases. In chapter 10, this was used implicitly when it was intuited that $\mathbb{Z}[i]$ forms a nice square lattice in $\mathbb{Q}^2$. More generally, ring of integer $\mathcal{O}_K$ will form a lattice when embedded into an appropriate vector-space. To better visualize this lattice, it is easer to see it embedded into a direct sum of $\mathbb{R}$ and $\mathbb{C}$[8]. This has the extra advantage that these spaces have a well-defined notion of *measure* that gives us an area. We shall take advantage of tools from analysis to further enlighten the structure of rings of integers, however as a consequence an elementary level of topology and analysis will be assumed.

Let $n = [K : \mathbb{Q}]$ so that $\mathrm{Hom}_{\mathbb{Q}}(K, \mathbb{C}) = n$. Then every embedding of K into $\mathbb{C}$ is either embeds into $\mathbb{R}$ or does not (that is, either $K \cap (\mathbb{C} \setminus \mathbb{R}) = \emptyset$ or $\neq 0$); let's call the former *real embeddings* and the latter *complex embeddings*. For a complex embedding $\tau \in \mathrm{Hom}_{\mathbb{Q}}(K, \mathbb{C})$, we may always find another

[7] $b \in B$ can be appropriately shrunk to a nonzero size so that its in A (since the dimensions are the same), and then we use closure of scalar multiplication

[8] For those who may prefer $\overline{\mathbb{Q}}$ over $\mathbb{R}$ or $\mathbb{C}$, I would stipulate $\mathbb{R}$ and $\mathbb{C}$ are is a very natural object in this context: $\mathbb{R}$ is perhaps the only "natural" object for which a finite extension is an algebraic closure, this will have algebraic and geometric consequence when counting the number of embeddings of a field into the algebraically closed field $\mathbb{C}$

embedding $\overline{\tau}$ since for any α where $p(\alpha) = 0$ for appropriate $p(x) \in \mathbb{Q}[x]$

$$0 = \overline{\tau}(0) = \overline{\tau}(p(\alpha)) = \tau(\overline{p(\alpha)}) = \tau(p(\overline{\alpha}))$$

showing that $\overline{\tau}$ maps roots to roots. Hence the degree n may be split into $n = r_1 + 2r_2$ where r_1 represents the amount of embeddings where the image σ_i is real:

$$\alpha_1, \alpha_2, ..., \alpha_{r_1}$$

while r_2 represents the embeddings that embedded into $\mathbb{C}$ is not strictly in $\mathbb{R}$, giving us a total list of embeddings:

$$\sigma_1, \sigma_2, ..., \sigma_{r_1}, \tau_1, \tau_2, ..., \tau_{r_2}, \overline{\tau_1}, ..., \overline{\tau_{r_2}}$$

Definition 22.1.42: Archimedean Embedding

Let $K/\mathbb{Q}$ be a number field. Then its *Archimedean embedding* is defined as

$$\begin{aligned} \sigma :& K \hookrightarrow \mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2} \\ & \alpha \mapsto (\sigma_1(\alpha), ..., \tau_{r_2}(\alpha)) \end{aligned}$$

The space $\mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2}$ is called the *Minkowski space*, and will be labeled $K_{\mathbb{R}}$.

Example 22.1.13: Archimedean Embedding

1. Trivially, $\mathbb{Q}_{\mathbb{R}} = \mathbb{R}$, and $\mathbb{C}_{\mathbb{R}}$ is not defined since $\mathbb{C}/\mathbb{Q}$ is not a finite extension.
2. Take $K = \mathbb{Q}(i)$ so that $\mathcal{O}_K = \mathbb{Z}[i]$. Then $\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(i)) = \{\mathrm{id}, c\}$ where c is the conjugate. Then K is simply the embedding of K into $\mathbb{C}$

   $$\sigma : K \hookrightarrow \mathbb{C}^1 \qquad k \mapsto k$$

   Notice the gaussian integers form the usual grid in $\mathbb{C}$.
3. Take $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$ Then $\mathrm{Gal}_{\mathbb{Q}}(K) \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ corresponds to the conjugating each root separately. Then

   $$\begin{gathered} \sigma : K \mapsto \mathbb{R}^4 \\ (a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}) \\ \mapsto \\ (a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}, a - b\sqrt{2} + c\sqrt{3} + d\sqrt{6}, a + b\sqrt{2} - c\sqrt{3} + d\sqrt{6}, a - b\sqrt{2} - c\sqrt{3} + d\sqrt{6}) \end{gathered}$$

Our goal is to show that $\sigma(\mathcal{O}_K)$ behaves like the just like $\mathbb{Z}[i]$ when embedded in $\mathbb{C}$. To characterize a lattice in higher dimensions, we define it in the following way:

Definition 22.1.43: [Integral] Lattice

Let V be a finite dimensional vector space over$\mathbb{R}$. Then a *lattice* in V is a $\mathbb{Z}$-submodule satisfying any two of the following 3 properties;

1. Λ spans V as a k-vector space ($\langle \Lambda \rangle_{\mathbb{R}} = V$)
2. Λ is discrete in the topological space V (there is an open ball near the origin that intersects only the origin) and the rank of the lattice is the dimension of V[a]
3. Λ is a free $\mathbb{R}$-module on n generators

[a] some authors call lattices where $n = m$ *complete lattices*, and if $n < m$ a *lattices*

Lemma 22.1.44: Equivalence Of Conditions

Any two of these conditions implies the third

Proof **:**
Let $L \cong_{\mathbf{Ab}} \mathbb{Z}^n$. We'll show that (2) $\Leftrightarrow$ (3). For (2) $\Rightarrow$ (1), namely

(2) L is discrete

(3) $\langle L \rangle_{\mathbb{R}} = \mathbb{R}^n$

let $\ell_1, \ell_2, ..., \ell_n$ be a $\mathbb{Z}$-basis for L. Then there exists an invertible matrix $A \in \mathrm{GL}_n(\mathbb{R})$ such that $A = (a_{ij})$ where for each e_i of the standard basis for $\mathbb{R}^n$,

$$e_i = \sum_j a_{ij} \ell_j$$

Let $v = (v_1, v_2, ..., v_n) \in \mathbb{Z}^n$. Then

$$\sum_i^n v_i \ell_i = \sum_i^n w_i e_i$$

where $w = A^{-1} v$. We have now converted our number into one in the standard basis, hence we have a notion of size we may borrow. In particular, we have that if $v = Aw$, then

$$\|w\| \geq \frac{\|v\|}{n\|A\|}$$

where we can choose our favourite matrix norm, say the max norm on entries. hence, as v grows, w get's further from 0.

Conversely, we'll do the contrapositive, so suppose $\langle \ell_1, \ell_2, ..., \ell_n \rangle_{\mathbb{R}} \neq \mathbb{R}^n$. Then for some $a_i \in \mathbb{R}$ not all zero

$$\sum_i^n a_i \ell_i = 0$$

By the Pigeon hole Principle[a] (also known as Dirichlet's box principle), there exists $N \in \mathbb{N}, N > 0$ such that $|Na_i - [Na_i]| < \epsilon$ where $[Na_i]$ is the closest integer. Thus

$$0 = \sum Na_i \ell_i = \underbrace{\sum [Na_i]\ell_i}_{\in L} + \underbrace{\sum \{Na_i\}\ell_i}_{\in B_\epsilon(0)}$$

where $\{Na_i\}$ denote the fractional part, completing the proof.

[a] see section 0.1.3

The discrete condition need only be checked at the origin due to linearity[9]. An example of a free $\mathbb{Z}$-module in $\mathbb{R}^2$ that is not discrete is the one elaborated in example 22.1.12

$$\langle (1,1), (\sqrt{2}, \sqrt{2}) \rangle \subseteq \mathbb{R}^2$$

In particular, we may get as close as we want to $(0,0)$ by adding integer-multiples these two elements.

Proposition 22.1.45: Archimedean Embedding of $\mathcal{O}_K$ is Lattice

Let $K/\mathbb{Q}$ be a number field and σ the Archimedean embedding. Then $\sigma(\mathcal{O}_k)$ is a lattice in $\mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2}$, namely

$$\langle \sigma(\mathcal{O}_K) \rangle_{\mathbb{R}} = \mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2}$$

Proof **:**
We may prove this result in using any of the equivalent definitions.

1. Let's show $\sigma(\mathcal{O}_K)$ is discrete. Let $\alpha \in K$ so that

$$\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) = \prod_i^{r_1} \sigma_i(\alpha) \cdot \prod_j^{r_2} \|\tau_j(\alpha)\|^2 \in \mathbb{Q}$$

   Then if $\alpha \in \mathcal{O}_K$, $\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) \in \mathbb{Z}$. Therefore, if $\alpha \in \mathcal{O}_K \setminus \{0\}$, we have $|\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)| \geq 1$. But then $|\sigma(\alpha)| \geq 1$. Thus, taking $B \subseteq \mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2}$,

$$B = \{\alpha \mid |\alpha_i| < 1,\ 1 \leq i \leq r_1 + r_2\}$$

   Then $\sigma(\mathcal{O}_K) \cap B = \{0\}$

2. Let's show that the $\mathbb{R}$-span of $\sigma(\mathcal{O}_K)$ is our entire space. Since $\mathcal{O}_K$ is $\mathbb{Z}$-finitely generated we have

$$\langle \alpha_1, \alpha_2, ..., \alpha_n \rangle_{\mathbb{Z}} = \mathcal{O}_K$$

   We want to show that

$$\mathrm{span}_{\mathbb{R}}(\sigma_1(\alpha_i), ..., \sigma_{r_1}(\alpha_i), .., \tau_{r_2}(\alpha_i))_{i \in \{1,...,n\}} = \mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2}$$

[9] This is reminiscent to how many conditions for topological groups need only be checked at the origin, including many proofs in real and complex analysis

Then the determinant of these values would be nonzero:

$$0 \neq \det \begin{pmatrix} \sigma_1(\alpha_1) & \cdots & \sigma_{r_1}(\alpha_1) & \operatorname{Re}\tau_1(\alpha_1) & \operatorname{Im}(\tau_1(\alpha_1)) & \cdots & \operatorname{Im}\tau_{r_2}(\alpha_1) \\ \sigma_1(\alpha_2) & \cdots & \sigma_{r_1}(\alpha_2) & \operatorname{Re}\tau_2(\alpha_2) & \operatorname{Im}(\tau_2(\alpha_2)) & \cdots & \operatorname{Im}\tau_{r_2}(\alpha_2) \\ & & & \vdots & & & \\ & & & \vdots & & & \\ & & & \vdots & & & \end{pmatrix}$$

By doing some elementary operations, we get the matrix to be of the form:

$$0 \neq \det \begin{pmatrix} \sigma_1(\alpha_1) & \cdots & \sigma_{r_1}(\alpha_1) & \tau_1(\alpha_1) & \overline{\tau_1(\alpha_1)} & \cdots & \overline{\tau_{r_2}(\alpha_1)} \\ \sigma_1(\alpha_2) & \cdots & \sigma_{r_1}(\alpha_2) & \tau_2(\alpha_2) & \overline{\tau_2(\alpha_2)} & \cdots & \overline{\tau_{r_2}(\alpha_2)} \\ & & & \vdots & & & \\ & & & \vdots & & & \\ & & & \vdots & & & \end{pmatrix}$$

Now, choose a primitive element $\alpha \in K$ so that $K = Q(\alpha)$. Then we know that $\mathbb{Z}[\alpha] = \langle 1, \alpha, ..., \alpha^{n-1} \rangle_{\mathbb{Z}} \subseteq \mathcal{O}_K$. We know this is of finite index. Letting $\alpha_i = \alpha^{i-1}$, we get a Vandermonde matrix. If M is the above matrix:

$$\det(M) = \prod_{\substack{\sigma \neq \sigma' \\ \sigma, \sigma' \in \operatorname{Emb}_K(\mathbb{C})}} (\sigma(\alpha) - \sigma'(\alpha)) \neq 0$$

But then this implies that the volume of $\sigma(\mathcal{O}_K)$ is non-zero, as we sought to show.

Corollary 22.1.46: Sign Of Discriminant

Let $K/\mathbb{Q}$ be a number field, $\mathcal{O}_K$ its ring of integers, its discriminant. Then:

$$\operatorname{sign}(d_K) = (-1)^{r_2}$$

Proof **:**
Simply keep track of the elementary operations when transitioning between both matrices in

proposition: 22.1.45

$$\det \begin{pmatrix} \sigma_1(\alpha_1) & \cdots & \sigma_{r_1}(\alpha_1) & \tau_1(\alpha_1) & \overline{\tau_1(\alpha_1)} & \cdots & \overline{\tau_{r_2}(\alpha_1)} \\ \sigma_1(\alpha_2) & \cdots & \sigma_{r_1}(\alpha_2) & \tau_2(\alpha_2) & \overline{\tau_2(\alpha_2)} & \cdots & \overline{\tau_{r_2}(\alpha_2)} \\ & & & \vdots & & & \\ & & & \vdots & & & \\ & & & \vdots & & & \end{pmatrix}$$

$$= \frac{1}{(2i)^{r_2}} \det \begin{pmatrix} \sigma_1(\alpha_1) & \cdots & \sigma_{r_1}(\alpha_1) & \operatorname{Re}\tau_1(\alpha_1) & \operatorname{Im}(\tau_1(\alpha_1)) & \cdots & \operatorname{Im}\tau_{r_2}(\alpha_1) \\ \sigma_1(\alpha_2) & \cdots & \sigma_{r_1}(\alpha_2) & \operatorname{Re}\tau_2(\alpha_2) & \operatorname{Im}(\tau_2(\alpha_2)) & \cdots & \operatorname{Im}\tau_{r_2}(\alpha_2) \\ & & & \vdots & & & \\ & & & \vdots & & & \\ & & & \vdots & & & \end{pmatrix}$$

completing the proof.

Hence, any ring of integers $\mathcal{O}_K$ forms a lattice in $K_{\mathbb{R}}$, and any ideal $\mathfrak{a} \subseteq \mathcal{O}_K$ will also form a lattice in $K_{\mathbb{R}}$. Since $K_{\mathbb{R}}$ is an inner product and a lattice is discrete, it is meaningful to define the volume of a lattice:

Definition 22.1.47: Volume of Lattice

let $\mathcal{O}_K$ be a ring of integers. Then if $\Gamma = (v_1, v_2, ..., v_n)$ is an integral basis for $\mathcal{O}_K$, then the *volume* of $\mathcal{O}_K$ is defined to be

$$\operatorname{vol}(\mathcal{O}_K) := \operatorname{vol}(K_{\mathbb{R}}/\Gamma) = \operatorname{coVol}(\Gamma) = |\det(\langle v_i, v_j \rangle)|^{1/2} =$$

where $\frac{1}{2^{r_2}}$ comes from the extra complex embeddings, as seen in the proof of corollary 22.1.46.

If you know some differential geometry, we may equivalently take the canonical differential form on $\mathbb{R}^n \cong \mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2}$:

$$\omega = dx_1 \wedge \cdots \wedge dx_n$$

and write:

$$\operatorname{vol}(\mathcal{O}_K) = \operatorname{coVol}(K_{\mathbb{R}}/\mathcal{O}_K) = \frac{1}{2^{r_2}} \left| \int_{K_{\mathbb{R}}/\mathcal{O}_K} \omega \right| = \frac{1}{2^{r_2}} \left| \int_P \omega \right|$$

where the $1/2^{r_2}$ comes from eliminating the extra complex dimension, and:

$$P = \left\{ \sum a_i v_i \;\middle|\; 0 \leq a_i \leq 1 \right\}$$

This also motivates the vocabulary of covolume. Note that any different choice of a basis will produce the same result since the change of basis matrix B will have determinant ± 1 which in the absolute value becomes 1, and hence it is well-defined. Note that every ideal $\alpha \subseteq \mathcal{O}_K$ will also produce a lattice, since it's a sublattice of $\mathcal{O}_K$, and hence we may find the volume of any sublattice. The way we shall find the volume of the lattice is by taking advantage of the discriminant and the index of the ideal:

Proposition 22.1.48: Volume of Ideal Lattice

Let $0 \neq \mathfrak{a} \subseteq \mathcal{O}_K$ be an ideal. Then $\Gamma_{\mathfrak{a}} = \sigma(\mathfrak{a})$ is a [complete] lattice in K_R and its fundamental mesh has volume

$$\mathrm{vol}(\Gamma_{\mathfrak{a}}) = \frac{1}{2^{r_2}}\sqrt{|\operatorname{disc}(K)}[\mathcal{O}_K : \mathfrak{a}]$$

As a consequence:

$$\mathrm{vol}(\mathcal{O}_K) = \frac{1}{2^{r_2}}\sqrt{|\operatorname{disc}(\mathcal{O}_K)}$$

Thus, the discriminant provides volume information for the lattice given by the ring of integers.

The theorem can equivalently be stated as:

$$\mathrm{vol}(\Gamma_{\mathfrak{a}}) = \left|\int_{\mathbb{R}^{r_1}\oplus\mathbb{R}^{r_2}/\Gamma_{\mathfrak{a}}} \omega\right|$$

Proof **:**
Let $\alpha_1, \alpha_2, ..., \alpha_n$ be a $\mathbb{Z}$-basis of $\mathfrak{a}$ so that $\Gamma_{\mathfrak{a}} = \mathbb{Z}\sigma(\alpha_1) + \cdots + \mathbb{Z}\sigma(\alpha_n)$ is a lattice. Numbering the embeddings $\tau K \to \mathbb{C}$ as $\tau_1, \tau_2, ..., \tau_n$, define the matrix $A = (\tau_\ell \alpha_i)$ so that we get by proposition 22.1.31,

$$\operatorname{disc}(\mathfrak{a}) = \operatorname{disc}(\alpha_1, \alpha_2, ..., \alpha_n) = \det(A)^2 = [\mathcal{O}_K : \mathfrak{a}]^2 \operatorname{disc}(\mathcal{O}_K) = [\mathcal{O}_K : \mathfrak{a}]^2 d_K$$

On the other hand, we have:

$$(\langle\sigma(\alpha_i), \sigma(\alpha_j)\rangle)_{ij} = \left(\sum_{\ell=1}^{n} \tau_\ell(\alpha_i)\overline{\tau}_\ell\alpha_j\right) = A\overline{A}^t$$

Thus:

$$\mathrm{vol}(\Gamma_{\mathfrak{a}}) = \left|\det\left(\langle\sigma(\alpha_i), \sigma(\alpha_j)\rangle\right)_{ij}\right|^{1/2} = |\det(A)| = \sqrt{|d_K|}[\mathcal{O}_k : \mathfrak{a}]$$

as we sought to show.

Thus, if $K = \mathbb{Q}(\sqrt{m})$ where m is square-free:

$$\mathrm{vol}(\mathcal{O}_K) = \begin{cases} \sqrt{m} & m \equiv 1 \bmod 4 \\ 2\sqrt{m} & m \equiv 2, 3 \bmod 4 \end{cases}$$

Absolute Norm and Geometric Inequalities

Besides introducing a geometric intuition for number rings and their ideals, the key reason to introduce volumes is that they will allow us make many bounding arguments on ideals using the lattice structure of $\mathcal{O}_K$. To do this, we introduce a new "norm" on the ideals of $\mathcal{O}_K$ that shall represent the size of the ideal with respect to the lattice. This will be helpful in many algebraic arguments, but also in upgrading many clasical results about primes (such as their distribution[10],

[10] the number of primes p where $|p| \leq X$ tends asymptotically to $\frac{X}{\log(X)}$

see CITE:CFT).

Definition 22.1.49: Absolute Norm

Let $0 \neq \mathfrak{a} \subseteq \mathcal{O}_K$ be a nonzero ideal of $\mathcal{O}_K$. Then the *absolute norm* of $\mathfrak{a}$ is given by

$$\mathfrak{N}_K(\mathfrak{a}) := [\mathcal{O}_K : \mathfrak{a}]$$

If unambiguous, the notation $\mathfrak{N}(\mathfrak{a})$ is also used.

The absolute norm is inspired by the fact that ideals of $\mathcal{O}_K$, when thought of as representing a sublattice of a certain sparsity as compared to the lattice of $\mathcal{O}_K$. The norm of an ideal is always finite by proposition 22.1.31. Note that the notation $\mathfrak{N}_{K/\mathbb{Q}}$ should not be used, for it shall have a different meaning than $\mathfrak{N}$[11]

Lemma 22.1.50: Norm Of Principal Ideal

Let $(\alpha) \subseteq \mathcal{O}_K$ be a principal ideal. Then:

$$\mathfrak{N}((\alpha)) = \left|N_{K/\mathbb{Q}}(\alpha)\right|$$

Proof **:**
Let $\beta_1, \beta_2, ..., \beta_n$ be a $\mathbb{Z}$-basis for $\mathcal{O}_K$. Then $\alpha\beta_1, \alpha\beta_2, ..., \alpha\beta_n$ is a $\mathbb{Z}$-basis for $(\alpha) = \alpha\mathcal{O}_K$. Then if $A = (a_{ij})$ is the change of basis formula so

$$\alpha w_i = \sum_j a_{ij}\omega_j$$

by proposition 22.1.31 $|\det(A)| = [\mathcal{O}_K : \alpha]$. But by definition of the basis of A, namely it is left multiplication, we get the definition of the norm:

$$|\operatorname{Nm}_{K/\mathbb{Q}}(\alpha)| = \det(A)$$

giving us equality, as we sought to show.

Lemma 22.1.51: Ideal Divides Absolute Norm

Let $I \subseteq \mathcal{O}_K$ be an ideal. Then:

$$I \big| (\mathfrak{N}(I))$$

Proof **:**
If we show $\mathfrak{N}(I) = [\mathcal{O}_K : I] \in I$, then $(\mathfrak{N}(I)) \subseteq I$, which by corollary 22.2.17 implies we're done. But this is simply some ring theory: Take $1 \in \mathcal{O}_K$. Then

$$\overline{n} = \underbrace{\overline{1 + \cdots + 1}}_{n \text{ times}} \in \mathcal{O}_K/I$$

[11] see definition 22.3.11

Since $|\mathcal{O}_K/I| = n$, it must be that $\overline{n} = \overline{0}$, in particular $n \in I$. Since $n = [\mathcal{O}_K : I]$, we have the desired result.

Lemma 22.1.52: Volume Form Reformulation

Let $I \subseteq R$ be an ideal of a number field K. Then:

$$\operatorname{vol}(I) = \mathfrak{N}(I) \operatorname{vol}(\mathcal{O}_K)$$

Proof **:**
This is just reformulating proposition 22.1.48.

Theorem 22.1.53: Bounding Ideals

Let $K/\mathbb{Q}$ be a number field and $\mathcal{O}_K$ its number ring. Then there exists $\lambda \in \mathbb{R}_{>0}$ such that for every $0 \neq I \subseteq \mathcal{O}_K$, there exists a $\alpha \in I$ such that

$$|\operatorname{Nm}_{K/\mathbb{Q}}(\alpha)| \leq \lambda \mathfrak{N}(I)$$

Proof **:**
Let $\alpha_1, \alpha_2, ..., \alpha_n$ be an integral basis for R, and let $\sigma_1, \sigma_2, ..., \sigma_n$ denote the embeddings of K into $\mathbb{C}$. I claim that:

$$\lambda = \prod_{i=1}^{n} \left(\sum_{j=1}^{n} |\sigma_i(\alpha_j) \right)$$

Let I be any ideal, and m the unique positive integer satisfying:

$$m^n \leq \mathfrak{N}(I) < (m + 10^n$$

Consider the $(m+1)^n$ elements of $\mathcal{O}_K$ where:

$$\sum_{j=1}^{n} m_j \alpha_j \qquad m_j \in \mathbb{Z} \qquad 0 \leq m_j \leq m$$

Then two of these *must* be congruent $\bmod I$ since there are more than $\mathfrak{N}(I)$ of them. Taking the difference of these two elements, we get a nonzero member of I of the form:

$$\alpha = \sum_{j=1}^{n} m_j \alpha_j \qquad m_j \in \mathbb{Z} \qquad |m_j| \leq m$$

Hence:

$$\left|\operatorname{Nm}_{K/\mathbb{Q}}(\alpha)\right| \leq \prod_{i=1}^{n} |\sigma_i(\alpha)| \leq \prod -i = 1^n \left(\sum_{j=1}^{n} m_j |\sigma_i(\alpha_j)| \right) \leq m^n \lambda \leq \mathfrak{N}(I)\lambda$$

as we sought to show.

Proposition 22.1.54: Image Of Norm Is Finite Up To Units

Let K be a field, $\mathcal{O}_K$ its ring of integers, and $a \in \mathbb{Z}_{>0}$. Then up to multiplication by units, there are only finitely many element $\alpha \in \mathcal{O}_K$ such that

$$\mathrm{Nm}_{K/\mathbb{Q}}(\alpha) = a$$

Proof **:**
Let $a \in \mathbb{Z}_{>0}$. Recall that $\mathcal{O}_K/a\mathcal{O}_K$ is finite. I claim that, up to multiplication by units, there is at most one element α in each coset such that $|\mathfrak{N}(\alpha)| = |\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)| = a$. For, if $\beta = \alpha + a\gamma$ where $\gamma \in \mathcal{O}_K$, then since $N(\beta) = a$, we can write $\beta = \alpha + N(\beta)\gamma$. Dividing both sides by β then re-arranging we get:

$$\frac{\alpha}{\beta} = 1 \pm \frac{\mathrm{Nm}_{K/\mathbb{Q}}(\beta)}{\beta}\gamma$$

since $\mathrm{Nm}_{K/\mathbb{Q}}(\beta) = \beta \prod_{\tau \neq \mathrm{id}} \tau(\beta)$, we have that $\alpha/\beta \in \mathcal{O}_K$. By diving by α, rearranging, and substituting $a = N(\alpha)$, we also get:

$$\frac{\beta}{\alpha} = 1 \pm \frac{\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)}{\alpha}\gamma \in \mathcal{O}_K$$

hence, α and β are associates. Thus, up to multiplication by units, there is at most $[\mathcal{O}_K : a\mathcal{O}_K]$ elements of norm $\pm a$.

For computational purposes, we add a tighter bound on the norms of ideals. We start with the following lemma:

Lemma 22.1.55: Minkowski's Lattice Point Theorem

Let Γ be a discrete complete lattice in the Euclidean vector space V of dimension n and let X be a centrally symmetric convex subset of V. Then if

$$\mathrm{vol}(X) > 2^n \,\mathrm{vol}(\Gamma)$$

then X contains at least one nonzero lattice point $\gamma \in \Gamma$

Before proving this result, consider $\sigma(\mathbb{Z}[i])$. The smallest centrally symmetric set would

it is clear this square have volume arbitrarily close to 2^2, any large and it would intersect $\sigma(\mathbb{Z}[i])$. Similarly, if there was a long stretched area that tried to avoid the points of $\sigma(\mathbb{Z}[i])$, do to the size constraint, it will necessarily be to large to avoid intersecting the lattice:

Proof **:**
It suffice to show that there exists two distinct lattice points $\gamma_1, \gamma_2 \in \Gamma$ such that

$$\left(\frac{1}{2}X + \gamma_1\right) \cap \left(\frac{1}{2}X + \gamma_2\right) \neq \emptyset$$

To show this, let's first say there exists such a point so that

$$\frac{1}{2}x_1 + \gamma_1 + \frac{1}{2}x_2 + \gamma_2$$

Then we may re-write to obtain:

$$\gamma = \gamma_1 - \gamma_2 = \frac{1}{2}x_1 - \frac{1}{2}x_2$$

Notice that this is in the center o the line segment joining $-x_1$ and x_2, and hence belongs in $X \cap \Gamma$.

Now, if the set $\frac{1}{2}X + \gamma$, $\gamma \in \Gamma$ were all pairwise disjoint, then the same is the case for the intersections $\Phi \cap (\frac{1}{2}X\gamma)$ for fundamental mesh Φ of Γ, hence we would have

$$\mathrm{vol}(\Phi) \geq \sum_{\gamma \in \Gamma} \mathrm{vol}\left(\Phi \cap \left(\frac{1}{2}X + \gamma\right)\right)$$

However, translation so $\Phi \cap (\frac{1}{2}X + \gamma)$ by $-\gamma$ creates the set $(\Phi - \gamma) \cap \frac{1}{2}X$ of equal volume, implying $\Phi - \gamma$ covers all of V, and hence $\frac{1}{2}X$ It follows from our assumption that

$$\mathrm{vol}(\Phi) \geq \sum)\gamma \in \Gamma\, \mathrm{vol}\left((\varphi - \gamma) \cap \frac{1}{2}X\right) = \mathrm{vol}(\frac{1}{2}X) = \frac{1}{2^n}X$$

contradicting the hypothesis.

Theorem 22.1.56: Minkowski's Bound

Let $[\mathfrak{a}] \in Cl(K)$ be an ideal class of degree n. Then there exists an integral ideal $\mathfrak{a}$ such that

$$\mathfrak{N}(\mathfrak{a}) \leq \frac{n!}{n^n}\left(\frac{4}{\pi}\right)^{r_2}\sqrt{|d_K|}$$

Proof **:**
Let $\mathbb{R}^N \cong_{\mathbb{R}} \mathbb{R}^{r_1} \oplus \mathbb{R}^{r_2}$. Define the following norm on $\mathbb{R}^n$:

$$N(x) = x_1 \cdots x_r(x_{r_1+1}^2 + x_{r_1+2}^2)\cdots(x_{n-1}^2 + x_n^2)$$

(actually, type this up later, Marcus p.95)

Notice that is we re-arrange, we get:

$$\frac{n^n}{n!}\left(\frac{\pi}{4}\right)^{r_2} \leq \sqrt{|d_K}$$

Thus, we get a result that may seem obvious, but is in general rather difficult to prove:

Corollary 22.1.57: Discriminant Nonzero

Let $\mathcal{O}_K \neq \mathbb{Z}$. Then:

$$|d_K| > 1$$

Proof **:**
immediate consequence of above since $n > 1$

22.2 Ring Structure and Dedekind Domains

So far, we have shown that rings of integers are lattices whose rank corresponds to the dimension of their corresponding field extension. Since they are rings, we may naturally ask what type of ring are they. Using example 22.1.5, we can look at the particular example of $\mathcal{O}_{\mathbb{Q}(\sqrt{-5})} = \mathbb{Z}[\sqrt{-5}]$. Is this ring a Euclidean Domain, Principal Ideal Domain, or Unique Factorization Domain? It is certainly Noetherian, being a finitely generated $\mathbb{Z}$-module (where $\mathbb{Z}$ is Noetherian). However, not all irreducible elements are prime, for example:

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

showing that 2 is not prime. This may seem troubling since we may have to keep track of multiple factorizations showing there is more structure or tools that would need to be added/used to keep track of this information. Fortunately, there is a way to recover factorization. The goal of this section is to define a new type of ring for which the notion of Unique factorization is recovered, and show that all rings of integers are this type of ring. First, we may use the norm to see that a factorization into irreducible is always certainly possible in $\mathcal{O}_k$. For, if $\alpha = \beta\gamma$, then if β, γ are non-units:

$$1 < |N_{K/\mathbb{Q}}(\beta)| < |N_{K/\mathbb{Q}}(\alpha)| \qquad 1 < |N_{K/\mathbb{Q}}(\gamma)| < |N_{K/\mathbb{Q}}(\alpha)|$$

showing that it is a factorization domain. However this factorization need not be unique as we have demonstrated with $\mathcal{O}_{\mathbb{Q}(\sqrt{-5})} = \mathbb{Z}[\sqrt{-5}]$, another example being $21 = 3 \cdot 7 = (1 + 2\sqrt{-5})(1 - 2\sqrt{-5})$. We may see these are irreducible elements by using the norm. Doing the case of 3, if $3 = \alpha\beta$ for non-units α, β, then $9 = N_{K/\mathbb{Q}}(\alpha)N_{K/\mathbb{Q}}(\beta)$ would give us that $N_{K/\mathbb{Q}}(\alpha) = \pm 3$. Given $\alpha = a + b\sqrt{-5}$ we get

$$N_{K/\mathbb{Q}}(a + b\sqrt{-5}) = a^2 + 5b^2 = \pm 3$$

But this certainly has no solutions in $\mathbb{Z}$, and hence 3 is irreducible in $\mathcal{O}_K$. A similar argument can be made for $7, 1 \pm 2\sqrt{-5}$. Furthermore, the numbers 3 and 5 are not associated to $1 \pm 2\sqrt{-5}$, for we would require that

$$\frac{1 \pm 2\sqrt{-5}}{3}, \frac{1 \pm 2\sqrt{-5}}{7} \in \mathcal{O}_K$$

both of which we can check are not using the norm. Hence, we certainly have different factorizations.

The failure of unique factorization has lead Eduard Kummer to try to embed the ring of integers of K into a bigger domain of "ideal numbers" where unique factorization into "ideal prime numbers" would hold (which was inspired by the discovery of complex numbers somehow being "ideal" for roots polynomial roots). As an example, our original decomposition in $\mathbb{Z}[\sqrt{-5}]$

$$21 = 3 \cdot 7 = (1 + 2\sqrt{-5})(1 - 2\sqrt{-5})$$

would factor into ideal prime numbers $\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{p}_3, \mathfrak{p}_4$ with the relations

$$3 = \mathfrak{p}_1\mathfrak{p}_2 \qquad 7 = \mathfrak{p}_3\mathfrak{p}_4 \qquad 1 + 2\sqrt{-5} = \mathfrak{p}_1\mathfrak{p}_3 \qquad 1 - 2\sqrt{-5} = \mathfrak{p}_2\mathfrak{p}_4$$

Then we would have fixed non-uniqueness since we would get the factorization:

$$21 = (\mathfrak{p}_1\mathfrak{p}_2)(\mathfrak{p}_3\mathfrak{p}_4) = (\mathfrak{p}_1\mathfrak{p}_3)(\mathfrak{p}_2\mathfrak{p}_4)$$

Some key properties that these ideals numbers should have is that for $a, b, \lambda \in \mathcal{O}_K$, an ideal number $\mathfrak{a}$ should satisfy the following natural divisibility restrictions:

$$\mathfrak{a}\big|a \text{ and } \mathfrak{a}\big|b \Rightarrow \mathfrak{a}\big|a \pm b \qquad a\mathfrak{a}\big|a \Rightarrow \mathfrak{a}\big|\lambda a$$

An furthermore an ideal number $\mathfrak{a}$ ought to be determinable by the set of its divisors in $\mathcal{O}_K$, that is

$$\underline{\mathfrak{a}} = \left\{a \in \mathcal{O}_K \;\middle|\; \mathfrak{a}\big|a\right\}$$

But given these rules, this set is simply an ideal of $\mathcal{O}_K$ as we know the notion of ideal. It was Richard Dedekind that took the notion of ideal numbers that Kummer invented and invited the notion of *ideals* of $\mathcal{O}_K$ giving the modern definition of ideals. Then division $\mathfrak{a}\big|a$ is replaced by belonging $a \in \mathfrak{a}$, and divisibility $\mathfrak{a}\big|\mathfrak{b}$ is replaced by containment $\mathfrak{b} \subseteq \mathfrak{a}$.

Hence, we shall seek a unique factorization of *ideals* into *prime ideals* in $\mathcal{O}_K$. The 3 key properties of $\mathcal{O}_K$ that require this factorization was singled out by Dedekind and form a class of rings larger than rings of integers and are worth taking a look at on their own:

Definition 22.2.1: Dedekind Domain

Let R be an integral domain. Then if R is:

1. a Noetherian ring
2. every nonzero prime ideal is maximal
3. R is normal ($\overline{R}^{K(R)} = R$)

then R is called a *Dedekind Domain*

The Noetherian condition can be thought of as a natural eliminator of any infinites lurking around, while the prime implies maximal condition can be thought of as making sure that there can be a prime ideal that further contains prime ideals. In particular, we shall see that $(a) \subseteq (b)$ if and only if $b|a$, which doesn't make sense if we allow for prime ideals being contained in larger prime ideals.

58: Foreshadowing in Algebraic Geometry

Notice that we have already seen a ring with these conditions before, namely if V is a smooth 1-dimensional variety, then the associated k-algebra is Noetherian, 1 [Krull] dimensional and is normal (normal implies smoothness). In Algebraic Geometry, Dedekind domains can be thought of as representing 1-dimensional smooth schemes.

Lemma 22.2.2: PID's, UFD's, and Dedekind Domain

Let R be a ring. Then:

1. If R is a PID, R is a Dedekind Domain
2. If R is a UFD, R need not be a Dedekind Domain
3. If R is a Dedekind domain and a UFD, it is a PID

The following image demonstrates the realation:

Figure 22.1: Venn diagram

Proof **:**
R is trivially Noetherian and all prime ideals are maximal. For normalness, since R is a Unique Factorization Domain, $\overline{R}^{K(R)} = R$, hence R is a Dedekind Domain.

For the counter-example, take $\mathbb{Z}[\sqrt{-13}]$ and find two ways of decomposing 14. The last is an exercise.

For algeometric purposes, the following characterizes Dedekind domains locally:

Lemma 22.2.3: Dedekind Domain As DVR

Let R be commutative ring. Then R is a Dedekind domain if and only if for each prime P, R_P is a DVR. Thus, Dedekind domains are local DVR's.

Proof **:**
exercise.

Lemma 22.2.4: Dedekind Domains And Localization

Let R be a Dedekind domain and $S \subseteq R \setminus \{0\}$ is a multiplicative subset. Then $S^{-1}R$ is also a Dedekind Domain.

Proof **:**
The localization of Noetherian rings is Noetherian, and by the correspondence of ideals in localizations, we have that prime implies maximal in $S^{-1}R$. Finally, $S^{-1}R$ is integrally closed since if $x \in K = K(R)$ satisfies:

$$x^n + \frac{a_1}{s_1}x^{n-1} + \cdots + \frac{a_n}{s_n} = 0 \qquad \frac{a_i}{s_i} \in S^{-1}R$$

Then multiplying with the nth power of $s_1 \cdots s_n$ gives that sx is integral over R, hence $sx \in R$, and so $x \in S^{-1}R$, as we sought to show.

Importantly for number theory, all rings of integers are Dedekind domains:

Theorem 22.2.5: Ring Of Integers Are Dedekind Domains

Let $\mathcal{O}_K$ be a ring of integers. Then it is a Dedekind domain.

Proof **:**
By proposition 22.1.26, $\mathcal{O}_K$ is Noetherian, and by definition $\mathcal{O}_K$ is normal, namely $\mathcal{O}_K = \overline{\mathbb{Z}}^K$ meaning by corollary 20.5.4:

$$\overline{\mathcal{O}_K}^K = \overline{\overline{\mathbb{Z}}^K}^K = \overline{\mathbb{Z}}^K$$

What's left to show is that every prime ideal is maximal, which is equivalent to showing $\mathcal{O}_K/P$ is a field. In particular, since $\mathcal{O}_K/P$ is always an integral domain, if we show that it is finite then by corollary 9.1.15 we shall have that it's a field. The key intuition comes from the fact that $\mathcal{O}_K$ and its ideals form a lattice, meaning when we quotient we necessarily get a finite ring. Furthermore, the volume of the ideal is determined by the discriminant and the index, which will determine the cardinality of the quotient ring[a]

Let $0 \neq \alpha \in P$. Then $(p) = P \cap \mathbb{Z}$ is also a prime ideal. Since $P \subseteq \mathcal{O}_K$, we know α is associated to a minimal polynomial $m_{\alpha,\mathbb{Z}}(x)$ where

$$m_{\alpha,\mathbb{Z}}(\alpha) = \alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0 \qquad a_i \in \mathbb{Z}$$

Then $a_0 = -(\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha) \in P \subseteq \mathcal{O}_K$, and since $m_{\alpha,\mathbb{Q}}(x)$ is irreducible with integer coefficients, $a_0 \neq 0$ and $a_0 \in (p) = p\mathbb{Z}$. From here, we shall give two proofs: first, we may directly appeal to the fact that

$$(\mathcal{O}_K/P)/(\mathbb{Z}/p\mathbb{Z})$$

is an integral extension (see example 20.5.1) where the base ring is a field, hence $\mathcal{O}_K/P$ is a field by proposition 20.5.6. Another is the following: we have a non-trivial ideal with chains of inclusions $(a_0) \subseteq P \subseteq \mathcal{O}_K$ implying

$$|\mathcal{O}_K/P| \leq |\mathcal{O}_K/(a_0)| = |\mathcal{O}_K/a_0\mathcal{O}_K|$$

Now, since we are quotienting ideals, this is a $\mathbb{Z}$-module quotient, and since we are quotienting n-rank free $\mathbb{Z}$-modules, we get that

$$|\mathcal{O}_K/(a_0)| = p^n$$

But then $\mathcal{O}_K/P$ is finite (even better, its order divides $|a_0|^n$). Thus, P must be maximal, and since P was arbitrary every prime ideal is maximal, completing the proof.

[a] We shall soon see that the index can be thought of as the "norm" of a general ideal; this will mean the constant term of a minimal polynomial represent the "size", sheading some light on the proof

Since in general $\mathcal{O}_K$ is not a Unique Factorization Domain, rings of integers represent a class of rings where prime implies maximal, but factorization is not unique. In this way, we see that the integral closure of $\mathbb{Z}$ with respect to K may be far from having the nice element-wise properties we are used to in PIDs, but this shall be salvaged through having an understandable ideal structure and a way to measure how far away our ring is from being a PID.

Another important example of a Dedekind domain that is not a number ring is the ring is $k[x]$ (since it's a PID). We shall later see that integral extensions of Dedekind domains are Dedekind domains, an so these types of rings will form a basis for another large class of Dedekind domains that are not number rings.

22.2.1 Fractional Ideals in Rings

Following from the idea that ideals are the generalisation of numbers, we may define the following divisibility notions usually used for the [algebraic] lattice of integers induced by divisibility. These concepts shall be defined for general commutative rings:

Definition 22.2.6: GCD, LCM, Divisibility Of Ideals

let R be a commutative ring, $\mathfrak{a}, \mathfrak{b} \subseteq R$ ideals of R. Then:

$$\gcd(\mathfrak{a}, \mathfrak{b}) = (\mathfrak{a}, \mathfrak{b}) = \mathfrak{a} + \mathfrak{b} \qquad \operatorname{lcm}(\mathfrak{a}, \mathfrak{b}) = \mathfrak{a}\mathfrak{b}$$

And we can say that $\mathfrak{a}, \mathfrak{b}$ are relatively prime when $\gcd(\mathfrak{a}, \mathfrak{b}) = (1)$. Divisibility $\mathfrak{a}\big|\mathfrak{b}$ is defined as $\mathfrak{b} \subseteq \mathfrak{a}$, and hence we have an order and the ideals similar to that given by divisibility.

As we've seen in chapter 10, gcd and lcm are usually well-behaved in special classes of rings, usually UFDs. We shall show they are well-behaved in Dedekind domains since the ideals shall have unique factorizations. Let's now prove some lemmas for this result:

Lemma 22.2.7: Ideal Contains Product Of Prime Ideals

Let R be a Noetherian ring and $I \subseteq R$ is an ideal. Then I contains a product of prime ideals

Proof **:**
This was already proven in corollary 21.2.22(3)

For the following lemmas, we bring back the notion of fractional ideal that was defined in section 20.2. These ideals were at the time a way of extending ideals into the ring of fractions, and were shown to work well with localization. Due to all the properties of Dedekind domains, the product of ideals can be made into a *group*, even an *abelian group*. Since the fractional ideal structure will have such a strong restriction, this will produce incredible results on the relations between ideals. To begin, the following lemma shows that no prime ideal acts like the identity:

Lemma 22.2.8: Fractional Ideals

Let $P \subseteq R$ be a prime ideal of the Dedekind domain R, $K = K(R)$. Define

$$(R :_K P) = P^{-1} = \{x \in K \mid xP \subseteq R\}$$

Then

$$AP^{-1} = \left\{ \sum_i a_i x_i \;\middle|\; a_i \in A,\ x_i \in P^{-1} \right\} \neq A$$

for every ideal $0 \neq A \subseteq R$

Note that by definition $R \subseteq P^{-1}$ and so P^{-1} is not in general an ideal. The simplest "fractional ideal" would be those of $\mathbb{Z}$, namely $(n)^{-1} = \frac{1}{n}\mathbb{Z}$.

Proof **:**
Note that when we work with the inverse of elements, we are considering them as being part of $K(R)$.

We shall first show the special case of $P^{-1} \neq R$. Let $a \in P$ with $a \neq 0$, and consider $P_1 P_2 \cdots P_r \subseteq (a) \subseteq P$ with r minimal. Then one of the P_i, is containing in P, without loss of generality say $P_1 = P$ (since P_l is maximal). Indeed, if non of the P_i were contained in P, then for every i there would exist $a_i \in P_i \setminus P$ such that $a_1 \cdots a_r \in P$, but P is maximal. Next, by minimality $P_2 \cdots P_r \nsubseteq (a)$, so there exists $b \in P_2 \cdots P_r$ such that $b \notin (a) = aR$, that is $a^{-1}b \notin R$. On the other hand, since $P = P_1$, $P_1 b \in P_1 P_2 \cdots P_n r$, hence we have $bP \subseteq (a) = aR$ that is $a^{-1}bP \subseteq R$, and so $a^{-1}b \in P^{-1}$. It follows that $P^{-1} \neq R$, namely $P^{-1} \supsetneq R$.

Next, let $0 \neq A \subseteq R$ be an ideal. Since R is Noetherian, let

$$A = (\alpha_1, \alpha_2, ..., \alpha_n)$$

For the sake of contradiction, say $AP^{-1} = A$. Then for every $x \in P^{-1}$

$$xa_i = \sum_j a_{ij}\alpha_j \qquad a_{ij} \in R$$

Letting M be the matrix $(x\delta_{ij} - a_{ij})$, we get $M(\alpha_1, \alpha_2, ..., \alpha_n)^t = 0$. By Cramers rule, if $d = \det(M)$, then $d\alpha_1 = \cdots = d\alpha_n = 0$, and so $d = 0$. Then x is integral over R, since its' the zero of the monic polynomial

$$f(X) = \det(X\delta_{ij} - a_{ij}) \in R[x]$$

Thus, since R is integrally closed $x \in R$. But then $P^{-1} = R$; a contradiction to your first part, as we sought to show.

Definition 22.2.9: Fractional Ideal

Let R be an integral domain and $K = K(R)$. Then a *fractional ideal* J of K is a finitely generated non-trivial R-submodule of K where there exists a $r \in R$ such that $rJ \subseteq R$ is an ideal

Fractional ideals are multiplied in the similar way ideals are multiplied (treat them as sub-algebra multiplication). Since all ideals of R are rank 1 submodules of R, they are all fractional ideals. For those who are motivated by algebraic geometry, fractional ideals shall be shown to form line bundles over the smooth $\mathbb{Z}$-curves $\mathcal{O}_K$ [12]. Fractional ideals that are ideals are given a name:

Definition 22.2.10: Integral Ideals

Let R be an integral domain and $K = K(R)$. Then a fractional ideal $I \subseteq K$ that is of rank 1 is called an *integral ideal*, and $I \subseteq R$.

Example 22.2.1: Fractional Ideal

1. Since $\mathbb{Z}$ is a Principal Ideal Domain, every ideal is of the form $a\mathbb{Z}$ for $a \in \mathbb{Z}$. Then $(a)^{-1} = a^{-1}\mathbb{Z}$. Hence, the fractional ideals are in bijective correspondence with $\mathbb{Q}^\times_{>0}$. Naturally, all ideals of $\mathbb{Z}$ are fractional ideals, in particular they are integral ideals.
2. More generally, if $a \in K^\times$, then we get fractional principal ideals $(a) = aR$, and we have a bijective correspondence between the fractional ideals and $K^\times/R^\times$.

The condition that there must exist an $r \in R$ such that $rJ \subseteq R$ is superfluous in Noetherian domains:

[12] See section 22.7 to see how to interpret rings of integers as smooth $\mathbb{Z}$-curves, and ref:HERE for interpreting fractional ideals as line bundles

Corollary 22.2.11: Fractional Ideal Representative

Let R be a Noetherian Domain, $K = K(R)$, and $\mathfrak{a} \subseteq K$ a fractional ideal. Then there exists a $r \in R \setminus \{0\}$ and ideal $I \subseteq R$ such that

$$\mathfrak{a} = r^{-1} I$$

Stated differently, $\mathfrak{a} \cong_R I$

If $R = \mathcal{O}_K$, we may take $r \in \mathbb{Z}$. Furthermore, the above is an if and only if statement.

Proof **:**
Use the fact that R is Noetherian. Let $\mathfrak{a} = (\alpha_1, \alpha_2, ..., \alpha_n)$. In particular, since $\mathfrak{a}$ is a finitely generated $\mathbb{Z}$-module:

$$\mathfrak{a} = \sum R\alpha_i \qquad \alpha_i \in K(R)$$

by definition, $\alpha_i = a_i/b_i$ for $a_i, b_i \in R$. Taking $b = b_1 b_2 \cdots b_n$, we get

$$b\mathfrak{a} \subseteq R$$

Hence, $b\mathfrak{a}$ is an ideal, say $b\mathfrak{a} = I$ Then $\mathfrak{a} \subseteq b^{-1} I$.

This immediately implies that $\mathfrak{a} \cong_R I$, for if $\mathfrak{a} = (\alpha_1, \alpha_2, ..., \alpha_n)$ so that $I = (r^{-1}\alpha_1, ..., r^{-1}\alpha_n)$, then define $\varphi(\alpha_i) = r^{-1}\alpha_i$ is an R-module isomorphism.

Notice that this generalizes the definition of ideals, if R is a Dedekind ring over itself, then the ideals of R are the finitely generated sub-modules. Then if $K = \mathrm{Frac}(R)$, the fractional ideals are the finitely generated K-modules.

Corollary 22.2.12: Inverse of Prime Ideals

Let R be a Dedekind domain and $\mathfrak{p} \subseteq R$ be a prime ideal. Then:

$$\mathfrak{p}\mathfrak{p}^{-1} = (1)$$

Proof **:**
If $x \in \mathfrak{p}\mathfrak{p}^{-1}$, then $x = \sum p_i k_i$ for $p_i \in \mathfrak{p}$ and $k_i \in \mathfrak{p}^{-1}$. But by definition of $\mathfrak{p}^{-1}$, we have that $p_i k_i \in R$, hence x element is in R, and so $\mathfrak{p}\mathfrak{p}^{-1} \subseteq R$. Next, certainly $\mathfrak{p} \subseteq \mathfrak{p}\mathfrak{p}^{-1} \subseteq R$. Since $\mathfrak{p}$ is maximal, either $\mathfrak{p}\mathfrak{p}^{-1} = \mathfrak{p}$ or $\mathfrak{p}\mathfrak{p}^{-1} = R$. The former is a contradiction by lemma 22.2.8, and hence $\mathfrak{p}\mathfrak{p}^{-1} = (1)$, completing the proof.

Lemma 22.2.13: Invariance Number Ring Multiplication

Let $t \in K$, $\mathfrak{a} \subseteq K$ a fractional ideal. If $tI = I$, then $t \in R$

Proof **:**
Since I is a finitely generated R module, there exists a surjective R-module homomorphism $\varphi : R^n \to I$. Next, since $tI = I$, there exists an R-module homomorphism $t : I \to I$ sending $i \mapsto ti$. Then $t \in \mathrm{Hom}_R(I, I)$. Since R^n is free, if we consider the following lift:

$$\begin{tikzcd} R^n \arrow[r, two heads] & I \\ R^n \arrow[u, dashed, "\text{lift}"] \arrow[r, two heads, "\varphi"] \arrow[ur] & I \arrow[u, "t"'] \end{tikzcd}$$

Then $xt \in \mathrm{Hom}_R(R^n, I)$. Taking $\alpha \in \mathrm{Hom}_R(R^n, R^n)$:

$$\det(xI - \alpha)(\alpha) = 0 \Rightarrow \det(xI - \alpha)(t) = 0$$

now using that R is a Dedekind domain, namely that it's integrally closed, we get that $t \in R$, completing the proof.

22.2.2 Prime Decomposition in Dedekind Domain

With these properties, we may show the factorization of prime ideals in Dedekind domains.

Theorem 22.2.14: Unique Factorization Of Ideals In Dedekind Domains

Let R be a Dedekind Domain and $\mathfrak{a} \subseteq R$ a non-trivial proper ideal. Then there exists unique $\mathfrak{p}_1, \mathfrak{p}_2, ..., \mathfrak{p}_n$ such that

$$\mathfrak{a} = \mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_n$$

Proof **:**
We start with the existence of such a factorization. Let $\mathfrak{M}$ be the set of all non-trivial proper ideals which do not admit a prime ideal decomposition. If $\mathfrak{M}$ is non-empty, then since R is Noetherian there exists a maximal element, say $\mathfrak{m}$. Then this ideal is contained in some maximal ideal $\mathfrak{p}$. Since $R \subseteq \mathfrak{p}^{-1}$, we get the following chain of inclusions:

$$\mathfrak{m} \subseteq \mathfrak{m}\mathfrak{p}^{-1} \subseteq \mathfrak{p}\mathfrak{p}^{-1} \subseteq R$$

by lemma 22.2.8, we have $\mathfrak{m} \subsetneq \mathfrak{m}\mathfrak{p}^{-1}$ and $\mathfrak{p} \subsetneq \mathfrak{p}\mathfrak{p}^{-1} \subseteq R$. By maximality of $\mathfrak{p}$, $\mathfrak{p}\mathfrak{p}^{-1} = R$, and by maximality of $\mathfrak{m}$ in $\mathfrak{M}$, $\mathfrak{m}\mathfrak{p}^{-1}$ admits a prime ideal decomposition, say

$$\mathfrak{m}\mathfrak{p}^{-1} = \mathfrak{p}_1 \cdots \mathfrak{p}_r$$

But then, since $\mathfrak{p}\mathfrak{p}^{-1} = R$, $\mathfrak{m}R = \mathfrak{m}$ and so

$$\mathfrak{m} = \mathfrak{m}\mathfrak{p}^{-1}\mathfrak{p} = \mathfrak{p}_1 \cdots \mathfrak{p}_r\mathfrak{p}$$

contradicting the definition of $\mathfrak{m}$. Hence $\mathfrak{M}$ is an empty-set

For uniqueness, let

$$\mathfrak{a} = \mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_r = \mathfrak{q}_1\mathfrak{q}_2 \cdots \mathfrak{q}_s$$

be two prime ideal factorizations of $\mathfrak{a}$. Then by the repeated application of property of prime ideals ($\mathfrak{a}\mathfrak{b} \subseteq P \Rightarrow \mathfrak{a} \subseteq P$ or $\mathfrak{b} \subseteq P$), we have that $\mathfrak{p}_1$ divides a factor of the right hand side, say

$\mathfrak{q}_1$, and by maximality is thus equal to $\mathfrak{q}_1$. Then we can multiply both sides by $\mathfrak{p}_1^{-1}$ and since $\mathfrak{p}_1 \neq \mathfrak{p}_1\mathfrak{p}_1^{-1} = R$ we get

$$\mathfrak{p}_2 \cdots \mathfrak{p}_r = \mathfrak{q}_2 \cdots \mathfrak{q}_s$$

Then proceeding inductively, we will get that $r = s$ and $\mathfrak{p}_i = \mathfrak{q}_i$, up to renumbering, completing the proof.

Note: It is in fact equivalent to define a Dedekind domain as ring in which each ideal factors uniquely, showing that this theorem characterizes Dedekind domains, see exercise ref:HERE.

From now on, we shall group the prime ideals to write

$$\mathfrak{a} = \mathfrak{p}_1^{n_1} \cdots \mathfrak{p}_r^{n_r} \qquad n_i > 0$$

If $\mathfrak{a} = (a)$ is a principal ideal, we shall use the slight abuse of notation and write

$$a = \mathfrak{p}_1^{n_1} \cdots \mathfrak{p}_r^{n_r} \qquad n_i > 0$$

Similarly, we shall often write $\mathfrak{a}\big|a$ instead of $\mathfrak{a}\big|(a)$. An important result worth emphasizing is the structure of R/I for an ideal $I \subseteq R$. First, note that if $\mathfrak{p}_i + \mathfrak{p}_j = (1)$, then $\mathfrak{p}_i^{n_i} + \mathfrak{p}_j^{n_j} = (1)$ (this was exercise ref:HERE (in CRT section)); the trick is to consider $1 = 1^{n_i+n_j+1} = (ax+by)^{n_i+n_j+1}$, use the binomial theorem, and group appropriate elements. Then by the Chinese Remainder Theorem:

$$R/I = \bigoplus_i R/P_i^{n_i}$$

By the 4th isomorphism theorem, $\overline{I} \subseteq R/P_i^{n_i}$ if and only if $P_i^{n_i} \subseteq I$, that is $I\big|P_i^{n_i}$. But by unique factorization, that means that $I = P_i^k$ for $1 \leq k \leq n_i$. Hence, every quotient has only finitely many ideals, each ideal I corresponding to a set of prime power of the decomposition of I via the isomorphism given by the CRT. This matches with the result that $R/(a)$ is a finite ring, showing there can only be finitely many ideals, and the prime factorization theorem for ideals gives us the structure of the ideals of $R/(a)$ through the correspondence given by the 4th isomorphism theorem. This also immediately shows that R/I is *Artinian*[13].

Using the fact that ideals factor and "behave" like the positive rational numbers ($-\mathfrak{p} = \mathfrak{p}$ since -1 is a unit), we can conclude some powerful results. We have shown that every ideal has a unique inverse. This immediately tells us that:

$$\mathfrak{a}\mathfrak{b} = \mathfrak{a}\mathfrak{c} \quad \Longrightarrow \quad \mathfrak{b} = \mathfrak{c}$$

by the Cancellation law. Next, it is easy to see is that we may generalize the notion of finding GCD and LCM of UFD's to Dedekind domains:

[13] Equivalently, $\operatorname{kdim}(R/I) = 0$, and hence it is Artinian

Corollary 22.2.15: GCD, LCM, and Divisibility In Dedekind Domains

Let $\mathfrak{a}, \mathfrak{b} \subseteq R$ be ideals in a Dedekind domain R. Then

1. $\operatorname{lcm}(\mathfrak{a}, \mathfrak{b}) = P_1^{\max(a_1,b_1)} \cdots P_k^{\max(a_k,b_k)}$
2. $\gcd(\mathfrak{a}, \mathfrak{b}) = P_1^{\min(a_1,b_1)} \cdots P_k^{\min(a_k,b_k)}$
3. $\mathfrak{a} \big| \mathfrak{b}$ if and only if there exists a $\mathfrak{c}$ such that $\mathfrak{a}\mathfrak{c} = \mathfrak{b}$

***Proof* :**
For the first two, work with the factorization and property of ideals.

For 3, if $\mathfrak{a} \big| \mathfrak{b}$, we have that $\mathfrak{b} \subseteq \mathfrak{a}$. In a Dedekind domain, multiplying by $\mathfrak{a}^{-1}$, we get $\mathfrak{a}\mathfrak{a}^{-1} = (1)$ by extending lemma 22.2.8 with the above theorem. Hence:

$$\mathfrak{b}\mathfrak{a}^{-1} \subseteq \mathfrak{a}\mathfrak{a}^{-1} \subseteq R$$

showing that $\mathfrak{b}\mathfrak{a}^{-1} = \mathfrak{c} \subseteq R$ is an ideal, and $\mathfrak{b} = \mathfrak{c}\mathfrak{a}$, which properly generalizes the notion factorizing elements to the level of ideals.

We know that every ideal has an inverse. However, that inverse is a fractional ideal. Fortunately, there always exists an integral ideal that is "almost" the inverse, in that the product is a principal ideal[14]:

Corollary 22.2.16: Exists J such that IJ is Principal

Let R be a Dedekind Domain and $I \subseteq R$ an ideal. Then there exists an ideal $J \subseteq R$ such that $IJ = (a)$ for some element $a \in I$. In other words, there exists an ideal J such that the product IJ is principal.

***Proof* :**
Let $I = P_1^{n_1} \cdots P_k^{n_k}$. For each i, choose a nonzero element $r_i \in P_i^{n_i} \setminus P_i^{n_i+1}$. By the Chinese Remainder Theorem, there exists an $a \in R$ such that

$$a \equiv r_i \bmod P_i^{n_i+1} \qquad 1 \leq i \leq k$$

Thus, $a \in P_i^{n_i} \setminus P_i^{n_i+1}$ for each i, and so since $a \in P_i^{n_i}$, the power of P_i in the factorization is n_i. Now, taking the prime factorization of (a), we get:

$$(a) = P_1^{n_1} \cdots P_k^{n_k} P_{n+1}^{n_{k+1}} \cdots P_m^{n_m}$$

Letting $J = P_{n+1}^{n_{k+1}} \cdots P_m^{n_m}$ gives us

$$(a) = IJ$$

This same process can be repeated for fractional ideals; for any fractional ideal I, take $r \in R$ such that $rI = J \subseteq R$. Then proceed with the proof above, giving us an ideal L such that $(a) = JL = rIL$. Then $(r^{-1}a) = IL$, completing the proof.

[14] this almost inverse can be made into a concrete inverse when we show the principal ideals form a normal subgroup formed by the group of fractional ideals

In section 22.3.1, we shall find a more systematic way of finding the prime ideals that multiply to get a principal ideal using Galois groups. As seen with rings of integers, not all Dedekind domains are PIDs since they are not all UFDs. With the factorization theorem, we get that they are the next best thing:

Corollary 22.2.17: PIR, and Generation Of Ideals In Dedekind Domains

Let R be a Dedekind domain and $I \subseteq R$ an ideal. Then R/I is a Principal Ideal Ring, and I is generated by *at most* 2 elements.

Proof **:**
For any ideal $I \subseteq R$, by the Chinese remainder theorem and prime factorization of I:

$$R/I = \bigoplus_i R/P_i^{n_i}$$

We shall show that each $R/P_i^{n_i}$ is a principal ideal ring, from which we may conclude by the isomorphism that R/I is a principal ideal ring. This shall give the desired result, for if $I \subseteq R$, then for any $J \subseteq I$ we may take R/J where $\overline{I} \subseteq R/(\alpha)$. Then since all ideals of $R/(\alpha)$ are principal, $\overline{I} = (\overline{\beta})$, that is $I = J + (\beta) = (\alpha, \beta)$. If we take $\alpha \in I$, then we would get $(\alpha) \subseteq I$ and $R/(\alpha)$, and so $I = (\alpha) + (\beta) = (\alpha, \beta)$, showing each ideal is generated by at most 2 elements.

Hence, we prove $R/P_i^{n_i}$ is a Principal ideal ring (PIR). If $\overline{J} \subseteq R/(P_i^{n_i})$, then $P_i^{n_i} \subseteq J \subseteq R$, meaning $J | P_i^{n_i}$, and so by unique factorization it must be that $J = P_i^k$ for some $1 \leq k \leq n_i$, which has image $\overline{P_i^k}$ (i.e., there are finitely many ideals).

Now, if $\overline{P_i} = \overline{P_i^2}$, then $\overline{P_i} = \overline{P_i^n} = \overline{0}$, and so $P_i = P_i^n$ implying that $R/P_i^{n_i}$ is a field, and it then automatically a PIR. If $\overline{P_i} \neq \overline{P_i^2}$, take $\alpha \in \overline{P_i} \setminus \overline{P_i^2}$. Then $\overline{\alpha} \notin \overline{P_i^2} \supseteq \overline{P_i^3} \supseteq \cdots$. Thus it must be that

$$\overline{P_i} \big| (\overline{\alpha})$$

But since the only ideals of $R/P_i^{n_i}$ are $\overline{P_i}, ..., \overline{P_i^{n_i - 1}}$, it must follow that $(\overline{\alpha}) = \overline{P_i}$. From that it follows that $(\overline{\alpha}^k) = \overline{P_i^k}$, showing that $R/P_i^{n_i}$ is a PIR. But then R/I is a PIR since the isomorphism maps principal ideals to principal ideals, which then as we saw completes the proof.

As a consequence of this, if $\mathfrak{p} \big| (p)$ for some $p \in \mathbb{Z}$, then by using the construction of the proof above, we may always find an element $\alpha \in \mathfrak{p}$ such that $\mathfrak{p} = (p, \alpha)$. Hence, factorization of a prime $p \in \mathbb{Z}$ can always by recovered by adding an appropriate element to (p). More generally, if $I \subseteq J$, then $J = I + (a)$ (can you think of any conditions a needs to satisfy?). Our final result shows all pairs of ideals are, up to a multiple of $K(R)$, relatively prime. This mirrors the property that if $\gcd(a, b) = c$, we can divide a by c to get $\gcd(a/c, b) = 1$:

Corollary 22.2.18: Always Almost Relatively Prime

Let A, B be fractional ideals of the Dedekind Domain R. Then there exists elements $x, y \in K$ such that xA and yB are relative prime.

Proof **:**
By multiplying A, B by appropriate scalars, we may assume they are integral ideal. We shall find an element c such that $cA + B = R$. First, let $a \in A$, and I be the ideal such that $AI = (a)$. Then since $BI \subseteq I$, consider:

$$\overline{I} \subseteq R/BI$$

Then since R/BI is principal, $I = BI + (b)$ for some $b \in R$. Multiplying by A, dividing by a, and letting $c = b/a$, we get

$$cA + B = R$$

as we sought to show.

With the division of ideals defined and mirroring the structure of $K^\times/R^\times$ ($\mathbb{Q}^\times_{>0}$ when working with $\mathbb{Z}$), it is tempting to say that they form a group structure, and this is indeed the case and is the focus in the remainder of this section.

One of the main metrics to measure is how many more "ideal numbers" need to be added to recover unique factorization into primes. If $\mathfrak{p}$ and $\mathfrak{q}$ are added, then if for some $t \in K$, $t\mathfrak{p} = \mathfrak{q}$, then $\mathfrak{p}$ and $\mathfrak{q}$ don't actually differ much in structure: it can be shown that $I = aJ$ ($a \in K(R)$) if and only if $I \cong_{\mathcal{O}_K} J$). Hence, we may classify "different" classes of ideals up to multiplication by a principal ideal. The more classes there are, the more a ring R fails to be a unique factorization domain. When R is a number ring, we shall show that there are only finitely many such classes.

The first result to prove is that the collection of fractional ideals of a Dedekind domain does indeed form a group:

Theorem 22.2.19: Ideal Group

Let R be a Dedekind domain, $K = K(R)$, and let J_K be the collection of fractional ideals. Then $(J_K, \cdot)$ has an abelian group structure via fractional ideal multiplication, and is called the *ideal group* of K

Proof **:**
The operation is certainly well-defined (the product of fractional ideals is a fractional ideal), associativity and commutativity come for free, and the identity is $R = (1)$ since for any $\mathfrak{a} \in J_K$

$$\mathfrak{a}(1) = (1)\mathfrak{a} = \mathfrak{a}$$

What's left to check for is the existence of an inverse. For prime ideals, we have that $\mathfrak{p} \subsetneq \mathfrak{p}\mathfrak{p}^{-1}$, an hence by maximality we have $\mathfrak{p}\mathfrak{p}^{-1} = (1)$. For an integral ideal, if $\mathfrak{a} = \mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_r$, then its inverse is

$$\mathfrak{b} = \mathfrak{p}_1^{-1} \cdots \mathfrak{p}_r^{-1}$$

We want to show that $\mathfrak{b} = \mathfrak{a}^{-1}$. Since $\mathfrak{b}\mathfrak{a} = (1)$, then by definition of fractional ideals $\mathfrak{b} \subseteq \mathfrak{a}^{-1}$. Conversely, to show $\mathfrak{a}^{-1} \subseteq \mathfrak{b}$, if $x\mathfrak{a} \subseteq (1)$, then $x\mathfrak{a}\mathfrak{b} \subseteq \mathfrak{b}$. Since $\mathfrak{a}\mathfrak{b} = R$, we have $aR \subseteq \mathfrak{b}$, that is $(a) \subseteq \mathfrak{b}$, and so $x \in \mathfrak{b}$. Hence, we have $\mathfrak{b} = \mathfrak{a}^{-1}$. Finally, if $\mathfrak{a}$ is an arbitrary fractional ideal and $0 \neq c \in R$ such that $c\mathfrak{a} \subseteq R$, then $(c\mathfrak{a})^{-1} = c^{-1}\mathfrak{a}^{-1}$ is the inverse of $c\mathfrak{a}$, so then $\mathfrak{a}\mathfrak{a}^{-1} = R$, completing the proof.

Corollary 22.2.20: Fractional Ideal Prime Decomposition

Every fractional ideal $\mathfrak{a}$ admits a unique representation as the product

$$\mathfrak{a} = \prod_{\mathfrak{p}} \mathfrak{p}^{n_{\mathfrak{p}}}$$

where $n_p \in \mathbb{Z}$ and $n_p = 0$ for almost all $\mathfrak{p}$. Thus, J_K is the free abelian group generated by the set of nonzero prime ideals $\mathfrak{p}$ of R

Proof **:**
Notice that every fractional ideal $\mathfrak{a}$ is the quotient of two integral ideals,

$$\mathfrak{a} = \mathfrak{b}/\mathfrak{c}$$

which both have a prime decomposition. Thus, $\mathfrak{a}$ has a prime decomposition as stated in the corollary. Uniqueness follows the same argument as for integral ideals.

Definition 22.2.21: Ideal Group

Let R be a Dedekind Domain. Then J_R denotes the *ideal group* given by ideal multiplication.

We shall often write J_R interchangeably with J_K

Example 22.2.2: Ideal Group

1. The canonical example would be $J_{\mathbb{Z}} \cong \mathbb{Q}_+^\times$ with the bijection

$$q\mathbb{Z} = (q) \mapsto q$$

   Note that it is important to think of the multiplicative structure of $\mathbb{Q}$, $(\mathbb{Q}, +)$ is not a free abelian group, having no two distinct elements be linearly dependent and not being cyclic.

2. More generally, let R be a Principal Ideal Domain. Then:

$$J_R \cong K^\times / R^\times$$

   This isomorphism does not work in general, namely since not all ideals of R are principal.

A natural subgroup of J_K is formed by the principal ideal (a) where $a \in K^\times$. This subgroup can be thought of as representing our usual notion of factorization into elements instead of ideals As some foreshadowing, notice that the principal ideal The quotient of J_K by the subgroup of principle ideals can be thought of as measuring how far away the ideal group is from being a PID:

Definition 22.2.22: Ideal Class Group

Let J_K be the class group of K. Then

$$Cl(K) = J_K/P_K$$

is called the *ideal class group*, *class group*, or *Picard group* of K.

59: Note on Nomenclature

The term Picard group shall not be used in this book, as the nomenclature is more so associated within algebraic geometry. It is pointed out here that some call the class group the Picard group as the class group is a special case of the Picard group.

These naturally fit in the following exact sequence:

$$1 \to R^\times \to K^\times \to J_K \to Cl(K) \to 1$$

where $K^\times \to J_K$ is the map $a \mapsto (a)$. Hence, we may think of the class group $Cl(K)$ as a measure of the expansion that happens when we pass from numbers to ideals, where the unit group $R^\times$ measure the contraction of the same process. Thus, we may want to understand the groups $R^\times$ and $Cl(K)$ more thoroughly. In general, these can be arbitrary abelian groups[15], but for rings of integers $\mathcal{O}_K$ of number fields K, we get that the size of the class group is always finite and that $\mathcal{O}_K^\times$ is a free $\mathbb{Z}$-module along with roots of unity contained in K, as will soon be shown.

22.2.3 Class Group of Ring of Integer

Recall that ideals $\mathfrak{a} \subseteq \mathcal{O}_K$ form a sub-lattice of $\mathcal{O}_K$ in the Archimedean embedding. The following theorem is key bound on the covolume of lattice. Using theorem 22.1.53, we may rather easily show that the class group is finite. The way we shall find the size of the class group is by showing that every element in a class group will have absolute norm less than a bound (say M), and that only finitely many ideals can have norm equal to $m < M$. Then, by considering the principal ideals that have norm less than M, we shall be able to find all prime ideals that compose these principal ideals, which shall give us an estimate on the size of the group.

Lemma 22.2.23: Absolute Norm of Prime

Let $0 \neq \mathfrak{p} \subseteq \mathcal{O}_K$ be a prime ideal of $\mathcal{O}_K$ and let $\mathcal{O}_K$ be rank n. Then

$$\mathfrak{N}(\mathfrak{p}) = p^n$$

Proof **:**
By theorem 20.5.13, $\mathfrak{p} \cap \mathbb{Z} = p\mathbb{Z}$. Since $\mathcal{O}_K/\mathbb{Z}$ is an integral extension, $\mathcal{O}_K/\mathfrak{p}$ is a finite field extension of $\mathbb{Z}/p\mathbb{Z}$ of degree n by example 20.5.1(3). Hence:

$$\mathfrak{N}(\mathfrak{p}) = p^n$$

[15]There are "enough" 1-dimensional smooth schemes to make arbitrary class groups on Dedekind domains

Proposition 22.2.24: Absolute Norm for Ring of Integers Is Multiplicative

Let $\mathfrak{a} = \mathfrak{p}_1^{n_1} \cdots \mathfrak{p}_k^{n_s}$ be the prime Factorization of the nonzero ideal $\mathfrak{a}$. Then:

$$\mathfrak{N}(\mathfrak{a}) = \mathfrak{N}(\mathfrak{p}_1)^{n_1} \cdots \mathfrak{N}(\mathfrak{p}_k)^{n_s}$$

From this, we may conclude that

$$\mathfrak{N}(\mathfrak{a}\mathfrak{b}) = \mathfrak{N}(\mathfrak{a})\mathfrak{N}(\mathfrak{b})$$

***Proof* :**
By the Chinese remainder theorem:

$$\mathcal{O}_K/\mathfrak{a} \cong \mathcal{O}_K/\mathfrak{p}_1^{n_1} \oplus \cdots \oplus \mathcal{O}_K/\mathfrak{p}_s^{n_s}$$

Hence, we see that the index $[\mathcal{O}_K : \mathfrak{a}]$ is the product of the indexes $[\mathcal{O}_K : \mathfrak{p}_k^{n_i}]$. To find this, notice that the chain

$$\mathfrak{p}^n \subseteq \mathfrak{p}^{n-1} \subseteq \cdots \mathfrak{p} \subseteq \mathcal{O}_K$$

is a sequence of integral extensions[a], so we may use the fact that the index is multiplicative, namely:

$$[\mathcal{O}_K : \mathfrak{p}^n] = [\mathcal{O}_K : \mathfrak{p}][\mathfrak{p}^2 : \mathfrak{p}] \cdots [\mathfrak{p}^n : \mathfrak{p}^{n-1}]$$

We must now show each index has the same size. We shall take advantage of the fact that $\mathcal{O}/\mathfrak{p}$ is finite, and show that each $\mathfrak{p}^i/\mathfrak{p}^{i+1}$ is isomorphic to $\mathcal{O}/\mathfrak{p}$ (meaning they are all of the same size, hence the same index). I claim that $\mathfrak{p}^i/\mathfrak{p}^{i+1}$ is an $(\mathcal{O}_K/\mathfrak{p})$-vector space of dimension 1 with the natural left-multiplication action; this will come down to the multiplicative structure of J_K. To see this, let $a \in \mathfrak{p}^i \setminus \mathfrak{p}^{i+1}$ and take $\mathfrak{b} = (a) + \mathfrak{p}^{i+1} = \gcd((a), \mathfrak{p}^{i+1})$ so that $\overline{\mathfrak{b}}_{\mathfrak{p}^{i+1}} = \overline{(a)}_{\mathfrak{p}^{i+1}}$. Then by construction

$$\mathfrak{p}^{i+1} \subsetneq \mathfrak{b} \subseteq \mathfrak{p}^i$$

From this, we must conclude that $\mathfrak{p}^i = \mathfrak{b}$; this can be seen immediately by realising that $\mathfrak{b} = \gcd((a), \mathfrak{p}^{i+1}) = \mathfrak{p}^i$ by definition of (a). Therefore $\overline{\mathfrak{b}} = a \bmod \mathfrak{p}^{i+1}$ is a basis for the $\mathcal{O}_K/\mathfrak{p}$-vector-space $\mathfrak{p}^i/\mathfrak{p}^{i+1}$, hence $\mathfrak{p}^i/\mathfrak{p}^{i+1} \cong \mathcal{O}_K/\mathfrak{p}$. This can also be demonstrated by taking

$$\mathcal{O} \to \mathfrak{p}^i/\mathfrak{p}^{i+1} \qquad \alpha \mapsto \overline{\alpha a}$$

and seeing the map is surjective since $\mathfrak{p}^i = \gcd(\mathfrak{b}, \mathfrak{p}^{i+1})$ and has kernel $\mathfrak{p}$. Since this is true for all i, $1 \leq i \leq n$:

$$\mathfrak{N}(\mathfrak{p}^n) = [\mathcal{O}_K : \mathfrak{p}^n] = [\mathcal{O}_K : \mathfrak{p}][\mathfrak{p} : \mathfrak{p}^2] \cdots [\mathfrak{p}^{n-1} : \mathfrak{p}^n] = \mathfrak{N}(\mathfrak{p})^n$$

Thus more generally:

$$\mathfrak{N}(\mathfrak{a}\mathfrak{b}) = \mathfrak{N}(\mathfrak{a})\mathfrak{N}(\mathfrak{b})$$

as we sought to show.

[a] This is a *rng* integral extension. Show every $x \in \mathfrak{p}^i$ is a solution to a $\mathfrak{p}^{i+1}$ polynomial

You may ask what is the absolute norm of a prime $\mathfrak{p}$? If $\mathfrak{p}$ were principal, we know it is p^f for appropriate f. For the non-principal case, in corollary 22.3.14, we shall show that if $\mathfrak{p}_k$ lies over p, then $\mathfrak{N}(\mathfrak{p}_k) = p^f$ for an appropriate number f that we shall determine later.

Since the absolute norm is multiplicative and the index is always positive, and fractional ideals also form lattices, we may extend it to be defined on fractional ideals and get a group homomorphism:

$$\mathfrak{N} : J_K \to \mathbb{R}^{\times}_{>0} \qquad \mathfrak{a} \mapsto \mathfrak{N}(\mathfrak{a})$$

where $\mathfrak{a} = \prod_{\mathfrak{p}} \mathfrak{p}^{n_{\mathfrak{p}}}$ with $n_{\mathfrak{p}} \in \mathbb{Z}$.

Corollary 22.2.25: Volume Form Reformulation

Let $I \subseteq K$ be a fractional ideal of a number field K. Then:

$$\operatorname{vol}(I) = \mathfrak{N}(I) \operatorname{vol}(\mathcal{O}_K)$$

Proof **:**
This is just reformulating proposition 22.1.48 or lemma 22.1.52

Perhaps include the proof in MIT p.18?

Theorem 22.2.26: Class Number

Let K be a number field and let $Cl(K) = J_K/P_k$ be the class group of K. Then $Cl(K)$ is finite.

Proof **:**

Let $0 \neq \mathfrak{p} \subseteq \mathcal{O}_K$ be a prime ideal of $\mathcal{O}_K$. By lemma 22.2.23,$\mathfrak{N}(\mathfrak{p}) = p^n$, where $n = [\mathcal{O}_K/\mathfrak{p} : \mathbb{Z}/p\mathbb{Z}]$. Given p, there are only finitely many prime ideals $\mathfrak{p}$ such that $\mathfrak{p} \cap \mathbb{Z} = p\mathbb{Z}$ by proposition 22.1.54 (equivalently, because $p\mathcal{O}_K$ must have a number of terms in it's factorization and $\mathfrak{p} \big| p\mathcal{O}_K$). Thus, there are only finitely many prime ideal of bounded absolute norm. Since every integral ideal factors into prime ideals, we have

$$\mathfrak{N}(\mathfrak{a}) = \mathfrak{N}(\mathfrak{p}_1)^{n_1} \cdots \mathfrak{N}(\mathfrak{p}_k)^{n_k}$$

It follows that there are only a finite number ideals $\mathfrak{a}$ of $\mathcal{O}_K$ with bounded absolute norm $\mathfrak{N}(\mathfrak{a}) \leq M$. Thus, it suffices to show that each class $[\mathfrak{a}] \in Cl(K)$ contains an integral ideal $\mathfrak{a}_1$ satisfying

$$\mathfrak{N}(\mathfrak{a}_1) \leq \lambda$$

for if there were infinitely many classes, then there would be infinitely many distinct ideals with absolute norm less than λ, contradicting our earlier discussion. By theorem 22.1.53, for any ideal $I \in [\mathfrak{a}]^{-1}$, pick $\alpha \in \mathfrak{a}$ such that

$$|\operatorname{Nm}_{K/\mathbb{Q}}(\alpha)| \leq \lambda \mathfrak{N}(I)$$

By corollary 22.2.16 there exists a J such that $(\alpha) = IJ$. By definition, $J \in [\mathfrak{a}]$. Then:

$$\mathfrak{N}(I)\mathfrak{N}(J) = \mathfrak{N}(\alpha) = |\operatorname{Nm}_{K/\mathbb{Q}}(\alpha)| \leq \lambda \mathfrak{N}(I) \implies \mathfrak{N}(J) \leq \lambda$$

but this now gives that there are only finitely many classes, and hence the class group is finite, as we sought to show.

Definition 22.2.27: Class Number

Let $Cl(K)/ = J_K/P_k$ be the class group of K. Then $h_K = [J_K : P_K]$ is called the *class number* of $Cl(K)$ or K.

A consequence of there being a finite class number is that for every ideal $I \subseteq \mathcal{O}_K$, there exists a minimal number n such that $I^n = (\alpha)$ for some $\alpha \in \mathcal{O}_K$. There is also a minimal number n' such that for *any* ideal of $\mathcal{O}_K$, the ideal to the power of n' is principal. This may encourage thoughts of converting the problem on ideals into one on principal ideals, given there is a way of going back to the original ideals. We shall see that this shall sometimes be the case.

Example 22.2.3: Class Groups

1. Naturally, $|Cl(\mathbb{Q})| = 1$ since $\mathbb{Z}$ is a PID
2. Similarly, $|Cl(\mathbb{Q}(i)| = |Cl(\mathbb{Q}(\sqrt{-3})| = 1$ since $\mathbb{Z}[i]$ and $\mathbb{Z}\left[\frac{1+\sqrt{-3}}{2}\right]$ are PID.
3. Every DVR has ideals of the form (π^n) for appropriate uniformizer π. Thus, the ideal group is isomorphic to $\mathbb{Z}$, and the class group is trivial.
4. Consider $Cl(\mathbb{Q}(\sqrt{-5})$. Since $\mathbb{Z}[\sqrt{-5}]$ is not a Unique Factorization Domain (hence Principal Ideal Domain), we will get a non-trivial class group. Let $K = \mathbb{Q}(\sqrt{-5})$. Choose an integral basis
$$\mathcal{O}_K = \langle 1, \sqrt{-5}\rangle_{\mathbb{Z}}$$
so that $D_k = -20$. Hence, the Minkowski bound gives us:
$$M_k = \sqrt{20}\frac{2}{2^2}\frac{4}{\pi} < 4$$
In fact, we have < 3, but we shall purposefully miss-guess the bound to show how to determine when two ideals belong in the same class group. Consider the ideal $(2), (3)$. Assume that $\mathfrak{p}_2 \big| (2)$. Then we know $\mathfrak{N}(\mathfrak{p}_2)\big|\mathfrak{N}((2)) = 4$. To find $\mathfrak{p}_2 \subseteq \mathcal{O}_K$ consider the map $\varphi : \mathcal{O}_K \to \mathcal{O}_K/(2)$. Since $\mathcal{O}_K/(2)$ is finite the pre-image of a prime ideal of a surjective map is prime, we may find prime ideals of $\mathcal{O}_K/(2)$. By the 4th isomorphism theorem, if $\overline{I} \subseteq \mathcal{O}_K/(2)$ if and only if $(2) \subseteq I$, that is $I\big|(2)$, and hence any prime ideal corresponding to prime ideal of $\mathcal{O}_K/(2)$ divides (2); thus we look for $\varphi^{-1}(P) = \mathfrak{p}_2$. To find such a $\mathfrak{p}_2$, think of $\mathcal{O}_K$ as $\mathcal{O}_K \cong \mathbb{Z}[x]/(x^2+5)^a$ so that we can take:
$$\mathcal{O}_K/(2) \cong \mathbb{Z}[x]/(x^2+5, 2) \cong \mathbb{F}_2[x]/(x^2-1) = \mathbb{F}_2[x]/(x+1)^2)$$
Recall that the prime ideals of a ring are in corresponding with the prime ideals of the ring quotiented by the nilradical (see section 21.2.2). But then we immediately see that the only prime ideal in $\mathcal{O}_K/(2)$ is $(1+x)$. Hence
$$\mathfrak{p}_2 = (1+\sqrt{-5}, 2)$$
Since $\mathfrak{N}(\mathfrak{p}_2)\big|4$, the only possibility is that $\mathfrak{p}^2 = 2$, and indeed:
$$\mathfrak{p}_2^2 = (4, 2+2\sqrt{-5}, -4+2\sqrt{-5}) = (2)$$

Is $\mathfrak{p}_2$ principal? If it was, then $\mathfrak{p}_2 = (a + b\sqrt{-5})$. Then:

$$2 = \mathfrak{N}(\mathfrak{p}_2) = |\operatorname{Nm}_{K/\mathbb{Q}}(a + b\sqrt{-5})| = a^2 + 5b^2$$

but it's evident this is not possible, hence it is not principal! Similarly with (3), we get:

$$\mathcal{O}_K/(3) \cong \mathbb{Z}[x]/(x^2 + 5, 3) = \mathbb{F}_3[x]/(x^2 - 1) = \mathbb{F}_2 \oplus \mathbb{F}_3$$

From this, we can see that $3 = \mathfrak{p}_3\mathfrak{p}_3'$ where

$$\mathfrak{p}_3 = (3, 1 + \sqrt{-5}) \qquad \mathfrak{p}_3' = (3, 1 - \sqrt{-5})$$

Finally, notice that $[\mathfrak{p}_2] = [\mathfrak{p}_3]$. To see this, notice their product is in fact principal:

$$\mathfrak{p}_2\mathfrak{p}_3 = (2, 1 + \sqrt{-5})(3, 1 + \sqrt{-5}) = (6, 1 + \sqrt{-5}, (1 + \sqrt{-5})^2) = (1 + \sqrt{-5})$$

Hence, $\mathfrak{p}_3 = (1 + \sqrt{-5})^{-1}\mathfrak{p}_2$. Evidently, $\mathfrak{p}_3, \mathfrak{p}_3'$ are in the same class. Since there are only two elements, it must be that $Cl(K) \cong \mathbb{Z}/2\mathbb{Z}$.

5. Let $K = \mathbb{Q}(\sqrt{-23})$. Choose $\mathcal{O}_K = \langle 1, \frac{1+\sqrt{-23}}{2}\rangle_{\mathbb{Z}}$ so that it's discriminant is $d_k = -23$ and the Minkowski bound is $M_K = \sqrt{23}\frac{4}{\pi}\frac{1}{2} < 4$. Hence, we again check the prime ideals (2) and (3), noting that the minimal polynomial of the generator is $x^2 - x + 6$:

$$\mathcal{O}_K/(2) \cong \mathbb{F}_2[x]/(t^2 - t)$$

Thus, if $t = \frac{1+\sqrt{-23}}{2}$:

$$(2) = (2, t - 1)(2, t)$$

and similarly

$$(3) = (3, t - 1)(3, t)$$

Certainly, each prime multiple of (2) and (3) are in the same equivalence class. Is $[(2, t)] = [(3, t - 1)]$? Choosing good representatives and multiplying, we get

$$(3, t)(2, t) = (6, t) = (t)$$

hence, we again get that $Cl(K) = \mathbb{Z}/2\mathbb{Z}$.

6. Let $K = \mathbb{Q}(\sqrt{15})$ and $\mathcal{O}_K = \langle 1, \sqrt{15}\rangle_{\mathbb{Z}}$. Then doing the same procedure, we get:

$$(3, \sqrt{15})^2 = (3)$$

How do we know that $(3, \sqrt{15})$ is principal or not? We can see that $a^2 - 15b^2 = 3$ has no solutions using modular arithmetic's, but it will not always be feasible to use this method, so we introduce another. Let's say $(3, \sqrt{15}) = (\alpha)$. Then

$$(\alpha^2) = (3)$$

meaning they are associate, hence $\alpha^2 = 3u$ for $u \in \mathcal{O}_K^\times$. Does there exist such a unit? In section 22.2.4, we shall find this structure. For now, it can be taken for granted that $\mathcal{O}_K^\times = \pm\langle 4 + \sqrt{15}\rangle$ (notice the norm of all such numbers is 1). Then from this we see that no such unit exist, hence $(3, \sqrt{15})$ is not principal.

Notice that all of these examples are for quadratic extensions. It is much more difficult to find prime-factorization for non-quadratic extension since the ring of integers is not easy to find. In the next section, we shall see how we may get some information on the factorization of primes when $\mathcal{O}_K$ is not available

[a] Note how this relied on knowing the structure of $\mathcal{O}_K$! See comment at the end of this example

In general, we have $h_K > 1$. If $h_K = 1$ then we can think of the prime factorization of ideals being the same as the prime factorization of elements as is already well-understood. If $K = \mathbb{Q}(\sqrt{d})$ where d is square-free and negative, then the quadradic extensions in which the class number is 1 is

$$-1, -2, -3, -7, -11, -19, -43, -67, -163$$

For real quadratic extensions, there are a lot more, for example:

$$2, 3, 5, 6, 7, 11, 13, 14, 17, 19, 22, 23, 29, 31, 33, \cdots$$

it is unknown where there are infinitely many real quadratic fields of class number 1, or even if there are infinitely many algebraic number fields (of any degree) with class number 1. In general, it is in fact very difficult to determine some structure in the class group given some class of number fields, with an important notable exception cyclotomic extensions which due to the work of Kenkichi Iwasawa we know behave in a regular way.

The importance of studying cyclotomic extensions links back to Fermat's last theorem which the reader surely recalls states that the equation $x^p + y^p = z^p$ has no nonzero integer solutions for $p \geq 3$. For $p = 2$, we take the cyclotomic extension $\mathbb{Q}(\zeta_4) = \mathbb{Q}(i)$ to get

$$x^2 + y^2 = (x + iy)(x - iy)$$

and then we get the Fermat's twin prime theorem that fully solved the problem in this case. If p is prime, then considering the equation $y^p = z^p - x^p$ over $\mathbb{Q}(\zeta_p)$, we may factor this in $\mathbb{Z}[\zeta_p]$ as

$$y \cdot yy \cdots y = (z - x)(z - \zeta_p x) \cdots (z - \zeta^{p-1} x)$$

Thus, if we assume an existence of a solution, we obtain two multiplicative decompositions of the same number in $\mathbb{Z}[\zeta_p]$. Using this, we can show it contradicts unique factorization *provided* that $\mathbb{Z}[\zeta_p]$ is a unique factorization domain. It was due to some mathematician erroneously assuming unique factorization (equivalently for us, that the class number is always 1) that lead to many fake proofs of Fermat's last theorem[16]. Kummer fixed this by introducing the notion of ideal primes, and translated the statement into one where we must show that $p \nmid h_p$ instead of assuming $h_p = 1$. If this is the case, such a prime is called *regular*, and *irregular* otherwise. Unfortunately, not all primes are regular. Of the first 25 primes less than 100, the primes 37, 59, and 67 are irregular. As of writing this book, it is unknown whether there are infinitely many regular primes.

The regular and irregular prime are going to play important roles in our coming theory, but they do not play much of a role in the eventual proof of Fermat's Last theorem. Due to the work of Gerhard Frey, there was a reformulation of Fermat's last theorem into a statement about elliptic curves. This lead to a new development of mathematics which is beyond the scope of this book, but the interested reader may feel free to search up the *Taniyama-Shimura-Weil Conjecture* to see how Fermat's Last Theorem got translated into a problem of elliptic curves.

We end this section with showing a certain Diophantine equation has no integer solution as it combines nicely all the machinery we have done, and some generalizations that is motivation for class field theory. Starting with the Diophantine equation:

[16] It is speculated that this is the proof Fermat was referring to as being too long for the margins of his book

Theorem 22.2.28: No Solution To Diophantine Equation

The equation

$$y^2 = x^3 - 5$$

has no integer solutions

***Proof* :**
Suppose the pair (x, y) is an integer solution. If x is even, then $x^3 - 5$ is odd, and we get

$$y^2 \equiv -1 \bmod 4$$

which is impossible. Next, x and y must be coprime, for if they weren't then their common factor must be 5 (since then $p \big| y^2 - x^3 = -5$), but then 5^2 divides the left hand side, but not the right hand side.

Now, factoring in $\mathbb{Z}[\sqrt{-5}]$, we get

$$(y - \sqrt{-5})(y + \sqrt{-5}) = x^3$$

These two ideals are coprime; for is $\mathfrak{p}$ divides both, then $\mathfrak{p} \big| (x)^3$, so $\mathfrak{p} \big| (x)$, and $y - \sqrt{-5} + y + \sqrt{-5} = 2y \in \mathfrak{p}$ so $\mathfrak{p} \big| (2y)$. Since x is odd and $\mathfrak{p} \big| (x)$, $\mathfrak{p} \big| (y)$. But then this contradicts that (x) and (y) are relatively prime.

Hence, they are co-prime. Thus, by unique factoriations into ideals, we get that

$$(y - \sqrt{-5}) = \alpha^3 \qquad (y + \sqrt{-5}) = \beta^3$$

Finally, since the class group is of size 2 and a third power produces a principal ideal, it must be taht α, β are both principal. Given that the only units are ± 1 (something we shall prove in the next section), we have that:

$$y + \sqrt{-5} = (a + \sqrt{-5})^3$$

Manipulating and comparing coefficients, we get that:

$$1 = 3ba^2 - 5b^3 = b(3a^2 - 5b^2)$$

implying $b = \pm 1$ since $a, b \in \mathbb{Z}$. But then:

$$3a^2 - 5 = \pm 1$$

but now this can be inspected to see that it's impossible, hence no such integer solution exists.

Exercise 22.2.1

1. Let $A, B \subseteq R$ with $B \subseteq A$ and $A \neq R$. Show there exists a $\gamma \in K$ such that $\gamma B \subseteq R$, but $\gamma B \nsubseteq A$
2. Show that $IJ = (I + J)(I \cap J) = \gcd(I, J) \operatorname{lcm}(I, J)$
3. Let $\alpha \in \mathcal{O}_K \setminus \{0\}$. Show that there exists a $\beta \in \mathcal{O}_K$ st $\alpha\beta \in \mathbb{Z} \setminus \{0\}$

4. Let $K = \mathbb{Q}(\sqrt{-3})$. Show that $\mathcal{O}_K \neq \mathbb{Z}[\sqrt{-3}]$. Show that

22.2.4 Dirichlet's Unit Theorem

As we saw in example 22.2.3, it is important to know the unit group of $\mathcal{O}_K$ to find elements in the class group. We start this section by showing that we may embed $\mathcal{O}_K^\times$ to form a lattice, and then use this to find the structure of $\mathcal{O}_K$.

The group $\mathcal{O}_K^\times$ may have some torsion elements, namely all roots of unity that lie in K; this collection is labeled as $\mu(K)$. This group will in fact be determined by the number of embeddings into Minkowski space. On a high-level, the structure of this group is within this diagram:

$$\begin{tikzcd}
K^\times \arrow[r, "\sigma"] \arrow[rr, bend left, "\sigma^\times"] \arrow[d, "N_{K/\mathbb{Q}}"] & K_{\mathbb{R}}^\times \arrow[r, "\log"] \arrow[d, "\mathrm{nm}"] & {[\prod_\tau \mathbb{R}]^+} \arrow[d, "\mathrm{Tr}"] \\
\mathbb{Q}^* \arrow[r, hook] & \mathbb{R}^* \arrow[r, "\log|\ |"] & \mathbb{R}
\end{tikzcd}$$

where we shall write $K_{\mathbb{R}}^\times = (\mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2})^\times$. In particular, the top row of the above commutative diagram can be replaced by:

$$\begin{aligned}
\mathcal{O}_K^\times &= \left\{ r \in \mathcal{O}_K \;\middle|\; N_{K/\mathbb{Q}(n)=\pm 1} \right\} \\
S &= \left\{ y \in K_{\mathbb{R}}^\times \;\middle|\; N(y) = \pm 1 \right\} && \text{norm-one surface} \\
H &= \left\{ x \in \left[\prod_\tau \mathbb{R} \right]^+ \;\middle|\; \mathrm{tr}(x) = 0 \right\} && \text{trace-zero hyperplane}
\end{aligned}$$

Then we get the homomorphism:

$$\mathcal{O}_K^\times \xrightarrow{\sigma} S \xrightarrow{\log} H$$

where $\sigma^\times = \log \circ \sigma$.

Proposition 22.2.29: Roots Of Unity In Group Of Units

Let $\Gamma_K^\times = \sigma^\times(\mathcal{O}_K^\times) \subseteq H$. Then the sequence

$$1 \to \mu(K) \to \mathcal{O}_K^\times \xrightarrow{\sigma^\times} \Gamma_K^\times \to 0$$

is exact.

Proof **:**
Certainly $\mu(K) \subseteq \ker(\sigma^\times)$. Conversely, say $r \in \mathcal{O}_K^\times$ where $\sigma^\times(r) = 0$. Then $|\tau(r)| = 1$ for each embedding $\gamma : K \to \mathbb{C}$, hence $\sigma(r) = (\tau(r))$ is in a bounded domain of the $\mathbb{R}$-vector space $K_{\mathbb{R}}$. On the other hand, $\sigma(r)$ is a point of the lattice $\sigma(\mathcal{O}_K)$ of $K_{\mathbb{R}}$. Thus, the kernel of $\sigma^\times$ can contain only a finite number of elements, and so since it's a finite group, it contains only the roots of unity of $K^\times$.

We now build towards the structure of $\Gamma_K^\times$. Under the right adjustments $\mathcal{O}_K^\times$ can be embedded in a similar way to $\mathcal{O}_K$. Take $\sigma(\mathcal{O}_K^\times) \subseteq \mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2}$. the way a mesh is formed is by taking the logarithm, for $\log(ab) = \log(a) + \log(b)$

$$\sigma^\times : K^\times \to \mathbb{R}^{r_1} \oplus \mathbb{R}^{r_2} \qquad \sigma^\times = \log \circ \sigma$$

Proposition 22.2.30: Multiplicative Group Forms Lattice

Let H be the trace-zero hyperplane. Then $\sigma^\times(\mathcal{O}_K^\times) \subseteq W$ and it is discrete.

Proof **:**
First, the map $(\mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2})^\times \to \mathbb{R}^{r_1+r_2}$ mapping

$$(a_1, a_2, ..., a_{r_1}, b_1, b_2, ..., b_{r_2}) \mapsto (\log(a_1), ..., \log(a_{r_1}), ..., \log(b_{r_2}))$$

is a proper topological map, in particular the image of a discrete set is discrete. Since $\sigma(\mathcal{O}_K) \subseteq \mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2}$ is discrete. $\sigma(\mathcal{O}_K \setminus \{0\}) \subseteq (\mathbb{R}^{r_1 \oplus r_2})^\times$ is discrete, Next, for $\epsilon > 0$, if $\sigma^\times(\alpha) \subseteq \overline{B_\epsilon}$, then $\forall \varphi \in \mathrm{Hom}_{\mathbb{Q}}(K, \mathbb{C})$, $e^{-\epsilon} < |\varphi(\alpha)| < e^\epsilon$. Now, since the ball is compact, by properness there are only finitely many α, but that completes the proof.

Theorem 22.2.31: Dirichlet Unit Theorem

The group of units $\mathcal{O}_K^\times$ is the direct product of the finite cyclic group $\mu(K)$ and a free abelian group of rank $r_1 + r_2 - 1$. The free-generators of $\mathcal{O}_K^\times$ are called the *fundamental units* of $\mathcal{O}_K^\times$:

$$\mathcal{O}_K \cong_{\mathbb{Z}} \mathbb{Z}^{r_1+r_2-1} \oplus \mu(K)$$

We may also characterize $\mathcal{O}_K^\times$ by the following short exact sequence:

$$1 \to \mu(K) \to \mathcal{O}_K^\times \to \Gamma_K^\times \to 0$$

where since the right term is a free $\mathbb{Z}$-module, the sequence splits giving us $\mathcal{O}_K^\times \cong \mu(K) \oplus \Gamma_K^\times$.

Proof **:**
(Marcus p.100 looks good)

Thus, there exist units $u_1, u_2, ..., u_t$, $t = r_1 + r_2 - 1$ such that for any other unit $u \in \mathcal{O}_K^\times$, it can be written as

$$u = \zeta u_1^{n_1} \cdots u_t^{n_t}$$

where $\zeta \in \mu(K)$ is a root of unity.

Example 22.2.4: Unit Group And Class Group

1. Let $K = \mathbb{Q}(\sqrt{2})$. Then $1 + \sqrt{2}$ generates the free part of the group of units. Similarly, for $K = \mathbb{Q}(\sqrt{3})$, $2 + \sqrt{3}$ generates the free part of the group of units.
2. Let us compute the class group and unit group of $\mathbb{Q}(\sqrt[3]{7})$ Let $K = \mathbb{Q}(\sqrt[3]{7})$ and $s = \sqrt[3]{7}$.

We again look for the Minkowski bound. Doing a similar computation as before, we may workout

$$\mathcal{O}_K = \langle 1, \sqrt[3]{7}, \sqrt[3]{7}^2 \rangle$$

Hence, the discriminant is:

$$\det \begin{pmatrix} 1 & \sqrt[3]{7} & \sqrt[3]{7}^2 \\ 1 & \sqrt[3]{7}\zeta_3 & \sqrt[3]{7}^2\zeta_3^2 \\ 1 & \sqrt[3]{7}\zeta_3^2 & \sqrt[3]{7}^2\zeta_3 \end{pmatrix} = -1323$$

Thus:

$$M_K = \sqrt{1323}\frac{6}{27}\frac{4}{\pi} < 11$$

Hence, we check prime ideal $(2), (3), (5), (7)$ of $\mathbb{Z}$. Since the minimal polynomial of $\sqrt[3]{7}$ is $x^3 - 7$, we see that in $\mathbb{F}_p[x]/(\overline{x^3 - 7}_p)$ we get:

(2) $(x+1)(x^2+x+1)$, hence

$$\mathfrak{p}_2 = (2, s+1) \qquad \mathfrak{p}_2' = (2, s^2+s+1)$$

(3) $(x-1)^3$, hence

$$\mathfrak{p}_3 = (3, s-1)$$

(5) $(x+2)(x^2+3x+4)$, hence

$$\mathfrak{p}_5 = (5, s+2) \qquad \mathfrak{p}_5' = (5, s^2+3s+4)$$

(7) x^3, hence

$$\mathfrak{p}_7 = (7, s) = (s)$$

Thus, just like before we have that $[\mathfrak{p}_2]$, $[\mathfrak{p}_3]$, and $[\mathfrak{p}_5]$ generate $Cl(K)$ (since $[\mathfrak{p}_7] = [1]$). Next, we see that

$$\mathfrak{p}_2\mathfrak{p}_3 = (6) \qquad \mathfrak{p}_3^2\mathfrak{p}_5 = (15)$$

Hence, $[\mathfrak{p}_3] = [\mathfrak{p}_2^{-1}]$ and $[\mathfrak{p}_3^2] = [\mathfrak{p}_5^{-1}]$. Then, with some computation we see that:

$$[\mathfrak{p}_3^3] = [\mathfrak{p}_3]^3 = [\mathrm{id}]$$

Thus, we see that $Cl(K)$ is generated by $[\mathfrak{p}_3]$ and has order 3, hence $CL(K) \cong \mathbb{Z}/3\mathbb{Z}$.

However, what if $\mathfrak{p}_3$ is in fact principal? For the sake of contradiction, let's say

$$\mathfrak{p}_3 = (\alpha) \implies \mathfrak{p}_3^3 = (3) = (\alpha^3)$$

Then $\alpha^3 = 3u$ for some unit u. To find u, we find the units of the ring of integers. To find its units, first notice that if:

$$1 = x^3 - 7 = (x-s)(x^2+sx+s^2)$$

Then $x - s$ is a unit. But then, $x = 2$ works since $8 - 7 = 1$. Then by Dirichlet's Unit Theorem, we have that the rank of the unit group is:

$$r_1 + r_2 - 1 = 1 + 1 - 1 = 1$$

Hence, the group of units is

$$\langle 2 - \sqrt[7]{2} \rangle$$

which is the second object asked of us to find. Using this, we can finally conclude that there is no unit u that satisfies the equation $\alpha^3 = 3u$ (if $\alpha = a + bs + cs^2$, in the process of finding the coefficients a, b, c we shall be forced to have $b = c = 0$). Hence, we may properly conclude that

$$Cl(K) \cong \mathbb{Z}/3\mathbb{Z}$$

completing the proof.

In general, there is an algorithm for computing the ring of units that involved continued fractions. It is not covered in this book, but for those interested see *Borevich and Shafarevich, Number Theory* section 7.3 of chapter 2.

Proposition 22.2.32: Volume Of Unit Lattice

Let $\mathcal{O}_K$ be a ring of integers with unit group $\mathcal{O}_K^*$. Then the volume of the unit lattice $\sigma^\times(\mathcal{O}_K^\times)$ in the hyper-plane H is:

$$\operatorname{vol}\left(\sigma^\times(\mathcal{O}_K^\times)\right) = \sqrt{r_1 + r_2}R$$

where R is the absolute value of the determinant of an arbitrary minor rank $t = r + s - 1$ of the following matrix:

$$\begin{pmatrix} \sigma_1^\times(u_1) & \cdots & \sigma_1^\times(u_t) \\ \vdots & \ddots & \ddots \\ \sigma_{t+1}^\times(u_1) & \cdots & \sigma_{t+1}^\times(u_t) \end{pmatrix}$$

Proof **:**
computations

Definition 22.2.33: Regulator

The value of R in the above proof is called the *regulator* of $\mathcal{O}_K$.

Exercise 22.2.2

1. What number rings $\mathcal{O}_K$ have finiately many units?

22.3 Extension of Dedekind Domains and Number Fields

In section 20.5.2, we covered how prime ideals of a base ring correspond to prime ideals in an integral extension. Since all extensions that we are working with are integral extensions, the same theory applies. Working over Dedekind domains, we may further ask if the integral extension of a Dedekind domain is a Dedekind domain. This will be the case, and so we may further factor prime ideals in the extended ring, namely if a prime ideal $\mathfrak{p} \subseteq R$ factors in S where S/R is an integral extension. When

focusing on rings of integers $\mathcal{O}_L$, the exploration of how primes factor into larger primes has a lot of consequences on the structure of $\mathcal{O}_L$ itself, and perhaps a bit surprisingly at first sight on the field extension L/K itself. We shall see in the next section that we can use the theory of factorization of prime ideals to classify all abelian extension over $\mathbb{Q}$, this is the *Kronecker-Weber Theorem* promised in example 18.1.8.

First, if R is a Dedekind domain with field of fraction K, $I \subseteq R$ an ideal, through them map $K \hookrightarrow L$ we get $I^e \subseteq \overline{R}^L$. We further generalize this for fractional ideals:

Definition 22.3.1: Extension Of Fractional Ideals

Let R be a Dedekind domain, K it's field of fractions, and L/K be a finite field extension. Then if $I \subseteq K$ is a fractional ideal and $\mathcal{O} = \overline{R}^L$,

$$I_L := I\mathcal{O} = \left\{ \sum a_i b_i \;\middle|\; a_i \in I,\ b_i \in \mathcal{O} \right\}$$

When necessary, the notation $I_{L/K}$ will be used to specify the base field over which I is taken.

It can be verified that $I_L \subseteq L$ is the minimal fractional ideal containing I, that if $M/L/K$, then $(I_L)_M = I_M$, and that $I_L = I \otimes_{\mathcal{O}_K} \mathcal{O}$. The results of lemma 20.5.9 follow quickly giving all the same results. Furthermore since $(I+J)_L = I_L + J_L$:

$$(\underset{K}{\gcd}(I,J))_L = \underset{L}{\gcd}(I_L, J_L)$$

However, by lemma 20.5.9, we so far only have

$$(\mathrm{lcm}_K(I,J))_L \subseteq \mathrm{lcm}_M(I_L, J_L)$$

Note that for $\alpha \in K$, $(\alpha)_L = (\alpha\mathcal{O}_K)_L = \alpha\mathcal{O}_L$, hence principal ideals extend to principal ideals. For reasons that shall become clearer later[17], the map

$$J_K \to J_L \qquad I \mapsto I_L$$

is called the *base change map*. To study the fractional ideals of L, the first immediate result we may want is that $\overline{R}^L$ remains a Dedekind domain:

Lemma 22.3.2: Extending Dedekind Domains

Let R be a Dedekind domain with field of fraction K and let L/K be a finite field extension of K. Then $\mathcal{O} = \overline{R}^L$ is a Dedekind domain.

Proof **:**
By definition, $\mathcal{O}$ is integrally closed, and if $\mathfrak{p} \subseteq \mathcal{O}$ is prime, then $P = \mathfrak{p} \cap R$ is also prime (the contraction of a prime ideal is prime), but P is then maximal in R; hence $(\mathcal{O}/\mathfrak{p})/(R/P)$ is an integral extension over a field meaning $\mathcal{O}/\mathfrak{p}$ is a field, thus $\mathfrak{p}$ is maximal. It remains to show that $\mathcal{O}$ is Noetherian.

For now, we shall prove this for the separable case (the non-separable case has an easier proof with

[17] we shall possibly increase the number of primes, hence change the "basis"

more machinery, and the separable case is all that is needed for number fields). Let $\alpha_1, \alpha_2, ..., \alpha_n$ be a K -basis for L. By multiplying by an appropriate constant, we may let $\alpha_1, \alpha_2, ..., \alpha_n \in \mathcal{O}$ (see proposition 22.1.5). By proposition 22.1.26,

$$\mathcal{O} \subseteq \frac{R\alpha_1}{d} + \cdots + \frac{R\alpha_n}{d}$$

Now, for any ideal $I \subseteq \mathcal{O}$, by the above it is contained in a finitely generated R-module, hence since R is Noetherian I (as a R-module) is finitely generated. But then *a fortiori*, it is a finitely generated $\mathcal{O}$-module, hence a finitely generated ideal, making $\mathcal{O}$ Noetherian.

Example 22.3.1: Dedekind Domains Via Extension

The result above shows that all rings of integers of a finite extension $L/K/\mathbb{Q}$ is a ring of integers. For an example that is not a number ring, take $R = k[x]$ for some field k. As R is a PID, it is a Dedekind domain. Then if $L/k(x)$ is an integral extension, the integral closure of R over L, $S = \overline{R}^L$, is a Dedekind domain. As $L \cong k(x)[y]/(p(y))$ for some irreducible polynomial over $k[x]$ L represents 1-dimensional curves. In section ref:HERE, we shall briefly show these curves are all 1-dimensional smooth curves.

Extending Number Fields

An immediate corollary of lemma 22.3.2 is the following:

Corollary 22.3.3: Extending Number Fields

Let L/K be a finite extension of numbers fields so that $\mathcal{O}_K = \overline{\mathbb{Z}}^K$ and $\mathcal{O} = \overline{\mathcal{O}_K}^L$. Then

$$\mathcal{O} = \mathcal{O}_L = \overline{\mathbb{Z}}^L$$

***Proof* :**

We have to show that $\mathcal{O}_L = \overline{\mathbb{Z}}^L \overset{!}{=} \overline{\mathcal{O}_k}^L = \mathcal{O}$, namely the $\overset{!}{=}$ equality. The $\subseteq$ direction is trivial, so consider the other direction: let $\alpha \in \mathcal{O}$. Then since $\mathcal{O}$ is Dedekind, it is Noetherian, it is finitely generated as a $\mathcal{O}_K$-module. Since $\mathbb{Z}[\alpha] \subseteq \mathcal{O}$:

$$\mathbb{Z}[\alpha] \subseteq \mathcal{O}_K a_1 + \cdots + \mathcal{O}_K a_n$$

Since each $\mathcal{O}_K$ is a finitely generated free $\mathbb{Z}$-module, we have that $\mathbb{Z}[\alpha]$ is a Noetherian $\mathbb{Z}$-module, hence finitely generated. But then by proposition 20.5.3 α is integral over $\mathbb{Z}$, and so $\alpha \in \overline{\mathbb{Z}}^L$, completing the equality.

Thus, when working with finite extensions of number ring L/K, there is a corresponding integral

extensions are those of the ring of integers $\mathcal{O}_L/\mathcal{O}_K$ creating the diagram:

$$\begin{tikzcd} K \arrow[r, "\iota"] \arrow[d] & L \arrow[d] \\ \mathcal{O}_K \arrow[r, "\iota"] & \mathcal{O}_L \end{tikzcd}$$

As we saw, principal ideals extend to principal ideals. As it turns out, all non-principal ideals extend to be principal given a large enough extension. This result shall underpin a lot of future simplification and is good to see early on:

Lemma 22.3.4: Number Fields, Extension To Principal

Let $I \subseteq K$ be a fractional ideal where $K/\mathbb{Q}$ is a number field. Then in some finite extension L/K, I_L becomes a principal ideal.

Proof **:**
As $K/\mathbb{Q}$ is a number fields, by theorem 22.2.26, $Cl(K)$ is finite. Thus, for every $I \in J_K$, there exists an $n \in \mathbb{N}$ such that $I^n = \alpha\mathcal{O}_K$. Let $L = K(\sqrt[n]{\alpha})$. letting $\beta = \sqrt[n]{\alpha}$, we get

$$(I_L)^n = (I_L^n) = (I^n\mathcal{O}_L) = (\alpha\mathcal{O}_L) = (\beta\mathcal{O}_L)^n$$

Thus, by unique factorization:

$$I_L = \beta\mathcal{O}_L$$

as we sought to show.

This result is quite useful; in the right cases we can pass to the field in which all our ideals become principal ideals, proof what we want for principal ideals, then try to come back down (coming back down is often called the descent argument). The following lemma gives a very common descent argument, and is important for computing the absolute norm of extended ideals: since the norm of an ideal gives a measure how space a given ideal is within its Dedekind, extending the domain shall make the given ideal sparser[18]:

Lemma 22.3.5: Number Fields, Extension And Absolute Norm

Let $L/K/\mathbb{Q}$ be a finite extension of number fields and $I \subseteq K$ a fractional ideal. Then:

$$\mathfrak{N}_{L/\mathbb{Q}}(I_L) = \mathfrak{N}_{K/\mathbb{Q}}(I)^{[L:K]}$$

Proof **:**
If $I = (\alpha) = \alpha\mathcal{O}_K$ is principal, then $\mathrm{Nm}(I) = \left|\mathrm{Nm}_{K/\mathbb{Q}}(\alpha)\right|$ and $\mathrm{Nm}(I_L) = \left|\mathrm{Nm}_{L/\mathbb{Q}}(\alpha)\right|$. Since $\alpha \in K$, by proposition 22.1.9

$$\mathrm{Nm}_{L/\mathbb{Q}}(\alpha) = \mathrm{Nm}_{K/\mathbb{Q}}(\alpha)^{[L:K]}$$

[18]This means that the ideal doesn't "extend enough" to cancel the added density given by extending the ring

In general, $I^n = \alpha\mathcal{O}_K$, thus using the multiplicity of the absolute norm:

$$\mathfrak{N}_{L/\mathbb{Q}}(I_L)^n = \mathfrak{N}_{L/\mathbb{Q}}(I_L^n) = \mathrm{Nm}_{L/\mathbb{Q}}(\alpha\mathcal{O}_L) = \mathrm{Nm}_{L/\mathbb{Q}}(\alpha\mathcal{O}_K)^{[L:K]} = \mathfrak{N}(I^n)^{[L:K]} = \mathfrak{N}_{K/\mathbb{Q}}(I)^{n[L:K]}$$

taking the nth root, and we're done

22.3.1 Ideal Correspondence in Extension

Let L/K be a finite extension, J_K and J_L be the corresponding ideal group. Then we have a map $J_K \to J_L$ given by $I \mapsto I_L$. Do we have a converse map, and if so is it an inverse of the natural extension map? If not, to what degree is it an inverse map? Recall in proposition 20.5.8 that $I \subseteq SI \cap R$, in other words $I \subseteq I_S \cap R$. This can be made an equality for when extending number fields, and further generalized to fractional ideals. Recall that this is not true in general (a counter-example for the general case is $(2) \subsetneq \mathbb{Z} = (2)_{\mathbb{Q}} \cap \mathbb{Z}$), hence this is special for extension of Dedekind domains:

Proposition 22.3.6: Ideal Correspondence In Extensions Of Dedekind Domains

let L/K be a finite field extension with Dedekind domain $\mathcal{O}$ and $\mathfrak{o}$ respectively. Let $I \subseteq K$ a fractional ideal. Then:

$$I_L \cap K = (I^e)^c = I$$

Hence, unlike for general integral extensions, the contraction is an inverse map to the extension.

Proof **:**
Let $\beta \in J = I\mathcal{O} \cap K$ so that $\beta = \sum_i a_i k_i$ where $a_i \in I_L$ and $k_i \in \mathfrak{o}$ (since $\mathcal{O} \cap K = \mathfrak{o}$). Then the following equality holds:

$$\beta^n = \mathrm{Nm}_{L/K}(\beta) = \prod_\tau \tau(\beta) = \prod_\tau \left(\sum_i k_i \tau(a_i) \right)$$

where the first equality comes from $\beta \in K$, the second comes from the equivalence definition of norm (given an appropriate normal closure N/L), and the third is expanding the definition of β. On the right hand side, we have that the sum is invariant under the Galois group $\mathrm{Gal}_K(N)$ for appropriate normal closure N/L from which we took each τ. Thus, the sum is an element in $\mathfrak{o}$, so the product is in $\mathfrak{o}$. Then we have $\beta^n \in I^n$, so $\beta \in I$, completing the proof.

Due to the importance of this new paradigm of thinking, two other proofs shall be presented. Notice that $IJ^{-1} \subseteq \mathfrak{o}$ and:

$$\mathfrak{N}(IJ^{-1}) = 1$$

giving that $IJ^{-1} = \mathfrak{o}$, i.e. $I = J$.

Another proof goes as follows: Since $I \subseteq J$, $I_L \subseteq J_L$. Since $I_L \subseteq J_L \subseteq I_L$ by properties of extensions, $I_L = J_L$. Thus:

$$\mathfrak{N}(J) = \mathfrak{N}(I_L)^{1/[L:K]} = \mathfrak{N}(I)$$

and since $I \subseteq J$, the norms being equal means the index between the two are the same, hence $I = J$.

Note that in the case where $\mathfrak{p}$ is prime, the result can be essentially immediately proven since by proposition 20.5.8 $\mathfrak{p} \subseteq \mathfrak{p}\mathcal{O}_L \cap K$, but since $\mathfrak{p}$ is maximal we get equality.

Corollary 22.3.7: Homomorphism between Extensions of Ideal Group

Let J_K be the ideal group of a Dedekind domain R and let L/K be a finite field extension. Then the map $J_K \to J_L$ mapping $I \mapsto I_L$ is an injective group homomorphism that preserves gcd and lcm.

On the other hand, the map $Cl(K) \to Cl(L)$ is a homomorphism that need not be injective.

For class groups of number fields:

$$\varinjlim_{\substack{K \\ [K:\mathbb{Q}]<\infty}} (Cl(K)) = 0$$

***Proof* :**
By the previous proposition, $I_L = J_K$ if $I = J$, hence the map $\varphi : \mathscr{F}(K) \to \mathscr{F}(L)$ mapping $I \mapsto I_L$ is injective and is certainly a homomorphism by a simple generalization of lemma 20.5.9 (recall that $(\mathfrak{a}\mathfrak{b})^e = \mathfrak{a}^e\mathfrak{b}^e$ for ideals).

For the map between class groups of number fields, since the extension for every fractional ideal $I \subseteq \mathbb{Q}$, there exists an extension in which it is principle, we see that the colimit can only be satisfied by the trivial group, hence $\varinjlim_{\substack{K \\ [K:\mathbb{Q}]<\infty}} (Cl(K)) = 0$, as we sought to show.

A side consequence of the colimit definition is that we may define

$$Cl(\overline{\mathbb{Q}}) := \varinjlim(Cl(K)) = 0$$

This shows that to some degree, the lack of principal ideals is tied to what algebraic extension we choose. Though the extension map is injective, it is not in general surjective:

Example 22.3.2: Lack of Surjectivity
Let $L = \mathbb{Q}(\sqrt{-5})$ and $K = \mathbb{Q}$ so $\mathcal{O}_L = \mathbb{Z}[\sqrt{-5}]$. Then the (non-principal) ideal $I = (1+\sqrt{-5}, 1-\sqrt{-5})$ are all numbers of the form $m + n\sqrt{-5}$ where m, n are even, so $I \cap \mathbb{Q} = (2)$, hence I lies over (2). However, $I \neq (2)_L$ since I is not principal (recall $(-)_L$ maps principals to principals), and no other ideal can map to I or else I would lie over some different prime, hence I is not an extension. The ideal (2) is still related to I, namely $I^2 = (2)$ (this can be used to show I is not principal).

With these results, we have better control over the extension of primes and primes lying over (resp. lying under) relate:

Proposition 22.3.8: Dedekind Domain, Lying Over And Lying Under

Let $\mathcal{O}/o$ be a finite extension of Dedekind domains given by L/K, $0 \neq \mathfrak{p} \subseteq o$ a prime, and $\mathcal{O} = \overline{o}^L$, $\mathfrak{p}$, where $\mathfrak{p}$ may factor further. Then $\mathfrak{p}$ will never be trivialized in an extension, namely

$$\mathfrak{p}\mathcal{O} \neq \mathcal{O} = (1)$$

Proof **:**
Let $p \in \mathfrak{p} \setminus \mathfrak{p}^2$. Then

$$(p) = \mathfrak{p}\mathfrak{a} \qquad \mathfrak{p} \nmid \mathfrak{a}$$

Hence, $\mathfrak{p} + \mathfrak{a} = (1)$ ($\gcd(\mathfrak{p}, \mathfrak{a}) = (1)$), and so $a + b = 1$ for $a \in \mathfrak{p}$ and $b \in \mathfrak{a}$. Certainly $b \notin \mathfrak{p}$ so $b\mathfrak{p} \subseteq \mathfrak{p}\mathfrak{a} = (p)$. Now, if $\mathfrak{p}\mathcal{O} = \mathcal{O}$, then $b\mathcal{O} = b\mathfrak{p}\mathcal{O} \subseteq p\mathcal{O}$. hence $b\mathcal{O}_K \subseteq \mathfrak{p}$ so $b = px$ for some $x \in \mathcal{O} \cap K = \mathcal{O}_K$, implying $b \in \mathfrak{p}$ – a contradiction.

Proposition 22.3.9: Equivalence Of Prime Division In Dedekind Domain

Let $\mathfrak{p} \subseteq R$, $\mathfrak{B} \subseteq \mathcal{O}$, $\mathcal{O}/R$ an integral extension where R is a Dedekind Domain. Then the following are equivalent:

1. $\mathfrak{B} \big| \mathfrak{p}\mathcal{O}$
2. $\mathfrak{B} \supseteq \mathfrak{p}\mathcal{O}$
3. $\mathfrak{B} \supseteq \mathfrak{p}$
4. $\mathfrak{B} \cap \mathcal{O} = \mathfrak{p}$
5. $\mathfrak{B} \cap K = \mathfrak{p}$

Proof **:**
Since we are in a Dedekind domain, $A \big| B$ if and only if $A \supseteq b$, giving us $1 \Leftrightarrow 2$. Since $\mathfrak{B}$ is an ideal of $\mathcal{O}_L$, certainly $2 \Leftrightarrow 3$. We got $3 \Leftarrow 4 \Leftrightarrow 5$ by the above injectivity argument. For $3 \implies 4$, since $\mathfrak{p}$ is maximal and $\mathfrak{B} \cap \mathcal{O}_K \supseteq \mathfrak{p}$, we have either equality or $\mathcal{O}_K$, if $\mathfrak{p} \cap \mathcal{O}_K = \mathcal{O}_K$, then $1 \in Q$, implying $Q = \mathcal{O}_L$, contradicting it's maximality.

Corollary 22.3.10: Lying Over Unique Prime

Let $\mathfrak{p}, \mathfrak{q} \subseteq \mathcal{O}$, $\mathfrak{p} \neq \mathfrak{q}$. Let:

$$\mathfrak{p}\mathcal{O}_L = \mathfrak{B}_1^{e_1} \cdots \mathfrak{B}_n^{e_n} \qquad \mathfrak{q}\mathcal{O}_L = \mathfrak{C}_1^{f_1} \cdots \mathfrak{C}_m^{f_m}$$

Then:

$$\{\mathfrak{B}_1, ..., \mathfrak{B}_n\} \cap \{\mathfrak{C}_1, ..., \mathfrak{C}_n\} = \emptyset$$

that is, extended prime ideals lie over a single prime

***Proof* :**
Let's say the intersection is non-empty, so that $\mathfrak{a}$ is a common factor. Then write:

$$\mathfrak{p}\mathcal{O}_L = \mathfrak{a}^{e_1} \cdots \mathfrak{B}_n^{e_n} \qquad \mathfrak{q}\mathcal{O}_L = \mathfrak{a}^{f_1} \cdots \mathfrak{C}_m^{f_m}$$

Then:

$$(\mathfrak{p})_L^{-1}\mathfrak{p}^{e_1-1}\mathfrak{B}_2^{-e_2} \cdots \mathfrak{B}_n^{-e_n} = \mathfrak{a} = (\mathfrak{q})_L^{-1}\mathfrak{p}^{f_1-1}\mathfrak{C}_2^{-f_2} \cdots\cdots \mathfrak{C}_m^{-f_m}$$

Thus we get:

$$\mathfrak{q}\mathcal{O}_L = (\mathfrak{p})_L\mathfrak{p}^{e_1-1}\mathfrak{B}_2^{-e_2} \cdots \mathfrak{B}_n^{-e_n}\mathfrak{p}^{f_1-1} \cdots \mathfrak{C}_m^{f_m}$$

And in particular:

$$\mathfrak{C}_1^{f_1} \cdots \mathfrak{C}_m^{f_m} = q\mathcal{O}_L \subseteq p\mathcal{O}_L = \mathfrak{B}_1^{e_1} \cdots \mathfrak{B}_n^{e_n}$$

But then, since $\mathfrak{p}\mathcal{O}_L \cap K = \mathfrak{p}$ and $\mathfrak{q}\mathcal{O}_L \cap K = \mathfrak{q}$, we have by corollary 22.3.7 that:

$$\mathfrak{q} \subseteq \mathfrak{p}$$

which since $\mathfrak{q}$ is maximal and $\mathfrak{p} \neq \mathcal{O}_K$ we have $\mathfrak{q} = \mathfrak{p}$, contradicting the fact that $\mathfrak{p} \neq \mathfrak{q}$, hence, it must be that they share no common factors.

Another proof goes as follows: if $\mathfrak{p} \subseteq \mathcal{O}_K$ is a prime and $\mathfrak{p}_L = \mathfrak{B}_1^{e_1} \cdots \mathfrak{B}_k^{e_k}$ is the factorization of $\mathfrak{p}_L$ in $\mathcal{O}_L$, then if there was a prime ideal $\mathfrak{q}$ such that $\mathfrak{q} \mapsto \mathfrak{B}_i$ so that by the above for any $1 \leq i \leq k$, then since $\mathfrak{p} \subseteq \mathfrak{B}_i \cap \mathcal{O}_k$, by maximality $\mathfrak{p} = \mathfrak{B}_i \cap \mathcal{O}_K$. Hence if $\mathfrak{q} = \mathfrak{B}_I \cap \mathcal{O}_k$, by injectivity $\mathfrak{p} = \mathfrak{q}$.

Another consequence of the above theorem is that if two prime factorizations of a prime share a factor in an extended ring of integers, they are equal.

Relative Norm

Since the ideals of Dedekind domains share factorization structure to numbers, we may hope there is an equivalent of the field norm $\mathrm{Nm}_{L/K}$ for ideals. by lemma 22.3.4, we may hope that our new relative norm map gives us some information on the power to which we may rise our ideal to get a principal ideal. We now endeavour to find a map $J_L \to J_K$ that behaves more like the field norm and generalizes the absolute norm, which can be thought of as a map:

$$\mathfrak{N}_{L/\mathbb{Q}} : J_L \to \mathbb{Q}_L \cong \mathbb{Q}_{>0}^{\times}$$

In particular, if we work of *Galois* extensions L/K, then we may hope to translate some of the fact that

$$\mathrm{Nm}_{L/K}(\alpha) = \prod_{\sigma} \sigma(\alpha)$$

to prime ideals! In particular, this allows us to bring in the Galois theory in the study of prime ideals!

We start by showing the case for Galois extensions, the general case can be inferred from the Galois case. Given a Galois extension L/K with Galois group $G = \mathrm{Gal}_K(N)$, then we can recover L from K using G by taking $L^G = K$, hence $\mathcal{O}_L^G = \mathcal{O}_K$. Certainly, if $\tau \in G$ and $\mathfrak{a} \subseteq L$, then $\tau(\mathfrak{a}) \subseteq L$ Since L/K is normal, and hence $\tau(\mathfrak{a})$ is well-defined. It may be tempting to then say that $J_L^G = J_K$ and hence defining $J_L \to J_K$ by taking

$$I \mapsto \prod_{\tau} \tau(I)$$

However, it is not the case that $J_L^G = J_K$; the ideal in example 22.3.2 is invariant under the Galois group, but certainly an ideal of $\mathbb{Z}$. We fix this by the following:

Definition 22.3.11: Relative Norm

Let L/K be a finite field extension and $I \subseteq K$ a fractional ideal in L. Then the *relative norm* or *ideal norm* of L with respect to K is:

$$\mathfrak{N}_{L/K}(I) = \left(\left\{\mathrm{Nm}_{L/K}(\alpha) \;\middle|\; \alpha \in I\right\}\right) \in J_K$$

namely, $\mathrm{Nm}_{L/K}(I)$ is the ideal generated by all norms of elements in I.

Note that this is well-defined since $\mathrm{Nm}_{L/K}(\alpha) \in K$ as every minimal polynomial $m_{\alpha,K}(x) \in K[x]$. The norm as defined above gives us a map $\mathfrak{N}_{L/K} : J_L \to J_K$. The field norm can be thought of as the ideal norm, for example take

$$\mathfrak{N}_{\mathbb{Q}(\sqrt{2})/\mathbb{Q}}((\sqrt{2})) = \left(\left\{\mathrm{Nm}_{L/K}\left(\sqrt{2}b\right) \;\middle|\; \sqrt{2}b \in \sqrt{2}\mathbb{Z}\right\}\right) = (2)$$

suggesting that $(\mathrm{Nm}_{L/K}(\alpha)) = \mathfrak{N}_{L/K}((\alpha))$, which will indeed be the case and shall be proved after the next theorem. An important reason to choose this definition is that it works well with extension: Let $J_L \xrightarrow{\mathrm{Nm}_{L/K}} J_K \xrightarrow{I_L} J_L$. What will be the image of an ideal in this chain? In the simplest case, we may take principal ideal $\alpha\mathcal{O}_L \subseteq \mathcal{O}_L$ where $\alpha \in \mathcal{O}_K$. Then it's image is $\alpha\mathcal{O}_K$, and it's extension is $\alpha\mathcal{O}_L$. Thus for the simplest type of integral principal ideal, the map is a right-inverse. For more general cases, it is not necessarily a right inverse (it cannot be, since I_L is not surjective). However, the behavior of the image is well-behaved, and in fact tied with the "almost" definition of the ideal norm presented earlier:

Theorem 22.3.12: Ideal Correspondence in Extension

Let $N/L/K$ be finite field extension where N/K is a normal extension so that $G = \mathrm{Gal}_K(N)$. Then:

$$(\mathfrak{N}_{L/K}(I))_L = \prod_{\tau \in G} (\tau I)_{L/\tau K} =: \mathrm{Nm}'_{L/K}(I)$$

In particular

$$\left\{\begin{matrix}\text{Fractional ideals of} \\ \mathcal{O}_K\end{matrix}\right\} \underset{\mathfrak{N}_{L/K}}{\overset{I_L}{\rightleftarrows}} \left\{\begin{matrix}\text{Fractional Ideals of} \\ \mathcal{O}_L\end{matrix}\right\}$$

Proof **:**
Let $I \subseteq L$ be a fractional ideal. Without loss of generality, we may make I integral. Then, certainly an ideal is a subset of it's extension (given the natural embedding) $\sigma_i I \subseteq (\sigma_i I)_{L/\sigma_i K}$, and so by definition of $\mathrm{Nm}'_{L/K}(I)$, forall $\alpha \in I$:

$$\mathrm{Nm}_{L/K}(\alpha) = \prod_{\tau}(\tau(\alpha)) \in \mathrm{Nm}'_{L/K}(I)$$

Hence $\mathrm{Nm}_{L/K}(I) \subseteq \mathrm{Nm}'_{L/K}(I)$. Extending both sides, we get $(\mathrm{Nm}_{L/K}(I))_L = (\mathrm{Nm}'_{L/K}(I))_L =$

$\mathrm{Nm}'_{L/K}(I)$. Hence, since we are in a Dedekind domain, there exists an [integral] ideal J such that

$$\mathrm{Nm}_{L/K}(I)_L = J\,\mathrm{Nm}'_{L/K}(I)$$

We want to show that $J = (1)$. We shall use lemma 22.6.1, namely that there exists an ideal $B \subseteq K$ relatively prime to $J \cap K$ such that JB is principal, say $JB = (b)$ $(b \in K)$. Then since J is fixed under $\mathrm{Gal}_K(L)$, each $(\tau B)_{L/\tau K})$ is relatively prime to J for each τ. Then since $b \in J$:

$$\begin{aligned}\mathrm{Nm}_{L/K}(I) &\supseteq \left(\prod_\tau \tau(b)\right) \\ &= \left(\prod_\tau (\tau(B))_{L/\tau K)}\right)\left(\prod_\tau (\tau(J))_{L/\tau K}\right)\end{aligned}$$

Thus, $\prod_\tau (\tau B)_{L/\tau K}$ is a multiple of J. However, $(\tau B)_{L/\tau K}$ is relatively prime to J, hence it must be that $J = (1)$, as we sought to show.

Lemma 22.3.13: Ideal Norm Properties

Let $M/L/K$ be algebraic number fields, $\mathfrak{a}, \mathfrak{b} \subseteq M$ be ideals, $I \subseteq K$ an ideal, and $a \in M$ an element. Then:

1. $\mathfrak{N}_{M/K}(a) = (\mathrm{Nm}_{M/K}(a)) = \left(\prod_{\sigma \in \mathrm{Gal}_K(M)} \sigma(a)\right)$
2. $\mathfrak{N}_{L/K}(\mathfrak{N}_{M/L}(\mathfrak{a})) = \mathfrak{N}_{M/K}(\mathfrak{a})$
3. $\mathfrak{N}_{M/k}(\mathfrak{a}\mathfrak{b}) = (\mathrm{Nm}_{M/k}(\mathfrak{a}))\,\mathrm{Nm}_{M/k}(\mathfrak{b})$
4. $I^{[M:K]} = \mathfrak{N}_{M/K}((I)_M)$
5. $(\mathfrak{N}_K(I)) = \mathfrak{N}_{K/\mathbb{Q}}(I)$

Proof **:**
Note that for all of these equalities, we immediately have $\supseteq$, hence we want to show $\subseteq$ for each of these.

1. comes straight from the definition
2. exercise
3. exercise
4. Use lemma 22.3.5 and the finiteness of the class group to make an argument on principal ideals, we see that if we take $J_K \xrightarrow{I_L} J_L \xrightarrow{\mathrm{Nm}_{L/K}} J_K$, we pick up a $[L:K]$ exponent.
5. use part (1)

Corollary 22.3.14: Relative Norm Of Prime Ideal

Let L/K be a finite extension of number fields and let $\mathfrak{B}$ lie above $\mathfrak{p}$. Then if $f = [\mathcal{O}_L/\mathfrak{B} : \mathcal{O}_K/\mathfrak{p}]$, then:

$$\mathfrak{N}_{L/K}(\mathfrak{B}) = \mathfrak{p}^f$$

Proof **:**
Using the above lemma:

$$\begin{aligned}(\mathfrak{N}_L(\mathfrak{p}_L)) &= \mathfrak{N}_{L/\mathbb{Q}}(\mathfrak{p}_L) \\ &= \mathfrak{N}_{K/\mathbb{Q}}(\mathfrak{p}^{[L:K]}) \\ &= (\mathfrak{N}_K(\mathfrak{p}))^{[L:K]}\end{aligned}$$

Thus, $\mathfrak{N}_{L/K}(\mathfrak{B})$ is a power of $\mathfrak{p}$ by the 4th point in the above lemma. and by the 2nd we get:

$$\mathfrak{N}_{K/\mathbb{Q}}(\mathfrak{N}_{L/K}(\mathfrak{B})) = \mathfrak{N}_{K/\mathbb{Q}}(\mathfrak{p}^f) \implies \mathfrak{N}_{L/K}(\mathfrak{B}) = \mathfrak{p}^f$$

as we sought to show.

The relative norm can in fact be defined to be the unique group homomorphism where

$$\mathfrak{N}(\mathfrak{B}) = \mathfrak{p}^{[\mathcal{O}_L/\mathfrak{B}:\mathcal{O}_K/\mathfrak{p}]}$$

which is left as an exercise. By multiplicity, we get that in general:

$$\mathfrak{N}_{L/K}(\mathfrak{p}\mathcal{O}_L) = \prod_i \mathfrak{N}_{L/K}\mathbb{F}_{L/K}(\mathfrak{B}_i)^{e_i} = \prod_i \mathfrak{p}^{f_i e_i}$$

Corollary 22.3.15: Ideal Norm Correspondence Not Galois

Let $M/L/K$ were field extensions, M/K normal but not necessarily L/K. Then

$$(\mathrm{Nm}_{L/K}(I))_M = \prod_{\tau \in \mathrm{Gal}_K(M)/\mathrm{Gal}_L(M)} \tau(I_M)$$

Proof **:**
exercise

With the relative norm defined, it is perhaps natural to think there is a relative discriminant:

Definition 22.3.16: Relative Discriminant

Let L/K be a finite extension of number fields. Then:

$$\mathrm{disc}'_{L/K}(\mathcal{O}_L) = \left(\left\{\mathrm{disc}_{L/K}(\alpha_1, \alpha_2, ..., \alpha_n) \;\middle|\; \alpha_1, \alpha_2, ..., \alpha_n \text{ basis for } \mathcal{O}_L \right\}\right)$$

is the *relative discriminant* of L with respect to K.

This won't come up much, but is useful in a couple of tower arguments later on.

Though we have characterized a correspondence of the ideals, it is more difficult to characterize I_L. In the following section, we characterize prime extensions, and take advantage of the fact that the factorization to give us a characterization of I_L.

22.4 Prime Factorization in Extension

Since $\mathcal{O}$ is a Dedekind domain by lemma 22.3.2, we have that $\mathfrak{p}\mathcal{O}$ decomposes uniquely into a product of prime ideals:

$$\mathfrak{p}\mathcal{O} = \mathfrak{B}_1^{e_1} \cdots \mathfrak{B}_r^{e_r}$$

If it is clear from context, we shall often write $\mathfrak{p}$ instead of $\mathfrak{p}\mathcal{O}$. Sometimes, we shall also write $\mathfrak{p}_L$ for $\mathfrak{p}\mathcal{O}_L$. We shall *almost always* find exactly how $\mathfrak{p}\mathcal{O}$ decomposes, sometimes without requiring knowledge of the ring of integers. When the extension L/K is Galois, we shall have even more control, and if L/K is not Galois, we can take $M/L/K$ to be a Galois extension, work within the Galois, and see if the results can be translated back to L/K.

Let's say $\mathfrak{B}$ lies over $\mathfrak{p}$ and that $\mathfrak{B}$ is in the prime decomposition of $\mathfrak{B}$ [19]. Then we have two important quantities associated to $\mathfrak{B}$: it's power in the prime decomposition, and the index $[\mathcal{O}/\mathfrak{B} : \mathcal{O}_K/\mathfrak{p}]$, that is the degree of the extension of the residue field. These quantities are given a name:

Definition 22.4.1: Ramification Index and Inertia Degree

Let R be a Dedekind domain, $K = \mathrm{Frac}(R)$, L/K be a finite field extension, $\mathcal{O} = \overline{R}^L$ and $\mathfrak{p} \subseteq R$ a prime ideal. If $\mathfrak{B}_i$ is a prime in $\mathcal{O}$ lying above $\mathfrak{p}$, then it's degree is called the *ramification index* and is often denoted

$$e(\mathfrak{B}_i/\mathfrak{p}) \qquad \text{where} \qquad \mathfrak{B}^{e(\mathfrak{B}_i/\mathfrak{p})} \| \, \mathfrak{p}\mathcal{O}$$

or e_i for short, where $\|$ means *maximally divides*. The degree of the field extension

$$f(\mathfrak{B}_i/\mathfrak{p}) = [\mathcal{O}/\mathfrak{B}_i : R/\mathfrak{p}]$$

or f_i for short, and is called is called the *inertia degree* of $\mathfrak{B}_i$ over $\mathfrak{p}$.

For future reference, it will be important to keep track to "how much" a prime ideal splits. This will share strong similarities to the splitting of polynomials[20] (recall how "rare" an irreducible polynomial is inseparable)? This prompts the following vocabulary:

[19] we shall soon show all primes lying above $\mathfrak{p}$ will be part of the prime decomposition

[20] Though there shall be important failures when this comparison will *not* hold! This shall be elaborated soon

Definition 22.4.2: Split, Unramified, and Totally Ramified Primes

Let $\mathfrak{p} \subseteq R$ be a prime ideal of a Dedekind domain R, $K = K(R)$, L/K a finite extension, and $\mathcal{O} = \overline{R}^L$. Then given

$$\mathfrak{p}\mathcal{O} = \mathfrak{B}_1^{e_1} \cdots \mathfrak{B}_g^{e_g}$$

If $g = [L : K]$, then $\mathfrak{p}$ is said to *split completely* ($e_i = f_i = 1$), giving:

$$\mathfrak{p}\mathcal{O} = \mathfrak{B}_1 \cdots \mathfrak{B}_n$$

If $e_i = 1$ for all i (but some f_i are not necessarily 1), then we sy that $\mathfrak{p}$ is *unramified* and (if the residue field is separable):

$$\mathfrak{p}\mathcal{O} = \mathfrak{B}_1 \cdots \mathfrak{B}_g$$

for some $g \leq n$. If $g = 1$, then $\mathfrak{p}$ is said to be *totally ramified*, *indecomposable*, or *nonsplit* if:

$$\mathfrak{p} = \mathfrak{B}_1^{e_1}$$

and we know there is only a single prime ideal $\mathfrak{B}_i$ lying over $\mathfrak{p}$.

Most of the time, we shall work in the case of character zero extensions, hence L/K is separable. In this case, the values e_i and f_i are connected; if $\mathfrak{p}$ factors in an extension, then there is a simple equation the ramification and inertia degree must satisfy which equals the degree of the extension:

Proposition 22.4.3: Fundamental Identity

Let L/K be a finite separable extension, $n = [L : K]$, $\mathcal{O} = \overline{\mathcal{O}_K}^L$, and $\mathfrak{p} \subseteq \mathcal{O}_K$ a prime ideal. If

$$\mathfrak{p}\mathcal{O} = \mathfrak{B}_1^{e_1} \cdots \mathfrak{B}_g^{e_g}$$

is the prime decomposition of $\mathfrak{p}\mathcal{O}$ and f_i is the inertia degree of $\mathfrak{B}_i$ over $\mathfrak{p}$, then

$$\sum_i^g e_i f_i = n$$

***Proof* :**
Using absolute norms, this theorem becomes almost immediate, since:

$$\mathfrak{N}(\mathfrak{p})^{[L:K]} \overset{\text{lem } 22.3.5}{=} \mathfrak{N}(\mathfrak{p}\mathcal{O}) = \mathfrak{N}\left(\prod_i \mathfrak{B}_i^{e_i}\right) \overset{\text{prop } 22.2.24}{=} \prod_i \mathfrak{N}(\mathfrak{B}_i)^{e_i} \overset{\text{cor } 22.3.14}{=} \prod_i \mathfrak{N}(\mathfrak{p})^{e_i f_i} = \mathfrak{N}(\mathfrak{p})^{\sum_i e_i f_i}$$

taking log of both sides gives us:

$$n = [L : K] = \sum_I e_i f_i$$

This can also be proven directly: first by the CRT we have

$$\mathcal{O}/\mathfrak{p}\mathcal{O} \cong \bigoplus_i^r \mathcal{O}/(\mathfrak{B}_i^{e_i})$$

Next, we have that $\mathcal{O}/\mathfrak{p}\mathcal{O}$ and $\mathcal{O}/\mathfrak{B}_i^{e_i}$ are vector-spaces over $k = \mathcal{O}_K/\mathfrak{p}$. Hence, it suffices to show that

$$\dim_k(\mathcal{O}/\mathfrak{p}\mathcal{O}) = n \qquad \dim_k(\mathcal{O}/(\mathfrak{B}_i^{e_i})) = e_i f_i$$

To show the first equality, since $\mathcal{O}$ is a finitely generated $\mathcal{o}$-module, pick $x_1, x_2, ..., x_m \in \mathcal{O}$ such that $\overline{x_1}, ..., \overline{x_n}$ is a k-basis for $\mathcal{O}/\mathfrak{p}\mathcal{O}$. If $x_1, x_2, ..., x_m$ is a basis for L/K, then we're done. Let's first show they are linearly independent. For the sake of contradiction, let's say they weren't so that $x_1, x_2, ..., x_m$ are linearly dependent over K. Then they are linearly dependent over $\mathcal{O}_K$ by Gauss's lemma. Hence, there are elements $a_1, a_2, ..., a_m \in \mathcal{O}_K$ that are not all zero such that

$$a_1 x_1 + \cdots + a_m x_m = 0$$

We want to use this to show that the congruent mod $\mathfrak{p}$ is also linearly dependent. To that end, consider $\mathfrak{a} = (a_1, a_2, ..., a_m)$. Take $a \in \mathfrak{a}^{-1}$ such that $a \notin \mathfrak{a}^{-1}\mathfrak{p}$ so that $a\mathfrak{a} \not\subseteq \mathfrak{p}$. Hence, the element $aa_1, ..., aa_m$ lie in $\mathcal{O}_K$, but are not all in $\mathfrak{p}$. Hence, we have:

$$aa_1 x_1 + \cdots + aa_m x_m = 0 \bmod \mathfrak{p}$$

giving linear dependence of the elements $\overline{x_1}, ..., \overline{x_m}$ over k – a contradiction.

Next, for spanning, consider the $\mathcal{O}_K$-module

$$M = \mathcal{O}_K x_1 + \cdots \mathcal{O}_K x_m$$

We shall show we can "upgrade" this to $L = Kx_1 + \cdots + Kx_m$. Consider $N = \mathcal{O}/M$. Since $\mathcal{O} = M + \mathfrak{p}\mathcal{O}$, $\mathfrak{p}N = N$. Since $\mathcal{O}$ is finitely generated (being a Dedekind domain), N is finitely generated. If $\alpha_1, \alpha_2, ..., \alpha_s$ are generators of N, then since $\mathfrak{p}N = N$:

$$\alpha_i = \sum_j a_{ij}\alpha_j \qquad a_{ij} \in \mathfrak{p}$$

Now, define the matrix $A = (a_{ij}) = I$ where I is the identity matrix of rank s. Let $B = A^*$ be the adjoint of A whose entries are the minors of rank $(s-1)$ of A. Then

$$A(\alpha_1, ..., \alpha_s)^t = 0 \qquad BA = \det(A)I = dI$$

Thus:

$$0 = BA(\alpha_1, \alpha_2, ..., \alpha_s)^t = d\alpha_1, ..., d\alpha_s)^t$$

thus, $dN = 0$, implying $d\mathcal{O} \subseteq M = \mathcal{O}_K x_1 + \mathcal{O}_K x_2 + \cdots + \mathcal{O}_K x_m$. Importantly, $d \neq 0$ since $d = \det((a_{ij}) = I) \equiv (-1)^s \bmod p$ since $a_{ij} \in \mathfrak{p}$. Thus:

$$L = dL = Kx_1 + Kx_2 + \cdots + Kx_m$$

showing that it spans L/K, and hence it is a basis, establishing the first equality.

For the second, recall in proposition 22.2.24 that each $\mathfrak{B}_i^{n+1}/\mathfrak{B}_i^n \cong_{\mathcal{O}/\mathfrak{B}_i} \mathcal{O}/\mathfrak{B}_i$. Hence, since $f_i = [\mathcal{O}/\mathcal{B}_i : \mathcal{O}_k/\mathfrak{p}]$, we get

$$n = \dim_k(\mathcal{O}/\mathfrak{p}\mathcal{O}) \sum_i \dim_k(\mathcal{O}/(\mathfrak{B}_i^{e_i})) = \sum_i^r e_i f_i$$

as we sought to show.

Corollary 22.4.4: Multiplicity of Ramification And Inertia

Let $M/L/K$ be a finite extension of number fields. Then:

$$e_{M/K} = e_{M/L}e_{L/K} \qquad f_{M/K} = f_{M/L}f_{L/K}$$

Proof **:**
immediate consequence of the arguments in the above theorem

Hence, we have a measure of how much a prime ideal splits in an extension.

Example 22.4.1: Decomposing Prime

1. It is possible that $\mathfrak{p}$ is totally ramified where $e_1 > 1$ and $f_1 = 1$: take $L = \mathbb{Q}(\sqrt{3})$ and $K = \mathbb{Q}$ with $I = (3)$. Then $(3)\mathcal{O} = (\sqrt{3})^2$

2. Similarly, it is possible that $\mathfrak{p}$ is totally ramified and $e_1 = 1$ and $f_1 > 1$: take $K = \mathbb{Q}$, $L = \mathbb{Q}(\sqrt{3})$, but $I = (5)$).

3. Let us consider the inertia degree to get some structure of the Dedekind domain. Let $L = \mathbb{Q}(\sqrt[3]{2})$ and $L/\mathbb{Q}$ be a finite extension. Since it is always the case that $\mathbb{Z}[\sqrt[3]{2}] \subseteq \mathcal{O}_L$[a], Take $I = (\sqrt[3]{2}) \subseteq \mathcal{O}_L$, $I \cap \mathbb{Q} = (2)$, and $\mathfrak{N}((\sqrt[3]{2})) = |\operatorname{Nm}_{L/\mathbb{Q}}(\sqrt[3]{2})| = 2$, and $2 = (\sqrt[3]{2})^3$. Hence, $f_1 = 1$, meaning
$$[\mathcal{O}_L/\sqrt[3]{2}\mathcal{O}_L : \mathbb{Z}/2\mathbb{Z}] = 1$$
implying both fields only have two elements, hence
$$\mathcal{O}_L/\sqrt[3]{2}\mathcal{O}_L \cong \mathbb{Z}/2\mathbb{Z}$$
Notice how this also gave information on the ring of integers,

4. Let $\mathcal{O}_K = \mathbb{Z}[i]$ so that $2 = (1+i)(1-i)$. Then $\mathcal{O}_K/2\mathcal{O}_K$ has order 4 (look at it'snorm), and certainly $2\mathcal{O}_K \subseteq (1-i)$, hence $|\mathcal{O}_K/(1-i)|\big|4$, which by direct computation we may see that $|\mathcal{O}_K/(1-i)| = 2$, hence $f = 1$. On the other hand, we can see that 3 is prime in $\mathcal{O}_K$, and $|\mathcal{O}_K/3\mathcal{O}_K| = 9$, giving $f_{3\mathcal{O}_K/3} = 2$.

[a] In fact, they are equal in this case, which is not needed at the moment

Since the factorization of $\mathfrak{p} \subseteq K$ in L/K is tied to the degree of L/K, we may venture into hoping that the structure of L/K may reflect something about the factorization of $\mathfrak{p}$. A natural question we may ask here is how do we find the prime factors in the extension? Finding them is split into two cases: the "good" primes and the "bad' primes. We shall see that the power of the exponentials has a lot to do with it. This is because we shall see that most primes shall decompose with ramification index 1, since it will correspond to the splitting of a polynomial in separable extension (note that all number field extensions are separable). However, as we saw it is possible that prime may have higher ramification index, however it is impossible that a polynomial splits with irreducible with multiplicity higher than 1 over a separable extension. Hence, we may think of such prime as "bad" primes. In the following we shall show how to find primes whose factorization mirrors that of a minimal polynomial, and how to detect when it is not the case.

let L/K be a finite separable extension. Then by the primitive extension theorem, there exists a $\theta \in K$

$$L = K(\theta)$$

Usually, we shall let L be Galois so that we may computationally find this element, namely it must not be fixed by any element of $\mathrm{Gal}_K(L)$ (hence, not in any intermediate extension). In fact, we may find θ so that $\theta \in \mathcal{O}_K$: if $L = K(\alpha)$, since $L = \overline{\mathcal{O}_K}^L$ we may multiply out the denominator by d to get $\alpha = \theta$. Now we have a single element, θ with minimal polynomial $m_{K,\theta}(x)$ which characterizes L. It would be nice to say something about the nature of the decomposition of $\mathfrak{p}$ in $\mathcal{O}_L$ given this information. In the good case, we shall be able to do so, and in the bad case we shall require a bit more work. In particular, a finite number of primes ideals will be exclude; only those that are relatively prime to an ideal known as the *conductor* of the ring $\mathcal{O}_K[\theta]$ will be considered:

Definition 22.4.5: Conductor

Let $\mathcal{O}_K[\theta]$ be the ring of integers extended by θ. Then the conductor of $\mathcal{O}_K[\theta]$, is the largest ideal $\mathscr{F}_\theta \subseteq \mathcal{O}_L$ which is contained in $\mathcal{O}_K[\theta]$:

$$\mathscr{F}_\theta = \{\alpha \in \mathcal{O}_L \mid \alpha\mathcal{O}_L \subseteq \mathcal{O}_K[\theta]\}$$

In the case where $K = \mathbb{Q}$ and $\mathcal{O}_\mathbb{Q} = \mathbb{Z}$, then $\mathscr{F}_\theta = \{\alpha \in \mathcal{O}_L \mid (\alpha)_L \subseteq \mathbb{Z}[\theta]\}$. Since $\mathcal{O}_L$ is a finitely generated $\mathcal{O}_K$ module, $\mathscr{F}_\theta \neq 0$. The conductor is a measure of the "bad primes" (hence the ramification). In our case, the conductor is never 0, but it may be that $\mathscr{F}_\theta = (1)$, in which case the bad primes will be "tame" (or the field is "tamely ramified")

Theorem 22.4.6: Dedekind-Kummer Theorem

Let L/K be a separable extension, $L = K(\theta)$ where θ is a primitive element where $m_{\theta,K}(x) \in \mathcal{O}_K[x]$. Let $\mathfrak{p}$ be a prime ideal of $\mathcal{O}_K$ which is relatively prime to the conductor $\mathscr{F}$ of $\mathcal{O}_K[\theta]$ and let $m(x) = m_{\theta,K}(x)$ and consider:

$$\overline{m}(x) = \overline{m}_1(x)^{e_1} \cdots \overline{m}_r(x)^{e_r} \bmod \mathfrak{p}$$

be the factorization of the polynomial $\overline{m}(x) = m(x) \bmod \mathfrak{p}$ into irreducible over the residue field $\mathcal{O}_K/\mathfrak{p}$. Then

$$\mathfrak{B}_i = \mathfrak{p}\mathcal{O} + m_i(\theta)\mathcal{O} = (\mathfrak{p}\mathcal{O}, m_i(\theta))_L$$

are different prime ideals of $\mathcal{O}$ over $\mathfrak{p}$. The inertia degree f_i of $\mathfrak{B}_i$ is the degree of $\overline{m}_i(x)$, and

$$\mathfrak{p} = \mathfrak{B}_1^{e_1} \cdots \mathfrak{B}_r^{e_r}$$

Proof **:**

Let $\mathcal{O}' = \mathcal{O}_K[\theta]$ and $\overline{\mathcal{O}_K} = \mathcal{O}_K/\mathfrak{p}$. The key isomorphism that we can establish since $\mathfrak{p}$ is relatively prime to the conductor $\mathscr{F}_\theta$ is:

$$\mathcal{O}_L/\mathfrak{p}\mathcal{O}_L \overset{!}{\cong} \mathcal{O}_K[\theta]/\mathfrak{p}\mathcal{O}_K[\theta] \cong \overline{\mathcal{O}_K}[x]/(\overline{m}(x))$$

where the $\overset{!}{\cong}$ isomorphism is the hard one to prove; the 2nd one comes from:

$$\mathcal{O}_K[\theta]/\mathfrak{p} \cong \mathcal{O}_K[x]/(\mathfrak{p}, m(x)) \cong \overline{\mathcal{O}}_K[x]/(\overline{m}(x))$$

For the $\overset{!}{\cong}$ isomorphism, from the relative primality of $\mathfrak{p}$ and $\mathscr{F}$ we get $\mathfrak{p}\mathcal{O}_L + \mathscr{F} = \mathcal{O}_L$. Since $\mathscr{F} \subseteq \mathcal{O}' = \mathcal{O}_k[\theta]$, we get that $\mathcal{O}_L = \mathfrak{p}\mathcal{O}_L + \mathcal{O}'$. From this, we get that the homomorphism

$$\mathcal{O}' \to \mathcal{O}_L/\mathfrak{p}\mathcal{O}_L$$

is surjective (every element of $\mathcal{O}_L$ is a linear combination of an element in $\mathcal{O}'$ and $\mathfrak{p}\mathcal{O}_L$) with kernel $\mathfrak{p}\mathcal{O}_L \cap \mathcal{O}'$. What's left to show is that this kernel equals $\mathfrak{p}\mathcal{O}'$. $\mathfrak{p}\mathcal{O}' \subseteq \mathfrak{p}\mathcal{O}_L \cap \mathcal{O}'$ is immediate, so for the other direction since $\mathfrak{p} + \mathscr{F} \cap \mathcal{O}_K = \gcd(\mathfrak{p}, \mathscr{F} \cap \mathcal{O}_K) = (1)$, $1 = pq + fs$ for some $p \in \mathfrak{p}$, $f \in \mathscr{F} \cap \mathcal{O}_K$. Then certainly $1 = pq + fs \in \mathfrak{p} + \mathscr{F}$, hence $\mathcal{O}' = \mathfrak{p} + \mathscr{F}$. Then:

$$\begin{aligned}
\mathfrak{p}\mathcal{O}_L \cap \mathcal{O}' &= (\mathfrak{p} + \mathscr{F})(\mathfrak{p}\mathcal{O}_L \cap \mathcal{O}') \\
&= \mathfrak{p}(\mathfrak{p}\mathcal{O}_L \cap \mathcal{O}') + \mathscr{F}(\mathfrak{p}\mathcal{O}_L \cap \mathcal{O}') \\
&\subseteq \mathfrak{p}\mathcal{O}'
\end{aligned}$$

where the last inclusion comes from the fact that any element of in the 2nd equality has the form $p(p\ell + f\ell')$ where $p\ell, f\ell' \in \mathcal{O}'$ by definition (namely the intersection), and so $\mathcal{O}_L/\mathfrak{p}\mathcal{O}_L \cong \mathcal{O}'/\mathfrak{p}\mathcal{O}'$. Then by our above argument we have $\mathcal{O}'/\mathfrak{p}\mathcal{O}' \cong \overline{\mathcal{O}}_K/(\overline{m}(x))$, hence we have:

$$\mathcal{O}_L/\mathfrak{p}\mathcal{O}_L \cong \overline{\mathcal{O}_K}[x]/(\overline{m}(x))$$

Now looking at the ring on the right, since in $\bmod \mathfrak{p}$ we have $\overline{m}(x) = \prod_i^r \overline{m}_i(x)^{e_i}$, by the CRT we get:

$$\overline{\mathcal{O}}_K[x]/(\overline{m}(x)) \cong \bigoplus_i^r \overline{\mathcal{O}}_K[x]/(\overline{m}_i(x))^{e_i}$$

But now we know the structure of each direct summand, $\overline{\mathcal{O}}_k[x]/(\overline{m}_i(x))^{e_i}$ is a principal ideal ring, and by the isomorphism we may deduce the prime ideal of $R = \overline{\mathcal{O}_K}[x]/(\overline{m}(x))$ are the principal ideal

$$(\overline{m_i}) = (\overline{m}_i(x) \bmod \overline{m}(x)) \qquad 1 \leq i \leq r$$

By ring extension theory, the degree of $[R/(\overline{m}_i) : \overline{\mathcal{O}}_K]$ equals the degree of the polynomial $\overline{m}_i(x)$, and by definition of the $(\overline{m}_i)$:

$$(0) = \bigcap_i^r (\overline{m}_i)^{e_i}$$

Now, through the isomorphism $\overline{\mathcal{O}}_K[x]/(\overline{m}(x)) \cong \mathcal{O}_L/\mathfrak{p}\mathcal{O}_L$ given by $\overline{f}(x) \mapsto \overline{f}(\theta)$, we get the same result for the primes ideals of $\overline{\mathcal{O}}_L = \mathcal{O}_L/\mathfrak{p}\mathcal{O}_L$. Namely, the prime ideals $\overline{\mathfrak{B}}_i$ of $\overline{\mathcal{O}}_L$ correspond to the prime ideals

$$\overline{\mathfrak{B}}_i = (\overline{m}_i) = (m_i(\theta) \bmod \mathfrak{p}\mathcal{O}_L)$$

with degrees $[R/\overline{\mathfrak{B}}_i : \mathcal{O}_K/\mathfrak{p}]$ where $R = \overline{\mathcal{O}_K}[x]/(\overline{m(x)})$ equalling the degrees of the polynomials $\overline{m}_i(x)$. Furthermore:

$$(0) = \bigcap_i^r \overline{\mathfrak{B}}_i^{e_i}$$

Now, given the canonical projection $\mathcal{O}_L \to \mathcal{O}_L/\mathfrak{p}\mathcal{O}_L$, let $\mathfrak{B}_i = \mathfrak{p}\mathcal{O}_L + m_i(\theta)\mathcal{O}_L$, (the preimage of $\overline{\mathfrak{B}}_i$). Then since the pre-image of a prime ideal is prime, each $\mathfrak{B}_i$ is prime, and

$$\mathfrak{B}_i \cap \mathcal{O}_K = \mathfrak{p}$$

since, when written in element form, we have that

$$p\ell + m_i(\theta)\ell' \in \mathcal{O}_K$$

which is true if and only if $\ell' = 0$ and $p\ell \in \mathcal{O}_K$, showing that $\mathfrak{B}_i$, $i \leq i \leq r$, varies over the prime ideals of $\mathcal{O}_L$ above $\mathfrak{p}$, with $f_i = [\mathcal{O}_L/\mathfrak{B}_i : \mathcal{O}_K/\mathfrak{p}]$ being the degree of the polynomial $\overline{m}_i(x)$. Furthermore, $\mathfrak{B}_i^{e_i}$ is the pre-image of $(\overline{\mathfrak{B}}_i^{e_i})$ (since $e_i = \#\left\{\overline{\mathfrak{B}}^n \;\middle|\; n \in \mathbb{N}\right\}$) and

$$(0) = \cap_{i=1}^r \mathfrak{B}_i^{e_i} = \prod \mathfrak{B}_i^{e_i} \subseteq \mathfrak{p}\mathcal{O}_L$$

so that

$$\mathfrak{p}\mathcal{O}_L \Big| \prod_{i=1}^r \mathfrak{B}_i^{e_i}$$

since $\deg(m(x)) = [L : K] = n = \sum e_i f_i$, we in fact has no other factors or powers, showing us that in fact:

$$\mathfrak{p}\mathcal{O}_L = \prod_{i=1}^r \mathfrak{B}_i^{e_i}$$

as we sought to show.

Corollary 22.4.7: Factorization For Number Rings (Dedekind-Kummer Theorem)

Let L/K be a finite extension of number fields, $L = K(\theta)$ for a primitive element θ where $m_{\theta,K}(x) \in \mathcal{O}_K[x]$. Let $\mathfrak{p} \subseteq \mathcal{O}_K$ a prime ideal lying over $p \in \mathbb{Z}$. If $p \nmid [\mathcal{O}_L : \mathcal{O}_K[\alpha]]$, then

$$\mathfrak{B}_i = \mathfrak{p}\mathcal{O}_L + m_i(\theta)\mathcal{O}_L = (\mathfrak{p}\mathcal{O}_L, m_i(\theta))_L$$

are different prime ideals of $\mathcal{O}$ over $\mathfrak{p}$. The inertia degree f_i of $\mathfrak{B}_i$ is the degree of $\overline{m}_i(x)$, and

$$\mathfrak{p} = \mathfrak{B}_1^{e_1} \cdots \mathfrak{B}_r^{e_r}$$

Proof **:**
In the proof of theorem 22.4.6 we used the conductor to show that:

$$\mathcal{O}_L/\mathfrak{p}\mathcal{O}_L \cong \mathcal{O}'/\mathfrak{p}\mathcal{O}' \cong \overline{\mathcal{O}_K}[x]/(\overline{m}(x))$$

In particular, we could have used the CRT earlier and shown that:

$$\frac{\mathcal{O}_L}{\mathfrak{B}} \cong \frac{\overline{\mathcal{O}_K}[x]}{(\overline{m_i(x)})}$$

Hence, if we can show this condition, we are done. Take the map $\varphi : \mathcal{O}_K[x] \to \mathcal{O}_L/\mathfrak{B}_i$ mapping $p(x) \mapsto p(\theta)$. We claim that this map is surjective. To see this, it suffices to show that

$$\mathcal{O}_L = \mathcal{O}_K[\theta] + \mathfrak{B}_i$$

We know that $p \in \mathfrak{p} \subseteq \mathfrak{B}_i$, hence $p\mathcal{O}_L \subseteq \mathfrak{B}_i$. We now claim the slightly stronger:

$$\mathcal{O}_L = \mathcal{O}_K[\theta] + p\mathcal{O}_L$$

Now, by assumption $p \nmid |\mathcal{O}_L : \mathcal{O}_K[\theta]|$. Since the index of $\mathcal{O}_K[\theta] + p\mathcal{O}_L$ in $\mathcal{O}_L$ is the common divisor of $[\mathcal{O}_L : \mathcal{O}_K[\theta]]$ and $[\mathcal{O}_L : p\mathcal{O}_L]$, and these *must* be relatively prime since

$$[\mathcal{O}_L : p\mathcal{O}_L] = p^k$$

for some $k \in \mathbb{N}$. Thus $\mathcal{O}_K[x] \to \mathcal{O}_L/\mathfrak{B}_i$ is surjective. Now Certainly $(\mathfrak{p}, m_i) \subseteq \ker \varphi$. Furthermore, it is clear that $(\mathfrak{p}, m_i) \subseteq \mathcal{O}_K[x]$ is a maximal ideal hence $\ker \varphi$ is either $(\mathfrak{p}, m_i)$ or $\mathcal{O}_K[x]$. since the map is surjective, then either $\mathfrak{B}_i = \mathcal{O}_L = (1)_L$ (i.e. $\mathfrak{B}_i$ is *not* a prime, a contradiction), or it is indeed the desired kernel, meaning:

$$\frac{\mathcal{O}_L}{\mathfrak{B}} \cong \frac{\overline{\mathcal{O}_K}[x]}{(\overline{m_i(x)})}$$

but now the proof follows from the rest of the proof of theorem 22.4.6, as we sought to show.

Corollary 22.4.8: Monogenic Number Rings

Let L/K be a finite extension of number rings such that $\mathcal{O}_L = \mathcal{O}_K[\alpha]$ for some $\alpha \in \mathcal{O}_L$. Then all primes split as given in the above

***Proof* :**
Choose $\theta = \alpha$. Then $[\mathcal{O}_L : \mathcal{O}_K[\alpha]] = [\mathcal{O}_K[\alpha] : \mathcal{O}_K[\alpha]] = 1$. Since certainly no prime divides 1, we have that all primes factor in given their minimal polynomials and modular factorization, as we sought to show.

Note that not all number rings are of the form $\mathcal{O}_L = \mathcal{O}_K[\alpha]$, that is not all number rings are monogenic. We shall show a counter-example in example 22.4.4. this contrasts to finite extension of field where every separable finite extension of is a simple extension by theorem 17.6.17. Note how this is dependent on finding a θ and $\mathfrak{p}$ being relatively prime to $\mathscr{F}_\theta$ (or $[\mathcal{O}_L : \mathcal{O}_K[\theta]]$ in the case of number fields).

As it turns out, the power of an extended prime has important geometric and algebraic consequences; it can be thought of as the measure of how much the Galois group moves a field around. We hence give it an name:

Definition 22.4.9: Ramified and Unramified Extensions

Let $\mathfrak{p} \subseteq R$ be a prime ideal, $K = K(R)$, L/K a finite extension, and $\mathcal{O} = \overline{K}^L$. Then given

$$\mathfrak{p}\mathcal{O} = \mathfrak{B}_1^{e_1} \cdots \mathcal{B}_r^{e_r}$$

Then the prime ideal $\mathfrak{B}_i$ is called *unramified* over $\mathcal{O}_k$ (or over K) if $e_i = 1$ and if the residue fields $(\mathcal{O}/\mathfrak{B}_i)/(\mathcal{O}_K/\mathfrak{p})$ is separable[a]:

$$\mathfrak{p}\mathcal{O} = \mathfrak{B}_1 \cdots \mathfrak{B}_r$$

Otherwise, $\mathfrak{B}_i$ is called *ramified*, and *totally ramified* if furthermore $f_i = 1$. The prime ideal $\mathfrak{p}$ is called *unramified* if all $\mathfrak{B}_i$ are unramified, and $\mathfrak{p}$ is ramified otherwise. The extension L/K is called *unramified* or an *unramified extension* if all prime ideals $\mathfrak{p}$ of K are unramified.

[a] which for number fields this is always the case

It is in fact rare that a prime ideal $\mathfrak{p}$ of K is ramified in L, however every [nontrivial] extension of $\mathbb{Q}$ is ramified (i.e. some primes ramify). Later, we shall see that there do exist [nontrivial] unramified extension L/K where $K \neq \mathbb{Q}$. To understand ramification in extension, we first show how we can factor most $\mathfrak{B}_i$. The term "ramification" comes from complex analysis, where if a holomorphic function f has a root, $f(a) = 0$, the order of the root is the ramification index[21].

Theorem 22.4.10: Counting Ramified Primes

Let L/K be a separable extension where $K = \mathrm{Frac}(R)$ or a Dedekind domain R. Then there are only finitely many prime ideals of K which are ramified in L, in particular the ideals $\mathfrak{p} \subseteq K$ that are *not* relatively prime to the discriminant of $m_{K,\theta}(x)$ and $\mathscr{F}_\theta$.

Note that it is still possible that primes factor as given in theorem 22.4.6 since a prime may be relatively prime to $\mathscr{F}_\theta$ but not the discriminant of $m_{K,\theta}(x)$.

Proof **:**
Let $\theta \in \mathcal{O}$ be a primitive element for L, $L = K(\theta)$, and let $m(x) = m_{K,\theta}(x) \in R[x]$ be the minimal polynomial. Take the discriminant of $m(x)$

$$\mathrm{disc}(m_{\theta,K}(x)) = \mathrm{disc}_{L/K}(1, \theta, ..., \theta^{n-1}) = \prod_{i<j} (\theta_i - \theta_j)^2 \in R$$

We shall show every prime (fractional) ideal $\mathfrak{p} \subseteq K$ which is relatively prime to d and to the conductor $\mathscr{F}_\theta$ of $R[\theta]$ is unramified. Note that there are finitely many primes $\mathfrak{p} \subseteq K$ that are not relatively prime to the conductor and the discriminant since both have finitely many factors, hence this proofs the result.

By theorem 22.4.6, $\mathfrak{p}\mathcal{O}$ factors in the same way $\overline{m}_{\theta,K}(x)$ factors (mod $\mathfrak{p}$). Then the ramification indices e_i equal 1 when $\overline{m}(x)$ has no multiple roots. But this is the case since $\overline{0} \neq \overline{d} = d$ mod $\mathfrak{p}$ is nonzero by assumption of being relatively prime, and the discriminant is nonzero if and only if all the roots are distinct. Thus, the residue class field extension $(\mathcal{O}/\mathfrak{B}_i)/(R/\mathfrak{p})$ are generated by

[21] if $f(a) = 0$ has order n, there exists a g such that $f = g^n$ in some neighbourhood, which tells us how much f "winds" around the root

$\overline{\theta} \equiv \theta \bmod \mathfrak{B}_i$, and thus are separable, showing $\mathfrak{p}$ is unramified, as we sought to show.

We can in fact be more precise than the above result: the ramified primes ideals will be given by the *relative discriminant* of $\mathcal{O}_L/\mathcal{O}_K$, which will be given by the ideal $\mathfrak{o} \subseteq \mathcal{O}_K$ which is generated by the discriminants $d(\omega_1, \omega_2, ..., \omega_n)$ for all bases $\omega_1, \omega_2, ..., \omega_n$ of L/K contained in $\mathcal{O}_L$. We shall cover these in more detail in section **??**. A special case is that of the base-field being $K = \mathbb{Q}$. Though this is a simple corollary of the results in section **??**, the proof techniques are enlightening enough to deserve their own proposition:

Proposition 22.4.11: Detecting Ramification In Number Fields

Let p be a prime in $\mathbb{Z}$. Suppose p is ramified. Then:

$$p \Big| \operatorname{disc}(\mathcal{O}_K)$$

In proposition 22.4.14, we will show the converse, showing this is an iff.

***Proof* :**
Let $\mathfrak{B}$ lie over $\mathfrak{p}$ and $e_{\mathfrak{B}/\mathfrak{p}} > 1$. Thus

$$p\mathcal{O}_K = \mathfrak{B}I$$

where I is divisible by *all* primes of $\mathcal{O}_K$ lying over p. Let $\sigma_1, \sigma_2, ..., \sigma_n$ denote the embeddings of K in $\mathbb{C}$, and extend each σ_i to automorphisms of some extension of L/K which is normal over $\mathbb{Q}$. Let $\alpha_1, \alpha_2, ..., \alpha_n$ be some integral basis for R. What we shall do is replace one of the α_i be a suitable element that will show that $p \Big| \operatorname{disc}(\mathcal{O}_K)$.

Take $\alpha \in I - p\mathcal{O}_K$. Then certainly α is in every prime ideal lying over p but *not* in $p\mathcal{O}_K$. writing

$$\alpha = m_1\alpha_1 + \cdots m_n\alpha_n$$

for appropriate $m_i \in \mathbb{Z}$, then since $\alpha \notin p\mathcal{O}_K$, we have that p does not divide each m_i. Re-arrange the terms so that $p \nmid m_1$. Then if d is the discriminant of $\mathcal{O}_K$, through direct computation, we can see that

$$\operatorname{disc}(\alpha, \alpha_2, ..., \alpha_n) = m_1^2 d$$

Since $p \nmid m_1$, we shall show that $p \Big| \operatorname{disc}(\alpha, \alpha_2, ..., \alpha_n)$

Given $L/K/\mathbb{Q}$, since α is in every prime of $\mathcal{O}_K$ lying over p, it certainly is in every prime in $\mathcal{O}_L$ lying over p. Fixing any $\mathfrak{B}$ of S lying over p, we see that for any automorphism σ_i, $\sigma(\alpha) \in \mathfrak{B}$. To see this ,$\sigma^{-1}(\alpha)$ is a prime of $\sigma^{-1}(S) = S$ lying over p, hence must contain α. In particular, this gives $\sigma_i(\alpha) \in \mathfrak{B}$ for all i. But then $\mathfrak{B}$ contains $\operatorname{disc}(\alpha, \alpha_2, ..., \alpha_n)$. Since the discriminant is in $\mathbb{Z}$, we get that:

$$\mathfrak{B} \cap \mathbb{Z} = p\mathbb{Z}$$

but then $p \Big| \operatorname{disc}(\mathcal{O}_K)$, as we sought to show.

Corollary 22.4.12: Ramification Over Number Fields

Let $\alpha \in R$ so that $K = \mathbb{Q}(\alpha)$ and let f be a monic polynomial over $\mathbb{Z}$ such that $f(\alpha) = 0$. Then if p is a prime such that

$$p \not| \operatorname{Nm} f'(\alpha)$$

Then p is unramified in K

Proof **:**
This follows from the above an the fact that if α is an algebraic integer and $f \in \mathbb{Z}[x]$ monic such that $f(\alpha) = 0$, then by proposition 22.1.24:

$$\operatorname{disc}(\alpha) \Big| \operatorname{Nm}_{\mathbb{Q}(\alpha)/\mathbb{Q}}(f'(\alpha))$$

Using this result, we have an easier verification for ramification of primes of $\mathbb{Z}$ to number fields, since they must divide the discriminant, but the discriminant may only have finitely many factors. In fact, it is easy to generalize this result for extension of number fields

Corollary 22.4.13: Ramification In Number Fields

Let L/K be an extension of number fields Then only finitely many primes of $\mathcal{O}_K$ are ramified in $\mathcal{O}_L$.

Proof **:**
Let $\mathfrak{p}$ be a prime of $\mathcal{O}_K$ which is ramified in $\mathcal{O}_L$. Then $\mathfrak{p} \cap \mathbb{Z} = p\mathbb{Z}$ is ramified in S. Hence there are only finitely many possible primes p that ramify, meaning there are only finitely many possible $\mathfrak{p}$.

go over these notes: here.

Proposition 22.4.14: Characterizing Ramification In Number Fields

Let p be a prime in $\mathbb{Z}$. Suppose $p \Big| \operatorname{disc}(\mathcal{O}_K)$. Then p is ramified.

Proof **:**
p.79 markus

22.4.1 Application: Quadratic Extensions and Quadratic Reciprocity

We now take a moment to solve a classical number theory problems with the developed tools. Let $p \in \mathbb{Z}$ be a prime number. In section 10.3, it was shown that p can be written as the sum of two squares if and only if $p \equiv 1 \bmod 4$. Let us now ask another question that may shed further light on the nature of primes: what primes p have the property that there exists an n such that $n^2 \equiv p \bmod q$ for another prime q? In other words, which primes have squares roots mod q. This is equivalent to asking what primes solve the equation $n^2 - p \equiv 0 \bmod q$, which is asking about the factorization of

p in $\mathbb{Q}(\sqrt{q})$. We can thus ask this question in a slightly more general framework where $K = \mathbb{Q}(\sqrt{m})$ where m is square-free. To start off, we have found the integral basis of the ring of integers of quadratic extension in 22.1.2[4], namely we have integral basis

$$\begin{aligned} \{1, \sqrt{m}\} & \qquad m \not\equiv 1 \bmod 4 \\ \left\{1, \frac{1+\sqrt{m}}{2}\right\} & \qquad m \equiv 1 \bmod 4 \end{aligned}$$

and $\operatorname{disc}(\mathcal{O}_K) = 4m$ in the first case and $\operatorname{disc}(\mathcal{O}_K) = m$ in the second. Now, take $(p) \subseteq \mathbb{Z}$ a prime ideal. Then we know that either p can split as:

$$p\mathcal{O}_K = \mathfrak{B} \qquad p\mathcal{O}_K = \mathfrak{B}^2 \qquad p\mathcal{O}_K = \mathfrak{B}_1\mathfrak{B}_2$$

Theorem 22.4.15: Primes In Quadratic Extensions

Let $p \subseteq \mathbb{Z}$ be a prime ideal and $K = \mathbb{Q}(\sqrt{m})$ a quadratic extension where m is square-free. Then:

1. If $p \big| m$:
$$p\mathcal{O}_K = (p, \sqrt{m})^2$$
2. if m is odd, then 2 splits as:
$$2\mathcal{O}_K = \begin{cases} (2, 1+\sqrt{m}^2) & m \equiv 3 \bmod 4 \\ \left(2, \frac{1+\sqrt{m}}{2}\right)\left(2, \frac{1-\sqrt{m}}{2}\right) & m \equiv 1 \bmod 8 \\ (2) & m \equiv 5 \bmod 8 \end{cases}$$
and if p is odd and $p \nmid m$:
$$p\mathcal{O}_K = \begin{cases} (p, n+\sqrt{m})(p, n-\sqrt{m}) & m = n^2 \bmod p \\ (p) & m \neq n^2 \bmod p \end{cases}$$

***Proof* :**

Most of these equations follow a similar pattern: If $p \big| m$, then $(p, \sqrt{m})^2 = (p^2, p\sqrt{m}, m)$. Since $p \big| m$, this is contained in $p\mathcal{O}_K$. Conversely, it contains $\gcd(p^2, m) = p$, hence it contains $\mathcal{O}_K$.

This type of reasoning works for all except when $p\mathcal{O}_K$ remains prime, in this case we shall show that if $\mathfrak{B}$ is any prime lying over p, then $\mathcal{O}_K/p\mathcal{O}_K$ is not isomorphic to $\mathbb{Z}/p\mathbb{Z}$ (hence will have $f_{\mathfrak{B}/p} = 2$). First, assuming p is odd, $p \nmid m$, and that m is not a square $\bmod p$, the consider $x^2 - m$. This has a root in $\mathcal{O}_K$, and hence in $\mathcal{O}_K/p\mathcal{O}_K$. But, by assumption it has no root in $\mathbb{Z}/p\mathbb{Z}$, hence $\mathcal{O}_K/p\mathcal{O}_K$ cannot be isomorphic to $\mathbb{Z}/p\mathbb{Z}$, giving us our desired result.

Finally, if $p = 2$, the case is similar by considering:

$$x^2 - x + \frac{1-m}{4}$$

The fact that a prime splits if it's a square mod p is important enough to be given a name:

Definition 22.4.16: Quadratic Residue

Let $n \in \mathbb{Z}$. Then if there exists an $m \in \mathbb{Z}$ such that $m^2 \equiv n \bmod p$, then n is called a *quadratic residue* mod p.

The next application is sometimes considered the first result of modern algebraic number theory, which was first proven by Gauss, but discovered by Euler. The question is to fully characterize which elements have quadratic residues mod p. The following notational shortcut shall be used in the proof of the result:

Definition 22.4.17: Legendre Symbol

Let p be an odd prime in $\mathbb{Z}$. Then for $n \in \mathbb{Z}$, $p \nmid n$, the *Legendre symbol* is defined to be:

$$\left(\frac{n}{p}\right) = \begin{cases} 1 & \exists m,\ m^2 \equiv n \bmod p \\ -1 & \text{otherwise} \end{cases}$$

Theorem 22.4.18: Quadratic Reciprocity

Let p, q be two distinct odd prime numbers. Then:

$$\left(\frac{q}{p}\right)\left(\frac{p}{q}\right) = (-1)^{\frac{q-1}{2}\frac{p-1}{2}}$$

In other words, if:

$$p^* = \begin{cases} 1 & p \equiv 1 \bmod 4 \\ -p & p \equiv -1 \bmod 4 \end{cases}$$

Then q is a quadratic residue modulo p if and only if p' is a quadratic residue modulo p

Proof **:**
Let $q \neq p$ be odd primes. We want to show that $n^2 \equiv q \bmod p$ if and only if $m^2 \equiv p' \bmod q$ for some numbers $n, m \in \mathbb{Z}$. We shall first prove a lemma: $n^2 \equiv q \bmod p$ if and only if q splits in $\mathbb{Q}(\sqrt{p'}) = L$. First,

$$p \nmid [\mathcal{O}_L : \mathbb{Z}[\sqrt{p}]]$$

since $[\mathcal{O}_L : \mathbb{Z}[\sqrt{p}]] \in \{1, 2\}$. Thus, if $n^2 \equiv q \bmod p$,

$$x^2 - q \equiv (x - n)(x + n) \bmod p$$

Thus, by theorem 27 in Marcus[a], we have:

$$q = \mathfrak{B}_1 \mathfrak{B}_2$$

Conversely, if q splits completely, consider $x^2 - q \in \mathbb{Z}[x]$. Certainly this is irreducible over $\mathbb{Z}$, and $p \nmid [\mathcal{O}_L : \mathbb{Z}[\sqrt{q}]]$, hence since we know q splits it must be that:

$$x^2 - q \equiv (x - n)(x + n) \bmod n$$

Namely, $q \equiv n^2 \bmod p$. Thus q splits completely iff $n^2 \equiv q \bmod p$.

Thus it suffices to show hat q splits if and only if $m^2 \equiv p' \bmod q$. Let's first say that q splits, so we have $\mathfrak{B}$ lying above q, and $(q)_L \subsetneq \mathfrak{B}$. Since $p' \equiv 1 \bmod 4$, we know the number ring of L is:

$$\mathbb{Z}\left[\frac{1+\sqrt{p'}}{2}\right]$$

Hence, we have that there is an element

$$a + b\sqrt{p'} \in \mathfrak{B} - (q) \qquad a, b \in \frac{1}{2}\mathbb{Z} \tag{22.4}$$

and since q is odd, we may without loss of generality choose a, b to be in $\mathbb{Z}$ by multiplying by 2. Now, Consider the norm of the element:

$$(a + b\sqrt{p'})(a - b\sqrt{p'}) = a^2 - b^2p' \in \mathfrak{B} \cap \mathbb{Z} = (q)$$

Hence, $q \big| a^2 - b^2p'$. Notice that q cannot divide a or b by our choice in equation (22.4). Hence, get that:

$$a^2 - b^2p' \equiv 0 \bmod q$$

re-arranging (and dividing since we in the residue field), we get:

$$(ab^{-1})^2 \equiv p' \bmod q$$

showing one direction of the implication.

Now, for the other direction, it suffices to show by our above lemma that q splits in L. By assumption:

$$m^2 \equiv p' \bmod q \qquad \Leftrightarrow \qquad m^2 - p' = (m - \sqrt{p'})(m + \sqrt{p'}) = cq$$

for some constant $c \in \mathbb{Z}$. I claim that $(m + \sqrt{p'}, q)$ divides q in L. Certainly

$$(q) \subsetneq (m + \sqrt{p'}, q)$$

since $m + \sqrt{p'} \notin (q)$ (since $m + \sqrt{p'}$ is not divisible by q). Furthermore, $(m + \sqrt{p'})$ is not a unit, for if it were, then if $\alpha = (m - \sqrt{p'})$ and $u = (m + \sqrt{p'})$:

$$u\alpha = (m + \sqrt{p'}, q)(m - \sqrt{p'}) = (x^2 - p', q) = (cq, q)$$

hence $q \big| u\alpha$, implying $q \big| \alpha$, but $q \not\big| m - \sqrt{p'}$. Thus, it must be that q splits.

But then, given our lemma this implies that q is a quadratic residue modulo p if and only if p' is a quadratic residue modulo q, as we sought to show.

[a] This is a slight generalization of the same result shown in class on when can we use the factorization of polynomial to factor primes

22.4.2 Example: Eisenstein Polynomials

We say $f \in \mathbb{Z}[x]$ is *Eisenstein* with respect to p is p divides each coefficient but p^2 does not divide the constant coefficient. Then by Eisenstein's criterion (proposition 11.3.7), such a polynomial is irreducible in $\mathbb{Q}[x]$. We shall use these to quickly determine a few useful ring of integer.

Lemma 22.4.19: Eisenstein Polynomial Modular Roots

Let $K/\mathbb{Q}$ be a number field of degree n and assume $K = \mathbb{Q}(\alpha)$ for $\alpha \in \mathcal{O}_K$ with minimal polynomial $m_{\alpha,\mathbb{Q}}$ Eisenstein at p. Then then for some $f(x) \in \mathbb{Z}[x]$ of degree at most $n-1$ with coefficients $a_0, a_1, ..., a_{n-1}$, if:

$$f(\alpha) \equiv 0 \bmod p\mathcal{O}_K$$

Then:

$$a_i \equiv 0 \bmod p\mathbb{Z} \qquad 1 \leq i \leq n-1$$

Proof **:**
we shall do an argument by induciton. Certainly if $a_0 \equiv 0 \bmod p\mathcal{O}_K$ then $a_0 \equiv 0 \bmod p\mathbb{Z}$ (since $a_0 \in \mathbb{Z}$). Thu,s asume for $i < j$, $j \in \{0, 1, ..., n-1\}$ that $a_i \equiv 0 \bmod p\mathbb{Z}$. Then for $i < j$:

$$a_j\alpha^j + a_{j+1}\alpha^{j+1} + \cdots + a_{n-1}\alpha^{n-1} \equiv \bmod p\mathcal{O}_K$$

Multiply the above by α^{n-1-j} to make all but the first torm $a_j\alpha^{n-1}$ a multiple of α^n. Since α is a root of an Eisenstein polynomial at p, we have

$$\alpha^n \equiv 0 \bmod p\mathcal{O}_K$$

Hnece:

$$a_j\alpha^{n-1} \equiv p\mathcal{O}_K$$

Then $a_j\alpha^{n-1} = p\gamma$, for $\gamma \in \mathcal{O}_K$. Taking the norm, we get:

$$a_j^n \operatorname{Nm}_{K/\mathbb{Q}}(\alpha)^{n-1} = p^n \operatorname{Nm}_{K/\mathbb{Q}}(\gamma) \in \mathbb{Z}$$

the left hand side is divisible by the constant term of the minimal polynomial (since that is the value of $\operatorname{Nm}_{K/\mathbb{Q}}(\alpha)$). By Eisenstein, p divides $\operatorname{Nm}_{K/\mathbb{Q}}(\alpha)$ exactly once, so since p^n must divide the left hand side gives that $p \big| a_j^n$, thus $p \big| a_j$, thus $a_i \equiv 0 \bmod p\mathbb{Z}$ for $i < j+1$, giving us the result.

Lemma 22.4.20: Eisenstein Polynomial divisibility by p

Let $K/\mathbb{Q}$ be a number field of degree n, $K = \mathbb{Q}(\alpha)$ where $\alpha \in \mathcal{O}_K$ with minimal polynomial Eisenstein at p. Then if

$$r_0 + r_1\alpha + \cdots + r_{n-1}\alpha^{n-1} \in \mathcal{O}_K \qquad r_i \in \mathbb{Q}$$

then each r_i has no p in its denominator

Proof **:**
For the sake of contradiction, assume some r_i has a p in its denominator. Let d be the lcm of the denominators so the r_i's so $p \big| d$ so that we may write $r_i = a_i/d$ for appropriate $a_i \in \mathbb{Z}$ where $a_i \nmid p$. Then:

$$r_0 + r_1\alpha + \cdots + r_{n-1}\alpha^{n-1} \in \mathcal{O}_K \implies \frac{a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}}{d} \in \mathcal{O}_K$$

multiplying by d,w e get:

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} \in d\mathcal{O}_K \subseteq p\mathcal{O}_K$$

Then by lemma 22.4.19 we have that $a_i \in p\mathbb{Z}$ for each i, a contradiction!

Theorem 22.4.21: Eisenstein Polynomial Divisibility Test

Let $K = \mathbb{Q}(\alpha)$ where $\alpha \in \mathcal{O}_K$ a root of an Eisenstain polnomial at p of degree n. Then:

$$p \nmid [\mathcal{O}_K : \mathbb{Z}[\alpha]]$$

Proof **:**
For the sake of contradiction, suppose $p \big| [\mathcal{O}_K : \mathbb{Z}[\alpha]]$. Then $\mathcal{O}_K/\mathbb{Z}[\alpha]$ has an element of order p, that is there is some $\gamma \in \mathcal{O}_K$ such that $\gamma \notin \mathbb{Z}[\alpha]$ but $p\gamma \in \mathbb{Z}[\alpha]$. Now, given basis $\{1, \alpha, ..., \alpha^{n-1}\}$ for $K/\mathbb{Q}$, then

$$\gamma = r_0 + r_1\alpha + \cdots + r_{n-1}\alpha^{n-1}$$

where $r_i \in \mathbb{Q}$. Since $\gamma \notin \mathbb{Z}[\alpha]$¡ not all $r_i \in \mathbb{Z}$¿ Since $p\gamma \in \mathbb{Z}[\alpha]$, $pr_i \in \mathbb{Z}$. Thus, r_i has p in its denominator, but that contradicts lemma 22.4.20, as we sought to show.

Example 22.4.2: Using Eisenstein Test

1. Let $K = \mathbb{Q}(\sqrt[3]{2})$. We shall quickly find $\mathcal{O}_K$. Recall that:

   $$-2^2 3^3 = -108 = \operatorname{disc}(\mathbb{Z}[\sqrt[3]{2}]) = [\mathcal{O}_K : \mathbb{Z}[\sqrt[3]{2}]]^2 \operatorname{disc}(\mathcal{O}_K)$$

   Since $\sqrt[3]{2}$ is a root of $x^3 - 2$, which is eisenstien at 2, we see that $2 \nmid [\mathcal{O}_K : \mathbb{Z}[\sqrt[3]{2}]]$. Furthermore, $1 + \sqrt[3]{2}$ is the root of

   $$(x-1)^3 - 2 = x^3 - 3x^2 + 3x - 3$$

   which is eisenstein at 3 s o

   $$3 \nmid [\mathcal{O}_K : \mathbb{Z}[1 + \sqrt[3]{2}]] = [\mathcal{O}_K : \mathbb{Z}[\sqrt[3]{2}]$$

   since $\mathbb{Z}[\sqrt[3]{2}] = \mathbb{Z}[1 + \sqrt[3]{2}]$. THus, the index must be 1, giving us:

   $$\mathcal{O}_K = \mathbb{Z}[\sqrt[3]{2}]$$

2. Let $K = \mathbb{Q}(\sqrt[4]{2})$. Then:

   $$-2^{11} = \operatorname{disc}(\mathbb{Z}[\sqrt[4]{2}]) = [\mathcal{O}_K : \mathbb{Z}[\sqrt[4]{2}]]^2 \operatorname{disc}(\mathcal{O}_K)$$

and $\sqrt[4]{2}$ is a root of $x^4 - 2$ which is eisenstein at 2, hence by the eisenstein divisibility test the idnex is 1 and

$$\mathcal{O}_K = \mathbb{Z}[\sqrt[4]{2}]$$

Let $K = \mathbb{Q}(\sqrt[5]{2})$. Then

$$2^4 5^4 = \operatorname{disc}(\mathbb{Z}[\sqrt[5]{2}]) = [\mathcal{O}_K : \mathbb{Z}[\sqrt[5]{2}]]^2 \operatorname{disc}(\mathcal{O}_K)$$

Certainly 2 doesn't divide the index. Next, the minimal polynomial of $\sqrt[5]{2} - 2$ is

$$(x+2)^5 - 2 = x^5 + 10x^4 + 40x^3 + 80x^2 + 80x + 30$$

which is Eisenstein at 5 and $\mathbb{Z}[\sqrt[5]{2} - 2] = \mathbb{Z}[\sqrt[5]{2}]$ Hence:

$$\mathcal{O}_K = \mathbb{Z}[\sqrt[5]{2}]$$

3. show that $\mathcal{O}_{\mathbb{Q}(\sqrt[3]{3})} = \mathbb{Z}[\sqrt[3]{3}]$ and $\mathcal{O}_{\mathbb{Q}(\sqrt[3]{5})} = \mathbb{Z}[\sqrt[3]{5}]$.

4. Recall the three cubic extensison K_1, K_2, K_3 in example 22.1.10 with the same discriminant, but it was claimed that they where not isomorphic as fields. The three polynomials in question where:

$$f_1(x) = x^3 - 18x - 6 \qquad f_2(x) = x^3 - 36x - 78 \qquad f_3(x) = x^3 - 54x - 150$$

each having discriminant $22345 = 5 \cdot 41 \cdot 109$. It can be verified that these three polynomials have real roots (by either graphing or doing some guess-work with derivative test and evaluating), and from the above methods we can check that

$$\mathcal{O}_{K_i} = \mathbb{Z}[\alpha_i]$$

for appropriate root α_i. Since the ring of integers is Monogenic, by Dedekind-Kummers theorem the factorization of primes p mirrors the factorization of $f_i(x)$ mod p. Then it can be verified that

(a) f_1, f_2 are irreducible mod 5, but

$$f_3(x) \equiv x(x-2)(x-3) \bmod 5$$

(b) f_2, f_3 are irreducible mod 11, but

$$f_1(x) \equiv (x-3)(x-9)(x-10) \bmod 11$$

Thus, we see that these fields cannot be isomorphic, or else the prime factorization would have been equivalent (recall that ring homomorphisms map primes to primes)

It may be asked if $K = \mathbb{Q}(\sqrt[n]{2})$

$$\mathcal{O}_K = \mathbb{Z}[\sqrt[n]{2}]$$

This is true for $n \leq 1000$, but is *not always true*. For those interested, they may search for *Wieferich primes* (see Keith Conrad *The ring of integers in radical extensions*).

Eisenstein polynomials also allow us to identify when a polynomial totally ramifies:

Theorem 22.4.22: Eisenstein Polynomial And Total Ramification

Let $K = \mathbb{Q}(\alpha)$. Then α is a root of a $\mathbb{Z}$-polynomial that is Eisenstein at p iff p is totally ramified in K.

Note how there is no reference to the ring of integers

Proof **:**
(*must* add this here! this link

Example 22.4.3: Dececting Ramification Using Eisenstein

Since $\alpha = \sqrt[3]{10}$ is a root of $x^3 - 10$, which is eisenstein at 2 and 5, then in $K = \mathbb{Q}(\alpha)$, $(2) = \mathfrak{p}^3$ and $(5) = \mathfrak{q}^3$. Note however htat

$$\mathcal{O}_K \neq \mathbb{Z}[\sqrt[3]{10}]$$

It is in fact equal to

$$\mathcal{O}_K = \mathbb{Z}[\beta] \qquad \beta = \frac{1}{3}\left(1 + \sqrt[3]{1-} + \sqrt[3]{100}\right)$$

which is a root of $x^3 - x^2 - 3x - 3$.

22.4.3 Primes in Galois Extensions

The question of prime decomposition in extension becomes more interesting when L/K is also a Galois extension since $\mathrm{Gal}_K(L)$ can act on the set of primes.

Lemma 22.4.23: Galois Group Acts On Primes

Let L/K be a finite galois extension with galois group $G = \mathrm{Gal}_K(L)$, $\mathfrak{p} \subseteq \mathcal{O}_K$ a prime ideal of $\mathcal{O}_K$, and $\mathfrak{P} = \{\mathfrak{B}_1, \mathfrak{B}_2, ..., \mathfrak{B}_n\}$ the set of primes lying above $\mathfrak{p}$. Then:

$$G \curvearrowright \mathfrak{P} \qquad \text{via} \qquad \sigma \cdot \mathfrak{B}_i = \sigma(\mathfrak{B}_i)$$

Proof **:**
let $G = \mathrm{Gal}_K(L)$. Then for each $a \in \mathcal{O}_L$, $\tau(a) \in \mathcal{O}_L$ for all $\tau \in G$ since K is invariant under τ and all conjugates have an integral minimal polynomial (i.e. the same minimal polynomial with coefficients in $\mathbb{Z} \subseteq K$), so there is a natural action of G on $\mathcal{O}_L$; $G \curvearrowright \mathcal{O}_L$. Next, given that τ is a field isomorphism, $\tau : \mathcal{O}_L \to \mathcal{O}_L$ is a ring isomorphism, importantly it sends primes to primes. If $\mathfrak{B}$ is a prime ideal of $\mathcal{O}_L$ above $\mathfrak{p}$, then $\tau(\mathfrak{B})$ is also a prime ideal above $\mathfrak{p}$ for each prime $\mathfrak{B}$ since

$$(\tau(\mathfrak{B})) \cap \mathcal{O}_K = \tau(\mathfrak{B} \cap \mathcal{O}_K) = \tau(\mathfrak{p}) = \mathfrak{p}$$

since τ is K-invariant, hence $\mathcal{O}_K \subseteq K$ is invariant.

Definition 22.4.24: Prime Ideal Conjugate

Let $\mathcal{O}_L/\mathcal{O}_K$ be an extension of rings of integers where L/K is Galois, and $\mathfrak{B}$ lie above $\mathfrak{p}$. An ideal $\tau(\mathfrak{B})$ is called a *prime ideal conjugate* to $\mathfrak{B}$.

As we saw, most primes split similarly to a minimal polynomial. The following is further parallel between prime numbers and polynomials: just like conjugate elements, the Galois group acts transitively:

Proposition 22.4.25: Galois Group Transitive Action On Ideals

Let $G = \mathrm{Gal}_K(L)$. Then G acts transitively on the set of all ideal $\mathfrak{B}$ of $\mathcal{O}$ lying above $\mathfrak{p}$, that is the prime ideals are all conjugates of each other.

Proof **:**
Say $\mathfrak{B}, \mathfrak{B}'$ are two primes ideals lying above $\mathfrak{p}$ where, for the sake of contradiction, $\sigma(\mathfrak{B}) \neq \mathfrak{B}'$ for any $\sigma \in G$. By the CRT, there exists an $x \in \mathcal{O}$ such that

$$x \equiv 0 \bmod \mathfrak{B}' \qquad x \equiv 1 \bmod \sigma(\mathfrak{B}) \qquad \forall \sigma \in G$$

Then

$$\mathrm{Nm}_{L/K}(x) = \prod_{\sigma \in G} \sigma(x) \overset{!}{\in} \mathfrak{B}' \cap \mathcal{O}_K = \mathfrak{p}$$

On the other hand, $x \notin \sigma(\mathfrak{B})$ for all $\sigma \in G$, hence $\sigma(x) \notin \mathfrak{B}$ for all $\sigma \in G$ since if $\sigma(x) = b \in \mathfrak{B}$, then $x = \sigma^{-1}(b) = \tau(b)$ for some $\tau \in G$. Thus,

$$\prod_{\sigma \in G} \sigma(x) \overset{!}{\notin} \mathfrak{B} \cap \mathcal{O}_K = \mathfrak{p}$$

Hence, $\prod_{\sigma \in G} \sigma(x)$ is both in and not in $\mathfrak{p}$, a contradiction.

Using the Galois group, we may encode in group-theoretic terms the number of different primes ideals a prime ideal $\mathfrak{p} \subseteq \mathcal{O}_K$ decompose to in $\mathcal{O}_L$. One of the properties of an element being acted on by a [transitive] group is it *stabalizer*:

Definition 22.4.26: Decomposition Group And Field

Let $G = \mathrm{Gal}_K(L)$. Let $\mathfrak{B} \subseteq \mathcal{O}$ be a prime ideal. Then the subgroup

$$D_{\mathfrak{B}/\mathfrak{p}} = D_{\mathfrak{B}} = \{\tau \in G \mid \tau(\mathfrak{B}) = \mathfrak{B}\} = \mathrm{stab}_G(\mathfrak{B})$$

is called the *decomposition group* of $\mathfrak{B}$ over K. The fixed field

$$L^{D_{\mathfrak{B}}} = \{x \in L \mid \tau(x) = x \ \forall \tau \in G_{\mathfrak{B}}\}$$

is called the *decomposition field* of $\mathfrak{B}$ over K, and is often labeled as $L_{\mathfrak{B}}$ or $L_{\mathfrak{B}}$ if unambiguous.

Notice that the less primes that are fixed, the larger the decomposition field is, hence there is a contravariant relation between the size of the group and the extension of the field, just like in Galois theory. Due to transitivity and the introduction of Galois theory, we immediately get many counting arguments. If $\mathfrak{B}$ is a prime ideal lying over $\mathfrak{p}$, and we let σ vary over a set of representatives from the cosets in $G/D_{\mathfrak{B}}$, then by transitivity $\sigma(\mathfrak{B})$ varies over the different primes ideals above $\mathfrak{p}$, each occurring exactly once, hence the number of prime ideals is equal to the index $[G : D_{\mathfrak{B}}]$. For reference, we shall write:

Definition 22.4.27: Prime Counter

Let L/K Then the number of prime numbers is equal to:

$$g_{L/K}(\mathfrak{p}) = [G : D_{\mathfrak{B}}]$$

for any prime $\mathfrak{B}$ over $\mathfrak{p}$. If clear from context, we shall write $g(\mathfrak{p})$.

Thus, if $D_{\mathfrak{B}} = 1$, then no $\sigma \in G$ fixes $\mathfrak{B}$ unless it's the identity and $[G : D_{\mathfrak{B}}] = |G| = [L : K]$, and so by the fundamental identity (proposition 22.4.3):

$$D_{\mathfrak{B}} = 1 \quad \iff \quad L_D = L \quad \iff \quad \mathfrak{p} \text{ totally splits}$$

On the other hand, if $D_{\mathfrak{B}} = G$, then every $\mathfrak{B}$ lying over $\mathfrak{p}$ is fixed by σ, and so there must only be one prime ideal above $\mathfrak{p}$ by transitivity giving:

$$D_{\mathfrak{B}} = G \quad \iff \quad L_D = K \quad \iff \quad \mathfrak{p} \text{ totally ramifies}$$

If $1 \subsetneq D_{\mathfrak{B}} \subsetneq G$, then we know $\mathfrak{p}$ splits to some degree with one of the factors being $\mathfrak{B}$, but we do not know how many factors or the ramification of $\mathfrak{B}$; this shall be given with the inertia group in proposition 22.4.36. Choosing a different prime does change the resulting index: if we were to choose a different prime $\mathfrak{B}$ lying over $\mathfrak{p}$, then $D_{\mathfrak{B}}$ is conjugate to $D_{\sigma(\mathfrak{B})}$:

$$D_{\sigma(\mathfrak{B})} = \sigma D_{\mathfrak{B}} \sigma^{-1}$$

In particular:

$$\begin{aligned} \tau \in D_{\sigma(\mathfrak{B})} &\iff \tau\sigma\mathfrak{B} = \sigma\mathfrak{B} \\ &\iff \sigma^{-1}\tau\sigma\mathfrak{B} = \mathfrak{B} \\ &\iff \sigma^{-1}\tau\sigma \in D_{\mathfrak{B}} \\ &\iff \tau \in \sigma D_{\mathfrak{B}} \sigma^{-1} \end{aligned}$$

This sould make sense, for by the proceeding theorem, the inertia and ramifcation of each prime lying ove $\mathfrak{p}$ in a galois extension is equal. At the end of this section, we shall show how to partially recover these results when working over non-Galois extensions. A key result when working with Galois extensions is that each prime has the same behavior in it's decomposition:

Proposition 22.4.28: Galois Extension, Inertia Degree And Ramification Index Independent

Let $\mathfrak{p} \subseteq K$ be a prime ideal, L/K a Galois extension, and

$$\mathfrak{p}\mathcal{O}_L = \mathfrak{B}_1^{e_i} \cdots \mathfrak{B}_r^{e_r}$$

Then

$$f_1 = \cdots = f_r = f \qquad e_1 = \cdots = e_r = e$$

Proof **:**
Let $\mathfrak{B} = \mathfrak{B}_1$ and $\mathfrak{B}_i = \sigma_i(\mathfrak{B})$ (which goes through all primes by transitivity). Then the isomorphism $\sigma_i : \mathcal{O}_L \to \mathcal{O}_L$ induces an isomorphism:

$$\mathcal{O}_L/\mathfrak{B} \xrightarrow{\sim} \mathcal{O}_L/\mathfrak{B}_i \qquad a \bmod \mathfrak{B} \mapsto \sigma_i(a) \bmod \sigma_i(\mathfrak{B})$$

But then:

$$f_i = [\mathcal{O}_L/\sigma_i(\mathfrak{B}) : \mathcal{O}_K/\mathfrak{p}] = [\mathcal{O}_L/\mathfrak{B} : \mathcal{O}_K/\mathfrak{p}]$$

Next, since $\sigma_i(\mathfrak{p}\mathcal{O}_L) = \mathfrak{p}\mathcal{O}_L$ (since $\mathfrak{p} \subseteq K$), we get

$$\mathfrak{B}^n \big| \mathfrak{p}\mathcal{O} \quad \iff \quad \sigma_i(\mathfrak{B}^n) \big| \sigma_i(\mathfrak{p}\mathcal{O}) \quad \iff \quad (\sigma_i(\mathfrak{B}))^n \big| \mathfrak{p}\mathcal{O}$$

showing that $e_1 = \cdots = e_r$.

Hence, the prime decomposition of $\mathfrak{p}$ can be written as :

$$\mathfrak{p} = \left(\prod_{\substack{\sigma \in S \\ S \subseteq G/D_{\mathfrak{B}}}} \sigma(\mathfrak{B}) \right)^e$$

where $S \subseteq G/D_{\mathfrak{B}}$ represents a set of representatives of $G/D_{\mathfrak{B}}$ and $|S| = g_{L/K}(\mathfrak{p})$. This information is already very powerful, the following is a use-case of such information:

Example 22.4.4: Biquadratic Extension That Is Not Monogenic

Let $K = \mathbb{Q}(\sqrt{7}, \sqrt{13})$. Without computing $\mathcal{O}_K$, it can be shown that $\mathcal{O}_K$ is *not* monogenic, that is there does not exists $\alpha \in \mathcal{O}_K$ such that $1, \alpha, \alpha^2, \alpha^3$ generates $\mathcal{O}_K$, in other words $\mathcal{O}_K$ is not isomorphic to a quotient of $\mathbb{Z}[x]$. The key is to look at the behavior of factorizations of small primes (for they can only have so many rood $\bmod p$)

What we shall do is consider the factorization of (3) in $\mathcal{O}_K$, which will be our setup for a

contradiction. First, notice that K has 3 intermediary fields[a]:

$$\begin{tikzcd}
 & K & \\
\mathbb{Q}(\sqrt{7}) \arrow[ur, "2"] & \mathbb{Q}(\sqrt{91}) \arrow[u, "2"] & \mathbb{Q}(\sqrt{13}) \arrow[ul, "2"'] \\
 & \mathbb{Q} \arrow[ul, "2"] \arrow[u, "2"] \arrow[ur, "2"'] &
\end{tikzcd}$$

Since we know the ring of integers of quadratics, by direct computation, we see that (3) splits into two primes in each quadratic extension. For example, in $\mathbb{Q}(\sqrt{7})$, we have:

$$\mathcal{O}_{\mathbb{Q}(\sqrt{7})}/3\mathcal{O}_{\mathbb{Q}(\sqrt{7})} = \frac{\mathbb{F}_3[x]}{(x-1)(x+1)}$$

Hence:

$$(3) = (3, 1+\sqrt{7})(3, 1-\sqrt{7})$$

Doing a similar computation for the other two computation, we get:

$$(3) = (3, 1+\sqrt{13})(3, 1-\sqrt{13})$$
$$(3) = (3, 1+\sqrt{91})(3, 1-\sqrt{91})$$

Thus, over K, there must be at least 2 prime ideals, and since

$$g_{K/\mathbb{Q}}(3) = [G : D_{\mathfrak{B}}] \qquad [G : D_{\mathfrak{B}}] \;\Big|\; |\operatorname{Gal}_{\mathbb{Q}}(K)| \qquad |\operatorname{Gal}_{\mathbb{Q}}(K)| = |(\mathbb{Z}/2\mathbb{Z})^2| = 4$$

for any prime ideal $\mathfrak{B}$ over (3), we get that $g_{K/\mathbb{Q}}(3)\big|4$. I claim that $g_{K/\mathbb{Q}}(3) = 4$. For, if $g_{K/\mathbb{Q}}(3) = 2$, then for any of the two primes lying over (3) in $\mathcal{O}_K$, say $\mathfrak{B}$:

$$\{e\} \neq D_{\mathfrak{B}} = \{\sigma \in \operatorname{Gal}_{\mathbb{Q}}(K) \mid \sigma(\mathfrak{B}) = \mathfrak{B}\} \subsetneq \operatorname{Gal}_{\mathbb{Q}}(K)$$

Since $D_{\mathfrak{B}}$ non-trivial proper subgroup acting transitively on the two primes of K lying over (3), we have $|D_{\mathfrak{B}}| = 2$, implying there is one non-trivial automorphism $\rho \in \operatorname{Gal}_{\mathbb{Q}}(K)$ fixing the prime ideals (since there are two prime ideals, if $\rho(\mathfrak{B}) = \mathfrak{B}$, then $\rho(\mathfrak{B}') = \mathfrak{B}'$, or else ρ wouldn't be bijective), and two non-trivial automorphisms σ, τ not fixing the primes. Consider now $\sigma \in \operatorname{Gal}_{\mathbb{Q}}(K)$ and take the fixed field $K^{\langle\sigma\rangle}$, which by some Galois theory we see must be one of the 3 quadratic extensions found above. Then by the 3rd fundamental theorem of Galois theory:

$$G := \operatorname{Gal}_{\mathbb{Q}}(K^{\langle\sigma\rangle}) \cong \frac{\operatorname{Gal}_{\mathbb{Q}}(K)}{\operatorname{Gal}_{K^{\langle\sigma\rangle}}(K)}$$

Importantly, since $G \cong \mathbb{Z}/2\mathbb{Z}$ and $[\sigma] = [\mathrm{id}]$, the map $\varphi : \operatorname{Gal}_{\mathbb{Q}}(K) \to \operatorname{Gal}_{\mathbb{Q}}(K^{\langle\sigma\rangle})$ maps $\rho, \tau \mapsto \rho|_{K^{\langle\sigma\rangle}}$. Since ρ is an automorphism and fixes both prime ideals in K over (3) (say $\mathfrak{B}, \mathfrak{B}'$), and all prime ideal of $K^{\langle\sigma\rangle}$ over (3) lie below $\mathfrak{B}, \mathfrak{B}'$:

$$\rho|_{K^{\langle\sigma\rangle}}(\mathfrak{B} \cap K^{\langle\sigma\rangle}) = \rho|_{K^{\langle\sigma\rangle}}(\mathfrak{B}) \cap K^{\langle\sigma\rangle} = \mathfrak{B} \cap K^{\langle\sigma\rangle}$$
$$\rho|_{K^{\langle\sigma\rangle}}(\mathfrak{B}' \cap K^{\langle\sigma\rangle}) = \rho|_{K^{\langle\sigma\rangle}}(\mathfrak{B}') \cap K^{\langle\sigma\rangle} = \mathfrak{B}' \cap K^{\langle\sigma\rangle}$$

Hence, G fixes all primes; but we saw that all intermediate fields of K have two primes, thus this contradicts the transitivity of action of the Galois group on the prime ideals. Thus, it must be that $g_{K/\mathbb{Q}}(3) = 4$.

Now, for the sake of contradiction let's say

$$\mathcal{O}_K \cong \frac{\mathbb{Z}[x]}{m_\alpha(x)}$$

for some $\alpha \in \mathcal{O}_K$ and minimal polynomial $m_\alpha(x)$ of degree 4. Then

$$\mathcal{O}_K/3\mathcal{O}_K \cong \frac{\mathbb{F}_3[x]}{(m_\alpha(x) \bmod 3)} \tag{22.5}$$

By our previous calculations, we see that there must be 4 prime ideals in the right hand side. Decomposing $m_\alpha(x)$ mod 3 into irreducible, by the CRT we get that the right hand side is isomorphic to the direct sum of $\mathbb{F}_3[x]/(m_i(x) \bmod 3)$ for each irreducible $m_i(x)$ mod 3. Since each $m_i(x)$ mod 3 is irreducible, each direct summand is a field, and hence by some ring theory we see that the number of primes on the right hand side of equation 22.5 is equal to the number of direct summands, implying $m_\alpha(x)$ splits into linear factors giving us:

$$\frac{\mathbb{F}_3[x]}{(m_\alpha(x) \bmod 3)} \cong \bigoplus_i^4 \mathbb{F}_3$$

However, this is impossible: any degree 4 polynomial over $\mathbb{F}_3$ that breaks down into linear factors has the form:

$$(x-1)^\alpha (x+1)^\beta x^\gamma$$

where one of α, β, γ is 2 and the other two are 1. But then $\mathbb{F}_3[x]/((x-1)^\alpha(x+1)^\beta x^\gamma)$ is not isomorphic to $\bigoplus_i^4 \mathbb{F}_3$ (the quotient contains nilpotent elements), showing that our assumption that $\mathcal{O}_K$ is monogenic is not possible.

[a] this can be seen by taking the fixed fields $\sqrt{7}, \sqrt{13}$, and $\sqrt{91}$, corresponding to the subgroups of $\mathrm{Gal}_{\mathbb{Q}}(K) = \mathbb{Z}_2 \oplus \mathbb{Z}_2$

Our next goal is to break down extension L/K into intermediary extensions which will "capture" the inertial degree and ramification index, similarly to how any field extension F/k may be broken down into intermediary extensions into separable, inseparable, and transcendental extensions. In our case, we shall get the tower:

$$K \xrightarrow[1]{1} L_D \xrightarrow[f]{1} L_I \xrightarrow[1]{e} L$$

where the ramification index of each field extension is on top, and the inertia degree is on the bottom.

Proposition 22.4.29: Prime Ideals in Decomposition Field

Let L/K be a finite field extension, $\mathfrak{p} \subseteq \mathcal{O}_K$ a prime ideal and $\mathfrak{B} \subseteq \mathcal{O}_L$ lying over $\mathfrak{p}$. Let $\mathfrak{B}_Z = \mathfrak{B} \cap L_D$ be the prime ideal of L_D below $\mathfrak{B}$. If $\mathfrak{B}$ has inertia degree f and ramification index e, then:

1. $\mathfrak{B}_Z$ is nonsplit in L, that is $\mathfrak{B}$ is the only prime ideal of L above $\mathfrak{B}_D$
2. $\mathfrak{B}$ over L_D has a ramification index e and inertia index f
3. The ramification index and inertia index of $\mathfrak{B}_D$ over K are both 1

Visually, we have:

$$\begin{tikzcd} \mathfrak{B}^e J \\ \mathfrak{B}_D I \arrow[u, "e"', "f"] \\ \mathfrak{p} \arrow[u, "1"', "1"] \end{tikzcd}$$

where I and J are the remainder of the factorization. Note that the factorization of each prime factor of J lying above $\mathfrak{p}$ has the same e and f since L/K is Galois, but L_D/K is not necessarily Galois, and so the factorization of I may be quite different (for ex. another prime $\mathfrak{B}'_D$ in L_D lying ove $\mathfrak{B}$ may have ramification and inertia degree)

Proof **:**

1. Notice that $\mathrm{Gal}_{L_{\mathfrak{B}}}(L) = D_{\mathfrak{B}}$: the automorphisms of L that fix $L_{\mathfrak{B}}$ are by definition the elements of $D_{\mathfrak{B}}$. Hence, the prime ideals above $\mathfrak{B}_D$ are the $\sigma(\mathfrak{B})$ for $\sigma \in \mathrm{Gal}_{L_{\mathfrak{B}}}(L)$, but $\sigma(\mathfrak{B}) = \mathfrak{B}$ by definition. Hence, $\mathfrak{B}_D$ is nonsplit in L.

2-3. The fundamental identity gives us:

$$n = efr$$

where $n = |\mathrm{Gal}_K(L)| = |G|$ and $r = [G : D_{\mathfrak{B}}]$. hence,

$$|D_{\mathfrak{B}}| = [L : L_{\mathfrak{B}}] = ef \tag{22.6}$$

Now, let e', e'' be the ramification index of $\mathfrak{B}$ over $L_{\mathfrak{B}}$ and $\mathfrak{B}_D$ over K respectively. Then $\mathfrak{p}$ is:

$$\mathfrak{p}\mathcal{O}_{L_{\mathfrak{B}}} = \mathfrak{B}_D^{e''} I$$

for some I and

$$\mathfrak{B}_D \mathcal{O}_L = \mathfrak{B}^{e'}$$

in L. Then

$$\mathfrak{p}\mathcal{O}_L = \mathfrak{B}^{e''e'} J$$

for some J. By the tower property, $e = e'e''$ and $f = f'f''$. Then the fundamental identity for the decomposition of $\mathfrak{B}_D$ in L gives us that the degree of L/L_D is:

$$[L : L_{\mathfrak{B}}] = e'f'$$

in particular, by equation (22.6), we get,

$$e'f' = ef = e'e''f'f''$$

Hence $e' = e$, $f' = f$, and $e'' = f'' = 1$, as we sought to show

This gives us:

$$K \underset{1}{\overset{1}{\to}} L_{\mathfrak{B}} \underset{f}{\overset{e}{\to}} L$$

Next, we see how we can separate out an extension splitting the ramification and the inertia degree into different groups/fields. Since $\sigma(\mathcal{O}_L) = \mathcal{O}_L$ and $\sigma(\mathfrak{B}) = \mathfrak{B}$ for every $\sigma \in D_{\mathfrak{B}}$, there is an induced automorphism on the residue field $\mathcal{O}_L/\mathfrak{B}$:

$$\begin{gathered}\overline{\sigma} : \mathcal{O}_L/\mathfrak{B} \to \mathcal{O}_L/\mathfrak{B} \\ (a \bmod \mathfrak{B}) \mapsto (\sigma(a) \bmod \mathfrak{B})\end{gathered}$$

Letting $\kappa(\mathfrak{B}) = \mathcal{O}_L/\mathfrak{B}$ and $\kappa(\mathfrak{p}) = \mathcal{O}_K/\mathfrak{p}$, we have that

$$f = [\kappa(\mathfrak{B}) : \kappa(\mathfrak{p})]$$

Recall that in Dedekind-Kummer's theorem, when $\mathfrak{p}$ is relatively prime to the conductor, the inertia degree of $\mathfrak{B}$ is given by the degree of $\overline{m_i}(x)$ which within the proof is derived by $\left[R/(\overline{m_i(x)}) : \overline{\mathcal{O}}_K\right]$. This shows that the above map has information about the inertia degree. The following links this quantity to $D_{\mathfrak{B}}$. To understand the result, the reader may see that $\mathrm{Gal}_K(L)$ would naturally surjects onto $\mathrm{Gal}_{\kappa(\mathfrak{p})}(\kappa(\mathfrak{B}))$ via restriction if the restriction was well-defined. The domain of this map can be restricted to the decomposition group to make it well-defined:

Proposition 22.4.30: Decomposition and Inertia Group Relation

The extension $\kappa(\mathfrak{B})/\kappa(\mathfrak{p})$ is normal and admits a surjective homomorphism:

$$D_{\mathfrak{B}} \twoheadrightarrow \mathrm{Gal}_{\kappa(\mathfrak{p})}(\kappa(\mathfrak{B})) = \mathrm{Gal}_{\mathcal{O}_K/\mathfrak{p}}(\mathcal{O}_L/\mathfrak{B}) \cong \mathrm{Gal}_{\mathbb{F}_p}(\mathbb{F}_q)$$

Proof **:**

As we are working over finite fields, we know that $\kappa(\mathfrak{B}) = \kappa(\mathfrak{p})(\overline{\theta})$ for some θ with minimal polynomial $\overline{g}(x)$ over $\kappa(\mathfrak{p})$ ($\overline{g}(x) \in \kappa(\mathfrak{p})[x]$)). Let $\theta \in \mathcal{O}_L$ be a representative of $\overline{\theta}$, and let $f(x)$ be its minimal polynomial over $\mathcal{O}_K$. Then by definition $\overline{g}(\overline{\theta}) \equiv 0 \bmod \mathfrak{B}$. Since $f(\theta) = 0$, $\overline{f}(\overline{\theta}) \equiv 0 \bmod \mathfrak{B}$. Hence, taking $\overline{f}(x) \in \kappa(\mathfrak{p})[x]$, we have by minimality that:

$$\overline{g}(x) \big| \overline{f}(x)$$

Next, we know that finite field extensions are normal as we saw in section 17.7.2, however here is another way of seeing this worth pointing out: since L/K is normal, $f(x)$ splits over $\mathcal{O}_L$ into linear factors. Thus, $\overline{f}(x)$ splits into linear factors over $\kappa(\mathfrak{B})$, and similarly for $\overline{g}(x)$. But then, $\kappa(\mathfrak{B})/\kappa(\mathfrak{p})$ is a normal extension.

Now, consider:

$$\overline{\sigma} \in \mathrm{Gal}_{\kappa(\mathfrak{p})}(\kappa(\mathfrak{B})) = \mathrm{Gal}_{\kappa(\mathfrak{p})}\left(\kappa(\mathfrak{p})(\overline{\theta})\right)$$

By corollary 17.2.15, these homomorphisms are determined by where they send the root θ. Then $\overline{\sigma}(\overline{\theta})$ is a root of $\overline{g}(x)$ and hence of $\overline{f}(x)$, that is there exists a root θ' of $f(x)$ such that

$\theta' \equiv \overline{\sigma}(\overline{\theta}) \bmod \mathfrak{B}$. θ' is a conjugate of θ, i.e. $\theta' = \sigma(\theta)$ for some $\sigma \in D_{\mathfrak{B}}$ (as the automorphism in $D_{\mathfrak{B}}$ fix $\mathfrak{B}$ and make the map well-defined). Since $\sigma(\theta) \equiv \overline{\sigma}(\overline{\theta}) \bmod \mathfrak{B}$, the automorphism σ is mapped by the homomorphism in question to $\overline{\sigma}$, proving surjectivity, as we sought to show.

Definition 22.4.31: Inertia Group And Field

The kernel $I_{\mathfrak{B}} \subseteq D_{\mathfrak{B}}$ of the homomorphism

$$D_{\mathfrak{B}} \to \mathrm{Gal}_{\kappa(\mathfrak{p})}(\kappa(\mathfrak{B}))$$

is called the *inertia group* of $\mathfrak{B}$ over K which is a normal subgroup of $D_{\mathfrak{B}}$ (as it is the kernel of a homomorphism). Equivalently:

$$I_{\mathfrak{B}} = \{\sigma \in \mathrm{Gal}_K(L) \mid \sigma(\alpha) \equiv \alpha \bmod \mathfrak{B}\}$$

The fixed field:

$$L_I = L^{I_{\mathfrak{B}}} = \{x \in L \mid \sigma(x) = x,\ \forall \sigma \in I_{\mathfrak{B}}\}$$

is called the *inertia field* of $\mathfrak{B}$ over K.

We thus get the following tower:

$$K \subseteq L_D \subseteq L_I \subseteq L$$

and short exact sequence:

$$1 \to I_{\mathfrak{B}} \to D_{\mathfrak{B}} \to \mathrm{Gal}_{\kappa(\mathfrak{p})}(\kappa(\mathfrak{B})) \to 1$$

Corollary 22.4.32: Isomorphism Of Two Galois Groups

Let $L/K/\mathbb{Q}$ be a finite galois extension, and let $\mathfrak{B}$ be a prime lying over $\mathfrak{p} \subseteq \mathcal{O}_K$. Then:

$$\mathrm{Gal}_{L_D}(L_I) \cong \mathrm{Gal}_{\kappa(\mathfrak{p})}(\kappa(\mathfrak{B}))$$

Proof **:**
exercise.

Corollary 22.4.33: Quotient Group of Decomposition Group

Let L/K be a finite extension. Then the group $D_{\mathfrak{B}}/I_{\mathfrak{B}}$ is cyclic of order f (hence $[D_{\mathfrak{B}} : I_{\mathfrak{B}}] = f$), and the generator $[\varphi] \in D_{\mathfrak{B}}/I_{\mathfrak{B}}$ that generates the group (and hence $\mathrm{Gal}_{\kappa(\mathfrak{p})}(\kappa(\mathfrak{B})) \cong \mathbb{Z}/m\mathbb{Z}$) is called the *Frobenius Element*. The Frobenius element $\varphi \in D_{\mathfrak{B}}$ can be represented in the simple form

$$\varphi(x) = x^{\mathfrak{N}}(\mathfrak{p})$$

Proof **:**
This is an immediate consequence of the surjectivity of $D_{\mathfrak{B}}$ onto the Galois group of finite field extensions, which is always cyclic.

The order of $\kappa(\mathfrak{B})/\kappa(\mathfrak{p})$ is f, giving the order of the quotient. As the galois group of a finite field is

cyclic, it is generated by a Frobenius element. The Frobenius element can be explicitly described, namely

$$\varphi(x) = x^{\mathfrak{N}(\mathfrak{p})} = x^f$$

giving us the rest of the results.

Corollary 22.4.34: Equivalent Characterization Of Inertia Group

L/K be a finite Galois extension, $I_{\mathfrak{B}} \leqslant D_{\mathfrak{B}}$. Then:

$$I_{\mathfrak{B}} = \{\sigma \in \mathrm{Gal}_K(L) \mid \sigma(\mathfrak{B}) = \mathfrak{B}, \text{ and } \forall \alpha \in \mathcal{O}_L, \sigma(\alpha) - \alpha \in \mathfrak{B}\}$$

Proof **:**
Let $\sigma \in I_{\mathfrak{B}}$ so that $\overline{\sigma\overline{\alpha}} = \overline{\alpha}$. Then $\sigma(\alpha) = \alpha + b$ for $b \in \mathfrak{B}$. But then:

$$\sigma(\alpha) - \alpha = b \in \mathfrak{B}$$

as we sought to show.

Corollary 22.4.35: Finding Element Of Galois Group

Let K be a number field, $f(x) \in \mathcal{O}_K[x]$ an irreducible polynomial over $\mathcal{O}_K$, L/K the splitting field extension of $f(x)$. Let $\mathfrak{p} \subseteq \mathcal{O}_K$ such that

$$f(x) \bmod \mathfrak{p} \equiv g_1(x) \cdots g_m(x)$$

for distinct irreducibles with degree $\deg(g_i(x)) = d_i$. Then in $\mathrm{Gal}_K(L) \lesssim S_n$, there exists an element α with cycle type

$$(d_1, d_2, ..., d_m)$$

Proof **:**
Since L is the splitting field, let $\alpha_1, \alpha_2, ..., \alpha_m$ be the roots of f so that

$$f(\alpha_i) = 0$$

Let $\mathfrak{B}$ in L lie above $\mathfrak{p}$. Then:

$$f(\alpha_i) \bmod \mathfrak{p} \equiv 0$$

Since each each g_i are distinct, each $\alpha_i \bmod \mathfrak{p}$ are disticnt. Hence, $D_{\mathfrak{B}}$ acts on the roots of $f \bmod \mathfrak{p}$. Now, taking $g \in D_{\mathfrak{B}}$ that maps to the generator of $\mathrm{Gal}_{\kappa(\mathfrak{p})}(\kappa(\mathfrak{B}))$, we shall get the desired cycle type, completing the proof.

Proposition 22.4.36: Inertia Group Properties

The extension L_I/L_D is normal, and

$$\mathrm{Gal}_{L_D}(L_I) \cong \mathrm{Gal}_{\kappa(\mathfrak{p})}(\kappa(\mathfrak{B}))$$

and $\mathrm{Gal}_{L_I}(L) = I_{\mathfrak{B}}$. If the residue field extension $\kappa(\mathfrak{B})/\kappa(\mathfrak{p})$ is separable, then:

$$|I_{\mathfrak{B}}| = [L : L_I] = e \qquad [D_{\mathfrak{B}} : I_{\mathfrak{B}}] = [L_I : L_D] = f$$

and we can find the prime ideals $\mathfrak{B}_I$ of L_I bellow $\mathfrak{B}$:

1. The ramification index of $\mathfrak{B}$ over $\mathfrak{B}_I$ is e and the inertia degree is 1
2. The ramification index of $\mathfrak{B}_I$ over $\mathfrak{B}_D$ is 1, and the inertia degree if f

Hence:

$$\mathfrak{B}_Z\mathcal{O}_{L_I} = \mathfrak{B}_T I \qquad \mathfrak{B}_T\mathcal{O}_K = \mathfrak{B}^e J$$

for some appropriate ideals I, J.

***Proof* :**
The fact that:

$$|I_{\mathfrak{B}}| = [L : L_I] = e \qquad [D_{\mathfrak{B}} : I_{\mathfrak{B}}] = [L_I : L_D] = f$$

follow from $|D_{\mathfrak{B}}| = [L : L_D] = ef$ and Galois theory (namely $[L : L_I] = |\mathrm{Gal}_{L_I}(L)| = |I_{\mathfrak{B}}|$), hence we need to show $(1) - (2)$. Using the fundamental identity, it all follows from $\kappa(\mathfrak{B}_I) = \kappa(\mathfrak{B})$. Since the inertia group $I_{\mathfrak{B}}$ of $\mathfrak{B}$ over K is also the inertia group of $\mathfrak{B}$ over L_I, it follows from proposition 22.4.30 applied to L/L_I that $\mathrm{Gal}_{\kappa(\mathfrak{B}_I)}(\kappa(\mathfrak{B})) = 1$, thus

$$\kappa(\mathfrak{B}_I) = \kappa(\mathfrak{B})$$

but then the rest follows.

To sum up the results in a diagram

$$K \underset{1}{\overset{1}{\rightarrow}} L_D \underset{f}{\overset{1}{\rightarrow}} L_I \underset{1}{\overset{e}{\rightarrow}} L$$

with a corresponding [contravariant] diagram of groups:

$$G \leftarrow D_{\mathfrak{B}} \leftarrow I_{\mathfrak{B}} \leftarrow \{e\}$$

where the ramification index of each field extension is on top, and the inertia degree is on the bottom. If the residue field extension $\kappa(\mathfrak{B})/\kappa(\mathfrak{p})$ is separable, then:

$$I_{\mathfrak{B}} = 1 \qquad \Longleftrightarrow \qquad L_I = L \qquad \Longleftrightarrow \qquad \mathfrak{p} \text{ is unramified in } L$$

and in this case the Galois group $\mathrm{Gal}_{\kappa(\mathfrak{p})}(\kappa(\mathfrak{B})) \cong D_{\mathfrak{B}}$, so the galois group of residue field extension can be viewed as a subgroup of $\mathrm{Gal}_K(L)$. This means that the inertia group can be thought of as the "ramification group". In fact, the generalizations of the inertia group (giving more nuanced information on the ramificaion) are called the *higher order ramification group* (which we shall cover soon). The

reason the inertia group is not called the 1st ramification group is because the corresponding fixed field only effects the inertia of a prime; ultimately this is due to the contravariant relation the galois correspondence gives, hence we must make a choice on naming convention depending if fields or groups are prioritised. In this case, fields were prioritised.

From the previous proposition, we may think that L_D and L_I are the *maximal* fields with these properties with respect to $\mathfrak{B}$, and this is indeed the case, making them "natural" with respect to these properties:

Proposition 22.4.37: Intermediary Fields Are Maximal

Let L/K be a finite extension of number fields, $\mathfrak{B}$ live over $\mathfrak{p}$. Then:

1. $L_{D_{\mathfrak{B}}}$ is the smallest intermediary field such that $\mathfrak{B}$ is the only prime of $\mathcal{O}_L$ lying over $\mathfrak{B}' = \mathfrak{B} \cap \mathcal{O}_{L_{D_{\mathfrak{B}}}}$
2. $L_{D_{\mathfrak{B}}}$ is the largest intermediate such that $e(\mathfrak{B}_D/\mathfrak{p}) = f(\mathfrak{B}_I/\mathfrak{p}) = 1$
3. $L_{I_{\mathfrak{B}}}$ is the smallest intermediary field such that $\mathfrak{B}$ is totally ramified over $\mathfrak{B}_I$ (that is $e(\mathfrak{B}/\mathfrak{B}_I) = [L : L_I]$)
4. $L_{I_{\mathfrak{B}}}$ is the largest intermediary field such that $e(\mathfrak{B}_I/\mathfrak{p}) = 1$

***Proof* :**
First, L_D and L_I have these properties as we've in proposition 22.4.29 and proposition 22.4.36, hence we prove they are maximal/minimal.

Let $K' = L^H$ for some $H \leqslant \mathrm{Gal}_K(L)$ where only $\mathfrak{B} \subseteq L$ is lying over $\mathfrak{p}' \subseteq K'$. Since H acts transitively on the primes of L over K' but there is only one prime $H \subseteq D_{\mathfrak{B}}$, so $L_D \subseteq K'$, giving (1).

Next, assume $e_{\mathfrak{p}'/\mathfrak{p}} = f_{\mathfrak{p}'/\mathfrak{p}} = 1$. Consider the following diagram:

$$\begin{tikzcd}
L & \\
L_I \arrow[u, "e"] & L_I K' \arrow[ul, "e'"'] \\
L_D \arrow[u, "f"] \arrow[ur] & L_D K' \arrow[u, "f'"'] \\
K \arrow[u, "g"] \arrow[ur] & K' \arrow[u, "g'"']
\end{tikzcd}$$

Then by multiplicativity of towers, $e = e'$ and $f = f'$. Then by the above diagram, L_D and $L_D K'$ voth ahve the same index in L. Since one contains the other, this means they must be equal, hence

$$K' \subseteq L_D$$

giving (2). A similar arguments works for (3) and (4): if $e_{\mathfrak{p}'/\mathfrak{p}} = 1$, then $e = e'$, so $L_I = L_I K'$, so $K' \subseteq L_I$, and if $\mathfrak{B}$ is totally ramified over $\mathfrak{p}'$, then $[L : K'] = e'$, then by the above diagramwe get $K' = L_I K'$, so $L_I \subseteq K'$, as we sought to show.

Ramification

An importance consequence of this is that we may consider primes that are completely ramified/totally unramified in their galois closure. we first require the following result:

Theorem 22.4.38: Composition Field And Primes

Let L/K and M/K be finite extensions of number fields, and let $\mathfrak{p}$ be a prime of K. Then if $\mathfrak{p}$ is unramified in both L and M, then $\mathfrak{p}$ is unramified in LM. If $\mathfrak{p}$ splits completely in L and M, then $\mathfrak{p}$ splits completely in LM

***Proof* :**
Let $\mathfrak{p}$ be unramified in L and M, and take $\mathfrak{B}$ to be any prime in LM lying over $\mathfrak{p}$. We want to show $e(\mathfrak{B}/\mathfrak{p}) = 1$. Let F be the normal closure of K containing LM, and let $\mathfrak{A}$ be any prime in F lying over $\mathfrak{B}$, namely it lies over $\mathfrak{p}$. Let $I_{\mathfrak{A}/\mathfrak{p}}$ be the inertial group corresponding to the field F_I. Then we see that F_I contains both L and M since the primes $\mathfrak{A} \cap L$ and $\mathfrak{A} \cap M$ are necessarily unramified over $\mathfrak{p}$. Hence, F_I contains LM, implying that $\mathfrak{A} \cap LM$ is unramified over $\mathfrak{p}$.

The argument for splitting follows exactly the same pattern, replacing $I_{\mathfrak{B}}$ with $D_{\mathfrak{B}}$.

Corollary 22.4.39: Primes In Normal Closure

Let L/K be a finite extension of number fields and $\mathfrak{p}$ a prime in K. Then:

1. If $\mathfrak{p}$ is unramified in L, then it is unramified in the normal closure of M of L over K
2. If $\mathfrak{p}$ splits completely in L, then it splits completely in the normal closure of M of L over K

***Proof* :**
If $\mathfrak{p}$ is unramified in L, then it is unramified in every $\sigma_i(L)$ where i runs over all the automorphism. But since $M = L\sigma_1(L) \cdots \sigma_n(L)$, we use induction on theorem 22.4.38 to get our desired result (similarly for splitting completely).

Example 22.4.5: More Decomposition Of Primes

Let $K = \mathbb{Q}(i, \sqrt{2}, \sqrt{5})$, which is abelian of degree 8. Take the prime $5 \in \mathbb{Z}$. Then it splits into tow primes in $\mathbb{Q}(i)$, stays inert in $\mathbb{Q}(\sqrt{2})$, and becomes a square in $\mathbb{Q}(\sqrt{5})$. Hence in L 5 must have ramification index at least 2, and split into at least 2 primes. Since the inertia field must be a field of degree 4 over $\mathbb{Q}$ in which 5 is unramified, the only choice is $\mathbb{Q}(i, \sqrt{2})$. Now, $(2+i)$ and $(2-i)$ remain prime in $\mathbb{Q}(\sqrt{i}, \sqrt{2})$, and become squares in L, hence 5 must have exactly 2 primes and ramification index 2.

Since normal subgroups over characteristic 0 fields are galois, it is meaningful to ask the behavior of normal subgroups of the galois subgroups:

Proposition 22.4.40: Decomposition Subgroup Is Normal

Suppose $D_{\mathfrak{B}} \trianglelefteq \mathrm{Gal}_K(L)$. Then $\mathfrak{p}\mathcal{O}_L$ splits into r distinct primes in L_D. If $I_{\mathfrak{B}} \trianglelefteq G$, then each prime remains as primes in L_I (i.e. they are inert). Furthermore, each become an eth power in L.

In general, we can only hope for:

$$\mathfrak{B}_D \mathcal{O}_{L_I} = \mathfrak{B}_I I \qquad \mathfrak{B}_I \mathcal{O}_K = \mathfrak{B}^e J$$

for some ideals, while in the normal case of D and I, we get:

$$\begin{aligned} \mathfrak{p}\mathcal{O}_{L_D} &= \mathfrak{B}_{L_1} \cdots \mathfrak{B}_{L_e} \\ \mathfrak{p}\mathcal{O}_{L_I} &= \mathfrak{B}_{L_1} \cdots \mathfrak{B}_{L_e} \\ \mathfrak{p}\mathcal{O}_L &= \mathfrak{B}_1^e \cdots \mathfrak{B}_g^e \end{aligned}$$

which also shows better why the groups $D_{\mathfrak{B}}$ and $I_{\mathfrak{B}}$ are called the "decomposition" and "inertia" groups respectively.

Proof **:**

This is taking advantage of the fact that $L^{D_{\mathfrak{B}}}/K$ would now be a Galois extension since $D_{\mathfrak{B}} \trianglelefteq \mathrm{Gal}_K(L)$. Then we know that

$$f(\mathfrak{B}_D) = e(\mathfrak{B}_D) = 1$$

and that (since $L^{D_{\mathfrak{B}}}$ is Galois), each other prime must also have the same ramification and inertia degree. Thus, $\mathfrak{p}$ splits into exactly g distinct primes in $L^{D_{\mathfrak{B}}}$. Then by a simple counting argument, we see that each $\mathfrak{B}_D$ must lie under a unique $\mathfrak{B}_I$. It may be that some of these ramify. However, if $I_{\mathfrak{B}} \trianglelefteq \mathrm{Gal}_K(L)$, then we have that each prime must have ramification index 1, showing they are all inert. Finally, a counting argument gives that each $\mathfrak{B}_I$ becomes an eth power in L

Corollary 22.4.41: Normal Subgroup And Completely Split Prime

Let $D_{\mathfrak{B}} \trianglelefteq \mathrm{Gal}_K(L)$. Then $\mathfrak{p}$ splits completely in K' if and only if $K' \subseteq L_D$

Proof **:**
exercise.

In the non-Galois case L/K, take $M/L/K$ where M/K is the Galois closure of L. Then M/L is Galois. Let $G = \mathrm{Gal}_K(M)$ and $H = \mathrm{Gal}_L(M)$ so that $H \leqslant G$. Take the following tower of primes:

$$\begin{tikzcd} M \arrow[r] & \mathfrak{A} \\ L \arrow[u] \arrow[r] & \mathfrak{B} \arrow[u] \\ K \arrow[u] \arrow[r] & \mathfrak{p} \arrow[u] \end{tikzcd}$$

Then by definition:

$$D_{\mathfrak{A}/\mathfrak{p}} \cap H = D_{\mathfrak{A}/\mathfrak{B}}$$
$$I_{\mathfrak{A}/\mathfrak{p}} \cap H = I_{\mathfrak{A}/\mathfrak{B}}$$

Proposition 22.4.42: Prime Decomposition Over Finite Extension

Let

$$\mathfrak{p}_M = \prod_1^{g_{M/K}} \mathfrak{A}_i^{e_i}$$

Let $\mathfrak{O}_1, \mathfrak{O}_2, ..., \mathfrak{O}_{g_{L/K}}$ be the orbits of H acting on the primes $\{\mathfrak{A}_1, \mathfrak{A}_2, ..., \mathfrak{A}_{g_{M/K}}\}$. Then:

$$\mathfrak{p}_L = \prod_i^{g_{L/K}} \mathfrak{B}_i^{e_{\mathfrak{B}_i/\mathfrak{p}}}$$

where each $\mathfrak{B}_i$ comes form an $\mathfrak{O}_i$. Furthermore:

$$e_{\mathfrak{B}_i/\mathfrak{p}} = \frac{|I_{\mathfrak{A}_i/\mathfrak{p}}|}{|\operatorname{stab}_H(\mathfrak{A}_i) \cap I_{\mathfrak{A}_1/\mathfrak{p}}|} \qquad f_{\mathfrak{B}_i/\mathfrak{p}} = \frac{|D_{\mathfrak{A}_i/\mathfrak{B}_i}|}{|e_{\mathfrak{A}_i/\mathfrak{p}}|} = \frac{|\operatorname{stab}_H(\mathfrak{A}_i)|}{|e_{\mathfrak{A}_i/\mathfrak{p}}|}$$

Proof **:**
exercise

It is also enlightening to have a second perspective here: for subgroup $U, V \leqslant G$, we may consider the equivalence class $\sigma \sim \sigma' \iff \sigma' = u\sigma v$ giving us the double-coset[22]:

$$U\sigma V = \{u\sigma v \mid u \in U, v \in V\}$$

This partition will give us the similar information in the non-Galois case. Let $N/L/K$ be two separable extensions where N/K is Galois with Galois group G. Let $H = \mathrm{Gal}_L(N) \leqslant \mathrm{Gal}_K(N)$. Let

[22]warning: double cosets need not all have the same size!

$\mathfrak{p}$ be a prime ideal of K and $P_{\mathfrak{p}}$ the set of all prime ideal of L above $\mathfrak{p}$. Then if $\mathfrak{B}$ is a prime ideal of N above $\mathfrak{p}$, then we get a well-defined bijection:

$$H \setminus G/D_{\mathfrak{B}} \to P_{\mathfrak{p}} \qquad H\sigma D_{\mathfrak{B}} \mapsto \sigma\mathfrak{B} \cap L$$

Frobenius Automorphism

Assume $\mathfrak{B}$ over $\mathfrak{p}$ is unramified in $L/K/\mathbb{Q}$, so that $I_{\mathfrak{B}}$ is trivial (which by proposition 22.4.11 and proposition 22.4.14 can be shown to be the case for all put finitely many primes). Then $D_{\mathfrak{B}} \cong \mathrm{Gal}_{\kappa(\mathfrak{p})}(\kappa(\mathfrak{B})) \cong (\mathbb{Z}/n\mathbb{Z})$. By finite field theory, we know that there exists the *Frobenius generator* sending $x \mapsto x^{\mathfrak{N}(\mathfrak{B})}$ since $\mathfrak{N}(\mathfrak{B}) = f$ and the order of this group is f. We thus have a corresponding element in the decomposition group, and hence the Galois group:

Definition 22.4.43: Frobenius Automorphism of a Prime

Let $\sigma \in D_{\mathfrak{B}}$. Then σ is called the *Frobenius* if $\sigma(\alpha) \equiv \alpha^f \mod \mathfrak{B}$ where f is the inertia degree of $\mathfrak{B}$ over $\mathfrak{p}$. The element is sometimes denoted:

$$\mathrm{Frob}_{\mathfrak{B}/\mathfrak{p}}(\alpha) := \sigma(\alpha) \equiv \alpha^f \mod \mathfrak{B}$$

Note that if $\mathfrak{B}$ were ramified, then there would be more than one element with this property, namely the appropriate elements in coset $\sigma + I_{\mathfrak{B}}$. If another $\mathfrak{B}' = \sigma(\mathfrak{B})$[23] is considered, the Frobenius element differ conjugation:

$$\mathrm{Frob}_{\sigma(\mathfrak{B}/\mathfrak{p})} = \sigma\,\mathrm{Frob}_{\mathfrak{B}/\mathfrak{p}}\,\sigma^{-1}$$

Furthermore, if $D_{\mathfrak{B}}$ is abelian, then the Frobenius element is *only* determined by the prime $\mathfrak{p}$. In this case, we have:

$$\sigma(\alpha) = \alpha^{\mathfrak{N}(\mathfrak{p})} \mod \mathfrak{p}\mathcal{O}_L$$

because $\mathfrak{p}\mathcal{O}_L$ is the product of all primes $\mathfrak{B}$ over $\mathfrak{p}$. Another immediate observation is that if $\mathfrak{p}$ spits completely, then $f = 1$, and the Frobenius element for $\mathfrak{p}$ shall be the identity, and hence the Frobenius element captures behaviour in the case where a prime doesn't completely split.

Exercise 22.4.1

1. Let L/K be a galois extension of Number fields, $G = \mathrm{Gal}_K(L)$, $\mathfrak{B}$ lie over $\mathfrak{p}$.
   a) If $\mathfrak{B}$ is inert, then G Is cyclic
   b) If $\mathfrak{p}$ is totally ramified in each intermediary field but not totally ramified in L, then there does not exist non-trivial proper intermediary field, hence G is cyclic (of prime order)
   c) Simiarly, Suppose every intermediate field contains a unique prime lying over $\mathfrak{p}$, but L does not. Show the same as the previous
   d) Let $\mathfrak{p}$ be unramified in every proper non-trivial intermediate field, but ramified in L. Show that G has unique smallest subgroup H and that $H \trianglelefteq G$. Use this to show that $|G| = p^k$ for some k and prime p, that $|H| = p^l$ for some l, and $H \leqslant Z(G)$.
   e) Similarly, show the same for when $\mathfrak{p}$ splits compeltely in every proper non-trivial intermediate field, but not in L.

[23] by transitivity and definition of the decomposition group, any $\mathfrak{B}'$ is of this form

f) Let $\mathfrak{p}$ be inert in every proper non-trivial intermediate field but not in L. Prove that G is a cyclic where $|G| = p^k$.

2. (the non-normal case for Frobenius element by taking normal closure, p.78 Marcus)

Wild and Tame Ramification

(see this! here)

(This section is essentially here for the proof of Kronecker-Weber Theorem. May choose to move it down to that section).

We end off by exploring some more about the inertia group, namely we shall like at *wild* and *tame* inertia.

Definition 22.4.44: Wild Inertia Group

Let L/K be an extension of number fields, $I_{\mathfrak{B}} \leqslant D_{\mathfrak{B}}$ be the inertia group. Then $W_{\mathfrak{B}} \trianglelefteq I_{\mathfrak{B}}$ is defined to be:

$$W_{\mathfrak{B}} := \left\{ g \in I_{\mathfrak{B}} \;\middle|\; \forall \overline{\alpha} \in \mathcal{O}_L/\mathfrak{B}^2,\ g(\overline{\alpha}) = \overline{\alpha} \right\}$$

the wild inertia group can be thought of as a parallel to "local coordinates". In particular if we have an extension $K/\mathbb{Q}$ of number fields, then we have an extension $K/\mathbb{Q}_p$ then $W_{\mathfrak{B}}$ is the unique sylow p-subgroup of the inertia group. If you know some algebraic geometry, an intuitive reason to see why the quotient $\mathfrak{B}^2$ shows up is since the tangent space of a variety/scheme is given by $\mathfrak{m}/\mathfrak{m}^2$.

Lemma 22.4.45: Wild Inertia And Uniformizer

Let $\pi \in \mathcal{O}_L$ be the uniformizer for Q (i.e. $Q \| \pi$). Let $\sigma \in I_Q$. Then:

$$\sigma \in W_Q \qquad \Leftrightarrow \qquad \sigma(\pi) - \pi \in Q^2$$

that is, $\sigma(\pi) = \pi \bmod Q^2$

Proof **:**
The $\Rightarrow$ direction is by definition. For $\Leftarrow$, note that we have calculated before that:

$$\left|(\mathcal{O}_L/Q^2)^\times\right| = \left|(\mathcal{O}_L/Q)^\times\right| \left|Q\mathcal{O}_L/Q^2\right| = (q-1)q$$

where $q = |\mathbb{F}_{p^r}|$ for appropriate p^r. Hence, there exists $H \leqslant (\mathcal{O}_L/Q^2)^\times$ such that $|H| = q-1$ and $H \xrightarrow{\sim} \mathbb{F}_Q^\times$. Hence

$$(\mathcal{O}_L/Q^2)^\times \cong H \oplus (1 + \pi\mathcal{O}_L)_{Q^2} \bmod Q^2$$

Hence $\sigma \in I_G$, then σ acts trivially on $\mathbb{F}_Q^\times$, and hence acts trivially on H¿ Thus $\sigma \in W_Q$ if and only if σ fixes $(1 + \pi\mathcal{O}_L)_{Q^2}$ if and only if $(\pi\mathcal{O}_L)_{Q^2}$ if and only if σ fixes $\pi \bmod Q^2$

Example 22.4.6: Example from Class for wild inertia

$K = \mathbb{Q}(\sqrt[8]{1})$, $i^4 = 1$, $\left(\frac{1+i}{\sqrt{2}}\right)^2 = i$. $[K : \mathbb{Q}] = |\mathbb{Z}/8\mathbb{Z}^\times| = 4$. and $(\mathbb{Z}/8\mathbb{Z})^\times \cong (\mathbb{Z}/2\mathbb{Z}) \oplus \mathbb{Z}/2\mathbb{Z}$. So K is biquadratic. Note that

$$K = \mathbb{Q}\left(\frac{1+i}{\sqrt{2}}\right) = \mathbb{Q}(\alpha) = \mathbb{Q}(i, \sqrt{2})$$

(diagram of the 3 quadratic extensions). Notice that 2 is the only prime that ramifies. In $\mathbb{Q}(\sqrt{d})$ where $d = -1, 2, 2$, so 2 ramifies

$$(e, f, g) = (2, 1, 1)$$

It may not be immediately clear, but in K, $2\mathcal{O}_K = \mathfrak{B}^4$, i.e. 2 totally ramifies (we have not seen cases where e increases when it's coprime). This cannot happen with unramified fields. What is $\mathfrak{B}$? Explicitly, we know that

$$\mathfrak{B} = (1 - \alpha)$$

which we can directly see because $\mathfrak{B}^2 = (1 - 2\alpha + \alpha^2) = (1 + i - 2\alpha) = (1 + i)(1 - (1 - i)\alpha)$, and we see that $(1 - (1 - i)\alpha)$ is a unit.

Let $\pi = 1 - \alpha$, which is a uniformizer of $\mathfrak{B}$. Recall that we looked at $\sigma(\pi) - \pi$ for $\sigma] in \operatorname{Gal}_{\mathbb{Q}}(K)$. Computing it's valuation at $\mathfrak{B}$

$$v_{\mathfrak{B}}(\sigma(\pi) - \pi)$$

we can do this by independence of valuation on the prime. First

$$\sigma(\alpha) = \alpha^3 \implies (1 - \alpha^3) - (1 - \alpha) = v_{\mathfrak{B}}(\alpha^2 - 1) = 1 + v_{\mathfrak{B}}((\alpha - 1) + 2) = 2$$

Next, we do

$$\sigma(\alpha) = \alpha^7 = \alpha^{-1} \implies v_{\mathfrak{B}}(\alpha^7 - \alpha) = v_{\mathfrak{B}}(\alpha^{-1} - \alpha) = v_{\mathfrak{B}}(\alpha^2 - 1) = 2$$

Finally, we have

$$\sigma(\alpha) = \alpha^5 \implies v_{\mathfrak{B}}(\alpha^5 - \alpha) = v_{\mathfrak{B}}(\alpha^4 - 1) = v_{\mathfrak{B}}(\alpha - 1) + v_{\mathfrak{B}}(\alpha + 1) + v_{\mathfrak{B}}(\alpha^2 + 1) = 1 + 1 + 2 = 4$$

Hence, we have a chain:

$$\langle 1 \rangle \subseteq \langle 1, 5 \rangle \subseteq (\mathbb{Z}/\gamma\mathbb{Z})^\times = \operatorname{Gal}_{\mathbb{Q}}(K)$$

It may not jump out, but notice that 2 divides the difference (linked to a theorem called *Hasse-Arf* Theorem).

Theorem 22.4.46: Abelian Extension Wild Map

Let L/K Galois and abelian, $G = \operatorname{Gal}_K(L)$, $P \subseteq \mathcal{O}_K$ a prime ideal of $\mathcal{O}_K$, Q a prime lying over P W_Q the wild inertia group, and π a uniformizer so that $Q \| \pi$. Then:

$$\varphi : I_Q/W_Q \hookrightarrow \mathbb{F}_Q^\times \qquad [\sigma] \mapsto \frac{\sigma(\pi)}{\pi} \bmod Q$$

Assume G is abelian. Then $\operatorname{im} \varphi \subseteq \mathbb{F}_P^\times$.

Proof **:**
Look at $\varphi(\sigma) = \varphi(\tau^{-1}\sigma\tau)$ (because abelian), so:

$$\frac{\tau^{-1}\sigma\tau(\pi)}{\pi} \bmod Q$$

and write it as:

$$\frac{\tau{-}1(\sigma(\tau(\pi)))}{\tau(\pi)} \bmod Q$$

Now, it doesn't matter what uniformizer we pick. If $\tau \in D_Q$, then we can mod by Q before or after, so

$$\tau^{-1}\varphi(\sigma)$$

What kind of element can be written as τ^{-1}? We know the decomposition group surjects onto $\mathrm{Gal}_{\mathbb{F}_p}(\mathbb{F}_Q)$, we can pick τ^{-1} to be the Frobenius operator, that is

$$\varphi(\sigma) = \varphi(\sigma)^{\mathrm{Nm}(P)} \implies \varphi(\sigma) \in \mathbb{F}_p$$

Theorem 22.4.47: Wild Inertia Group Is Sylow Subgroup

Take $W_Q \leqslant \mathrm{Gal}_{\mathbb{Q}}(L)$ where Q is a prime lying over p. Then W_Q the unique p-sylow subgroup, hence $W_Q \trianglelefteq I_Q$.

Proof **:**
check notes for now

22.4.4 Example: Cyclotomic Fields

As we saw in example 22.1.8, the discriminant of $\mathbb{Z}[\zeta_{p^n}]$ is a prime power of p. We now generalize this result for arbitrary n, and find how p, and will find the discriminant is a power of p. Thus, p is the only prime that ramifies in K, and we shall find how p ramifies:

Lemma 22.4.48: Cyclotomic Fields Lemma

Let $n = p^k$ be a prime power, ζ_{p^k} a root of unity, and $\lambda = 1 - \zeta_{p^k}$. Then the principal ideal $(\lambda) \subseteq \mathcal{O}$ in the ring of integers of $\mathbb{Q}(\zeta_{p^k})$ is a prime ideal of degree 1, and:

$$p\mathcal{O} = (\lambda)^d$$

where $d = \varphi(n) = [\mathbb{Q}(\zeta_{p^k}) : \mathbb{Q}]$. Furthermore, the basis $1, \zeta_{p^k}, ..., \zeta_{p^k}^{d-1}$ of $\mathbb{Q}(\zeta_{p^k})/\mathbb{Q}$ has discriminant:

$$d_{\mathbb{Q}(\zeta_{p^k})/\mathbb{Q}}(1, \zeta_{p^k}, ..., \zeta_{p^k}^{d-1}) = \pm p^s$$

where $s = p^{k-1}(kp - k - 1)$

Note that in order for the above discriminant to be the discriminant of the ring of integers, we would have to insure the ring of integers has the above basis up to change of coordinates, which shall be

the result of the next theorem.

***Proof* :**
Recall that the minimal polynomial of ζ_{p^k} over $\mathbb{Q}$ is:

$$m_{\zeta_{p^k},\mathbb{Q}}(x) = \frac{x^{l^n}-1}{x^{l^{n-1}}-1} = x^{l^{n-1}(l-1)} + \cdots + x^{l^{n-1}} + 1$$

Letting $x = 1$, we get

$$n = p^k = \prod_{r\in(\mathbb{Z}/n\mathbb{Z})^\times} (1-\zeta_{p^k}^r)$$

Next, we know that

$$(1-\zeta_{p^k}^r) = (1-\zeta_{p^k})(1+\zeta_{p^k}+\cdots+\zeta_{p^k}^{r-1}) = (1-\zeta_{p^k})\epsilon_r$$

If r' is an integer such that $rr' = 1 \bmod n$, then

$$\frac{1-\zeta_{p^k}}{1-\zeta_{p^k}^r} = \frac{1-(\zeta_{p^k}^r)^{r'}}{1-\zeta_{p^k}^r} = 1+\zeta_{p^k}^r+\cdots+(\zeta_{p^k}^r)^{r'-1}$$

is integral as well, namely ϵ_r is a unit. Thus,

$$p = \epsilon(1-\zeta_{p^k})^{\varphi(n)}$$

with unit $\epsilon = \prod_r \epsilon_r$. Thus:

$$p\mathcal{O} = (\lambda)^{\varphi(n)}$$

Since $[\mathbb{Q}(\zeta_{p^k}):\mathbb{Q}] = \varphi(n)$, the fundamental identity ($\sum e_i f_i$) shows that (λ) is a prime ideal of degree 1.

Now, take $\zeta_{n,1},...,\zeta_{n,d}$ be the conjugates of ζ_{p^k} (i.e. powers of ζ_{p^k} that are coprime with the order of ζ_{p^k}). Then the cyclotomic polynomial:

$$\Phi_{p^k}(x) = \prod_{i=1}^{d}(x-\zeta_{n,i})$$

and

$$\pm\operatorname{disc}(1,\zeta_{p^k},...,\zeta_{p^k}^{d-1}) = \prod_{i\neq j}(\zeta_{n,i}-\zeta_{n,j}) = \prod_{i=1}^{d}\Phi'_{p^k}(\zeta_{n,i}) = \operatorname{Nm}_{\mathbb{Q}(\zeta_{p^k})/\mathbb{Q}}(\Phi'_{p^k}(\zeta_{p^k}))$$

Differentiating the equation:

$$(x^{p^{k-1}}-1)\Phi_{p^k}(x) = x^{p^k}-1$$

and substituting ζ_{p^k} for x gives:

$$(\zeta_{p^k}^{p^{k-1}}-1)\Phi'_{p^k}(\zeta_{p^k}) = (\eta-1)\Phi'_{p^k}(\zeta_{p^k}) = n\zeta_{p^k}^{-1}$$

But we know $\operatorname{Nm}_{\mathbb{Q}(\eta)/\mathbb{Q}}(\epsilon-1) = \pm p$, thus

$$\operatorname{Nm}_{\mathbb{Q}(\zeta_{p^k})/\mathbb{Q}}(\eta-1) = \operatorname{Nm}_{Q(\eta)/\mathbb{Q}}(\eta-1)^{p^{k-1}} = \pm p^{p^{k-1}}$$

since $\zeta_{p^k}^{-1}$ has norm ± 1, we get:

$$d(1,\zeta_{p^k},..,\zeta_{p^k}^{d-1}) = \pm\operatorname{Nm}_{\mathbb{Q}(\zeta_{p^k})/\mathbb{Q}}\left(\Phi'_{p^k}(\zeta_{p^k})\right) = \pm p^{k^{p^{k-1}(k-1)-p^{k-1}}} = \pm p^s$$

where s is the exponent

For completeness of this section, here is a quick proof of the ring of integers of cyclotomic fields:

Proposition 22.4.49: Ring Of Integers Of Cyclotomic Fields

Let $K = \mathbb{Q}(\zeta_n)$ be a cyclotomic field. Then

$$\mathcal{O}_{K_n} \cong \mathbb{Z}[\zeta_n]$$

With discriminant:

$$\operatorname{disc}(\mathcal{O}_{K_n}) = (-1)^{\varphi(n)/2} \frac{n^{\varphi(n)}}{\prod_{p|n} p^{\frac{\varphi(n)}{p-1}}}$$

***Proof* :**
Let's start with n being a prime power p^k. Since $\operatorname{disc}(1, \zeta, ..., \zeta^{d-1}) = \pm p^s$, we have:

$$p^s \mathcal{O}_K \subseteq \mathbb{Z}[\zeta_n] \subseteq \mathcal{O}$$

Putting $\lambda = 1 - \zeta_n$, by lemma 22.4.48 we have $\mathcal{O}_K/\lambda\mathcal{O}_K \cong \mathbb{Z}/p\mathbb{Z}$, thus $\mathcal{O}_K = \mathbb{Z} + \lambda\mathcal{O}_K$, hence we *a fortiori* get

$$\lambda\mathcal{O}_K + \mathbb{Z}[\zeta] = \mathcal{O}$$

Multiplying by λ, and substituting $\lambda\mathcal{O}_K = \lambda^2\mathcal{O}_K + \lambda\mathbb{Z}[\zeta]$, we get

$$\lambda^2\mathcal{O}_K + \mathbb{Z}[\zeta] = \mathcal{O}$$

Iterating this procedure, we get:

$$\lambda^t\mathcal{O}_K + \mathbb{Z}[\zeta] = \mathcal{O} \qquad \forall t \geq 1$$

When $t = s\varphi(p^k)$, then given $p\mathcal{O}_K = \lambda^{\varphi(p^k)}\mathcal{O}_K$, we get

$$\mathcal{O} = \lambda^t\mathcal{O}_K + \mathbb{Z}[\zeta] = p^s\mathcal{O}_K + \mathbb{Z}[\zeta] = \mathbb{Z}[\zeta]$$

For the general case, let $n = p_1^{k_1} \cdots p_r^{k_r}$. Then $\zeta_i = \zeta^{n/p_i^{k_i}}$ is a primitive $p_i^{k_k}$-th root of unity, and hence:

$$Q(\zeta) = \mathbb{Q}(\zeta_1) \cdots \mathbb{Q}(\zeta_r)$$

and $\mathbb{Q}(\zeta_1) \cdots \mathbb{Q}(\zeta_{i-1}) \cap \mathbb{Q}(\zeta_i) = \mathbb{Q}$. Then for each $i \in \{1, ..., r\}$, the elements $1, \zeta_i, ..., \zeta_i^{d_i-1}$, where $d_i = \varphi(p_i^{k_i})$ form an integral basis of $\mathbb{Q}(\zeta_i)/\mathbb{Q}$. Since the discriminants $d(1, \zeta_i, ..., \zeta_i^{d_i-1}) = \pm p_i^{s_i}$, are relatively prime, we get from successive uses of (Proposition 2.11 in Neukirch, I should add it to my notes!!) we get that

$$\zeta_1^{j_1} \cdots \zeta_r^{j_r}$$

where $j_i \in \{0, ..., (d_i - 1)$ form an integral basis of $\mathbb{Q}(\zeta)/\mathbb{Q}$. But then each of these elements is a power of ζ. Thus, every $\alpha \in \mathcal{O}$ may be written as a polynomial $\alpha = f(\zeta)$ with coefficients in $\mathbb{Z}$. Since ζ has degree $\varphi(n)$ over $\mathbb{Q}$, the degree of $f(\zeta)$ may be reduced to $\varphi(n) - 1$. But then, we may obtains:

$$\alpha = a_1 + a_1\zeta + \cdots + a_{\varphi(n)-1}\zeta^{\varphi(n)-1}$$

Thus, $1, \zeta, ..., \zeta^{\varphi(n)-1}$ is an integral basis, as we sought to show.

by **??**, we know that the primes that ramify are those that divide $\operatorname{disc}(\mathcal{O}_{K_n})$, hence by the above the primes that ramify are the factors of n. The following gives exactly how they factor:

Proposition 22.4.50: Prime Decomposition In Cyclotomic Field

Enumerates all the primes $p_1, p_2, ...$, then let $n = \prod_n p_n^{n_k}$ be a prime factorization of $n \in \mathbb{Z}$. Let f_k be the smallest positive integer such that

$$p_k^{f_k} \equiv 1 \bmod (n/p_k^{n_k})$$

Then in $\mathbb{Q}(\zeta_n)$, we have:

$$p_k = (\mathfrak{p}_1 \cdots \mathfrak{p}_r)^{\varphi(p_k^{n_k})}$$

where $\mathfrak{p}_1, ..., \mathfrak{p}_r$ are distinct prime ideals all of degree f_k.

Note that if $p_k \nmid n$, then $n_k = 0$, so the condition becomes

$$p_k^{f_k} \equiv 1 \bmod n$$

and $\varphi(p_k^{n_k}) = 1$, giving us:

$$p_k = \mathfrak{p}_1 \cdots \mathfrak{p}_r$$

***Proof* :**
Since $\mathcal{O}_K = \mathbb{Z}[\zeta]$, we have that the conductor of $\mathbb{Z}[\zeta]$ is 1, and hence we may apply theorem 22.4.6 to any prime number p, in particular every p decompose into prime ideals in the exact same way that $\Phi_n(x)$ (the minimal polynomial of ζ_n) factors into irreducible mod p. Thus, we are reduced to showing that:

$$\Phi_n(x) \equiv (p_1(x) \cdots p_r(x))^{\varphi(p^{\nu_p})} \bmod p$$

where $p_1(x), p_2(x), ..., p_r(x)$ are distinct irreducible polynomials over $\mathbb{Z}/p\mathbb{Z}$ of degree f_p. Write n as $p^{\nu_p} m$ ($p \nmid m$). Let ξ_i and η_j vary over the primitive roots of unity over order m and p^{ν_p} respectively. Then the product $\xi_i \eta_j$ vary over the nth root of unity, namely we have the decomposition:

$$\Phi_n(x) = \prod_{i,j} (x - \xi_i \eta_j)$$

Since $x^{p^{\nu_p}} - 1 \equiv (x-1)^{p^{\nu_p}} \bmod p$, we have that $\eta_j \equiv 1 \bmod \mathfrak{p}$ for any prime ideal where $\mathfrak{p} \big| p$. This, we get:

$$\Phi_n(x) \equiv \prod_i (x - \xi_i)^{\varphi(p^{\nu_p})} = \Phi_n(x)^{\varphi(p^{\nu_p})} \bmod \mathfrak{p}$$

Hence, we have when modding by p:

$$\Phi_n(x) \equiv \Phi_n(x)^{\varphi(p^{\nu_p})} \bmod p$$

Now, notice that f_p is the smallest positive integer such that $p^{f_p} \equiv 1 \bmod m$, hence the congruence reduces to th case where $p \nmid n$, and then $\varphi(p^{\nu_p}) = \varphi(1) = 1$.

Now, since the characteristic of $\mathcal{O}_K/\mathfrak{p}$ (i.e. p) does not divide n, the polynomial $x^n - 1$ and nx^{n-1} have no common roots in $\mathcal{O}_K/\mathfrak{p}$ by the derivative test. Thus, $x^n - 1 \bmod p$ has *no multiple roots*. Thus, we get that $\mathcal{O}_K \to \mathcal{O}_K/\mathfrak{p}$ maps the group of nth roots of unity μ_n bijectivity onto the nth roots of unity of $\mathcal{O}_K/\mathfrak{p}$. In particular ,the primitive nth roots of unity modulo $\mathfrak{p}$ ($\zeta \bmod \mathfrak{p}$) remain

primitive nth roots of unity. Since the extension field of $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ containing it is the field $\mathbb{F}_{p^{f_p}}$, since the multiplicative group $\mathbb{F}^*_{p^{f_p}}$ is cyclic of order $p^{f_p} - 1$. Thus, $\mathbb{F}_{p^{f_p}}$ is the field decomposition of the reduced cyclotomic polynomial:

$$\overline{\Phi}_n(x) = \Phi_n(x) \bmod p$$

Since it is a divisor of $X^n - 1 \bmod p$, this polynomial has no multiple roots, and if

$$\overline{\Phi}_n(x) = \overline{p}_1(x) \cdots \overline{p}_r(x)$$

i it's factorization into irreducible over $\mathbb{F}_p$, then every $\overline{p}_i(x)$ is the minimal polynomial of a primitive nth root of unity $\overline{\xi}\ in \mathbb{F}^*_{p^{f_p}}$. It's degree is thus f_p, as we sought to show.

Corollary 22.4.51: Special Case Of Factorization

A prime number p is ramified in $\mathbb{Q}(\zeta_n)$ iff

$$n \equiv 0 \bmod p$$

except in the case where $p = 2 = (4, n)$. A prime number $p \neq 2$ is totally split in $\mathbb{Q}(\zeta)$ iff

$$p \equiv 1 \bmod n$$

Proof **:**
direct consequence of last proposition

Corollary 22.4.52: Subfield Of Cyclotomic Field

Let $K \subseteq \mathbb{Q}(\zeta_n) = L$ with corresponding subgroup $H \leqslant \mathrm{Gal}_K(L) \cong (\mathbb{Z}/n\mathbb{Z})^\times$ considered as a subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$. For prime p where $p \nmid n$, let f be the least positive integer such that $\overline{p^f} \in H$. Then for prime $\mathfrak{p}$ lying above p in K:

$$f = f_{\mathfrak{p}/p}$$

Proof **:**
Notice hat $f_{\mathfrak{p}/p}$ is the order of the Frobenius automorphism.

Overall, the ring of integers of a cyclotomic field is monogenic, the only primes that ramify are those that divide n (where n is the order of the root of unity) since the divisors of the discriminant are completely dependent on n, the discriminant is a product of the powers of primes of n, and we know exactly how the each element $n \in \mathbb{Z}$ factors in cyclotomic fields!

22.5 Application: Kummer Theory and Kronecker Weber

We now apply our finding to Field Theory and find some important consequences' on *Abelian Extension*. This shall also lead to proving one of the central theorems in field theory, The *Kronecker-Weber Theorem*, which states that all finite abelian extensions show up as sub-extensions of Cyclotomic extensions over $\mathbb{Q}$! Since Cyclotomic extensions are generated by roots of unity, every algebraic integers whose Galois group is abelian can be written as a $\mathbb{Q}$-linear combination of roots of unity! For example:

$$\sqrt{5} = e^{2\pi i/5} - e^{4\pi i/5} - e^{6\pi i/5} + e^{8\pi i/5}$$

namely, all square roots of integers may be written as $\mathbb{Q}$-linear combinations of roots of unity!

22.5.1 Kummer Theory

As we saw, we may form abelian extensions from $\mathbb{Q}$ by taking cyclotomic extensions, since their Galois group is $(\mathbb{Z}/n\mathbb{Z})^\times$. Naturally, any sub-extension of a cyclotomic extension is also a Galois group. It turns out that the converse of this last statement is true, and it is goal of this section to show this.

We first give the setting in which we work in:

Definition 22.5.1: Universal Cyclotomic Field and Abelian Closure

1. Let $K^{\mathbf{cycl}} = \bigcup_{n\in\mathbb{N}} K(\zeta_n)$. If $K = \mathbb{Q}$, then:

$$\mathrm{Gal}(\mathbb{Q}^{\mathbf{cycl}}) = \varprojlim \mathrm{Gal}(\mathbb{Q}(\zeta_n)) = \varprojlim (\mathbb{Z}/n\mathbb{Z})^\times = \hat{\mathbb{Z}}^\times$$

   is the pro-finite group given by the limit of units of cyclic groups.

2. Let K be a field. Then

$$K^{\mathbf{ab}} \subseteq \overline{K}$$

   is the maximal extension for which $\mathrm{Gal}_K(K^{\mathbf{ab}})$ is abelian.

Proposition 22.5.2: Properties of Abelian Extensions

1. If $L_1, L_2 \subseteq K^{\mathrm{ab}}$ have finite abelian groups, then $\mathrm{Gal}_K(L_1L_2)$ is a finite abelian group.
2. If $G_K = \mathrm{Gal}_K(\overline{K})$ and $H = [G, G]$ (the commutator [closed] subgroup), then

$$K^{\mathrm{ab}} = \overline{K}^H$$

   with Galois group G_K/H.

Proof **:**

1. See proposition 18.1.20
2. This comes from the universal property of abelian quotients, see proposition 5.2.2

Note that in general, $K^{\mathbf{cycl}} \subsetneq K^{\text{ab}}$:

Example 22.5.1: Abelian, But Not Cyclotomic

Let $k = \mathbb{Q}(i)$ and $F = \mathbb{Q}(\sqrt{1+2i})/\mathbb{Q}(i)$. Since every quadratic extension is abelian, F is an abelian extension. However, it is not a cyclotomic extension: for if it was a sub-cyclotomic extension, then since $\mathbb{Q}(i)$ is a cyclotomic extension over $\mathbb{Q}$, then $\mathbb{Q}(\sqrt{1+2i})$ would be a cyclotomic extension, but that is not possible since $\mathbb{Q}(\sqrt{1+2i})/\mathbb{Q}$ is not Galois (but all sub-extensions of a cyclotomic extension are necessarily Galois).

It is thus quite special that if $K = \mathbb{Q}$, we have the following:

Theorem 22.5.3: Kronecker Weber's Theorem

Let $K = \mathbb{Q}$. Then:

$$\mathbb{Q}^{\mathbf{cycl}} = \mathbb{Q}^{\text{ab}}$$

In other words, every abelian extension of $\mathbb{Q}$ is contained in a cyclotomic extension.

We shall delay the proof till page 1005. We first require some more Kummer Theory. Kummer's theory is essentially the study of extensions that contain "enough" roots of unity. As we shall see, this is a rather powerful condition, and is hence given a name:

Definition 22.5.4: Kummer Extensions and G-Extensions

Let L/K be a field extension. Then L/K is called a *Kummer Extension* if:

1. for some n, K contains μ_n (the roots of $x^n - 1$)
2. L/K is an Abelian extension with exponent n[a]

If in particular we want extensions with Galois groups G, we shall call such extensions G-extensions.

[a] The least common multiple of the orders of the elements

Example 22.5.2: Kummer Extensions

1. Any quadratic extension of $\mathbb{Q}$ is a Kummer extension of degree 2, since $\mathbb{Q}$ have $1, -1$ and has abelian extension. More generally, any multi-quadratic extension is a Kummer extension of degree 2
2. There is no Kummer extensions of degree 3 over $\mathbb{Q}$ since $\mathbb{Q}$ does not have a 3rd root of unity, since it's complex.
3. More generally if K contains μ_n (meaning the characteristic of K does not divide n), then adjoining any nth root of $a \in K$, $L = K(\sqrt[n]{a})$ is a Kummer extension (of degree m dividing n, since the minimal polynomial divides $x^n - a$)

Theorem 22.5.5: Hilbert's 90th Theorem

Let K be a field of characteristic 0 and L/K be an abelian extension such that

$$\mathrm{Gal}_K(L) \cong \mathbb{Z}/n\mathbb{Z}$$

If σ generate $\mathrm{Gal}_K(L)$, then:

1. For $\beta \in L^\times$ such that $\mathrm{Nm}_{L/K}(\beta) = 1$, there exists an $\alpha \in L^\times$ such that

$$\beta = \frac{\sigma(\alpha)}{\alpha}$$

2. Let $\beta \in L$ such that $\mathrm{tr}(\beta) = 0$. Then there exists an $\alpha \in L$ such that

$$\beta = \alpha - \sigma(\alpha)$$

Proof **:**
Since L/K is a finite separable extension, pick t so that $L = K(t)$. Let

$$S = \beta + \beta\sigma(\beta) + \beta\sigma(\beta)\sigma^2(\beta) + \cdots = \sum_{i=1}^{n-1} \prod_{j}^{i} \sigma^j(\beta)$$

If $S \neq 0$, then:

$$\frac{\sigma(S)}{S} = \beta^{-1} \implies \frac{\sigma(S^{-1})}{S^{-1}} = \beta$$

Thus, we would need to guarantee such a nonzero S exists.

To see this, notice that solving $\beta = \frac{\alpha}{\sigma(\alpha)}$ is equivalent to $\beta\sigma(\alpha) = \alpha$. Notice that

$$\beta\sigma(-) : L \to L$$

mapping $\alpha \mapsto \beta\sigma(\alpha)$ is a K-linear map. If we show that it has an eigenvector with eigenvalue 1, then we solved our problem, for if α is such an eigenvector then (if we represent our function as f)

$$\alpha = f(\alpha) = \beta\sigma(\alpha)$$

We shall verify this by extending scalars from K to L. Recall by corollary 14.3.6 that extending scalars does not change the eigenvectors. We may thus extend our map to:

$$\begin{aligned} \mathrm{id} \otimes_K \beta\sigma(-) : L \otimes_K L &\to L \otimes_K L \\ \ell \otimes_K \ell' &\mapsto \ell \otimes_K \beta\sigma(\ell') \end{aligned}$$

Since L/K is finite and separable, there exists a γ such that $L = K(\gamma)$. Then γ has minimal polynomial $m_{\gamma,K}(x) \in K[x]$, and we have the isomorphisms:

$$(L \otimes_K L) \cong \left(L \otimes_K \frac{K[x]}{m_{\gamma,K}(x)}\right) \cong \frac{L[x]}{m_{\gamma,K}(x)} \cong L^n$$

Since any $\ell' \in L$ can be written as $\ell' = p(\gamma)$ for some polynomial $p(x) \in K[x]$, we have:

$$(\ell \otimes_K p(\gamma)) \mapsto \ell(p(\gamma), p(\sigma(\gamma)), ..., p(\sigma^{n-1}(\gamma)))$$

Thus, the map $\beta\sigma(-)$ becomes interpreted as $L^n \to L^n$ as:

$$\ell(p(\gamma), ..., p(\sigma^{n-1}(\gamma))) \mapsto \ell\left(\beta p(\sigma^{n-1}(\gamma)), \sigma(\beta p(\gamma)), ..., \sigma^{n-1}(\beta p(\sigma^{n-2}(\gamma)))\right)$$

or perhaps a bit more clearly:

$$(\ell_1, \ell_2, ..., \ell_n) \mapsto (\beta\ell_n, \sigma(\beta\ell_1), ..., \sigma^{n-1}(\beta\ell_{n-1}))$$

Since $\mathrm{Nm}_{L/k}(\beta) = 1$,

$$\left(1, \sigma(\beta), \sigma(\beta)\sigma^2(\beta), ..., \sigma(\beta)\sigma^2(\beta)\cdots\sigma^{n-1}(\beta)\right)$$

is an eigenvector with eigenvalue 1, hence our original transformation had an eigenvector of eigenvalue 1, as we sought to show.

For the trace argument, say $\beta \in L$ such that $\mathrm{tr}(\beta) = 0$. Take

$$S = \beta 2\sigma(\beta) + 3\sigma^2(\beta) + \cdots n\sigma^{n+1}(\beta)$$

Then:

$$\sigma(S) = \sigma(\beta) + 2\sigma^2(\beta) + \cdots + n\sigma^n(\beta)$$

since $\sigma^n = \mathrm{id}$, $\sigma^n(\beta) = \beta$, so we get:

$$S - \sigma(S) = -n\beta$$

Thus, set $\alpha = -\beta/n$, completing the proof.

60: Foreshadowing

In part VII, we shall reformulate this theorem in purely Homological terms, namely it says that $H^1(G, L^*) = 0$

We now look at some properties of Kummer extensions. First, by some Galois theory, we see that if $\mathbb{Q}(\zeta_n) \subseteq K$, $L = K(\alpha)$ where $\alpha^n = \beta \in K$, then L/K is an abelian Galois extension, for:

$$\varphi : \mathrm{Gal}_K(L) \xrightarrow{\sim} \mu_n \qquad \tau \mapsto \frac{\tau(\alpha)}{\alpha}$$

which is well-defined, and any nth root of β gives the same isomorphism. Note how important $\mu_n \subseteq K$ is for this isomorphism to work! This is certainly not true in general (is $\mathbb{Q}(\sqrt[n]{2})/\mathbb{Q}$ galois for $n > 2$?). Using Hilbert's 90th problem, the converse of this statement is also true:

Theorem 22.5.6: Primitive Element For Cyclic Kummer Extension

Let $\mathrm{char}(K) = 0$, L/K be a Galois extension and $\mathrm{Gal}_K(L) \cong \mathbb{Z}/n\mathbb{Z}$. Assume K contains μ_n, the group of roots of unity generated by ζ_n. Then there exists an $\alpha \in L^\times$ such that

$$L = K(\alpha) \qquad \alpha^n \in K^\times$$

Proof **:**
Let ζ_n be a primitive nth root of unity. Recall that $\mathrm{Nm}_{L/K}(\zeta_n) = \zeta_n^n = 1$. Thus, by Hilbert's 90th, there exists $\alpha \in L$ such that $\sigma(\alpha) = \zeta_n \alpha$. Since σ is a generator, we see that α is not fixed by any σ, hence:

$$L = K(\alpha)$$

Furthermore,

$$\sigma(\alpha^n) = \sigma(\alpha)^n = \zeta_n^n \alpha^n = \alpha^n$$

Hence, each σ fixes α^n, meaning $\alpha^n = \beta \in K$, hence

$$L = k(\sqrt[n]{\beta})$$

as we sought to show.

Corollary 22.5.7: Cyclic Kummer Extension Correspondence

Let $\mu_n \subseteq K$. Then we have the following bijective correspondence

$$\left\{ \begin{matrix} L/K \\ \mathrm{Gal}_K(L) \cong \mathbb{Z}/n\mathbb{Z} \end{matrix} \right\} \underset{\psi}{\overset{\varphi}{\rightleftarrows}} \left\{ \begin{matrix} G \subseteq K^\times/(K^\times)^n \\ G \cong \mathbb{Z}/n\mathbb{Z} \end{matrix} \right\}$$

Proof **:**
First finding ψ, pick $\overline{\beta} \in K^\times/(K^\times)^n$. Pick $\beta \in K^\times$ which represents $\overline{\beta}$. Then define

$$\psi(\langle \overline{\beta} \rangle) = K(\sqrt[n]{\beta}) = L$$

this is well-defined for if we have any other coset representative, $\beta' = \beta\gamma^n$, $K(\sqrt[n]{\beta}) = K(\sqrt[n]{\beta'})$. Certainly $\mathrm{Gal}_K(L) \cong \mathbb{Z}/m\mathbb{Z}$ and $m \big| n$. We need to show that $n = m$. For the sake of contradiction, suppose not. Then if σ generates $\mathrm{Gal}_K(L)$, Then if we let $\alpha = \sqrt[n]{\beta}$, we have

$$\sigma(\alpha) = \zeta_m \alpha \implies \sigma(\alpha^m) = \zeta_m^m \alpha^m = \alpha^m$$

hence $\alpha^m \in K^\times$. Then $\beta = \alpha^n = (\alpha^m)^{\frac{n}{m}} \in (K^\times)^{n/m}$, a contradiction.

Conversely, to define φ, suppose we have an L where $\langle \sigma \rangle = \mathrm{Gal}_K(L)$. By theorem 22.5.6, pick $\alpha \in L^\times$ such that $\sigma(\alpha)/\alpha = \zeta_n$. Then α is well-defined up to being an element of $K^\times$. In particular, $[\alpha^n] \in K^\times/(K^\times)^n$ is well-defined. Thus, take the map:

$$\varphi(L) = \langle [\alpha^n] \rangle \subseteq K^\times/(K^\times)^n$$

It remains to show these maps are inverses of each other (important exercise!)

Corollary 22.5.8: Pontryagin Dual For Kummer Extensions

Let $\mathbb{Q}(\zeta_n) \subseteq K$, and $K_n \subseteq \overline{K}$ the maximal finite Kummer extension, that is the maximal finite abelian extension with $\operatorname{ord}(\operatorname{Gal}_K(K_n))\big|n$ (every element of the group has order that divides n). Then there exists a natural isomorphism:

$$\operatorname{Gal}_K(K_n) \cong \operatorname{Hom}(K^\times/(K^\times)^n, \mu_n)$$

$$\sigma \mapsto \left(\overline{\beta} \mapsto \frac{\sigma(\sqrt[n]{\beta})}{\sqrt[n]{\beta}}\right)$$

Proof **:**
exercise

22.5.2 Kronecker-Weber Theorem

We know have enough to prove Kronecker-Weber. Recall we want to show that $\mathbb{Q}^{\mathrm{ab}} = \mathbb{Q}^{\mathbf{cycl}}$, that is if $K/\mathbb{Q}$ is abelian where K is a number field, then $K \subseteq \mathbb{Q}^{\mathbf{cycl}}$. We first present some important lemmas to reduce the proof. We'll first show we can assume $\operatorname{Gal}_{\mathbb{Q}}(K)$ is a p-group for some prime p. For if $\operatorname{Gal}_{\mathbb{Q}}(K)$ is abelian, we have

$$\operatorname{Gal}_{\mathbb{Q}}(K) \cong \bigoplus_p G_p$$

where G_p is a p-group.

Proof. **Reduction #1**:
Define $K_{p'} = K^{(\bigoplus_{p \neq p'} G_p)}$. Then $K = K_2 K_3 \cdots$ is the join of all these fields. Hence, if all such fields is inside a Cyclotomic extension, so is K. This can naturally be extended to reduce to cyclic groups in the same manner. □

Next, we show that p is the *only* prime which ramifies in K

Proof. **Reduction #2**:
Assume not. Assume ℓ ramifies in K as well. Let I be the inertia group (doesn't matter what prime we pick). Then the wield inertia group W is trivial (by group theory). Then by theorem 22.4.46:

$$I \hookrightarrow \mathbb{F}_\ell^\times \qquad I \cong \mathbb{Z}/\ell\mathbb{Z}, \quad e\big|\ell - 1$$

Let $L \in \mathbb{Q}(\mu_\ell)$ be the (unique) extension such that $[L : \mathbb{Q}] = e$. So ℓ totally ramifies in L. Then take:

$$\begin{tikzcd}
 & M = KL & \\
L \arrow[ur] & & K \arrow[ul] \\
 & \mathbb{Q} \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

K and L are abelian, thus so is M. Let I_M be the inertia group of a prime over ℓ. Then

$$I_M \hookrightarrow \mathbb{Z}/e\mathbb{Z} \oplus \mathbb{Z}/\ell\mathbb{Z}$$

Hence

$$I_M \cong \mathbb{Z}/e\mathbb{Z}$$

Let $K' = M^{I_M}$. Then $K' \cap L = \mathbb{Q}$ since nothing ramifies in $K' \cap L$.

Now, $K'L/K'$ has degree at least e, in particular$\geq e$, since ℓ has to ramify (ℓ doesn't ramify in K', and ramifies in L by e). But then,

$$K'L = M = KL$$

Note that we have not introduced any new ramifications, since no other primes ramify that ramify in L, K. We can afterward inductively remove (if it ramifies, it has to have some inertia group, but it will be trivial in the intermediate extensions).

With this, let's now prove the $p = 2$ case.

Assume $K/\mathbb{Q}$ is abelian $[K : \mathbb{Q}] = 2^n$ and 2 is the only prime which ramifies. I claim that K is cyclotomic.

Without loss of generality, let's assume $\mathbb{Q}(i) \subseteq K$ (just adjoin it if need be). Then $K_0 = K \cap \mathbb{R}$ (fixed field under conjugation). This is still abelian since K is abelian, and $[K : K_0] = 2$, so $K = K_0(i)$. Now, let $K_n/\mathbb{Q}$ be the real cyclotomic field where $\mathrm{Gal}(K_n/\mathbb{Q}) \cong \mathbb{Z}/2^n\mathbb{Z}$ (done it in comments). Consider now

$$M = KK_n$$

Then $\mathrm{Gal}_{\mathbb{Q}}(M) \subseteq \mathbb{Z}/2^n\mathbb{Z} \oplus \mathbb{Z}/2^n\mathbb{Z}$. If $K_n \neq K$, then this Galois group is *not* cyclic (too large order). Therefore, M has ≥ 3 real quadratic subfields (real, since everything is real). But there is only 2 with ramified primes – a contradiction ($\mathbb{Q}(\sqrt{2})$ is the only real quadratic extension in which 2 ramifies). □

Now, for the other primes, we have the following key proposition:

Proposition 22.5.9: Ramification Of Primes in $\mathbb{Z}/n\mathbb{Z}$-Extensions

Let p be an odd prime. Then there exists a unique field $K/\mathbb{Q}$ which is Galois with Galois group $\mathbb{Z}/p\mathbb{Z}$ in which p is the only ramified prime.

Proof **:**

Let ζ be a pth primitive root of 1, let $L = \mathbb{Q}(\zeta)$. $\pi = 1 - \zeta$. So that $(\pi)^{p-1} = (p)$. Let $\gamma \in \mathcal{O}_L$ be a non pth power. Assume $\gamma \equiv 1 \bmod \pi^p$. Take the field $M = L(\sqrt[p]{\gamma})/L$. Then π is unramified.

Suppose not. Then $\pi\mathcal{O}_M = Q^p$ for some prime Q where $\mathbb{F}_Q \cong \mathbb{F}_{(\pi)} \cong \mathbb{F}_p$. Let δ be such that $\delta^p = \gamma$. Let $r = \frac{\delta - 1}{\pi}$. Claim that $r \in \mathcal{O}_M$. Take the conjugates of r over L, i.e. they are

$$\frac{\zeta^i\delta - 1}{\pi} \qquad i \in \mathbb{Z}/p\mathbb{Z}$$

So the minimal polynomial is the product:

$$\prod\left(x - \frac{\zeta^i\delta - 1}{\pi}\right) = \frac{1}{\pi^p}\prod_{i=1}^{p-1}((\pi x - 1) - \zeta^i\delta) = \frac{(\pi x - 1)^p - \gamma}{\pi^p}$$

So this is inside $\mathcal{O}_L[x]$, so r is inside of integers. Thus, there exists some $m \in \mathbb{Z}$ such that $r - m \in \mathbb{Q}$ (because $\mathbb{F}_Q$ is just $\mathbb{F}_p$). But then for all $\sigma \in \mathrm{Gal}_K(M)$, $\sigma(r-m) \in \mathbb{Q}$ since it's the inerta group. But

$$\frac{\zeta\gamma - 1}{\pi} - m \in \mathbb{Q}$$

and so is $\frac{\gamma-1}{\pi} \in \mathbb{Q}$. Subtracting we get

$$\frac{1-\gamma}{\pi}\gamma = \frac{\pi}{\pi}\gamma = \gamma \in \mathbb{Q}$$

But this is a contradiction of our $\gamma \equiv 1 \bmod Q$, as we sought to show.

With this, we can continue the second reduction

Proof. **Reduction #2 cont.**:

By reduction # 2, we only need to work with $K/\mathbb{Q}$ abelian of prime power order p^n, and without loss of generality $K/\mathbb{Q}$ may be assumed to be cyclic since every field is a compositum which have cyclic subgroups.

Now consider $K_n \subseteq \mathbb{Q}(\zeta_{n+1})$ be the field such that

$$[K_n : \mathbb{Q}] = p^n \qquad \mathrm{Gal}(K_n/\mathbb{Q}) \cong \mathbb{Z}/p^n\mathbb{Z}$$

Let $M + KK_n$. In general $\mathrm{Gal}(M/\mathbb{Q}) \hookrightarrow \mathbb{Z}/p^n\mathbb{Z} \oplus \mathbb{Z}/p^n\mathbb{Z}$. If $K \neq K_n$, then like for the $p = 2$ case, this Galois group is not cyclic. Thus, M contains at least $p+1$ subfields which have Galois group $\mathbb{Z}/p\mathbb{Z}$ over $\mathbb{Q}$ in which only p ramifies, a contradiction $\square$

Proof **:**

Of Kronecker-Weber's Theorem

We shall show that if p is an odd prime, $K/\mathbb{Q}$ is a $\mathbb{Z}/p\mathbb{Z}$ extension, and p is the only prime that ramified in K, then $K \subseteq \mathbb{Q}(\zeta_{p^2})$, which by our earlier reduction shall complete the proof.

Let $\mathrm{Gal}_{\mathbb{Q}}(M) = \langle \sigma \mid \sigma^{p(p-1)} = \mathrm{id} \rangle$, $K = M^{\langle \sigma^p \rangle}$ and $L = M^{\langle \sigma^{p-1} \rangle}$. Consider:

$$\begin{tikzcd}
 & M = KL & \\
L = \mathbb{Q}(\zeta_p) \arrow[ur] & \mathbb{Z}/p(p-1)\mathbb{Z} \arrow[u] & K \arrow[ul] \\
 & \mathbb{Q} \arrow[ul, "\mathbb{Z}/(p-1)\mathbb{Z}"] \arrow[u] \arrow[ur, "\mathbb{Z}/p\mathbb{Z}"'] &
\end{tikzcd}$$

We hat that p is the only prime that ramifies in M. Thus, by Kummer theory, we know that $M = L(\alpha)$ where $\alpha^p = \beta \in L^\times$. Next, we know that

$$\frac{\sigma(\alpha)}{\alpha}$$

is a non-trivial rot of unity. Without loss of generality say that $\frac{\sigma^{p-1}(\alpha)}{\alpha} = \zeta_p$ (so $\langle \sigma^{p-1} \rangle = \mathrm{Gal}_L(M)$). Note that $\sigma(\zeta_p) = \zeta_p^n$, where $\overline{n} \in (\mathbb{Z}/p\mathbb{Z})^\times$ (i.e. is a primitive root).

I claim that

$$M = L\left(\frac{\sigma(\alpha)}{\alpha}\right)$$

For suppose not. Then $\frac{\sigma(a)}{\alpha} \in L$, and

$$\frac{\sigma(\alpha)}{\alpha} = \sigma^{p-1}\left(\frac{\sigma(\alpha)}{\alpha}\right) = \frac{\sigma^p(\alpha)}{\sigma^{p-1}(\alpha)} = \frac{\sigma(\zeta_p\alpha)}{\zeta_p\alpha} \implies 1 = \frac{\sigma(\zeta_p)}{\zeta_p}$$

but that is a contradiction, hence $M = L\left(\frac{\sigma(\alpha)}{\alpha}\right)$.

For the rest of this proof, we write

$$\alpha \mapsto \frac{\sigma(\alpha)}{\alpha} = \alpha' \qquad \beta \mapsto \frac{\sigma(\beta)}{\beta} = \beta'$$

Let P_k, P_L, P_M be the primes ideals above $p\mathbb{Z}$. Take the prime decomposition:

$$\beta\mathcal{O}_L = \prod Q^{n_Q} = P_L^r \cdot \prod_{Q \neq P} Q^{n_Q} \qquad n_Q \in \mathbb{Z}$$

Then:

$$\sigma(\beta)\mathcal{O}_L = P_L^r \prod_{Q \neq P_L} \sigma(Q)^{n_Q}$$
$$\frac{\sigma(\beta)\mathcal{O}_L}{\beta} = \prod_{Q \notin P_L} Q^{m_Q}$$

Hence, P_L does not occur in $\beta\mathcal{O}_L$. Take $\beta'' = \beta' N^P$ and $\alpha'' = \alpha' N$ for $N \in \mathbb{Z}$ such that $\beta'' \in \mathcal{O}_L$ Then:

$$M = L(\alpha'') \qquad (\alpha'')^P = \beta''$$

Now let $\pi = 1 - \zeta_p \in \mathcal{O}_L$ so that $P_L = (\pi)$. Then $\mathbb{F}_{P_L} \cong \mathbb{F}_P$ implies $\beta'' \bmod P_L = m \bmod P_L$ for $m \leq N$. Let $m_0 = m^{-1} \bmod p$, so $\beta'' \to m_0\beta''$ and $\alpha'' \to m_0\alpha''$. Without loss of generality say $\beta'' = 1 \bmod P_L$. Then for some $c \in \mathbb{Z}/p\mathbb{Z}$

$$\beta'' = 1 + c\pi \bmod P_L^2 \qquad \zeta_p = 1 - \pi \bmod P_L^2 \qquad \zeta_p^r = 1 - r\pi \bmod P_L^2 \implies \beta'' = \zeta_p^{-c}\nu$$

where $\nu \equiv 1 \bmod P_L^2$. I claim that $\nu \in (L^\times)^P$, which will imply $M = \mathbb{Q}(\zeta_{p^2})$ and complete the

proof. First:

$$\begin{aligned}
(\alpha'')^p &= \beta'' = \zeta_p^{-c}\nu \\
\sigma^{p-1}(\alpha'') &= \alpha''\zeta_p \\
\implies \sigma(\zeta_p) &= \zeta_p^n && n \in (\mathbb{Z}/p\mathbb{Z})^\times \\
\implies \sigma^p(\alpha'') &= \sigma(\sigma^{p-1}(\alpha'')) = \sigma(\alpha''\zeta_p) = \sigma(\alpha'')\cdot\zeta_p^n \\
\implies \sigma^{p-1}((\alpha'')^n) &= (\alpha'')^n\zeta_p^n \\
\implies \sigma^{p-1}\left(\frac{\sigma(\alpha'')}{(\alpha'')^n}\right) &= \frac{\sigma(\alpha'')}{(\alpha'')^n} \\
\implies \frac{\sigma(\alpha'')}{(\alpha'')^n} &\in M^{\langle\sigma^{p-1}\rangle} = L \\
\implies \frac{\sigma(\beta'')}{(\beta'')^n} &\in (L^\times)^P \\
\implies \frac{\sigma(\nu\zeta_p^{-c})}{\nu^n\zeta_p^{-nc}} &\in (L^\times)^p
\end{aligned}$$

Thus $\frac{\sigma(\nu)}{\nu^n} \in (L^\times)^p$.

Next, we claim that $\nu = 1 \bmod P_L^p$. Suppose not. Then let $p^M \| \nu - 1$ where $1 \leq m \leq p-1$, and $\nu = 1 + c\pi^m \bmod P_L^{m+1}$ (where $c \in (\mathbb{Z}/p\mathbb{Z})^\times$). Then by this:

$$\nu^n = 1 + nc\pi^m \bmod P_L^{m+1}$$

hence,

$$\sigma(\pi) = \sigma(1-\zeta_p) = 1 - \zeta_p^n = (1-\zeta_p)(1+\zeta_p+\cdots+\zeta_p^{n-1}) = \pi n \bmod P_L^2$$

implying

$$\sigma(\pi^m) = \pi^m n^m \bmod P_L^{m+1} \implies \sigma(\nu) = 1 + cn^m\pi^m \bmod P_L^{m+1}$$

Note that $n^m \not\equiv n \bmod p$ for $n \in (\mathbb{Z}/p\mathbb{Z})^\times$. Now as a corollary:

$$\frac{\sigma(\nu)}{\nu^m} \equiv 1 + d\pi^m \bmod P_L^{m+1} \qquad d \in \mathbb{Z}/p\mathbb{Z}^\times$$

However,

$$\frac{\sigma(\nu)}{n\nu} = S^p \qquad S \in L$$

where P_L does not occur in S, and:

$$S^p \equiv 1 \bmod P_L \implies S = 1 \bmod P_L$$

Giving us $S = 1 + t$, $P_L \big| t$, so

$$S^P = 1 + pt + \cdots + t^p \equiv 1 \bmod P_L^P$$

a contradiction!

Finally, $L(\sqrt[p]{\nu})/L$ is unramified above P_L. Let

$$F = L(\sqrt[p]{\nu}),\ F \subseteq M(\zeta_{p^2}) \qquad F/Q \text{ abelian}$$

and $I \leqslant \mathrm{Gal}_{\mathbb{Q}}(\mathbb{F})$ where I is the inertia group above P, and $|I| = p-1$, so $F^I/\mathbb{Q}$ is unramified everywhere, namely $F^I = \mathbb{Q}$, so $F = L$. This implies

$$\nu \in (L^\times)^P$$

But then

$$M = \mathbb{Q}(\zeta_{p^2})$$

as we sought to show.

Let us use it in an actual example:

Example 22.5.3: Using Kronecker-Weber

Let $K/\mathbb{Q}$ be a galois extension with Galois group $\mathbb{Z}/5\mathbb{Z}$ whose discriminant is only divisible by $5, 11$, and 31. How many such number fields are there?

Proof. By Kronecker Weber, we know that K is an intermediary field of a Cyclotomic extension:

$$K \subseteq \mathbb{Q}(\zeta_n)$$

We may without loss of generailty assume n is minnimal for which this containment makes sense. Since only $5, 11, 31$, divide the discriminant, we know the discriminant of K is of the form:

$$\mathrm{disc}(\mathcal{O}_K) = 5^a 11^b 31^c$$

which also mean only $5, 11, 31$ ramify in K. Now, I claim that only $5, 11, 31$ divide n and no other number those. For the sake of contradiction, let's say $p \big| n$ but $p \notin \{5, 11, 31\}$. By minimality, we have:

$$\begin{tikzcd}
 & \mathbb{Q}(\zeta_n) \\
\mathbb{Q}(\zeta_{n/p}) \arrow[ur] & K \arrow[u] \\
 & \mathbb{Q} \arrow[u] \arrow[ul]
\end{tikzcd}$$

Let $I_{\mathfrak{B}}$ be the inertia group (for any prime $\mathfrak{B}$ lying above p). Since $p \nmid \mathrm{disc}(\mathcal{O}_K)$, p is unramified. If p is unramified in $\mathbb{Q}(\zeta_{n/p})$, then since $\mathbb{Q}(\zeta_{n/p}) \subseteq \mathbb{Q}(\zeta_n)$, we get that $\mathbb{Q}(\zeta_{n/p}) \subseteq \mathbb{Q}(\zeta_n)^{I_p} \subseteq \mathbb{Q}(\zeta_n)$

$$K \subseteq \mathbb{Q}(\zeta_n)^{I_{\mathfrak{B}}} \subseteq \mathbb{Q}(\zeta_{n/p})$$

contradicting minimality of n. Hence, we get $n = 5^a 11^b 31^c$. Thus:

$$\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(\zeta_n)) \cong (\mathbb{Z}/n\mathbb{Z})^\times \cong \mathbb{Z}/4\mathbb{Z} \times (\mathbb{Z}/5^{a-1}\mathbb{Z}) \times \mathbb{Z}/10\mathbb{Z} \times (\mathbb{Z}/11^{b-1}\mathbb{Z}) \times \mathbb{Z}/30\mathbb{Z} \times (\mathbb{Z}/31^{c-1}\mathbb{Z})$$

which now become a counting argument of how many subgroups of index 5 there are, counting the number of intermediary subfields, thus completing the proof. $\square$

22.6 *Modules Over Dedekind Domains

(Useful for alg-geo, theory itself is also really nice. Also $\mathcal{O}_K$-modules appear in the next section)

Lemma 22.6.1: all Ideal co-maximal up to Module Isomorphism

Let $I, J \subseteq R$ be fractional ideals of a Dedekind domain R. Then $I \oplus J \cong_R R$

Proof **:**
By corollary 22.2.11, we may choose in fact choose I, J to be integral ideals. We shall show that $I + j^{-1}J = R$ for some appropriate j^{-1}, which since $j^{-1}J \cong_R J$, we have $I \oplus J \cong_R R$

In particular, we shall show that for any ideal J, there exists an element $j \in J$ such that $I + jJ^{-1} = R$ (that is, every ideal is "almost" comaximal with any fractional ideal). It is equivalent to show that $IJ = (j) = J$. By corollary 22.2.17, IJ is generated by at most 2 elements, say $IJ = (a, b)$. Then since $a \in IJ \subseteq J$, we may choose a c such that $(a, c) = J$ (see proof of corollary 22.2.17 if you're not convinced we can choose such a c). Then since $b \in J$ (since $b \in IJ$)

$$IJ + (c) = (a, b) + (c) = (a, b, c) = (a, c) = J$$

hence, $I + jJ^{-1} = R$, which by corollary 22.2.11 we get $I + J \cong_R R$, completing the proof.

Proposition 22.6.2: Ideals Are Projective

Let R be a Dedekind domain and $I \subseteq R$ an ideal. Then I is a projective R-module

Proof **:**
By lemma 22.6.1, we may choose any J such that $I + jJ^{-1} = R$, or equivalently $IJ + (j) = J$. We shall define an isomorphism $I \oplus J \to IJ \oplus R$ via

$$(x, y) \mapsto (iy - jx, x + ky)$$

This map is certainly well-defined since $iy, jx \in IJ$, so $iy - jx \in IJ$, and it is certainly R-linear. It is bijective, since it's inverse is $IJ \oplus R \to I \oplus J$

$$(x, y) \mapsto (jx + y, ix - ky)$$

Hence, we have $I \oplus J \cong IJ \oplus R$. Finally, by corollary 22.2.16, choose J so that $IJ = (a)$ is principal. Then it is a free R-module of rank 1, and so

$$I \oplus J \cong IJ \oplus R \oplus R^2$$

as we sought to show.

Proposition 22.6.3: IBN equivalent for Dedekind Domains

Let $I_1, ..., I_n$ and $J_1, ..., J_m$ be fractional ideals of a Dedekind domain R. Then

$$I_1 \oplus I_2 \oplus \cdots \oplus I_n \cong_r J_1 \oplus J_2 \oplus \cdots \oplus J_m$$

if and only if $n = m$ and

$$I_1 I_2 \cdots I_n = (a) J_1 J_2 \cdots J_n$$

that is, the product differs by a principal ideal (they belong to the same equivalence class in the class group). In particular,

$$I_1 \oplus I_2 \oplus \cdots \oplus I_n \cong_R \underbrace{R \oplus R \oplus \cdots \oplus R}_{n-1 \text{ times}} \oplus (I_1 I_2 \cdots I_n)$$

and $R^n \oplus I \cong R^n \oplus J$ if and only if $I = (a)J$ for some $a \in K(R)$.

Proof **:**
First, let $I, J \subseteq R$ be two fractional ideals. By lemma 22.6.1, we may choose I, J to be integral ideals that are up to R-isomorphism relatively prime: $I + J = R$. Then $I \cap J = IJ$, and $\varphi : I \oplus J \to R$ given by $(x, y) \mapsto x + y$ has kernel $I \cap J = IJ$. Then:

$$0 \to IJ \to I \oplus J \to R \to 0$$

is a short exact sequence. Since R is a free R-module, it is projective, and hence the sequence splits, giving us

$$I \oplus J = IJ \oplus R$$

Theorem 22.6.4: Classifying Torsion-Free Modules Over Dedekind Domain

Let M be a torsion-free R-module over a Dedekind domain R of rank n. Then

$$M \cong R^{n-1} \oplus I$$

where we I may be replaced with any ideal J such that $I = (a)J$ (they belong in the same element in the class group). As a consequence, any torsion-free module over a Dedekind domain is projective.

Proof **:**
If there exists a non-trivial map $\varphi : M \to R$, then since R-submodules of R are projective, we may iteratively split M as $M \cong N \oplus I$ to get a direct sum of ideals of R, which we may then use to conclude by using proposition 22.6.3.

Since M is torsion-free of rank n, we may extend scalars we get a n-dimensional $K = K(R)$ vector space $K \otimes_R M$ where $M \hookrightarrow K \otimes_R M$ naturally embeds injectively. Let $\beta_1, \beta_2, ..., \beta_n$ be a basis for $K \otimes_R M$ and $m_1, m_2, ..., m_k$ be a set of generators for M. Then

$$m_i = \sum_j (a_{ij}/b_{ij})\beta_j$$

Let $d = \prod_{ij} b_{ij}$ and $y_j = \beta_j/d$. Then

$$M = Ry_1 + \cdots + Ry_n \subseteq K\beta_1 + \cdots + K\beta_n$$

Hence, M is a R-submodule of a free R-module! Importantly, for each $m \in M$, it can be written as

$$m = \sum a_i y_i \qquad a_i \in R$$

hence, we may define $\varphi : M \to R$ by $\varphi(m) = a_n$. Then $\varphi(M) \subseteq R$ is an ideal of R, and so

$$M = \ker(M) \oplus I$$

Notice that $\mathrm{Tor}(\ker(M)) = 0$, and so we may inductively repeat this procedure until we have products of prime ideals!

since the product of projective module is projective, any torsion-free module over a Dedekind domain is projective, as we sought to show.

22.7 *Geometric Idea: 1-Dimensional Schemes

See Neukirch, and put orders here too for their geometric interpretation

22.7.1 Localization of Dedekind Domains

For the motivation to CFT

22.7.2 Orders: Curves with Singularities

This entire section is completely inspired by Neukirch p.72

All smooth 1-dimensional curves correspond to Dedekind domains. However, there are some curves with singularities. These curves are associated to sub-rings of Dedekind domains called *orders*

Definition 22.7.1: Order

Let $K/\mathbb{Q}$ be a number field of degree n. Then an order of K is a subring of $\mathcal{O}_K$ which contains an integral basis of length n. The ring $\mathcal{O}_K$ is called a *maximal order*.

By the above definition, an order is of the form:

$$\mathbb{Z}[\alpha_1, ..., \alpha_n]$$

where $K = \mathbb{Q}(\alpha_1, ..., \alpha_n)$, which motivates why $\mathcal{O}_K$ is the maximal order; it is the maximal ring with this property. This also implies that orders need not be normal, hence an order is a one-dimensional Noetherian integral domain (prime remain maximal as if $a \in \mathfrak{p} \cap \mathbb{Z}$ then $a\mathcal{O} \subseteq \mathfrak{p} \subseteq \mathcal{O}$ that is $\mathfrak{p}$ and $\mathcal{O}$ have the same rank n, and thus $\mathcal{O}/\mathfrak{p}$ is a finite integral domain, showing $\mathfrak{p}$ is maximal) . An example of an order is $\mathbb{Z} = \mathbb{Z}(\sqrt{5})$. Many order's naturally are as "rings of multipliers. For example if $\alpha_1, .., \alpha_n$ is a basis for $K/\mathbb{Q}$ and $M = \sum_i \mathbb{Z}\alpha_i$, then

$$\mathcal{O}_M = \{\alpha \in K \mid \alpha M \subseteq M\}$$

is an order.

It is important that we re-prove the Chinese remainder

Here

As orders are not Dedekind domain in general, *not* all ideals are invertible. This also means that the collection of fractional ideals does not form a group.

Proposition 22.7.2: Invertible Fracitonal Ideals

Let I be a fractional ideal of a Noetherian domain R. Then I is invertible iff its localization at every maximal ideal of R is invertible, equivalently, iff its localization at every prime ideal of R is invertible.

Proof **:**
MIT notes p.17

Example 22.7.1: non-invertible ideal of an order
put this example here

(I want this build up here)

Definition 22.7.3: Picard Group Of An Order

Let $J(\mathcal{O})$ be the collection of invertible ideals and $P(\mathcal{O})$ the fractional principal ideals. Then

$$\mathrm{Pic}(\mathcal{O}) = \frac{J(\mathcal{O})}{P(\mathcal{O})}$$

A lot more to be said.

The Picard group ignore s ideals that are not invertible. Another approach that gives some useful results is to define the following for A Dedekind domain:

$$\mathrm{Div}(\mathcal{O}) = \mathcal{O}_{\mathfrak{p}}\mathbb{Z}\mathfrak{p}$$

Then $\mathrm{Div}(\mathcal{O})$ is isomorphic to the ideal group. For an order, the divisor group shall be a quotient of this group. We would then need to define the divisor elements, which gives the divisor class group/chow group.

Part VI

Representation Theory

There are many domains of mathematics which have made huge amounts of progress, having very-well understood objects and morphisms, theorems, results, and properties. One of the most well-understood area's of matheamtics has become linear algebra, where vector-spaces and linear transformations. One of the goals of this chapter will be do be able to study the following 3 objects where the operaions becomes matrix addition and multiplicatoin

1. (Finite) groups as a subgroup of $\mathrm{GL}(V)$, where the group operation becomes the matrix multiplication
2. Associative algebas
3. Lie Algebras

The furhter advantage of bringing these objects into linear algebra is that many field intersect with it: If the vector-spaces are infinite dimensional, we can make them Hilbert spaces and use tools from analysis. Vector-spaces are at the heart of differential geometry allowing us to use such theories to study groups! There have even been strides to try to connect number theory[24]

(The reader may want to go back to section 12.3 to review about direct sums of modules)

(Also, this entire link is really cool: from wiki. Also has some cool examples of duality!!!)

[24]see Langland's program

23

Semisimple Rings and Modules

As mentioned in the introduction, representation theory is all about decomposing structures. In part I, we showed in chapter 3, section 3.7, that all finite groups can uniquely be represented as a series of extensions with simple factors (i.e. G_i/G_{i+1} is simple). The simplest type of composition series would be when G is a direct product of simple groups, which would be in the terminology of exact sequences a series of trivial extensions with simple factors: such a group is sometimes *semi-simple* group[1]. When it came to modules, we found that we may decompose it in different ways depending on what rings acts on it, in particular we saw in chapter 13 that if $R = k$ is a field k, then any k-module is free, and in chapter 14 we saw that if R is a Principal Ideal Domain, where we got that any R-module decompsoed into a free part R-isomorphic to R^k, and a torsion part $R/(a_1) \times \cdots \times R/(a_n)$. It is evident that the ring that acts on the abelian group can have a large impact on it's structure. In this section, we define a new type of ring, which will come to be called a *semi-simple* ring, that will force the module M to *always* be a direct product of *simple modules* (so that we are analysing the "building blocks" of modules).

To define this ring, we first shift focus and start by work with the module itself to find the "simplest" structure it may be[2], and explore in a manner which at the beginning parallels groups: we will look for *simple R-modules*, and direct product of simple modules (since they are the simplest extension of simple module).[3]. After defining such a strucutre, we will define which rings acting on moduels will *always* make the module semi-simple

[1] the term "semi-simple group" is not a completely standard terminology, as direct product of simple groups is not generally very heavily studied, see this

[2] "simple" here means no non-trivial sub-structures that can be used to build an extension, just like for simple groups or simple rings

[3] It might be tempting to say that this means we are exploring projective modules. However, we have yet to establish whether simple modules are free-modules. Are they?

23.1 Decomposing Modules

Definition 23.1.1: Decomposition Terminology

Let M be a nonzero R-module.

1. M is said to be *simple* (or *irreducible*) if the only submodules of M are 0 and M. Otherwise, M is *reducible*
2. M is said to be *indecomposable* if there do not exist submodules $M_1, M_2 \subseteq M$ such that $M = M_1 \oplus M_2$. Otherwise, M is said to be *decomposable*.
3. The module M is said to be *semisimple* (or *completely reducible*) if it is the direct summand of simple modules
4. The module M is said to be semi-indecomposable (or completely indecomposable) if it is a direct summand of indecomposable modules.
5. If M is semisimple, any direct summand of M is called a *constituent* of M (i.e. N is a constituent of M if there exists a submodulde N' such that $M = N \oplus N'$)

A moment should be taken to find the relations between these definitions:

1. It is clear that simple modules are semi-simple and indecomposable. The prototypical example of a simple module is a field k over itself.
2. Semi-Simple $\not\Rightarrow$ Simple: take k^2 over k (the 2 dimensional vector-space over k). The cannonical example of a semi-simple module whould be vector-spaces, since they are all direct sums of the underlying field, which is a simple k-module.
3. Indecomposable $\not\Rightarrow$ Simple: this statement is equivalent to saying that not all short exact sequence spilt (i.e., not all modules are *projective*, see section 16.2). For a concrete example, let $\mathbb{Z}^2$ be a module over $\mathbb{Z}$ with the action:

   $$r \cdot (a, b) = (a + rb, b)$$

   Then it should be checked that it is not simple since $\mathbb{Z} \times \{0\}$ is a submodule, but it is indecomposable[4].
4. Every semi-simple module is a semi-indecomposable module, however the converse is not true: $\mathbb{Z}$ is a semi-indecomposable $\mathbb{Z}$-module (more genreally any module]oTj/Regular Func

   semi-indecomposable due to the Decomposition Theorem from chapter 14), however $\mathbb{Z}$ is not semi-simple a semi-simple $\mathbb{Z}$-module as you should try to show (this will be immediate after lemma 23.1.7)[5]

As stated in the introduction, we want to determine a semi-simple module or semi-simple modules in terms of the ring that is acting on it. The following defintion was made for this purpose:

[4] Hint: Is it possible to write $\mathbb{Z} = n\mathbb{Z} \oplus m\mathbb{Z}$?

[5] A mention of Krull-Schmidt's theorem for the decomposition of Noetherian or Artinian modules into the direct sum of indecomposable

Definition 23.1.2: Semi-Simple Ring

Let R be a ring. Then R is *semi-simple* if every R-module is semi-simple. Recall that R is *simple* if the only two-sided ideals are (0) and R.

Notice how a semi-simple ring was *not* defined as the direct product of simple rings. In the following proofs, we will show that the definition of semi-simple ring implies that R is the direct product of simple rings, however the converse is not true: there exists simple rings that are *not* semi-simple modules over themselves (and hence not simple R-modules). The problem is that of dimension; we will soon explore the relation between being semi-simple and Noetherian/Artinian. If we impose the condition that R be Noetherian and Artinian, then we are able to say that R is a semi-simple ring if and only if R is the direct product of simple rings.

We next establish many propertie of simple and semi-simple modules, and show how closely related they are to the ring acting on them:

Lemma 23.1.3: Simple Module Characterization

Let M be an R-module. Then M is a simple R-module if and only if there exists a two-sided maximal ideal $I \subsetneq R$ such that

$$M \cong_R R/I$$

Proof **:**

($\Rightarrow$) Let M be a simple [non-trivial] R-module so that the only sub-modules of M are $\{0\}$ and M. Consider any $0 \neq x \in M$ and the R-module homomorphism:

$$\varphi : R \to M \qquad r \mapsto r \cdot x$$

Since $\varphi(R) \subseteq M$ is a sub-module and M is a non-trivial R-module, it must be that $\varphi(R) = M$. Hence:

$$R/\ker(\varphi) \cong M$$

But then, $R/\ker(\varphi)$ cannot have any submodules, since M doesn't, and so $\ker(\varphi)$ must be maximal.

($\Leftarrow$) Let $R/I \cong M$ as R-module where I is maximal. If $0 \subsetneq N \subsetneq M$, then there would exist an ideal $\subsetneq J \subsetneq R/I$, but that would make $I \subseteq J$, contradicting maximality of I.

If R is commutative, then R/I is a field. For the non-commutative case, R/I need not be a field, not even a division ring (see the remark ref:HERE or go over section 9.4.3). Do not confuse the isomorphism in the above statement as a ring isomorphism: the structure of M is isomorphic to the R-module structure of left-multiplication on R/I. Notice how this also shows that a simple R-module M is cyclic since $R \cdot \varphi(1) = \varphi(R \cdot 1) = \varphi(R/I) = M$[6]. If $R = k$ is a field, then M is isomorphic to a 1-dimensional vector-space over k.

[6] In fact, any element must generate it, can you see why?

Using this result, we see that any semi-simple R-module is of the form:

$$M \cong_R \bigoplus_{i \in I} R/I_i$$

for maximal ideals $I_i \subseteq R$. In fact, we can further characterize I_i:

Corollary 23.1.4: Simple Module's And Annihilators

Let M be a simple left (or right) R-module. Then $M \cong R/\operatorname{Ann}(m)$ for any m, and $\operatorname{Ann}(m)$ is a maximal ideal.

Proof **:**
Since M is a simple left R-module, for any $m \in M$, $M \cong_R Rm$. Then we have a surjective R-module homomorphism

$$R \to Rm \qquad r \mapsto r \cdot m$$

Thus, $R/I \cong_R Rm$ for some ideal I (note that the R-submodules of R are ideals). In particular, $\varphi(I) = Im = 0$, so $I = \operatorname{Ann}(m)$ by definition. Thus, we have the isomorphism chian:

$$M \cong_R Rm \cong_R R/\operatorname{Ann}(m)$$

Finally, since Rm is simple, $R/\operatorname{Ann}(m)$ must have no other two-sided ideal,s i.e. $\operatorname{Ann}(m)$ is a two-sided maximal ideal, completing the proof.

Since we've seen the limitations of the structure of simple modules, it is also natural to ask if there are any limitations on the structure of the hom-set given two simple modules:

Lemma 23.1.5: Schur's Lemma

Let M and N be simple R-modules, $\varphi : M \to N$ an R-module homomorphism. Then either $\varphi = 0$ or φ is an isomorphism. As a consequence, $\operatorname{End}_R(M)$ is a division ring.

Proof **:**
Since N is simple, either $\varphi(M) = N$ or $\varphi(M) = 0$. If $\varphi(M) = N$, it has to be that $\ker(\varphi) = 0$, since M is simple, and hence φ would be an isomorphism.

As a consequence, every element in $\operatorname{End}_R(M)$ is either 0 or is a unit, i.e. has a multiplicative inverse, which by definition makes $\operatorname{End}_R(M)$ a divising ring under $(+, \circ)$.

Since $M \cong_R R/I$ for some maximal ideal $I \subsetneq R$, we have that $\operatorname{End}_R(R/I)$ is always a division ring, whether R is commutative or not. In a sense, this can be thought of as "recovering" a similar result to proposition 9.4.11, where it was shown that R/I is a field if R is commutative.

Notice that if $M = V$ was an 1-dimensional vector-space over k, then $\operatorname{End}_R(V)$ is a known division ring, in particular it's ring-isomorphic to k If V is n-dimensional, then $\operatorname{End}_R(V)$ is ring-isomorphic to $M_n(k)$. We might hope for something similar for the structure of $\operatorname{End}_R(M)$ is M is semi-simple. Take a moment to ponder this before reading the next lemma (hint: recall proposition 12.3.22)

Lemma 23.1.6: Hom's of Semisimple Modules

Let M be a semisimple R-module, so that:

$$M = S_1^{\oplus n_1} \oplus S_2^{\oplus n_2} \oplus \cdots \oplus S_k^{\oplus n_k}$$

where each S_i are simple R-modules, $S_i \ncong S_j$. Then as a ring:

$$\mathrm{End}_R(M) \cong_{\mathbf{Ring}} \mathrm{End}_R(S_1^{\oplus n_1}) \oplus \cdots \oplus \mathrm{End}_R(S_k^{\oplus n_k})$$

Furthermore, for each $S_i^{\oplus n_i}$, we have that direct sums "pull out" to become matrices:

$$\mathrm{End}_R(S_i^{\oplus n_i}) \cong_{\mathbf{Ring}} M_{n_i}(\mathrm{End}_R(S)) = M_{n_i}(D_i)$$

where $D = \mathrm{End}_R(S_i)$ and $M_{n_i}(D_i)$ is the matrix ring over the division ring $\mathrm{End}_R(S_i)$.

***Proof* :**
We invoke proposition 12.3.22 to get the result in terms of groups, and then use Schur's lemma and check that the ring operation is compatible:

$$\begin{aligned}
\mathrm{End}_R(M) &= \mathrm{Hom}_R\left(\bigoplus_i S_i^{\oplus n_i}, \bigoplus_j S_j^{\oplus n_j}\right) \\
&\cong_{\mathbf{Grp}} \bigoplus_{i,j} \mathrm{Hom}_R\left(S_i^{\oplus n_i}, S_j^{\oplus n_j}\right) \\
&\cong_{\mathbf{Grp}} \bigoplus_{i,j} \mathrm{Hom}_R\left(S_i, S_j\right)^{\oplus n_i n_j}
\end{aligned}$$

where when $S_i \neq S_j$, by Shur's lemma we must have:

$$\mathrm{Hom}_R(S_i, S_j)^{n_i n_j} = 0$$

Hence we get that $\mathrm{End}_R(M) \cong_{\mathbf{Grp}} \oplus_i \mathrm{Hom}_R(S_i^{n_i}, S_i^{n_i}) = \oplus_i \mathrm{End}_R(S_i^{n_i})$. Next, this map is compatible with the ring structure, since the map:

$$\varphi \to (\varphi|_{S^{n_i}} : S_i^{n_i} \to S_i^{n_i} \subseteq M)$$

is clearly clearly compatible with composition, and hence

$$\mathrm{End}_R(M) \cong_{\mathbf{Ring}} \bigoplus_i \mathrm{End}_R(S_i^{n_i})$$

For part 2, we follow the same steps as for part (1): re-indexing so that $S^n = S_1 \oplus S_2 \oplus \cdots \oplus S_n$ so that $S_i = S_j = S$, we get

$$\mathrm{Hom}_R\left(\bigoplus_i S_i, \bigoplus_j S_j\right) \cong_{\mathbf{Grp}} \bigoplus_{i,j} \mathrm{Hom}_R(S_i, S_j) \overset{!}{=} \bigoplus_{i,j} D \cong_{\mathbf{Grp}} M_n(D)$$

where $\overset{!}{=}$ comes from Shur's lemma with $D = \mathrm{End}_R(S)$ since $S_i = S_j = S$. The last equality comes from the fact that the matrix $M_n(D)$ when treated like a group is simply re-arranging where the addition is happening in $\bigoplus_{i,j} D$. To show that the ring structure is compatible, observe that

$$\varphi \mapsto \varphi_{ji}$$

is the desired map, where

$$\varphi \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} = \begin{pmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1n} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2n} \\ \vdots & \vdots & \ddots & \cdots \\ \varphi_{n1} & \varphi_{n2} & \cdots & \varphi_{nn} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix}$$

which is simply a formalization of what the group-isomorphism give us, showing us that composition is compatible, completing the proof. (see this for a more detailed explanation)

Hence, for any semi-simple R-module M, we get that:

$$\mathrm{End}_R(M) \cong_{\mathbf{Ring}} \prod_i M_{n_i}(D_i) \cong_{\mathbf{Ring}} \prod_i \mathrm{End}_R(S_i^{\oplus n_i})$$

and since each S_i is a simple R-module, $S \cong_R R/I_i$ for appropriate maximal ideal $I_i \subseteq R$, $\mathrm{End}_R(S_i) \cong \mathrm{End}_R(R/I_i)$. Thus:

$$\mathrm{End}_R(M) \cong_{\mathbf{Ring}} \prod_i M_{n_i}(\mathrm{End}_R(R/I_i))$$

If $R = k$ is a field making $M = V$ a vector-sapce over k (say n-dimensonal), then the only simple k-submodule up to isomorphism is k, and so this result reduces to:

$$\mathrm{End}_R(V) \cong_{\mathbf{Ring}} M_n(\mathrm{End}_k(k/(0))) = M_n(k)$$

where $\mathrm{End}_k(k) \cong_{\mathbf{Ring}} k$, aligning with the results we found in secion 13.2. This shows that when generlizing from linear algebra, due to the fact that a ring is non-commutative, we may have multiple non-isomorphic maximal ideals giving us multiple products of matrix rings. With a better understanding of simple modules and their endomorphisms, we continue by characterizing semi-simple modules:

Lemma 23.1.7: Semi-Simple Module Equivalence

Let M be a semi-simple module. Then the following are equivalent:

1. M is an external direct sum of simple modules:
$$M \cong \bigoplus_{i \in I} S_i$$
2. M is is the internal direct sum of simple modules:
$$M = \bigoplus_{i \in I} S_i$$
3. If $N \subseteq M$, there exists a complimentary module, that is there exists another submodule $N' \subseteq M$ such that
$$M = N \oplus N'$$
i.e., every semisimple R-module M is *projective* and *injective*.

***Proof* :**
(1) $\Rightarrow$ (2) is proposition 12.3.3.

For (2) $\Rightarrow$ (3), Let $N \subseteq M$. Then using Zorn's lemma, we take a maximal submodule $N' \subseteq M$ ordered by inclusion with the property that $N \cap N' = 0$ (upper-bounds naturally exist in this set, and 0 is in this set, and hence Zorn's lemma does apply). If $M = N + N'$, we're done, since we can use the isomorphism from proposition 12.3.3. For the sake of contradiction, let's say $N + N' \subsetneq M$. Then since $M = \sum_i S_i$, there must exist some $i \in I$ and $s_i \in S_i$ such that $s_i \notin N + N'$, since $0 \in S_j$ for all $j \in I$. Hence $S_i \subsetneq N + N'$, so $S_i \cap (N + N') \subsetneq S_i$, however S_i is simple, and hence:
$$S_i \cap (N + N') = 0$$
But then:
$$(S_i + N') \cap N \subseteq N' \cap N = 0$$
Thus, by maximality of N', $S_i + N' = N'$, so $S_i \subseteq N' \subseteq N + N'$, contradicting the assumption that $s_i \notin N + N'$, as we sought to show.

For (3) $\Rightarrow$ (1), Let $\mathcal{S}$ be the collection of simple submodule of M where their direct sum maps injectively into M. Since $0 \in \mathcal{S}$ and taking the sum's of all the modules is an upper bound, Let N be the maximal submodule with this property. By proposition 12.3.3,
$$N = \sum_{i \in I} S_i \cong \bigoplus_{i \in I} S_i$$
If the map is surjective, we are done. For the sake of contradiction, suppose it is not. Then by part (3), there exists an N' such that $M = N + N'$. If $N' = 0$, we found a contradiction. So, let's say $N' \neq 0$. Then it is enough to show that N' contains a simple module, say S, since then we can make the collection $\{(S_i)_{i \in I}, S\}$ which will map injectively into M, contradicting maximality.

(here is what the professor posted Suppose every submodule N of M has a complement (i.e. a submodule N' of M such that $M = N \oplus N'$ (internal direct sum)). Then every nonzero submodule N' of M contains a simple submodule.

Proof: first we show that every submodule K of M has the same property (every submodule has a complement): suppose L is a submodule of K. by assumption, $M = L \oplus L'$ for some submod L' of M. then note that $K = L \oplus (L' \cap K)$.

now suppose N' is nonzero submod of M. take a nonzero $x \in N'$, so Rx is a nonzero submodule and by 1st isom thm we have $Rx \cong R/I$ for some proper left ideal I. by the previous paragraph, every submodule of Rx has a complement, i.e. every submod of R/I has a complement. take any maximal left ideal m containing I. then m/I is a submodule of R/m with complement $L' in R/I$, i.e. $R/I = m/I \oplus L'$, so $L' \cong (R/I)/(m/I) \cong R/m$ is simple, i.e. L' is a simple submodule of N' (via the isom $Rx \cong R/I$))

This theorems shows that every vector-space over k is a semisimple k-module, however by remark internalSumVsSumRmk, $\mathbb{Z}$ is *not* a semi-simple $\mathbb{Z}$-module, since $2\mathbb{Z}$ is a submodule with no compliment (showing that not all free modules are semi-simple).We can characterize semi-simple rings by saying that R is semi-simple if and only if every R-module is projective. This also shows us that $\mathbb{Z}$ is not a semi-simple ring; though $\mathbb{Z}$ is projective over itself, any finite abliean group is a non-projective $\mathbb{Z}$-module. Thus, a module of a semi-simple ring is stronger then being a projective module. Hence being semi-simple is stronger (in a sense much stronger) than being projective, but may not be free: even though many example of semi-simple modules are free-modules, we would require the underlying ring to be semi-simple over itself. In fact, it is enough that the underlying ring be a semi-simple module over itself, giving us an easier condition for defining a semi-simple ring! We first intorduce an important result we will use:

Corollary 23.1.8: Extension Closure Of Semisimple Modules

Let M an R-module. Then submodule and quotient modules of M are semisimple if and only if M is semi-simple. Furthermore, if $(M_i)_{i \in I}$ is a collection of semi-simple modules, then $\bigoplus_{i \in I} M_i$ is a semi-simple module.

Proof **:**
The direct sum of semi-simple modules is semi-simple by definition, and so we check the other two statements.

Let M be a semi-simple R-module so that $M = \bigoplus_{i \in I} S_i$ is by lemma 23.1.7. First, consider a surjection $M \twoheadrightarrow N$.

$$\varphi(M) = \varphi\left(\bigoplus_{i \in I} S_i\right) = \bigoplus_{i \in I} \varphi(S_i)$$

then by Schur's lemma, either $\varphi(S_i) = 0$ or $\varphi(S_i) \cong S_i$, and hence N is semisimple.

If $N \subseteq M$, then by lemma 23.1.7, we have that $M = N \oplus N'$, and so by proposition 12.3.3

$$M/N' \cong N$$

which we just showed makes N semi-simple.

Using these two facts, it is not difficult to show that if $N \subseteq M$ and $M/N \cong N'$ are semi-simple, then so is M, where if $x \in M$, $x = \sum_i n_i + \sum_j n'_j$.

These properties of semi-simple modules allow us to characterize semi-simple rings much more easily:

Corollary 23.1.9: Semi-Simple Rings Condition

Let R be a ring. Then R is a semi-simple ring if and only if R is a semi-simple R-module, that is we only need to check that R is a semi-simple R-module for one R-module structure (namely R itself)

Proof **:**
By definition, R if R is a semi-simple ring then any R-module is a semi-simple R-module, including R over itself.

Now R is given some R-action becomes a semi-simple R-module, then by corollary 23.1.8 so is $\oplus_{i\in I} R$. We will use this to show that any R-module M is semi-simple, which will make R semi-simple by definition.

Let M is any other R-module, by by the universal property of the direct sum and free-modules:

$$M = \bigoplus_i Rm_i$$

for appropriate m_i (for ex. we can pick all nonzero elemetnts of M). Then we get a surjective map:

$$\bigoplus_{i\in I} R \to M \qquad (r_i)_{i\in I} \mapsto \sum r_i m_i$$

hence, M is the quotient of a semi-simple module, and hence semi-simple. Thus, R is semi-simple, as we sought to show.

Due to this, we get a nice description of semi-simple rings which is even stronger than semi-simple modules:

Lemma 23.1.10: Decomposing Semi-Simple Rings

Let R be a semi-simple ring. Then:

$$R = I_1 \oplus \cdots \oplus I_n$$

for a finite collection of I_k, where I_k are simple submodules of R, i.e. *minimal* left ideals of R

Proof **:**
Since R is a semi-simple ring, it is a semi-simple R-module and so by definition:

$$R = \bigoplus_{i\in\mathcal{C}} I_i$$

for simple submodules I_i (i.e. minimal ideal I_i). To make this a finite sum, since $1 \in R$, we get for an appropraite finite subset $A \subseteq \mathcal{C}$

$$1 \in I_{\alpha_1} + \cdots + I_{\alpha_n}$$

But then, since these are all ideals,

$$R = R \cdot 1 \subseteq I_{\alpha_1} \oplus \cdots \oplus I_{\alpha_n} \subseteq R$$

and hence $R = I_{\alpha_1} \oplus \cdots \oplus I_{\alpha_n}$, completing the proof.

61: Remark

As a consequence of this, any semi-simple ring R is a left-Noetherian/Artinian ring. Interestingly, we will soon show that this also implies that R is a *right*-Noetherian/Artinian ring (see rightNoethArtRemark)
Furthermore, this shows how it is necessary for the condition that R be left-Noetherian/Artinian so that the notion of being semi-simple coincide between the two definitions.

Since we characterized simple R-modules for any ring R to be R-isomorphic to R/I, and we've given a structure of R if R is a semisimple ring, we can make lemma 23.1.3 more precise:

Corollary 23.1.11: Structure Of Simple Modules Of Semi-Simple Ring

Let R be a semi-simple ring so that $R = I_1 \oplus I_2 \oplus \cdots \oplus I_n$ for minimal left ideal I_k (equivalently, simple left R-modules). Then if M is a simple left R-module,

$$M \cong_{R\text{-}\mathbf{Mod}} I_i$$

for some $1 \leq i \leq n$.

***Proof* :**
Let M be a simple nonzero left R-module, and consider the R-module homomorphism:

$$\varphi : R \to M \qquad r \mapsto r \cdot x$$

where $x \neq 0$. Since M is simple, it is nonzero, so φ is surjective. Since $R = \bigoplus_i^n I_i$, for some i's

$$\varphi|_{I_i} \neq 0$$

and since $\varphi|_{I_i} = \pi_i \circ \varphi$ we have the R-module homomorphism:

$$\varphi|_{I_i} I_i \to M$$

and since both modules are simple and this is a nonzero homomorphism, by Schur's lemma this map must be an isomorphism, as we sought to show.

Note that there can be multiple I_i's that are isomrophic to M, which we will show in a following theorem.

This shows that any simple R-module over a semi-simple ring R is in fact isomorphic to a simple R-module over R (i.e. R-isomorphic to a minimal ideal of R). Thus, studying simple R-modules over semi-simple rings comes down to studying the ring R itself, and that there are only *finitely many* simple R-modules. Furthermore, we have that $R/(I_1 \oplus \cdots I_{n-1}) \cong_R I_n$. Here is a good example of where the ring structure of the left hand side cannot exist in the right hand side. For example, if R is commutative, then R/I is a field that is R-module isomrophic to a minimal ideal, but ideals

cannot be fields (over the same identity[7]). Thus, in the case where R is semi-simple, for a maximal ideal I, R/I is R-isomorphic to a minimal ideal of R.

Exercise 23.1.1

1. Let S be a simple module. Show that any nonzero submodule of the injective enveloppe $E(S)$ is indecomposable, but not simple.
2. Show that if M is a simple R-module, then it necessarily has to be cyclic (the converse, however, is not true).
3. Show that $A = \mathbb{Q}[x, y]/(xy - yx - 1)$ is a simple ring, but no a semi-simple ring (this is a special example of an algebra called a *Weyl* Algebra)

23.2 Artin-Wederburn

We just showed how studying simple and semi-simple R-modules comes down to stuying the ring R, especially if the ring R is semi-simple. Thus, it is fruitful to classify semi-simple rings, which leads us to the main classification result for semi-simple rings, namely we build on lemma 23.1.10 and show that we can find more precisely how to repesent each minimal ideal I_i, and furthemore show that this clasification is exhaustive, and that any semi-simple ring uniquely breaks down into the result we'll show. We will then finish by showing that if we upgrade our ring to k-algebras (that is, we can put a vector-space structure on our ring), we get some stronger results due to the fact we can count dimensions.

First, we try to relate what we've shown for the hom-sets of simple modules with the strucure of simple modules. You may perhaps recall that we commented in exercise ref:HERE (in 9.1.1) and example 9.1.5 that R is note necessarily isomorphic to R^{op}. There is in fact a very nice way of linking R^{op} and R, which will be demonstrated after formally define R^{op} for reference

Definition 23.2.1: Opposite Ring

Let R be a ring. Then we define a new ring labelled R^{op} that has the same underlying set and addition operation, whith the multiplication operation $*$ being:

$$x * y = yx$$

where the multiplication yx happens in R

If R is commutative, then $R = R^{\text{op}}$, and so we only get non-trivial results when R is not commutative. An example of $R \ncong R^{\text{op}}$ would be a free-algebra over two variables (say $k\langle x, y\rangle$).

[7] $k \times \{0\} \subseteq k \times k$ is an ideal and a field with identity $(1, 0)$

Proposition 23.2.2: Properties Of Opposite Ring

Let R be a ring and R^{op} its oppposite ring. Then:

1. $(R^{\text{op}})^{\text{op}} = R$
2. $(R \times S)^{\text{op}} = R^{\text{op}} \times S^{\text{op}}$
3. $M_n(R)^{\text{op}} = M_n(R^{\text{op}})$ where the isomorphism is $A \mapsto A^t$ (essentially since $(AB)^t = B^t A^t$)

Proof **:**
exercise

With this, we can make a link between R and R^{op}:

Lemma 23.2.3: Yonneda's Lemma for opposite Rings

Let R be a ring and consider the ring $\text{End}_R(R)$ to be the set of all R-module homomorphisms from R (as an R-module) to itself. Then $\varphi : R^{\text{op}} \to \text{End}_R(R)$, $r \mapsto (x \mapsto xr)$ is a ring isomorphism:

$$R^{\text{op}} \cong_{\textbf{Ring}} \text{End}_R(R)$$

Proof **:**
Let φ_r be the given map. Let's first check that $\varphi_r \in \text{End}_R(R)$. Then we see that:

$$\begin{aligned}\varphi_r(\lambda x + \lambda y) &= (\lambda x + \lambda y)r \\ &= \lambda(xr) + \lambda(yr) \\ &= \lambda\varphi_r(x) + \lambda\varphi_r(y)\end{aligned}$$

showing that $\varphi_r \in \text{End}_R(R)$. Next, we check that φ is ineed a ring-homomorphism. By the properties of modules, it is easy to see that:

1. $\varphi_1 = \text{id}$, and hence φ sends the identity to the identity.
2. $\varphi_{r+r'} = \varphi_r + \varphi_{r'}$
3. $\varphi_{rr'} = \varphi_{r'} \circ \varphi_r$ (don't forget we are working with R^{op})

Finally, showing bijectivity, for injectivity we see that if $\varphi_r = 0$, then $0 = \varphi_r(1) = r$, and hence $\varphi_r = \varphi_0$. For surjectivity, let $\varphi \in \text{End}_R(R)$. Let's say $\varphi(1) = r$. Then $\varphi(x) = x\varphi(1) = xr$, so $\varphi = \varphi_r$, completing the proof.

In the special case where R is commutative, then $R \cong R^{\text{op}}$, showing that in this case $R \cong \text{End}_R(R)$. Using this, we will use lemma 23.2.3 along with lemma 23.1.6 to upgrade lemma 23.1.10:

Theorem 23.2.4: Artin-Wedderburn, one Direction

Let R be a semisimple ring. Then:

$$R \cong_{\mathbf{Ring}} M_{n_1}(D_1) \times \cdots \times M_{n_k}(D_k)$$

where each D_i is a division ring and $M_{n_i}(D_i)$ is the matrix ring over D_i

The key realization about this theorem is that while the multiplication on the left hand side can "send you" all over the place, the multiplication on the right hand side only sends you back into your particular component. To see this more clearly, let's say R was a k-algebra for a field k so that it has a basis. If R was a semi-simple ring, then that means that when multiplying elements of R (so figuring out where the basis elements go while multiplying), we can isolate the problem down to the components on the right hand side, and figure out where they go through matrix multiplication!

***Proof* :**
First, by lemma 23.1.10, for we get simple R-submodule of R, i.e. minimal ideals of R,

$$R = I_1 \oplus \cdots \oplus I_n$$

Now, we can always re-order the sums (i.e. $I_1 \oplus I_2 \cong I_2 \oplus I_1$, since $R \cong I_1 \oplus \cdots \oplus I_n$ by proposition 12.3.3). Hence, we will group all the I_i's that belong to the same isomrophism class (as R-modules), letting us re-write R as:

$$R \cong_{\mathbf{Ring}} I_1^{n_1} \oplus \cdots \oplus I_k^{n_k}$$

where $I_i \ncong I_j$ for $i \neq j$. Hence by lemma 23.2.3 and lemma 23.1.6:

$$\begin{aligned} R^{\mathrm{op}} &\cong_{\mathbf{Ring}} \mathrm{End}_R(R) \\ &\cong_{\mathbf{Ring}} M_{n_1}(D_1) \times \cdots \times M_{n_k}(D_k) \end{aligned}$$

where $D_i = \mathrm{End}_R(I_i)$ is a divison ring. Finally, Taking the opposite ring on both sides, we get:

$$\begin{aligned} R &\cong_{\mathbf{Ring}} (R^{\mathrm{op}})^{\mathrm{op}} \\ &\cong_{\mathbf{Ring}} M_{n_1}(D_1)^{\mathrm{op}} \times \cdots \times M_{n_k}(D_k)^{\mathrm{op}} \\ &\cong_{\mathbf{Ring}} M_{n_1}(D_1^{\mathrm{op}}) \times \cdots \times M_{n_k}(D_k^{\mathrm{op}}) \end{aligned}$$

and it is clear the D_i^{op} is still a division ring, hence completing the proof.

Thus, if we have a simple-ring then the ring breaks down into the product of matrix rings over division rings. In the special case where $R = k$ is a field, then k has only one minimal ideal, namely itself, and $\mathrm{End}_k(k) \cong M_1(k) \cong k$, and hence we reduce to $k \cong k$. If $R = k^n$ is a product of the same field, then the minimal ideals of R are $R = k \oplus \cdots \oplus k$, so that we get back $k^n \cong_{\mathbf{Ring}} k^n$.

We will now show that the converse of this theorem is true, that the product of matrix rings over division rings is semi simple. We start by showing that matrix ring are simple modules[8].

[8]This will be an example of R being both a simple R-module and a simple ring

Proposition 23.2.5: Matrix Ring over Division Rings are semi-simple

Let $R = M_n(D)$ be a matrix ring over a divison ring. Let $C = D^{\oplus n}$ be the left $M_n(D)$-module where the action is left matrix multiplication (imagine C as a column). Then C is a simple $M_n(D)$-module, and

$$M_n(D) \cong_{M_n(D)\text{-}\mathbf{Mod}} C^{\oplus n} = (D^{\oplus n})^{\oplus n}$$

showing that $M_n(D)$ is semisimple ring and $M_n(D) \cong_{\mathbf{Ring}} C^{\oplus n}$

***Proof* :**
Let $R = M_n(D)$ and $C = D^{\oplus n}$ (C for *column*). For notational simplicity, let $E_{ij} \in R$ be the elementary matrix which is 1 in position ij and zero everywehre else. Let $e_i \in C$ be the column where it is 1 at position i and zero everywehre else:

$$E_{ij} = \begin{pmatrix} 0 & & & & 0 \\ & \ddots & & \iddots & \\ & & 1 & & \\ & \iddots & & \ddots & \\ 0 & & & & 0 \end{pmatrix} \qquad e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

Then $E_{ij}e_k = \delta_{jk}e_i$ where δ_{jk} is the Kronecker Delta:

$$\delta_{jk} = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$$

Now, we will start by showing that C is a simple $M_n(D)$-module. Since $C = D^{\oplus n}$,

$$C = \bigoplus_i De_i$$

for some basis $e_1, e_2, ..., e_n$. To show that C is simlpe, it is enough to show that for any $0 \neq x \in C$, $C = Rx$ since then any nonzero submodule would in fact be C. Given our above notation, we have that:

$$x = \sum_k d_k e_k$$

for appropriate $d_k \in D$, where at least one of these is nonzero, let's say $d_j \neq 0$. Then consider $E_{ij}x \in Rx$:

$$E_{ij}x = \sum_k \delta_{jk} d_k e_k = d_j e_i$$

since $d_j \neq 0$. Since D is a division ring, we get that $d_j^{-1}e_i \in Rx$, for all $1 \leq i \leq n$. But then, we get:

$$C = \sum De_i \subseteq Rx \subseteq C$$

showing that $C = Rx$ for any $0 \neq x \in C$, showing that C is indeed simple.

Now, we can re-write this ring as the internal direct sum of matrices I_i which are zero outisde the ith column:

$$R = M_n(D) = I_1 \oplus \cdots \oplus I_n$$

$$I_k = \begin{pmatrix} 0 & \cdots & a_{1,k} & \cdots & 0 \\ 0 & & a_{2,k} & & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & & a_{2,n-1} & & 0 \\ 0 & \cdots & a_{n,k} & \cdots & 0 \end{pmatrix}$$

This decomposition works since each I_i is stable under *left* multiplication by elements of $M_n(D)$ (since left multiplication is a *column* transformation, right multiplication would be *row* transformation). Finally, it is clear that $C \cong_{R\text{-}\mathbf{mod}} I_i$ by simply moving the column to the right place, and so each I_i is simple. Hence, we have decomposed $M_n(D)$ as a dierct product of simple left $M_n(D)$-modules, and hence $M_n(D)$ is a semisimple $M_n(D)$-module, which by corollary 23.1.9 makes $M_n(D)$ a semi-simple ring. Finally, since $C^{\oplus n}$ represents the n rows of n-length columns, we see that the action of $M_n(D)$ on $C^{\oplus n}$ is identical to that of matrix multiplying elements of $C^{\oplus n}$, which gives us that $M_n(D) \cong_{\mathbf{Ring}} C^{\oplus n}$, as we sought to show.

62: Remark

In fact, $M_n(D)$ is a simple ring, meaning the only *two-sided* ideals of R are 0 and R; the above proof can be modified to show this. This gives us a nice class of non-commutative simple rings (since all simple commutative rings are fields).

Notice that it is quite sepcial that the ring opration and the module operation coincided, this because of the special way in which we decompoed the $M_n(D)$-module $M_n(D)$ and by the fact that it is a module over itself. In the previous proposition, we showed that $D^{\oplus n}$ is a simple $M_n(D)$-module. In fact, all simple modules of $M_n(D)$ are isomorphic to $D^{\oplus n}$:

Corollary 23.2.6: Structure of Simple $M_n(D)$-Module

Let $R = M_n(D)$ be a matrix ring over a division ring. Then any simple $M_n(D)$-module is isomorphic to $D^{\oplus n}$.

***Proof* :**
Let M be a simple $M_n(D)$-module. By proposition 23.2.5

$$M_n(D) \cong_{\mathbf{Ring}} D^{\oplus n} \oplus \cdots_{\mathbf{Ring}} \oplus D^{\oplus n}$$

And by corollary 23.1.11,

$$M \cong D^{\oplus n}$$

Corollary 23.2.7: Endomorphism Of Simple $M_n(D)$-Module

Let S be a simple $M_n(D)$-module. Then

$$\mathrm{End}_{M_n(D)}(S) \cong_{\mathbf{Ring}} D^{\mathrm{op}}$$

***Proof* :**
Since S is simple, $S \cong_R D^{\oplus n}$ for some n. Let $\varphi_d : S \to S$, $\varphi_d(x) = x \cdot d$ where $x \cdot d$ is x times dI. Then:

$$\varphi_d(Ax) = (Ax) \cdot d = A(x \cdot d) = A\varphi_d(x)$$

and

$$\varphi_d(x+y) = (x+y) \cdot d = (x+y)dI = xdI + ydI = x \cdot d + y \cdot d = \varphi_d(x) + \varphi_d(y)$$

hence $\varphi_d \in \mathrm{End}_R(S)$. Conversely, if $\varphi \in \mathrm{End}_R(S)$, Since $S \cong D^{\oplus n}$, and

$$D^{\oplus n} \cong \begin{pmatrix} D^{\oplus n} & 0^{\oplus n} & \cdots & 0^{\oplus n} \end{pmatrix}$$

we will treat S as a $M_n(D)$ submodule of $M_n(D)$. Now, since φ is a group homomorphism, we know that:

$$\varphi \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \varphi_{11} & \cdots & \varphi_{n1} \\ \vdots & \ddots & \vdots \\ \varphi_{n1} & \cdots & \varphi_{nn} \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \varphi_{11}(1) + \cdots + \varphi_{n1}(1) \\ \vdots \\ 0 \end{pmatrix}$$

hence let $d = \sum_i^n \varphi_{i1}(1)$, and let the right-action on $\mathrm{End}_{M_n(D)}(S)$ be right multiplication by the identity matrix dI. Then: Then for all $x \in S$, we have:

$$\begin{aligned} \varphi \begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ x_n & \cdots & 0 \end{pmatrix} &= \varphi \left(\begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ x_n & \cdots & 0 \end{pmatrix} \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \right) \\ &\overset{!}{=} \begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ x_n & \cdots & 0 \end{pmatrix} \varphi \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \\ &= \begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ x_n & \cdots & 0 \end{pmatrix} \cdot d \end{aligned}$$

where $\overset{!}{=}$ comes from the fact that we can treat that matrix as an element of $M_n(D)$, and since φ is an $M_n(D)$-module homomorphism, we can pull it out. But then, the action is completely dependent on on d, an so we have a bijection between right multiplication by $d \in D$ an $M_n(D)$-endomrophisms. It is an easy taks to check that:

$$\varphi_{dd'} = \varphi_{d'} \circ \varphi_d$$

showing that

$$D^{\mathrm{op}} \cong_{\mathbf{Ring}} \mathrm{End}_{M_n(D)}(S)$$

as we sought to show.

The next thing we will need is to show that if R, S are semisimple rings, then so is $R \times S$. Since R and S are semi-simple R-modules and S-modules, if we can show that $R \times S$ is a semi-simple $(R \times S)$-module, we are done. For this to be the case, we must be able to decompose $R \times S$ into simple $(R \times S)$-submodules. Forunately, many universal propeties play in our favour to be able to

easily identify simple $(R \times S)$-modules of $R \times S$ given we know the simple R-modules and S-modules. To see the key idea of the following proof, we define some terms

Definition 23.2.8: Idempotent And Orthogonal Elements

Let R be a ring. Then $e \in R$ is said to be *idempotent* if

$$e^2 = e$$

Furthermore, if e_1, e_2 are idempotent elements in R, if

$$e_1 e_2 = e_2 e_1 = 0$$

then they are said to be *orthogonal*. If e is idempotent and cannot be written as the sum of two (commuting) idempotent elements, then e is said to be a *primitive* idempotent element.

Example 23.2.1: Idempotent Elements

Naturally, for any ring R, $1 \in R$ is an idempotent element.

For a more interesting example the product ring $R \times S$, and take $e = (1, 0) \in R \times S$, $f = (0, 1) \in R \times S$. Then:

1. $e^2 = e$ and $f^2 = f$ (i.e. they are both *idempotent*)
2. $ef = fe = 0$ (implying also that $ef \in Z(R \times S)$)
3. $e + f = 1$ (i.e. e and f are *not* primitive)

What happens if the number of terms in the product increases?

Next, if M is an R-module an N is an S-module, we can M and N to be $(R \times S)$-module by ignoring what happens to S (resp. R), i.e. think of the action being defined by the following function

$$R \times S \twoheadrightarrow R \to \mathrm{End}_{\mathbf{ab}}(M)$$

and similarly for a S-module N. Hence, we can define $M \oplus N$ as an $(R \times S)$-module. Conversely, if M is an $(R \times S)$-module, then using what we've shown in example 23.2.1, given $e = (1, 0)$ and $f = (0, 1)$:

$$M = eM + fM$$

Notice that eM is killed by Rf, so it's a $A \cong A \times B/(\{0\} \times B)$-module, and similarly for fM. In other words, idempoent elements characterize the splitting of modules! This gives us the following:

Proposition 23.2.9: Direct Product Decomposition

If A-**Mod** and B-**Mod** are the categoies of A-modules and B-modules, and A-**Mod**$\times B$-**Mod** is the category where every object and morphism is a product of objects or morphisms. Then there is an equivalence:

$$A\text{-}\mathbf{Mod} \times B\text{-}\mathbf{Mod} \underset{\psi}{\overset{\varphi}{\rightleftarrows}} (A \times B)\text{-}\mathbf{Mod}$$

sending:

$$(M, N) \mapsto M \oplus N$$
$$(eM, fM) \leftarrow\!\shortmid M$$

where we further get: $\varphi(\psi(M \oplus N)) \cong M \oplus N$ and $\psi(\varphi(M)) \cong M$

Proof **:**
easy to check

Proposition 23.2.10: Classifying Simple $(A \times B)$-Modules

Let P be a simple $(A \times B)$-module. Then

$$P \cong_{A\times B\text{-Mod}} M \oplus (0) \qquad \text{or} \qquad P \cong_{A\times B\text{-Mod}} (0) \oplus N$$

where M is a simple A-module and N is a simple B-module.

Proof **:**
By proposition 23.2.9, $P \cong_{A\times B\text{-Mod}} M \oplus N$ where M is an A-module and N is a B-module. BUt then:

$$P \cong_{A\times B\text{-Mod}} (M \oplus (0)) \oplus ((0) \oplus N)$$

since P is simple, one of these has to be zero, and the other one has to be simple.

Corollary 23.2.11: Product Of Semi-Simple Rings

Let R and S be semi-simple rings. Then $R \times S$ is a semi-simple ring. Furtheromre, if $R_1, R_2, ..., R_n$ are the simple R-module up to isomorphism and $S_1, S_2, ..., S_m$ are the simple S-modules, then the simple $R \times S$-modules are of the form:

$$R_i \oplus (0) \qquad (0) \oplus (S_j)$$

Proof **:**
By proposition 23.2.9, since R is a semi-simple R-module and S is a semi-simple S-module, $R \times S$ is a semi-simple $(R \times S)$-module, and so $R \times S$ is a semi-simple ring. Then by proposition 23.2.10,

if $P \subseteq R \times S$ is a simple $(R \times S)$-module, then they it must be of the form:

$$R_i \oplus (0) \qquad (0) \oplus (S_j)$$

Corollary 23.2.12: Structure Of $\prod_i M_{N_i}(D_i)$: Artin Wedderburn part II

Let $R = M_{n_1}(D_1) \times \cdots \times M_{n_k}(D_k)$. Then

1. R is semi-simple
2. There exist exactly k simple R-modules up to isomorphism, each of the form:

$$0 \oplus \cdots \oplus D_i^{\oplus n_i} \oplus \cdots \oplus 0$$

***Proof* :**
The ring R is semi-simpleby corollary 23.2.11 using induction. Furthermore, each $M_{n_i}(D_i)$ has $D_i^{\oplus n_i}$ is it's unique simple $M_{n_i}(D_i)$-module, since there are k matrix rings in the productt, there must be exactly k simple R-modules, as we sought to show.

63: Remark

We can as before get that

$$R^{\text{op}} \cong M_{n_1}(D_1^{\text{op}}) \times \cdots \times M_{n_k}(D_k^{\text{op}})$$

and hence R^{op} is left Noetherian/Artinian, which we know now is true if and only if R is right Noetherian/Artinian (Since a left R^{op})-module is a right R-module). By leftNoethArtRemark, we have that any semisimple ring is left-Noetherian/Artinian, and hence any semisimple ring is both left and right Noetherian/Artinian, even though R is not necesarily commutative!

Thus, we have the following statemnt, were the last thing we need to prove is uniquness

Theorem 23.2.13: Artin-Wedderburn

Let R be a ring. Then R is a semi-simple ring if and only if

$$R \cong_{\textbf{Ring}} M_{n_1}(D_1) \times \cdots \times M_{n_k}(D_k)$$

where each D_k is a divison ring and $M_{n_i}(D_i)$ is the matrix ring over D_i, and each D_i n_i are unique up to isomorphism

***Proof* :**
All that is left to prove is the uniqueness. Let:

$$M_{n_1}(D_1) \oplus \cdots \oplus M_{n_s}(D_s) \cong_{\textbf{Ring}} R \cong_{\textbf{Ring}} M_{n_1'}(D_1') \oplus \cdots \oplus M_{n_s}(D_s')$$

By corollary 23.2.12, R has exactly s and t simple module, and so $s = t$. Hence, for any $D_i^{\oplus n_i}$,

there must exist a $(D_i')^{\oplus n_i'}$ such that

$$D_i^{\oplus n_i} \cong_R (D_i')^{\oplus n_i'}$$

Since $D_i^{\oplus n_i} \cong_R (D_i')^{\oplus n_i'}$, $\mathrm{End}_{M_{n_i}(D_i)}(D_i^{\oplus n_i}) \cong_{\mathbf{Ring}} \mathrm{End}_{M_{n_i'}(D_i')}((D_i')^{\oplus n_i'})$. By corollary 23.2.7,

$$D_i^{\mathrm{op}} \cong_{\mathbf{Ring}} \mathrm{End}_{M_{n_i}(D_i)}(D_i^{\oplus n_i}) \cong_{\mathbf{Ring}} \mathrm{End}_{M_{n_i'}(D_i')}((D_i')^{\oplus n_i'}) \cong_{\mathbf{Ring}} (D_i')^{\mathrm{op}}$$

which, taking op on both sides, we see that:

$$D_i \cong_{\mathbf{Ring}} D_i'$$

Finally, we need to show that $D^{\oplus n} \cong_R D^{\oplus m}$ implies $n = m$. Treating $D^{\oplus n}$ and $D^{\oplus m}$ as D-modules, then we get that each have a basis of dimension n and m. Importantly, the results of the Replacement Theorem (theorem 13.1.7) applies to divions rings (notice that commutativity was not necessary), and so corollary 13.1.9 and theorem 13.1.4 give us that $n = m$, completing the proof.

This theorem also shows us that if R is a [finitely-generated] semi-simple ring, then R is indeed the direct product of simple rings (since $M_{n_i}(D_i)$ is a simple ring). Notice too that this theorem just gave use the answer to the classification of all *finite simple rings*: namely they would have ot be of the form $M_n(\mathbb{F}_q)$[9] for some finite field of order q (which is a power of some prime p). Contrast this to how hard it is to find all finite simple groups! Furthermore, the finite dimensional assumption was key: the *Weyl Algebra*:

$$W = k\langle x, y\rangle/(xy - yx - 1)$$

is a simple algebra that is not the matrix ring of a division algebra.

Example 23.2.2: Artin Wedderburn
$\mathrm{End}_{M_2(k)}(M_2(k)) = \mathrm{End}_{M_2(k)}((k^2)^{\oplus 2}) = M_2(\mathrm{End}_{M_2(k)}(k^2)) = M_2(k)$

Corollary 23.2.14: Artin-Wedderburn For k-Algebras

Let R be a semi-simple k-algebra for some field k. Then

$$R \cong_{k\text{-}\mathbf{Alg}} M_{n_1}(D_1) \times \cdots \times M_{n_l}(D_l)$$

where each D_i is a divsion k-algebra, finite dimensional as a k-vector space:

$$\dim_k R = \sum n_i^2 \dim_k(D_i)$$

and has exactly l simple R-modules as we've classified before.

***Proof* :**
By the Artin-Wedderburn theorme, we have that

$$R \cong_{\mathbf{Ring}} M_{n_1}(D_1) \times \cdots \times M_{n_l}(D_l)$$

[9] Why is the division ring a field? The answer lies in proposition 9.1.11

Let's now further assume R is a k-algebra for some field k, so that there is a map $k \hookrightarrow Z(R)$. Then by the above isomorphism:

$$\begin{aligned} k \mapsto Z(R) &\cong Z(M_{n_1}(D_1) \times \cdots \times M_{n_l}(D_l)) \\ &\cong Z(M_{n_1}(D_1)) \times \cdots \times Z(M_{n_l}(D_l)) \\ &\cong Z(D_1) \times \cdots \times Z(D_l) \end{aligned}$$

where the last one comes form the fact that $Z(M_{n_i}(D_i))$ are the diagonal elements, and $z \mapsto zI$. This shows that D_i each have a natural k-algebra strucure induced by R, and so the above isomrophsim was already a k-algebra isomorphism:

$$R \cong_{k\text{-}\mathbf{Alg}} M_{n_1}(D_1) \times \cdots \times M_{n_l}(D_l)$$

Finally, treating $M_{n_i}(D_i)$ as k-vector spaces we have:

$$M_{n_i}(D_i) \cong_{k\text{-}\mathbf{Vect}} D^{\oplus n_i^2}$$

which gives the desired result, completing the proof.

Corollary 23.2.15: Special Case: k Algebraically Closed

Let $k = \overline{k}$ be an algebraically clsoed field and D a k-algebra that is finite-dimensional as a k-vector-space. Then $D = k$

Proof **:**
Let's say $k \subsetneq D$. Then pick any $\delta \in D \setminus k$. Then

$$k \subseteq k[\delta] \subseteq D$$

the key thing to observe is that $k[\delta]$ is still commutative!! This is essentially the same reasoning for why any cyclic group is commutative! Since D is a finite-dimensional vector-sapce over k, some power of δ^n must be a k-linear combination of every smaller term. Since D is a division ring, every δ has an inverse, and so every element of $k[\delta]$ is in fact a unit. Hence $k[\delta]/k$ is an algebraic field extension. However, k is an *algebraically cloased field*, and hence $k[\delta] = k$. Since δ was arbitary, we must have that $D = k$, as we sought to show.

Corollary 23.2.16: Artin-Wedderburn For $\overline{k}$-Algebras

Let R be a semi-simple k-algebra for some algebraically closed field $k = \overline{k}$. Then

$$R \cong_{k\text{-}\mathbf{Alg}} M_{n_1}(k) \times \cdots \times M_{n_l}(k)$$

showing that R is a finite dimensional vector-space over R with dimension:

$$\dim_k(R) = \sum_i^l n_i^2$$

and has exactly l simple R-modules as we've classified before.

As a consequence of this, if R is a semi-simple $\mathbb{C}$-algebra, the simple R-modules are all R-isomorphic to $\mathbb{C}^{n_i}$ for appropriate power n_i. As a final word of warning, it is natural to wonder if all simple rings are of the form $M_n(D_i)$

Exercise 23.2.1

1. Let k be a field, k^n be a direct product of fields a k^n-module over itslef. Let $0 \oplus \cdots \oplus k \oplus \cdots \oplus 0 = k_i \subseteq k^n$. Then $\mathrm{End}_{k^n}(k_i) \cong_{\mathbf{Ring}} k$, $\mathrm{End}_{k^n}(k^n) \cong_{\mathbf{Ring}} k^n$, and $\mathrm{End}_{k^n}(k^n) \cong_{\mathbf{Ring}} M_n(\mathrm{End}_{k^n}(k_i)) \cong_{\mathbf{Ring}} M_n(k)$. But that would imply $k^n \cong_{\mathbf{Ring}} M_n(k)$ – what's the contradiction?

2. Let k be an algebraically closed field, and let R be a finite-dimensional k-algebra. Let M be a simple R-module. Prove that if r is an element of the centre $Z(R)$ of R, there exists $c \in k$ such that $rm = cm$ for all $m \in M$

24

Representation of Finite Groups

All modules will be left R-modules for a not necessarily commutative ring R. Furthermore, all actions will be non-trivial actions unless specified otherwise.

24.1 Representations of Finite Groups

Similarly to how we studied groups and by how they acted on themselves through group-actions (ex. how a group acted on itself by conjugation gave us Sylow's Theorem), we will now try to study groups using *linear algebra*. Many really difficult results in classical group theory will become much easier when studied through this new "representation" of groups (and hence why it's called *represenation theory*). Furthermore, vector-spaces afford many more structures than a module structures (ex. a complex vector-space is the starting point for the Nullstellensatz, Complex Geometry, Analysis, and much more!) Hence, one large advantage of Representation theory will be the intersection of many mathematical diciplines which contribute towards our study (an they all feedback on each other)

In this part of the group, G will always be a finite group[1], k will always be a field (usually $\mathbb{C}$, but can be any field of any character), $\mathrm{Aut}_{k\text{-}\mathbf{Vect}}(V) = \mathrm{GL}(V)$ represents all automorphisms from the k-vector space V to itself, which when $\dim(V) < \infty$ is equivalent to the group of all nonsingular linear transformations from V to itself $\mathrm{GL}_n(k)$ where $n = \dim(V)$.

Definition 24.1.1: Representation

Let G be a finite group and V a vector space over k. Then A *linear representation* is any homomorphism $\varphi : G \to \mathrm{GL}(V)$. The *degree* of the representation is the dimension of V. A *matrix representation* of G is any homomorphism $\varphi : G \to \mathrm{GL}_n(k)$

[1] Not sure about where to address infinite groups

Usually, when we say representation we will mean *linear representation*. If a vector-space V has a representation, we will often call it a G-representation. To be as rigorous as possible we say would say we have a representation of G on V over k, if the context requires it. Furthermore, most of the time we will have map from G to $\mathrm{GL}(V)$ to be *injective*. Just like for group actions, A representation is *faithful* if it is injective.

As usual, we can curry or uncurry an action[2]: we can take $\varphi : G \to \mathrm{GL}(V)$ and make it $G \times V \to V$ where $g : V \to V$ with $v \mapsto g \cdot v$ is linear over k for all g. It is left for the curious to see that this is indeed the proper uncurrying of this action.

Using the definition of $\varphi : G \to \mathrm{GL}(V)$, it is easy to define a representation homomorphism mirroring the definition of group-actions and modules (which are ring actions on abelian groups):

Definition 24.1.2: Representation Homomorphism

Let V, W be two vectors spaces over k with a linear represenation by G. Then $\varphi : V \to W$ a *representation homomorphism* if φ k-linear and for any $g \in G$:

$$\varphi(g \cdot v) = g \cdot \varphi(v)$$

such a homomorphism will be called G-linear.

From this perspective, it might be hard to see how our previous discussion on semi-simple rings might give us any insight. We thus introduce a perspective better suited to the results of chapter 23: every representation of G on V over k is in bijective correspondence with all $k[G]$-modules over V, where $k[G]$ represents the *group ring* as seen in section 9.2. This second perspective will also make it easier to define a couple of examples, and so we will delay giving examples till the correspondence has been established. As a reminder, a group ring $k[G]$ over a ring k (which in our case will always be a field) is the set of all formal sums:

$$\sum_{g_i \in G} a_i g_i$$

where addition and multiplication are defined in the natural way[3]:

$$\sum_i^n a_i g_i + \sum b_i g_i = \sum_i (a_i + b_i) g_i$$

$$\left(\sum_i^n a_i g_i\right)\left(\sum b_i g_i\right) = \sum_k^n \left(\sum_{\substack{i,j \\ g_i g_j = g_k}} a_i b_j\right) g_k$$

Another common notation for groups rings is kG, where the brackets are dropped. A couple of important acts about group rings should be recalled (or the execises ref:HERE can be completed)

1. The group G naturally injects into $k[G]$ by identifying g_i with $1g_i$ which is the formal sum $\sum_j \delta_{ij} g_i$ (where δ_{ij} is the Kronecker delta).
2. The field k naturally injects in $k[G]$ by corresonding an element $a \in k$ with $ag_1 = ae$ where $g_1 = e$ is the identity. If it is clear from context, we will sometimes simply write $a \in k[G]$ if it is understood $a = ae = ag_1$.

[2] See curryingRmk
[3] A word on this really being a natural way?

3. There is a natural k action on $k[G]$ defined by

$$b \cdot \left(\sum_i a_i g_i \right) = \sum_i (ba_i) g_i$$

due to this, $k[G]$ is a vector space over k with the elements of G as a basis, and hence is finite dimensional if $|G| < \infty$.

4. By definition, The elements of k identified as kg_1 commute with all the elements of $k[G]$, hence it is in the center of $k[G]$: $k \subseteq Z(k[G])$. Thus, $k[G]$ is in fact a k-algebra.

5. A group ring the adjoint given by the forgetful functrom from **Ring** to **Grp**, hence is free with respect to groups, justifying the the name "formal sum" for it's elements.

Using these, we show how representations of G on V over k correspond to $k[G]$-modules: let $\varphi : G \to \mathrm{GL}(V)$ be a representation, with $G = \{g_1, g_2, ..., g_n\}$. Thus, each $\varphi(g_i)$ is linear transformation (in particular an automorphism). Define the $k[G]$ action on V by defining:

$$\left(\sum_i^n a_i g_i \right) \cdot v = \left(\sum_i^n a_i \varphi(g_i) \right)(v) = \sum_i^n a_i \varphi(g_i)(v)$$

Then this defines a $k[G]$-module (as should be verified if it doesn't seem like it). A particular reminder for those who are checking that $\varphi(g_i g_j) = \varphi(g_i) \circ \varphi(g_j)$, since φ is a group homomorphism onto $\mathrm{GL}(V)$ under composition. Notice how this action in fact extends the action of k on $k[G]$ since $a \cdot v = av$ (similarly to how the $F[x]$ action extended the action of F on V in section 14.3).

Conversely, let's say V is an $k[G]$-module. Seaking to make V a vector space over k and finding a group homomorphism $\varphi : G \to \mathrm{GL}(V)$, the frist thing is that V is already a K vector-space by limiting to the action of $k \subseteq k[G]$. Furthermore, for each $g \in G$, we can obtain a map $\varphi : V \to V$ by defining:

$$\varphi(g)(v) = g \cdot v$$

Then it is easy to check that $\varphi(g)$ is a linear transformation: for $\alpha, \beta \in k$, $v, w \in V$

$$\begin{aligned} \varphi(g)(\alpha v + \beta w) &= g \cdot (\alpha v + \beta w) \\ &= \alpha(g \cdot v) + \beta(g \cdot w) \\ &= \alpha\varphi(g)(v) + \beta\varphi(g)(w) \end{aligned}$$

and by the associativity axiom of modules, we get

$$\varphi(g_i g_j)(v) = (\varphi(g_i) \circ \varphi(g_j))(v)$$

Thus, $\varphi : G \to \mathrm{GL}(V)$ is indeed a group homomorphism, and we have:

$$\left\{ V \text{ an } FG\text{-module} \right\} \rightleftarrows \left\{ \begin{array}{c} V \text{ a vector space over } k \\ \text{and} \\ \varphi : G \to \mathrm{GL}(V) \text{ a representation} \end{array} \right\}$$

Note that this also means that this also shows that linear representations also have kernel objects, since modules have kernels. Hence, the kernel of a G-representation homomorphism between V and W is G-invariant.

Definition 24.1.3: Affording Representation

Let V be a $k[G]$-module. Then given the previous discussion, we say that V *affords* the representation φ of G.

Given this new perspective, we can also characterize sub-actions. If we have a $k[G]$-module V, then a $k[G]$-submodule $W \subseteq V$ would have to be a k-submodule, and hence a subspace, but it also has to be closed under the action of g on W, i.e. $g \cdot W \subseteq W$. Hence, any $k[G]$-submodule is a G-stable subspace of V. Conversely, if $W \subseteq V$ is G-stable, then it is closed under the action of G and is a subspace, and hence a $k[G]$-submodule. Hence

$$\left\{ W \subseteq V \text{ a } k[G]\text{-submodule} \right\} \rightleftarrows \left\{ W \subseteq V \text{ a } G\text{-stable subspace} \right\}$$

Next, we should know when two representations are equivalent. Recall that when studying a linear transformation T on V to itself, a basis for V can be found that gives a "canonical form" to the matrix representation of T. Changing basis for V kept V the same, but changed the matrix representation of T up to similarity (i.e. changed the isomorphism from $\mathrm{GL}(V)$ to $\mathrm{GL}_n(k)$). The following is the analogous term:

Definition 24.1.4: Equivalent Representation

Let φ and ψ be two representations of G over k. Then we say these two represetnations are *equivalent* (or *similar*) if the $k[G]$-module affording them are isomorphic as modules. If two representations are not equivalent, they are called *inequivalent* (or *not similar*).

When the field is known[a], if V and W are equivalent (i.e. isomorphic as $k[G]$-modules), then we write $V \cong_G W$.

[a] In most cases, the field will be $\mathbb{C}$

Note that two different representations can mean two different vector-spaces over k. If $\varphi : G \to \mathrm{GL}(V)$ and $\psi : G \to \mathrm{GL}(W)$ are equivalent representations, it would make sense if there is a relation between V and W. If $T : V \to W$ is the $k[G]$-module isomorphism between them, It is a k-module isomorphism, and hence $V \cong W$ as vector spaces, implyg $\dim(V) = \dim(W)$. Furthermore, by definition of T, $T(g \cdot v) = g \cdot T(v)$. By definition of the action, we have that $T(\varphi(g)(v)) = \psi(g)(T(v))$, i.e.

$$T \circ \varphi(g) = \psi(g) \circ T \quad \Leftrightarrow \quad T \circ \varphi(g) \circ T^{-1} = \psi(g) \qquad \forall g \in G$$

That is, T is the *simultaneous change of basis* for all $\varphi(g)$, $g \in G$. In terms of matrices, there must exist a fixed invertible matrix P such that for all $g \in G$,

$$P\varphi(g)P^{-1} = \psi(g)$$

The function T (or matrix P) are said to *intertwine* φ and ψ (it gives the rules on how to relate φ and ψ).

Example 24.1.1: Representations

1. Let G be any finite group, V any 1 dimensional vector-space over k. Define the $k[G]$-module structure on V by
$$g \cdot v = v$$
for all $g \in G$, $v \in V$. This called the *trivial representation* of G. This module affords the representation $\varphi : G \to \mathrm{GL}(V)$ by $\varphi(g) = I$ (the corresponding matrix representation from G to $\mathrm{GL}_1(k)$ sends $g \mapsto I$).

2. Let G be any finite group and k any field. Let $V = k[G]$ and consider it as a $k[G]$-module. Then $\dim_k(V) = |G| = n$, with the basis being the elements of G. Then we see that each $g \in G$ permute the basis elements:
$$g \cdot g_i = gg_i = g_j \in G$$
Thus, the matrix representation for the action by the element g is an $n \times n$ matrix with 1 in the ij component if $g \cdot g_j = g_i$ (hence n 1's) and zero everywhere else. This representation of g is called the *regular representation* of G. Note how the underlying field didn't matter: the coefficients in the matrices can only ever by 1 or 0, and we would simply say "Let φ be the regular representation of G" without specifying which field the group-ring is.

3. Let $G = S_n$ ($n \in \mathbb{N}_{>0}$) and k be any field. Then let V be a vector-space over k of dimension n, $(e_1, e_2, ..., e_n)$ a bassis for V, and define the action:
$$\sigma \cdot e_i = e_{\sigma(i)}$$
that is, σ permute the basis of V. Through this, we see that we afford a faithful representation $\varphi : S_n \hookrightarrow \mathrm{GL}(V)$. Furthermore, since any finite group is a subgroup of S_n for an appropriate n, we see that a group of order n we can define $\varphi : G \to S_n \to \mathrm{GL}(V)$, and so any finite group of order n can be embedded in a vector-space of dimension $n!$ notice too how we made a $k[G]$-module given the action of a $k[S_n]$-module. This is called the *premutation representation* of G.

4. Generlizing the last result, if $\psi : H \to \mathrm{GL}(V)$ is any representation and $\varphi : G \to H$ is a group homomorphism, then $\psi \circ \varphi$ is a representation of G on V.

5. Recall that in section 18.1.3 we defined a *group character* to be a homomorphsim from G to $L^\times$. This is in fact a 1-dimensional representation of G, since $\mathrm{GL}_1(L) = L^\times$. For a concrete example, let $G = C_n$ be the cyclic group of order n and $\zeta_n \in \mathbb{C}^\times$ the nth root of unity. Then the map:
$$\chi : G \to \mathbb{C}^\times \qquad a^i \mapsto \zeta_n^i$$
is a faithful representation of G.

6. Let $G = D_8$. We already have the regular represenation of dimension 8 and premutation represetntation of dimension 8, but can we do better? In fact we can. In order to see this, let's first write out D_8 in terms of generators and relations:
$$D_8 = \langle r, s \mid r^4 = s^2 = 1,\ rs = sr^{-1} \rangle$$

Let $k = \mathbb{R}$ and consider $\mathrm{GL}_2(\mathbb{R})$. If we can find matrices that satisfy these relations, than we would have a faithful representation of degree 2. This can be laborious, but with some work we can find:

$$r \mapsto \begin{pmatrix} \cos(2\pi/n) & -\sin(2\pi/n) \\ \sin(2\pi/n) & \cos(2\pi/n) \end{pmatrix} \qquad s \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Let $G = Q_8$. We will show how different fiels can produce different represetnations. First, recall the following two generator and relation set for Q_8:

$$Q_8 = \langle i, j \mid i^4 = j^4 = 1,\ i^2 = j^2,\ i^{-1}ji = j^{-1} \rangle$$

$$Q_8 = \langle i, j, k \mid i^2 = j^2 = k^2 = ijk = 1 \rangle$$

If $k = \mathbb{C}$, and we work over $\mathrm{GL}_2(\mathbb{C})$ we can use the first representation an find:

$$i \mapsto \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix} \qquad j \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

can you find one if $k = \mathbb{R}$ and $\mathrm{GL}_4(\mathbb{R})$? (Hint: think of the Hamiltonians $\mathbb{H}$)

7. Let $H \trianglelefteq G$ which is further an abelian p-group for some prime p. Then $V = H$ is a vector-space over $\mathbb{F}_p$ with the action of $a \cdot v = v^a$. Then define the action of G on H by conjugation: $g \cdot h = ghg^{-1}$. This action is well-defind since $gh^a g^{-1} = (ghg^{-1})^a$. The kenrel of this action is $C_G(H)$. Such an action is calld a *modulr representation*.

8. Let V be the $k[S_n]$-module which afford's the permutation representation. We will find 2 S_n-invariant subspaces: Let $N \subseteq V$ be the subspace defined by:

$$N = \{\alpha_1 e_1 + \cdots + \alpha_n e_n \mid a_1 = \cdots = \alpha_n\}$$

Then this a subspace that is also S_n-invariant. As practice, show that if $n \geq 3$, then N is the unique 1-dimension subspace which is S_n-stable. This submodule is called the *trace* submodule of V.

Another natural S_n-stable subspace is the following: take $I \subseteq V$ such that

$$I = \{\alpha_1 e_1 + \cdots + \alpha_n e_n \mid a_1 + \cdots + \alpha_n = 0\}$$

notice that $I = \ker(\varphi)$ where $\varphi : V \to \mathbb{C}$, $\varphi(\alpha_1 e_1 \alpha_2 e_2 \cdots \alpha_n e_n) = \sum_i \alpha_i$, and so I is $n-1$ dimensional. Since φ is called the *augmentation map* (was in section 7.3 of D&F), I is called the *augmentation module*.

9. If V is a regular representation of G so that $V = k[G]$, then V has the same two natural G-invariant subspaces:

$$N = \{\alpha_1 e_1 + \cdots + \alpha_n e_n \mid a_1 = \cdots = \alpha_n\}$$
$$I = \{\alpha_1 e_1 + \cdots + \alpha_n e_n \mid a_1 + \cdots + \alpha_n = 0\}$$

these are both in fact *two-sided ideals*, even if G is not abelian. Since they appear naturally, they are usually given a name: N is caleld the *trace ideal* and I is called the *augment ideal*.

Corollary 24.1.5: Schur's Lemma For Representations

Let V, W be irreducible G-representations (over $\mathbb{C}$) with represetnations ρ_V and ρ_W. Then:

1. If $V \ncong_G W$, then there does not exist a non-trivial G-linear map between V and W
2. Let's say $V = W$, are finite dimensional vecor-spaces with $\rho_V = \rho_W$, then the only non-trivial G-linear maps from V to W are scalars: $\lambda I : V \to W$, $\lambda I(v) = \lambda v$.

***Proof* :**

1. Since V and W are irreducible G-representations, they are simple $k[G]$-modules, and hence we apply the original Schur's lemma
2. Let $V = W$ of the same dimension with matching actions $\rho_V = \rho_W$, and let $f : V \to W$. Since f is a linear transformation over $\mathbb{C}$, there always exists an eigen-value, say λ so that $f(x) = \lambda x$. Considser:
$$\varphi = f - \lambda I$$
so that $\varphi(x) = 0$. Since the difference of two G-linear maps is G-linear, then $\ker(\varphi)$ is a G-invariant subspace. But V is irreducible, and $\ker(\varphi)$ is non-empty, and hence must be all of V, and so $\varphi(V) = 0$, telling us that $0 = f(x) - \lambda I(x)$, i.e.
$$f(x) = \lambda x$$
completing the proof.

Recall that if V is a simple R-module over a semi-simple ring R, then $\mathrm{End}_R(V) \cong_{\mathbf{Ring}} D^{\mathrm{op}}$ is a division ring, and that any finite-dimensional division $\mathbb{C}$-algebra is in fact equal to $\mathbb{C}$. This lemma now gives an explicit isomorphism between these two!

Since we are working over a vector-space, it is often convinient when possible to bring in an inner product that give further relations between our objects. The key the following inner product is that it is invariant on G-actions:

Definition 24.1.6: G-Invariant Inner Product

Let V be $k[G]$-module. Then a *G-invariant inner prouct* is an inner product $\langle \cdot, \cdot \rangle$ along with the additional property of:
$$\langle g \cdot v, g \cdot w \rangle = \langle v, w \rangle$$

Instead of giving examples of this inner product, we will soon show if V is a semi-simple $k[G]$-module, we can transform any inner product into a G-invariant inner product (hint: can you see how to use the "averaging" idea in Machke's Theorem to do so? The answer is given in corollary 24.1.10, page 1049).

Now, the study of linear representations in themseves can difficult; we are essentially studying all possible $\mathbb{C}[G]$-module. We would rather be able to decompose them into smaller pieces, and show that any represetnation of G is some how build form these simpler pieces (similarly to our exploration of semi-simple modules). The rest of this chapter is dedicated to seeing if we can bring

Artin-Wedderburn's theorem and all of it's consequences to understanding linear representations of finite groups:

Definition 24.1.7: Decomposing Representations

A representation is called *simple, reducible, indecomposable, decomposable*, or *semisimple* if the $k[G]$-module affording it has the corresponding property

Recall that indecomposable in general does not imply irreducible. The example under definition 23.1.1 could is in fact a representation of $\mathbb{Z}$ defined by:

$$n \mapsto \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$$

which is reducible (since it is in block upper-triangular form) but not indecomposable (since it's not in block-diagonal form). Notice that any 1-dimensional representation is automatically an irreducible, since any $k[G]$-submodule of k would be a k-subspace of k, but k is already irreducible as vector-space over k.

To get an idea of how $k[G]$-submodules interact with the definitions above, let V be an n-dimensional representation of G affording the represetnation ρ, and $W \subseteq V$ an m-dimensional $k[G]$-submodule (i.e. m-dimensional G-invariant subspace) with $0 \neq m \neq n$. Then choosing a basis $\{w_1, w_2, ..., w_m\}$ for W, we can extend it to a basis $\{w_1, w_2, ..., w_m, v_{m+1}, ..., v_n\}$. Since W is G-invariant, for any $w \in W$, $\rho(g)(w) \in W$. In matrix represetntation, $\varphi(g)(w^i) = \rho^i_j w^j$, and the coefficeints of $v^{m+1}, ..., v_n$ are zero. Hence we get the matrix representation of $\rho(g)$ as:

$$\begin{pmatrix} \alpha(g) & \beta(g) \\ 0 & \gamma(g) \end{pmatrix}$$

Notice too that:

$$\begin{aligned} \begin{pmatrix} \alpha(gh) & \beta(gh) \\ 0 & \gamma(gh) \end{pmatrix} &= \rho(gh) \\ &= \rho(g)\rho(h) \\ &= \begin{pmatrix} \alpha(g) & \beta(g) \\ 0 & \gamma(g) \end{pmatrix} \begin{pmatrix} \alpha(h) & \beta(h) \\ 0 & \gamma(h) \end{pmatrix} \\ &= \begin{pmatrix} \alpha(g)\alpha(h) & \alpha(g)\beta(h) + \beta(g)\gamma(h) \\ 0 & \gamma(g)\gamma(h) \end{pmatrix} \end{aligned}$$

which shows the the converse is also true: if we find a basis so that the matrices are block upper triangular, then the subspace spanned by the first m are G-invariant. As an example of this, let V be the permutation representation of S_3. Then the set of all vectors that are invariant to the action of $G = S_3$ is $V^G = \text{span}(e_1 + e_2 + e_3)$. Let $W = \text{span}(e_1 + e_2 + e_3)$. Since $G \curvearrowright W$ is of the form $g \cdot w = w$, we see that W is in fact the *trivial representation*. Since V is semi-simple, we have that $V = V^G \oplus V^\perp$. Making some new basis for V, say $(e_1 + e_2 + e_3, e_2, e_3)$, we see that $S_3 \curvearrowright V$ will have matrix representations that are of the above form:

$$\begin{pmatrix} 1 & * & * \\ 0 & * & * \\ 0 & * & * \end{pmatrix}$$

If V were irreducible, then no such basis would exist. If $V = U \oplus W$, then we would get a block diagonal form. It turns out that all representations of groups over either a character zero ring or fields where $|G| \nmid \mathrm{char}(k)$ are block-diagonal, i.e. are semisimple!

Theorem 24.1.8: Maschkle's Theorem

Let G be a finite group and let k be a field whose characteristic does not divide $|G|$ or $\mathrm{char}(k) = 0$. Then if V is a $k[G]$-module, it is a semi-simple $k[G]$-module.

Note that the condition that $|G| \nmid \mathrm{ch}(k)$ is equivalent to $|G| \in k^\times$.

Proof **:**
We will use lemma 23.1.7[3] to show that for any $U \subseteq V$, there exists a compliment W such that $V = U \oplus W$

We will find a projection onto U: $\pi : V \to U$. Thus, we need to show it has the following properties:

1. $\pi(u) = u$ for all $u \in U$
2. $\pi(\pi(v)) = \pi(v)$ for all $v \in V$[a]

As a reminder on why this is sufficient[b], let π be such that the function defined on the $k[G]$-module that satisfies these properties. Let $W = \ker(\pi)$. Then $W \subseteq V$ is a submodule, and W is the comliment of U. First, let $v \in U \cap W$. Then since $v \in U$, $\pi(v) = v$ and since $v \in W$, $\pi(v) = 0$, hence $v = 0$, meaning $U \cap W = 0$. Next, let $v \in V$ be an arbitrary element, and write $v = \pi(v) + (v - \pi(v))$. By definition, $\pi(v) \in U$, and by the 2nd property:

$$\pi(v - \pi(v)) = \pi(v) - \pi(\pi(v)) = \pi(v) - \pi(v) = 0$$

and hence $v - \pi(v) \in W$. Hence $v \in U + W$, and so $V = U + W$, hence $V \cong U \oplus W$. Hence, it suffices to find an $k[G]$-module projection π.

Let $U \subseteq V$ be a $k[G]$-submodule. Then it is already a k-subspace, and so it has a vector-space direct sum compliment $W_0 \subseteq V$ (take a basis β for U, extend it to a basis $\mathcal{B}$ for V, then take $\mathcal{B} - \beta$) Thus, $V = U \oplus W_0$ as a vector-space. However, W_0 need not be G-stable (i.e. a $k[G]$-submodule). We will show how to find such the appropriate $k[G]$-module given W_0. First, let $\pi_0 : V \to U$ be the vector-space projection map:

$$\pi_0(u + w_0) = u$$

Here is the kay to finding a proper compliment: we will "average" π_0 over G as follows:

$$g\pi_0 g^{-1} : V \to U \qquad g\pi_0 g^{-1}(v) = g \cdot \pi_0(g^{-1} \cdot v)$$

Since π_0 maps V to U, and U is a $k[G]$-submodule, U is stable under this G-action, hence $g\pi_0 g^{-1}$ is well-defined. Note that both g and g^{-1} act as k-linear transformations, so $g\pi_0 g^{-1]}$ is a k-linear transformation. Furthermore, if $u \in U$, then so must $g^{-1}u$, so by definition of π_0, $\pi_0(g^{-1}u) = g^{-1}u$. Thus, for all $g \in G$:

$$g\pi_0 g^{-1}(u) = u$$

hence, $g\pi_0 g^{-1}$ is also a vector-space projection of V onto U!

Let $n = |G|$, and viewing n as an element of k ($n = 1 + \cdots + 1$), since $\mathrm{ch}(k) \nmid |G|$ $n \neq 0$, i.e. $n \in k^\times$. Thus, n has an inverse in k, meaning we can define:

$$\pi = \frac{1}{n} \sum_{g \in G} g \pi_i g^{-1}$$

We can immediately deduce that π has the following properties:

1. Then by definition, π is still k-linear from V to U.
2. Restricting π to U would restrict each summand in the definition of π, and so as we've pointed out earlier each summand would be the identity map. Hence, we get $1/n$ times n copes of u, meaning $\pi(U) = U$
3. π is well-defined, hence $\pi(v) \in U$, so $\pi(v)^2 = \pi(v)$

It remains to show that π is in fact a $k[G]$-module homomorphism (i.e. is $k[G]$ linear). It suffices to show that for any $h \in G$, $\pi(hv) = h\pi(v)$. This is simply a computation:

$$\begin{aligned}
\pi(hv) &= \frac{1}{n} \sum_{g \in G} g \pi_0 (g^{-1} hv) \\
&= \frac{1}{n} \sum_{g \in G} h(h^{-1} g) \pi_0 (g^{-1} hv) \\
&= \frac{1}{n} \sum_{\substack{k = h^{-1} g \\ g \in G}} h(k \pi_0 (k^{-1} v)) \\
&\overset{!}{=} h \frac{1}{n} \sum_{\substack{k = h^{-1} g \\ g \in G}} (k \pi_0 (k^{-1} v)) \\
&= h\pi(v)
\end{aligned}$$

where $\overset{!}{=}$ comes form the fact that as g goes over all elements of G, then so will $k = h^{-1} g$ (i.e. groups acting on themselves are always transitive actions). Hence, π is indeed a $k[G]$-homomorphism, and so it is a projection map, completing the proof.

[a] In fact, this is implied by 1

[b] see ref:HERE for the original development of this

Thus, if we take any representation V over k of G, and consider the $k[G]$-module, then we know through $k[G]$ and G being finite and k a field (so V is a vecor-space) that there are only *finitely many* simple $k[G]$-modules (i.e. finitely many irreducible representations of G), meaning we only have a finite number of "building blocks" from which we can build) any representation for G! Even better, the way we put together irreducible representation is through direct sums the simplest of ways of putting togegher algebaric objects! Then through the Artin-Wedderburn decomposition of V, we can find all *irreducible* represetnations that make up the representation V of G, giving us a simple finite list of irreducible representations per finite group! Note that if $p = \mathrm{ch}(k) > 0$ and $p \big| |G|$, then $k[G]$ is *not* semisimple, or if G is infinite, it need not be semi-simple

Example 24.1.2: Non-Simple Group-Ring

For the infinite case, we've presented $\mathbb{Z} \curvearrowright \mathbb{Z}^2$ above. For the finite case where $p \big| |G|$,

1. let $G = C_n$ be the cyclic group of order n. Show that $k[C_p] \cong k[a]/(a^n - 1)$
2. Show that if $\operatorname{char} k \nmid |G|$, then $k[C_n]$ is isomorphic product of fields. Since $M_1(k) = k$, by the uniqueness of Artin Wedderburn, this is the Artin-Wedderburn decomposition. Conclude that since $k[C_n]$ is abelian, any Artin-Wedderburn decomposition *has* to be a product of fields.
3. Show that if $\operatorname{char}(k) = p$, then $k[C_p]$ has nilpotent elements, and hence cannot be a product of fields.

If $k = \mathbb{C}$, then Maschkle's Theorem always applies, and we can further use Artin Wedderburn over closed k-algebras (corollary 23.2.16) to deduce a relatively simple structure! Hence, from now on, we will always have $k = \mathbb{C}$, meaning all of our representations of finite groups will be over the complex vector-field of finite dimension.

Corollary 24.1.9: Matrix Representation Consequences

Let $\rho : G \to \mathrm{GL}(V)$ be a representation over a field k where $\operatorname{char}(k) \nmid |G|$. Then there exists a basis for V such that for all $g \in G$, the matrix $\rho(g)$ is of block-digonal form:

$$\begin{pmatrix} \rho_1(g) & & & \\ & \rho_2(g) & & \\ & & \ddots & \\ & & & \rho_m \end{pmatrix}$$

where ρ_i is an irreducible representaiton of G, $1 \leq i \leq m$.

Proof **:**

Let V be the $k[G]$-module representing ρ. By Maschkle's Theorem, we may write $V = U_1 \oplus \cdots \oplus U_m$. where U_i is simple $k[G]$-module. For each U_i, choose a basis β_i. Let $\beta = \bigcup_i^m \beta_i$. then each matrix representation of $\rho(g)$ is of the form in the corollary, with $\rho(g)|_{U_i} = \rho_i(g)$, completing the proof.

Corollary 24.1.10: Defining G-Invariant Inner Product

Let V be a $k[G]$-module and $\langle \cdot, \cdot \rangle$ be an inner product. Then this inner product induces a G-invariatn inner product (commented the place I found it)

Proof **:**

If $\langle \cdot, \cdot \rangle$ is an inner product, then:

$$\langle v, w \rangle_G = \frac{1}{|G|} \sum_{g \in G} \langle gv, gw \rangle$$

is a well-defined inner product (as can be checked), and for any $h \in G$

$$\begin{aligned}\langle hv, hw\rangle_G &= \frac{1}{|G|}\sum_{g\in G}\langle ghv, ghw\rangle \\ &= \frac{1}{|G|}\sum_{k=gh\in G}\langle kv, kw\rangle \\ &= \langle v, w\rangle_G\end{aligned}$$

In other words, there is always an inner product on our space in which the group actions keeps all elements of the same length and the same angle between them, showing how we can interpret a group as somehow rotating and reflecting an object in a vector-space. Now that we know any represetnation of G is composed in some way by finitely many irreducible representations, it would be useful to be able to quickly deduce the Artin-Wedderburn decomposition of $\mathbb{C}[G]$:

Proposition 24.1.11: Properties Of Representations

Let $k = \mathbb{C}$ and G be finite. Let $S_1, S_2, ..., S_r$ be the irreducible G-representations (up to isomorphism). Then:

1. the number of irreducible G-representations is equal to the number of conjugacy classes in G
2. For every representation V of G,
$$V \cong S_1^{\oplus \ell_1} \oplus \cdots \oplus S_r^{\ell_r}$$
where $\ell_i > 0$ and is unique
3. $\mathbb{C}[G] \cong \oplus_{i=1}^n \left(S_i^{\oplus \dim_k S_i}\right)$ as regular representation
4. $|G| = \sum(\dim_k S_i)^2$

Proof **:**
part (3) and (4) come from corollary 23.2.16.

For part (1), we first use Artin-Wedderburn to conclude that $Z(\mathbb{C}[G]) = \prod_i^r \mathbb{C}$, so $\dim_{\mathbb{C}}(Z(\mathbb{C}[G])) = r$, so the dimension of the center is equal to the number of irreducible G-representations.

Now, let $\mathcal{K}_1, ..., \mathcal{K}_s$ represent the conjugacy classes (recall they partition G). Define:

$$X_i = \sum_{g\in\mathcal{K}_i} g$$

Then any X_i and X_j have no common terms, since the conjugacy classes partition G. Furthermore, by definition, for any $g \in g$, $gX_ig^{-1} = X_i$, and so $X_i \in Z(\mathbb{C}[G])$. If the X_i's form a basis for $Z(\mathbb{C}[G])$, then we would have $s = t$. As a $\mathbb{C}$-vector-space, they are already linearly independent since the group elements from a basis, so it suffices to show they span. If $\sum_{g\in G}\lambda_g a_g \in Z(\mathbb{C}[G])$.

Then by definition for all $h \in G$,

$$\begin{aligned}
\Leftrightarrow\ h\left(\sum_{g\in G}\lambda_g g\right) &= \left(\sum_{g\in G}\lambda_g g\right)h \\
\Leftrightarrow\ h\left(\sum_{g\in G}\lambda_g g\right)h^{-1} &= \sum_{g\in G}\lambda_g g \\
\Leftrightarrow\ \left(\sum_{g\in G}\lambda_g (hgh^{-1})\right) &= \sum_{g\in G}\lambda_g g \\
\Leftrightarrow\ \left(\sum_{g\in G}\lambda_{hgh^{-1}} g\right) &= \sum_{g\in G}\lambda_g g \\
\Leftrightarrow \lambda_{hgh^{-1}} &= \lambda_g
\end{aligned}$$

since h arbitrary, we see that the coefficients of elements in the same conjugacy class remain with an element that's in that conjugacy class. Equivalently, the function $G \to \mathbb{C}$, $g \mapsto \lambda_g$ is constant on conjugacy classes. Thus:

$$Z(\mathbb{C}[G]) = \left\{\sum_{g\in G} f([g])g\right\}$$

where f is a function from the conjugacy classes of G to $\mathbb{C}$. Hence $r = s$.

for (2), Since $\mathbb{C}[G]$ is a semi-simple ring, V is a semi-simple $\mathbb{C}[G]$-module, giving us existence of the decomposition by AW. For uniqueness, first note that:

$$\mathrm{End}_{\mathbb{C}[G]}(S_i) = \mathbb{C}$$

since $\mathbb{C}$ is the only [finite-dimensional?] division $\mathbb{C}$-algebra. Then, we get by Schur's lemma:

$$\mathrm{Hom}_{\mathbb{C}[G]}(S_i, V) \overset{\text{Schur}}{\cong} \mathrm{Hom}_{\mathbb{C}[G]}(S_i, S_i)^{\oplus l_i} \cong \mathbb{C}^{l_i}$$

Corollary 24.1.12: Abelian Groups And Degree 1 Decomposition

1. If A is an abelian group, then every irreducible complex representation of A is 1-dimensional (i.e. a homomorphism from A to $\mathbb{C}^\times$), and A has $|A|$ inequivalent irreducible complex representations.
2. The number of inequivalent (irreducible) 1-dimensional complex representations of any finite group G is equal to $|G/G'|$

Proof **:**

1. Since A is abelian, so is $\mathbb{C}[A]$. By the Artin-Wedderburn decomposition, it must be a product of matrices, which since $\mathbb{C}[A]$ abelian is so must each matrix ring, which is only true if and

only if $n_i = 1$. Then, by corollary 23.2.16, $\mathbb{C}[A]$ is a product of fields. It follows that any irreducible representation (i.e. simple $\mathbb{C}[A]$-module) is 1-dimensional. Conersely, if A were not abelian, $\mathbb{C}[A]$ cannot be a product of fields, hence one of the dimensions of the matrix rings *must* be greater than 1

2. Let G be an arbitrary group. we want to show that there is a bijective correspondenc between the 1-dimensiona representaitons of G and the 1-dimensional representaitons of G/G'. $\rho : G/G' \to \mathbb{C}^\times$ is a 1-dimensional representaitons of G/G', then we have the natural map $G \to G/G'$, which shows that each 1-dimensional representaitons of G/G' is also a 1-dimensional representaitons of G. Converlsy, if f $\rho : G \to \mathbb{C}^\times$ is a 1-dimensional representaitons of G, then by the universal property, this lifts to:

$$\begin{tikzcd} G/G' \arrow[r] & \mathbb{C}^\times \\ G \arrow[u] \arrow[ur] & \end{tikzcd}$$

showing that any 1-dimensional representaiton of G is associated to a 1-dimensional representaitons of G/G', showing there can't be more 1-dimensional representaitons of G than there are of G/G', completing the proof.

A more general technique can be adpoted from this proof: If $H \trianglelefteq G$ and V is an irreducible representation of G/H, then V is an irreducible representaiton of G.

Example 24.1.3: Decomposition Of Modules

1. DnF p. 847, and a bit about $\dim(V) = 1$ example right under definition

Exercise 24.1.1

1. If χ_1 and χ_2 are both 1 dimensioanl representations, then $\chi_1 \cong \chi_2 \Leftrightarrow \chi_1 = \chi_2$. Use the property of equivalent representaitons to see this.
2. Let V be a non-trivail irreducible G-representation (i.e. not $\chi(g) = 1$ for all $g \in G$). show that there are $\dim_{\mathbb{C}}(V^G) = 0$
3. By Caley's Theorem, all finite groups are linear. Do there exists infinite non-linear groups? (Hard: Do there exists finitely generated non-linear groups?)

24.2 Character Theory

We know now we can decompose a G-representation V into the direct sum of r different types of irreducible G-modules. How do we know proceed to find these? These can be quite unwiedly: if we have a 10 or 100-dimensional represenation of a group G, then we have a matrix with 100 and $10'000$ entries! We'd then have to multiply these and figure out what relation's they have between each other, quite a hassle! Furthermore, how do we do this in such way that is independent of equivalent representations? A convinient basis might be nice, but to find a nice basis for every $\rho(g) \in \rho(G)$,

we'd have to find a basis that can *simultaneously* diagonalize every element in $\rho(G)$! If there was some invariant we can find on a representation that we can use to keep track of the decomposition information between representations that is invariant of representation, that would be useful to this endevaour. The following does exactly this (Move the discussion of trace here to build-up to it? Give some of D&F examples to demonstrate it does get quite ugly sometimes?)

Definition 24.2.1: Class Function and Character

Let V be a representation of G over an arbitrary field F. Then

1. A function $f : G \to F$ that is constant on conjugacy classes is called a *class function* of G
2. If φ is a representation of G afforded by the $F[G]$-module V, then a function:

$$\chi_V : G \to \mathbb{C} \qquad g \mapsto \operatorname{tr}(g : V \to V) = \operatorname{tr}(\varphi(g))$$

is called the *character* of φ, where $\operatorname{tr}(\varphi(g))$ is the trace of φ with respect to some basis for V. A character is called *irreducible* if φ is irreducible (and reducible oherwise). The *degree* of a character is the degree of any representation affording it.

For ease of reference, recall these properties of the trace:

1. $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$ an $\operatorname{tr}(cA) = c\operatorname{tr}(A)$
2. $\operatorname{tr}(A) = \operatorname{tr}(A^t)$
3. $\operatorname{tr}(AB) = \operatorname{tr}(BA)$

Note that in general, a character is *not* a homomorphism from G into the additive or multiplicative structure of k. Since $\operatorname{tr}(AB) = \operatorname{tr}(BA)$, $\operatorname{tr}(ABA^{-1}) = \operatorname{tr}(B)$, we see that the trace (and hence the character) is independent of choice of basis, and so equivalent representations have the same character, meaning the character is indeed an invariant. Furthermore, for any representation φ of G,

$$\varphi(g^{-1}xg) = \varphi(g)^{-1}\varphi(x)\varphi(g)$$

hence, we can take the trace and use the 3rd property above to show that characters an an example of *class functions*. Furthermore, since the trace is linear, we can extend the domain of a character to be the trace over any element in $\mathbb{C}[G]$. If $F = \mathbb{C}$, then the collection of class function's form a complex vector-space, since the sum of two class functions is a class function, and multiplying by a scalar $z \in \mathbb{C}$ does not change that a class function is invariant.

Example 24.2.1: Character

1. Given the trivial representation of G (i.e. any 1-dim vector space where $g \mapsto I$), then if φ is the representating affording the trivial representation, it has $\chi_\varphi(g) = 1$ for all g. This is called the *trivial character*.
2. If we have a degree 1-representation, then the character and the representation are the same. Thus for abelian group, complex representations and their characters are the same. For

example, take the representation $\varphi : C_p \to \mathbb{C}^\times$, $\varphi(a^i) = \zeta_p^i$. Then:

$$\chi(a^i) = \zeta_p^i$$

3. Let $\Pi : G \to S_n$ be a permutation representation of G and φ the linear of G of degree n on V where:

$$\varphi(g)(e_i) = e_{\Pi(g)(i)}$$

with respect to the given basis, $\varphi(g)$ has 1 in the ii-entry if $\Phi(g)$ fixes i, and zero otherwise. Thus, we can say that the character χ is:

$$\chi(g) = \#\{\text{number of fixed points of } g \text{ on } \{1, 2, ..., n\}\}$$

If in particular the permuation representation of G into S_n is by left multiplication, then the character is called the *pemutation character*.

4. Consider any regular representation of G, φ. Then the character is:

$$\chi(g) = \begin{cases} 0 & g \neq 1 \\ |G| & g = 1 \end{cases}$$

This character is called the *regular character* of G (note how this character is not a group homomorphism into k or $k^\times$)

5. Take the representation $\varphi : D_n \to \mathrm{GL}_2(\mathbb{R})$ from example 24.1.1. Then:

$$\chi(r) = 2\cos(2\pi/n) \qquad \chi(s) = 0 \qquad \chi(1) = 2$$

where the last one is since $1 \in G$ must map to the identity matrix $I \in \mathrm{GL}_2(\mathbb{R})$.

From now on, all our representation will be complex.

Lemma 24.2.2: Properties Of Character

Let $\rho_V : G \to \mathrm{GL}(V)$ be a complex representation. Then:

1. $\chi_V(1) = \dim_{\mathbb{C}}(V)$
2. $\rho_V(g)$ is diagonalizable for all $g \in G$ and $\chi(g)$ is integral over the integers: $\chi(g) \in \overline{\mathbb{Z}}^{\mathbb{C}}$
3. If $g \in G$ has order n, then $\chi_V(g)$ is a sum of $\dim(V)$ number of nth roots of 1
4. $|\chi_V(g)| \leq \chi_V(1)$ for all $g \in G$
5. $\chi_V(g^{-1}) = \overline{\chi_V(g)}$

Proof **:**

1. We showed this in the above examples.
2. The criterion we need from linear algebra is that $A \in M_n(\mathbb{C})$ is diagonlizable if and only if

the minimal polynomial has no repeated roots, which can be implied if $M^k = I$ for some k, which it certainly is in our case since every element is finite.

3. If $g^n = 1$, then $\rho_V(g)^n = 1$, so by (2)

$$\rho_V(g) \sim \begin{pmatrix} z_1 & & \\ & \ddots & \\ & & z_k \end{pmatrix}$$

where $z_i^n = 1$ for all $1 \leq i \leq k$. Furthermore, this tells us that these trace values in the matrix representation are always integral over the integers, being the sum of integral elements.

4. $$|\chi_V(g)| = \left|\sum z_i\right| \leq \sum |z_i| = \dim(\chi_V(1))$$

5. Since $\rho_V(g^{-1}) = \rho_V(g)^{-1}$ and

$$\rho_V(g)^{-1} \sim \begin{pmatrix} z_1^{-1} & & \\ & \ddots & \\ & & z_k^{-1} \end{pmatrix}$$

$\chi_V(g^{-1}) = \sum_i z_i^{-1}$, but $\|z_i\| = 1$, so $z_i^{-1} = \overline{z_i}$, so

$$\chi_V(g^{-1}) = \sum_i z_i^{-1} = \sum_i \overline{z_i} = \overline{\chi_V(g)}$$

Now, if we have some G-representation V, it has some character χ. In the following, we will show how we can always break-down a character to be the sums of characterse on the irreducible G-representations:

Proposition 24.2.3: Linearly Independent Characters

Let G be a finite group and V a semi-simple $k[G]$-module, and φ the representation that V affords. Then:

1. The character of V is the sum of the characters of the componens appearing in the direct sum decomposition.
2. The irreducible characters of G are linearly independent set in the set of class functions. Furthermore, the irreducible characters of G form a basis of the vector-space of class functions.

Proof **:**
For the 1st point, By the Artin-Wedderburn decoposition, we have:

$$V \cong (S_1)^{n_1} \oplus \cdots \oplus (S_k)^{n_k}$$

where S_i are simple $k[G]$-modules (irreducible G-representations). If we focus on the case where

$M = S_1 \oplus S_2$, then we can choose a basis for S_1 and S_2 which will give us:

$$\varphi(g) = \begin{pmatrix} \varphi_1(g) & 0 \\ 0 & \varphi_2(g) \end{pmatrix}$$

where φ_1, φ_2 are the representations afforded by M_1 and M_2 respectively. Then if ψ_1 and ψ_2 are the characters of φ_1 and φ_2 an χ is the character of φ then we get:

$$\chi = \psi_1 + \psi_2$$

using induction get's us the result.

For the 2nd point, first notice that the class functions defined by being 1 on a conjugacy class and 0 otherise forms a basis for class functions of dimension r. It thus suffices to sho that the r irreducible characters are independent. To that end, define:

$$\widetilde{\chi_V} : \mathbb{C}[G] \to G \qquad x \mapsto \operatorname{tr}(x : V \to V)$$

Note that this map is $\mathbb{C}$-linear since $\mathbb{C}$ commutes with $\mathbb{C}[G]$, and $\widetilde{\chi_V}|_G = \chi_V$. Now, say $S_1, ..., S_r$ are simple $\mathbb{C}[G]$-modules and:

$$\sum_i^r \lambda_i \chi_{S_i} = 0 \qquad \lambda_i \in \mathbb{C}$$

Then this is equivalent to:

$$\sum_i^r \lambda_i \widetilde{\chi_{S_i}}(g) = 0 \qquad \forall g \in G$$

and since the maps are $\mathbb{C}$-linear, by linearity:

$$\sum_i^r \lambda_i \widetilde{\chi_{S_i}}(x) = 0 \qquad \forall x \in \mathbb{C}[G]$$

which by plugging into the definition we get:

$$\sum_i^r \lambda_i \operatorname{tr}(x : S_i \to S_i) = 0 \qquad \forall x \in \mathbb{C}[G]$$

Next, by Artin Wederburn, we have the decomposition:

$$C[G] \cong M_{n_1}(\mathbb{C}) \times \cdots \times M_{n_r}(\mathbb{C})$$

where each $M_{n_i}(\mathbb{C}) \cong S_i^{\oplus n_i}$, with $S_i = \mathbb{C}^{\oplus n_i}$. Using this, take the element:

$$x_i = (0, ..., I, ..., 0)$$

where I is the identity matrix[a]. Then $x_i : S_j \to S_j$ is 0 if $i \neq j$, and the identity marix if $i = j$. Thus, the trace is:

$$\operatorname{tr}(x_j : S_j \to S_j) = \begin{cases} 0 & i \neq j \\ n_i & i = j \end{cases}$$

and since n_j is at least 1, we must have that

$$\lambda_1 = \cdots = \lambda_r = 0$$

where we took advantage of the fact we are in characeristic 0

[a] If the field had finite characteristic, it is better to take the elementary matrix E_{11}

64: Remark

We can prove this directly without using Artin-Wedderburn, which we will

Corollary 24.2.4: Detecting Isomorphic Representations Using Characters

Let V_1, V_2 be representations that are semi-simple modules.

1. Every character of G is uniquely of the form

$$\sum_{i}^{r} \lambda_i \chi_i \qquad \lambda_i \geq 0$$

   where χ_i are irreducible characters.

2. the representations V_1 and V_2 are isomorphic if and only if $\chi_{V_1} = \chi_{V_2}$

Proof **:**

1. This comes straight from the fact that irreducible characters form a basis, and that any G-representation V will be some direct sum of a finite list of known simple G-modules

2. If $V_1 \cong_G V_2$, there is a matrix P that simultaneously diagonalizes so that if A and B are basis matrices, $A = PBP^{-1}$. But the trace is independent of conjugation, and so the characters are the same. For the othe way around, treating a representation as a $k[G]$-module, then by Machke's theorem, we *know* the list simple $k[G]$-modules and the actions of $k[G]$ on them. Since any $k[G]$-module can be broken down into a direct summand of simlpe $k[G]$-modules, and the two chracters are the same, the list of $k[G]$-modules and their multiplicity are the same.

Thus, we have translated the problem of finding the irreducible representations of a group G to finding the irreducible characters of G! Since a character of a representation is determined by the irreducible characters, we give a way to systematize the tracking of irreducible characters:

Definition 24.2.5: Character Table

Let $\varphi : G \to \mathrm{GL}(V)$ be a representaiton of G an χ_V it's character. Then we call the *character table* a table with the rows being irreducible characters and columns being an element in the conjugacy class. Common practice is to make the first column the identity conjugacy class and the first row the trivial representation

Example 24.2.2: Character Table

Consider $G = S_3$ Then we can start creating it's character table:

	e	$(1\ 2)$	$(1\ 2\ 3)$
1	1	1	1
χ_{sgn}	1	-1	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Since G has 3 conjugacy classes, exactly 3 irreducible characters of G. How would we go about finding it? The comning propositions work towards this.

To be able to find the missing character, and more generally find new characters for a character table, we may look to construct new character given old one, and hope that we break down those characters into a linear combination of irreducible characters. In a sense, we are finding a way to create represetnations without doing it "by hand". We will do this in two way: the first is by taking advatange of group-actions, since are relatively easy to define and their representations are matrices with 1's and 0's. Since all groups embedd into a symmetric group, this can be done systematically. We then see how to work directly with known representation to create new represenations.

Definition 24.2.6: Premuation Represenation

Let G act on a finite set X. Define $\mathbb{C}X$ to be the set of formal sums:

$$\mathbb{C}X = \left\{ \sum_{x \in X} \lambda_x x \;\middle|\; \lambda_x \in \mathbb{C} \right\}$$

So that we get a $\mathbb{C}$-algebra with $\dim_{\mathbb{C}}(\mathbb{C}X) = |X|$. Then we can make this a G-represetnation by defining:

$$g \cdot \left(\sum \lambda_x x \right) = \sum \lambda_x (g \cdot x)$$

The 3 usual premutation representations is $G \curvearrowright G$ by left multiplication (which gives the regular representation) or conjugation, and $G \curvearrowright \{*\}$ trivially (which gives the trivial representation).

Lemma 24.2.7: Character Of Permutation Representation

Let G act on X, $G \curvearrowright X$. Then

$$\chi_{\mathbb{C}X}(g) = \#\{x \in X \mid gx = x\}$$

Proof **:**
FIx $g \in G$. Then the set of x's is a basis for the representation. To make the computation simpler, we will find a convinient order for them. In particular, by choosing any $x_1 \in X$, we can find the orbit of the element:

$$\begin{tikzcd} x_1 \arrow[r, "g"] & x_2 \arrow[r] & \cdots \arrow[r] & x_k \arrow[lll, bend left] & x_{k+1} \arrow[r, "g"] & \cdots \end{tikzcd}$$

so the basis representation will look like:

$$\begin{pmatrix} 0 & & & \\ 1 & 0 & & \\ 0 & 1 & \ddots & \\ & & 1 & 0 \end{pmatrix}$$

so we get that the trace is 0 if the order of the orbit is greater than 1, hence:

$$\chi_{\mathbb{C}X}(g) = \#\{g\text{-orbits of size } 1\}$$

An immediate consequence of this is that we have another way of computing the character of the regular representation

$$\chi_{\mathbb{C}G}(g) = \begin{cases} |G| & g = e \\ 0 & g \neq e \end{cases}$$

which makes sense, since the $\dim_{\mathbb{C}}(\mathbb{C}G) = |G|$.

Next, we turn towards a more systematic way of constructing characters by creating them out of pervious ones, that is, given some representations V and W of a group G, we can make new representations by creating new representations from V and G:

Definition 24.2.8: Constructing Representations from smaller ones

Let V, W be G-representations. Then

1. we can make a $V \oplus W$ representation by:

$$g \cdot (v, w) = (g \cdot v, g \cdot w) \ [(\rho_V(g)v, \rho_W(g)w)]$$

2. we can make $V \otimes_{\mathbb{C}} W$ a representations by:

$$g \cdot (v \otimes w) \mapsto (gv) \otimes (gw)$$

3. we can make $\mathrm{Hom}_{\mathbb{C}}(V, W)$ a representations by:

$$g \cdot f : V \to W \quad v \mapsto g \cdot f(g^{-1} \cdot v)$$

4. we can make $V^* = \mathrm{Hom}_{\mathbb{C}}(V, \mathbb{C})$ a representaiton by:

$$g \cdot f : V \to \mathbb{C}, \ v \mapsto f(g^{-1} \cdot v)$$

Note that the last construction is the same as the 3rd if we take $\mathbb{C}$ as the trivial representation. Before we show what the character's of these new representations are, we first find a relatively simpler way of representing the hom representation of V and W:

Lemma 24.2.9: Hom-Tensor Representation Dual

Let V, W be G representations. Then:

$$\mathrm{Hom}_{\mathbb{C}}(V, W) \cong_G V^* \otimes_{\mathbb{C}} W$$

as G-representations.

Note that this is not true in general, even for vector-spaces. It is important that the spaces are finite-dimensional. In this way, this is a stronger result then the hom-tensor adjoint presented in theorem 16.4.1 for the case of finite-dimensional vector-spaces.

Proof **:**
By the universal property:

$$\theta : V^* \otimes_{\mathbb{C}} W \to \mathrm{Hom}_{\mathbb{C}}(V, W) \qquad \lambda \otimes w \mapsto (v \mapsto \lambda(v)w)$$

you should check that this is well-define (note that it is bilinear).

Now, pick basis $e_1, e_2, ..., e_n$ for V, $f_1, f_2, ..., f_m$ for W, and the natural dual $e_1^*, ..., e_n^*$ for V^*. Then we get a basis for $V^* \otimes W$:

$$(e_i^* \otimes f_j) \qquad 1 \leq i \leq n,\ 1 \leq j \leq m$$

Now, notice that:

$$\theta(e_i^* \otimes f_j) \qquad v \mapsto e_i^*(v) f_j \sim \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & 0 \\ 0 & & 1 & & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} = E_{ij}$$

which is a matrix with 1 in the ij position and zero evywehre else. This is clearly a basis, hence θ sends a basis to a basis, and hence θ is a vector-space isomorphism. We next check that θ is G-linear:

$$\begin{aligned} \theta(g \cdot (\lambda \otimes w)) &= \theta(g\lambda \otimes gw) \\ &\equiv v \mapsto (g\lambda)(v)gw \\ &= v \mapsto (g\lambda)(g^{-1}v)w \\ &= g(\lambda(g^{-1}v)w) \\ &\equiv g(\theta(\lambda \otimes w)g^{-1}v) \\ &= g \cdot \theta(\lambda \otimes w) \end{aligned}$$

Lemma 24.2.10: Creating New Characters

1. $\chi_{V \oplus W} = \chi_V + \chi_W$
2. $\chi_{V \otimes W} = \chi_V \cdot \chi_W$
3. $\chi_{\mathrm{Hom}_{\mathbb{C}}(V,W)} = \overline{\chi_V}\chi_W$
4. $\chi_{V^*} = \overline{\chi_V}$

Proof **:**

1. obvious
2. It is enough to show that if $f : V \to V$ and $g : W \to W$ are linear functions, then:
$$\mathrm{tr}(f \otimes h) = \mathrm{tr}(f)\,\mathrm{tr}(h)$$
Now if we pick basis (e_i) for V and (f_j) for W, then $f(e_i) = \sum a_{ij}e_j$ and $h(f_k) = \sum b_{lk}f_l$. Then:
$$(f \otimes h)(e_i \otimes f_j) = \left(\sum_j a_{ij}e_j\right) \otimes \left(\sum_l b_{kl}f_l\right) = \sum_{j,l} a_{ij}b_{lk}(e_j \otimes f_l)$$
Notice the cofficient of $e_i \otimes f_j$ is $\sum_{i,k} a_{ii}b_{kk} = \left(\sum_i a_{ii}\right)\left(\sum_k b_{kk}\right) = \mathrm{tr}(f)\,\mathrm{tr}(h)$
3. Recall that for $f \in V^*$,
$$(gf)(v) = f(g^{-1} \cdot v)$$
that is
$$\rho_{V^*}(g)(f) = f \circ \rho_V(g^{-1})$$
that is, it is the transpose:
$$\rho_{V^*}(g) = (\rho_V(g^{-1}))^t$$
So, since the transpose matrix has the same trace, we have
$$\chi_{V^*}(g) = \chi_V(g^{-1}) = \overline{\chi_V(g)}$$
we then apply lemma 24.2.9 and part (3)to get:
$$\chi_{\mathrm{Hom}_{\mathbb{C}}(V,W)} = \overline{\chi_V} \cdot \chi_W$$
4. $\chi_{V^*} = \chi_{\mathrm{Hom}(V,\mathbb{C})} = \overline{\chi_V} \cdot 1 = \overline{\chi_V}$.

With this, we can create new characters! However, we have yet to establish how we may find the irreducible components of these chracters. For example, if we take $V \otimes V$, we may find that $\chi_{V \otimes V}$ is a linear combination of some characters already present in our table, but then we have a left-over part. How would we determine if the left-over is irreducible? We will show this by defining an *inner product* on the vector-space of class equations, and show that the irreducible form an *orthogonal basis*!

Definition 24.2.11: Inner Product On Class functions

Let $\chi_1, \chi_2 : G \to \mathbb{C}$ be class functions. Then define:

$$\langle \chi_1, \chi_2 \rangle = \frac{1}{|G|} \sum_{g \in G} \chi_1(g)\overline{\chi_2(g)} \in \mathbb{C}$$

Then this is a [Hermitian] *inner product*.

It should be verified that this is indeeed an inner product. Since class functions are by definition constant on conjugacy classes, we only need to run through representatives of conjgacy classe times the order of each conjugacy class:

$$\langle \chi_1, \chi_2 \rangle = \frac{1}{|G|} \sum_{\substack{\mathcal{K} \in \operatorname{conj}(G) \\ g \in \mathcal{K}}} |\mathcal{K}| \chi_1(g)\overline{\chi_2(g)} \in \mathbb{C}$$

where we pick a single represetnative $g \in \mathcal{K}$ and sum over the number of conjugacy classes, i.e. $|\operatorname{conj}(G)|$. Looking a the inner product like this is desirable to make computation easier (ex. if taking the inner product of a character with itself, we would simply take the squrae of each element in a row multiplied by the size of it's conjugcay class). Notice how this inner product mirrors the G-invariant inner product we induced in corollary 24.1.10 if we let $\langle , \cdot, \cdot \rangle$ be the usual conjugate dot product on $\mathbb{C}^n$. (TBD)

Lemma 24.2.12: Inner Product With Trivial Character

Let V be a G-representation. Then if $\mathbb{1}$ is the trivial representation,

$$\langle \chi_V, \chi_{\mathbb{1}} \rangle = \dim_{\mathbb{C}}(V^G)$$

where $V^G = \{v \in V \mid g \cdot v = v \ \forall g \in G\}$ is the subspace fixed by G.

Proof **:**
Define the linear map $\pi : V \to V$ where

$$v \mapsto \frac{1}{|G|} \sum_{g \in G} g \cdot v$$

then:

$$\pi|_{V^G} = \mathrm{id}$$

and

$$\operatorname{im}(\pi) \subseteq V^G$$

which implies $\pi^2 = \pi$, i.e. a projection. By linear linear algebra:

$$V = \ker(\pi) \oplus \operatorname{im}(\pi)$$

where on the kernel, $\pi = 0$, and on the image $\pi = \mathrm{id}$. Thus,

$$\frac{1}{|G|} \sum_g \chi_V(g) = \frac{1}{|G|} \sum_g \operatorname{tr}(g : V \to V) = \operatorname{tr}(\pi) = \dim_{\mathbb{C}}(V^G)$$

but the left hand side is $\langle \chi_V, \chi_{\mathbb{1}} \rangle$, completing the proof.

Corollary 24.2.13: Row Orthogonality

If V and W are any representations of G, then

$$\langle \chi_V, \chi_W \rangle = \dim_{\mathbb{C}}(\hom_{\mathbb{C}[G]}(W, V))$$

In particualr, if V an W are irreducible, then

$$\langle \chi_V, \chi_W \rangle = \begin{cases} 1 & V \cong W \\ 0 & \text{otherwise} \end{cases}$$

that is, irreducible character's are orthogonal to each other.

Proof **:**
We reduce to the previous lemma:

$$\begin{aligned} \langle \chi_V, \chi_W \rangle &= \frac{1}{|G|} \sum_{g \in G} \chi_V(g) \overline{\chi_W(g)} \\ &= \frac{1}{|G|} \sum_{g \in G} \chi_{\mathrm{Hom}_{\mathbb{C}}(W,V)}(g) \\ &= \langle \chi_{\mathrm{Hom}_{\mathbb{C}}(W,V)}, 1 \rangle \\ &= \dim_{\mathbb{C}}(\mathrm{Hom}_{\mathbb{C}}(W, V)^G) \end{aligned}$$

Note that $f \in \mathrm{Hom}_{\mathbb{C}}(W, V)$ is G-invariant, so $g \cdot f = gfg^{-1} = f$ for all $g \in G$ (recall the definition of $g \cdot f$ on $\mathrm{Hom}_{\mathbb{C}}(W, V)$). Thus this is true iff

$$gf = fg$$

Thus

$$gf(w) = f(gw) \qquad \forall g \forall w \in W$$

so f is $\mathbb{C}[G]$-linear so f is in the $\mathrm{Hom}_{\mathbb{C}[G]}(W, V)$.

There is also a more concrete proof worth remarking: since characters are linear combination's of irreducible characters, the direct product of irreducible modules translates to the sum of an irreducible character with the multiplicity of the irreducible module afording the representation being the multiplicity of the ireducibile characters. Hence:

$$\begin{aligned} \langle \chi_V, \chi_W \rangle &= \langle \sum_i m_i \chi^{(i)}, \sum_j m_j \chi^{(j)} \rangle \\ &= \sum_{i,j} m_i m_j \langle \chi^{(i)}, \chi^{(j)} \rangle \end{aligned}$$

where we have now reduced it to $\langle \chi^{(i)}, \chi^{(j)} \rangle$. Expanding the inner product and using the hom-set definition, we can apply Schur's Lemma and get:

$$\langle \chi^{(i)}, \chi^{(j)} \rangle = \delta_{ij}$$

Thus, our formula becomes:

$$\begin{aligned}\langle \chi_V, \chi_W \rangle &= \langle \sum_i m_i \chi^{(i)}, \sum_j m_j \chi^{(j)} \rangle \\ &= \sum_{i,j} m_i m_j \langle \chi^{(i)}, \chi^{(j)} \rangle \\ &= \sum_{i,j} m_i m_j \delta_{ij}\end{aligned}$$

To find $\langle \chi^{(i)}, \chi^{(i)} \rangle$ for an irreducible character, seeing it through multiplicites, then we know that an irreducible character corresponds to an irreducible representation, which necessarily has to then have multiplicity one, so if in particular:

$$\langle \chi^{(i)}, \chi^{(i)} \rangle = m_i^2 = 1 \cdot 1 = 1$$

Thus, the irreducible character produce and orthogonal basis of class functions!

Corollary 24.2.14: Consequence of Orthogonality

Let $V = \bigoplus_i^r S_i^{l_i}$ where each S_i irreducible and $S_i \ncong S_j$. Then

1. the multipliity of l_i of S_i is: $l_i = \langle \chi_V, \chi_{S_i} \rangle$
2. $\langle \chi_V, \chi_V \rangle = \sum_i l_i^2$
3. $\hom_G(W, V) \cong \hom_G(V, W)^*$

Proof **:**

1. Use the Artin-Wedderburn decomposition and the fact that direct sums of modules becomes sums of characters
2. this is just taking advantage of the summing formula
3. This is taking advantage of the fact that $\langle \chi, \psi \rangle = \overline{\langle \psi, \chi \rangle}$.

The second line is useful enough to be given it's own name:

Definition 24.2.15: Character Norm

Let V be a G-representation and let $\langle \cdot, \cdot \rangle$ be the inner product on the class functions of G. Then for any class function $\psi = \sum_i \lambda_i \chi_i$,

$$\|\psi\| = \left(\sum_i \lambda_i^2 \right)^{1/2}$$

We will use all this informaiton to deduce new characters:

Example 24.2.3: Orthogonality Relation
Let's go back to our table for S_3:

	e	$(1\ 2)$	$(1\ 2\ 3)$
1	1	1	1
χ_{sgn}	1	-1	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Since S_3 has 3 conjugacy classes, it can at most have one more irreducible character. What can this character be? We may start by looking at the regular represetnation of S_3 and take it's character χ_V. Then:

$$\langle \chi_V, 1\rangle = 1$$
$$\langle \chi_V, \chi_{\text{sng}}\rangle = 0$$

showing the regular character χ_V has one copy of the trivial representation (which makes sense, since $\chi_V(e) = 1$). Since

$$\chi_V(e) = 3 \qquad \chi_V((1\ 2)) = 1 \qquad \chi_V((1\ 2\ 3)) = 0$$

then we can take a "compliment" character by subtracting the trivial character, which defines a character which gives:

$$\chi^{\perp}(e) = 2 \qquad \chi^{\perp}(1\ 2) = 0 \qquad \chi^{\perp}(1\ 2\ 3) = -1$$

So, if there were to be another irreducible character, it would be of this form. But does there exist a degree 2 irreducible representation with this character? In fact, there has to: we saw that the number of irreducible representations is equal to the number of conjugacy classes, meaning the number of characters must be equal to the number of irreducible representations.

More generally, there can be a few approaches to finding it. One is to tensor two characters we know and see if it is irreducible, or one of it's linear components is irreducible. Another is to find the compliment of a represetnation of an existing character (since $k[G]$-modules are semi-simple). This compliment method is in general less desirable, and usually harder, but is managable if working with permutation representations.To demonstrate, let V be the degree three permutation representation of $S_3{}^a$ with basis $\{e_1, e_2, e_3\}$. Then if $\mathbb{1}$ is the trivial action, $V^{\mathbb{1}} = \text{span}(e_1+e_2+e_3)$, Thus, by Machke's theorem, $V^{\mathbb{1}}$ has a compliment. To see it, we simply have to extend $e_1+e_2+e_3$ to an orthogonal basis for V, for example, choose $(e_1 + e_2 + e_3, e_2 - e_1, e_3 - e_1)$. Then we can calculate what the action ρ does on this basis. For example:

$$\begin{aligned}\rho(1\ 2)(e_1 + e_2 + e_3) &= e_1 + e_2 + e_3\\ \rho(1\ 2)(e_2 - e_1) &= e_1 - e_2 = -(e_2 - e_1)\\ \rho(1\ 2)(e_3 - e_1) &= e_3 - e_2 = -(e_2 - e_1) + (e_3 - e_1)\end{aligned}$$

If we write out ρ for an element in each conjugacy class, we will find that it's trace will be

$$\chi^{\perp}(e) = 2 \qquad \chi^{\perp}((1\ 2)) = 0 \quad \chi^{\perp}((1\ 2\ 3)) = -1$$

showing we have found the desired irreducible representation! Finally, by Artin-Wedderburn, we have that $\mathbb{C}S_3 \cong \mathbb{1} \oplus \text{sgn} \oplus V^{\perp}$

Now that we know the character table for S_3, we can construct any other character of representations of S_3, and hence any other representation of S_3. For example, we know that $V^\perp \otimes V^\perp$ has as character $\chi_{V^\perp} \cdot \chi_{V^\perp}$. Using this, we can compute that:

$$\langle V^\perp \otimes V^\perp, \mathbb{1} \rangle = 1$$
$$\langle V^\perp \otimes V^\perp, \text{sgn} \rangle = 1$$
$$\langle V^\perp \otimes V^\perp, V^\perp \rangle = 1$$

showing that

$$\chi_{V^\perp \otimes V^\perp} = \chi_{\mathbb{1}} + \chi_{\text{sgn}} + \chi_{V^\perp}$$

[a]Which in this case, is just the regular representation

Recall from linear algebra that if the rows are orthogonal in a square matrix, then so would be the column (and vice-versa). We bring to light this relation for the character table. As in linear algebra, it comes down to proper normalizing of a different basis: recall that the class equations of the form $\delta_{\mathcal{K}} : G \to \mathbb{C}^\times$ where:

$$\delta_{\mathcal{K}}(g) = \begin{cases} 1 & g \in \mathcal{K} \\ 0 & \text{otherwise} \end{cases}$$

this naturally forms a basis for the class equations too. The following theorem will also show how to go between this basis and the irreducible character basis:

Theorem 24.2.16: Column Orthogonality

If $\chi_1, \chi_2, ..., \chi_r$ are the irreducible characters of G and $g_1, g_2, ..., g_r$ represent the conjugacy classes of G, then

$$\sum_i \chi_i(g_s)\overline{\chi_i(g_t)} = \delta_{st} \frac{|G|}{|\operatorname{conj}(g_s)|}$$

which by Orbt-stabalizer,

$$\sum_i \chi_i(g_s)\overline{\chi_i(g_t)} = \delta_{st} |C_G(g_s)|$$

If $g_s, g_t \in \operatorname{conj}(g)$, then we get that

$$\sum_i \chi_i(g_s)\overline{\chi_i(g_t)} = \sum_i |\chi_i(g_s)|^2 = |C_G(g_s)|$$

in particular, we must have:

$$\sum_i |\chi_i(e)|^2 = |G|$$

Proof **:**
Define $f_s : G \to \mathbb{C}$,

$$g \mapsto \begin{cases} 1 & g \in \operatorname{conj}(g_s) \\ 0 & \text{otherwise} \end{cases}$$

which is natural basis for the class functions. since it's a class function, we know we can write it

as the sum of orhtogonal ireducibile characters:

$$f_s = \sum_i^r \lambda_{is}\chi_i \qquad \lambda_{is} \in \mathbb{C}$$

since these are orthornormal, we know that:

$$\lambda_{is} = \langle f_s, \chi_i \rangle$$

so computing:

$$\begin{aligned}
\lambda_{is} &= \langle f_s, \chi_i \rangle \\
&= \frac{1}{|G|} \sum_{g \in G} f_s(g)\overline{\chi_i(g)} \\
&= \frac{1}{|G|} \sum_{g \in \operatorname{conj}(g_s)} \overline{\chi_i(g)} \\
&= \frac{|\operatorname{conj}(g_s)|}{|G|} \overline{\chi_i(g_s)}
\end{aligned}$$

so

$$f_s = \frac{\operatorname{conj}(g_s)|}{|G|} \sum_i^r \overline{\chi_i(g_s)}\chi_i$$

we know evaluate at g_t.

Since both the rows and columns of a character table are all orthogonal to each other, the character table is in fact invertible!

Example 24.2.4: Using Column Orthogonality

Let $G = S_4$. We want to find it's character table. We first pick what conjugacy classes to index the columns by, say:

conj. class	1	(1 2)	(1 2 3)	(1 2)(3 4)	(1 2 3 4)
size	1	6	8	3	6

Next, we can immediately deduce that $\chi_{\mathbb{1}}$ and χ_{sgn} are irreducible characters, giving us the current begining of a character table:

	1	6	8	3	6
	1	(1 2)	(1 2 3)	(1 2)(3 4)	(1 2 3 4)
$\chi_{\mathbb{1}}$	1	1	1	1	1
χ_{sgn}	1	-1	1	1	-1

To find more characters, we can first take the pemutation character of $\mathbb{C}X$. The character of each conjugacy class is the number of fixed points, giving us:

	1	6	8	3	6
	1	(1 2)	(1 2 3)	(1 2)(3 4)	(1 2 3 4)
$\chi_{\mathbb{C}X}$	4	2	1	0	0

We can see that the trivial character is part of linear decomposition of $\chi_{\mathbb{C}X}$, since:

$$\langle \chi_{\mathbb{C}X}, \chi_{\mathbb{1}} \rangle = (1(4 \cdot 1) + 6(2 \cdot 1) + 8(1 \cdot 1))/24 = 24/2 = 1$$

and so $\chi_{\mathbb{C}X} - \chi_{\mathbb{1}}$ is a 3 dimensional character, let's label this χ_3. Then $\|\chi_3\|^2 = 1$, showing that it's irreducible, hence we expand our table:

	1	6	8	3	6
	1	(1 2)	(1 2 3)	(1 2)(3 4)	(1 2 3 4)
$\chi_{\mathbb{1}}$	1	1	1	1	1
χ_{sgn}	1	-1	1	1	-1
$\chi_{\mathbb{C}X} - \chi_{\mathbb{1}} = \chi_3$	3	2	0	-1	-1

To find the next character, we can multiply χ_{sgn} and χ_3 (note we can do χ_3 with itself, but it will be 9 dimensional, which we'll show in the next proposition cannot be an irreducible representation). Let$\chi_4 = \chi_{\text{sgn}} \cdot \chi_3$. Then $\|\chi_4\|^2 = 1$, meaning it's irreducible, and so we add another row to our table:

	1	6	8	3	6
	1	(1 2)	(1 2 3)	(1 2)(3 4)	(1 2 3 4)
$\chi_{\mathbb{1}}$	1	1	1	1	1
χ_{sgn}	1	-1	1	1	-1
χ_3	3	2	0	-1	-1
$\chi_{\text{sng}} \cdot \chi_3 = \chi_4$	3	-1	0	-1	1

Finally, since $|G| = \sum_i \chi_i(e)^2$, we can deduce that $\chi_5(e) = 2$. We can't do this same trick exactly for the other columns since we will get the value up to a complex number z with $\|z\| = 1$. To find the rest of the character values, we can either use column orthogonality For example, to find $\chi_5(1\ 2)$, noticet that the inner product of the first and second column is:

$$1 \cdot 1 + (-1) \cdot 1 + 3 \cdot 1 + 3 \cdot (-1) = 0$$

and so $\chi_5(2)$ must be 0. For the next column, we'd get :

$$1 \cdot 1 + 1 \cdot 1 + 3 \cdot 0 + 3 \cdot 0 = 2$$

and so, since $2 \cdot -1 = -2$, we get that $\chi_5(1\ 2) = -1$. Continuing this way, we get that the rest of the table will be:

	1	6	8	3	6
	1	(1 2)	(1 2 3)	(1 2)(3 4)	(1 2 3 4)
$\chi_{\mathbb{1}}$	1	1	1	1	1
χ_{sgn}	1	-1	1	1	-1
χ_3	3	2	0	-1	-1
χ_4	3	-1	0	-1	1
χ_5	2	0	-1	2	0

We can verify this is indeed irreducible by taking $\|\chi_5\|^2 = 1$.

Another method would be to use the regular representation, where we get:

$$\chi_{\mathbb{C}[G]} = \sum_i \chi_i(1)\chi_i$$

These can be verified by taking advantage of the fact that the "column norm" should give you $C_G(g)$ (i.e. the number on top of the table).

As a fun aside, it can be shown that the represetned by χ_4 is the action of the rotation of a cube that is, if $C \subseteq \mathbb{R}^3$ are the vertex points of the cube, then χ_3 is the character of $S_4 \curvearrowright \mathrm{GL}_3(\mathbb{R})$ of all rotations C.

Notice how the dimension of each irreducible representation divided the order of the group. This is in fact true in general. It is not a trivial fact, the only "order" information we've shown so far is that the sums of the squares of the dimensions of the ireducibile representation sum to $|G|$. It is also relatively eays to show that if $W \subseteq V$ is an irreducible $k[G]$-module then the dimension of W is $< |G|$. The following is thus a much stronger result:

Proposition 24.2.17: Dimension Theorem of Irreducible Representations

Let V be an irreducible G-representation. Then:

$$\dim_{\mathbb{C}}(V) \mid |G|$$

that is, the dimension of V divides the order of G.

Proof **:**
Let $g_1, g_2, ..., g_r$ be representatives of the conjugacy class. Recall the following trick from proposition 24.1.11:

$$e_i = \sum_{g \in \mathrm{conj}(g)} g \in Z(\mathbb{C}[G])$$

Note that e_i in fact lies in $e_i \in Z(\mathbb{Z}[G])$, since $ge_ig^{-1} = e_i$ (by definition), and so extending linearly all $xe_ix^{-1} = e_i$. Now we have a commutative ring that is a finitely generated $\mathbb{Z}$-module (one way to see this is as we've done in proposition 24.1.11, another way is to notice that $\mathbb{Z}[G]$ is a finitely generated $\mathbb{Z}$-module and $\mathbb{Z}$ is Noetherian, so the center is finitely generated)

Thus, by proposition 20.5.3, $\mathbb{Z} \subseteq Z(\mathbb{Z}[G])$ is an *integral extension*. Next, we have a ring homomorphism:

$$\varphi : Z(\mathbb{C}[G]) \to \mathrm{End}_{\mathbb{C}[G]}(V) \overset{!}{\cong} \mathbb{C} \qquad z \mapsto (z : V \to V)$$

where $\overset{!}{=}$ comes from corollary 23.2.15 and Schur's lemma, and $z(v) = zv$. Hence, each element of $\varphi(Z(\mathbb{C}[G]))$ is $\mathbb{C}$-linear, and must be invertible, and so φ is in fact a map from $Z(\mathbb{C}[G]) \to \mathrm{Aut}_{\mathbb{C}}(V) = \mathrm{GL}(V)$, which is simply a restriction of the origional G-action ρ to $Z(\mathbb{C}[G])$. Hence $\rho(e_i) = \lambda_i I$ for appropriate $\lambda_i = \varphi(e_i)$, meaning $\chi(e_i) = \mathrm{tr}(e_i : V \to V) = n\varphi(e_i)$ where $n = \dim(V)$ since by what we've just shown, $\varphi(e_i)$ acts as scalar multiplication on the representation V.

Now, since $e_i \in Z(\mathbb{C}[G])$ is integral over $\mathbb{Z}$, $\varphi(e_i) \in \mathbb{C}$ is integral over $\mathbb{Z}$, that is $\varphi(e_i) \in \overline{\mathbb{Z}}^{\mathbb{C}} \subseteq \mathbb{C}$ (i.e. $\varphi(e_i)$ is algebraic). Next, we can express $\chi(e_i)$ in a different way:

$$\sum_{g \in \mathrm{conj}(g_i)} \chi_V(g) = |\mathrm{conj}(g_i)| \cdot \chi_V(g_i) = \chi(e_i)$$

so:

$$\frac{\mathrm{conj}(g_i)}{n} \chi_V(g_i) = \varphi(e_i) \in \overline{\mathbb{Z}}^{\mathbb{C}}$$

which is telling us that if we divide it by n, we still get an algebraic integer! Taking a nice linear combination and multiplying by $\overline{\chi_V(g_i)}$ we get:

$$\sum_i^r \frac{\text{conj}(g_i)}{n} \chi_V(g_i)\overline{\chi_V(g_i)}$$

Then we've just shown that $\frac{\text{conj}(g_i)}{n}\chi_V(g_i)$ is an algebraic integer, and $\overline{\chi_V(g_i)}$ is an algebraic integer (being the sum of roots of unity). Now, notice this is the form of the inner product, giving us:

$$\frac{|G|}{n}\langle \chi_V, \chi_V \rangle = \frac{|G|}{n} \in \overline{\mathbb{Z}}^{\mathbb{C}}$$

as a V irreducible. Thus, we have that $\frac{|G|}{n}$ is some algebraic integer. We further have that it is certainly a rational number ($|G|, n \in \mathbb{N}$), but then we have:

$$\frac{|G|}{n} \in \overline{\mathbb{Z}}^{\mathbb{C}} \cap \mathbb{Q} \overset{!}{=} \mathbb{Z}$$

where $\overset{!}{=}$ comes from the fact that if R is a Unique Factorization Domain, $\alpha \in \text{Frac}(R)$, α is integral over R if and only if $\alpha \in R$, completing the proof.

This will make the computation of characters of abelian groups paticularly simple! (see rep theory of finite groups: An introductory Appraoch p. 47!)

Exercise 24.2.1

1. Let G act on the n-dimensional vector-space V by basis permutation. Show that $\dim(V^G)$ equals the number of orbits of G.

24.2.1 More Character Tables

In this section, we go over a few more exmaples of character tables in order to see some techniques used to find them. For any abelian, group, the character table is straight forward. As a really simple example, consider $\mathbb{Z}/3\mathbb{Z}$:

	1	1	1
	1	2	3
χ_1	1	1	1
χ_2	1	ζ_3	ζ_3^2
χ_3	1	ζ_3^2	ζ_3

If we take $\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$, we will get a 9 by 9 table. The table bellow fills in a few of the characters (note that these were found by taking $\chi_i(g_i)\chi_j(g_j)$)

	1	1	1	1	1	1	1	1	1
$\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$	$(1,1)$	$(2,1)$	$(3,1)$	$(1,2)$	$(2,2)$	$(3,2)$	$(1,3)$	$(2,3)$	$(3,3)$
$\chi_{(1,1)}$	1	1	1	1	1	1	1	1	1
$\chi_{(2,1)}$	1	ζ_3	ζ_3^2	1	ζ_3	ζ_3^2	1	ζ_3	ζ_3^2
$\chi_{(2,2)}$	1	ζ_3	ζ_3^2	ζ_3	ζ_3^2	1	ζ_3^2	1	ζ_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Note that the root ζ_3 might change if we take the direct product of cyclic groups of differnet orders, for example the character table of $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$ has roots of unity of order 3 and 6 (as you can verify).

More generally, if G is abelian, we break G down into a product of cyclic groups and construct the table like above. Thus, we focus on non-abelian groups. We have already found the character table for S_3:

S_3	e	(1 2)	(1 2 3)
$\chi_{\mathbb{1}}$	1	1	1
χ_{sgn}	1	-1	1
χ_3	2	0	-1

and the character table for S_4:

	1	6	8	3	6
S_4	1	(1 2)	(1 2 3)	(1 2)(3 4)	(1 2 3 4)
$\chi_{\mathbb{1}}$	1	1	1	1	1
χ_{sgn}	1	-1	1	1	-1
χ_3	3	2	0	-1	-1
χ_4	3	-1	0	-1	1
χ_5	2	0	-1	2	0

Next, let $G = D_4$. Then: $D' = \{e, r^2\}$, so G has 4 degree 1-irreducible representations. We also found in example 24.1.1 that D_4 has a degree 2 irreducible representation. These are all of the representations, sinc $1 + 1 + 1 + 1 + 2^2 = 8$. To find the 1-dimensional irreducible representations, notice that $D_4/D_4' \cong \mathbb{Z}_2 \times \mathbb{Z}_2$, meaning we can find the irreducible representations of $\mathbb{Z}/2\mathbb{Z}$, which are:

$\mathbb{Z}/2\mathbb{Z}$	0	1
χ_0	1	1
χ_1	1	-1

and take the tensor product to create:

$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
$\chi_{(0,0)}$	1	1	1	1
$\chi_{(1,0)}$	1	-1	1	-1
$\chi_{(0,1)}$	1	-1	-1	1
$\chi_{(1,1)}$	1	1	-1	-1

and using this, we can complete the table for D_4:

D_4	e	r	r^2	σ	σr
χ_1	1	1	1	1	1
χ_2	1	1	1	-1	-1
χ_3	1	-1	1	1	-1
χ_4	1	1	1	-1	1
χ_5	2	0	-2	0	0

More generally, for D_n, the character table will conssite of characters of degree 1 and 2.

Let's now let $G = Q_8$. Taking the presentation:

$$Q_8 = \langle i, j | i^4 = 1,\ i^2 = j^2,\ i^{-1}ji = j^{-1} \rangle$$

with $k = ij$ and $i^2 = -1$, we get the conjugacy classes can be represented by $1, -1, i, j, k$, with sizes 1, 1, 2, 2, 2. Since $Q_8/Q_8' \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, there are 4 degree 1 characters. The final irreducible character thus *must* have degree 2; let's label if χ_5. We have in fact explicitely found this character before in example 24.1.1, but we can find it without actually manually finding a representation. Let φ be an irreducible representation of degree 2. Since $-1 \in Z(Q_8)$, by Schur's lemma $\varphi(-1)$ is a 2×2 scalar matrix of order 2, hence it is $\pm I$, and so $\chi_5(-1) = \pm 2$. Next, if $\chi_5(i) = a$, $\chi_5(j) = b$, and $\chi_5(k) = c$, then since χ_5 is orthogonal, we have that:

$$1 = \langle \chi_5, \chi_5 \rangle = \frac{1}{8}(2^2 + (\pm 2)^2 + 2a\overline{a} + 2b\overline{b} + 2c\overline{c})$$

and since $a\overline{a}, b\overline{b}, c\overline{c} \in \mathbb{R}^+$, they must all be zero. Furthermore, we must have that:

$$0 = \langle \chi_1, \chi_5 \rangle = \frac{1}{8}(2(\pm 2) + 0 + 0 + 0 + 0)$$

implying that $\chi_5(-1) = -2$. Thus, we have that the character table is:

Q_8	1	i	-1	j	k
χ_1	1	1	1	1	1
χ_2	1	1	1	-1	-1
χ_3	1	-1	1	1	-1
χ_4	1	1	1	-1	1
χ_5	2	0	-2	0	0

Notice that this character table is identical to the character table of D_8. Hence, identical character tables does not imply isomorphic groups! Finally, we could have found the last row of the above table by applying column orthogonality to find the missing values, once we knew $\chi_5(1) = 2$.

Next, let $G = A_4$. Since we found the character table for S_4, we can use it it to find the character table of A_4. We first note that A_4 has 4 conjugacy classes (represented by 1, (1 2)(3 4), (1 2 3) and (1 3 2)), and $[A_4 : A_4'] = 3$, meaning A_4 has 3 chacacters of degree 1, and the only other character must be of degree 3. Next, using the orthogonality relation, notice that χ_3 in the table for S_4 when restricted to A_4 is still irreducible (this is certainly not generally the case, for ex. the degree 2 representaiton of S_3 will become reducible since any proper subgroup of S_3 is abelian). Overall, we get:

A_4	1	(1 2)(3 4)	(1 2 3)	(1 3 2)
χ_1	1	1	1	1
χ_2	1	1	ζ_3	ζ_3^3
χ_3	1	1	ζ_3^2	ζ_3
χ_5	3	-1	0	0

Finally, we will do the character table for S_5:

S_5	1	(1 2)	(1 2 3)	(1 2 3 4)	(1 2 3 4 5)	(1 2)(3 4)	(1 2)(3 4 5)
χ_1	1	1	1	1	1	1	1
χ_2	1	-1	1	-1	1	1	-1
χ_3	4	2	1	0	-1	0	-1
χ_4	4	-2	1	0	-1	0	1
χ_5	5	-1	-1	1	0	1	-1
χ_6	5	1	-1	-1	0	1	1
χ_7	6	0	0	0	1	-2	0

TBD

24.2.2 Burnside's Theorem

We now move towards Burnsides Theorem, which tells us that any group of order $p^a q^b$ for primes p, q and numbers $a, b \in \mathbb{N}$ is a simple group! This proof is relatively simple in the context of representation theory, but it took over 70 years before a proof was found using just group theory! This is a good example of the power perspective: simply switching to the context of representation theory and seeing groups as embedded into $GL(V)$ let us prove a claim that from another perspective is much much harder!

Lemma 24.2.18: Burnside build-up Lemma

Suppose that V is an irreducible representation is of dimensin n (i.e. $\chi_1(V) = n$). If $\gcd(n, |\operatorname{conj}(g)|) = 1$, then:

$$|\chi_V(g)| = 0 \text{ or } n$$

Proof **:**
By the above proof, we have:

$$\frac{\operatorname{conj}(g)}{n}\chi_V(g) \in \overline{\mathbb{Z}}^{\mathbb{C}}$$

where $\chi_V(g) \in \overline{\mathbb{Z}}^{\mathbb{C}}$ as well, so by Bézout $(1 = am + b|\operatorname{conj}(g)|)$

$$\frac{1}{n}\chi_V(g) \in \overline{\mathbb{Z}}^{\mathbb{C}}$$

Let $\alpha = \frac{1}{n}\chi_V(g)$. Furthermore, notice that:

$$|\alpha| = \left|\frac{\chi_V(g)}{\chi_V(1)}\right| \leq 1$$

Now, let N be the Galois closure of $\mathbb{Q}(\alpha)/\mathbb{Q}$ so that $G = \mathrm{Gal}(N/\mathbb{Q})$ is the glaois group of N. Let

$$\beta = \prod_{\sigma \in G} \sigma(\alpha)$$

Then certainly β is in the fixed field of the galois group, $\beta \in N^G = \mathbb{Q}$. Notice that each $\sigma(\alpha)$ satisfies the same monic polynomials as that of α, and so $\sigma(\alpha) \in \overline{\mathbb{Z}}^{\mathbb{C}}$, and so $\beta \in \overline{\mathbb{Z}}^{\mathbb{C}}$, but then $\beta \in \overline{\mathbb{Z}}^{\mathbb{C}} \cap \mathbb{Q} = \mathbb{Z}$.

Now, recall that:

$$\alpha = \frac{\chi_V(g)}{n} = \frac{\zeta_1 + \zeta_2 + ... + \zeta_n}{n}$$

which are roots of unity, which means that:

$$\sigma(\alpha) = \frac{\sigma(\zeta_1) + \cdots + \sigma(\zeta_n)}{n}$$

so $|\sigma(\zeta_1)| \leq 1$, so $|\beta| \leq 1$. Since β is an integer, we either get that $|\beta| \in \{0, 1\}$, If $|\beta| = 0$, then $\alpha = 0$, and if $|\beta| = 1$, then $|\alpha| = 1$, and so plugging back into our original equation we get that:

$$\chi_V \in \{0, n\}$$

as we sought to show.

We can also treat this lemma as another check that we have given the correct character table (verify it for the character table of S_4 we found in example 24.2.4).

Theorem 24.2.19: Burnside's Theorem

Let $p \neq q$ be prime numbers, $a, b \in \mathbb{N}$. If $|G| = p^a q^b$, then G is solvable.

Notice how this immedaitely implies that the smallest possible order of a non-abelian simple group is 30, since $30 = 2 \cdot 3 \cdot 5$, which matches with the fact that the smallest non-abelian group is A_5 which has order 60! Recall also that if $|G| = p^a$, then we would need that $a = 1$ (namely, any p-group has a non-trivial center)

Proof **:**
Take the composition series:

$$1 = G_0 \trianglelefteq G_1 \trianglelefteq \cdots \trianglelefteq G_k = G$$

where G_i/G_{i-1}. Then we'll show that if G is simple, then G is *abelian*. For the sake of contradiction, suppos that G is non-abelian and simple with order $p^a q^b$ (where we may as well assume $a, b \in \mathbb{N}_{>0}$). Let $\chi_1, \chi_2, ..., \chi_r$ be the irreducible representation for the group G, and $\chi_1 = \chi_{\mathbb{1}}$. Let $Z_i = \{g \in G \mid |\chi_i(g)| = \chi_i(1),\ \forall i\}$. The condition in side doesn't happen so often, we can equivanelty say that $\rho_i(g)$ is a scalar by using the triangle inequality. Thus,

$$Z_i \trianglelefteq G$$

If $Z_i = G$, then G acts by scalars on V_i. But then every subspace of V_i is a subrepresentation, so the dimension of all the irreducible is 1. And so each character is of the form:

$$\chi_i : G \to \mathbb{C}^\times$$

Since G is simple, each each i is injective or trivial. If each i is injective, then G is abelian, a contradiction, and if it is trivial we get $i = 1$ (the trivial representation). Thus, it must be that:

$$Z_i = 1 \qquad \forall i > 1$$

Now, take the sylow q-subgroup Q: (notice we are using that $a, b > 0$)

$$1 \subsetneq Q \subsetneq G$$

So Q is a p-group, and so has a non-trivial center: $Z(Q) \neq 1$. Take $h \in Z(Q) \setminus \{e\}$. Hence $Q \leqslant C_G(h)$. , so $[G : C_G(h)] \big| [G : Q] = p^a$, where $[G : C_G(h)] = |\operatorname{conj}(h)$. Now recall we've shown that:

$$\chi_1 \big| |G| = p^a q^b$$

Now, if $p \nmid \chi_i(1)$, then by the previous line we have that $\chi_1(1) \big| q^b$. So, $\gcd(\chi_i(1), |\operatorname{conj}(h)) = 1$, and so by lemma 24.2.18

$$\chi_i(h) = 0 \qquad \text{or} \qquad \chi_i(1)$$

But if $|\chi_i(h)| = \chi_i(1)$, $h \in Z_i$, and so $i = 1$

So, if $p \nmid \chi_i(1)$ and $i > 0$, then $\chi_i(h) = 0$. Using this, we can get our contradiction: as h is not the trivial element, we get:

$$\begin{aligned}
0 &= \sum_i \overline{\chi_i(h)} \chi_i(1) \\
&= \sum_i \chi_i(h) \chi_i(1) \\
&\overset{!}{=} 1 + \sum_{\substack{i > 1 \\ p \big| \chi_i(1)}} \underbrace{\chi_i(1)}_{\in p\mathbb{Z}} \underbrace{\chi_i(h)}_{\in \overline{\mathbb{Z}}^{\mathbb{C}}} \\
&= 1 + p\alpha && \text{some } \alpha \in \overline{\mathbb{Z}}^{\mathbb{C}}
\end{aligned}$$

where the one in the $\overset{!}{=}$ equality came from when $i = 1$. Thus,

$$a = -1/p \in \overline{\mathbb{Z}}^{\mathbb{C}} \cap \mathbb{Q} = \mathbb{Z}$$

a contradiction, completing the proof.

(huge list of character tables: commented)

24.3 Induced Representation

This is another way of constructing representations. We saw that if $H \leqslant G$ and V is a G represetnation, then we can make V an H-representation through the inclusion map $H \hookrightarrow G \to \mathrm{GL}(V)$. We may ask if we may go the other way: given an H repreesentation W, can we find a G-represetnation? If φ is the represetnations afforeded by H, can we find a Φ such that $\Phi|_H = \varphi$? It turns out we cannot: take $A_3 \subseteq S_3$ and consider a faithful representation of degree 1 (recall it's a cyclic group). Then notice that any 1 dimensional representations of S_3 will have to have A_3 in it's kernel, meaning

we can't extend the action of A_3 to an action of S_3. A little more generally, if we let G be a simple group, $H \leqslant G$, and φ a representation of H where $\ker(\varphi)$ is a proper nontrivial subgroup of H, then if we somehow extend the representation φ of H to Φ of G, then $\ker(\Phi)$ would have a proper kernel too, contradicting that G is simple. So instead, we will not be extending representation from H to G, but *inducing* a representation in a natural way. This method of inducing a representation will work even in the two above cases, meaning an induced representation will not always be an extension.

Let $H \leqslant G$, and consider some H-representation V, which can be represented as a $k[H]$-module V. Recall that we can "extend scalars" and define a new $k[G]$-module:

$$k[G] \otimes_{k[H]} V$$

This is (the outer tensor product for representatives, which I haven't talked about yet, but probably should have it here to motivate the idea)

Definition 24.3.1: Induced Representation

Let $H \leqslant G$ of a finite group G and let V be a $k[H]$-module affording the representation φ. Then the $k[G]$-module $k[G] \otimes_{k[H]} V$ is called the *induced module* of V, the representation is called the *induced represetnation*, and if χ is the character of φ, then the character of the induced representation is called the *induced character*, and is written as $\mathrm{Ind}_H^G(\psi)$.

Antoher intuition behind defining the induced module is the follwoing: If W is an H-representation, then let our new module be:

$$\mathrm{Ind}_H^G(W) = \{f \in \mathrm{Hom}_{\mathbf{Set}}(V, W) \mid f(hg) = hf(g),\ \forall g \in G, h \in H\}$$

Then $\mathrm{Ind}_H^G(W)$ is clearly a compelx vector space under pointwise addition and scalar multiplication, and the G-action on this space would be:

$$\gamma \cdot f : V \to W \qquad g \mapsto f(g\gamma)$$

It should be verified that this is indeed G-linear. In this case, we get that:

$$\mathrm{Ind}_H^G(W) = \mathrm{Hom}_{\mathbb{C}[H]}(\mathbb{C}[G], W)$$

where we may use a variant of lemma 24.2.9 to show that this is the same thing as we've just defined. It is also easy to see the dimension of $\mathrm{Ind}_H^G(W)$, namely:

$$\dim_{\mathbb{C}}(\mathrm{Ind}_H^G(W)) = [G : H] \dim_{\mathbb{C}}(W)$$

with the isomorphism being:

$$\mathrm{Ind}_H^G(W) \to W^{\oplus r} \qquad f \mapsto f(g_1), ..., f(g_r)$$

where we treat $G = \bigsqcup_i Hg_i$ for the cosets Hg_i.

Example 24.3.1: Induced Representations

1. $\mathrm{Ind}_G^G(W) = W$
2. $\mathrm{Ind}_{\mathbb{1}}^G(\mathbb{C}) = \mathbb{C}[G]$

3. $\mathrm{Ind}_H^G(\mathbb{1}) = \mathbb{C}[G/H]$

This motivates the following adjoint of the restriction functor:

Theorem 24.3.2: Frobenius Reciprocity

Let $H \leqslant G$. Then if$V$ is a G-representation and W and H-representation, then:

$$\mathrm{Hom}_G(V, \mathrm{Ind}_H^G(W)) \cong_{\mathbb{C}} \mathrm{Hom}_H(V|_H, W)$$

where $V|_H$ is the restriction of the action to H

Proof **:**
We will explicitly construct inverse bijections Φ from left to right and Θ from right to left.
Let $\alpha : \mathrm{Ind}_H^G(W) \to W$, $f \mapsto f(1)$. Then α is H-linear. Then let

$$\Phi(f) = \alpha \circ f$$

Then $\Phi(f)$ is H-linear (check $\mathbb{C}$-linearity of Φ). For the othe way, let $\psi : V|_H \to W$ be H-linear,

$$\Theta(\psi) : V \to \mathrm{Ind}_H^G(W) \qquad v \mapsto (g \mapsto \psi(gv))$$

Then this is well-defined. What's left is to check they are inerses of each other

Corollary 24.3.3: Frobenius Reciprocity On Characters

Let $H \leqslant G$, χ be the character of a G representation and ψ a character of an H-representation. Then:

$$\langle \chi|_H, \psi \rangle_H = \langle \chi, \mathrm{Ind}_H^G(\psi) \rangle_G$$

Proof **:**
`https://unapologetic.wordpress.com/2010/12/03/dual-frobenius-reciprocity/`

Another way of thinking of this result is that if V is an irreducible represetnation of G and W is an irreducible representation of W, then the number of times W occurs in $\mathrm{Ind}_H^G(W)$ is equal to the number of times W occurs in $V|_H$.

Proposition 24.3.4: Properties Of Induced Representation

Let $K \leqslant H \leqslant G$, W be an K-representation, $\mathrm{Ind}_K^H(W)$ the induced H rpresentation, $\mathrm{Ind}_H^G(\mathrm{Ind}_K^H(W))$ the induced G-representation. Then:

1. If χ, χ' are two characters of W,
$$\mathrm{Ind}_K^H(\chi + \chi') = \mathrm{Ind}_K^H(\chi) + \mathrm{Ind}_K^H(\chi')$$
2. Induction is transitive:
$$\mathrm{Ind}_H^G(\mathrm{Ind}_K^H(W)) = \mathrm{Ind}_K^G(W)$$

Proof **:**

Theorem 24.3.5: Computing Induced Representation

Let $H \leqslant G$ of a finite group G, and $g_1, g_2, ..., g_m$ be representatives for distinct left cosets of H in G. Let V be an H-representation of degree n. Then the induced G-representation $W = k[G] \otimes_{k[H]} V$ is an nm degree representation, and if Φ repersented the induced representation, then there exists a basis for W such that W affords the matrix representation defined for each $g \in G$ by:
$$\Phi(g) = \begin{pmatrix} \varphi(g_1^{-1}gg_1) & \cdots & \varphi(g_1^{-1}gg_m) \\ \vdots & \vdots & \cdots \\ \varphi(g_m^{-1}gg_1) & \cdots & \varphi(g_m^{-1}gg_m) \end{pmatrix}$$
where each $\varphi(g_i^{-1}gg_j)$ is an $n \times n$ block, with $\varphi(g_i^{-1}gg_j) = 0$ if $g_i^{-1}gg_j \notin H$.

Proof **:**
First, $k[G]$ is a free right $k[H]$-module with decomposition:
$$k[G] = g_1k[H] \oplus \cdots \oplus g_mk[H]$$
Since tensor products commute with direct sums, we get:
$$W = k[G] \otimes_{k[H]} V \cong (g_1 \otimes V) \oplus \cdots \oplus (g_m \otimes V)$$
Since k is in the center of $k[G]$, the above isomorphism is also a k-vector space isomorphism. Thus, if $v_1, v_2, ..., v_n$ is a basis for V, then $\{g_i \otimes v_j \mid 1 \leq i \leq m,\ 1 \leq j \leq m\}$ is a basis for W, and hence $\dim_k(W) = nm$.

To show $\Phi(g)$ is of the given form, order the basis into m of size n as:
$$g_1 \otimes v_1, g_1 \otimes v_2, ..., g_1 \otimes v_n, g_2 \otimes v_1, ..., g_m \otimes v_n$$
Using this order, we can compute the matrix $\varphi(g)$ for each g acting on W given this basis. Fixing j and g, let $g_j = g_ih$ for some index i and $h \in H$. Then for every k:
$$g(g_j \otimes v_k) = (gg_j \otimes v_k) = g_i \otimes v_k = \sum_{t=1}^{n} a_{tk}(h)(g_i \otimes v_t)$$

where a_{tk} is the t, k coefficient of the matrix h acting on V with respect to the basis $v_1, v_2, ..., v_n$. Notice that the action of g on W maps the jth block of n basis vectors of W to the ith block of basis vectors, and then has the matrix $\varphi(h)$ as that block. Since $h = g_i^{-1} g g_j$, this fully describes $\Phi(g)$, as we sought to show.

Corollary 24.3.6: Computing The Induced Character

Let $H \leqslant G$ where G is a finite group, V be an H-representation with character χ, and $\mathrm{Ind}_H^G(\chi)$ the induced character. Then:

$$\mathrm{Ind}_H^G(\chi) = \sum_{i=1}^{m} \chi(g_i^{-1} g g_i)$$

where $\chi(g_i^{-1} g g_i) := 0$ if $g_i^{-1} g g_i \notin H$. Hence, $\mathrm{Ind}_H^G(\chi)(g) = 0$ if g is not conjugate of some element in H in G, in particular if $H \trianglelefteq G$ then $\mathrm{Ind}_H^G(\chi)$ is zero on all elements $G - H$.

Proposition 24.3.7: Induced Index Two Induced Representation

Let $H \leqslant G$ where $[G : H]$ has index two[a] (ex. $A_n \leqslant S_n$ or $C_n \leqslant D_{2n}$). Then if V is a finite-dimensional irreducible representation of G, either $V|_H$ is irreducible or $V|_H = W_1 \oplus W_2$ where $W_1 \ncong W_2$, are both ireducibile, and have the same dimension. Moreover, the second option implies that $V \cong V \otimes \chi$ where $\chi : G \twoheadrightarrow G/H \cong \{\pm 1\}$ (i.e. a 1-dimensional representation)

[a] Recall that this makes H a normal subgroup

In the case where $V|_H$ is not irreducible, we can imagine it splitting into a diagram like so:

$$\begin{tikzcd}[row sep=small] & W_1 \\ V \arrow[ur, dash] \arrow[dr, dash] & \\ & W_2 \end{tikzcd}$$

while in the case where $V|_H$ is irreducible, what we actually have (and didn't have time to prove), is that we have *two* irreducible representation's of G (V and V') that both get restricted to W:

$$\begin{tikzcd}[row sep=small] V \arrow[dr, dash] & \\ & W \\ V' \arrow[ur, dash] & \end{tikzcd}$$

and then the induced representation of W is the direct sum of V and V'.

***Proof* :**
Let $W \subseteq V|_H$ be an irreducible H-representation. Then Since $H \trianglelefteq G$, gW is is an H-stable space: gW is some subspace of V (it satisfies the subspace axioms since the action is linear), and for any $h \in H$, $h \cdot gW = hgW = g'hW$ that is $gW = hg'h^{-1}W$, and since $hg'h^{-1} \in H$, we have that $h \cdot gW = gW$. Now, gW is irreducible: since W is irreducible and $\varphi_g(W) = gW$ is a G-linear isomorphism (with inverse$\varphi_{g^{-1}}(gW) = W$), gW is irreducible. Finally, the structure of gW only depends on that of the coset gH

Thus, fix $\gamma \in G \setminus H$ so that

$$\sum_{g \in G} gW = W + \gamma W \subseteq V$$

where the $\subseteq$ comes from the fact that this is a sum of two subspaces. Now, notice that the left hand side is a G-stable space, but V is irreducible meaning it does not have a G-stable subspace, and therefore, we actually have the equality:

$$W + \gamma W = V$$

meaning the $H + gH = G$ is parallelled in the action! We know have two possibilities: either $W = \gamma W$, and so $V|_H = W$, and we have $\dim V = \dim W$ (which is case (a)).

on the otherhand, if $W \neq \gamma W$, then $W \cap \gamma W = (0)$ (since they are both irrreducible), and so $V = W \oplus \gamma W$ (i.e. direct sum as H-representations), and so V is the direct sum of two irreducible H-representations.

Now, by Frobeinus reciprocity, we get:

$$\langle W, V|_H \rangle_H = \langle \mathrm{Ind}_H^G(W), V \rangle_G$$

Since W occurs in $V|_H$ by definition, we get that $1 \leq \langle W, V|_H \rangle_H$, while since V is irreducible, meaning by a dimension argument we get $\langle \mathrm{Ind}_H^G(W), V \rangle_G \leq 1$, and hence:

$$1 = \langle W, V|_H \rangle_H = \langle \mathrm{Ind}_H^G(W), V \rangle_G = 1$$

Thus, from the first one, we get that $W \ncong \gamma W$ (or else the left hand side would be equal to 2), and we must have $V = \mathrm{Ind}_H^G(W)$ (or else the right hand side would be equal to something greater than 1)

If part (b) holds, then

$$V \oplus \chi|_H \cong V|_H$$

and

$$V \cong \mathrm{Ind}_H^G(W) \cong V \oplus \chi$$

Conversely, if $V \cong V \oplus \chi$, then $\chi_V(g) = \chi_V(g)\chi(g)$ for all $g \in G$. Thus, by simple arithmetics we see that $\chi(g) = 0$ for all $g \notin H$. And so:

$$\|\chi_{V|_H}\|_H^2 = 2\|\chi_V\|_G^2$$

and so it is a direct sum of irreducibles, as we sought to show.

24.4 *Application to Harmonic Analysis

(Perhaps assume Fourier theory?)

(Talk briefly how this is a special case of the Fourier Transform? How the fourier transform on locally compact topological groups take a function on the group to the function on the dual group, and transforms convolution to multiplication)

25

Representation of Lie Algebra's and Quiver Theory

(Currently, this chapter really focuses on Quiver theory. I will put my exploration of Lie Algebra Representations here)

(This motivation is if you know some differential geometry) Quiver's have their origin in the classification of Lie groups. A Lie group is a group $(G, \cdot)$ G is a *smooth manifold* and $\cdot : G \times G \to G$ is a *smooth map*. The classification of these types of groups (in particular, the continuous ones, since discrete lie groups is just the classification of all groups) is a problem completed by Killing. Out of his classification, there came out 5 exotic Lie groups lying outisde of the other classiffications. Why did these Lie groups come up? The answer of this quesiton might come from a change in persepctive: Any Lie group has a Lie Algebra – the tangent space at the identity of the group which also has an extra operation called the *lie braket* which allows to capture in some of the smooth properties of G into T_eG. These Lie brackets allow us to define a linear map $\mathrm{ad}_x : T_eG \to \mathbb{C}$, $\mathrm{ad}_x(y) = [x, y]$. The image of these operations are called the *ideals* of a lie algebra (not unlike rings, since an ideal $I \subseteq L$ is the set of all elements where for all $x \in L$ and $a \in I$, $[x, a] \in I$). Using this fact, we can ask for the classification of *simple lie algebras*. As an example of one:

$$\mathrm{sl}(3, \mathbb{C}) = \left\{ \begin{matrix} a & b & c \\ d & e & f \\ g & h & i \end{matrix} \middle| \; a + e + i = 0 \right\}$$

which is closed under Lie braked $[A, B] = AB - BA$ (but not closed under multiplication). Now, since these are linear maps, they can be diagonalized. If we take $h_1, ..., h_k \in \mathrm{sl}(3, \mathbb{C})$ to be diagonal elements (elements that are zero everywhere except possibly the diagonal), then $\mathrm{ad}_{h_1}, ..., \mathrm{ad}_{h_k}$ are all simultaneously diagonalizable. This leads us to be able to define *root systems*, which allows for *root decomposition* and *root systems*. This leads to two important theorems for classification of *complex lie algebras*:

(Cartan's Thorems)

1. Up to isomorphism, there is just one complex semisimple Lie algebra for each root system
2. With 5 exceptions, every finite dimensional semisimple Lie algebra over $\mathbb{C}$ is isomorphic to one of the classical Lie Algebras:

$$\mathrm{sl}(n, \mathbb{C}) \quad \mathrm{so}(n, \mathbb{C}) \quad \mathrm{sp}(2n, \mathbb{C})$$

and 5 excpetional Lie algebras, labeled: e_6, e_7, e_8, f_4, g_2

We will be building up to this result for the fisrt bit of this chapter.

25.1 Lie Algebras

(here)

25.2 Quiver's and their Decomposition

The following is just an accumulation of common terms in graph theory that we will need. The one notable exception is that we will rename oriented graphs to *quivers*

Definition 25.2.1: Quiver

A *quiver* Q is an oriented graph (V, E, s, t) where V represents the vertices (the vertex set), E is the set of edges (the "arrows") and $s : E \to V$ with $t : E \to V$ represents the orientation (s for source, t for terget), that is for any $e \in E$, if we have:

$$\overset{i}{\bullet} \xrightarrow{e} \overset{j}{\bullet}$$

then $s(e) = i$ and $t(e) = j$.

Usually, we will abuse notation say $Q = (V, E)$ is a quiver and ommit s and t. Most of the time, we will draw quivers instead of explicitely define them through a 4-tuple or 2-tuple

Example 25.2.1: Quivers

1. If $V = \{1\}$, $E = \{e\}$, and $s(e) = t(e) = 1$, then we can draw this quiver as

$$\begin{tikzcd} 1 \arrow[loop above] \end{tikzcd}$$

2. If $V = \{1, 2\}$, $E = \{e\}$, $s(e) = 1$, and $t(e) = 2$, then we can draw this quiver as

$$1 \xrightarrow{e} 2$$

3. If $V = \{1, 2\}, E = \{e, f\}$, with $s(e) = 1$, $t(e) = 2$ and $s(f) = 2$, $t(f) = 1$, then we can draw this quiver as

$$\begin{tikzcd} 1 \arrow[r, bend left, "e"] & 2 \arrow[l, bend left, "f"] \end{tikzcd}$$

If you are still uncomfortable, with quivers, try to come up with some as a challenge.

Definition 25.2.2: Path and Cycle

Let Q be a quiver.

1. A path γ is a sequence of edges

$$\gamma = e_k e_{k-1} \cdots e_1$$

where $t(e_i) = s(e_{i+1})$. for $1 \leq i \leq k - 1$.

2. If γ is a path where $s(e_1) = a$ and $t(e_k) = b$, then we say that γ is a path from a to b, and we write $s(\gamma) = a$, $t(\gamma) = b$.

3. The number of number of edges in a path is called the *length* of he path.

4. If γ is a path and $s(\gamma) = t(\gamma)$, we call γ a *cycle*.

5. A cycle of length 1 is called a *loop*

Definition 25.2.3: Sink And Source

Let Q be a quiver

1. Let $e \in E$ and $i \in V$. We say that e *contains* i if $s(e) = i$ or $t(e) = i$. We say that e is *incoming at* i if $t(e) = i$ and that e is *outgoing at* i if $s(e) = i$
2. We say a vertex $i \in V$ is a *sink* if all edges containing i are incoming. Intuitvely, this can be visualized as:

$$\begin{tikzcd}
\cdot \arrow[dr] & \cdot \arrow[d] & \cdot \arrow[dl] \\
\cdot \arrow[r] & i & \cdot \arrow[l] \\
\cdot \arrow[ur] & \cdot \arrow[u] & \cdot \arrow[ul]
\end{tikzcd}$$

3. We say a vertex $i \in V$ is a *source* if all edges containing i are outgoing. Intuitvely, this can be visualized as:

$$\begin{tikzcd}
\cdot \arrow[dl] & \cdot & \cdot \arrow[dr] \\
\cdot & i \arrow[l] \arrow[r] & \cdot \\
\cdot \arrow[ur] & \cdot \arrow[d] & \cdot \arrow[dl]
\end{tikzcd}$$

4. Let s be a sink or a source. Then define sQ to be a new quiver where all the arrows at s are reversed.

25.2.1 Quiver Representation and Varieties

Definition 25.2.4: Quiver Representation

Let Q be a quiver. Then a *quiver representation* V *of* Q consists of

1. a vector space $V(i)$ over a fixed field $\mathbb{F}$ for each vertex of Q
2. a linear transofmration $T_e : V(i) \to V(j)$ for each $e \in E$

Let $\mathrm{Rep}_{\mathbb{F}}(Q)$, or $\mathrm{Rep}(Q)$ be the collection of all quiver representations of Q.

Example 25.2.2: Quiver Representations

1. Let

$$1 \xrightarrow{e} 2$$

be a quiver. Then any linear transformation $T : V \to W$ is a representation of this quiver

2. Let

$$\begin{tikzcd}
 & \cdot \arrow[loop above, "f"] \\
1 \arrow[r, "e"] & 2
\end{tikzcd}$$

be a quiver. Then we can take any linear transformation $T : V \to W$ and an identity map $\mathrm{id} : W \to W$. Note that f need not be represented by the identity: any linear transformation from W to itself is valid.

3. Take

$$1 \xrightarrow{e} 2 \xrightarrow{f} 3$$

Then one class of possible representations of this quiver consits of vector-spaces $V(1)$, $V(2)$, $V(3)$ where the sequence is exact at $V(2)$. Another representation can have T_e and T_f be scalar multiplications. Yet another can be the set of vector spaces and linear transformations T_e and T_f where $T_e, T_f \neq \mathrm{id}$ but $T_f \circ T_e = \mathrm{id}$. There are many more possibilities

Since each vertex has a representation, we can capture a notion of dimension of a quiver representation

Definition 25.2.5: Quiver Dimension

Let Q be a quiver and V a representation of Q. Then the *dimension* or *quiver dimension* of V is the tuple that is denoted:

$$\dim_Q V = (\dim V(i))_{i \in I}$$

where each i is a vertex. If there are finitely many vertices, then we can biject them to natural numbers and have:

$$\dim_Q V = (\dim V(1), \dim V(2), ..., \dim V(n))$$

Definition 25.2.6: Quiver Subrepresentation

Let Q be a quiver and V be a quiver representation. Then a *subrepresentation* W of V consists of a subspace $W(i) \subseteq V(i)$ for each $i \in V$ and a linear transformations restricted to the new domain, $T_e|_{W(i)} = S_e$ for each $e \in E$, such that each S_e, $S_e(W_i) \subseteq W_j$

the last condition is so that each T_e is still well-defined. As an exercise, see if you can come up with quivers and subrepresentations (you can keep it simple and consider a quiver $\bullet \to \bullet$).

Definition 25.2.7: Quiver Morphisms

Let Q be a quiver and let V and V be representations of Q. Then we say that $\varphi : V \to W$ is a *quiver morphism* if for every edge e with representation $T_e : V(i) \to V(j)$ in V and representation $S_e : W(i) \to W(j)$ in W, there exists linear transformations $\varphi(i)$ and $\varphi(j)$ such that the following diagram commutes:

$$\begin{tikzcd} V(i) \arrow[r, "T_e"] \arrow[d, "\varphi(i)"] & V(j) \arrow[d, "\varphi(j)"] \\ W(i) \arrow[r, "S_e"] & W(j) \end{tikzcd}$$

collection of all quiver morphisms between V and W is denoted $\mathrm{Hom}_{\mathbf{Q}}(V, W)$ or $\mathrm{Hom}(V, W)$ if it is unambiguous in context.

Proposition 25.2.8: Structure Of $\mathrm{Hom}(V, W)$

Let $V, W \in \mathrm{Rep}(Q)$. Then $\mathrm{Hom}(V, W)$ forms a vector-space structure over the base field $\mathbb{F}$.

Proof **:**
left as an exercise

Example 25.2.3: Quiver Morphism

1. Let $V = \overset{\mathbb{F}^1}{\bullet} \xrightarrow{1} \overset{\mathbb{F}^1}{\bullet}$ and $W = \overset{\mathbb{F}^1}{\bullet} \xrightarrow{0} \overset{\{0\}}{\bullet}$ be two quiver representations. Then we can define a morphism:
$$\begin{tikzcd} V \arrow[d, "\varphi"] \\ W \end{tikzcd} \qquad \begin{tikzcd} \mathbb{F} \arrow[r, "1"] \arrow[d, "1"] & \mathbb{F} \arrow[d, "0"] \\ \mathbb{F} \arrow[r, "0"] & \{0\} \end{tikzcd}$$
Does there exist a nonzero morphism $\varphi : W \to V$?

2. Let $V = \overset{\mathbb{R}}{\bullet} \xrightarrow{1} \overset{\mathbb{R}}{\bullet} \xrightarrow{1} \overset{\mathbb{R}}{\bullet}$ and $W = \overset{\mathbb{R}^2}{\bullet} \xrightarrow{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}} \overset{\mathbb{R}^2}{\bullet} \xrightarrow{\begin{pmatrix} 1 & 1 \end{pmatrix}} \overset{\mathbb{R}}{\bullet}$. Then: (commented to compile) is a quiver morphism. Does there exist a nonzero morphism $\varphi : W \to V$?

What we have technically defined is a *quiver representation morphism*. A quiver morphisms would usually be between two direct graphs where if $Q = (V, E, s, t)$ and $R = (V', E', s', t')$ are two quivers, then a quiver morphism would consit two functions $m_v : V \to V'$ and $m_e : E \to E'$ such that $m_v \circ s = s' \circ m_e$ and $m_v \circ t = t' \circ m_e$. This is ostensibly the same concept with different objects. All mosrphisms between Q and R would be represented as $\mathrm{Hom}_{\mathbf{Quiv}}(Q, R)$. Most of the time, we will care about the representation of quivers, and hence leaving the name of "quiver morphism" to representations. In either case, the collection of all quivers along with the just-defined quiver morphism defines a category **Quiv**, and the collection of all representations of a quiver $\mathrm{Rep}(Q)$ along with the previously-defined quiver morphism defines a category **Quiv(Q)** (you can check these as exercises)

The category **Quiv(Q)** has a a 0, namely the representation where all vertices are assigned the zero vector-sapces and the linear transformations are the identity. Since 0 exists, we can define kernels. Images also exist in this category:

Definition 25.2.9: Kernels And Images

Let $\varphi : V \to W$ be a quiver morphism. Then $\ker(\varphi)$ is the subrepresetnation of V consisting of $\ker(\varphi(i)) \subseteq V(i)$, and $\mathrm{im}(\varphi)$ is the subrepresentation of W consisting of $\mathrm{im}(V(i)) \subseteq W_i$

Note that $\ker(\varphi)$ and $\mathrm{im}(\varphi)$ are indeed subrepresentations: the fact that the vertices exist is a direct consequence of linear algebra. It remains to show that the diagram still commutes.

Definition 25.2.10: Quiver Isomorphisms

Let $V, W \in \mathrm{Rep}(Q)$. Then V is isomorphic to W if there exists a quiver morphism $\varphi : V \to W$ and $\psi : W \to V$. We denote two isomorphic quivers as $V \cong W$

Example 25.2.4: Quiver Isomorphism
In Jamail's notes. Some pretty cool ones. In parituclar, the classification of quiver representations of the form $\bullet \xrightarrow{e} \bullet$ with each vector-space being finite dimensional.

Next, we define the dual of sub-representations. Since for every subspace of a vector space $W \subseteq V$ we can define V/W, the notion of a quotient quiver is defined similarly

Definition 25.2.11: Quotient Quiver

Let $V \in \mathrm{Rep}(Q)$ and $W \subseteq V$ a subrepresentation. Then V/W is defined as a new quiver where every vertex of V is replaced by the quotient of $V(i)/W(i)$.

Proposition 25.2.12: 1st Isomorphism for Quivers

Let $V, W \in \mathrm{Rep}(Q)$ and $\varphi : V \to W$ be a quiver morphism. Then:

$$\mathrm{im}(\varphi) = V/\ker(\varphi)$$

Proof **:**
left as an exercise

Definition 25.2.13: Product Quiver

Let $V, W \in \mathrm{Rep}(Q)$. Define $V \oplus W \in \mathrm{Rep}(Q)$ as

1. $(V \oplus W)(i) = V(i) \oplus W(i)$
2. $T_e : (V \oplus W)(i) \to (V \oplus W)(j)$ is given by $T_e = T_{e,v} \oplus T_{e,w}$ (recall that $T \oplus S(a, b) = (T(a), S(b))$)

Definition 25.2.14: Quiver Variety

Let $Q = (I, E, s, t)$ be a finite quiver. and $\underline{d} \in \mathbb{N}^n$. Then a *quiver variety* is:

$$\mathrm{Rep}(\underline{d}, Q) := \prod_{e \in E} M_{d_j, d_i}(\mathbb{F})$$

where $i = s(e)$ and $j = t(e)$. Every point in $\mathrm{Rep}(\underline{d}, Q)$ is a representation of Q with dimension vector $\underline{d}$.

A quiver variety let's us reduce a quiver where each vertex can have a different dimension (i.e. any possible total dimension) to only thinking about quivers with the same total dimension

Proposition 25.2.15: Action On Quiver Variety

Let $Q = (I, E, s, t)$ be a finite quiver with $I = \{1, ..., n\}$ and $E = \{e_1, ..., e_n\}$. Let $\underline{d} = (d_1, ..., d_n) \in \mathbb{N}^n$, and define the group $G(\underline{d}) = \prod_i GL_{d_i}(\mathbb{F})$. Then define the $G(\underline{d})$ action on $\text{Rep}(Q, \underline{d})$ as follows:

$$g = (\varphi_1, ..., \varphi_n)$$
$$(g \cdot V)_e = (\varphi_j A_e \varphi_i^{-1})_e$$

where $s(e) = i$ and $t(e) = j$.

Proof **:**
Clearly, if $g = (I)_e$ for each edge, then

$$(g \cdot V)_e = (I_j A_e I_i)_e = (A_e)_e$$

showing that $I \cdot V = V$. Associativity comes for free from assocaitivty of composition of linear maps.

I think the group in the theorme is called a *Weyl Group*.

Theorem 25.2.16: Isomorphism Equivalence

Let $\text{Rep}(Q, \underline{d})$ be a quiver variety and $G(\underline{d})$ be the group as defined in proposition 25.2.15. Then two representations $V, W \in \text{Rep}(Q, \underline{d})$ are isomorphic if and only if there exists a $g \in G(\underline{d})$ such that $g \cdot V$. That is, two quivers are isomorphic if they lie in the same orbit:

$$\{G(\underline{d})\text{-orbit of } \text{Rep}(Q, \underline{d})\} \rightleftarrows \{\text{iso classes of rep with dim } \underline{d}\}$$

Proof **:**
As we've seen in example 25.2.4, two quiver represenations (of the same dimension) of $\bullet \xrightarrow{e} \bullet$ are equivalent if T_e and S_e are equivalent, i.e., their range is of the same dimension, i.e. $\varphi_j A_e \varphi_i^{-1} = B_e$ for two invertible matrices φ_i and φ_j. However, this is exactly the condition for the orbits of this group-actions.

25.2.2 Decomposing Quivers

Similarly to all other algebraic structures we've encountered, we have a notion of simple objects:

Definition 25.2.17: Simple Representation

Let Q be a quiver. Then $V \in \text{Rep}(Q)$ is called a *simple* or *irreducible* representation if the only subrepresentations of V are $\{0\}$ and V

The notion of simple objects is essentially trivial for vectorspaces since $V \cong \mathbb{F}^n$. In other words, every vector-space is decomposable into simple vector-spaces. The analogous notion of decomposability exists for quivers, however, they can be more complicated for quivers

Definition 25.2.18: Indecomposable Quiver

Let Q be a quiver. Then $V \in \mathrm{Rep}(V)$ said to be *decomposable* if there exists non-trival subrepresentation $V_1, V_2 \subseteq V$ such that:

$$V \cong V_1 \oplus V_2$$

and is said to be *indecomposable* otherwise

Example 25.2.5: Simple Quivers and decomposability

Let $Q = \bullet \xrightarrow{e} \bullet$. Then:

$$0 \xrightarrow{0} \mathbb{F} \qquad \mathbb{F} \xrightarrow{0} 0$$

are both simple representations. Note that the left quivers is a subrepresentations of $V = \mathbb{F} \xrightarrow{\mathrm{id}} \mathbb{F}$, and along with $\mathbb{F} \xrightarrow{\mathrm{id}} \mathbb{F}$ and $0 \xrightarrow{0} 0$, $\mathbb{F} \xrightarrow{\mathrm{id}} \mathbb{F}$ has 3 sub-quivers. Thus, V is not a simple quiver.

However, V is *indecomposable*. To see this, notice that if V were decomposable, then we get:

$$\dim_Q(V_1 \oplus V_2) = \dim_Q(V_1) + \dim_Q(V_2)$$

and for the representation to be nontrival, we would have that one representation would have dimension $(1, 0)$, while the other has $(0, 1)$. However,V has no representation of dimension $(1, 0)$.

This means that the extension of quivers will be more elaborate than split exact sequences.

Theorem 25.2.19: Quiver Representation Decomposition

Let Q be a quiver and V be a representaiton where $\sum_i \dim V(i) < \infty$. Then there exists a finite set of indecomposable subrepesentations I_1, ..., I_n such that

$$V = I_1 \oplus I_2 \oplus \cdots \oplus I_n$$

***Proof* :**

Define the *total dimension* $\mathrm{Tdim}(V) = \sum_i \dim V(i)$. If $\mathrm{Tdim}(V) = 1$, then V must be only contain one vector-space $\mathbb{F}$ with all other vecor-spcaes being the trivial vector spaces. Assume for a moment that V has no oriented cycles. Then all maps from edges that are being the trivial map. Clearly, V is indecomposable, and so it is it's own decomposition.

On the other hand, if V has an oriented cycle, then without loss of generality we can consider

$$T = T_n \circ T_{n-1} \circ \cdots \circ T_1$$

so that we have $T : \mathbb{F} \to \mathbb{F}$. Since $\mathbb{F}$ has no non-trival subspace (being an indecomposable vector-space over $\mathbb{F}$), V is still indecomposable, and so is it's own decomposition.

Using strong induction, assume that all vector-spaces of total dimension $\mathrm{Tdim}(V) = n$ can be decomposed into a direct sum of indecomposable representations, and consider a representation of total dimension $n+1$. If V is indecomposable, then we're done, so let's assume V is decomposableso

that there exists two non-trival subrepresentations V_1, V_2 such that

$$V = V_1 \oplus V_2$$

Then since V_1, V_2 are both non-trivial, their total dimension must be strictly greater than and strictly less than that of V and nonzero. But then, by the inductive hypothesis, V_1 and V_2 can de decomposed into a finite direct sum of indecomposable modules. Hence:

$$V = V_1 \oplus V_2 = (I_{1,1} \oplus \cdots \oplus I_{1,n_1}) \oplus (I_{2,1} \oplus \cdots \oplus I_{2,n_2})$$

completing the proof.

Example 25.2.6: Decomposing Representations

1. Let $Q = \bullet \xrightarrow{e} \bullet$ be a quiver. Find all simple and indecomposable representations of Q. Using this knowledge, decompose any representation $V \xrightarrow{T_e} W$.
2. Let $Q = \bullet \xrightarrow{e} \bullet \xrightarrow{f} \bullet$. Find all simple and indecomposable representation of Q. Using this knoweldge, decompose any representation $V \xrightarrow{T_e} W \xrightarrow{T_f} L$.
3. Let $Q = \begin{tikzcd} \bullet \arrow[loop above] \end{tikzcd}$ be a quiver. Given $\mathbb{F}$ is algebraically closed, find all simple and indecomposable representations of Q (this is heavily reliant on Jordan canonical forms from chapter 14). Using this knowledge, decompose any representation $T : V \to V$.

25.2.3 Root Systems

Recall that if $(V, \langle \cdot, \cdot \rangle)$ ia an inner product space and $v \in V$, then

$$\frac{\langle w, v \rangle}{\langle v, v \rangle} v$$

is the projection of w onto v. Furthermore, define:

$$s_v(w) = w - 2\frac{\langle w, v \rangle}{\langle v, v \rangle} v$$

which reflects w across the hyperplane orthogonal to v.

Definition 25.2.20: Root System

Let V be an inner product space over $\mathbb{R}$. Then a *(reduced) root system* $R \subseteq V$ is a finite collection of non-zero vectors such that

1. $\text{span}(R) = V$
2. For all $\alpha, \beta \in R$, $\eta_{\alpha,\beta} = \frac{2\langle \alpha, \beta \rangle}{\langle \beta, \beta \rangle} \in \mathbb{Z}$. In terms of angles, recall that $\theta = \cos^{-1}\left(\frac{\langle \alpha, \beta \rangle}{\|\alpha\|\|\beta\|}\right)$, so $n_{\alpha\beta} = \|\alpha\| \cos(\theta)$
3. For each $\alpha \in R$, the corresponding reflection $s_\alpha : V \to V$ has the property that for all $\beta \in R$, $s_\alpha(\beta) \in R$, i.e., R is symmetric under all α reflections.
4. If both $\alpha, c\alpha \in R$ then $c = \pm 1$ (this is true for reduced systems only)

Example 25.2.7: Root Systems

Let $V = \{(x, y, z) \in V \mid x + y + z = 0\}$, where V inherits the usual inner product from $\mathbb{R}^3$. Let

$$R = \{\pm(e_1 - e_2), \pm(e_2 - e_3), \pm(e_1, e_3)\} = \{\alpha, \beta, \gamma\}$$

Then R is a reduced root system. This set spans V because HERE.

Furthermore:

$$\eta_{\pm\alpha,\pm\beta} = \frac{2\langle \pm(e_1 - e_2), \pm(e_2 - e_3) \rangle}{\pm\langle (e_2 - e_3), \pm(e_2 - e_3) \rangle}$$

(finish rest later)

Include these

Notice that any root system can be simplified:

Definition 25.2.21: Polarization And Simple

Let R be a root sytem. Then if there exists a vector v such that $\langle v, \alpha \rangle \neq 0$ for all $v \in R$ then $R_+ = \{\alpha \in R \mid \langle v, \alpha \rangle > 0\}$ along with $R_+ = \{\alpha \in R \mid \langle v, \alpha \rangle < 0\}$ is called the *polarization of* R, and $R = R_+ \sqcup R_-$. The elements of R_+ are called *positive roots*. A positive root is called *simple* if it cannot be expressed as the sume of other positive roots.

Example 25.2.8: Simple Roots

homework, exercise 8 in Jimails notes

Theorem 25.2.22: Simple Root Decomposition

Let R, V be a (reduced) root system with polarization $R = R_+ \sqcup R_-$. Let Π denote the set of simple roots for this polarization. Then:

1. Every $\alpha \in R_+$ is a sum of simple roots
2. If $\alpha, \beta \in R_+$, then $\langle \alpha, \beta \rangle \leq 0$
3. The set Π of simple roots froms a basis for V

Proof **:**
exercise 9

(words about root systems being isomorphic, which I'm guessing is a map from one vec sp from another s.t. all the properties of root systems are preserved)

Definition 25.2.23: Cartan Matrix

Let R, V be a (reduced) root system with polarization $R = R_+ \sqcup R_-$, let Π denote the set of simple roots for this polarization, and let $r = |\Pi|$. Then define *Cartan matrix* A to be the $r \times r$ matrix defined by:

$$A_{ij} = n_{\alpha_j \alpha_i} = \frac{2\langle \alpha_i, \alpha_j}{\alpha_i, \alpha_i}$$

Example 25.2.9: Cartan Matrix
exercise 10

(word to motivate the following visualization)

Definition 25.2.24: Dynkin Diagrams

Let R, V be a (reduced) root system with polarization $R = R_+ \sqcup R_-$, let Π denote the set of simple roots for this polarization, and identify Π with the set $\{1, ..., k\}$. Then we can construct a graph following these steps:

1. The vertices are given by simple roots Π (or equivalently, by indices $i \in \{1, ..., k\}$)
2. We join a vertex i and j by n_{ij} edges, which depends on the angle θ where

$$\theta = \frac{\pi}{2} \Rightarrow n_{ij} = 0 \qquad \theta = \frac{2\pi}{3} \Rightarrow n_{ij} = 1 \qquad \theta = \frac{3\pi}{4} \Rightarrow n_{ij} = 2 \qquad \theta = \frac{5\pi}{6} \Rightarrow n_{ij} = 3$$

3. Roots joined by a single edge have the same length. Roots joined by muliple edges have an arrowhead ">" on the edge pointing towards the root with sorther length (imagine this as an inequality between the length of the roots)

Example 25.2.10: Dynkin Diagrams

1. finding dynkin diagram for many root systems

We know proceed to classify root systems. As with other definition, if R is a root system and $R = R_1 \sqcup R_2$ where R_1, R_2 are proper subsets of R that are root systems, we say that R is reducible, and irreducible otherwise.

Theorem 25.2.25: Classification Of Root Systems

Let R be a (reduced) irreducible root system. Then it's dynkin Diagram is one of the following:

Proof **:**
here

(representing root systems via it's symmetries)

Definition 25.2.26: Weyl Group

Let $R \subseteq V$ be a root system. Let $W \subseteq \mathrm{GL}(V)$ be the group generated by the reflections s_α, $\alpha \in R$. Then W is called the *Weyl group* of R.

Since it is a group generated by reflections, every element acts linearly on V. Furthermore, it perserves lengths and angles, and so it is an element of $O(V)$ (definition 1.2.22)

Example 25.2.11: Weyl Group
p. 11 Jimail

Definition 25.2.27: Simple Reflection

Let $R \subseteq V$ be a root system with Weyl group W, and $R = R_+ \sqcup R_-$ is a polarizatoin with simple roots Π.
Then for each $\alpha_i \in \Pi$, the corresponding reflection $s_i = s_{\alpha_i} \in W$ is called a *simple reflection*.

Theorem 25.2.28: Weyl Groups Representing Root Systems

Let $R \subseteq V$ be a root system with Weyl group W, and $R = R_+ \sqcup R_-$ a polarization with simple roots Π. Then:

1. The simple roots s_i (for $i \in \Pi$) generate W
2. $W(\Pi) = R$. Hence, the root system can be recovered from the simple roots;we apply all Weyl group elements to Π to produce the full list of possible roots

Proof **:**
exercise?

(reduced expressions and longest elements)

Definition 25.2.29: Length Of Weyl Group Elements

Let W be the Weyl group of a root system R, and Π the simple roots. Then the *length* of $w \in W$ is the shortest length of word for w as a product of simple reflections. We denote the length of w by $l(w)$.

In other words, if $w = s_{i_\ell} \cdots s_{i_1}$ and no shorts sequence exists, then $l(w) = \ell$.

Theorem 25.2.30: Generating Roots Using Weyl Group

Let W be the Weyl group of a root system R with polarization $R = R_+ \sqcup R_-$ and Π the corresponding simple roots.

1. There exists a unique element $w_0 \in W$ such that $\ell(w_0) = |R_+|$
2. The element w_0 has the property that $w_0(R_+) = R_-$
3. If $w = s_{i_k} \cdots s_{i_1}$ is an expression for w_0 of length $l(w_0)$, then

$$\begin{aligned} \gamma_1 &= \alpha_{i_1} \\ \gamma_2 &= s_{i_1}(\alpha_{i_2}) \\ &\vdots \\ \gamma_k &= s_{i_k} \cdots s_{i_1}(\alpha_{i_k}) \end{aligned}$$

   is an enumeration of the positive roots

4. The elements γ_j defined above satisfy the propety that the roots γ_j, $s_{i_1}(\gamma_j)$, ..., $s_{i_1}(\gamma_j) \in R_+$ are positive, but the roots $s_{i_j} s_{i_{j-1}} \cdots s_{i_1}(\gamma_j) \in R_-$ are negative.

Proof **:**
Jimail didn't prove – we don't have enough tools yet

Now that we have a better undersatnding of Weyl groups, we seek to link some of their structure to structure of simple roots. As we saw, the relations between simple roots can be captured via Dynkin Diagrams. We will seek for elements of the Weyl group which are "compatible" with the Dynkin diagrams. (for some reaosn), let Γ be a dynkin diagram of type A, D, or E (i.e. a Dynkin diagram with no multiple edges). Let R denote the corresponding root system. For any orientation Ω of Γ, Let (γ, Ω) represent the correspondign quiver. Recall that if i is a sink or source in Ω, we have an operator $s_i^{\pm}$ that reverses the arrows at i. Denote thie operation by

$$(\Gamma, \Omega) \mapsto (\Gamma, s_i^{\pm}\Omega)$$

where we would put a $+$ if it is a sink-to-source operation, and a $-$ if it is a source-to-sink operation

Definition 25.2.31: Adapted Word

Let Γ be a Dynkin diagram of type A, D, or E (i.e. a diagram with no multiple edges, or "simply laced"), W the corresponding Weyl group, and Ω an orientation of Γ. Let $w \in W$. Then the reduced expression $w = s_{i_k} \cdots s_{i_1}$ for w is *adapted for* Ω if

1. The verex i_1 is a sink for (Γ, Ω)
2. The verex i_2 is a sink for $(\Gamma, s_{i_1}^{+}\Omega)$

   $\vdots$

3. The verex i_k is a sink for $(\Gamma, s_{i_{k-1}}^{+} \cdots s_{i_1}^{+}\Omega)$

i.e. at each step, we can apply the operator $s_{i_j}^{+}$ to the quiver $(\Gamma, s_{i_{j-1}}^{+} \cdots s_{i_1}^{+}\Omega)$

Example 25.2.12: Adapted Dynkin Diagrams
p. 15 Jaimal

Theorem 25.2.32: Lusztig's Theorem

Let R be an A, D, or E root system with dynkin diagram Γ and Weyl group W. For any orientation Ω of Γ, there exists an expession for the longest element $w_0 \in W$ which is adapted for Ω

Proof **:**
exercise 19

25.2.4 Gabriel's Theorem

In this section, we come back to Quiver's. Recall in example 25.2.6 it was asked to show how find the indecomposable quivers with underlying graphs of type A_1, A_2, and a cyclic quiver. The result was that A_2 and A_3 (with the given orientation) reduced to finitely many indecomposable representations. However, if there is a cycle, the quiver did not afford such a representation. A quiver which has finite decomposition is naturally much easier to study (TBD). The question then becomes:

> Which quivers Q can be decomposed into finitely many indecomposable representations? Can we classify all such quivers?

The answer of this question is *Gabriel's Theorem*; the goal of this section is to prove this result. Gabriel's Theorem is the key result linking the representation of quivers to Lie algebras, a pivotal tool in studying manifolds with group structures.

(a,b) To make the connection more precise, some terms need to be defined:

Definition 25.2.33: Collection Of Indecomposables

Let Q be a quiver, and let $\mathrm{Rep}(Q)$ be it's representations. Then the set $\mathrm{Ind}(Q) = \{[X] \mid X \in \mathrm{Rep}(Q) \text{ is indecomposable } \}$ is the set of isomrophism classes of indecomposable quivers of $\mathrm{Rep}(Q)$.
We say that a quiver is of *finite type* if $\mathrm{Ind}(Q)$ is finite.

Before continuing with proving he result, it is useful to at least see how such a connection might even be hypothesized:

Example 25.2.13: Motivating Connection

Consider the quiver $\overset{1}{\bullet} \longrightarrow \overset{2}{\bullet} \longrightarrow \overset{3}{\bullet}$. The underlying graph is A_3, so let α_1, α_2, and α_3 denote the simple roots for the root system A_3. Using these, we can deduce the positive roots are:

$$\{\alpha_1, \alpha_2, \alpha_3, \alpha_1 + \alpha_2, \alpha_2 + \alpha_3, \alpha_1 + \alpha_2 + \alpha_3\}$$

We furthermore saw in example 25.2.6 that the indecomposable representations are:

$$\mathbb{F} \xrightarrow{0} 0 \xrightarrow{0} 0 \qquad \mathbb{F} \xrightarrow{1} \mathbb{F} \xrightarrow{0} 0$$

$$0 \xrightarrow{0} \mathbb{F} \xrightarrow{0} 0 \qquad 0 \xrightarrow{0} \mathbb{F} \xrightarrow{1} \mathbb{F}$$

$$0 \xrightarrow{0} 0 \xrightarrow{0} \mathbb{F} \qquad \mathbb{F} \xrightarrow{1} \mathbb{F} \xrightarrow{1} \mathbb{F}$$

Here is the key insight: notice we can correspond the *dimension vector* of these repesenations to the positive roots!

Theorem 25.2.34: Gabriel's Theorem

Let Q be a quiver. then Q is of finite type if and only if the underlying graph of Q is a Dynkin Diagram of type A, D, or E. In this case, the dimension vector:

$$\dim_Q : \mathrm{Ind}(Q) \to R_+ \qquad [I] \mapsto \dim_Q(I)$$

bijects the indecomposable representations (up to isomorphism) and positive roots in the corresponding root system.

The proof will be relegated to page pageref:HERE

The theorem hints to us that indecomposable quivers somehow relate to the simple reflections of a Weyl group. However, a Dynkin Diagram is inherently unoriented, while Quivers are by definition an oriented graph. The first thing we need to do to fix is find the equivalent of a "reflection" for quivers.

The idea for the following definition is as follows: Take the quiver and the reflection

$$\overset{1}{\bullet} \xrightarrow{e} \overset{2}{\bullet} \qquad \overset{1}{\bullet} \xleftarrow{f} \overset{2}{\bullet}$$

Now, given a representation assigning $V(1)$, $V(2)$, and T_e, how should we assign values to the reflection? One way of doing it is by taking $T_e : V(1) \to V(2)$, and mapping it to $S_f : \ker(T_e) \to V(1)$ where $S_f(x) = 0$. Visually, we see that:

$$\left(\overset{V(1)}{\bullet} \xrightarrow{T_e} \overset{V(2)}{\bullet} \right) \mapsto \left(\overset{V(1)}{\bullet} \xleftarrow{S_f} \overset{\ker(T_e)}{\bullet} \right)$$

this map considers what happens at a sink. Another approach maps by considering the source:

$$\left(\overset{V(1)}{\bullet} \xrightarrow{T_e} \overset{V(2)}{\bullet} \right) \mapsto \left(\overset{\operatorname{coker}(T_e)}{\bullet} \xleftarrow{S_f} \overset{V(2)}{\bullet} \right)$$

hwere $S_f(x) = \overline{T_e(x)}$. Both of these asignements are in fact functorial, If we have a quiver morphism $\varphi : V \to W$ represented by:

$$\begin{tikzcd} V(1) \arrow[r, "T_e"] \arrow[d, "\varphi(1)"] & V(2) \arrow[d, "\varphi(2)"] \\ W(1) \arrow[r, "S_e"] & W(2) \end{tikzcd}$$

Then the appropriate morphim φ maps to is represented in these diagrams:

$$\begin{tikzcd} V(1) \arrow[d, "\varphi(1)"] & \ker(T_e) \arrow[l, "T_f|_{\ker(T_e)}"'] \arrow[d, "\varphi(1)|_{\ker(T_e)}"] \\ W(1) & \ker(S_e) \arrow[l, "S_f|_{\ker(S_e)}"'] \end{tikzcd} \qquad \begin{tikzcd} \operatorname{coker}(T_e) \arrow[d, "\overline{\varphi(2)}"] & V(2) \arrow[l, "\overline{T_f}"'] \arrow[d, "\varphi(2)"] \\ \operatorname{coker}(S_e) & W(2) \arrow[l, "\overline{S_f}"'] \end{tikzcd}$$

which both commute (where it was shown that $\varphi|_{\ker(T_e)}(\ker(T_e)) \subseteq \ker(S_e)$ when we showed the cateogry of quivers contains kernels). This will essentially represent how a reflection functor between two quivers will be defined – locally.

The different between these two functors is where there is a sink or a source. If i is a sink, then for any given quiver representatin $V \in \operatorname{Rep}(Q)$, let

$$\Psi_i : \bigoplus_{e:i\to j} V(j) \to V(i)$$

represent the collection of all linear transformation, where for each summand we have $T_e : V(i) \to V(j)$. If i is a source, then let:

$$\Phi_i : V(i) \to \bigoplus_{e:i\to j} V(j)$$

where each summand is mapped to T_e.

Definition 25.2.35: Reflection Functor

Let Q be a quiver. Define a functor from $\mathrm{Rep}(Q)$ to $\mathrm{Rep}(s_i^{\pm}Q)$ as follows:

1. For any $V \in \mathrm{Rep}(Q)$, define $S(V)_i^+ \in \mathrm{Rep}(s_i^+(Q))$ for a sink i, where at each vertex k,

$$S_i^+(V)(k) = \begin{cases} V(k) & i \neq k \\ \ker(\Psi_i) & i = k \end{cases}$$

and for each linear transformation T_e:

$$S_i^+(T_e) = \begin{cases} T_e & s(k) \neq i \\ \pi_k \circ \iota_i & \text{if } f : i \to k \text{ is in } s_i^+(Q) \end{cases}$$

where $\iota_i : \ker(\Psi) \to \bigoplus_{e:j\to i} V(j)$ is the inclusion map and $\pi_k : \bigoplus_{e:j\to i} V(j) \to V(k)$ is the projection map onto the kth component.

2. For any $V \in \mathrm{Rep}(Q)$, define $S(V)_i^- \in \mathrm{Rep}(s_i^n(Q))$ for a source i, where at each vertex k,

$$S_i^+(V)(k) = \begin{cases} V(k) & i \neq k \\ \mathrm{coker}(\Phi_i) & i = k \end{cases}$$

and for each linear transformation T_e:

$$S_i^+(T_e) = \begin{cases} T_e & s(k) \neq i \\ q_k \circ \iota_i & \text{if } f : i \to k \text{ is in } s_i^+(Q) \end{cases}$$

where $\iota_i : V(k) \to \bigoplus_{e:j\to i} V(j)$ is the inclusion map and $q : \bigoplus_{e:j\to i} V(j) \to \mathrm{coker}(\Phi_i)$ is the quotient map

Then these two functors, $S_i^{\pm} : \mathrm{Rep}(Q) \to \mathrm{Rep}(s_i^{\pm}Q)$ are called the *Bernstein-Gelfand-Ponomarev reflection functor*, or *reflection functor* for short.

Example 25.2.14: Reflection Functor

p. 70 Jaimal. The key idea is that these functors give different quiver representations, even if it on A_2 with the only orientation and doing S_1^+ then S_2^-.

A couple of properties of the functors needs to be established. For notational simplicity, let $\mathrm{Rep}^{i,-}(Q)$ denote the full subcategory consisting of representations of Q for which the map

$$\Psi_i : \bigoplus_{e:j\to i} V(j) \to V(i)$$

is surjective, while $\mathrm{Rep}^{i,+}(Q)$ denotes the full subcategory consisting of representations of Q for which the map

$$\Phi_i : V(i) \to \bigoplus_{e:i\to j} V(j)$$

is injective.

Proposition 25.2.36: Properties Of Reflection Functor

Let Q be a quiver.

1. If i is a sink, then for any $V \in \mathrm{Rep}(Q)$, $S_i^+ \in \mathrm{Rep}^{i,+}(s_i^+(Q))$ (similarly for source)
2. If i a sink, then for any $V \in \mathrm{Rep}(Q)$, $S_j^- S_i^+ \in \mathrm{Rep}^{i,-}(Q)$ (similarly for source)
3. The restriction of the reflection functors are inverse equivalences (adjoints?)

$$S_i^+ : \mathrm{Rep}^{i,-}(Q) \to \mathrm{Rep}^{i,+}(s_i^+ Q)$$
$$S_i^- : \mathrm{Rep}^{i,+}(Q) \to \mathrm{Rep}^{i,-}(s_i^- Q)$$

4. If V is simple at vertex i, then $S_i^{\pm}(V) = 0$.
5. If V is a non-zero indecomposable representation of Q that is not simple at i, then $S_i^{\pm}(V)$ is a nonzero indecomposable representation of $s_i^{\pm}(Q)$, and $V \cong S_j^{\mp} S_i^{\pm}(V)$.
6. $S_i^{\pm}(V)$ is a nonzero indecomposable representation of $s_i^{\pm}(V)$ if and only if $s_i(\dim_Q(V)) \geq 0$
7. If $V \in \mathrm{Rep}^{i,\pm}(Q)$ is a nonzero indecomposable representation, then $\dim_Q(S_i^{\pm}(V)) = s_i \dim_Q(V)$

Proof **:**
exercise

Finally, we define a new object with purpose HERE (Apparently, these definitions can be given if we use some homological algebra. Perhaps revamp this after you do the homological algebra section, or keep it at this level with the mention of the version of greater generality for those who want to skip Hom Alg)

Definition 25.2.37: Root Lattice

Let Q be a quiver with vertex set I. Then if the underlying graph of Q is a dynkin diagram, $\mathbb{Z}^I$ is called the *root lattice*. For $(v_i)_{i \in I} \in \mathbb{Z}^I$, we say that $v \geq 0$ if $v_i \geq 0$ for all $i \in I$. Furthermore, we let $\mathbb{Z})+^I := \left\{ v \in \mathbb{Z}^I \;\middle|\; v \geq 0 \right\}$

(The following theorem seems to make more sense with some homological algebra, where the domain of the function should actually be the Grothendick group of a $\mathrm{Rep}(Q)$) (Are objects in the following theorem representations?)

Theorem 25.2.38: Root Lattice Map

here exists a map $\dim_Q : \{\text{iso classes of Rep's}\} \to \mathbb{Z}_+^I$

***Proof* :**
not proved

Definition 25.2.39: Euler And Tits Form

Let $Q = (I, E)$ be a quiver. Define a bilinar form on $\mathbb{Z}^I$ as follows: for any $v, w \in \mathbb{Z}^I$:

$$\langle v, w \rangle_Q = \sum_i v_i w_i - \sum_e v_{s(e)} w_{t(e)}$$

This is called the *Euler form* of Q. The *symmetrized Euler form* of Q is given by

$$(v, w)_Q = \langle v, w \rangle_Q + \langle w, v \rangle_Q$$

Finally, the *tits form* of Q is the quadratic form $q_Q(v) = \langle v, v \rangle = \frac{1}{2}(v, v)$

Example 25.2.15: Euler And Tits Forms

Let $Q = \overset{1}{\bullet} \xrightarrow{e} \overset{2}{\bullet} \xrightarrow{f} \overset{3}{\bullet}$. Then the Euler's form is:

$$\langle v, w \rangle_Q = \sum_{i=1}^{3} v_i w_i - v_1 w_2 - v_2 w_3$$

$$q_Q(v) = \sum_{i=1}^{3} v_i v_i - v_1 v_2 - v_2 v_3$$

The tits form is also positive definite, since if it weren't, we would be able to pick 3 positive integer's satisfying:

$$v_2(v_1 + v_3) > v_1^2 + v_2^2 + v_3^2$$

However, this is no possible (we can show this with elementary calculus and optimization, or take a couple of space cases and induction).

Now let $Q' = \overset{1}{\bullet} \xleftarrow{e} \overset{2}{\bullet} \xrightarrow{f} \overset{3}{\bullet}$. Then:

$$\langle v, w \rangle_Q = \sum_i^{3} v_i w_i - v_2 w_1 - v_2 w_3$$

and:

$$q_Q(v) = \sum_i^{3} v_i v_i - v_2 v_1 - v_2 v_3$$

then we see that the tits form of Q and Q' agree.

Theorem 25.2.40: Positive Definite Forms And Dynkin Diagrams

Let Γ be a connected graph. The Tits form q_Γ is positive definite if and only if Γ is a Dynkin diagram of type A, D, or E

Proof **:**
ommited by Jimail

(exercise showing not true for not A, D, E)

Finally, we will state one more theorem we need for the proof

Theorem 25.2.41: Coxeter Element Lemma

Let R be a root system, $\Pi = \{\alpha_1, \alpha_2, ..., \alpha_n\}$ a set of simple roots and W its Weyl group. Let $s_i = s_{\alpha_i}$ denote the simple reflections corresponding to the simple root α_i. let $C = s_{i_n} \cdots s_{i_1}$ where $i_1, i_2, ..., i_n$ is some ordering of the simple roots (such an element is called a *Coxeter element*).

Then for any $v \in \mathbb{Z}^I$, there exists a $m \geq 0$ such that $C^m(v) \ngeq 0$

Proof **:**
ommited

Proof **:**
(of Gabriels theorem) here

25.2.5 Path Algebra of Quivers

(Okay, look at your final project for this course; there are some really important insights that feel like they should actually be the priority of this section!! Essentially, we build-up more on how to think about algebras in terms of quotients of free modules, build-up how "path algebras" naturally come into play in representing quivers, and finish off with the maurita equivalence between R-Mod and R/I-Mod where $I = \mathrm{Ann}_R(M)$).

Let $Q = (V, E)$ be a quiver. For vertices i, j of Q, Define the set $\mathrm{Path}(i, j) = \mathrm{span}\{\text{paths } \gamma \text{ from } i \text{ to } j\}$ (i.e. consider the path as a basis to generate a vector-space over k, where k is usually understood), which we will call the *path space*. the length of a particular path is the number of edges between each point. The length of the path from i to i is zero. If γ is a path, and if we let i, j represent a path from it to itself, then if γ is a path from i to j, we will say that γi and $j\gamma$ are the same path.

Lemma 25.2.42: Properties Of Path

final project of reading course

***Proof* :**

1. We want to show that P_i has the structure of a representation $\text{Rep}(Q)$, which consists of a collection of vector-spaces and linear transformation's between them. First, each $P_i(j)$ is a vector-space, as pointed out earlier. Next, we may define a "linear" map $T_e : P_i(j) \to P_i(k)$ to be $\gamma \mapsto e\gamma$ (i.e. append another edge to the path). Notice that since we are treating paths as basis elements to some freely generated vector-space $P_i(j)$, then we get by construciton that $T_e(c \cdot \gamma) = ec \cdot \gamma = ce\gamma = cT_e(\gamma)$. Next, taking $\gamma + \epsilon$ to be a formal sum of two paths, we can naturally extend the action of concatinating to the formal sum:

$$e(\gamma + \epsilon) = e\gamma + e\epsilon$$

which will mirror the linear maps within a represetnation $\text{Rep}(Q)$. Furthermore, if e, f are neigbhouring edges, then $T_{ef} = T_e T_f$. This now mirrors how $\text{Rep}(Q)$ is both a collection of vector-spaces and a collection of maps between them.

2. We are now going to treat P_i as a representation, since we considered them to be related to them in the above example. We want to prove that the collection of vector-spces P_i along with the linear transformations T_e are indecomposable, that is we cannot find subrepresentations. This comes from the construction of our map, in particular that we have that for any collection of paths that form a basis for $P_i(j)$, we have that they are eached mapped to by some paths through the appropriate T_e for an appropriate edge e. To make it more intuitive to see, consider for example P_i of the bellow quiver with a "cycle":

$$\begin{tikzcd} & j \arrow[dr, "b"] & \\ i \arrow[ur, "a"] \arrow[rr, "c"'] & & k \end{tikzcd}$$

Then $\dim(V(i)) = 1$, $\dim(V(j))) = 1$, and $\dim(V(k)) = 2$, with the path i mapping to ci in $V(k)$ and mapping to bi in $V(j)$, while the path bi of $V(j)$ maps to ibc, in $V(k)$. Then notice that $V(k)$ is "surjected" upon by both $V(i)$ and $V(j)$ combined, and all the maps are injective. If the above quiver were to split, then either the dimension of $V(i)$ or $V(j)$ would have to get smaller (since $V(k)$ would have to diminish a dimension): if $\dim(V(j)) = 0$, then that would break the surejctivity of T_a, while if $\dim(V(i)) = 0$, then the map $T_c \neq T_c|_A \oplus T_c|_B$ for the two subquivers A and B.

If the quiver has a proper cycle (i.e. $P_i(i)$ has more than $\{i\}$ as a basis), then since every map must be injective (mapping onto the appropriate basis) we have that each vector-space is 1-dimensional and an isomorphism, which we've shown in chapter 1 with a dimension argument that this representation is indecomposable.

If The quiver does not have a cycle:

$$\begin{tikzcd} j & & k \\ & i \arrow[ul, "a"] \arrow[ur, "b"'] & \end{tikzcd}$$

Then if for the sake of contradiction P_i were decomposable, then if $\dim(V(i)) = 0$ in one of the decomposibition, so would $V(j)$ and $V(k)$, making it a trivial decomposition, while if

$\dim(V(i)) = 1$, both $\dim(V(j))$ and $\dim(V(k))$ would have to be 1 or else the map wouldn't be well-defined, and hence this quiver is indecomposable.

More generally, any quiver Q can be broken down into one of these two cases: If the quiver has a "cycle", then since every map is injective and must surject onto the vertex where two edges meet, we can repeat the above argument to show that the two scenarios which we may have while trying to decompose a quiver lead to a contradiction, and if the quiver has no cycles, we repeat the argument we've done for the above quiver.

3. If P_i were not a sink, then if e is an edge where that makes i not a sink, say $s(e) = i$ and $t(e) = j$, then we can take a subquiver where the vector-space at $P_i(i) = 0$, and we have a zero-map from $P_i(i)$ to $P_i(j)$, giving us a subquiver.

   Conversely, if P_i were a sink, the only path from i would be to i, and so all other vertices would have a 0 dimensional vector-space, while $P_i(i)$ would be 1-dimensional, which makes P_i into a simple representation

4. For notational simplicity, if $\gamma = k_n k_{n-1} \cdots k_1$ is some path, let

$$f_\gamma(v) = f_{k_n}(f_{n-1}(\cdots f_1(v)))$$

   so that $f_{ab}(v) = f_a(f_b(v))$.

   For the proof, pick any $v \in V(i)$, and consider the following commutative diagram:

$$\begin{tikzcd} P_i(a) \arrow[r, "T_e"] \arrow[d, "\varphi_a"] & P_i(b) \arrow[d, "\varphi_b"] \\ V(a) \arrow[r, "S_e"] & V(b) \end{tikzcd}$$

   Then define $\varphi^v : P_i \to V$, $\varphi^v(\gamma) = S_\gamma(v)$. Then

$$S_e(\varphi_a^v(\gamma)) = S_e S_\gamma(v) = S_{e\gamma}(v) = \varphi_b^v(e\gamma) = \varphi_b^v(T_e(\gamma))$$

   proving commutativity of the diagram. If $v \neq w$, then since $\varphi_i^v(i) = v \neq w = \varphi_i^w(i)$, $\varphi^v \neq \varphi^w$. Next, if $\varphi \in \mathrm{Hom}_Q(P_i, V)$, then if $(\gamma_1, \gamma_2, ..., \gamma_n)$ is a basis for $V(i)$, consider

$$(\varphi_i(\gamma_1), ..., \varphi_i(\gamma_n)) = (v_1, ..., v_n)$$

   and define $v = \sum_i v_i$. Then extending linearly we get that $\varphi_i\left(\sum_i c_i\gamma_i\right) = \sum_i c_i v_i$, and then extending φ_i to any other φ_j through a chain of commutative diagrams, we get that all of φ is determined by where v maps to in any path in V. Thus, $\varphi = \varphi^v$, and so $\mathrm{Hom}(P_i, V)$ and $V(i)$ are in bijective correspondence. Finally, By linearly of each map in φ, and each map assigned to an edge in P_i and $V(i)$, it we see that:

$$\varphi^{v+cw} = \varphi^v + c\varphi^w \in \mathrm{Hom}(P_i, V)$$

   giving us that the map is in fact a linear map, completing the proof.

Definition 25.2.43: Path Algebra

Let $Q = (V, E)$ be a quiver. Then the *path algebra* $P(Q)$ is defined as follows:

1. As a vector-space $P(Q) = \text{span}\{\text{paths in } Q\}$, where each vertex $i \in V$ we include an element e_i representing the constnat path $i \to i$
2. multiplication is defined via concatenation when legal, that is:

$$p \cdot q = \begin{cases} pq & \text{if } t(q) = s(p) \\ 0 & \text{otherwise} \end{cases}$$

Part VII

Homological Algebra

Throughout this book, the reader has encountered many functors, some notable examples being $\mathrm{Hom}_R(-, N)$, $\mathrm{Hom}_R(M, -)$ and $M \otimes_R -$ among many others. Each of these functors give some information about the objects and the morphisms passing through the functor. As we saw for the Hom and $M \otimes_R -$ functors, certain appropriate objects preserve exactness, while most preserve it only partially. Perhaps a crowning achievement of mid 20th century mathematics is to be able to quantify this lack of exactness, and rather spectacularly use the information given by this quantification to extract a (sometimes copious) amount of information about the objects of study. The great effectiveness with which homological algebra can produce interesting information may seem a little miraculous at first, for example, using homological algebra the notion of *dimension* as it is understood in geometry is readily translatable to algebra, or that we may measure how far away a module is from being a vector-space (details are in K-theory)[1] How this is the case is perhaps best motivated by going through learning the subject to see for oneself, but it may be helpful for some readers to have some of my (the authors) intuition presented to stand on some ground before proceeding.

If we have two objects M_1 and M_2, we may wish to consider a way of constructing a new object M given M_1 and M_2. Ideally, we may form a short exact sequence like so:

$$0 \to M_1 \to M \to M_2 \to 0$$

More generally, we may have a collection of objects $\{M_i\}_{i \in I}$ which we use to construct an object via a long exact sequence. However, it is often not the case that such an exact sequence is too strong, that is, it is not the case that $\mathrm{im}(\varphi_i) = \ker(\varphi_{i+1})$. Even if such a construction exists, we often then want to "transform" this information via a functor which shall then lose us exactness (one can think of a functor as asking some sort of "question" whose output is the answer). In this case, it has been eerily common that the image of φ_i is contained in the kernel of φ_{i+1}:

$$\mathrm{im}(\varphi_i) \subseteq \ker(\varphi_{i+1}) \tag{25.1}$$

Conceptually, this can be thought as the object "being obstructed" from fully being an extension. This new type of sequence is called a *cochain complex*. The reader with a good memory may recall definition 12.2.10 where it was given that another name for a long exact sequence is an exact cochain complex. In this way, the condition given by equation 25.1 is a generalization of the exact condition. The theoretical heart of homological algebra can be thought of how to extract "quantitatively" how a cochain complex *fails* to be an exact cochain complex. For example, in chapter 16, it was have eluded that through the use of homology theory, the Ext functor gives us all the ways two modules M_1, M_2 may be extended. In this case, the failure of $\mathrm{Hom}(M, -)$ or $\mathrm{Hom}(-, N)$ preserving injectivity or surjectivity (equivalently, the "obstruction" to these functors being exact) gave rise to the Ext functor and the study of its homology.

The term "algebraic object" has been is a bit ambiguous. What type of algebraic objects can be used to in the theory of homological algebra? As it turns out, not all algebraic objects satisfy the right conditions for the theory of homological algebra (for example, **Grp** and **Fld** will be shown to be inappropriate for a theory of homology). However, there are non-algebraic objects for which a homological theory can be induced (famously, topological spaces and smooth manifolds admit a theory of homological algebra). This spurred the need to generalize the theory of homological algebra to the categorical level, where homological theory is well-defined on *abelian categories*. We thus shall start homological algebra with an overview of abelian categories. For those who are already familiar with some homological algebra, they may have studied homological algebra within the category of

[1] This is overviewed in chapter 28

R-**Mod**, which is a concrete abelian category. Doing so will in fact not loose any generality due to Freyd-mitchell Theorem (theorem 26.0.13) which states that any [small] abelian cateogry A admits a fully faithful exact functor to a category R-**Mod** for a suitable ring R. Working with the concrete category R-**Mod** brings it benefits as it may be easier to conceptualize the manipulation of elements, and so the reader may ask for the benefit in sticking to working with general abelian categories. As the author, I believe the benefits comes in being comfortable with more advanced cateogry theory. Future algebraists are going to encounter more complicated problems involving category theory (higher order categories, topoi's, and stacks, all examples the reader may see in the not too distant future). Hence, introducing a categorical perspective will familiarize the reader for more complicated categorical notions in their not too distant future[2].

The reader may wonder why a cochain complex is not called a chain complex. This is because we can work in the dual category, where all arrows are flipped. In the dual category, we shall get chain complexes. Homological algebra in the dual category is theoretically the same, however we shall see that cohomology is often computationally easier to work in, permitting more useful algebraic structures we can use to study. It is for this reason that most mathematicians work with cohomology rather than homology, even though every theorem we shall show for homology transaltes directly to cohomology[3]. In this book, we shall start with homology, and switch to cohomology when starting to work with more computational parts of the theory to get the reader accustomed to both modes of thinking.

Goal

Chapter 26 shall introduce the theory of abelian categories which shall be the types of categories in which homological algebra is well-defined.

The main result chapter 27 of the book is trying to convey is what information from R-modules give rise to certain properties in it's respective chain complex(es), and conversely what information from it's chain complex(es) gives us information about modules. We shall see that given an object M within an abelian category, there are an infinite number of ways we may associate a [co]chain complex to it. Thus, there will have to be a way to either associate a canonical [co]chain complex (in a functorial manner), or we assign [co]chain complexes to an object up to appropriate notion of isomorphism. The latter shall be the approach, where we shall see that we will want to consider [co]chain complexes up to something called *quasi-isomorphism* of [co]chain complexes. We shall then see what information we may deduce given different choices of [co]chain complex functors.

After having developed the ground work for exacting cohomological information, in chapter 28 we shall put this knowledge to good use and apply cohomology theory to G-invariant functors and see what this this can tell us about groups. This shall also finally address what was foreshadowed in box:HERE where it was said the units of a Galois field form a trivial 1st homology group. We shall show how to quantify the obstruction of a module to being free via a bit of K-theory, and define the notion of homological dimension which shall line up with the results presented in the section on commutative algebra.

The part will be concluded with the introduction chapter 29 which are the quintessential for any

[2] assuming, naturally, that the reader holds great interest in algebra and shall keep persuing branches within the field

[3] there is even an equivalent algebraic strucutre for the additional algebraic structure that is induced in cohomology, however it is less useful

concrete computations in homological algebra and are an indispensable tool for any serious work in many disciplines of algebra and geometry (interesting examples being algebraic topology, algebraic geometry).

26

Abelian Categories

We have extensively worked with categories which have a multitudes of different properties. Some important terms to recall are the following:

- *zero-object*: an object 0 in a category C $0 \in \mathrm{Obj}(C)$ that is both the initial and final object. Such an object is also called a terminal object. Similarly, a zero morphism is a morphism $f : X \to Y$ such that for any other morphisms $g, h : X \to Y$, $fg = fh$ and $hf = gf$.
- *Products* and *Coproducts*: Given two objects $a, b \in \mathrm{Obj}(C)$, if there exists an object $c \in \mathrm{Obj}(C)$ that is the limit (resp. colimit) of the diagram:

$$\overset{A}{\bullet} \qquad \overset{B}{\bullet}$$

  then c is the product (resp. coproduct) of a and b, and is usually written as $a \times b$ (resp. $a \amalg b$)
- *kernels* and *cokernels*: Given a morphism $f : a \to b$, the kernel (resp. cokernel) of f is an object k and a morphism $\iota : k \to a$ such that k is the limit (resp. colimit) of the diagarm:

$$\overset{a}{\bullet} \xrightarrow{f} \overset{b}{\bullet}$$

  with the added restriction that the map $\varphi : k \to y$ must be the zero morphism (namely, by commutativity, $f\iota = 0_{kb}$ must be the zero morphism). In terms of diagrams, we get that that for any $\iota' : K \to A$ such that $f\pi\iota' = 0$, that there exists a morphism φ st the following diagram commutes:

$$\begin{tikzcd}
K' \arrow[r, "\iota'"'] \arrow[rr, bend left, "0"] \arrow[dr, dashed, "\varphi"'] & A \arrow[r, "\varphi"'] & B \\
& K \arrow[u, "\iota"'] &
\end{tikzcd}$$

and simmilarly for cokernels:

$$\begin{tikzcd} & K' \arrow[d, "\pi'"] \arrow[dr, dashed, "\psi"] & \\ A \arrow[r, "\varphi"] \arrow[rr, bend right, "0"'] & B \arrow[r, "\pi"] & K \end{tikzcd}$$

Kernels and cokernels do not necessarily exist in a category, or not all morphisms have kernels (ex. the category of finitely generated R-modules for any ring R need has morphisms without kernels). Furthermore, monomorphism need not be kernels (not all subgroups are normal) and epimorphisms need not be cokernels (recall $\mathbb{Z} \to \mathbb{Q}$ is an epimorphism in **Ring**)

Not all categories we have worked with have these properties. For example, **Fld** doesn't have products (what would be the characteristic of the product of $\mathbb{F}_p$ and $\mathbb{F}_q$ for $p \neq q$?), some categories don't have zero morphisms (set being an example), some have monomorhpisms/epimorphisms that do not correspond to kernels (as showed above). We first limit our attention to categories which have these nice properties:

Definition 26.0.1: Additive Category

Let $\mathbf{A}$ be a category. Then $\mathbf{A}$ is said to be *additive* if it has a zero-object, has both finite products and coproducts, and each $\mathrm{Hom}_A(X, Y)$ can be given an abelian group structure such that the composition of maps:

$$\mathrm{Hom}_A(X, Y) \times \mathrm{Hom}_A(Y, Z) \to \mathrm{Hom}_A(X, Z)$$

is bilinear. A functor between additive categories is *additive* if it is a group homomorphism on hom-sets.

A functor between two additive categories is called an *additive functor* if the maps between the hom-sets is a homomorphism.

The term "additive" comes froms the fact that we may interpret the additive category as a monoidal category (a category which we may add a functor $\otimes : C \times C \to C$ that is given the property of a monoid operation, for an additive category this would be the direct product), where the hom-sets are "enriched" to also be abelian groups. Another way of looking at additive categories is in which the algebra of matrices is well-defined.

Example 26.0.1: Additive Categories

1. The category of R-modules is an example of an additive category.
2. **Grp** is not additive, for $\mathrm{Hom}_{\mathbf{Grp}}(A, B)$ doesn't have an abelian group strucuture satisfying the above condition.
3. The category of matrices over a ring is an additive category.
4. The category of finitely generated R-modules is an additive category

A key realisation is that in the absense of kernels and cokernels as objects monomorhpisms and epimorhpisms take their place:

Lemma 26.0.2: Additive Categories: Monomorphisms and Kernels I

Let A be an additive category. Then kernels are monomorhpissm and cokernels are epimorhpisms

Proof **:**
Aluffi p.562-563

Note that kernels and cokernels need not exist in an additive category, hence the converse of each accession need not be the case (in a vacuous way). If a morhpism φ has a kernel then we may identify exactly what this kernel is:

Lemma 26.0.3: Additive Categories: Monomorphisms and Kernels II

Let A be an additive category and $\varphi : X \to Y$ a morphism. Then:

1. If φ has a kernel, then φ is a monomorhpism if and only if $0 \to X$ is it's kernel
2. If φ has a cokernel, then φ is a epimorphism if and only if $Y \to 0$ is it's kernel

Proof **:**
Aluffi p.563

As mentioned, it is not guaranteed that kernels and cokernels exist. We shall shortly axiomatically add them to our category by adding them into the definition of a new category that supersedes additive categories, but first we show a few more important results about abelian categories. Namely, in an additive category, products and coproducts are equivalent:

Lemma 26.0.4: Additive Categories: Products and Coproducts

Let A be an additive category and $X, Y \in \mathrm{Obj}(A)$. Then if $X \times Y$ exists, so does $X \amalg Y$, and if $X \amalg Y$ exists, so does $X \times Y$. Furthermore,

$$X \times Y \cong X \amalg Y$$

Proof **:**
Stacks Project (see this link)

Hence, additive categories behave very much like abelian groups, however as we pointed out kernels and cokernels need not exist. We thus define a new category that forces these conditions:

Definition 26.0.5: Abelian Category

[Abelian Category] Let A be an additive category. Then A is an *abelian* category if kernel and cokernels exist in A.

By the above lemmas, that is every monomophrims is the kernel of some morphism, and every epimorphism is the cokernel of some morphism. Since each kernel is asociated to a monomorhpism, we may intuitively think of the association:

$$\text{kernel} \Leftrightarrow \text{submodule}$$

and similarly:

$$\text{cokernel} \Leftrightarrow \text{quotient module}$$

In many ways, an abelian category should be thought of as the minimum number of conditions necessary to define exact sequence[1]. To see this, we require the following:

Lemma 26.0.6: Abelian Category: Kernels And Cokernels

Let A be an abelian category. Then every kerne lis the kernel of its cokernel, and every cokernel is the cokernel of its kernel.

Proof **:**
Aluffi p.565

Lemma 26.0.7: Abelian Categories: Isomorphisms

[Abelian Categories, Isomorphisms] Let A be an abelian category and $\varphi : A \to B$ a morphism in A. Then if φ is a monomorphism and an epimorphism, it is an isomorphism

Note that this does not hold in additive categories (see ref:HERE)

Proof **:**
Aluffi p.566

Theorem 26.0.8: Abelian Category: Canonical Decomposition

[Abelian Category, Canonical Decomposition] Let A be an abelian category and $\varphi : A \to B$ a morphism. Then A can be decomposed in the following cannonical way:

$$\begin{tikzcd} A \arrow[r, two heads] \arrow[rrr, bend left, "\varphi"] & C \arrow[r, "\tilde{\varphi}"'] & K \arrow[r, tail] & B \end{tikzcd}$$

[1]Those familiar with some algebraic topology may find it insightful that an abelian category is the the category with ninimum number of condition so that the snake lemma (theorem 27.1.11) holds

Proof **:**
Aluffi p.570

From the above, we may think of ker(coker) : $K \to B$ as representing the "image" of φ:

Lemma 26.0.9: Abelian Category: Images

Let $\varphi : A \to B$ be a morphism in an abelian category, and let $\iota : K \to B$ be the kernel of the cokernel of φ. Then:

1. ι is a monomorhpism
2. φ factors through ι
3. ι is initial with these properties

Proof **:**
Aluffi p.572

Definition 26.0.10: Abelian Category: Image

Let $\varphi : A \to B$ be a morphism in an abelian category. Then the *iamge* of φ is $\ker(\operatorname{coker}(\varphi))$ and is denoted $\operatorname{im}(\varphi)$. Similarly, its co-image is $\operatorname{coker}(\ker(\varphi))$ and is dentoed $\operatorname{coim}(\varphi)$.

With our intuition from working with R-modules, given a morphism $\varphi : A \to B$ with image $\iota : K\iota B$, we may imagine there is an induced epimorphism $\overline{\varphi} : A \to K$. This is indeed the case:

Proposition 26.0.11: Abelian Category: sub object and quotient objects

Let $\varphi : A \to B$ be a morphism in an abelian category with image $\operatorname{im}\varphi : K \to B$ and coimage $\operatorname{coim}\varphi : AC \to A$. Then the induced morphisms $A \to K$ and $C \to A$ are (respectively) epimorphisms and monomorphisms.

Proof **:**
Aluffi p.573

From the cannonical decomposition, we may say that the image and coimage are "isomorphic" (which is not strictly true since these are morphisms, however by the Freyd-Mitchell Theorem this is a safe way of looking at it).

Theorem 26.0.12: Abelian Category: Exactness

Consider a sequence of objects in an abelian cateogy

$$\cdots \to A \xrightarrow{\varphi} B \xrightarrow{\psi} C \to \cdots$$

Then this sequence is *exact* if:

1. $\psi \circ \varphi = 0$
2. $\operatorname{coker} \varphi \circ \ker \psi = 0$

Example 26.0.2: Categorical Exactness

1. Aluffi p.577

Freyd-Mitchell Theorem

As mentioned, R-mod is an abelian category, and as we shall develop a bit later it the categories R-mod "represent" all abelian categories.

Theorem 26.0.13: Freyd-Mitchell Theorem

Let A be a small abelian category. Then there is a fully faithful exact functor $A \to R$-Mod for a suitable ring R.

Proof **:**
partial proof?

26.1 Category of Categories

(needed for derived categories proper definition)

27

Complexes and (co)Homology

Let $\mathbf{A}$ be a fixed abelian category, which by theorem 26.0.13 may without loss of generality it can be considered ${}_R\mathrm{Mod}$. In this chapter, we shall focus on studying (co)chains of such modules, that is sequence of modules that respect equation (25.1):

$$\mathrm{im}(d^n) \subseteq \ker(d^{n+1})$$

The collection of all chains shall be denoted $\mathrm{Ch}(\mathbf{A})$. We shall see that $\mathrm{Ch}(\mathbf{A})$ is an abelian category, the notion of extensions (and snakes lemma) is well defined, which will be explored. In the right categories, it is relatively easy to construct a concrete complex given the category has enough objects, for example $M \in \mathrm{Obj}(R\text{-}\mathbf{Mod})$ always has a complex known as a *free resolution* (see definition 12.3.17). There will be many different ways of creating complexes given an object, giving us different types of resolutions. We shall see that when a category has enough projective objects (resp. Injective objects), then "up to homotopy equivalence", projective resolutions shall be *initial* (resp. final) in the category of complexes $\mathrm{Ch}(A)$. Thus, we shall take the time to develop the homology theory in categories with enough projective or injective objects. We shall see that under appropriate circumstances, taking (co)homology of a projective resolution of an object can give us some important information about said object.

Homology and cohomology are duals of each other, which means on the theoretical level they are the same. As cohomology is the more computationally useful tool, we shall be focusing on cohomology on this chapter, with the idea that the reader shall be comfortable being able to switch indices when necessary. Taking the cohomology of a complex shall be shown to be an additive functorial process, meaning for a given morphism of complexes $\alpha^\bullet : A^\bullet \to B^\bullet$, we shall get a morphism each cohomology $H^i(\alpha) : H^i(A^\bullet) \to H^i(B^\bullet)$ which will be shown can be to form a new complex $H^\bullet(\alpha) : H^\bullet(A^\bullet) \to H^\bullet(B^\bullet)$.

The goal of this chapter is to show how to extract information given a complexes (co)homology. Given this, it is important to know what information a morphism of complexes preserves in cohomology, that is given a morphism of complexes, $\alpha^\bullet : A^\bullet \to B^\bullet$, when is $H^\bullet(\alpha) : H^\bullet(A^\bullet) \to H^\bullet(B^\bullet)$ an

isomorphism? This shall lead us to the notion of *quasi-isomorphisms*, and their "representation" through mapping cones. As it turns out, quasi-isomorphisms do not behave nicely in general. What we shall then seek out is a stricter concept known as a *homotopy equivalence* that will still be a quasi-isomorphism, but will be invertible in the appropriate category.

27.1 Basic Definition and Properties

Starting with defining the generalization of an exact sequence:

Definition 27.1.1: Chain Complex

A *chain complex* $(C_\bullet, d_\bullet)$ is a sequence of objects and morphisms:

$$\cdots \xleftarrow{d_i} M_i \xleftarrow{d_{i+1}} M_{i+1} \xleftarrow{d_{i+2}} \cdots$$

with the indices running through $\mathbb{Z}$, such that

$$d_i \circ d_{i+1} = 0$$

and each d_i is called a *boundary map*. Similarly, a *cochain complex* $(C^\bullet, d^\bullet)$ is a sequence of objects

$$\cdots \xrightarrow{d^i} M^i \xrightarrow{d_{i-1}} M^{i-1} \xrightarrow{d^{i+2}} \cdots$$

with the indices running through $\mathbb{Z}$, such that

$$d^i \circ d^{i-1} = 0$$

where $d^\bullet$ are is called the *differential*[a].

[a] Due to it's origin's in differential geometry and algebraic topology

Chains and cochains can be interchanged by taking

$$M_i = M^{-i}$$

Working with one or the other makes essentially no difference for homological algebra. In certain cases, some structures present themselves more evidently in the cohomological case (ex. a natural ring structure called the *cohomology ring* shall become important to us), and so we opt to work with the cohomology, even though at this stage this can be considered to be an arbitrary choice. The condition that $d^i \circ d^{i-1} = 0$ is the same as $\operatorname{im} d^{i-1} \subseteq \ker d^i$. If the complex is exact, then $\operatorname{im} d^{i-1} = \ker d^i$ for all i. We also often write $C^\bullet$ instead of $(C^\bullet, d^\bullet)$ if the mappings $d^\bullet$ are clear from context or implied. In some textbooks, even the bullet is dropped and M represents the complex. Sometime, for homology, the notation $\partial_\bullet$ shall be used instead of $d_\bullet$ due to it's connection to being the boundary map for topological spaces when creating complexes out of topological spaces[1].The condition that $d^i \circ d^{i-1} = 0$ (resp. $d_i \circ d_{i+1} = 0$) is often abbreviated to $d \circ d = 0$.

Due to historical origins in algebraic topology, the kernel and image of the differential

[1] Similarly, $d^\bullet$ stands for the *differential* due to it's link to the derivative, which would be explored in a class in algebraic topology, or see EYNTKA Algebraic Topology

Definition 27.1.2: Cycles and Boundaries

Let $C_\bullet$ (resp. $C^\bullet$) be a chain complex (resp. cochain complex). Then:

1. $\ker(d_n)$ is called the *n-cycles* of $C_\bullet$ and $\operatorname{im}(d_{n+1})$ is called the n-boundary of $C_\bullet$. The n-cycle is represented as $Z_n = Z_n(C_\bullet)$ and the *n-boundary* is represented as $B_n = B_n(C_\bullet)$
2. $\ker(d^n)$ is called the *n-cocycles* of $C^\bullet$ and $\operatorname{im}(d^{n-1})$ is called the *n-coboundary* of $C^\bullet$. The n-cocycle is represented as $Z^n = Z_n(C_\bullet)$ and the n-coboundary is represented as $B_n = B_n(C_\bullet)$

The name cycles and boundaries has it's origin in algebraic topology:

Example 27.1.1: Cycles and Boundaries
(Put that image from p.67 from alg top notes)

The collection of all complexes forms a Category:

Definition 27.1.3: Category Of Complexes

Let $\mathbf{A}$ be an abelian category. Then $\mathrm{Ch}^\bullet(\mathbf{A})$ is a new category called the *category of complexes of A* where

1. $\mathrm{Obj}(\mathrm{Ch}^\bullet(\mathbf{A})) = \{\text{cochain complexes in } \mathbf{A}\}$
2. For two complexes $C^\bullet$ and $N^\bullet$, a morphism $\alpha^\bullet \in \mathrm{Hom}_{\mathrm{Ch}(\mathbf{A})}(C^\bullet, N^\bullet)$ consists of morphisms st the following diagram commute:

$$\begin{tikzcd}
\cdots \arrow[r] & M^{i-1} \arrow[r, "d^{i-1}_{C^\bullet}"] \arrow[d, "\alpha^{i-1}"] & M^i \arrow[r, "d^i_{C^\bullet}"] \arrow[d, "\alpha^{i-1}"] & M^{i+1} \arrow[r] \arrow[d, "\alpha^{i-1}"] & \cdots \\
\cdots \arrow[r] & N^{i-1} \arrow[r, "d^{i-1}_{N^\bullet}"] & N^i \arrow[r, "d^i_{N^\bullet}"] & N^{i+1} \arrow[r] & \cdots
\end{tikzcd}$$

Similarly for $\mathrm{Ch}_\bullet(\mathbf{A})$ with chain complexes.

If unambiguous, the bullet is dropped giving $\mathrm{Ch}(\mathbf{A})$. It should be checked that $\mathrm{Ch}(A)$ is indeed a category. There are many variations of $\mathrm{Ch}(A)$ which will be important:

1. $C^+(A)$ is the full subcategory of $\mathrm{Ch}(A)$ of complexes which are bounded from bellow, so that $L^i = 0$ for $i \ll 0$
2. $C^-(A)$ is the full subcategory of $\mathrm{Ch}(A)$ of complexes which are bounded from above, so that $L^i = 0$ for $i \gg 0$
3. $C^{\geq 0}(A)$ (resp. $C^{\leq 0}$) for which $L^i = 0$ for $i \leq 0$ (resp. $i \geq 0$)
4. P, I are full subcategories of A consisting of projective and injective objects in A respectively, and hence we have $\mathrm{Ch}(P)$, $\mathrm{Ch}(I)$, and all its variants

There are also a few simple functors we may immediately define:

1. Take $\iota_r : \mathbf{A} \to \mathrm{Ch}(\mathbf{A})$ that sends an object A form $\mathbf{A}$ to a complex that is zero every and A in the rth degree. We shall commonly identify A with the complex given by ι_0.
2. Take $s_r : \mathrm{Ch}(\mathbf{A}) \to \mathrm{Ch}(\mathbf{A})$ which shifts over the complex, namely
$$C^\bullet \mapsto M[r]^\bullet$$
where $M[r]^i = M^{i+r}$ and $d^i_{M[r]^\bullet} = (-1)^r d^{ir+r}_{C^\bullet}$ (the $-1)^r$ shall be explained shortly).
3. the truncation functor shall truncate the complex in the natural way.

More importantly, $\mathrm{Ch}(A)$ is an abelian category, allowing the notion of exactness between complexes to be well-defined:

Lemma 27.1.4: Complex Category Is Abelian

Let A be an abelian category and $\mathrm{Ch}(A)$ the category of complexes of A. Then $\mathrm{Ch}(A)$ is an abelian category

Proof **:**
Left as an exercise, check that:

1. morphisms between two complexes form an abelian group, namely since $\alpha^i + \beta^i$ will be well-defined
2. Finite products and coporducts exist in $\mathrm{Ch}(A)$ as a consequence of them existing in A
3. kernels and cokernls will come from the commutativity of the diagram defining morphisms of complexes

Since it is an abelian category, we shall be able to consider sequences of complexes, and in particular exact sequences of complexes:
$$\cdots \to L^\bullet \to C^\bullet \to N^\bullet \to \cdots$$
which give a commutative diagram like so:

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
\cdots \arrow[r] & A_{n+2} \arrow[r] \arrow[d, "\iota"] & A_{n+1} \arrow[r] \arrow[d, "\iota"] & A_n \arrow[r] \arrow[d, "\iota"] & A_{n-1} \arrow[r] \arrow[d, "\iota"] & \cdots \\
\cdots \arrow[r] & B_{n+2} \arrow[r] \arrow[d, "\pi"] & B_{n+1} \arrow[r] \arrow[d, "\pi"] & B_n \arrow[r] \arrow[d, "\pi"] & B_{n-1} \arrow[r] \arrow[d, "\pi"] & \cdots \\
\cdots \arrow[r] & C_{n+2} \arrow[r] \arrow[d] & C_{n+1} \arrow[r] \arrow[d] & C_n \arrow[r] \arrow[d] & C_{n-1} \arrow[r] \arrow[d] & \cdots \\
 & 0 & 0 & 0 & 0 &
\end{tikzcd} \tag{27.1}$$

Example 27.1.2: Complexes

1. Given any object M, we may always create the trivial complex

$$\cdots 0 \to 0 \to M \to 0 \to 0 \to \cdots$$

   Verify that this is indeed a complex.

2. As mentioned, In R-**Mod**, the free resolution is an example of a complex.

3. Any short exact sequence can be thought of an (exact) complex if we add 0's that extend arbitrarily to the left and right.

4. Let $C^n = \mathbb{Z}/8\mathbb{Z}$ for $n \geq 0$ and $C^n = 0$ for $n < 0$. Let each d^n be defined as:

$$\overline{x} \mapsto \overline{4x}$$

   Show that this is a cochain complex.

5. Let $\{V_i, W_i\}_{i \in \mathbb{Z}}$ be a collection of vector spaces. Define:

$$A_n = V_n \oplus W_n \oplus V_{n-1}$$

   Combining the projection and inclusion maps $A_n \to V_{n-1} \to A_{n-1}$, show that $A_\bullet$ forms a chain complex.

6. Let $C^\bullet$ be a cochain complex, choose an object $A \in \mathbf{A}$, and apply $\mathrm{Hom}_A(-, A)$ to each object. Show that this is a cochain complex.

Since $\mathrm{Ch}(\mathbf{A})$ is an abelian category, kernels and cokernels exist and match with monomorphisms/epimorphisms, hence giving us the following:

Definition 27.1.5: Sub-complexes And Quotient Complexes

Let $C^\bullet$ be a (co)chain complex. Then a (co)chain complex $N^\bullet$ is called a *subcomplex* if each N_n is a sub-object of C_n and the differential on B_n is the restriction of the differential in C_n (categorically, the inclusion map $\iota : N^\bullet \to C^\bullet$ is a morphism of complexes).

Furthermore, C^n/B^n is a *quotient-complex*. Furthermore, by the canonical decomposition of morphisms in abelian categories, for any complex-morphism $\alpha^\bullet$, $\{\ker(\alpha^\bullet)\}$ and $\{\mathrm{coker}(\alpha^\bullet)\}$ is a sub-complex and quotient complex respectively.

27.1.1 Homology and Cohomology

Chain complexes need not be exact. The cohomology of a complex measures how far away a complex is from being exact, that is how far away a complex is form representing a sequence of extensions

Definition 27.1.6: Homology and Cohomology Of A Complex

Let $(C^\bullet, d^\bullet)$ be a complex. Then the ith homology and cohomology of the complex at M^i is, respectively:

$$H_i(C^\bullet) = \frac{\ker d_i}{\operatorname{im} d_{i+1}} \qquad H^i(C^\bullet) := \frac{\ker d^i}{\operatorname{im} d^{i-1}}$$

Though homology and cohomology are similar, it is not necessarily the case that $H^i(C^\bullet) = H_{n-i}(C^\bullet)$ for some appropriate n around which the chain can "pivot". We shall see such examples soon, and later also show under what conditions we may compare homology and cohomology[2]

Example 27.1.3: Cohomology Of Complex

Let's compute the cohomology of sequences in example 27.1.2.

1. The homology and cohomology at the 0th entry is:

$$\frac{\ker(d_0)}{\operatorname{im}(d_1)} = \frac{\ker(d^1)}{\operatorname{im}(d^0)} = M$$

   which shows that at this position, the chain is "off" by M in being exact.

2. As we have defined exact free resolutions, they are exact, meaning $\operatorname{im}(d^i) = \ker(d^{i-1})$, hence their homology and cohomology is always 0

3. Any exact sequence has trivial homology and cohomology

4. This sequence is not exact, and yet it's homoogy/cohomology is trivial (as the reader should verify). Such a sequence is called *acyclic*.

5. The reader should calculate this cohomology as an exercise

6. This cohomology shall be calculated later in this chapter.

Lemma 27.1.7: Cohomology Functor

For every $i \in \mathbb{Z}$, there exists a functor:

$$H^i : \operatorname{Ch}(A) \to A \qquad C^\bullet \mapsto H^i(C^\bullet)$$

which is additive and covariant.

This shows that taking cohomology is indeed a way of analyzing resolutions, for it preserves enough of the structures of modules and complexes. Notice it is not necessarily an exact functor, hence it will not in general preserve extensions.

[2]The universal coefficient theorem and Poincaré duality being two famous examples the reader may have heard of if they took algebraic topology

Proof **:**
Let $\alpha^\bullet : C^\bullet \to N^\bullet$ be a morphism of complexes and take $H^i(C^\bullet)$ and $H^i(N^\bullet)$. Let $x \in H^i(H^\bullet)$ so that $x = k + \mathrm{im}(d_M^{i-1})$ for d^{i-1}. Since $d^i(x) = 0$, by commutativity it must be that $\alpha^i(x) \in \ker(d_N^i)$. Furthermore, by commutativity we see that $\overline{\alpha^i}$ is zero on the image of $\mathrm{im}(d_M^{i-1})$, hence $H^i(\alpha^\bullet)$ is well-defined.

This functor is certainly covariant, and is additive since $\mathrm{Hom}_A(H^i(C^\bullet), H^i(N^\bullet))$ is an abelian group.

Notice that each H^i takes d^i to the zero map. Hence, we may consider the following functor:

$$H^\bullet : \mathrm{Ch}(\mathbf{A}) \to \mathrm{Ch}(\mathbf{A})$$

by taking $H^i(C^\bullet)$ for each i and connecting them by the zero-morphism. This is the cohomology of this complex then equals that of $C^\bullet$, namely:

$$H^\bullet(H^\bullet(C^\bullet)) = H^\bullet(C^\bullet)$$

Lemma 27.1.8: Cohomology And Products

Let $\{A_\alpha\}$ be a collection of objects in $\mathbf{A}$. Then:

$$H^n(\oplus \alpha A_\alpha) = \oplus_\alpha H^n(A_\alpha) \qquad H^n(\prod_\alpha A_\alpha) = \prod_\alpha H^n(A_\alpha)$$

Proof **:**
exercise.

Now, given a morphism of complexes $\alpha^\bullet : C^\bullet \to N^\bullet$, we may ask what (co)homological information such a morphism preserves. In other words, given a morphism $\alpha^\bullet$, when is $H^\bullet(\alpha^\bullet)$ an isomorphism? We give such morphisms a name:

Definition 27.1.9: Quasi-Isomorphism

Let $\alpha^\bullet : C^\bullet \to N^\bullet$ be a morphism of complexes. Then if $H^\bullet(\alpha^\bullet) : H^\bullet(C^\bullet) \to H^\bullet(N^\bullet)$ is an isomorphism, $\alpha^\bullet$ is called a *quasi-isomorphism*

One may wonder if the fact that the (co)homology is isomorphism is strong enough of a condition for the chains to be isomorphic. This turns out to be rather far from the case:

Example 27.1.4: Quasi-Isomorhpism

1. The following counter example shall come up with many different minor variations, and so the reader should keep this in mind. Take the following morphism of complexes:

$$\begin{tikzcd}
\cdots \arrow[r] & 0 \arrow[r] \arrow[d] & \mathbb{Z} \arrow[r, "\cdot 2"] \arrow[d] & \mathbb{Z} \arrow[r] \arrow[d, "\pi"] & 0 \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & 0 \arrow[r] & 0 \arrow[r] & \mathbb{Z}/2\mathbb{Z} \arrow[r] & 0 \arrow[r] & \cdots
\end{tikzcd}$$

This is a quasi-isomorphism, as the reader should check, however it does not have an inverse. More generally, it is not even an equivalence relation, failing to be symmetric (show there is no morphism the other way around that is a quasi-isomorphism).

2. The natural maps $H^\bullet(C^\bullet) \to C^\bullet$ and $C^\bullet \to H^\bullet(C^\bullet)$ are quasi-isomorphisms, given the property of $H^\bullet(H^\bullet(C^\bullet)) = H^\bullet(C^\bullet)$ discussed earlier.

3. Let $\iota(A) \in \mathrm{Obj}(\mathrm{Ch}(\mathbf{A}))$ be the complex with A at the 0th position and zero's everywhere else. Say that $C^\bullet$ is a resolution such that $C^\bullet \to \iota(A)$ is a quasi-isomorphism. Show that this implies $C^\bullet$ is exact everywhere except possibly at $H^0(M^0)$

The lack of symmetry means that quasi-isomorphisms do not form an equivalence relation, but a *weak equivalence relation* (a relation that is reflective and transitive, but not symmetric). This is rather a pity, since if it did form an equivalence relation, then the equivalence class of morphisms up to weak isomorphism would hold all the necessary (co)homological information. Alas, this was not mean to be. Thus, in the pursuit of finding (co)homological invariants, it productive to ask how (co)homological information can be detected between (co)chain complexes. One idea is to map $\mathrm{Ch}(\mathbf{A})$ into another category $D(\mathbf{A})$ in which quasi-isomorphisms become isomorphisms. If quasi-isomorphisms formed equivalence relationships, we could have formed a congruence category which would satisfy this condition, however as just mentioned this cannot be the case.

Such a functor has already been encountered: the (co)homology functor is already one such example. It is more fruitful to as if this can be done in a *univerasl* way, that is if there is a category $D(\mathbf{A})$ where for any category D for which there exists an additive functor $\mathscr{F} : \mathrm{Ch}(\mathbf{A}) \to D$ where every quasi-isomorphism in $\mathrm{Ch}(\mathbf{A})$ becomes an isomorphism in D, there exists a unique functor such that the following diagram commutes:

$$\begin{tikzcd} \mathrm{Ch}(\mathbf{A}) \arrow[r, "\mathscr{F}"] \arrow[d] & D \\ D(\mathbf{A}) \arrow[ur, "\exists!"'] & \end{tikzcd} \tag{27.2}$$

This category does indeed exist, and is called the *derived category* (hence the $D(-)$ notation). This category shall further be explored in the coming sections. In some nice categories $\mathbf{A}$, quasi-isomorphisms will line up with the nicer notion of *homotopy-equivalence*, in which case the derived category will be eminently understandable. We shall explore these all in greater detail section 27.2.

Though quasi-isomorphisms may not be ideal in computing (co)homological invariants, they are still important as the defining concept for (co)homological information. For example, the following shows the relationship between quasi-isomorphism and exactness:

Lemma 27.1.10: Quasi-isomorphisms And Exactness

Let $C^\bullet$ be a cochain complex. Then the following are equivalent:

1. $C^\bullet$ is exact
2. $C^\bullet$ is acyclic, that is $H^\bullet(C^\bullet) = 0$
3. The morphism of complexes $0 \to C^\bullet$ is a quasi-isomorphism

***Proof* :**
Good exercise.

27.1.2 Snake Lemma

We often have multiple ways of constructions a resolution for an object M. We shall often extend resolutions using smaller resolutions. We then get cohomology of this new extended resolution by simply applying the cohomology functor. However, is is this functor exact? More generally, let's say we have a short exact sequence of complexes, then after applying cohomology do we get a short exact sequence of homology complexes? The answer is no: if

$$0 \to L^\bullet \to C^\bullet \to N^\bullet \to 0$$

is a short exact sequence, then we get the (not necessarily exact) sequence:

$$H^i(L^\bullet) \to H^i(C^\bullet) \to H^i(N^\bullet)$$

In terms of morphisms, if we have:

$$\alpha^\bullet : C^\bullet \to N^\bullet$$

Then the morphism after applying the cohomology functor $H^\bullet(\alpha^\bullet) : H^\bullet(C^\bullet) \to H^\bullet(N^\bullet)$ need not be exact, i.e. it doesn't preserve kernels and cokernels. Can we "fix" this in some way by recovering some exact sequence from this and thus allowing our analysis of building more complicated resolutions for M to be independent of cohomology. The answer is yes, Namely, if we take the cohomology of complexes, we may "snake" like in the following diagram:

$$\begin{tikzcd}
 & 0 \arrow[d, dotted] & 0 \arrow[d, dotted] & 0 \arrow[d, dotted] & 0 \arrow[d, dotted] & \\
\cdots \arrow[r, dotted] & H_{n+2}(A) \arrow[r, dotted] \arrow[d, "\iota"] & H_{n+1}(A) \arrow[r, dotted] \arrow[d, "\iota"] & H_n(A) \arrow[r, dotted] \arrow[d, "\iota"] & H_{n-1}(A) \arrow[r, dotted] \arrow[d, "\iota"] & \cdots \\
\cdots \arrow[r, dotted] & H_{n+2}(B) \arrow[r, dotted] \arrow[d, "\pi"] & H_{n+1}(A) \arrow[r, dotted] \arrow[d, "\pi"] & H_n(B) \arrow[r, dotted] \arrow[d, "\pi"] & H_{n-1}(B) \arrow[r, dotted] \arrow[d, "\pi"] & \cdots \\
\cdots \arrow[r, dotted] & H_{n+2}(C) \arrow[r, dotted] \arrow[d, dotted] \arrow[ur] & H_{n+1}(C) \arrow[r, dotted] \arrow[d, dotted] \arrow[ur] & H_n(C) \arrow[r, dotted] \arrow[d, dotted] \arrow[ur] & H_{n-1}(C) \arrow[r, dotted] \arrow[d, dotted] & \cdots \\
 & 0 & 0 & 0 & 0 &
\end{tikzcd}$$

Theorem 27.1.11: Snake Lemma

Given a short exact sequence of chain complexes, we may form a long exact sequence of modules using cohomology:

$$\cdots \to H^n(A) \xrightarrow{i^*} H^n(B) \xrightarrow{\pi^*} H^n(C) \xrightarrow{\partial} H^{n+1}(A) \xrightarrow{i^*} H^{n+1}(B) \xrightarrow{\pi^*} H^{n+1}(C) \to \cdots$$

The following proof can be found (in the Homological variant) in EYNTKA Algebraic Topology

***Proof* :**
We first define the map $\partial : H_n(C) \to H_{n-1}(A)$. We should keep in mind the following diagram:

$$\begin{tikzcd} & A_{n-1} \arrow[d, "\iota"] \\ B_n \arrow[r, "\partial"] \arrow[d, "\pi"] & B_{n-1} \\ C_n & \end{tikzcd}$$

Let $c \in C_n$ be a cycle so that $c \in \ker \partial_n$. Since π is surjective, $c = \pi(b)$ for some $b \in B_n$. Then by definition $\partial : B_n \to B_{n-1}$ so $\partial(b) \in B_{n-1}$. Then by commutativity $j(\partial(b)) = \partial(j(b)) = \partial(c) = 0$ (since c is a cycle), thus $\partial b \in \ker(\pi)$. Since $\ker j = \operatorname{im} i$, $\partial b = i(a)$ for some $a \in A_{n-1}$. Since $i(\partial a) = \partial i(a) = \partial\partial b = 0$ since i is injective, $\partial a = 0$. Thus, define $\partial : H_n(C) \to H_{n-1}(A)$ by $[c] \mapsto [a]$. This is well defined since:

1. The element a is uniquely determined by ∂b since i is injective
2. for differnet choice b' for b, we get $j(b') = j(b)$, os $b' - b \in \ker(j)$ which is equal to $\operatorname{im}(i)$. Thus, $b' - b = i(a')$ for some a', ehcen $b' = b + i(a')$. But then this means we just chanage a to another representative (a homologous element) $a + \partial a'$, namely since
$$i(a + \partial a') = i(a) + i(\partial a') = \partial b + \partial i(a') = \partial(b = i(a'))$$
3. A difference choice of c within its homology class would have hte form $c = \partial c'$. To see this, since $c' = j(b')$ for some b',
$$c + \partial c' = c\partial j(b') = c + j(\partial b') = j(b + \partial b')$$
thus b is replaced by $b = \partial b'$, which gives ∂b and thus a is unchanged.

Finally, $\partial : H_n(C) \to H_{n-1}(A)$ is a homomorphism, for if $\partial[c_1] = [a_1]$ and $\partial[c_2] = [a_2]$ via elements b_1, b_2, then
$$j(b_1 + b_2) = j(b_1) + j(b_2) = c_1 = c_2$$
and
$$i(a_1 + a_2) = i(a_1) + i(a_2) = \partial b_1 + \partial b_2 = \partial(b_1 + b_2)$$
hence
$$\partial([c_1] + [c_2]) = [a_1] + [a_2]$$
giving us the map ∂ is indeed a homomorphism. What's left to show is the following inclusions:

1. $\operatorname{im} i_* \subseteq \ker j_*$. Since $ji = 0$, $j_* i_* = 0$
2. $\operatorname{im} j_* \subseteq \ker \partial$. By definition of ∂, $\partial b = 0$ so $\partial j_* = 0$.
3. $\ker j_* \subseteq \operatorname{im} i_*$. Let $b \in B_n$ with boundary represent a homology class in $\ker j_*$, so $j(b) = \partial c'$ for some $c' \in C_{n+1}$. Since j is surjectiv,e $c' = j(b')$ for some $b' \in B_{n+1}$. Then since $\partial j(b') = \partial c' = j(b)$,
$$j(b - \partial b') = j(b) - j(\partial b') = j(b) - \partial j(b') = 0$$

Thus, $b - \partial b' = i(a)$ for some $a \in A - n$. Since $i(\partial a) = \partial i(a) = \partial(b - \partial b') = \partial b = 0$ and i is injective, a is a cycle. Thus:

$$\iota_*[a] = [b - \partial b'] = [b]$$

showing that i_* maps onto $\ker j_*$

4. $\ker \partial \subseteq \operatorname{im} j_*$. By definition of ∂, if c represents a homology class in $\ker \partial$, then $a = \partial a'$ for some $a' \in A_n$. The element $b - i(a')$ is a cycle since $\partial(b - i(a')) = \partial b - \partial i(a') = \partial b - i(\partial a') = \partial b - i(a) = 0$. Since

$$j(b - i(a')) = j(b) - ji(a') = j(b) = c$$

we get that $j_*[b - i(a')] = [c]$

5. $\ker i_* \subseteq \operatorname{im} \partial$. Given a cycle $a \in A_{n-1}$ such that $i(a) = \partial b$ for seom $b \in B_n$¡ then $j(b)$ is a cycle since $\partial j(b) = j(\partial b) = ji(a) = 0$, hence $\partial[j(b)] = [a]$

Doing this shall completing the proof.

Example 27.1.5: Uses Of Snake Lemma
(from alg top?)

There is another way of looking at chain complexes which shall be important in elucidating triangle categories in section ref:HERE.

Mapping Cones

Using a particular total complex, we can detect when a quasi-isomorphism is an isomorphism. This result requires the snake lemma to prove, hence the need to put this section after the snake lemma.

Definition 27.1.12: Mapping Cone

Let $\alpha^\bullet : C^\bullet \to N^\bullet$ be a morphism of complexes. Then the *mapping cone* is the bi-complex denoted by $MC(\alpha^\bullet)$ defined by:

$$MC(\alpha)^i := L[1]^i \oplus M^i = L^{i+1} \oplus M^i$$

where the morphism $d^i_{MC(\alpha)} : MC(\alpha)^i \to MC(\alpha)^{i+1}$ are given by:

$$d^i_{MC(\alpha)}(l, m) := (-d_L i + 1(l), \alpha^{i+1}(l) + d^i_M(m))$$

The term mapping cone comes from the topological notion of a mapping cone of a function. Visually,

the mapping cone can be seen as the following commutative diagram:

$$\begin{tikzcd}
\cdots \arrow[r] \arrow[ddr] & L^{i+1} \arrow[r, "-d_L^{i+1}"] \arrow[ddr, "\alpha^{i+1}"] & L^{i+2} \arrow[r] \arrow[ddr] & \cdots \\
 & \oplus & \oplus & \\
\cdots \arrow[r] & M^i \arrow[r, "d_M^i"'] & M^{i+1} \arrow[r] & \cdots
\end{tikzcd}$$

Given $MC(\alpha)^\bullet$, we may define two natural chain maps $C^\bullet \to MC(\alpha)^\bullet$ and $MC(\alpha)^\bullet \to L[1]^\bullet$ induced by the natural exact morphisms $M^i \to L^{i+1} \oplus M^i \to L^{i+1}$ which implies that:

$$0 \to C^\bullet \to MC(\alpha)^\bullet \to L[1]^\bullet \to 0$$

is exact.

Proposition 27.1.13: Mapping Cone And Exact Triangles

There is an exact triangle

$$\begin{tikzcd}
 & H^\bullet(C^\bullet) \arrow[dr] & \\
H^\bullet(L[1]^\bullet) \arrow[ur, "+1\ \delta"] & & H^\bullet(MC(\alpha)^\bullet) \arrow[ll]
\end{tikzcd}$$

where δ is the morphism induced by cohomology

Proof **:**
Aluffi p.607

65: Remark

Not all categories have this triangle property, categories which will have this will be called *triangulated categories*

Proposition 27.1.14: Quasi-isomorphism And Isomorphism

Let $\alpha^\bullet : C^\bullet \to N^\bullet$. Then $H^\bullet(\alpha^\bullet)$ is an isomorphism if and only if $MC(\alpha)^\bullet$ is an exact complex.

Proof **:**
exercise

27.1.3 Chain Homotopies and Homotopic Categories

The notion of homotopies originated in topology as a weaker alternative to continuous functions[3]. The topological notion of homotopy settled on the level of continuous functions, where $f, g : X \to Y$ are homotopic (denoted $f \simeq g$) if there exists a continuous function $F : X \times [0,1] \to Y$ where $F(\cdot, t) = f_t$ is continuous, $f_0 = f$ and $f_1 = g$. If we represent X, Y as one dimensional lines, then we may visualize a homotopy as:

Figure 27.1: homotopy

In the development of homotopy theory, it was found that the (co)homology and homotopy of topological spaces (or their representations as chain complexes) is homology-invariant, hence homotopies are examples of quasi-isomorphisms. The proofs of these results turn out to be independent of the underlying topological structures and is purely algebraic, and hence the notion of homotopy has found a generalization as a purely algebraic notion between two (co)chain complexes.

Due to its origin in topology, the intuitions behind the following definitions will proceed more naturally with a more geometric build-up. Since homotopy may be interpreted purely algebraically, this geometrica build-up will be followed by a more algebraic interpretation. First, the reader may take for granted that given a topological space X, we may induce a chain $C_\bullet$, which is a topological object[4]. Given two continuous maps $f, g : X \to Y$, we may push forward the chain like so:

update image

[3]every homeomorphic is a homotopy equivalence, however not every homotopy equivalence is a homeomorphism, hence homotopy can be seen as weaker than continuity

[4]See EYTNKA algebraic topology or Hatcher as for more details

Figure 27.2: homotopy on chain

The space $h(C_\bullet)$ is a topological space, hence we may ask for it's border, $\partial h(C_\bullet)$. By inspecting the image, we see that:

$$\partial h(C_\bullet) = g(C_\bullet) - \delta_+ - f(C_\bullet) + \delta_- = g(C_\bullet) - f(C_\bullet) + h(\partial C_\bullet)$$

re-ordering, we get:

$$g(C_\bullet) - f(C_\bullet) = \partial h(C_\bullet) - h(\partial C_\bullet)$$

Abstracting notation, we get:

$$g - f = \partial h - h\partial \tag{27.3}$$

Now, recall that in a chain complex, the image of the function ∂ is called the boundary, and that by definition of homology the boundary becomes trivial. In a topological setting, this vocabulary reflects the concrete of the boundary map ∂: it maps a chain to it's boundary. What this means for us is that when taking homology on both sides, we get:

$$H^k(g|_{C_\bullet} - f|_{C_\bullet}) = \overline{0}$$

meaning the two maps are equivalent up to homology! In other words, this "geometric intuition" shows that a homotopy is an example of a quasi-isomorphism. Even better, homotopies shall form an equivalence relation, hence they are an example of a quasi-isomorphism that shall capture some (co)homological information (and in the case where all quasi-isomorphisms are homotopies, they could fully capture the (co)homological information!)

Though this is the origins of chain homotopies being (co)homology-invariant, since they may be interpreted in pure algebraic terms, it is worthwhile taking a moment to have an algebraic build-up as well. Algebraically, homotopies can be seen as forming equivalence classes of (co)chain complexes up to *obstruction from being split-exact*. Recall that a split exact sequence is a short exact sequence that is also split at each map, that is there exists a section morphism $s^n : C^n \to C^{n-1}$ such that $d^{n-1} \circ s^n \circ d^{n-1} = d$, or if we drop the indices and collect all the section maps:

$$d = dsd$$

In other words, it is possible to "inverse" the map d via the map s in such a way that it "cancels" the effect of d (for we may re-apply the map d after sd and get the same result). The simplest source of

split-exact (co)chains are exact (co)chains of vector-spaces, for any vector-space maps split[5]. Given a cochain, we may write:

$$C^n = Z^n \oplus B' \qquad (B^n)' \cong C^n/Z^n = d(C^n) = B^{n+1}$$

$$Z^n = B^n \oplus (H^n)' \qquad (H^n)' = Z^n/B^n = H^n(C)$$

Hence, $C^n \cong Z^n \oplus B^{n+1}$. This allows for the definition of the section map via the composition:

$$C^n \hookrightarrow Z^n \hookrightarrow B^n \cong d(C^n) \cong B^{n+1} \subseteq C^{n+1}$$

We may thus define the maps $d \circ s$ and $s \circ d$ where $d \circ s(C^n) = B^n$ and $s \circ d(C^n)(B^n)' \cong B^{n+1}$. Adding these maps together, we get a map $ds + sd : C^n \to C^n$ whose kernel is the homology:

$$\ker(ds + sd)_n = H^n(C^\bullet)$$

The kernel (and cokernel) of the chain map $ds + sd$ is naturally $H^\bullet(C^\bullet)$. Working over more general R-modules (and more generally in an arbitrary abelian category), we loose the natural section maps, for example $0 \to \mathbb{Z} \twoheadrightarrow \mathbb{Z}/2\mathbb{Z} \to 0$ does note have a section map for each map. Instead, we may define arbitrary maps $s^n : C^n \to D^{n-1}$ where we may put them in the following (*not* necessarily commutative) diagram:

$$\begin{tikzcd}
\cdots \arrow[r] & C^{n-1} \arrow[r, "d"] \arrow[dl] & C^n \arrow[r, "d"] \arrow[dl, "s^n"] & C^{n+1} \arrow[r] \arrow[dl, "s^{n+1}"] & \cdots \arrow[dl] \\
\cdots \arrow[r] & D^{n-1} \arrow[r, "d"] & D^n \arrow[r, "d"] & D^{n+1} \arrow[r] & \cdots
\end{tikzcd}$$

Given any such collection of morphisms, we may define $f^n = s^{n+1} \circ d^n + d^{n-1} \circ s^n$ (or, if dropping indices and abbreviate, $f = sd + ds$). Notice that even though the maps $s^\bullet$ do not commute, $f^\bullet$ *will* create a commutative diagram, in particular it is a chain map. To see this, notice that (after dropping indices):

$$df = d(ds + sd) = dsd = (ds + sd)d = fd$$

showing it is indeed a chain map:

$$\begin{tikzcd}
\cdots \arrow[r] & C^{n-1} \arrow[r, "d"] \arrow[d, "f^{n-1}"] \arrow[dl, dotted] & C^n \arrow[r, "d"] \arrow[d, "f^n"] \arrow[dl, dotted, "s^n"] & C^{n+1} \arrow[r] \arrow[d, "f^{n+1}"] \arrow[dl, dotted, "s^{n+1}"] & \cdots \arrow[dl, dotted] \\
\cdots \arrow[r] & D^{n-1} \arrow[r, "d"] & D^n \arrow[r, "d"] & D^{n+1} \arrow[r] & \cdots
\end{tikzcd}$$

Thus, given any collection of morphisms $s^\bullet$, we may naturally define a chain map using the differentials. Conversely, is every morphism of complexes representable as $ds + sd$. This is not in general the case; maps being representable like so are given a name:

Definition 27.1.15: Null-homotopic Maps

Let $f^\bullet : C^\bullet \to D^\bullet$ be a morphism of complexes. Then f is called *null-homotopic* if there exists a collection $\{s^i\}_{i \in \mathbb{Z}}$ of morphisms from C^i to C^{i-1} between two cochain complexes $C^\bullet$ and $D^\bullet$ such that

$$f^\bullet = ds + sd$$

The maps $s^\bullet$ are called the *chian contraction* of $f^b b$

[5]More generally, any maps of projective modules split

To show that not all maps are null-homotopic, consider the following lemma:

Lemma 27.1.16: Null-homotopic And Split Exact Complexes

Let $C^\bullet$ be a complex. Then $C^\bullet$ is split exact if and only if the identity map is null-homotopic.

Proof **:**
exercise

The name "null-homotopic" comes from how the (co)homology of such maps are trivial:

Proposition 27.1.17: Null-homotopic And Cohomology

Let $f^\bullet : C^\bullet \to D^\bullet$ be null-homotopic. Then $H^\bullet(f)$ is the zero map

Proof **:**
Since f is null-homotopic, let $f = ds + sd$. Take a n-cycle representative x of an equivalence class of $H^n(C^\bullet)$. Then:

$$f(x) = d(s(x))$$

Hence, $f(x)$ is an n-boundary in D^n, hence $\overline{f(x)} = \overline{0}$ in $H^n(D^\bullet)$, as we sought to show. (this is an elementary result, but a similar proof is given in Aluffi p.612 if needed)

Thus, maps that are null-homotopic induce zero (co)homology maps![6] Thus, we may think of maps that are null-homotopic as being the "zero" maps. This leads us to the following definition:

Definition 27.1.18: Chain-homotopy

Let $f^\bullet, g^\bullet : C^\bullet \to D^\bullet$ be two morphisms of complexes. Then we say that these two maps are *chain homotopic* or just *homotopic* if there exists maps $h^i : C^i \to D^{i-1}$ such that

$$g^i - f^i = (d^{i-1}_{C^b}) \circ h^i + h^i \circ (d^i_{D^\bullet})$$

or, dropping indices:

$$g^\bullet - f^\bullet = dh - hd$$

If $f^\bullet, g^\bullet$ are homotopic, we write $f^\bullet \sim g^\bullet$

The s notation has been replaced with h notation to emphasize now the "homotopy" aspect of it rather than the "section" property that was evident in the case of vector spaces. If we were working with chains, then we would have

$$g_\bullet - f_\bullet = \partial h - h\partial$$

which, the reader may recognize as equation (27.3). If we were working with chains, then $h_i : C_i \to D_{i+1}$ would shift up a degree, which would be interpreted in the geometrical setting as moving *up* a dimension, in the same way that homotopies are a "dimension more" than the original spaces X, Y.

[6] TBD How much can we say about the converse, if a map induces the zero morphism, then is it null-homotopic?

Furthermore, the reader should verify the "up to homotopy" *is* an equivalence relation, as the reader should verify!

Importantly, two chain-homotopic maps induce the same (co)homology:

Proposition 27.1.19: Chain-Homotopy and Cohomology

Let $f^\bullet, g^\bullet : C^\bullet \to D^\bullet$ be chain homotopic. Then

$$H^\bullet(f^\bullet) = H^\bullet(g^\bullet)$$

***Proof* :**
Since $f^\bullet$ and $g^\bullet$ are chain-homotopic, by proposition 27.1.17 for each element in $H^k(C^\bullet)$ (since it is an additive functor):

$$H^k(f^\bullet)(\overline{x}) - H^k(g^\bullet)(\overline{x}) = 0 \qquad \Rightarrow \qquad H^k(f^\bullet)(\overline{x}) = H^k(g^\bullet)(\overline{x})$$

as we sought to show.

The above result shows promise in using chain-homotopies to hold (co)homological information. The following definition is the important connection idea:

Definition 27.1.20: Homotopy Equivalent Chains

Let $f^\bullet : C^\bullet \to D^\bullet$ be a morphism of complexes. Then $f^\bullet$ is a *homotopy equivalence* if there exists a morphisms of complexes $g^\bullet : D^\bullet \to C^\bullet$ such that

$$f^\bullet \circ g^\bullet \sim 1_{D^\bullet} \qquad \text{and} \qquad g^\bullet \circ f^\bullet \sim 1_{C^\bullet}$$

If $f^\bullet : C^\bullet \to D^\bullet$ is a homotopy equivalence, then $C^\bullet$ and $D^\bullet$ are said to be *homotopy equivalent*

Proposition 27.1.21: Homotopy Equivalence Ans Cohomology

Let $f^\bullet : C^\bullet \to D^\bullet$ be a homotopy equivalence. Then $f^\bullet$ is a quasi-isomorphism, that is $H^\bullet(f^\bullet)$ is an isomorphism

***Proof* :**
Since $f^\bullet$ is a homotopy equivalence, we have $g^\bullet$ that is it's homotopic inverse. Then the maps $f^\bullet \circ g^\bullet$ and $g^\bullet \circ f^\bullet$ induce inverse morphisms of cohomology by proposition 27.1.19, and hence give an isomorphism, as we sought to show.

As demonstrated in example 27.1.4 not all quasi-isomorphisms are invertible even up to homotopy equivalence, hence quasi-isomorphisms are more general. As mentioned earlier, in nice enough categories these two concepts shall coincide.

Now that homotopy equivalences is established, can be asked if given an (additive) functor $\mathscr{F} : \mathbf{A} \to \mathbf{B}$, if there are two homotopic complexes in $\mathrm{Ch}(\mathbf{A})$, will the (co)homology be isomorphic in $\mathbf{B}$? In other

words, when trying to define the derived category that was mentioned earlier, can this category be defined at least up to homotopy? The answer is yes! The proof of this result essentially follows from the following lemma:

Lemma 27.1.22: Homotopy Equivalence And Additive Functors

Let $\mathscr{F} : \mathbf{A} \to \mathbf{B}$ be an additive functor between two abelian categories, $C(\mathscr{F})$ the induced functor on chains. Then if $f^\bullet \sim g^\bullet$ in $\mathrm{Ch}(\mathbf{A})$, $C(\mathscr{F})(f^\bullet) \sim \mathrm{Ch}(\mathbf{B})(g^\bullet)$ in $\mathrm{Ch}(\mathbf{B})$. Furthermore, if $C^\bullet$ and $D^\bullet$ are homotopy equivalence, $C(\mathscr{F})(C^\bullet)$ and $C(\mathscr{F})(D^\bullet)$ are homotopy equivalent

***Proof* :**
Given the first assertion, the second is an immediate consequence, hence we prove the first. Let $f^\bullet \sim g^\bullet$ so that :

$$f^i - g^i = d^{i-1}_{D^\bullet} \circ h^i + h^{i+1} \circ d^i_{C^\bullet}$$

Since $\mathscr{F}$ is an additive functor:

$$\mathscr{F}(f^i) - \mathscr{F}(g^i) = \mathscr{F}\left(d^{i-1}_{D^\bullet}\right) \circ \mathscr{F}(h^i) + \mathscr{F}\left(h^{i+1}\right) \circ \mathscr{F}\left(d^i_{C^\bullet}\right)$$

hence, the morphisms $\mathscr{F}(h^i)$ gives a homotopy between $\mathscr{F}(f^\bullet)$ and $\mathscr{F}(g^\bullet)$, as we sought to show.

Note this does *not* hold for quasi-isomorphisms (exercise 4.15 in Aluffi), needing the stronger condition of exact functors.

Theorem 27.1.23: Derived Category And Homotopy Equivalence

Let $\mathscr{F} : \mathbf{A} \to \mathbf{B}$ be an additive functor between two abelian categories. Then if $C^\bullet$ and $D^\bullet$ are homotopy-equivalent, then:

$$H^\bullet(\mathscr{F}(C^\bullet)) \cong H^\bullet(\mathscr{F}(D^\bullet))$$

that is, their cohomology is isomorphic.

***Proof* :**
exercise.

Hence, up to homotopy, the choice of cochain shall give the same cohomology information!

Homotopic Categories

With chain homotopies being a nicer sub-class of quasi-isomorphisms, their exploration is worth further pursuing. They are in fact almost the solution to the universal problem given in equation (27.2). Let us take for granted that a category $D(A)$ exists that satisfies the universal property. What necessary conditions can be deduced about $D(A)$?

Lemma 27.1.24: Derived Category And Additive Functors

Let $\mathbf{D}$ be an additive category, $\mathscr{F} : \mathrm{Ch}(\mathbf{A}) \to \mathbf{D}$ be an additive functor that takes quasi-isomorphisms to isomorphisms.

1. Let $C^\bullet$ be an exact complex in $\mathrm{Ch}(\mathbf{A})$. Then $\mathscr{F}(C^\bullet) = 0$, that is $\mathscr{F}(C^\bullet)$ is the zero object in $\mathbf{D}$.
2. Let $\alpha^\bullet : C^\bullet \to E^\bullet$ be a morphism of complexes that factors through an exact complex:

$$\begin{tikzcd} C^\bullet \arrow[rr, "\alpha^\bullet"] \arrow[dr] & & E^\bullet \\ & D^\bullet \arrow[ur] & \end{tikzcd}$$

where $D^\bullet$ is exact. Then $\mathscr{F}(\alpha^\bullet)$ is the zero morphism.

***Proof* :**

1. Since $C^\bullet$ is exact, the zero morphism $0^\bullet : C^\bullet \to C^\bullet$ is a quasi-isomorphism, and so $\mathscr{F}(0^\bullet)$ maps to an isomorphism (i.e. an invertible morphism in $\mathbf{D}$). Thus:

$$\begin{tikzcd} \mathscr{F}(C^\bullet) \arrow[r, "\mathscr{F}(0^\bullet)"'] \arrow[rr, bend left, "\mathrm{id}_{\mathscr{F}(C^\bullet)}"] & \mathscr{F}(C^\bullet) \arrow[r, "\mathscr{F}(0^\bullet)^{-1}"'] & \mathscr{F}(C^\bullet) \end{tikzcd}$$

Since $\mathscr{F}$ is additive functor, in particular a group homomorphism on hom-sets, $\mathscr{F}(0) = 0$. Thus:

$$\mathrm{id}_{\mathscr{F}(C^\bullet)} = 0$$

It follows that $\mathscr{F}(C^\bullet)$ must be the zero object (as the reader should verify, see exercise ref:HERE)

2. Applying $\mathscr{F}$ to the diagram in the question, we get the diagram:

$$\begin{tikzcd} \mathscr{F}(C^\bullet) \arrow[rr] \arrow[dr] & & \mathscr{F}(E^\bullet) \\ & \mathscr{F}(D^\bullet) \arrow[ur] & \end{tikzcd}$$

Since $\mathscr{F}(D^\bullet) = 0$ by the above result, we see that $\mathscr{F}(\alpha^\bullet)$ must be the zero morphisms, completing the proof.

Lemma 27.1.25: Derived Categories And Homotopies

Let $\mathscr{F} : \mathrm{Ch}(\mathbf{A}) \to \mathbf{D}$ be an additive functor that maps quasi-isomorphisms to isomorphisms. Let $f^\bullet, g^\bullet : C^\bullet \to D^\bullet$ be homotopic morphisms in $\mathrm{Ch}(\mathbf{A})$. Then

$$\mathscr{F}(f^\bullet) = \mathscr{F}(g^\bullet)$$

Note that this was shown explicitly when $\mathscr{F} = H^i(-)$.

Proof **:**
Since $\mathscr{F}$ is additive, it suffices to show that if $f^\bullet \sim 0$, then $\mathscr{F}(f^\bullet) = 0$. By lemma 27.1.24, it suffices to show $f^\bullet$ factors through an exact complex. The exact complex that shall be used is the mapping cone $MC(\mathrm{id}_{C^\bullet})$ for the identity map $\mathrm{id}_{C^\bullet} : C^\bullet \to C^\bullet$, namely

$$MC(\mathrm{id}_{C^\bullet})^i = C^{i+1} \oplus C^i$$

$$d^i_{MC(\mathrm{id}_{C^\bullet})}(a, b) = (-d^{i+1}_{C^\bullet}(a), a + d^i_{C^\bullet}(b))$$

By proposition 27.1.14, since the identity map $\mathrm{id}_{C^\bullet}$ is (trivially) a quasi-isomorphism, $MC(\mathrm{id}_{C^\bullet})$ is an exact complex.

Now, the goal is to find morphism of complexes where:

$$C^\bullet \to MC(\mathrm{id}_{C^\bullet}) \to D^\bullet$$

the morphism of complexes $C^\bullet \to MC(\mathrm{id}_{C^\bullet})$ is readily available: for each i:

$$b \mapsto (0, b)$$

On the other hand, it is not immediately clear there is a morphism of complexes $MC(\mathrm{id}_{C^\bullet}) \to D^\bullet$. To find such a morphism, we take advantage of the homotopy $f^\bullet \sim 0$, that is there exists maps:

$$h^i : C^i \to D^{i-1}$$

such that $f^\bullet = fh + hd$. With these, define the maps:

$$\pi^i : MC(\mathrm{id}_{C^\bullet})i \to D^i \qquad (a, b) \mapsto h^{i+1}(a) + f^i(b)$$

If this is indeed a morphism of complexes, then the map $f : C^\bullet \to D^\bullet$ factored through an exact complex and hence must be trivial. What must be verified is that:

$$d^{i-1}_{D^\bullet} \circ \pi^{i+1} = \pi^i \circ d^{i-1}_{MC(\mathrm{id}_{C^\bullet})} \tag{27.4}$$

The left hand side of equation (27.4) gives:

$$(a, b) \mapsto h^i(a) + f^{i-1}(b) \mapsto d^{i-1}_{D^\bullet}(h^i(a) + f^{i-1}(b))$$

while the right hand side gives:

$$(a, b) \mapsto (-d^i_{C^\bullet}(a), a + d^{i-1}_{L^\bullet}(b)) \mapsto -h^{i+1}(d^i_{C^\bullet}(a) + f^i(a + d^{i-1}_{C^\bullet}(b))$$

These two sides must be equal. Simplifying the equality, we get:

$$(d^{i-1}_{D^\bullet} \circ h^i + h^{i+1} \circ d^i_{C^\bullet} - f^i)(a) = (f^i \circ d^{i-1}_{C^\bullet} - d^{i-1}_{D^\bullet} \circ f^{-1})(a)$$

for all $a \in C^{i+1}, b \in C^i$. But notice that both sides are zero: the right hand side b $f^\bullet$ is a morphism of complexes, while the left hand side because h is a homotopy and $f^\bullet \sim 0$. But then this gives us our desired result, completing the proof.

Thus, behave well for the desired universal property presented in equation 27.2. We formalize the category in question:

Definition 27.1.26: Homotopy Category

Let $\mathbf{A}$ be an abelian category. Then the *homotopy category* of $\mathbf{A}$, denoted $K(\mathbf{A})$ is the category whose objects ar ethe (co)chain complexes in $\mathbf{A}$ (i.e. $\mathrm{Ch}(\mathbf{A})$), and whose morphisms are:

$$\mathrm{Hom}_{K(\mathbf{A})}(C^\bullet, D^\bullet) = \mathrm{Hom}_{\mathrm{Ch}(\mathbf{A})}(C^\bullet, D^\bullet)/\sim$$

where $\sim$ is the homotopy relation, that is, the morphisms are morphisms of complexes up to homotopy.

These are indeed morphisms, namely $\sim$ is transitive, it respects composition (exercise). Naturally, the equivalent bounded variations $K^-(\mathbf{A})$ and $K^+(\mathbf{A})$ are obtained from $\mathrm{Ch}^+(\mathbf{A})$ and $\mathrm{Ch}^-(\mathbf{A})$.

Lemma 27.1.27: Homotopy Category Is Additive

Let $\mathbf{A}$ be an abelian category. Then $K(\mathbf{A})$ is an *additive category*

Proof **:**
exercise

Note that in general homotopic categories are *not* abelian. Indeed, homotopic maps $f^\bullet \sim g^\bullet$ need not have the same kernel/cokernel, giving no natural definition of these concepts. The homotopic category is a type of category that is stronger than additive but weaker than abelian, known as a *triangulated category* (recall remark$\tilde{\text{t}}$riangulatedCat).

In $K(\mathbf{A})$, homotopy equivalences become isomorphisms. This comes straight from the definition, for if $f^\bullet \circ g^\bullet \sim \mathrm{id}$ in $\mathrm{Ch}(\mathbf{A})$, then $f^\bullet \circ g^\bullet = \mathrm{id}$ in $K(\mathbf{A})$, which can be thought of as formally inverting homotopy equivalences. Using this, we have the following almost universal property, or factorization

This really bugs me, this feels like the universal property, why isn't it

Proposition 27.1.28: Homotopy Category Factorization

Let $\mathscr{F} : \mathrm{Ch}(\mathbf{A}) \to D$ be an additive functor that maps quasi-isomorphisms to isomorphisms. Then ff factors uniquely through $K(\mathbf{A})$:

$$\begin{tikzcd} \mathrm{Ch}(\mathbf{A}) \arrow[r, "\mathscr{F}"] \arrow[d] & D \\ K(\mathbf{A}) \arrow[ur, "\exists!"'] & \end{tikzcd}$$

Proof **:**
exercise

Hence, in the case where $K(\mathbf{A})$ is *not* the universal object satisfying equation (27.2), there must be a unique functor from $K(\mathbf{A})$ to $D(\mathbf{A})$.

27.2 Projective and Injective Resolutions

It has been alluded to a few times that for special categories, homotopy equivalence and quasi-isomorphism coincide. This shall now be covered in this section. In particular, this shall linked to the idea of an abelian category $\mathbf{A}$ being bounded from above or bellow, and having *enough projective objects* or *enough injective objects*. These categories shall furthermore be shown to be universal in some appropriate way, making their choice not arbitrary. Recall that in $\mathbf{A} = R\text{-}\mathbf{Mod}$, an R-module M Is called projective if $\mathrm{Hom}_R(M, -)$ is exact, and injective if $\mathrm{Hom}_R(-, M)$ is exact. Recall that projective/injective modules nicely lend themselves to a splitting condition on short exact sequences for which they are one of the modules forming them (recall theorem 16.2.3 and theorem 16.3.3). This will be important, due to the link of chain homotopies with split exact sequences.

Definition 27.2.1: Projective And Injective Objects

Let $\mathbf{A}$ be an abelian category. Then

1. An object $P \in \mathbf{A}$ is said to be *projective* if the functor $\mathrm{Hom}_A(P, -)$ is exact.
2. An object $Q \in \mathbf{A}$ is said to be *injective* if the functor $\mathrm{Hom}_A(-, Q)$ is exact

theorem 16.2.3 and 16.3.3 hold for abelian categories more generally (in particular the splitting remarks), as the reader may venture to check. Since the dual of an abelian category is abelian (see exercise ref:HERE), the theory on projective objects is dual to that of injective object, and the functor $(-)^{\mathrm{op}}$ maps projective/injectives to injectives/projectives respectively. Thus, many proofs shall simply proof the result for projectives, the result for injectives coming automatically. Note that as we've seen in chapter 16, in a fixed [abelian] category $\mathbf{A}$, projective and injective may have very different features.

In what follows, let $\mathbf{P}$ denote the full subcategory of $\mathbf{A}$ determined by the projective objects, and let $\mathbf{I}$ denote the full subcategory of $\mathbf{A}$ determined by the injective objects. Let $K^-(\mathbf{P})$ denote the full subcategory of $K(\mathbf{A})$ consisting of bounded-above complexes of projective objects of $\mathbf{A}$, and $K^+(\mathbf{I})$ be the full subcategory of $K(\mathbf{A})$ consisting of bounded-below complexes of injective objects of $\mathbf{A}$. Note that these are generally not abelian categories, and furthermore projective and injective objects need not exist:

Example 27.2.1: Category With And Without Projective/Injective Resolutions

1. The category of finite abelian groups has no projective or injective, hence no projective/injective resolution.
2. The category of finitely generated abelian groups has projectives, but no injectives, namely since $\mathbb{Z}^n$ for every n are objects this category. Thus, it is possible to create a free, and hence projective, resolution.
3. the dual of the above category has enough injectives but not enough projectives.
4. The category of R-modules has both projective and injective resolutions (projective resolutions

comes from free resolutions, while injective resolutions comes from repeated application of theorem 16.3.17)

It is thus helpful to add the following condition to an abelian category

Definition 27.2.2: Enough Projective And Enough Injectives

Let $\mathbf{A}$ be an abelian category. Then:

1. we say that $\mathbf{A}$ as *enough projectives* if for every object A of $\mathbf{A}$, there exists a projective object P of $\mathbf{A}$ and an epimorphism $P \twoheadrightarrow A$.
2. we say that $\mathbf{A}$ as *enough injectives* if for every object A of $\mathbf{A}$, there exists a injective object Q of $\mathbf{A}$ and a monomorhpism $A \hookrightarrow Q$.

With the important definitions established, the main objective of this section may start going under way. The goal is to show that if we have a quasi-isomorphism between two bounded-above complexes of projectives $f^\bullet : P_0^\bullet \to P_1^\bullet$ (resp. quasi-isomorphism between two bounded-bellow complexes of injectives $f^\bullet : Q_0^\bullet \to Q_1^\bullet$), then $f^\bullet$ is in fact a homotopy-equivalence, from which it can be concluded that in $K^-(\mathbf{P})$ and $K^+(\mathbf{I})$ (homotopy-classes of) quasi-isomorphisms are isomorphisms, which shall give us the desired derived category $D^-(\mathbf{A})$ and $D^+(\mathbf{A})$ in the case where A has enough projectives (resp. enough injectives).

Let us now revisit this splitting condition. Recall that if $C^\bullet$ in $\mathrm{Ch}(\mathbf{A})$ is exact, then the identity map $\mathrm{id}_{C^\bullet}$ and the trivial map induce the same morphisms in (co)homology. However, there do exist exact complexes that are *not* homotopic to 0. As discussed earlier, this directly links to the splitting condition, namely the identity is homotopic to 0 if the sequence is split exact. Being projective is exactly the condition that causes splitting behavior. With this in mind, let us prove the first lemma:

Lemma 27.2.3: Zero-morphism In Cohomology And Homotopy

Let $\mathbf{A}$ be an abelian category, $P^\bullet \in \mathrm{Ch}^-(\mathbf{P})$ (i.e. $P^i = 0$ for $i > 0$), and let $C^\bullet$ be a complex in $\mathrm{Ch}(\mathbf{A})$ such that $H^i(C^\bullet) = 0$ for $i < 0$ (i.e. has trivial cohomology in negative coefficients). Let $f^\bullet : P^\bullet \to C^\bullet$ be a morphism of complexes that induces the zero-morphism in cohomology: $H^\bullet(f^\bullet) = 0$. Then

$$f^\bullet \sim 0$$

i.e., $f^\bullet$ is null-homotopic.

The injective version of this statement is comes from dualizing the appropriate terms.

Proof **:**
To show $f^\bullet$ is null-homotopic, we must show there exists maps $h^i : P^i \to C^{i-1}$ such that

$$f^i = d_{C^\bullet}^{i-1} \circ h^i + h^{i+1} \circ d_{P^\bullet}^i \tag{27.5}$$

as a diagram:

$$\begin{tikzcd}
\cdots \arrow[r] & P^{-2} \arrow[r] \arrow[d, "f^{-2}"] & P^{-1} \arrow[r] \arrow[d, "f^{-1}"] & P^0 \arrow[r] \arrow[d, "f^0"] & 0 \arrow[r] \arrow[d, "h^1=0"] & \cdots \\
\cdots \arrow[r] & C^{-2} \arrow[r] \arrow[ul, "h^{-2}"] & C^{-1} \arrow[r] \arrow[ul, "h^{-1}"] & C^0 \arrow[r] \arrow[ul, "h^0"] & C^1 \arrow[r] & \cdots
\end{tikzcd}$$

Naturally, $h^i = 0$ for $i > 0$. For the rest, we shall use induction. The base case is $i = 0$. Since the cohomology induced by $f^\bullet$ is zero, f^0 factors through the image of $d^{-1}_{C^\bullet}$:

$$\begin{tikzcd} & P^0 \arrow[d, "f^0"] & \\ C^{-1} \arrow[r, "d^{-1}_{C^\bullet}"] & \operatorname{im} d^{-1}_{C^\bullet} \arrow[r] & 0 \end{tikzcd}$$

Now, since P^0 is projective, there exists a morphism $h^0 : P^0 \to C^{-1}$ such that $f^0 = d^{-1}_{C^\bullet} \circ h^0$, as needed.

Next, assume h^i and h^{i+1} satisfies equation (27.5). The goal is to show h^{i-1} satisfies it too. By equation (27.5) and the fact that $P^\bullet$ is a complex:

$$d^{i-2}_{C^\bullet} \circ h^i \circ d^{i-1}_{P^\bullet} = (f^i - h^{i+1} \circ d^{i-1}_{P^\bullet}) = f^i \circ d^{i-1}_{P^\bullet} - h^{i+1} \circ 0 = f^i \circ d^{i-1}_{P^\bullet}$$

which can be seen by staring at the following diagram:

$$\begin{tikzcd} \cdots \arrow[r, dotted] & P^{i-1} \arrow[r, "d^{i-1}_{P^\bullet}"] \arrow[d, dotted] & P^i \arrow[r, "d^i_{P^\bullet}"] \arrow[d, "f^i"] \arrow[dl, "h^i"'] & P^{i+1} \arrow[r, dotted] \arrow[d, dotted] \arrow[dl, "h^{i+1}"] & \\ \cdots \arrow[r, dotted] & C^{i-1} \arrow[r, "d^{i-1}_{C^\bullet}"'] & C^i \arrow[r, dotted] & C^{i+1} \arrow[r, dotted] & \end{tikzcd}$$

Thus, since f^b is a morphism of cochain complexes:

$$d^{i-1}_{C^\bullet} \circ (f^{i-1} - h^i \circ d^{i-1}_{P^\bullet}) = d^{i-1}_{C^\bullet} \circ f^{i-1} - f^i \circ d^{i-1}_{P^\bullet} = 0$$

Importantly, $(f^{i-1} - h^i \circ d^{i-1}_{P^\bullet})$ factors through $\ker d^{i-1}_{C^\bullet}$. Since $C^\bullet$ is exact at C^{i-1} for $i \leq 0$, $\ker d_{C^\bullet} = \operatorname{im}^{i-2}_{C^\bullet}$. Thus, we get:

$$\begin{tikzcd} & P^{i-1} \arrow[d, "f^{i-1} - h^i \circ d^{i-1}_{P^\bullet}"] & \\ C^{i-2} \arrow[r, "d^{i-2}_{C^\bullet}"] & \operatorname{im}(d^{i-2}_{C^\bullet}) \arrow[r] & 0 \end{tikzcd}$$

Now, since P^{i-1} is projective, there exists a lift $h^{i-1} : P^{i-1} \to C^{i-2}$ such that

$$f^{i-1} - h^i \circ d^{i-1}_{P^\bullet} = d^{i-2}_{C^\bullet} \circ h^{i-1}$$

which, shuffling around, gives us

$$f^{i-1} = d^{i-2}_{C^\bullet} \circ h^{i-1} + h^i \circ d^{i-1}_{P^\bullet}$$

as we sought to show.

Corollary 27.2.4: Exactness Cohomology And Homotopy

Let $\mathbf{A}$ be an abelian category, $P^\bullet \in \mathrm{Ch}^-(\mathbf{A})$, and $C^\bullet \in \mathrm{Ch}(\mathbf{S})$ an exact complex. Then every morphism $P^\bullet \to L^\bullet$ is homotopic to 0

***Proof* :**
Since every morphism to an exact complex induces zero-morphism in cohomology, this is a direct application of lemma 27.2.3.

Note the importance of the bounding from above/bellow (depending on whether working with projectives or injective) in the inductive argument, without it there wouldn't be an appropriate base case. Using the above results, the first quasi-isomorphism to homotopy equivalence can be established:

Corollary 27.2.5: Projectives With Quasi-isomorphism And Zero Morphisms

Let $P^\bullet$ (resp. $Q^\bullet$) be a bounded-above exact complex of projectives (resp. a bounded bellow exact complex of injectives). Then $P^\bullet$ (resp. $Q^b b$) is homotopy-equivalent to the zero-complex.

***Proof* :**
exercise (Aluffi made if 5.12 in his book, see p. 623)

Another important result that will be used in the next section is that quasi-isomorphisms are "non-zero-divisors" up to homotopy wrt morphisms from commplexes of projectives:

Lemma 27.2.6: Quasi-isomorphisms Of Projectives And Zero-divisor

Let $\mathbf{A}$ be an abelian category, $f^\bullet : C^\bullet \to D^\bullet$ be a quasi-isomorphism. Let $P^\bullet$ be a bounded-above complex of projectives, and let $\alpha^\bullet : P^\bullet \to C^\bullet$ be a morphism of cochian complexes st the composition:

$$P^\bullet \xrightarrow{\alpha^\bullet} C^\bullet \xrightarrow{f^\bullet} D^\bullet$$

is homotopic to the zero-morphism. Then $\alpha^\bullet \sim 0$

***Proof* :**
Let $h^i : P^i \to D^{i-1}$ define the homotopy between $f^\bullet \circ \alpha^\bullet$ and 0, namely:

$$-f^i \circ \alpha^i = d^{i-1}_{C^\bullet} \circ h^i + h^{i+1} \circ d^i_{P^\bullet}$$

Take the mapping cone $MC(f)^\bullet$ of $f^\bullet$ and define the morphisms:

$$\beta^i = (\alpha^{\cdot} h^i) : P^i \to MC(f)^{i-1}$$

which may be visualized in the following diagram:

$$\begin{tikzcd}
\cdots \arrow[r] & P^{i-1} \arrow[r, "d^{i-1}_{P^\bullet}"] \arrow[d, dotted, "\beta^{i-1}"] & P^i \arrow[r, "d^i_{P^\bullet}"] \arrow[d, "\beta^i"'] & P^{i+1} \arrow[r] \arrow[d, "\beta^{i+1}"] & {} \\
\cdots \arrow[r] & C^{i-1} \oplus D^{i-2} \arrow[r, "d^{i-2}_{MC(f)^\bullet}"'] & C^i \oplus D^{i-1} \arrow[r, "d^{i-1}_{MC(f)^\bullet}"'] & C^{i+1} \oplus D^i \arrow[r] & {}
\end{tikzcd}$$

These give a morphism of cochain complexes $P^\bullet \to MC(f)[-1]^\bullet$. Indeed:

$$\begin{aligned}
\beta^{i+1} \circ d^i_{P^\bullet} &= (\alpha^{i+1} \circ d^i_{P^\bullet}, h^{i+1} \circ d^i_{P^\bullet}) \\
-d^{i-1}_{MC(f)} \circ \beta^i &= (d^i_{C^\bullet} \circ \alpha^i, -f^i \circ \alpha^i - d^{i-1}_{M^\bullet} \circ h^i)
\end{aligned}$$

Then by definition of h^i and by the morphism property of $\alpha^\bullet$, these are equal. Now, by hypothesis, $f^\bullet$ is a quasi-isomorphism. Then by proposition 27.1.14 $MC(f)^\bullet$ is an exact complex. Then by corollary 27.2.5 $\beta^\bullet$ is homotopic to 0. Since $\alpha^\bullet$ is the composition:

$$P^\bullet \xrightarrow{\beta^\bullet} MC(f)[-1]^\bullet = L^\bullet \oplus M[-1]^\bullet \to L^\bullet$$

it follows that $\beta^\bullet \sim 0$ implies $\alpha^\bullet \sim 0$, as we sought to show.

Though $\alpha^\bullet$ must be homotopic to 0, it need not be the case that $\alpha^\bullet = 0$. Take for example:

$$\begin{tikzcd}
\cdots \arrow[r] & 0 \arrow[r] \arrow[d] & \mathbb{Z} \arrow[r] \arrow[d, "\mathrm{id}"] & \mathbb{Z} \arrow[r] \arrow[d, "\cdot\, 2"] & 0 \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & 0 \arrow[r] \arrow[d] & \mathbb{Z} \arrow[r, "\cdot\, 2"] \arrow[d] & \mathbb{Z} \arrow[r] \arrow[d, "\pi"] & 0 \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & 0 \arrow[r] & 0 \arrow[r] & \mathbb{Z}/2\mathbb{Z} \arrow[r] & 0 \arrow[r] & \cdots
\end{tikzcd}$$

In this case, $f^\bullet$ is a quasi-isomorphism and $f^\bullet \circ \alpha^\bullet = 0$, and so $\alpha^\bullet \sim 0$, even though it is not equal to 0. With this, one more result is required before the desired goal:

Proposition 27.2.7: projectives, and one-sided inverses

Let $\mathbf{A}$ be an abelian category, $L^\bullet \in \mathrm{Ch}(\mathbf{A})$, $P^\bullet \in \mathrm{Ch}^-(\mathbf{P})$, and $f^\bullet : C^\bullet \to P^\bullet$ a quasi-isomorphism. Then there exists a morphism of complexes $g^\bullet : P^\bullet \to C^\bullet$ such that:

$$f^\bullet \circ g^\bullet \sim \mathrm{id}_{P^\bullet}$$

For injective, $Q^\bullet \in \mathrm{Ch}^+(\mathbf{I})$, $f^\bullet : Q^\bullet \to C^\bullet$ is a quasi-isomorphism, for which then there exists a morphism $g^\bullet : C^\bullet \to Q^\bullet$ such that $g^\bullet \circ f^\bullet \sim \mathrm{id}_{Q^\bullet}$. Notice that this is a one-sided inverse, hence is a slightly more general result than what is being built up to.

***Proof* :**
Since $f^\bullet : C^\bullet \to P^\bullet$ is a quasi-isomorphism, $MC(f)^\bullet = L[1]^\bullet \oplus P^\bullet$ is an exact complex. Define

$\rho^\bullet = (0, \mathrm{id}_{P^\bullet})$ to be the morphism:

$$\rho^\bullet : P^\bullet \to MC(f)^\bullet$$

By corollary 27.2.4, since $MC(f)^\bullet$ is exact, $\rho^\bullet$ is homotopic to the zero morphism. Thus, there exists morphisms:

$$\widehat{h}^i : P^i \to MC(f)^{i-1} = L^i \oplus P^{i-1}$$

such that

$$\rho^i = d^{i-1}_{MC(f)^\bullet} \circ \widehat{h}^i + \widehat{h}^{i+1} \circ d^i_{P^\bullet}$$

Visually, we may see this as:

$$\begin{tikzcd}
\cdots \arrow[r] & P^{i-1} \arrow[r, "d^{i-1}_{P^\bullet}"] \arrow[d, "\rho^{i-1}"] & P^i \arrow[r, "d^i_{P^\bullet}"] \arrow[d, "\rho^i"] \arrow[dl, "\widehat{h}^i"'] & P^{i+1} \arrow[r] \arrow[d, "\rho^{i+1}"] \arrow[dl, "\widehat{h}^{i+1}"'] & \cdots \\
\cdots \arrow[r] & C^i \oplus P^{i-1} \arrow[r, "d^{i-2}_{MC(f)^\bullet}"'] & C^{i+1} \oplus P^i \arrow[r, "d^{i-1}_{MC(f)^\bullet}"'] & C^{i+2} \oplus P^{i+1} \arrow[r] & \cdots
\end{tikzcd}$$

Now, write out $\widehat{h}^i$ out in components so that $\widehat{h}^i = (g^i, h^i)$ where $g^i : P^i \to C^i$ and $h^i : P^i \to P^{i-1}$. We'll show that the collection $\{g^i\}_{i\in\mathbb{Z}}$ is a morphism of cochian complexes (i.e. they form commutative squares $g^{i+1} \circ d^i_{P^\bullet} = d^i_{C^\bullet} \circ g^i$) such that $f^\bullet \circ g^\bullet \sim \mathrm{id}_{P^\bullet}$. Indeed, since

$$\begin{aligned}
(0, \mathrm{id}_{P^i}) &= \rho^i \\
&= d^{i-1}_{MC(f)^\bullet} \circ \widehat{h}^i + \widehat{h}^{i+1} \circ d^i_{P^\bullet} \\
&= (d_{MC}(f)^\bullet)^{i-1} \circ (g^i, h^i) + (g^{i+1}, h^{i+1}) \circ d^i_{P^\bullet} \\
&= (-d^i_{C^\bullet} \circ g^i, f^i \circ g^i + d^{i-1}_{P^\bullet} \circ h^i) + (\beta^{i+1} \circ d^i_{P^\bullet}, h^{i+1} \circ d^i_{P^\bullet})
\end{aligned}$$

It follows that:

$$\begin{aligned}
g^{i+1} \circ d^i_{P^\bullet} - d^i_{C^\bullet} \circ g^i &= 0 \\
\mathrm{id}_{P^i} - f^i \circ g^i &= d^{i-1}_{P^\bullet} \circ h^i + h^{i+1} \circ d^i_{P^\bullet}
\end{aligned}$$

which is exactly what we need, hence the morphism of complexes $g^\bullet : P^\bullet -> C^\bullet$ is the needed homotopy right-inverse of $f^\bullet$, as we sought to show.

With this, we can finally prove the desired result:

Theorem 27.2.8: Projective, Quasi-isomorphism Becomes Homotopy

Let $\mathbf{A}$ be an abelian category, and let $\alpha^\bullet : P^\bullet_0 -> P^\bullet_1$ be a quasi-isomorphism between bounded-above complexes of projectives (resp. quasi-isomorphism between two bounded-bellow complexes of injectives $f^\bullet : Q^\bullet_0 \to Q^\bullet_1$). Then $\alpha^b b$ is a homotopy-equivalence.

Proof **:**
The argument for projectives shall be given, the injective argument following by appropriate dualizing. By proposition 27.2.7, since $P_1^\bullet$ is in $C^-(\mathbf{P})$ and $\alpha^\bullet : P_0^\bullet \to P_1^\bullet$ is a quasi-isomorphismm, there exists a morhpism of complexes $\beta^\bullet : P_1^\bullet \to p_0^\bullet$ such that

$$\alpha^\bullet \circ \beta^\bullet \sim \mathrm{id}_{P_1}^\bullet$$

Then by proposition 27.1.19, $H^\bullet(\alpha^\bullet) \circ H^\bullet(\beta^\bullet) = \mathrm{id}$. Since $H^\bullet(\alpha^\bullet)$ is invertible, so is $H^\bullet(\beta^\bullet)$, and so $\beta^\bullet$ must be a quasi-isomorphism. Since $P_0^\bullet$ is also in $\mathrm{Ch}^-(\mathbf{P})$, we have that there exists a morphism $(\alpha^\bullet)' : P_0^\bullet \to P_1^\bullet$ such that $\beta^\bullet \circ (\alpha^\bullet)' \sim \mathrm{id}_{P_0^\bullet}$. Since:

$$(\alpha^\bullet)' \sim (\alpha^\bullet \circ \beta^\bullet) \circ (\alpha^\bullet)' = \alpha^\bullet(\beta^\bullet \circ (\alpha^\bullet)') \sim \alpha^\bullet$$

it follows that $\beta^\bullet \sim \alpha^\bullet \sim_{P_0^\bullet}$ (exercise 4.6 in Aluffi). Thus, $\alpha^\bullet$ is indeed a homotopy equivalence, as we sought to show.

Corollary 27.2.9: Projective In Homotopy Category

In $K^-(\mathbf{P})$ and $K^+(\mathbf{I})$, (homotopy classes of) quasi-isomorphissm are isomorphisms.

Proof **:**
follows immediately form theorem 27.2.8.

27.2.1 Resolutions

With the important equivalence between quasi-isomorphisms and homotopy-equivalence established in the case of complexes of projective (resp. projective) objects, we turn back our attention to the question of finding (co)homological invariants, focusing on categories with enough projective s(resp. Injectives). In particular, given an object A in an abelian category $\mathbf{A}$, it wouldb e desirable to associate to it a (co)chain of projective (resp. injective) objects in some universal way (up to homotopy) from which information can be extracted.

Let us first refine the definition of projective/injective resolutions of an object. Recall that there i a copy of $\mathbf{A}$ in $\mathrm{Ch}(\mathbf{A})$ given by mapping $A \mapsto \iota(A)$.

Definition 27.2.10: Projective And Injective Resolution

Let $A \in \mathrm{Obj}(\mathbf{A})$ be an object in an abelian category A. Then:

1. A *resolution* of A is a quasi-isomorphism $C^\bullet \to \iota(A)$ for some complex $C^\bullet \in \mathrm{Obj}(\mathrm{Ch}(\mathbf{A}))$.
2. A *projective resolution* of A is a quasi-isomorphism $P^\bullet \to \iota(A)$ where $P^\bullet$ is a complex in $\mathrm{Ch}^{\leq 0}(\mathbf{P})$
3. An *injective resolution* of A is a quasi-isomorphism $\iota(A) \to Q^\bullet$ where $Q^\bullet$ is a complex in $\mathrm{Ch}^{\geq 0}(\mathbf{I})$

Often, $P^\bullet$ and $Q^\bullet$ shall be referred to as the projective resolution and injective resolution respectively, omitting the quasi-isomorphism.

66: Remark

There is a point of confusion that may arise. It may seem that the resolutions *themselves* are projective/injective objects in the abelian category $\mathrm{Ch}(\mathbf{A})$. However, this is not the case (exercise ref:HERE).

Recall further that any resolution $C^\bullet$ will be exact everywhere except possibly at C^0 (exercise, or double-check example 27.1.4). Hence, the non-trivial cohomological information lies in $H^0(C^\bullet)$. Let us now focus on resolutions for a particular object A: consider the category of homotopy classes of quasi-isomorphisms with *target* $\iota(A)$. Then we shall show that, if they exist, projective resolution are initial in this category.

$$\begin{tikzcd} P^\bullet \arrow[d, "\exists!"] \arrow[dr] & \\ C^\bullet \arrow[r, "\text{q-iso}"'] & \iota(A) \end{tikzcd}$$

Analogously, injective resolutions are final in homotopic category of quasi-isomorphisms with *source* $\iota(A)$. Given this universal property, this means that each projective resolution maps uniquely (in $K(\mathbf{A})$) to *every* resolution of A. The following lemma proves this:

Lemma 27.2.11: Projective Resolution As Initial Objects

Let $A \in \mathbf{A}$, $C^\bullet$ a resolution of A (note that $H^0(C^\bullet) = A$), and $P^\bullet \in \mathrm{Ch}^{\leq 0}(\mathbf{P})$. Then any morphism

$$\varphi : H^0(P^\bullet) \to H^0(C^\bullet) \ [= A]$$

induces a morphism of complexes $\varphi^\bullet : P^\bullet \to C^\bullet$ that is unique up to homotopy equivalence.

Proof **:**
The goal is to define $\varphi^i : P^i \to C^i$ for each $i \in \mathbb{Z}$. Since $\varphi^i = 0$ for $i > 0$, it is safe to replace $C^\bullet$

with its truncated version. From this, consider the following diagram:

$$\begin{tikzcd}
\cdots \arrow[r] & P^{-2} \arrow[r, "d^{-2}_{P^\bullet}"] \arrow[d, dashed, "\varphi^2"] & P^{-1} \arrow[r, "d^{-1}_{P^\bullet}"] \arrow[d, dashed, "\varphi^1"] & P^0 \arrow[r, "\pi"] \arrow[d, dashed, "\varphi^0"] & H^0(P^\bullet) \arrow[r] \arrow[d, "\varphi"] & 0 \\
\cdots \arrow[r] & C^{-2} \arrow[r, "d^{-2}_{C^\bullet}"] & C^{-1} \arrow[r, "d^{-1}_{C^\bullet}"] & \ker d^0_{C^\bullet} \arrow[r, "\mu"] & H^0(C^\bullet)\ [= A] \arrow[r] & 0
\end{tikzcd}$$

Then by the amendment to the resolution, the bottom complex is now exact.

Now, P^0 is projective, and maps to A via $\varphi \circ \pi$. Hence, by theorem 16.2.3 there exists a lift, say φ^0

$$\begin{tikzcd}
P^0 \arrow[d, "\varphi^0"] \arrow[dr, "\varphi\circ\pi"] & \\
\ker(d^0_{C^\bullet}) \arrow[r, "\mu"] & A
\end{tikzcd}$$

Next, since $\ker(d^0_{C^\bullet}) \hookrightarrow M^0$, we have a map $\varphi^0 : P^0 : M^0$. Next, notice that

$$\mu \circ (\alpha^0 = d^{-1}_{P^\bullet} = \varphi \circ \pi \circ d^{-1}_{P^\bullet} = 0)$$

hence, $\alpha^0 \circ d^{-1}_{P^\bullet}$ factors through $\ker \mu$. Since the bottom is exact, we have $\operatorname{im} d^{-1}_{C^\bullet}$. Thus, we have the following diagram:

$$\begin{tikzcd}
 & P^{-1} \arrow[d, "\varphi^0 \circ d^{-1}_{P^\bullet}"] & \\
C^{-1} \arrow[r, "d^{-1}_{C^\bullet}"] & \operatorname{im}(d^{-1}_{C^\bullet} \arrow[r] & 0
\end{tikzcd}$$

hence, there exists a list φ^{-1}. From this, we may continue to inductively construct each φ^{-1} for $i > 1$, using exactness of the bottom. This then gives a morphism of complexes $\varphi^\bullet : P^\bullet \to M^\bullet$. Its uniqueness up to homotopy is given by lemma (ref;HERE), ctp.

Since these are initial (resp. final) objects in the category, we have a unique isomorphism beetween any two solutions to the universal problem. In this case, this may be stated as followed:

Proposition 27.2.12: Unique Projection Resolution

Let $\mathbf{A}$ be an abelian category. Then any two projective (resp. injective) resolutions of an object A in $\mathbf{A}$ are homotopy equivalent

Proof **:**
Follows the above remark, but can also be directly deduced using the above lemma.

A final important observation is that this lift is functorial!

Proposition 27.2.13: Projective Resolution Functoriality

Let $\mathbf{A}$ be an abelian category, and let A_0, A_1 be objects of $\mathbf{A}$. Let $P_0^\bullet$ and $P_1^\bullet$ be projective resolutions for A_0 and A_1 respectively. Then every morphism $\varphi : A_0 \to A_1$ in $\mathbf{A}$ induces a morphism $\alpha^\bullet : P_0^\bullet \to P_1^\bullet$ that is unique up to homotopy

***Proof* :**
Notice that φ can be interpreted as a morphism $\varphi : H^0(P_0^\bullet) \to H^0(P_1^\bullet)$. The complex $P_0^\bullet$ consists of projectives, and $P1^\bullet$ is a resolution of A_1. Thus, by lemma 27.2.11, there exists a list α^b that is unique up to homotopy, as we sought to show.

Corollary 27.2.14: Full Subcategory Of Homotopic Category

Let $\mathbf{A}$ be an abelian category with enough projectives (resp. enough injectives). Then $\mathbf{A}$ is equivalent to a full subcategory of $K^-(\mathbf{P})$ (resp. $K^+(\mathbf{I})$).

***Proof* :**
important exercise!

27.2.2 Derived Category with Enough Projectives

Let us finally confirm that the solution to the universal problem in equation 27.2. Recall that the derived category is a category $D(\mathbf{A})$ where for any category D for which there exists an additive functor $\mathscr{F} : \mathrm{Ch}(\mathbf{A}) \to D$ where every quasi-isomorphism in $\mathrm{Ch}(\mathbf{A})$ becomes an isomorphism in D, there exists a unique functor such that the following diagram commutes:

$$\begin{tikzcd} \mathrm{Ch}(\mathbf{A}) \arrow[r, "\mathscr{F}"] \arrow[d] & D \\ D(\mathbf{A}) \arrow[ur, "\exists!"'] & \end{tikzcd}$$

In the case where $\mathbf{A}$ has enough projectives (resp. enough injectives) and is bounded from above (resp. from bellow), then $K^-(\mathbf{P})$ (resp. $K^+(\mathbf{I})$) is the solution to the above universal problem. In particular, there is an equivalence of categories between $K^-(\mathbf{P})$ and $D^-(\mathbf{A})$. The formal definition of the derived category shall be given in section ref:HERE, hence this equivalence shall not be proven, however the universal property is readily provable, and once the derived category is formally defined it will be the equivalence essentially come to proper manipulation of universal properties.

To start, a general procedure for finding projective/injective resolutions for any (co)chain must be explored. The discussion so far has been for projective/injective resolution, that is quasi-isomorphisms $P^\bullet \to \iota(A)$ or $\iota(A) \to Q^\bullet$. Is it possible to replace $\iota(A)$ with any bounded complex? The answer is yes, though rather technical. The importance of this will be the existence of a functor that maps $\mathrm{Ch}^-(\mathbf{P})$ into $K^-(\mathbf{P})$, which shall be important for solving the universal problem above, as well as finding functors between derived categories. Such functors shall also be key in solving the "extension problem" that was mentioned all the way back in chapter 16.

Theorem 27.2.15: Resolution And Complexes

Let $\mathbf{A}$ be an abelian category with enough projectives and $C^\bullet$ a complex in $C^-(\mathbf{A})$. Then there exists a founded-above complex of projective $P^\bullet$ and a quasi-isomorphism $P^\bullet \to C^\bullet$, where $P^\bullet$ is uniquely defined up to homotopy equivalence. Furthermore, every morphism $\alpha^\bullet$ of complexes in $C^-(\mathbf{A})$ lifts to a morphism of the corresponding projective resolutions in $K^-(\mathbf{P})$, which is uniquely defined up to homotopy.

Proof **:**
Aluffi p.632

Thus, given an abelian category $\mathbf{A}$ with enough projective, there is a functor

$$\mathscr{P} : C^-(\mathbf{A}) \to K^-(\mathbf{P})$$

associating each (bounded-above) complex $C^\bullet$ to a projective resolution $\mathscr{P}(C^\bullet)$ in a functorial manner (so that morphisms $f^\bullet$, $g^\bullet$ are lifted and $\mathscr{P}(g^\bullet \circ f^\bullet) = \mathscr{P}(g^\bullet) \circ \mathscr{P}(f^\bullet)$ since both sides are classes of homotopy lifts of $g^\bullet \circ f^\bullet$). The uniqueness of $\mathscr{P}$ is up to natural transformation, that is if another functor $\mathscr{P}'$ is chosen that assigns a different choice of projective resolution, then by theorem 27.2.15 there is a homotopy equivalence between $\mathscr{P}(C^\bullet)$ and $\mathscr{P}'(C^\bullet)$, hence an isomorphism $\varphi_{C^\bullet} : \mathscr{P}(C^\bullet) \to \mathscr{P}'(C^\bullet)$ in $K^-(\mathbf{P})$. Since the identity on $C^\bullet$ lifts to a unique map up to homotopy, the isomorphism $\varphi_{C^\bullet}$ is unique. Furthermore, given a morphism $f^\bullet : C^\bullet \to D^\bullet$, the diagram:

$$\begin{tikzcd}
\mathscr{P}(C^\bullet) \arrow[r, "\mathscr{P}(f^\bullet)"] \arrow[d, "\varphi_{C^\bullet}"', "\sim"] & \mathscr{P}(D^\bullet) \arrow[d, "\varphi_{D^\bullet}", "\sim"'] \\
\mathscr{P}'(C^\bullet) \arrow[r, "\mathscr{P}'(f^\bullet)"'] & \mathscr{P}'(D^\bullet)
\end{tikzcd}$$

commutes (again by the uniqueness of lifts up to homotopy, can you see why)? Thus, we have the unique natural isomorphism for any two choices of functors giving a projective resolution.

Theorem 27.2.16: Universal Property Of Derived Category With Enough Projectives

Let $\mathbf{A}$ be an abelian category with enough projectives, and $\mathscr{P}$ a functor sending bounded-above cochains to projective resolutions. Then:

1. If $f^\bullet$ is a quasi-isomorphism in $C^-(\mathbf{A})$, then $\mathscr{P}(f^\bullet)$ is an isomorphism in $K^-(\mathbf{P})$
2. If $\mathscr{F} : C^-(\mathbf{A}) \to D$ is an additive functor that takes quasi-isomorphisms to isomorphisms, then there exists a unique functor $\widetilde{\mathscr{F}} : K^-(\mathbf{P}) \to D$ unique up to natural isomorphism such that the following diagram commutes up to natural isomorphism:

$$\begin{tikzcd} C^-(\mathbf{A}) \arrow[r, "\mathscr{F}"] \arrow[d, "\mathscr{P}"'] & D \\ K^-(\mathbf{P}) \arrow[ur, "\widetilde{\mathscr{F}}"'] & \end{tikzcd}$$

   where commuting up to natural transformation means there exists a natural transformation $\nu : \widetilde{\mathscr{F}} \circ \mathscr{P} \rightsquigarrow \mathscr{F}$ such that:

$$\nu_{C^\bullet} : \widetilde{\mathscr{F}} \circ \mathscr{P}(C^\bullet) \xrightarrow{\sim} \mathscr{F}(C^\bullet)$$

   is an isomorphism for every complex $C^\bullet$ in $C^-(\mathbf{A})$.

The fact that the commutative diagram commutes up to natural isomorphism reflects that the homotopic category is categorically equivalent to the derived category. If $\mathbf{A}$ has enough injective, then we may instead define the functor $\mathscr{I} : C^+(\mathbf{A}) \to K^+(\mathbf{I})$ that shall satisfy the dual universal property.

Proof **:**

1. Let $f^\bullet : C^\bullet \to D^\bullet$ be any morphism in $C^-(\mathbf{A})$. By theorem 27.2.15, $f^\bullet$ lifts to a morphism of resolutions which commutes up to homotopy in the following diagram:

$$\begin{tikzcd} \mathscr{P}(C^\bullet) \arrow[r, "\mathscr{P}(f^\bullet)"] \arrow[d, "\mu_{C^\bullet}"'] & \mathscr{P}(C^\bullet) \arrow[d, "\mu_{D^\bullet}"] \\ C^\bullet \arrow[r, "f^\bullet"] & D^\bullet \end{tikzcd}$$

   where $\mu_{C^\bullet}, \mu_{D^\bullet}$ are the quasi-isomorphisms of the projective resolutions. Since the diagram commute up to homootpy, it commutes after taking cohomology, thus since $f^\bullet$ is a quasi-isomorphism, so is $\mathscr{P}(f^\bullet)$. Then by corollary 27.2.9, $\mathscr{P}(f^\bullet)$ is an isomorphism.

2. Define $\widetilde{\mathscr{F}} : K^-(\mathbf{P}) \to D$ via:

$$\widetilde{\mathscr{F}}(P^\bullet) := \mathscr{F}(P^\bullet) \qquad \widetilde{\mathscr{F}}(f^\bullet) := \mathscr{F}(f^\bullet)$$

   where $\widetilde{\mathscr{F}}(f^\bullet)$ can be any representative of $f^\bullet$. This map is well-defined, since homotopic morphisms have the same image in D (lemma 27.1.25). Naturally, $\widetilde{\mathscr{F}}$ is a functor. We need

to check the commutativity up to morphism of the diagram and the uniqueness of $\widetilde{\mathscr{F}}$ up to natural isomorphism.

First, let $C^\bullet \in C^-(\mathbf{A})$. Then (by construction of $\mathscr{P}$, there exists a quasi-isomorphism

$$\mu_{C^\bullet} : \mathscr{P}(C^\bullet) \to C^\bullet$$

Since $\mathscr{F}$ maps quasi-isomorphisms to isomorphism, the induced morphism $\nu_{C^\bullet} := \mathscr{F}(\mu_{C^\bullet})$:

$$\nu_{C^\bullet} : \widetilde{\mathscr{F}}(\mathscr{P}(C^\bullet)) = \mathscr{F}(\mathscr{P}(C^\bullet)) \to \mathscr{F}(C^\bullet)$$

is an isomorphism. What's left to check is that $\nu_{C^\bullet}$ defines a natural transformation. Let $f^\bullet : C^\bullet \to D^\bullet$ in $C^-(\mathbf{A})$. We want to show the following diagram commutes:

$$\begin{tikzcd}
\widetilde{\mathscr{F}}(\mathscr{C}) \arrow[r, "\widetilde{\mathscr{F}}(\mathscr{P}(f^\bullet))"] \arrow[d, "\nu_{C^\bullet}"'] & \widetilde{\mathscr{F}}(\mathscr{P}(D^\bullet)) \arrow[d, "\nu_{D^\bullet}"] \\
\mathscr{F}(C^\bullet) \arrow[r, "\mathscr{F}(f^\bullet)"] & \mathscr{F}(D^\bullet)
\end{tikzcd}$$

This diagram is obtained by applying $\mathscr{F}$ to the diagram discussed above the theorem:

$$\begin{tikzcd}
\mathscr{P}(C^\bullet) \arrow[r, "\mathscr{P}(f^\bullet)"] \arrow[d, "\varphi_{C^\bullet}"', "\sim"] & \mathscr{P}(D^\bullet) \arrow[d, "\varphi_{D^\bullet}", "\sim"'] \\
\mathscr{P}'(C^\bullet) \arrow[r, "\mathscr{P}'(f^\bullet)"'] & \mathscr{P}'(D^\bullet)
\end{tikzcd}$$

This diagram commutes up to homotopy, hence the other diagram commutes directly (lemma 27.1.25).

For uniqueness up to isomorphism, say $\overline{\mathscr{F}} : K^-(\mathbf{P}) \to D$ is another functor satisfying the properties. Then there exists a natural isomorphism:

$$\overline{\nu} : \overline{\mathscr{F}} \circ \mathscr{P} \rightsquigarrow \mathscr{F}$$

Now, for any $P^\bullet \in K^-(\mathbf{P})$, take the composition:

$$\overline{ff}(P^\bullet) \underset{\sim}{\overset{\overline{ff}(\mu_{P^\bullet})^{-1}}{\to}} \overline{ff}(\mathscr{P}((P^\bullet)) \underset{\sim}{\overset{\overline{\nu}_{P^\bullet}}{\to}} = \widetilde{ff}(P^\bullet)$$

If φ is this composition, then it is easy to verify that $\varphi : \overline{\mathscr{F}} \to \widetilde{\mathscr{F}}$ is a natural isomorphism, completing the proof.

27.3 Derived Functors: Ext and Tor

Let $\mathscr{F} : \mathbf{A} \to \mathbf{B}$ be an additive functor. Often, we study information about $\mathbf{A}$ by passing it through $\mathbf{B}$ (think of $- \otimes N$ or $\mathrm{Hom}_R(N, -)$). As alluded to in section 16.5, such information may be the number of extensions two objects may have, the amount of torsion an object may have. In section 16.5, it was claimed that this comes from "fixing" the lack of exactness of a left- or right-exact functor using

homological algebra. We are finally in a position to make this precise. In this section, we shall show how to take a (not-necessarily left/right exact) additive functor between two abelian categories with $\mathbf{A}$ having enough projectives (resp. injectives) $\mathscr{F} : \mathbf{A} \to \mathbf{B}$ and get a *derived functor*:

$$D^-(\mathscr{F}) : D^-(\mathbf{A}) \to D^-(\mathbf{B}) \qquad (\text{resp. } D^-(\mathscr{F}) : D^-(\mathbf{A}) \to D^-(\mathbf{B}))$$

This functor shall be used to develop a long-exact sequence in $\mathbf{B}$, given a short exact sequence of objects in $\mathbf{A}$.

27.3.1 Derived Functor

Let $\mathscr{F} : \mathbf{A} \to \mathbf{B}$ be an additive functor between two abelian categories. Then $\mathscr{F}$ extends to a functor:

$$\mathrm{Ch}(\mathscr{F}) : \mathrm{Ch}(\mathbf{A}) \to \mathrm{Ch}(\mathbf{B})$$

namely, if $C^\bullet$ is a complex:

$$\mathscr{F}(d^{i+1}_{C^\bullet}) \circ \mathscr{F}(d^i_{C^\bullet}) = \mathscr{F}(d^{i+1}_{C^\bullet} \circ d^i_{C^\bullet}) = \mathscr{F}(0) = 0$$

hence, $\mathrm{Ch}(\mathscr{F})(C^\bullet)$ is indeed a complex. Then since homotopic morphisms are sent to homotopic morphisms (lemma 27.1.22), $\mathrm{Ch}(\mathscr{F})$ induces a morphism $K(\mathscr{F})$:

$$K(\mathscr{F}) : K(\mathbf{A}) \to K(\mathbf{B})$$

From the results we have proved, we have the following commutative diagram:

$$\begin{tikzcd} \mathbf{A} \arrow[r, "\mathscr{F}"] \arrow[d, hook, "\iota"'] & \mathbf{B} \arrow[d, hook, "\iota"] \\ K(\mathbf{A}) \arrow[r, "K(\mathscr{F})"] & K(\mathbf{B}) \end{tikzcd}$$

Furthermore, if the complexes are bounded, then certainly $\mathrm{Ch}(\mathscr{F})$ sends bounded complexes to bounded complexes, and hence the same holds for $K(\mathscr{F})$. Now, if $\mathbf{A}$ has enough projectives so that we may consider $K^-(\mathbf{P})$, can we restrict $K(\mathbf{A})$ to $K(P(\mathbf{A}))$, and map to $K(P(\mathbf{B}))$ (where $P(\mathbf{A})$ and $P(\mathbf{B})$ are the projectives of $\mathbf{A}$ and $\mathbf{B}$ respectively)? This is not the case in general; for example we saw that $- \otimes N$ in R-**Mod** does not preserve projectives. If $\mathbf{B}$ has enough projectives, then we may choose a projective resolution functor $\mathscr{P}_B : \mathrm{Ch}^-(B) \to K^-(\mathbf{P}^{-1})$ mapping each complex to a projective resolution in a functorial manner. Since $\mathscr{P}_B$ sends homotopic morphisms to homotopic morphisms (lemma 27.2.6), $\mathscr{P}_B$ descends to a functor $K(\mathscr{P}_B)$. Combining these facts, we may define the following functor:

Definition 27.3.1: Left-Derived (right-Derived) Functor

Let $\mathscr{F} : \mathbf{A} \to \mathbf{B}$ be an additive functor. Then the *left-derived functor* of $\mathscr{F}$ is the composition:

$$\begin{tikzcd} K^-(P(\mathbf{A})) \arrow[r, "\iota_A"'] \arrow[rrr, bend left, "\mathbf{L}\mathscr{F}"] & K^-(\mathbf{A}) \arrow[r, "K(\mathscr{F})"'] & K^-(\mathbf{B}) \arrow[r, "K(\mathscr{P}_B))"'] & K^-(P(\mathbf{B})) \end{tikzcd}$$

where ι_A is the inclusion of $K^-(P(\mathbf{A}))$ as a full subcategory of $K^-(\mathbf{A})$. The resulting functor is denoted $\mathbf{L}\mathscr{F}$.

It may seem that the natural notation would be $D^-(\mathscr{F})$, however the $\mathbf{L}\mathscr{F}$ notation was established before the abstraction given above, and has stuck. Similarly, if $\mathbf{A}$, $\mathbf{B}$ has enough injectives, we may define the right-derived functor $\mathbf{R}\mathscr{F}$.

It is natural to want to characterize $\mathbf{L}\mathscr{F}$ and $\mathbf{R}\mathscr{F}$ in terms of some universal property. Naturally, this universal property would be defined up to natural isomorphism, for $\mathscr{P}_A$ and $\mathscr{P}_B$ are chosen up to natural isomorphism. This turns to be a bit subtle. For example, there is no natural functor which shall always make the following diagram commute:

$$\begin{tikzcd} \hat{\mathbf{A}} \arrow[r, dashed, "??"] \arrow[d, hook] & \hat{\mathbf{B}} \arrow[d, hook] \\ K^-(P(\mathbf{A})) \arrow[r, "\mathbf{L}\mathscr{F}"] & K^-(P(\mathbf{B})) \end{tikzcd}$$

The reason is, it is not necessarily the case that a complex whose cohomology is concentrated at 0 would map via $\mathbf{L}\mathscr{F}$ to a complex with the same property. In fact, the reader may expect this, for the image of a functor $\mathscr{F}$ may present some non-trivial cohomological information, (think of $\mathscr{F} := \mathrm{Hom}_R(N, -)$ where N is non-projective). If the functor $\mathscr{F}$ is *exact*, then then there is essentially no "change in cohomology". This is captured in the following proposition:

Proposition 27.3.2: Derived Functors And Exactness

Let $\mathscr{F}$ be an additive *exact* functor. Then $\hat{\mathbf{A}}$ is in fact sent to $\hat{\mathbf{B}}$ via $\mathbf{L}\mathscr{F}$

***Proof* :**
Let $P^\bullet \in \hat{\mathbf{A}}$ so that $H^i(P^\bullet) = 0$ for all $i \neq 0$. Then $\mathbf{L}\mathscr{F}(P^\bullet)$ is given by choosing a projective resolution $P^\bullet_{\mathscr{F}(P^\bullet)}$ of $\mathscr{F}(P^\bullet)$. Since quasi-isomorphissm preserve cohomology and (key!) exact functors *commute* with cohomology (exercise ref:HERE):

$$H^i(\mathbf{L}\mathscr{F}(P^\bullet)) = H^i(P^\bullet_{\mathscr{F}(P^\bullet)}) \cong H^i(\mathscr{F}(P^\bullet)) \cong \mathscr{F}(H^i(P^\bullet)) = \mathscr{F}(0) \overset{i\neq 0}{=} 0$$

Thus, $\mathbf{L}\mathscr{F}(P^\bullet)$ does indeed map to an object of $\hat{\mathbf{B}}$.

Even more is true if $\mathscr{F}$ is exact: $H^i(\mathbf{L}\mathscr{F}(P^\bullet)) \cong \mathscr{F}(H^0(P^\bullet))$, saying that the restriction of $\mathbf{L}\mathscr{F}$ agrees with $\mathscr{F}$ (modulo equivalence of categories). Hence, in general we shall be interested in non-exact functors, usually functors that are only left or right exact.

Returning to finding an appropriate universal property for derived functors, another natural candidate would be given by taking a "step back" from the above diagram and consider instead if the following diagram commutes (up to natural isomorphism)

$$\begin{tikzcd} K^-(\mathbf{A}) \arrow[r, "K(\mathscr{F})"] \arrow[d, "\mathscr{P}_A"] & K^-(\mathbf{B}) \arrow[d, "\mathscr{P}_B"] \\ K^-(P(\mathbf{A})) \arrow[r, "\mathbf{L}\mathscr{F}"'] & K^-(P(\mathbf{B})) \end{tikzcd}$$

in particular, that there may be a natural isomorphism:

$$\mathbf{L}\mathscr{F} \circ \mathscr{P}_A \overset{\sim}{\rightsquigarrow} \mathscr{P}_B \circ K(\mathscr{F})$$

Unfortunately, this isn't in general the case

Example 27.3.1: Lacking Natural Isomorphism
(exercise 7.3 in Aluffi, should I put solution here?)

The following is the best we can hope for:

Proposition 27.3.3: Universal Property of (left-)Derived Functor

Let $\mathscr{F} : \mathbf{A} \to \mathbf{B}$ be an additive functor between abelian categories. Then $\mathbf{L}\mathscr{F}$ satisfies the following universal property:

1. There is a natural transformation:

$$\mathbf{L}\mathscr{F} \circ \mathscr{P}_A \overset{\sim}{\rightsquigarrow} \mathscr{P}_B \circ K(\mathscr{F})$$

2. for every functor $\mathscr{G} : K^-(P(\mathbf{A})) \to K^-(P(\mathbf{B}))$ and every natural tarnsformation $\gamma : \mathscr{G} \circ \mathscr{P}_A \rightsquigarrow \mathscr{P}_B \circ K(\mathscr{F})$, there is a unique (up to natural isomorhpism) natural transformation $\mathscr{G} \rightsquigarrow \mathbf{L}\mathscr{F}$ inducign a factorization:

$$\gamma : \mathscr{G} \circ \mathscr{P}_A \rightsquigarrow \mathbf{L}\mathscr{F} \circ \mathscr{P}_A \rightsquigarrow \mathscr{P}_B \circ K(\mathscr{F})$$

A similar result holds for right-derived functors.

***Proof* :**
Aluffi p.644

(a word on composition)

27.3.2 Long Exact Sequences of Derived Functors

With the derived functor established and being the best candidate for holding (co)homology information, we proceed with extracting this informaiton: we may take $\mathbf{L}_i\mathscr{F} := H^{-i} \circ \mathbf{L}\mathscr{F}$. The reason for taking H^{-i} instead of i is since the indexing works out better. This produced a functor $\mathbf{L}_i\mathscr{F} : D^-(\mathbf{A}) \to B$. Notice that B doesn't need to have enough projectives for:

$$\mathbf{L}_i\mathscr{F}(P^\bullet) = H^{-i}(\mathscr{P}_B(\mathrm{Ch}(\mathscr{F})(P^\bullet))) \cong H^{-i}(\mathrm{Ch}(\mathscr{F})(P^\bullet))$$

where the last isomorphism comes from $\mathscr{P}_B(\mathrm{Ch}(\mathscr{F})(P^\bullet)))$ being quasi-isomorphic to $\mathrm{Ch}(\mathscr{F})(P^\bullet)$. Though this abstract approach is useful, the following more concrete version of $\mathbf{L}_i\mathscr{F}$ shall prove more useful for computations It feels like I should just skip to this, idk what the above really gives me here

Definition 27.3.4: ith Left-Derived Functor

Let $\mathscr{F} : \mathbf{A} \to \mathbf{B}$ be an additive functor between two abelian categories where $\mathbf{A}$ has enough projectives. Then the *ith left-derived functor* $\mathbf{L}_i\mathscr{F} : A \to B$ is given by:

$$\mathbf{L}_i\mathscr{F} := H^{-1} \circ \mathbf{L}\mathscr{F} \circ \mathscr{P}_A \circ \iota_A$$

The complex in $\mathrm{Ch}(\mathbf{B})$ with $\mathbf{L}_i\mathscr{F}(M)$ in the $-i$th degree with vanishing differentials is denoted $\mathbf{L}_\bullet\mathscr{F}(M)$.

To be precise, if $M \in \mathbf{A}$, then $\mathbf{L}_i(M)$ is found by

1. finding any projective resolution $P_M^\bullet$ of M
2. Applying $\mathrm{Ch}(\mathscr{F})$ to $P_M^\bullet$, $\mathrm{Ch}(\mathscr{F})(P_M^\bullet)$
3. take the $(-i)$th cohomology of this complex, $H^{-i}(\mathrm{Ch}(\mathscr{F})(P_M^\bullet))$

With all the build-up done so far, the resulting cohomology is independent of choice, for:

1. Any two projective resolutions are homotopy-equivalent
2. the image of homotopy-equivalent complexes via additive functors have isomorphic cohomology

Naturally, the *ith right-derived functor* $\mathbf{R}^i\mathscr{F}$ of an additive functor $\mathscr{F}$ and the assocaited complex $\mathbf{R}^\bullet\mathscr{F}$ is defined simplify (given $\mathbf{A}$ has enough injectives), namely

$$\mathbf{R}^i\mathscr{F} := H^i \circ \mathbf{R}\mathscr{F} \circ \mathscr{Q}_A \circ \iota_A$$

A note should be said about *contravriant functors*: since they can always be viewed as *covariant* functor $\mathbf{A}^{\mathrm{op}} \to B$, $\mathbf{A}$ would instead have to have the appropriate injectives or projectives swapped. The right-derived functor of an additive *contravariant functor* would need $\mathbf{A}$ to have enough *projectives*, since $\mathbf{A}^{\mathrm{op}}$ would have to have enough injectives. The $\mathbf{L}_i$ and $\mathbf{R}_i$ notation can be thought of as "looking" at the cohomology provided by $\mathscr{F}$. As mentioned earlier, if $\mathscr{F}$ is exact, then the derived functors are *trivial* in terms of cohomology, hence the derived functors are only interesting if $\mathscr{F}$ is not exact, at which point these functors give you interesting information.

Example 27.3.2: ith Derived Functors

It is finally time to bring pack section 16.5

1. Let $\mathbf{A} = R\text{-}\mathbf{Mod}$ for a commutative ring R, pick an R-module N, and define

$$- \otimes_R N$$

The left-derived functor of $- \otimes_R N$ is labeled

$$- \overset{\mathbf{L}}{\otimes}_R N : D^-(R\text{-Mod}) \to D^-(R\text{-Mod})$$

The ith derived left-functor of $- \otimes_R N$ is denoted:

$$\mathrm{Tor}_i^R(- \otimes_R N) = H^{-i}((-)_\bullet \otimes N)$$

where the right hand side of the inequality comes from the result elaborated in section 16.5. It was mentioned that a flat resolution may be used instead of a projective resolution, however this was does not currently fit within the theory. This shall be addressed in the next section.

2. Similarly, $\mathrm{Hom}_R(-, N)$ and $\mathrm{Hom}_R(N, -)$ admits a right derived functors

$$\mathrm{Ext}^i_R(M, N) = H^i(\mathrm{Hom}_R(M, N^\bullet))$$

For $\mathrm{Hom}_R(-, N)$, don't forget that one must take the opposite category and a projective resolution.

The most important part of left-/right-derived functors is their use in fixing exactness after applying a not necessarily exact functor $\mathscr{F}$. On a high level, since the derived category there is not an abelian category (hence not necessarily containing kernels and cokernels, nor is it really feasible to investigate such categories), the notion of *triangulated categories* is useful for categories where given a "short exact sequence", we can create a long exact sequence (that is, a category in which snakes lemma can be applied). We shall not investigate the idea of triangulated categories, but the following is a concrete results needed for any rigorous exploration of triangulated categories.

Assuming the whole that that our category $\mathbf{A}$ has enough projectives, take a short exact sequence:

$$0 \to L \to M \to N \to 0$$

Since there is enough projectives, can the projective resolutions be arranged to create a short exact sequence? The answer is yes, and given the following apt name:

Lemma 27.3.5: Horseshoe Lemma

Let

$$0 \to L \to M \to N0 \to 0$$

be an exact sequence in an abelian category $\mathbf{A}$ with enough projectives, and let $P^\bullet_L, P^\bullet_N$ be projections resolutions of L, N respectively. Then there exists an exact sequence:

$$0 \to P^\bullet_L \to P^\bullet_M \to P^\bullet_N \to 0$$

where $P^\bullet_M$ is a projective resolution of M, inducing the original sequence in cohomology (i.e. applying H^0)

***Proof* :**
Visually, we want to find objects and maps to fill in this commutative diagram:

$$\begin{tikzcd}
 & \vdots \arrow[d] & \vdots \arrow[d, dotted] & \vdots \arrow[d] & \\
0 \arrow[r] & P_L^{-2} \arrow[r, dotted] \arrow[d] & P_M^{-2} \arrow[r, dotted] \arrow[d, dotted] & P_N^{-2} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & P_L^{-1} \arrow[r, dotted] \arrow[d] & P_M^{-1} \arrow[r, dotted] \arrow[d, dotted] & P_N^{-1} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & L \arrow[r] & M \arrow[r] & N \arrow[r] & 0
\end{tikzcd}$$

where the dotted lines and the P_M^i are the goal (the name of the lemma comes from this horseshoe shaped diagram). I claim that $P_M^i = P_L^i \oplus P_N^i$ (which the reader should recall or reprove is projective) with the natural inclusion and projection maps shall work. The argument is done via induction. Since P_N^0 is projective and $M \to N$ is an epimorphism, the morphism $P_N^0 \to N$ lifts to a isomorphism m $\beta : P_N^0 \to M$. On the other hand, P_L^0 maps to M via $\alpha : P_L^0 \to L \to M$. Thus, there is a morphism:

$$\pi_M := \alpha \oplus \beta : P_L^0 \oplus P_N^0 \to M$$

diving us the diagram

$$\begin{tikzcd}
 & \vdots \arrow[d] & \vdots \arrow[d, dotted] & \vdots \arrow[d] & \\
0 \arrow[r] & P_L^{-2} \arrow[r, dotted] \arrow[d] & P_M^{-2} \arrow[r, dotted] \arrow[d, dotted] & P_N^{-2} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & P_L^{-1} \arrow[r, dotted] \arrow[d] & P_M^{-1} \arrow[r, dotted] \arrow[d, dotted] & P_N^{-1} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & P_L^0 \arrow[r] \arrow[d] & P_L^0 \oplus P_N^0 \arrow[r] \arrow[d] & P_N^0 \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & L \arrow[r] & M \arrow[r] & N \arrow[r] & 0
\end{tikzcd}$$

The snake lemma and π_L being an epimorphism gives that the kernels of the vertical maps form an exact sequence, and hence we have an epimorphism from P_L^{-1} and P_N^{-1} to the respective kernels since the column of the original diagram are exact:

$$\begin{tikzcd}
 & P_L^{-1} \arrow[d, "\pi_L^{-1}"] & & P_N^{-1} \arrow[d, "\pi_N^{-1}"] & \\
0 \arrow[r] & \ker \pi_L \arrow[r] & \ker \pi_M \arrow[r] & \ker \pi_N \arrow[r] & 0
\end{tikzcd}$$

Now, repeat as before, defining $P_M^{-1} = P_L^{-1} \oplus P_N^{-1}$ and define $\pi_M^{-1} : P_L^{-1} \oplus P_N^{-1} \to \ker \pi_M \to P_L^0 \oplus P_N^0$ giving the diagram:

$$\begin{tikzcd}
0 \arrow[r] & P_L^{-1} \arrow[r] \arrow[d, "\pi_L^{-1}"] & P_L^{-1} \oplus P_N^{-1} \arrow[r] \arrow[d, "\pi_M^{-1}"] & P_N^{-1} \arrow[r] \arrow[d, "\pi_N^{-1}"] & 0 \\
0 \arrow[r] & P_L^0 \arrow[r] \arrow[d, "\pi_L"] & P_L^0 \oplus P_N^0 \arrow[r] \arrow[d, "\pi_M"] & P_N^0 \arrow[r] \arrow[d, "\pi_N"] & 0 \\
0 \arrow[r] & L \arrow[r] & M \arrow[r] & N \arrow[r] & 0
\end{tikzcd}$$

By then by the snake lemma, $P_L^{-1} \oplus P_N^{-1} \to \ker \pi_M$ is an epimorphism, and hence we get exactness at $P_L^0 \oplus P_N^0$. Continuing inductively constructs the rest of the diagram.

It is tempting to say from this result that the mapping $\mathbf{A} \to K(\mathbf{A})$ is exact, however $K(\mathbf{A})$ is not abelian and hence this is simply undefined. The above is should instead be interpreted as follows: $P_M^\bullet$ is the mapping cone of a morphism: $\rho^\bullet : P_N[-1]^\bullet \to P_L^\bullet$, giving the distinguished triangle:

$$\begin{tikzcd}
 & P_L^\bullet \arrow[dr] & \\
P_N^\bullet \arrow[ur, "+1"] & & MC(\rho)^\bullet[= P_M^\bullet] \arrow[ll]
\end{tikzcd}$$

This is as close as we can get to preserving exactness.

Corollary 27.3.6: Resolution Of Projective Resolutions

Let $M^\bullet$ be a complex bounded from above in a category $\mathbf{A}$ with enough projectives. Then there exists a complex $P_{M^\bullet}^b b$:

$$\cdots \to P_{M^{-3}}^\bullet \to P_{M^{-2}}^\bullet \to P_{M^{-1}}^\bullet \to\to P_{M^0}^\bullet \to 0$$

where each $P_{M^i}^\bullet$ is a projection resolution of M^i and inducing $M^\bullet$ in cohomology. If $M^\bullet$ is exact, then $P_{M^\bullet}^\bullet$ can be chosen to be exact.

***Proof* :**
Since $M^\bullet$ is a complex, it can be broken down into two short exact sequences:

$$0 \to K^i \to M^i \to I^{i+1} \to 0 \qquad 0 \to I^i \to K^i \to H^i \to 0$$

where K^i is the kernel of $M^i \to M^{i+1}$, I^{i+1} is its image, and H^i the cohomology at M^i. Chose arbitrary projective resolutions for each I^i and H^i for all i. Then by the horseshoe lemma, we get a projective resolution for K^i that fits within the short exact sequence:

$$0 \to P_{I^i}^\bullet \to P_{K^i}^\bullet \to P_{H^i}^\bullet \to 0$$

Applying the horseshoe lemma gain, we get a short exact sequence:

$$0 \to P_{K^i}^\bullet \to P_{M^i}^\bullet \to P_{I^{i+1}}^\bullet \to 0$$

Using these, the $P^\bullet_{M^i}$ can be used to construct a cochain complex, the differentials obtains by composition. If $M^\bullet$ is exact, then $I^i = K^i$, and so the construction gives $P^\bullet_{I^i} = P^\bullet_{K^i}$, so that $P^\bullet_{M^\bullet}$ is also exact.

Lemma 27.3.7: Projective Resolution and Functors

Let $\mathbf{A}$ be an abelian category and let

$$0 \to L^\bullet \to M^\bullet \to P^\bullet \to 0$$

be an exact sequence of complexes in $\mathbf{A}$, where P^i is projective for all i. Let $\mathscr{F} : \mathbf{A} \to \mathbf{B}$ be an additive functor between abelian categories. Then the sequence:

$$0 \to \mathscr{F}(L^b) \to \mathscr{F}(M^\bullet) \to \mathscr{F}(P^\bullet) \to 0$$

is exact

Note that there is nothing about $\mathscr{F}$ being an exact functor! This can be thought of as the generalization of the properties of projective modules which preserve exactness!!

Proof **:**
Since P^i is projective, the sequence

$$0 \to L^i \to M^i \to P^i \to 0$$

splits. Thus

$$0 \to \mathscr{F}(L^i) \to \mathscr{F}(M^i) \to \mathscr{F}(P^i) \to 0$$

is split exact for all i (as the reader should check). Since exactness of a sequence of complexes is given by exactness at each degree, this gives the desired result.

Theorem 27.3.8: Long Exact Sequence From Derived Functor

Let $\mathscr{F} : \mathbf{A} \to \mathbf{B}$ be an additive functor of abelian categories, and assume $\mathbf{A}$ has enough projective. Then every short exact sequence in $\mathbf{A}$

$$0 \to L \to M \to N \to 0$$

[covariantly] functorially induces a long exact sequence in $\mathbf{B}$:

$$\begin{tikzcd}
0 \arrow[r] & L_2\mathscr{F}(L) \arrow[r] & L_2\mathscr{F}(M) \arrow[r] & L_2\mathscr{F}(N) \arrow[dll, "\delta_2"] \\
 & L_1\mathscr{F}(L) \arrow[r] & L_1\mathscr{F}(M) \arrow[r] & L_1\mathscr{F}(N) \arrow[dll, "\delta_1"] \\
 & L_0\mathscr{F}(L) \arrow[r] & L_0\mathscr{F}(M) \arrow[r] & L_0\mathscr{F}(N) \arrow[r] & 0
\end{tikzcd}$$

***Proof* :**
The functorial assignment will be left as an exercise. Given the short exact sequence, we get short exact sequence of projective resolutions, which by lemma 27.3.7 creates a short exact sequence:

$$0 \to \mathscr{F}(P_L^\bullet) \to \mathscr{F}(P_M^\bullet) \to \mathscr{F}(P_N^\bullet) \to 0$$

Since the cohomology of these complexes is precisely the left-derifevd functors of $\mathscr{F}$, the long exact cohomology sequence is given by Snakes lemma, completing the proof.

The change from short exact sequence to long exact sequence can be unified under the notion of triangles, namely this theorem can be seen as the triangle:

$$\begin{tikzcd}
 & \iota(L) \arrow[dr] & \\
\iota(N) \arrow[ur, "+1"] & & \iota(M) \arrow[ll]
\end{tikzcd}$$

inducing the triangle

$$\begin{tikzcd}
 & \mathbf{L}_\bullet(L) \arrow[dr] & \\
\mathbf{L}_\bullet(N) \arrow[ur, "-1", "\delta"'] & & \mathbf{L}_\bullet(M) \arrow[ll]
\end{tikzcd}$$

where the vertices of this triangle are complexes in $C^{\leq 0}(\mathbf{B})$. Naturally, the dual situation happens for right derived functors for categories that have enough injectives, iin which case the above triangle induces a triangle:

$$\begin{tikzcd}
 & \mathbf{R}^\bullet(R) \arrow[dr] & \\
\mathbf{R}^\bullet(N) \arrow[ur, "-1", "\delta"'] & & \mathbf{R}^\bullet(M) \arrow[ll]
\end{tikzcd}$$

where the vertices of this triangle are complexes in $C^{\geq 0}(\mathbf{B})$.

These consideration were for general additive functors. In most cases we saw, there is usually a left-exact or right-exact functor (indeed, the nature of the left-derived and right-derived series almost force us to choose "a side"). If a functor is given to be either left- or right-exact, then it can be directly recovered from their derived functors, as we shall now show.

Proposition 27.3.9: Right-Exactness And Derived Functors

L $\mathbf{A}$ be an abelian category with enough projective and $\mathscr{F} : \mathbf{A} \to \mathbf{B}$ be a right-exact additive functor. Then

1. $\mathbf{L}_i\mathscr{F} = 0$ for $i < 0$
2. $L_0\mathscr{F}$ is naturally isomorphic to $\mathscr{F}$

The dual result holds for the category with enough injectives. Do note that the *right*-exact functor is associated to the *left* derived functors, and vice-versa for left-exact functors which would have the

associated *right* derived functors. This is because the derived functors are "fixing" the exactness on the other side.

***Proof* :**
Since projective resolution $P^\bullet$ of an object M in $\mathbf{A}$ is in $C^{\leq 0}(\mathbf{A})$, $C(\mathscr{F})(P^b b)$ is 0 in positive degree, and hence so is its cohomology. Since $H_i = H^{-i}$, the first claim is immediate.

For the second, an exact sequence:

$$P^{-1} \to P^0 \to M \to 0$$

will map to a exact sequence by $\mathscr{F}$ (due to right-exactness)

$$\mathscr{F}(P^{-1}) - > \mathscr{F}(P^0) - > \mathscr{F}(M) - > 0$$

This induces an isomorphism: $\nu_L : \mathbf{L}_0\mathscr{F}(M) = H^0(C(\mathscr{F})(P^\bullet)) \xrightarrow{\sim} \mathscr{F}(M)$. If $\varphi : M_0 - > M_1$ is a morphism, then there exists a morphism of projective resolutions $\alpha^\bullet : P_0^\bullet - > P_1^b b$:

$$\begin{tikzcd}
\cdots \arrow[r] & P_0^{-1} \arrow[r] \arrow[d, "\alpha^{-1}"] & P_1^0 \arrow[r] \arrow[d, "\alpha^0"] & M_0 \arrow[r] \arrow[d, "\varphi"] & 0 \\
\cdots \arrow[r] & P_1^{-1} \arrow[r] & P_1^0 \arrow[r] & M_1 \arrow[r] & 0
\end{tikzcd}$$

induced by φ where $H^0(\alpha^b b)$ agrees with φ. Applying $\mathscr{F}$ and truncating to $\alpha^\bullet$, we get the commutative diagram:

$$\begin{tikzcd}
\mathscr{F}(P_0^{-1}) \arrow[r] \arrow[d, "\mathscr{F}(\alpha^{-1})"] & \mathscr{F}(P_1^0) \arrow[r] \arrow[d, "\mathscr{F}(\alpha^0)"] & \mathscr{F}(M_0) \arrow[r] \arrow[d, "\mathscr{F}(\varphi)"] & 0 \\
\mathscr{F}(P_1^{-1}) \arrow[r] & \mathscr{F}(P_1^0) \arrow[r] & \mathscr{F}(M_1) \arrow[r] & 0
\end{tikzcd}$$

with exact rows since $\mathscr{F}$ is right-exact. But then the following diagram commutes:

$$\begin{tikzcd}
\mathbf{L}_0\mathscr{F}(M_0) \arrow[r, "\mathbf{L}_0\mathscr{F}(\varphi)"] \arrow[d, "\nu_{M_0}"] & \mathbf{L}_0\mathscr{F}(M_1) \arrow[d, "\nu_{M_1}"] \\
\mathscr{F}(M_0) \arrow[r, "\mathscr{F}(\varphi)"'] & \mathscr{F}(M_1)
\end{tikzcd}$$

completing the proof.

This finally gives the desired recovery of exactness promised way back. Given a short exact sequence

$$0 \to L \to M \to N \to 0$$

then applying a right-exact functor

$$\mathscr{F}(L) \to \mathscr{F}(M)\mathscr{F}(N) \to 0$$

will give an exact sequence, which can be stretched out to form a long exact sequence using the

derived functors:

$$\begin{tikzcd}
0 \arrow[r] & L_2\mathscr{F}(L) \arrow[r] & L_2\mathscr{F}(M) \arrow[r] & L_2\mathscr{F}(N) \arrow[dll, "\delta_2"] \\
& L_1\mathscr{F}(L) \arrow[r] & L_1\mathscr{F}(M) \arrow[r] & L_1\mathscr{F}(N) \arrow[dll, "\delta_1"] \\
& \mathscr{F}(L) \arrow[r] & \mathscr{F}(M) \arrow[r] & \mathscr{F}(N) \arrow[r] & 0
\end{tikzcd}$$

In this way, $\mathbf{L}_1\mathscr{F}$ measures the extent to which $\mathscr{F}$ fails to be left-exact, and each progressive $\mathbf{L}_i\mathscr{F}$ gives further measurement, further information, about this failure. When in a geometrical setting, this failure is often thought of as some form of obstruction (famously, in algebraic topology, this obstruction is interpreted as "'holes", in K-theory).

Similarly, $\mathbf{R}^-\mathscr{F}$ is naturally isomorphic to $\mathscr{F}$ if $\mathscr{F}$ is left-exact and the right-derived functors fit into the diagram:

$$\begin{tikzcd}
0 \arrow[r] & \mathscr{F}(L) \arrow[r] & \mathscr{F}(M) \arrow[r] & \mathscr{F}(N) \arrow[dll, "\delta_1"] \\
& R^1\mathscr{F}(R) \arrow[r] & R^1\mathscr{F}(M) \arrow[r] & R^1\mathscr{F}(N) \arrow[dll, "\delta_0"] \\
& R^2\mathscr{F}(R) \arrow[r] & R^2\mathscr{F}(M) \arrow[r] & R^2\mathscr{F}(N) \arrow[r] & \cdots
\end{tikzcd}$$

Thus, we have successfully captured the cohomological information that a functor may provide!

Acyclic Functors

There still remain some questions to be answered, one that was mentioned was that instead of using a projective resolution for Tor, a flat resolution may be used instead and give the same answer.

Definition 27.3.10: Acyclic Functor

Let $\mathscr{F}$ be an additive functor. An object M of $\mathbf{A}$ is *$\mathscr{F}$-acyclic* (with respect to left-derived functors) if $\mathbf{L}_i\mathscr{F}(M) = 0$ for all $i \neq 0$.

Similarly, an object M is $\mathscr{F}$-acyclic with respect to right-derived functors if $\mathbf{R}^i\mathscr{F}(M) = 0$ for $i \neq 0$. Naturally, projectives are acyclic for every right-exact functor, and injective acyclic for every left-exact functor, however such a functor may admit other $\mathscr{F}$-acyclic objects.

Example 27.3.3: Functor-acyclic Objects

Let R be a commutative ring. Recall that an R-module M is *flat* if $- \otimes_R M$ is exact (equivalently, by symmetry of $\otimes$ that $M \otimes_R -$ is exact). Then flat modules are *acyclic* with respect to tensor products, namely if N is flat:

$$\mathrm{Tor}_R^i(M, N) = 0 \qquad \forall i \neq 0,\ \forall M$$

and symmetrically if M is flat. Thus, flat modules are $(- \otimes_R N)$-acyclic for all modules N.

What shall be built up to now is that $\mathscr{F}$-acyclic resolutions can be used to compute $\mathbf{L}_i\mathscr{F}$.

Theorem 27.3.11: Derived Functors And Acylcic Functors

Let $\mathscr{F} : \mathbf{A} \to \mathbf{B}$ be a right-exact functor of abelian groups where $\mathbf{A}$ has enough projectives. Let $A^\bullet$ be a resolution of M where each A^i is $\mathscr{F}$-acyclic. Then for all i:

$$\mathbf{L}_i\mathscr{F}(M) \cong H^{-i}(\mathrm{Ch}(\mathscr{F})(A^\bullet))$$

Note that if each A^i is projective, the statement follows by definition.

***Proof* :**
The 0th and -1th position will be computed directly, and each successive position comes from induction. Take the exact sequence

$$A^{-1} \to A^0 \to M \to 0$$

which creates another exact sequence in $\mathbf{B}$ after applying the right-exact functor $\mathscr{F}$:

$$\mathscr{F}(A^{-1}) \to \mathscr{F}(A^0) \to \mathscr{F}(M) \to 0$$

Then:

$$H^0(\mathrm{Ch}(\mathscr{F})(A^\bullet)) \cong \mathscr{F}(M) \cong \mathbf{L}_0\mathscr{F}(M)$$

Now, if K is the kernel of $A^0 \to M$, then we have a short exact sequence:

$$0 \to K \to A^0 \to M \to 0$$

Thus applying the left-derived functors we form the long exact sequence:

$$\mathbf{L}_1\mathscr{F}(A^0) \to \mathbf{L}_1\mathscr{F}(M) \to \mathscr{F}(K) \to \mathscr{F}(A^0) \to \mathscr{F}(M) \to 0$$

Since A^0 is acyclic, $L_1\mathscr{F}(A^0) = 0$, hence by exactness

$$\mathbf{L}_1\mathscr{F}(M) \cong \ker(\mathscr{F}(K) \to \mathscr{F}(A^0)) \cong H^{-1}(\mathrm{Ch}(\mathscr{F})(A^\bullet))$$

where the last isomorphism is some manipulation of definition. For the rest, we do induction. Assume now that for all object of $\mathbf{A}$ and all indices $< i$. Truncating and shifting $\mathbf{A}^\bullet$ so that A^{-1} is in the 0 degree:

$$\cdots \to A^{-3} \to A^{-2} \to A^{-1} \to 0 \to \cdots$$

we get an $\mathscr{F}$-acyclic resolution of K, and so by the induction hypothesis $\mathbf{L}_{i-1}\mathscr{F}(K)$ can be computed by applying $\mathrm{Ch}(\mathscr{F})$ to the complex and taking cohomology, namely:

$$\mathbf{L}_{i-1}\mathscr{F}(K) \cong H^{-i}(A^\bullet)$$

On the other hand, the other terms in the long exact sequence obtained using the left derived functors gives:

$$\cdots \to \mathbf{L}_i\mathscr{F}(A^0) \to L_i\mathscr{F}(M) \to \mathbf{L}_{i-1}\mathscr{F}(K) \to \mathbf{L}_{i-1}\mathscr{F}(A^0) \to \cdots$$

since A^0 is $\mathscr{F}$-acyclic and $i > 1$, we get:

$$\mathbf{L}_i\mathscr{F}(M) \cong \mathbf{L}_{i-1}\mathscr{F}(K)$$

giving the desired result.

Thus, for the Tor functor, we may use a flat resolution to compute it the left-derived functors. Since $\mathscr{F}$-acyclic objects suffice in the computation, one can imagine that there is not need for $\mathbf{A}$ to have enough projectives to compute it. This is indeed the case, given that $\mathbf{A}$ has "enough $\mathscr{F}$-acyclic's". (this is an exercise in Aluffi, 8.14, with bifunctors).

27.4 Double Complexes

This final section shall be the cornerstone in being able to compute many complexes. It was mentioned in remark ref:HERE that $\mathrm{Ext}^i_R(M, N)$ may be computed by using either a projective resolution of M or an injective resolution on N, both give the same answer. More generally, we may compute the resolution of either M or N and get the same value. This section will now prove this result. This will make many computations much simpler, for usually one object in is readily understandable while the other is being analyzed.

To motivate this, switching from projective to $\mathscr{F}$-acyclic resolutions can be viewed in the following important alternative way. Consider the following diagram where the top row and right column are cochain:

$$\begin{tikzcd}
\cdots \arrow[r] & \mathscr{F}(A^{-2}) \arrow[r] & \mathscr{F}(A^{-1}) \arrow[r] & \mathscr{F}(A^{-0}) \arrow[r] & \mathscr{F}(M) \\
\cdots \arrow[r, dotted] & \cdot \arrow[r, dotted] \arrow[u, dotted] & \cdot \arrow[r, dotted] \arrow[u, dotted] & \cdot \arrow[r, dotted] \arrow[u, dotted] & \mathscr{F}(P_M^0) \arrow[u] \\
\cdots \arrow[r, dotted] & \cdot \arrow[r, dotted] \arrow[u, dotted] & \cdot \arrow[r, dotted] \arrow[u, dotted] & \cdot \arrow[r, dotted] \arrow[u, dotted] & \mathscr{F}(P_M^{-1}) \arrow[u] \\
\cdots \arrow[r, dotted] & \cdot \arrow[r, dotted] \arrow[u, dotted] & \cdot \arrow[r, dotted] \arrow[u, dotted] & \cdot \arrow[r, dotted] \arrow[u, dotted] & \mathscr{F}(P_M^{-2}) \arrow[u] \\
 & \vdots \arrow[u, dotted] & \vdots \arrow[u, dotted] & \vdots \arrow[u, dotted] & \vdots \arrow[u]
\end{tikzcd}$$

in some way, there was a reason to go from the top chain to the right chain, with the intermediate values "interpolating" the result. The natural candidate for this would be a complex of complex. Needing to bound it in one direction, say we are looking at $\mathrm{Ch}^{\leq 0}(\mathrm{Ch}^{\leq 0}(\mathbf{A}))$. Such an object would be of the form:

$$\cdots \to M^{-3,\bullet} \xrightarrow{d_h^{-3,\bullet}} M^{-2,\bullet} \xrightarrow{d_h^{-2,\bullet}} M^{-1,\bullet} \xrightarrow{d_h^{-1,\bullet}} M^{0,\bullet} \to 0 \to \cdots$$

The h in the subscript is to represent the fact we are going *horizontally*, and the respective morphism of each $d_h^{-i,\bullet}$ will be going vertically and are denoted $d_v^{i,j}$, or simply d_v to simply show the function's

direction:

$$\begin{tikzcd}
\cdots \arrow[r] & M^{-3,0} \arrow[r] & M^{-2,0} \arrow[r] & M^{-1,0} \arrow[r] & M^{0,0} \\
\cdots \arrow[r] & M^{-3,-1} \arrow[r] \arrow[u] & M^{-2,-1} \arrow[r] \arrow[u] & M^{-1,-1} \arrow[r] \arrow[u] & M^{0,-1} \arrow[u] \\
\cdots \arrow[r] & M^{-3,-2} \arrow[r] \arrow[u] & M^{-2,-2} \arrow[r] \arrow[u] & M^{-1,-2} \arrow[r] \arrow[u] & M^{0,-2} \arrow[u] \\
 & \vdots \arrow[u] & \vdots \arrow[u] & \vdots \arrow[u] & \vdots \arrow[u]
\end{tikzcd}$$

Given a complex of complex, it would be nice to reduce this to being a complex, a more familiar setting. One strategy the reader may come up with is to take:

$$\bigoplus_{i+j=n} M^{i,j}$$

with the natural morphisms between them. However these morphisms *almost* complexes: instead of commutativity for each square, they are *anti-commutative*. To fix this, create a new type of object that will satisfy this condition:

Definition 27.4.1: Double Complex

Let $A = \{A_{p,q}\}$ be a family of objects from $\mathbf{A}$. Then A along with maps:

$$d^h : C_{p,q} \to C_{p-1,q} \qquad \delta^v : C_{p,q} \to C_{p,q-1}$$

with the condition of $d^h \circ d^h = 0$, $\delta^v \circ \delta^v = 0$ and $d^h \circ \delta^v + \delta^v \circ d^h = 0$ is called a *double complex* or *bicomplex*

We may visualize a double complex like so:

$$\begin{tikzcd}
 & \cdots & \cdots & \cdots & \\
\cdots \arrow[r] & C^{p+1,q-1} \arrow[r] \arrow[u] & C^{p,q-1} \arrow[r] \arrow[u] & C^{p,q-1} \arrow[r] \arrow[u] & \cdots \\
\cdots \arrow[r] & C^{p+1,q} \arrow[r] \arrow[u] & C^{p,q} \arrow[r] \arrow[u] & C^{p,q} \arrow[r] \arrow[u] & \cdots \\
\cdots \arrow[r] & C^{p+1,q+1} \arrow[r] \arrow[u] & C^{p,q+1} \arrow[r] \arrow[u] & C^{p,q+1} \arrow[r] \arrow[u] & \cdots \\
 & \cdots \arrow[u] & \cdots \arrow[u] & \cdots \arrow[u] &
\end{tikzcd} \tag{27.6}$$

Note this is not a commutative diagram, but an *anti-commutative* diagra. We can geet a complex of complex by introducing a sign:

$$d^{p,q} = (-1)^p \delta^v_{p,q} : C^{p,q} \to C^{p,q+1}$$

As mentioned, we can form a total complex from this:

Proposition 27.4.2: Total Complex

Let $A = \{A^{p,q}\}$ be a double complex. Then define the total complex $\mathrm{Tot}^{\Pi}(A)$ and $\mathrm{Tot}^{\oplus}(A)$ by:

$$\mathrm{Tot}^{\Pi}(A)_n = \prod_{p+q=n} A_n \qquad \mathrm{Tot}^{\oplus}(A)_n = \oplus_{p+q=n} A_n$$

with $d^{tot} = d^v + d^h$ so that $d \circ d = 0$

***Proof* :**
Computing:

$$d_{tot} \circ d_{tot} = (d_h + \delta_v) \circ (d_h + \delta_v) = d_h \circ d_h + d_h \circ \delta_v + \delta_v \circ d_h + \delta_v \circ \delta_v = 0$$

Note that the total complex does not exist in all categories, namely it may not exist in categories which are not closed under infinite products and coproducts (ex. in the category of finite abelian groups). A category which is closed for infinite products is called *complete* and *cocomplete* if it is closed for infinite sums. We can visualize a total complex by taking the diagram in equation (27.6) and moving it like so:

$$\begin{tikzcd}
 & & & M^{0,0} \\
 & & M^{-1,0} \arrow[ur, "d_h"] & M^{0,-1} \arrow[u, "\delta_v"'] \\
 & M^{-2,0} \arrow[ur, "d_h"] & M^{-1,-1} \arrow[u, "\delta_v"] \arrow[ur, "d_h"] & M^{0,-2} \arrow[u, "\delta_v"'] \\
M^{-3,0} \arrow[ur, "d_h"] & M^{-2,-1} \arrow[u, "\delta_v"] \arrow[ur, "d_h"] & M^{-1,-2} \arrow[u, "\delta_v"] \arrow[ur, "d_h"] & M^{0,-2} \arrow[u, "\delta_v"'] \\
\vdots \arrow[u, "\delta_v"] & \vdots \arrow[u, "\delta_v"] & \vdots \arrow[u, "\delta_v"] & \vdots \arrow[u, "\delta_v"']
\end{tikzcd}$$

It is then easy to visualize how we get a total complex form this.

Example 27.4.1: Total Complexs

1. Consider a double complex:

$$\cdots \to 0 \to L^\bullet \xrightarrow{\alpha^\bullet} M^\bullet \to 0 \to \cdots$$

which we may represent as:

$$
\begin{tikzcd}
 & \vdots & \vdots & \vdots & \vdots & \\
\cdots \arrow[r] & 0 \arrow[r] \arrow[u] & L^{i+1} \arrow[r, "\alpha^{i+1}"] \arrow[u, "d_L^{i+1}"] & M^{i+1} \arrow[r] \arrow[u, "d_M^{i+1}"] & 0 \arrow[r] \arrow[u] & \cdots \\
\cdots \arrow[r] & 0 \arrow[r] \arrow[u] & L^{i} \arrow[r, "\alpha^{i}"] \arrow[u, "d_L^{i}"] & M^{i} \arrow[r] \arrow[u, "d_M^{i}"] & 0 \arrow[r] \arrow[u] & \cdots \\
\cdots \arrow[r] & 0 \arrow[r] \arrow[u] & L^{i-1} \arrow[r, "\alpha^{i-1}"] \arrow[u, "d_L^{i-1}"] & M^{i-1} \arrow[r] \arrow[u, "d_M^{i-1}"] & 0 \arrow[r] \arrow[u] & \cdots \\
 & \vdots \arrow[u] & \vdots \arrow[u] & \vdots \arrow[u] & \vdots \arrow[u] &
\end{tikzcd}
$$

we shall get the following:

$$
\begin{tikzcd}
 & \vdots & \vdots & \vdots & \vdots & & & \vdots & & \vdots \\
\cdots \arrow[ur] & 0 \arrow[u] & L^{i+2} \arrow[u] \arrow[ur, "\alpha^{i+2}"] & M^{i+1} \arrow[u, "d_L^{i+1}"] & 0 \arrow[u, "d_M^{i+1}"] & \cdots & & L^{i+2} \arrow[u] \arrow[urr] & \oplus & M^{i+1} \arrow[u] \\
\cdots \arrow[ur] & 0 \arrow[u] & L^{i+1} \arrow[u] \arrow[ur, "\alpha^{i+1}"] & M^{i} \arrow[u, "d_L^{i}"] & 0 \arrow[u, "d_M^{i}"] & \cdots & \Rightarrow & L^{i+1} \arrow[u] \arrow[urr] & \oplus & M^{i} \arrow[u] \\
\cdots \arrow[ur] & 0 \arrow[u] & L^{i} \arrow[u] \arrow[ur, "\alpha^{i}"] & M^{i-1} \arrow[u, "d_L^{i-1}"] & 0 \arrow[u, "d_M^{i-1}"] & \cdots & & L^{i} \arrow[u] \arrow[urr] & \oplus & M^{i-1} \arrow[u] \\
 & \vdots \arrow[ur] & \vdots \arrow[ur] & \vdots \arrow[ur] & \vdots \arrow[ur] & & & \vdots \arrow[ur] & & \vdots
\end{tikzcd}
$$

which is exactly the definition of a mapping cone!

2. The operators $\mathrm{Hom}_A(-,-)$ and $-\otimes-$ both give rise to double complexes. Focusing on the $\mathrm{Hom}_A(-,-)$ example, consider a chain in $L^\bullet \in \mathrm{Ch}^{\leq 0}(\mathbf{A})$ and $M^\bullet \in \mathrm{Ch}^{\geq 0}(\mathbf{A})$. These can be used to form a double complex, or isomorphically a complex in $\mathrm{Ch}^{\leq 0}(\mathrm{Ch}^{\leq 0}(\mathbf{A}))$:

$$\cdots \to \mathrm{Hom}_A(L^\bullet, M^0) \to \mathrm{Hom}_A(L^\bullet, M^1) \to \mathrm{Hom}_A(L^\bullet, M^2) \to \cdots$$

which creates the following upper-quadrant grid:

$$\begin{tikzcd}
\vdots & \vdots & \vdots & \\
\mathrm{Hom}_A(L^{-2}, M^0) \arrow[u] \arrow[r] & \mathrm{Hom}_A(L^{-2}, M^1) \arrow[u] \arrow[r] & \mathrm{Hom}_A(L^{-2}, M^2) \arrow[u] \arrow[r] & \cdots \\
\mathrm{Hom}_A(L^{-1}, M^0) \arrow[u] \arrow[r] & \mathrm{Hom}_A(L^{-1}, M^1) \arrow[u] \arrow[r] & \mathrm{Hom}_A(L^{-1}, M^2) \arrow[u] \arrow[r] & \cdots \\
\mathrm{Hom}_A(L^{0}, M^0) \arrow[u] \arrow[r] & \mathrm{Hom}_A(L^{0}, M^1) \arrow[u] \arrow[r] & \mathrm{Hom}_A(L^{0}, M^2) \arrow[u] \arrow[r] & \cdots
\end{tikzcd}$$

Flipping the above diagram across the main diagonal, we would get the double complex:

$$\cdots \to \mathrm{Hom}_A(L^0, M^\bullet) \to \mathrm{Hom}_A(L^{-1}, M^\bullet) \to \mathrm{Hom}_A(L^{-2}, M^\bullet) \to \cdots$$

These two double complexes are isomorphic (as the reader should check). Due to this, by slight abuse of notation the total complex of both can be denoted $TC(\mathrm{Hom}_A(L, M))^\bullet$, where the degree i term is:

$$\mathrm{Hom}_A(L^{-i}, M^0) \oplus \mathrm{Hom}_A(L^{-i+1}, M^1) \oplus \cdots \oplus 0\,\mathrm{Hom}_A(L^{-1}, M^{i-1}) \oplus 0\,\mathrm{Hom}_A(L^0, M^i)$$

The reader should verify that:

(a) $\ker d^i_{tot}$ is equal to morphisms of cochains $L^\bullet \to M[i]^\bullet$

(b) The image $\mathrm{im}\, d^{i-1}_{tit}$ consists of morphisms that are homotopy equivalent to 0

(c) Thus, the cohomology of $TC(\mathrm{Hom}_A(M, N))^\bullet$ "paramaterizes" the morphisms in the homotopic category

To ground this, consider $TC(\mathrm{Hom}_A(M, N))^0 = \mathrm{Hom}_A(L^0, M^0)$. Then $d_{tot}(\alpha) = 0$ means that both $\alpha \circ d^{-1}_{L^\bullet} \in \mathrm{Hom}_A(L^{-1}, M^0)$ and $d^0_{M^\bullet} \in \mathrm{Hom}_A(L^0, M^1)$ both vanish. The reader should draw out this diagram, and notice that this will make α have the condition to be a morphism of cochain complexes $L^\bullet \to M^\bullet$!

$$H^0(TC(\mathrm{Hom}_A(M, N))^\bullet) = \mathrm{Hom}_{\mathrm{Ch}(\mathbf{A})}(L^\bullet, M^\bullet) \overset{!}{=} \mathrm{Hom}_{K(\mathbf{A})}(L^\bullet, M^\bullet)$$

where the $\overset{!}{=}$ equality comes from the fact that there is no nontrivial homotopies in this case! For $H^i(TC(\mathrm{Hom}_A(M, N))^\bullet)$, with some care for signs, it can be identified with $\mathrm{Hom}_{K(\mathbf{A})}(L^\bullet, M[i]^\bullet)$

The cohomology of double complexes requires some book-keeping through the use of *spectral sequences*, which shall be explored in greater depth in chapter 29. For the current goal, it is enough to consider when a double-complex is exact:

Proposition 27.4.3: Exactness Of Complex Of Complex

Let $\mathbf{A}$ be an abelian category and let:

$$\cdots \to M^{-3,\bullet} \to M^{-2,\bullet} \to M^{-1,\bullet} \to M^{0,\bullet} - \cdots$$

be a complex in $\mathrm{Ch}^{\leq 0}(\mathrm{Ch}^{\leq 0}(\mathbf{A}))$. Then the corresponding total complex $^\bullet$ is exact

- The above chain is exact
- each complex $M^{i,\bullet}$ is exact

The same naturally holds for $\mathrm{Ch}^{\leq 0}(\mathrm{Ch}^{\leq 0}(\mathbf{A}))$. Note further that only one of the complexes need to be bounded, the boundedness being required for the proceeding inductive argument.

Proof **:**

1. Flip the double complex corresponding to the above cochain, which the reader should check gives an exact complex $T^\bullet$
2. Aluffi p.671 (it's book keeping on induction)

Example 27.4.2: Exactness Of Double Complex

Let $P^\bullet \in C^{\leq 0}(\mathbf{A})$ be a complex and $L^\bullet \in C^{\geq 0}(\mathbf{A})$ an exact complex. Since each P^i is projective, $\mathrm{Hom}_{\mathbf{A}}(P^i, L^\bullet)$ for each i. Hence, the complex:

$$0 \to \mathrm{Hom}_{\mathbf{A}}(P^0, L^\bullet) \to \mathrm{Hom}_{\mathbf{A}}(P^{-1}, L^\bullet) \to \mathrm{Hom}_{\mathbf{A}}(P^{-2}, L^\bullet) \to \cdots$$

is exact. Using example 27.4.1, we see that:

$$\mathrm{Hom}_{K(\mathbf{A})}(P^\bullet, L[i]^\bullet) = 0 \qquad \forall i \in \mathbb{Z}$$

This implies that every cochain morphism from a bounded-above projective complex to a bounded-bellow exact complex is homotopic to 0. This just proved corollary 27.2.4 (with an extra boundedness condition)! In some textbooks, due to the speed at which some results can be proved using double-complexes, they are introduced earlier. However, the more "down-to-earth" introduction has is favored in this chapter since it demonstrated key homological concepts.

Next, we look at resolutions of complexes.. As we saw in definition 27.2.10, a (projective) resolution is a quasi-isomorphism $M^\bullet \to \iota(A)$ for some complex $M^\bullet \in \mathrm{Obj}(\mathrm{Ch}(\mathbf{A}))$, and more generally a projective resolution for a general complex is a quasi-isomorphism $P^\bullet \to C^\bullet$ for some projective complex $P^\bullet$. Generalizing the definition of resolution of the first kind, $\iota(A)$, and say that resolution of a complex $N^\bullet \in \mathrm{Ch}^\bullet(A) = A'$ is a complex of complex $M^{\bullet,\bullet} \in \mathrm{Ch}^{\leq 0}(A')$ and a quasi-isomorphism

$$\begin{tikzcd}
\cdots \arrow[r] & M^{-2,\bullet} \arrow[r, "d_h"] \arrow[d] & M^{-2,\bullet} \arrow[r, "d_h"] \arrow[d] & M^{-2,\bullet} \arrow[r] \arrow[d] & 0 \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & 0 \arrow[r] & 0 \arrow[r] & N^\bullet \arrow[r] & 0 \arrow[r] & \cdots
\end{tikzcd} \tag{27.7}$$

where the diagram commutes. We shall show that this definition of a resolution induces a resolution as given in theorem 27.2.15. To see this, notice that in equation (27.7), $\alpha^\bullet \circ d_h = 0$, hence we can collapse the diagram like so:

$$\cdots \to M^{-2,\bullet} \xrightarrow{-d_h} M^{-1,\bullet} \xrightarrow{d_h} M^{0,\bullet} \xrightarrow{\alpha^\bullet} N^\bullet \to 0 \cdots$$

Call this complex $M_N^{\bullet,\bullet}$. Two important diagrams can be derived here. The first is by taking the total complex of $M^{\bullet,\bullet}$ and taking a morphism from this to $N^\bullet$:

$$\begin{tikzcd}
 & & & 0 \arrow[r] & 0 \\
 & & & M^{0,0} \arrow[r] \arrow[u] & N^0 \arrow[u] \\
 & & M^{-1,0} \arrow[ur] \arrow[r, phantom, "\oplus"] & M^{0,-1} \arrow[r] \arrow[u] & N^{-1} \arrow[u] \\
M^{-2,0} \arrow[ur] \arrow[r, phantom, "\oplus"] & M^{-1,-1} \arrow[ur] \arrow[u] \arrow[r, phantom, "\oplus"] & M^{0,-2} \arrow[ur] \arrow[u] \arrow[r] & N^{-2} \arrow[u] & \\
 & & & &
\end{tikzcd}$$

This gives a morphism:

$$TC(\alpha^\bullet) : TC(M^{\bullet,\bullet}) \to N^\bullet$$

where the morphism is defined since $TC(-)$ is a functorial. Another diagram we can create is by considering the total complex of $M_N^{\bullet,\bullet}$:

$$\begin{tikzcd}
 & & & N^0 \\
 & & M^{0,0} \arrow[ur] \arrow[r, phantom, "\oplus"] & N^{-1} \arrow[u] \\
 & M^{-1,0} \arrow[ur] \arrow[r, phantom, "\oplus"] & M^{0,-1} \arrow[ur] \arrow[u] \arrow[r, phantom, "\oplus"] & N^{-2} \arrow[u] \\
M^{-2,0} \arrow[ur] \arrow[r, phantom, "\oplus"] & M^{-1,-1} \arrow[ur] \arrow[u] \arrow[r, phantom, "\oplus"] & M^{0,-2} \arrow[ur] \arrow[u] \arrow[r, phantom, "\oplus"] & N^{-3} \arrow[u] \\
 & & & &
\end{tikzcd}$$

These two are related in the following way:

Proposition 27.4.4: Total Complex Of Resolution

Let $M^{\bullet,\bullet}$ be a resolution of $N^\bullet$ and let $M_N^{\bullet,\bullet}$ be defined as above Then:

$$TC(M_N)^\bullet = MC(TC(\alpha))^\bullet$$

***Proof* :**
exercise

Using this, we can see how the resolution in $\mathrm{Ch}^{\leq 0}(A')$ leads to a resolution as defined in theorem 27.2.15:

Theorem 27.4.5: Total Complexes And Resolution Of Complexes

Let $\mathbf{A}$ be an abelian category, $\mathbf{A}' = \mathrm{Ch}^{\leq 0}(\mathbf{A})$, and $N^\bullet \in A'$. Let $M^{\bullet,\bullet} \in \mathrm{Ch}^{\leq 0}(\mathbf{A}')$ with total complex $T^\bullet$.

1. Assume that $M^{\bullet,\bullet}$ is a resolution of $N^\bullet$ in $\mathrm{Ch}^{\leq 0}(\mathbf{A}')$. Then $T^\bullet$ is quasi-isomorphic to $N^\bullet$ in $\mathbf{A}'$.
2. Assuming the cohomology of $M^{i,\bullet}$ is concentrated at 0. Then $T^\bullet$ is quasi-isomorphic to the complex of cohomology objects induced by the complex $M^{\bullet,\bullet}$:

$$\cdots \to H^0(M^{-3,\bullet}) \to H^0(M^{-2,\bullet}) \to H^0(M^{-1,\bullet}) \to H^0(M^{0,\bullet}) \to 0 \cdots$$

Notice that this theorem is a direct extensions of proposition 27.4.3. This is also a special case of results we shall see in chapter 29

***Proof* :**
Notice that (2) follows from (1) by flipping the corresponding double-complex across the diagonal, hence we prove the first. For the first, if $M_N^{\bullet,\bullet}$ is quasi-isomorphic, then $TC(M_N)^\bullet$ is exact by proposition 27.4.3, which by proposition 27.4.4 is the mapping cone of the morphism $T^\bullet \to N^\bullet$, which by proposition 27.1.14 shows that $\alpha^\bullet : T^\bullet \to N^\bullet$ is a quasi-isomorphism, as we sought to show.

Using this theorem, a simple alternative to resolving complexes (theorem 27.2.15). By the horseshoe lemma, every bounded-above complex $N^\bullet$ in an abelian category with enough projectives may be viewed as a complex induced by H^0 from a complex:

$$\cdots \to P^{-3,\bullet} \to P^{-2,\bullet} \to P^{-1,\bullet} \to P^{0,\bullet} \to 0$$

where $P^{i,\bullet}$ are projective resolutions of N^i. Then by theorem 27.4.5(2) the total complex of $P^{\bullet,\bullet}$ is a projective resolution of $N^\bullet$. This then can give theorem 27.2.15. It is still worth presenting theorem 27.2.15 to firmly grasp the fundamentals of homological algebra.

27.4.1 Applications to Computations of Ext and Tor

Let us now show that we may compute $\mathrm{Ext}(M, N)$ or $\mathrm{Tor}(M, N)$ using a resolution of either term. Going back to the $\mathscr{F}$-acyclic resolution of an object M:

$$\cdots \to A^{-2} \to A^{-1} \to A^0 \to M \to 0 \to \cdots$$

We may create a complex of projective resolution by corollary 27.3.6:

$$\cdots \to P^{-2,\bullet} \to P^{-1,\bullet} \to P^{0,\bullet} \to P_M^\bullet \to 0 \to \cdots$$

where $P^{i,\bullet}$ is a projective resolution of A^{-i}. Since the resolution by projective resolution's is exact, the rows of the double-complex corresponding to $P^{\bullet,\bullet}$ are projective resolutions of P^i_M. Now, applying the functor $\mathscr{F}$, we get the diagram:

$$\begin{tikzcd}
\cdots \arrow[r] & \mathscr{F}(A^{-2}) \arrow[r] & \mathscr{F}(A^{-1}) \arrow[r] & \mathscr{F}(A^{0}) \arrow[r] & \mathscr{F}(M) \\
\cdots \arrow[r] & \mathscr{F}(P_M^{-2,0}) \arrow[r] \arrow[u] & \mathscr{F}(P_M^{-1,0}) \arrow[r] \arrow[u] & \mathscr{F}(P_M^{0,0}) \arrow[r] \arrow[u] & \mathscr{F}(P_M^{0}) \arrow[u] \\
\cdots \arrow[r] & \mathscr{F}(P_M^{-2,-1}) \arrow[r] \arrow[u] & \mathscr{F}(P_M^{-1,-1}) \arrow[r] \arrow[u] & \mathscr{F}(P_M^{0,-1}) \arrow[r] \arrow[u] & \mathscr{F}(P_M^{-1}) \arrow[u] \\
\cdots \arrow[r] & \mathscr{F}(P_M^{-2,-2}) \arrow[r] \arrow[u] & \mathscr{F}(P_M^{-1,-2}) \arrow[r] \arrow[u] & \mathscr{F}(P_M^{0,-2}) \arrow[r] \arrow[u] & \mathscr{F}(P_M^{-2}) \arrow[u] \\
 & \vdots \arrow[u] & \vdots \arrow[u] & \vdots \arrow[u] & \vdots \arrow[u]
\end{tikzcd}$$

Since each row and column of this diagram (excluding the top row and right column) is exact, by theorem 27.4.5 the total complex of:

$$\cdots \mathscr{F}(P^{-2,\bullet}) \to \mathscr{F}(P^{-1,\bullet}) \to \mathscr{F}(P^{0,\bullet}) \to 0 \to \cdots$$

is quasi-isomorphic to both $\mathrm{Ch}(\mathscr{F})(A^\bullet)$ and $\mathrm{Ch}(\mathscr{F})(P^\bullet)$, that is:

$$TC(\mathscr{F}(P))^\bullet \sim \mathrm{Ch}(\mathscr{F})(A^\bullet) \sim \mathrm{Ch}(\mathscr{F})(P^\bullet)$$

Thus, by definition of quasi-isomorphism

$$\mathbf{L}_i\mathscr{F}(M) \cong H^i(\mathrm{Ch}(\mathscr{F})(A^\bullet)) \cong H^i(\mathrm{Ch}(\mathscr{F})(P^\bullet))$$

Applying these results to Ext and Tor will finally give us the following results:

Theorem 27.4.6: Computation Of Tor and Ext

Let M, N be R-modules for a commutative ring R, and let $P^\bullet_M$ and $P^\bullet_N$ be projections resolutions of M and N respectively, and $Q^\bullet_N$ an injective resolution of N. Then:

$$\mathrm{Tor}_i^R(M,N) \cong H^i(P^\bullet_M \otimes_R N) \cong H^i(M \otimes_R P^\bullet_N)$$

and

$$\mathrm{Ext}_i^R(M,N) \cong H^i(P^\bullet_M \otimes_R N) \cong H^i(M \otimes_R Q^\bullet_N)$$

***Proof* :**
We shall show the result for Tor, the result following very similarly for Ext. Take the complex:

$$\cdots \to P_M^{-2} \otimes_R P^\bullet_N \to P_M^{-1} \otimes_R P^\bullet_N \to P_M^{0} \otimes_R P^\bullet_N \to 0 \cdots$$

Since each P_N^j is projective, the cohomology of the complex is concentrated at 0 and equals $M \otimes_R P_N^\bullet$. By theorem 27.4.5 we have that :

$$TC(P_M^\bullet \otimes_R P_N^\bullet)^\bullet \qquad \text{is quasi-isomorphic to} \qquad M \otimes_R P_N^\bullet$$

Similarly, since each P_M^i is projective, the complex $P_M^i \otimes_R P_N^\bullet$ is a resolution of $P_M^i \otimes N$, i.e. the cohomology of each $P_M^i \otimes_R P_N^\bullet$ is concentrated at degree 0 and equals $P_M^i \otimes_R N$. By theorem 27.4.5 we have that :

$$TC(P_M^\bullet \otimes_R P_N^\bullet)^\bullet \qquad \text{is quasi-isomorphic to} \qquad P_M^\bullet \otimes_R N$$

but then, the cohomologies are identical, as we sought to show.

27.5 *Formal Derived Category and Triangulated Category

In this bonus section, we shall introduce formally give the definition of a derived category and introduce the notion of triangulated categories.

27.5.1 Derived Categories, Formally

As defined in this chapter, derived categories required enough projectives/injectives. However, it was mentioned that derived categories may be defined without this restriction, allowing for a category that that satisfies the universal property for abelian categories more generally. To define $\mathrm{D}(\mathbf{A})$ more generally, more theoretical build-up is necessary and shall not be carried out in this book, however presenting the idea of how to construct a general derived category will give the reader a chance to see some interesting applications of the concept if *localization* of a category. This section is heavily inspired by *Aluffi Chapter 0* which presented these idea's very concisely.

We shall have the objects still be $\mathrm{C}(\mathbf{A})$, however we need the quasi-isomorphisms to be invertible. Recall lemma 27.2.6 where it was shown that quasi-isomorphisms of projective resolutions have the property of being acting like non-zero divisors. We shall take full advantage of this idea, and essentially "formally invert" quasi-isomorphisms in the same vein as the construction of localization's for rings and modules. In fact, from an abstract perspective, the study of homological algebra can be thought of as the study of the localization of the category of (co)chain complexes (localized at the class of quasi-isomorphisms). When there is enough projectives, the localization of $\mathrm{C}(\mathbf{A})$ becomes $\mathrm{K}(\mathbf{a})$.

For the more general case, consider the following *roof diagram*:

$$\begin{tikzcd} & Z \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ A & & B \end{tikzcd}$$

Then we may formally add a morphism $\beta \circ \alpha^{-1} : A \to B$:

$$\begin{tikzcd} & Z \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ A \arrow[rr, dashed] & & B \end{tikzcd}$$

This can be thought of as the fraction. There are many technicalities being swept under the rug, like how do we insure these are well-defined morphisms (ex. how would we compose roofs?), while more fundamental problems like the amount of morphisms shouldn't exceed a *set*-amount (note that **A** and **B** are classes). Furthermore, we may ask if this is possible on any possible category, or should we restrict to abelian categories? Can any morphism be invertible, or do we need to restrict to a special subset (similar to a multiplicative set)? Giving this technical difficulties, we may imagine that from this the diagram

$$\begin{tikzcd} & Z \arrow[dl, "\alpha"'] \arrow[dr, "\mathrm{id}_Z"] & \\ A \arrow[rr, dashed] & & Z \end{tikzcd}$$

would formally invert a morphism. For $\mathrm{D}(\mathbf{A})$, we would localize at the class of quasi-isomorphism of $\mathrm{K}(\mathbf{A})$. The category $\mathrm{D}(\mathbf{A})$ shall be the homotopic category of $\mathrm{C}(\mathbf{A})$ where a morphism $L^\bullet \to M^\bullet$ can be represented as a roof:

$$\begin{tikzcd} & N^\bullet \arrow[dl, "\text{quasi-iso}"'] \arrow[dr] & \\ L^\bullet & & M^\bullet \end{tikzcd}$$

(A word on how many abelian categories may induce the same derived categories. However, there is a special "triangulatedd structure", or t-structure which allows you to recover A)

27.5.2 Triangulated Categories

Homotopic categories are certainly almost never not abelian, for the kernels and cokernels of homotopic maps need not match (hence not being well-defined in the equivalence class). However, a derived functors $\mathrm{D}(\mathscr{F})$ always has a long exact sequence associated to it, which shows some weaker property these categories possess. These considerations are explored in this chapter. Take a short exact sequence of complexes:

$$0 \to L^\bullet \to M^\bullet \to N^\bullet \to 0$$

Then snakes lemma says that after taking homology, there is an interesting connection between $N^\bullet$ and $L^\bullet$. When working with the complex, there is no interesting connection between $N^\bullet$ and $L^\bullet$. We may connect these two complexes via the zero morphism: $N^\bullet \xrightarrow{0} L^\bullet$. From this, the short exact sequence of complexes may be interpreted as these three exact sequences:

$$\begin{aligned} &N^\bullet \xrightarrow{0} L^\bullet \to M^\bullet \\ &L^\bullet \to M^\bullet \to N^\bullet \\ &M^\bullet \to N^\bullet \xrightarrow{0} L^\bullet \end{aligned}$$

These three exact sequences are essentially can be combined if we fold any of them into the following triangle:

Definition 27.5.1: Exact Triangle

Let $L^\bullet, M^\bullet, N^\bullet \in \mathrm{Ch}(\mathbf{A})$ be three complexes. Then if there exists morphisms st the following diagram is exact (at each vertex):

$$\begin{tikzcd} & L^\bullet \arrow[dr] & \\ N^\bullet \arrow[ur] & & M^\bullet \arrow[ll] \end{tikzcd}$$

Then the diagram is called an *exact triangle* of $\mathrm{Ch}(\mathbf{A})$. This diagram is also sometimes denoted

$$L^\bullet \to M^\bullet \to N^\bullet \to$$

if it is easier to write it out flat.

Hence, the particular exact triangle given by folding any of the above exact sequences is:

$$\begin{tikzcd} & L^\bullet \arrow[dr] & \\ N^\bullet \arrow[ur, "+1"] & & M^\bullet \arrow[ll] \end{tikzcd}$$

where the $+1$ morphism represents the zero map. The $+1$ notation is to hint that this map will be a sort of shift from $N^\bullet$ to $L^\bullet[1]$. This can be stretched out into a single long sequence:

$$\cdots \to N^{i-1} \xrightarrow{0} L^i \to M^i \to N^i \xrightarrow{0} L^{i+1} \to \cdots$$

More explicitly, the triangle can be "unwrapped" to create the following nice diagram:

here, v good visual intuition

Applying the cohomology functor to the above long exact sequence and replacing the zero maps with the connecting morphism, we get:

$$\cdots \to H^{i-1}(N^\bullet) \xrightarrow{\delta^{i-1}} H^i(L^\bullet) \xrightarrow{0} H^i(M^\bullet) \xrightarrow{0} H^i(N^\bullet) \xrightarrow{\delta^i} H^{i+1}(L^\bullet) \to \cdots$$

Note that the maps δ^i did not come from the cohomology functor. In a sense, this is because $\mathrm{Ch}(\mathbf{A})$ is the wrong category to apply the cohomology functor, rather it is better to apply it in $\mathrm{D}(\mathbf{A})$. To make this connection, the notion of *distinguished triangles* would have to be developed in some level of detail. I do wish to do so since I'm interested in some of the connections they have, but for now I shall follow Aluffi's example and simply state the ingredients necessary on a shallow enough level to explain how the morphism δ would arise without needing any special considerations.

Let $\mathbf{A}$ be any Abelian category. We shall define the notion of a *distinguished triangle*. We first define a functor which is called the *translation functor*. Though it is called a translation functor, it shall be characterized by the axioms of a distinguished triangle. For the categories $\mathrm{K}(\mathbf{A})$ and $\mathrm{D}(\mathbf{A})$ this is the shift functor $L^\bullet \xrightarrow{+1} L^\bullet[1]$. Distinguished triangle are triangles that must satisfy 4 axioms:

1. The triangle:

$$\begin{tikzcd} & 0 \arrow[dr] & \\ A \arrow[ur, "+1"] & & A \arrow[ll, "\mathrm{id}_A"'] \end{tikzcd}$$

has to be part of the collection of distinguished triangles

2. every morphism $\alpha : A \to B$ must be the side of a distinguished triangle.
3. Triangles that are isomorphic (as triangle diagrams) to distinguished triangles must also be distinguished
4. Distinguished triangles must be able to rotate, namely if:

$$\begin{tikzcd} & A \arrow[dr, "\alpha"] & \\ C \arrow[ur, "\gamma^{+1}"] & & B \arrow[ll, "\beta"] \end{tikzcd}$$

is a distinguished triangle if and only if

$$\begin{tikzcd} & A \arrow[dr, "\beta"] & \\ A[1] \arrow[ur, "-\alpha^{+1}"] & & B \arrow[ll, "\gamma"] \end{tikzcd}$$

is a distinguished triangle (think of visual ref:HERE, the 3D view, to intuit this axiom)

5. Finally, distinguished triangles must satisfy an axiom called the *octahedral axiom* which will not be necessary to cover for the remarks this section endeavors to explicate.

A category that contains a collection of distinguished triangles is called a *triangulated category.* The category $\mathrm{K}(\mathbf{A})$ is an example of such a category where the translation functor is the shift functor and all distinguished triangles are isomorphic to:

$$\begin{tikzcd} & L^\bullet \arrow[dr, "\alpha^\bullet"] & \\ MC(\alpha^\bullet) \arrow[ur, "+1"] & & M^\bullet \arrow[ll] \end{tikzcd}$$

Note how this diagram also fulfills the 3rd axiom: any map $\alpha^\bullet : L^\bullet \to M^\bullet$ is the edge of the above distinguished triangle where the other morphisms are the natural ones. Note also how the axioms prevent $\mathrm{Ch}(\mathbf{A})$ from being a triangulated category. Since

$$\begin{tikzcd} & 0 \arrow[dr] & \\ A \arrow[ur, "+1"] & & A \arrow[ll, "\mathrm{id}_A"] \end{tikzcd}$$

needs to be a distinguished triangle, then since $L^\bullet$ is the mapping cone of the zero-morphism $0 \to L^\bullet$ the rotation would axiom would require

$$\begin{tikzcd} & L^\bullet \arrow[dr, "\mathrm{id}_{L^\bullet}"] & \\ 0 \arrow[ur, "+1"] & & L^\bullet \arrow[ll] \end{tikzcd}$$

The mapping cone of the identity is not 0, and hence it cannot be a distinguished category. However, the mapping cone of the identity is *homotopically-equivalent* to 0, and hence the above triangle is indeed distinguished in $\mathrm{K}(\mathbf{A})$. Furthermore, a triangulated category, the cohomology functor will map distinguished triangles to exact triangles (in the case of $\mathrm{K}(\mathbf{A})$, this is because all distinguished triangles are isomorphic to the triangles with mapping cones), and hence give long exact sequences. Two notable examples of cohomological functors that map quasi-isomorphisms to isomorphisms for any triangulated category $\mathbf{A}$ are $\mathrm{Hom}(A, -)$ and $\mathrm{Hom}(=, A)$ for any object $A \in \mathbf{A}$.

The derived category $\mathrm{D}(\mathbf{A})$ is naturally a triangulated category, and the natural functor $\mathrm{K}(\mathbf{A}) \to \mathrm{D}(\mathbf{A})$ maps distinguished triangles to distinguished triangles and maps the translation functor to the translation functor. Derived categories are less restrictive than homotopic categories in the following important way: if

$$0 \to L^\bullet \xrightarrow{\alpha^\bullet} M^\bullet \xrightarrow{\beta^\bullet} N^\bullet \to 0$$

is a short exact sequence, then in general there is no distinguished triangle

$$\begin{tikzcd} & L^\bullet \arrow[dr, "\alpha^\bullet"] & \\ N^\bullet \arrow[ur, "\gamma^\bullet\ {}^{+1}"] & & M^\bullet \arrow[ll, "\beta^\bullet"] \end{tikzcd}$$

In $\mathrm{K}(\mathbf{A})$ for any choice $\gamma^\bullet$ (even more is the case, $N^\bullet$ need not be homotopically equivalent to $MC(\alpha^\bullet)$). However, such a triangle *does* always exist in $\mathrm{D}(\mathbf{A})$), namely because there does exist a quasi-isomorphism $MC(\alpha)^\bullet \to N^\bullet$. Then applying the cohomology functor to this triangle shall give

$$\begin{tikzcd} & H^\bullet(L^\bullet) \arrow[dr] & \\ H^\bullet((N^\bullet) \arrow[ur, "+1"] & & H^\bullet((M^\bullet) \arrow[ll] \end{tikzcd}$$

where $+1$ is the connection homomorphism δ up to isomorphism. Furthermore, the 0 morphisms that connect the above in when applying the $H^\bullet(-)$ turned out to be particular choices in the more general case.

Finally, it should be noted that any functor $\mathscr{F} : \mathbf{A} \to \mathbf{B}$ between Abelian categories shall induce a functor $\mathrm{K}(\mathscr{F}) : \mathrm{K}(\mathbf{A}) \to \mathrm{K}(\mathbf{B})$, and this functor is in fact a functor of triangulated categories. Hence, the bounded equivalents, as well as $\mathrm{L}(\mathscr{F}), \mathrm{R}(\mathscr{F}) : \mathrm{D}^{\pm}(\mathbf{A}) \to \mathrm{D}^{\pm}(\mathbf{B})$ are functor of triangulated categories. From this perspective, the results proven about derived functors is a consequence of this fact.

28

Applications of (co)Homology

With cohomology theory more securely under our belt, it is time to use it!

1. Shall use it to find cohomological invariants of groups
2. Shall show some ways of intepreting Ext and Tor
3. Shall show that the notion of dimension from the geometric setting can be brought over to the algebraic!
4. Some K-theory will certainly interest the reader
5. Perhaps even Euler's identity?

Won't do sheaf cohomology, that is an exploration for EYNTKA Algebraic Geometry.

28.1 Group Cohomology

Given a (not necessarily abelian) group G, what would it mean to extract cohomological information from it? A (co)chain information would have to naturally be formed, and a functor must be found to derive. Groups themselves don't form an abelian category, it was shown that representation, that is $k[G]$-modules do form them. This is perhaps a first place in which we may think we can study groups, however this category is too nice (since we are working with vector-spaces). What is more useful is it's generalization:

Definition 28.1.1: G-Modules

Let G be an arbitrary group, and M an abelian group. Then M is a *G-module* if there exists a [left] group-action on M. Equivalently, if there exists a $\mathbb{Z}[G]$-module structure on M.

The reader is free to parse through the technicalities of the group action. All $k[G]$-modules presented in chapter 24 are $\mathbb{Z}[G]$-modules with the action restricted. An example of a G-module that is not a representation is the action of the group $\mathrm{Gal}_K(L)$ on the module L/K (a galois extension of L over K). All finite groups G have a non-trivial G-module given by the action of G on the group-ring $k[G]$ (note the addition is abelian). As we have seen many times at this point, the object acts on an abelian group can give lots of information on the object acting on it and the object being acted on. Given a G-module M, we can define the following:

Definition 28.1.2: G-Invariant subgroup

Let M be a G-module. Then the *G-invariant* set of M is the set:

$$M^G := \{m \in M \mid g \cdot m = m, \ \forall g \in G\}$$

which is the largest trivial G-submodule of M.

The following will help in computing cohomology

Corollary 28.1.3: Hom Representation of G-invariant

$$M^G \cong \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, M^G) \cong \mathrm{Hom}_G(\mathbb{Z}, M)$$

Proof **:**
exercise (what G-action should be put on $\mathbb{Z}$?)

The action G descends onto M^G and becomes the trivial action, and it is not too hard to see that any G-module homomorphism can descend functorially, hence $(-)^G$ is a covariant functor $G\text{-Mod} \to \mathbf{Ab}$. Importantly

Proposition 28.1.4: Invariants Is Left-exact

The functor $(-)^G$ is left-exact

Proof **:**
Prove that $(-)^G$ is right-adjoint to the trivial module functor.

The functor need not be right exact:

Example 28.1.1: Lack Of Right-exactness's
Let G act trivially on a coset $[m]$ of a quotient M/L. Can this action awlays be lifted so that G acts trivially on a representative m?

Thus, it is fruitful to take the ith right-derived functor of $(-)^G$:

$$\mathbf{R}^i(-)^G := H^i(G, -)$$

which would give us the long exact sequence of abelian groups:

$$0 \to L^G \to M^G \to N^G \to H^1(G, L) \to H^1(G, M) \to \cdots$$

Since $M^G = \mathrm{Hom}_G(\mathbb{Z}, M)$ by taking an injective resolution of M, or a projective resolution of $\mathbb{Z}$, we can write:

$$H^i(G, M) = \mathrm{Ext}^i_{\mathbb{Z}[G]}(\mathbb{Z}, M)$$

Example 28.1.2: Group Cohomologies

1. The trivial group has trivial cohomology
2. (Integers) Let $G = \mathbb{Z}$. Then $\mathbb{Z}[G] \cong \mathbb{Z}[x, x^{-1}]$. Then since $\mathbb{Z}[x, x^{-1}]/(1-x) \cong_G \mathbb{Z}$ with the trivial action, define the following free (hence projective) resolution:

   $$\cdots \to 0 \to \mathbb{Z}[x, x^{-1}] \xrightarrow{(\cdot(1-x))} \mathbb{Z}[x, x^{-1}] \to 0 \to \cdots$$

   Applying the (contravariant) $\mathrm{Hom}_{\mathbb{Z}[x,x^{-1}]}(-, M)$, we can compute the cohomology of

   $$\cdots \to 0 \to M \xrightarrow{(\cdot(1-x))} \to M \to 0 \to \cdots$$

   Then we can conclude either by direct computation or referencing ref:HERE that

   $$H^0(\mathbb{Z}, M) \cong \ker(M \xrightarrow{(\cdot(1-x))} M) = M^{\mathbb{Z}}$$

   and:

   $$H^1(\mathbb{Z}, M) = \frac{M}{(1-x)M}$$

   and furthermore $H^i(\mathbb{Z}, M) = 0$ for $i \notin \{0, 1\}$
3. (Finite Cyclic Groups)
4. (If you know some algebraic geometry) Take the category of varieties W over k such that when extending scalars to $\overline{k}$ we have $V/\overline{k} \cong W/\overline{k}$. Let's label it $\mathbf{Var}_{k \to \overline{k}}$, and say that the objects are:

   $$\mathrm{Obj}(\mathbf{Var}_{k \to \overline{k}}) = \left\{ W/k \;\middle|\; V/\overline{k} \cong W/\overline{k} \right\}$$

   and the morphism are regular morphisms. Then I claim that:

   $$\mathbf{Var}_{k \to \overline{k}} \cong H^1(\mathrm{Gal}_k(\overline{k}), \mathrm{Aut}(V/\overline{k}))$$

   To see this, first note that an element $\sigma \in \mathrm{Gal}_k(\overline{k})$ acts on a $\overline{k}$-variety $W/\overline{k} = W_k \times_{\mathrm{Spec}(k)} \mathrm{Spec}(\overline{k})$ via:

   $$\mathrm{id}_{W_k} \times \mathrm{Spec}(\sigma^{-1}) : W/\overline{k} \to W/\overline{k}$$

Or, for those less comfortable with fibered produced from algebraic geometry, they can think of this map as the "equivalent":

$$\mathrm{id}_{W_k} \otimes_k \sigma^{-1} : k[W] \otimes_k \overline{k}$$

With the action defined, let us look at the correspondence. Take $W_k \in \mathbf{Var}_{k \to \overline{k}}$ so that $\varphi : W_{\overline{k}} \xrightarrow{\sim} V_{\overline{k}}$ is an isomorphism. Then notice that this isomorphism induces a map $\varphi \circ \sigma^{-1} \circ \varphi^{-1}$, which is an automorphism on $W_{\overline{k}}$. The collection of all these maps as σ varies can uniquely define W_k (k is the fixed subfield over all σ). This is true up to a choice of $\mathrm{Aut}_k(V)$, and so we quotient out by it. Then we can see this creates the correspondence between these two categories, and shows how far away the category $\mathbf{Var}_{k \to \overline{k}}$ is from being invariant to the group-action of the galois group.

The first cohomology group for when $G = \mathbb{Z}$ is interesting, we can deduce from it that there is somehow some trivial action going on in some interesting way, and indeed:

Definition 28.1.5: G-Coinvariant subgroup

Let M be a G-module. Then the G-*coinvariant* of M is the set:

$$M_G := \frac{M}{\langle (gm - m) \mid g \in G, m \in M \rangle}$$

The same way M^G is the maximal invariant subgroup, M_G is the maximal invariant quotient module with trivial action. In fact, $(-)_G$ can create it's own functor that is right-exact. The reader should further verify that $M_G \cong_G \mathbb{Z} \otimes_{\mathbb{Z}[G]} M$ (with $\mathbb{Z}$ the trivial G-module structure). In this way, we may compute *left-derived functors* as well! This functor is best seen as studying *group homology*.

Let us now give a standard way of defining a projective resolution of $\mathbb{Z}$ by G-modules. First, let $\mathbb{Z}[G^n]$ be a G-module with the diagonal action given by multiplying on the left, or more specifically:

$$g \cdot (g_1, ..., g_n) = (gg_1, ..., gg_n)$$

These modules are free G-modules; clearly $\mathbb{Z}[G]$ is a $\mathbb{Z}$-module with basis given by the elements G. Then for $n \geq 0$:

$$Z[G^{n+1}] \cong \bigoplus_{(g_1,...,g_n) \in G^n} \mathbb{Z}[G](1, g_1, ..., g_n)$$

which shows that these are free. From this, we can consider the following projective (even free) resolution:

Definition 28.1.6: Standard Resolution

Let G e a group. Then the *standard resolution* of $\mathbb{Z}$ by G-modules is the exact sequence of G-module homomorphism:

$$\cdots \to \mathbb{Z}[G^{n+1}] \xrightarrow{d_n} \mathbb{Z}[G^n] \xrightarrow{d_{n-1}} \cdots \xrightarrow{d_1} Z[G] \xrightarrow{d_0} \mathbb{Z} \to 0$$

where on the group G we define:

$$d_n(g_0, ..., g_n) = \sum_{i=1}^{n} (-1)^i (g_0, ..., \hat{g}_i, ..., g_n)$$

where the term $\hat{g}_i$ is omitted from the tuple. The map d_0 is also called the *augmentation map* and is often denoted ϵ.

Note that each $\mathbb{Z}[G^{n+1}]$ is a We claimed this is an exact sequence, however this ought to be verified.

Lemma 28.1.7: Standard Resolution is Exact

The standard resolution of $\mathbb{Z}$ by G-modules is exact

Proof **:**
This a proof that is about keeping track a lot of technical details, but it is all eminently doable, and so left to the willing reader.

Hence, after taking the contravariant functor $\mathrm{Hom}_{\mathbb{Z}[G]}(-, M)$, we get the cochain complex:

$$0 \to \mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G], M) \to \cdots \to \mathrm{Hom}_{\mathbb{Z}[G^n]}(-, M) \xrightarrow{d_n^*} \mathrm{Hom}_{\mathbb{Z}[G^{n+1}]}(-, M) \to \cdots$$

where d_n^* maps $\varphi \mapsto \varphi \circ d_n$. Similarly, we can take the covariant functor

$$\cdots \to \mathbb{Z}[G^3] \otimes_{\mathbb{Z}[G]} A \xrightarrow{d_{2*}} \mathbb{Z}[G^2] \otimes_{\mathbb{Z}[G]} A \xrightarrow{d_{1*}} \mathbb{Z}[G] \otimes_{\mathbb{Z}[G]} A \xrightarrow{\epsilon} 0$$

where d_{n*} is given by $(g_0, ..., g_n) \otimes a \mapsto d_n(g_0, ..., g_n) \otimes a$, the map ϵ is the augmentation map given by $\sum a_g g \mapsto \sum a_g$, and we view the modules as *right* $\mathbb{Z}[G]$-modules wit the action

$$(g_1, ..., g_n) \cdot g = (g - 1g, ..., g_n g)$$

Let us now better understand these hom-sets, for their structure turns out to be rather simple. Let

$$C^n(G, M) = (\mathrm{Hom}_{\mathbf{Set}}(G^n, M), +)$$

where with group operation defined on the maps given by point-wise addition and is called the n-cochain group. From this, we can define a boundary map:

$$\begin{aligned} d^n &: C^n(G, M) \to C^{n+1}(G, M) \\ d^n(f)(g_0, ..., g_n) &= g_0 f(g_1, ..., g_n) - f(g_0 g_1, g_2, ..., g_n) \\ &\quad + f(g_0, g_1 g_2, ..., g_n) - \cdots + (-1)^n f(g_0, ..., g_{n-2}, g_{n-1} g_n) + (-1)^{n+1} f(g_0, ..., g_{n-1}) \end{aligned}$$

Then $d^{n+1} \circ d^n = 0$, giving us a cochain complex. We can then have:

$$H^n(G, M) = \frac{\ker d^n}{\operatorname{im} d^{n-1}} = \frac{Z^n(G, M)}{B^n(G, M)}$$

Call $Z^n(G, M)$ the set of *n-cycles*, and $B^n(G, M)$ the set of *n-boundaries*. Notice that:

1. $C^0(G, M) = \{1\}$, and so $C^0(G, M) \cong M$
2. $d^0 : C^0(G, M) \to C^1(G, M)$ is given by $d^0(a)(g) = ga - a$
3. $H^0(G, M) \cong \ker d^0 = M^G$
4. $\operatorname{im} d^0 = \{f : G \to M \mid \exists a \in M \text{ such that } f(g) = ga - a,\ \forall g \in G\}$ (principal crossed homomorphism)
5. $d^1 : C^1(G, M) \to C^2(G, M)$ given by $d^1(f)(g, h) = gf(h) - f(gh) - f(g)$
6. $\ker d^1 = \{f : G \to M \mid f(gh) = f(g) + gf(h)\}$ (crossed homomorphisms)
7. If $0 \to A \to B \to C \to 0$ is exact, we get an exact sequence of cochains for every n:

$$0 \to C^n(G, A) \to C^n(G, B) \to C^n(G, C) \to 0$$

We shall show that the collection of complexes given by the above is isomorphic to those given by the standard resolution

Proposition 28.1.8: Characterizing Hom of Resolution

Let M be a G-module. Then for every $n \geq 0$, we there is an isomorphism:

$$\Phi^n : \operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^{n+1}], M) \xrightarrow{\cong} C^n(G, M)$$

where the map $\varphi : \mathbb{Z}[G^{n+1}] \to M$ goes to $f : G^n \to M$ given by:

$$f(g_1, ..., g_n) = \varphi(1, g_1, g_1g_2, g_1g_2g_3, ..., g_1g_2g_3 \cdots g_n)$$

The isomorphism Φ^n commutes with boundary maps, that is

$$\Phi^{n+1} \circ d^*_{n+1} = d^n \circ \Phi^n$$

Hence we have an isomorphism from the category of cochain complexes with the standard resolution to the one with boundary-cycle maps.

***Proof* :**
very doable, and worth doing at least once.

Due to the relevance of the standard representation, we have the following definition boed

Definition 28.1.9: Induced and Coinduced G-module

Let G be a group and A an abelian group. Then:

$$\mathrm{CoInd}^G(A) = \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}[G], A)$$

with the G-action defined as $(g\varphi)(z) = \varphi(zg)$ is called the *coinduced G-module* associated to A. Similarly

$$\mathrm{Ind}^G(A) = \mathbb{Z}[G] \otimes_{\mathbb{Z}} A$$

with G-action defined by $g(z \otimes a) = (gz) \otimes a$ is the is the induced G-module associated to A.

A result that is common in many Cohomologies theories is some version of the following:

Lemma 28.1.10: Induced and Coinduced Equivalence

Let G be a finite group and A an abelian group. Then

$$\mathrm{Ind}^G(A) \cong \mathrm{CoInd}^G(A)$$

As a consequence, if $B = \mathrm{Ind}^G(A)$ or $B = \mathrm{CoInd}^G(A)$, then

$$H_0(G, B) \cong H^0(G, B) \cong A$$

and they are both zero for $n \geq 1$.

Proof **:**

Define the map: $\alpha : \mathrm{Coind}^G(A) \to \mathrm{Ind}(G)(A)$ given by:

$$\varphi \mapsto \sum_{g \in G} g^{-1} \otimes \varphi(g)$$
$$(g^{-1} \mapsto a) \leftarrow\!\shortmid g \otimes a$$

with $(g^{-1} \mapsto a)(h) = 0$ for $h \in G \setminus \{g^{-1}\}$. These are certainly inverses, and easily verifiable to be G-module homomorphisms (this is a similar technique used in Maschkle's Theorem (see theorem 24.1.8).

The last part is left for exercise ref:HERE.

Exercise 28.1.1

1. Let A, B be any G-modules. Show that $H^n(G, A \oplus B) \cong H^n(G, A) \oplus H^n(G, B)$
2. Show that $H^0(G, \mathrm{CoInd}^G(A)) \cong A$ and $H^n(G, \mathrm{CoInd}^G(A)) \cong 0$ for all $n \geq 1$ (hint: consider $\zeta : \mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^n], \mathrm{CoInd}^G(S)) \to \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}[G^n], A)$ given by $\varphi \mapsto (z \mapsto \varphi(z)(1))$. What is the inverse of this map?)
3. Show that $H_0(G, \mathrm{Ind}^G(A)) \cong A$ and $H_n(G, \mathrm{Ind}^G(A)) \cong 0$ for all $n \geq 1$.

28.2 K-Theory

(I'm not sure if I should keep my K-theory notes separate or in here. I may just make the basics of K-theory prerequisite and expand some of the cohomological considerations here)

28.3 Homological Dimension

29

Spectral Sequences

It is often the case that the (co)homology of an object is difficult to calculate, but the object can be broken down into simpler objects whose (co)homology is more readily computed. Perhaps the simplest type decomposition of an object is that of a filtration. The reader that read section 20.4.2 may recognize the notion of a filtration for an R-modules, which is a sequence:

$$M = M_0 \supseteq M_1 \supseteq \cdots \supseteq M_n \supseteq \cdots$$

For the case of homology, the sequence usually is increasing. In the category R-**Mod**, each respective M_i will have an associated $P_i^\bullet$ associated to it creating a natural grid

$$\begin{tikzcd}
 & \vdots \arrow[d] & \vdots \arrow[d] & \\
\cdots \arrow[r] & P_{i,j} \arrow[r] \arrow[d] & P_{i,j+1} \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & P_{i+1,j} \arrow[r] \arrow[d] & P_{i+1,j+1} \arrow[r] \arrow[d] & \cdots \\
 & \vdots & \vdots &
\end{tikzcd}$$

From this grid, we will be able to extract (co)homological information connecting (co)homology of M_i and M_{i+1}. The spectral sequence then will be a way of combining this information to give cohomological information about M. The goal of this chapter is to go about the formalization of these concepts and give some interesting examples. As it turns out, there are many ways of creating grid of information, and so the definition of spectral sequences takes this into account and may at first seem a little separated from this original intuition, however examples will be given to make the connection.

29.1 Definition and Examples

On a high-level, a spectral sequence arises from a special type of exact triangle:

Definition 29.1.1: Exact Couple

Let A, E be two object in an abelian category A. Then the exact triangle

$$\begin{tikzcd} & A \arrow[dr, "\alpha"] & \\ E \arrow[ur, "\gamma"] & & A \arrow[ll, "\beta"] \end{tikzcd}$$

is called an *exact couple*. Notice that the object A is repeated twice.

No assumption are done on the function γ. Define the function $d : E \to E$ where $d = \beta \circ \gamma$. Since an exact couple is an exact triangle, $\gamma \circ \beta = 0$, and so

$$\delta \circ \delta = \beta \circ \gamma \circ \beta \circ \gamma = \beta \circ 0 \circ \gamma = 0$$

Hence the (co)homology of E with respect to d is well-defined and is

$$E = \frac{\ker d}{\operatorname{im} d}$$

We may furthermore define:

- $A' = \operatorname{im} \alpha$
- $\alpha' : A' \to A'$ by restricting α to A'
- $\beta' : A' \to E'$ by defining $\beta(\alpha(a)) = [\beta(\alpha)]$
- $\gamma' : E' \to A'$ by defining $\gamma([e]) = \gamma(e)$ (which by exactness is well-defined)

What is interesting is the following

Lemma 29.1.2: Derived Couple

Given an exact couple, then the E' and A' along with the above map give rise to another exact couple, namely

$$\begin{tikzcd} & A' \arrow[dr, "\alpha'"] & \\ E' \arrow[ur, "\gamma'"] & & A' \arrow[ll, "\beta'"] \end{tikzcd}$$

is an exact triangle

Proof **:**
This is diagram chasing (p.523 of Hatcher chap 5 if you want to verify)

The resulting exact couple is called a derived couple, where the nomenclature arose separately from the nomenclature for derived categories (it arose in originally in algebraic topology). Thus, since the result is another exact triangle, the process can repeat itself indefinitely, giving a sequence:

Definition 29.1.3: Spectral Sequence

Let $\mathbf{A}$ be an abelian category. Then a *spectral sequence* is a sequence of tuples $\{(E_i, d_i)\}_i$ where $d_i : E_i \to E_i$ are maps with the property that $d_i \circ d_i = 0$ and $E_{i+1} = \ker d_i / \operatorname{im} d_i$

Part VIII

Category Theory

for adjoints eventually

(I'm starting to think that this part should actually be a chapter in the Appendix going over

1. All the category theory that was build up as needed through the book, so categories, duality, functors and natural transformations, equivalence of categories, diagram chasing, universal properties, limits and colimits, adjoints.
2. Then, in a second section, go over Yonneda's lemma and Kan Extensions.

)

In this last part of this book, we will take a deeper dive into category theory. I want to write about my experiences with Category theory because it really feels like if you don't do this categorically, you don't actually understand what's going on. At best, you only see half the picture!! The key idea in Category theory is that when defining objects, we are not defining them by *intrinsic* properties objects might have, but rather, we will define objects by *how they relate to every other object*. It's sometimes said that category is 10% about objects [1], and 90% about the relation's between them. This intuition is actually rigorous in one of the (if not the) central theorem in Category theory, called *Yoneda's* Lemma, which we will cover soon.

Also, since category theory has application beyond algebra, I will be slipping in some examples from graph theory, set theory, topology, and logic. These can be skipped, but if you know these domains, this will be even more rewarding a section to read

In the algebraic structures we've studied (Groups, Abelian Groups, Rings, Fields, Group-Actions, Modules, Vector Spaces, etc), we found a lot of parrallels between concepts

1. groups, rings, modules, each have practically identical isomorphism theorems
2. all *group-actions* can be represented a a special case of a *group*
3. All free *modules* can be seen as a direct some of the underlying *ring*
4. The Fundamental Theorem of Galois theory links the *group* lattice of a galois group to a galois *field*

(Also a word on how categories are used far and wide a field, touching essentially everything from real and complex analysis to combinatorics to logic to geometry and topology and so on. Such examples will be included in this section)

[1] 10% because all objects can be thought of as identity morphisms!

30

Summary of Category Theory

In this chapter, we summarize the basic terminology introduced sporadically in this book.

Definition 30.0.1: Category

A ***category*** consists of[a]

1. A colletion of *objects*
2. A collection of *morphisms*

where

- Each morphism has a specified ***domain*** and ***codomain*** object. Notationally, this is written as $f : X \to Y$, where f is the morphism with domain X and codomain Y [b]
- Eah object has an ***identity morphism*** $1_X : X \to X$
- For each pair of morphisms f, g with the codomain of f equal to the domain of g, there exists a specified ***composite morphism*** gf (or $g \cdot f$) whose domain is equal to the domain of f and whose codomain is equal to the codomain of g, i.e.

$$f : X \to Y \qquad g : Y \to Z \qquad \Rightarrow \qquad gf : X \to Z$$

such that

1. For any $f : X \to Y$, the composites $1_Y f$ and $f 1_X$ are both equal to f (i.e. unital with the identity homomorphism)
2. For any composable triple of morphisms f, g, h, the composites $h(gf)$ and $(hg)f$ are equal (i.e. associative), and so we can write hgf unambigously.

[a] The word *collection* being used instead of *set* was purposeful; we will get to this
[b] Note that f **need not be a function**! We will give examples of this shortly

Since identity morphisms are unique, they are in bijective correspondence to objects, meaning we can technically replace objects from our definition with the identity morphisms. Some authors do, but we stick to the more conventional object/morphism distinction. Another convention some some authors take is to interchange the word morphisms and *arrows*, i.e. the word arrow and morphism both mean the same thing.

Most of Category theory is about studying the *morphisms* in a category, since the major concern of category theory is how can we relate objects in or between categories, so it might be a bit counter-intuitive that many categories are named by the *object* that are inside it. This is simply convention, with most of the times the morphisms being "clear" in most contexts.

Example 30.0.1: Algebraic examples

1. The **Set** Category has *sets* as objects, and *funtions* (with specified domain and codomain) as morphisms. This is for most the first category you work with, and what you are introduced as the foundation of mathematics. At some point in your mathematical career, you might have been uncomfortable at how "strong" the definition of a function is. To me, it felt like there were too many restriciton given to them to truly be the way to describe mathematics generally. This feeling is ratified in Category theory, where it is simply one category you can work in. Many categories will build off this one (think of how group and all further

algebraic structures start with sets, or topologies having an underlying set), but there are also categories which will not!

2. The **Group** Category (sometimes denoted **Grp**) has *groups* as objects, and *group homomorphisms* as morphisms. It's from the word *homomorphism* that the term *morphism* etymologically originates.

3. The **Ring** Category has *rings* as objects, and *ring homomorphisms* as morphisms.

4. The **Field** Category has *field* as objects, and *ring homomorphisms* as morphisms, since fields are [unital] commutative division rings. This shows how you can have change what objects you have to defined a different category.

5. Similarly, we can define a category with *groups* as objects, and morphisms being the *isomorphisms*, keeping the object, but changing the morphism.

6. The $\mathbf{Mod}_R$ Category has *Left Modules* as objects and and *R-module homomorphisms* as morphissm.

   (a) When $R = K$ is a field, then we denote this category $\mathbf{Vect}_K$

   (b) When $R = \mathbb{Z}$, then we dneote this category **Ab**, since $\mathbb{Z}$-modules are precicely the abelian groups.

Categoreis are used in *much* more contexts than algebra. For example,

Example 30.0.2: Non-Algebraic Examples

1. The **Top** Category has *topological spaces* as objects and *continuous functions* as morphisms.

2. The $\mathbf{Top}_*$ Category has *topological spaces with specified basepoint* as objects and *base-point preserving continuous functions* as morphisms. We can similarly define $\mathbf{Set}_*$.

3. The **Meas** Category has *space* as objects and *measurable functions* (or meseomorphism) as morphisms. This is an example of

4. The **Man** Category has *manifolds* as objects and and *smooth maps* as morphisms

5. The **Graph** Category has *graphs* as objects and *graph morphisms* as morphisms (functions carrying vertices to vertices, and edges to edges preserving incidence relations). As a special case, **DirGraph** as *directed graphs* as objects, and *directed graph morphissm* as morphisms, whcih also preserves the soruces and target

6. The **Poset** Category has *partly ordered sets* as objects and *orer-preserving functions* as morphissm

7. Cool example generalizing the category of group, rings, fields, with the $\mathbf{Model}_{\mathbb{T}}$

All of these examples are in some sense related to **Set**. Each object have an underlying set, *and* each morphism was a function with added properties; the "structure-preserving" morphism. Categories of relying on **Set** are called *small categories*.

Example 30.0.3: : Non Set Examples

1. Let R be a unital ring. Then the $\mathbf{Mat}_R$ Cateogry has *positive integers* as objects, and the morphisms between the integer n and m is the set of $m \times n$ matrices with values in R. Composition is by matrix multiplication of every element. If A is an $m \times n$ matrices and B is a $k \times m$ matrices, then their composition is:

$$n \xrightarrow{A} m, \qquad m \xrightarrow{B} k \qquad \Rightarrow \qquad n \xrightarrow{B \cdot A} k$$

   with identity matrices serving as the identity morphisms. The morphisms in this category can no longer be written as functions between two positive integers, as it doens't make sense to say the morphism in this category takes elements from the "set" n and maps it to the "set" m (n and m aren't even sets)

2. A poset $(P, \leq)$ (partially ordered set) is a category, where the objects of P are the elements, and there exists a unique morphim $x \to y$ if and only if $x \leq y$. Transitivity and reflextivity again make sure that the unique morphism exists, and assocaitive composition is well-defined.

3. The **Htpy** Category (Homotopy Category) has *topological spaces* as it's objects and morphisms are *homotopy class of continuous maps*. We can define $\mathbf{Htpy}_*$ similarly to $\mathbf{Set}_*$.

4. The **Cat** is the Category of categories!! However, we need to be a little precise in order to avoid Russel's paradox (as we'll explain after this ample). We will make the objects of **Cat** *small* categories. Though **Group**, **Ring**, etc. Are sub-categories of **Cat**, none of them are objects in **Cat**. If we are more generous, we can define **CAT** to be the category of *locally small categories*. The morphisms in this category are *functors*. We will get back to these in much more detail soon! Interestingly, once we get there, we will actually be interesting in one more generalization where the morphisms of this category become the objects, and the morphism being a morphism between functor!

As noted in a footnote of the defining, I made sure to say *collection* instead of *set* when thinking of all objects or morphisms. This is to to avoid *Russell's paradox*. Russel's paradox has a technical phrasing, but we will give a easier and more common way of understanding Russell's paradox:

67: Russell's Paradox

Fix S to be the set of all sets (it contains $\{\}$, $\mathbb{N}$, $\{\{\mathbb{R}\}\}$, and any other set you can imagine as a subset). Since we have defined S, we now have another set (that is, S is a set). Since S contains all sets, that means S contains S, that is $S \subset S$. But just a moment ago, we fixed S to have all the sets, meaning it cannot be that S missed a set. But the moment we define the set S, we have that S should have already contained it *before* it was even defined.

Thus, we usually don't set the set of all sets, but the *collection* of all sets (simply to avoid the term set). Similarly, Russell's paradox comes in when dealing with the category of all categories, and so we often limit the size of categories: Because of this, we can make the distinction of categories form which we can make a *set* of it's morphisms or not

Definition 30.0.2: Small Category

A category is ***small*** if it has only a set's worth of morphisms (arrows)

Since there always exists the identity morphism, this implies a sets worth of objects too.

As noted, none of the algebraic categories are small. However, when taking an appropriate subcollection, they will resemble small categories. We do this often, and so we label this:

Definition 30.0.3: Locally Small Category

A category is **locally small** if between any pari of objects, theer is only a set's worht of morphisms, usually denoted

$$C(X,Y) \qquad \text{or} \qquad \mathrm{Hom}(X,Y) \qquad \text{or} \qquad \mathrm{Hom}_C(X,Y)$$

The set of all moprhisms between particular objects is called a **hom-set**. Note that it's not necessarily a homomorphism (though the name was taken form homomorphism between particular Modules). Note that For the Hom_C notation, the C stands for *composition*, and is the place-holder for what the composition in this category is.

Though we used the notation for locally small categories, it has become popular enough to be used even if the category is *not* locally small.

Example 30.0.4: Hom-Set

1. This is already known notation! For example, $\mathrm{Hom}(G, S_A)$ is all homomorphisms form some group G to S_A, and it represented a group action. Thus, a group-action is simply $\mathrm{Hom}(*, S_{|G|})$, where $*$ represents a group in **Group**. More cool insights here.

Now, we define the notion of two objects being equal in a category

Definition 30.0.4: Isomorphism

An **isomorphism** in a category is a morphism $f : X \to Y$ for which there exists a morphism $g : Y \to X$ so that $gf = 1_X$ and $fg = 1_Y$. The objects X and Y are **isomrophic** whenever there exists an isomorphism between X and Y, in which case we write $X \cong Y$
An **endomorphism** (morphism with equal domain and codomain) that is an isomrophism is called an **automorphism**. We will usually denote the collection of all automrophism as $\mathrm{Aut}(*)$ (where $*$ is the object in question)

Example 30.0.5: Isomorphisms in Different Categories

1. The isomorphisms in **Set** are precisely the bijective functions! If you have $f : X \to Y$ that is bijective, then there exists an inverse such that $f^{-1} \circ f = id_X$, and $f \circ f^{-1} = id_Y$. This means that in **Set**, they key feature of a set is it's cardinality. It also makes the intuition that the elements of a set "don't matter", since a set with the same cardinality of with different elements is isomorphic!

2. The isomorphisms in **Group**, **Ring**, **Field**, $\mathbf{Mod}_R$, are the bijective homomorphisms

3. The isomorphisms in **Top** are the *homeomorphisms*, that is, the bicontinuous bijective functions (which is stronger than being a continuous bijective function)! This is perhaps the first example you've where a "homomorphism" which is bijective did *not* imply isomorphism.

4. The isomorphism of Posets as a category are the identities. Since we need that $x \leq y$ and $y \leq x$, this implies that $x = x$, leaving us only with the identity isomrophsim.

As fun aside, we will give a name to categories in which every morphism is an isomorphism, since it will lead to a categorical definition of a group!

Definition 30.0.5: Groupoid

A **groupoid** is a category in which every morphism is an isomrophism

Example 30.0.6: Rethinking a Groups

A **group** is a groupoid with one object! Let $*$ be the one object in the Category with every morphism being an isomorphism. Then

$$\mathrm{Hom}(*, *) = \mathrm{Aut}(*)$$

since every morphism is an isomrophism, and there is only one object, meaning we can only have isomorphism to itself, making it a homormophism. Let's say that $G = \mathrm{Aut}(*)$. Then by definition of a category, the morphisms in G are associative. Since $*$ is an object in a category, there exists a unique identity morphism, let's call it, e_*, for which the proof of uniqueness follows exactly like in proposition 1.1.8. Furthermore, since every morphism in $g \in G$ is an isomorphism, by the definition of an isomorphism, there exists a g^{-1} such that $gg^{-1} = g^{-1}g = 1_*$ (since there is only one object, there is only one identity morphism, forcing the two to be equal). Well, we have *exactly* proven the axioms of a group!

In Catagory theory, this becomes the actual definition of a group! The reason for this is that we have a clear idea of why those axioms are followed. In fact, many algebraic structures with a long list of axioms of some kind have become the consequence of working through some Categorical construction like we just did. This can make not only remembering the aximos easier (they're not just random axioms, but have real reasons to exist), but also bring deeper intuition on really how to think about those algebraic objects more broadly – every category, when restricted to a sub-category and isomrophism morphisms, will can be though of as groups!

We can continue this further to define rings! Consider G to be the category as before, but also make sure that

$$ab = ba \qquad a, b \in G$$

that is, composing morphisms from the left and right produce the same result (the abelian criterion as we needed for defining distributivity in a ring). Then we shall define a new category:

$$\mathrm{End}(G)$$

which has the morphisms of G as objects and its morphisms are morphisms between the morphisms of G. Take all the morphisms of G as the objects. Then this will actually define a ring (with identity)!

The really cool thing about this construction is that we are building up from the axioms of categories to construct other known algebraic structures!

We will leave with one more way of looking at smaller parts of a category, this time the more global approach:

Definition 30.0.6: Subcategory

A subcategory D of a category C is defined by restricting to a subcollection of objects and subcollection of morphisms subject to the requirements that the subcategory D contains the domain and codomain of any morphsim in D, the identity morphisms of any object in D, and the composite of any composable pair of morphisms in D

Example 30.0.7: Sub-Categories
Simple example,

$$\mathbf{CRing} \subset \mathbf{Ring} \subset \mathbf{Rng}$$

where **CRing** denotes the comutative ring ring with identity, **Ring** a ring with identity, and **Rng** a ring without idnteity.

Lemma 30.0.7: Maximal Groupoid Sub Category

Any Category C contains a **maximal groupoid**, the subcategory containing all of the objects and only those morphisms that are isormophisms

Proof **:**
exercice in the book! Should do! p.9 – Category Theory in Context

30.1 Duality

Category theory's philosophy is to look at relation between objects. The order of these relations matter, hence why we draw arrows, and not lines, for morphisms $X \to Y$. This is a very fundamental part of Category theory. So one might inquire, what happens if we turn the arrows the other way? Will we get different results? If not, then $X - Y$ might be a more natural notation. But it turns out that in many categories we *do* get dirrerent results. This leads to the following notion:

Definition 30.1.1: Opposite (Dual) Category

Let C be *any* Category. The **opposite category** C^{op} has

- the same objects as in C
- a morphisms f^{op} in C^{op} for each morphism f in C so that the domain of f^{op} is defined to be the codomain of f, and the codomain of f^{op} is defined to be the domain of f

$$f^{op} : X \to Y \quad \in C^{op} \qquad \leftrightarrow \qquad f : Y \to X \quad \in C$$

such that

1. For each object X, the morphism 1_X^{op} serves as its identity in C^{op}
2. when f, g are composable, $g \cdot f$, we will define $g^{op} \cdot f^{op}$ to be $(f \cdot g)^{op}$

will put commutative diagram here (p.9)

Example 30.1.1: Duality

1. If we take any poset as a category, it's possite poset will have a morphism between x and y if and only if $y \leq x$. If we take $(\mathbb{Z}, \leq)$ as a category, then

$$\cdots \longrightarrow 0 \longrightarrow 1 \longrightarrow 2 \longrightarrow 3 \longrightarrow \cdots$$

would be a diagram we can draw to illustrate some of the morphisms between objects. In the opposite category, this diagram would become

$$\cdots \longleftarrow 0 \longleftarrow 1 \longleftarrow 2 \longleftarrow 3 \longleftarrow \cdots$$

2. Let G be a group, thought of as a groupoid with 1 element, and so $G = \mathrm{Aut}(*)$. This collection on it's own actually constitues a category. Then since for any morphism $g \in G$, g is an isomorphism, it is easy to see how $g : * \to *$ defines $g^{op} : * \to *$ for any g, meaning that the morphisms of the opposite category act like morphisms in a group, meaning the oppossite of a group is also a group. Furthermore, if we take $g, h \in G$, then we need that $g^{op} \cdot h^{op} = (h \cdot g)^{op}$, or, translated into axioms, $g \cdot^{op} h = h \cdot g$, therefore, this morphism defines an operation which *switches* the order of g and h!

   Every opposite group is naturally isomprhic to group with $\varphi(g) = g^{-1}$.

   The place in which this is perhaps most intuitvely understood is when related to group-acitons. Let X be an object of some category, and $\psi : G \to \mathrm{Aut}(X)$ define a *right-action*. Then $\psi^{op} : G^{op} \to \mathrm{Aut}(X)$ defines a *left-action* by $\psi^{op}(g) \cdot x = x\psi(g) = xg$!

   There is also a natural generalization of this. Instead of taking all isomorphisms, take all morphisms! This will produce an algebraic structure called a *monoid*, where there need not be inverses!

3. If we take the **Set** category, then it's opposite category $\mathbf{Set}^{op}$ consists of the same objects, but every function $f : A \to B$ get's mapped to the "inverse function" $f^{-1} : B \to A$. Note

that this is *not* necessarily a function. Functions are right-unique ($a = b \Rightarrow f(a) = f(b)$), left-total (all elements of A need to map somewhere to B) binary relations between sets. For $f^{op} \cdot g^{op} = (f \cdot g)^{op}$, we will have that these "functions" are *left-unique* ($f^{-1}(a) = f^{-1}(b) \Rightarrow a = b$), *right-total* (for every elements in the co-domain, there is an element in the right)

There are many theorem's in mathematics that claim a duality between two objects, four notable examples being:

1. (Gelfand Duality) There is a "duality" between Compace Hausdorff spaces and commutative unital C^*-algebras[a]. Generalization of this duality works for LCH[b].
2. (Stone Duality) would love to put it here
3. (Zariski Duality) The duality between affine schemes and commutative algebra's.
4. (Pontryagin duality) The duality between a group G and the its' dual group $\hat{G}$ with dualizing object being the cirlce group S^1, $\hat{G} := \hom(G, S^1)$. This is a generalization of the Fourier Transfrom from Fourier analysis and is often studied in Harmonic Analyis.

[a] For those who know some Harmonic analysis, there is another duality also known as Gelfand Duality which gives a way of representing commutative Banach algebras as algebras of continuous functions. This duality is a generalization of the Pontryagin Duality

[b] TerryTaoBlock

The duality of categories means that proving something about a category also gets you an [equivalent] proof for it's dual category! In the next lemma, we will demonstrate this.

Lemma 30.1.2: Isomorphism And Relation To Every Other Object

The following are equivalent

1. $f : x \to y$ is an isomorphism in C
2. For all objects $c \in C$, post-composition with f defines a bijection
$$f_* : \mathrm{Hom}(c, x) \to \mathrm{Hom}(c, y) \qquad f_*(h) = fh$$
3. For all objects $c \in C$, pre-composition with f defines a bijection
$$f^* : \mathrm{Hom}(y, c) \to \mathrm{Hom}(x, c) \qquad f^*(h) = hf$$

***Proof* :**
Let's assume (1). let $f : x \to y$ be an isomorphism. We want ot show that f_* is a bijection. By definition, it thus has an inverse isomrophism $g : y \to x$. Thus, with g, we can construct g_*, which will be the two-sided inverse of f_*:

$$g_* : \mathrm{Hom}(c, y) \to (c, x)$$

With g_*, consider

$$g_* f_* : \mathrm{Hom}(c, x) \to \mathrm{Hom}(c, x) \qquad \text{and} \qquad f_* g_* : \mathrm{Hom}(c, y) \to \mathrm{Hom}(c, y)$$

To show it's a two sided inverse, take some $h \in \mathrm{Hom}(c, x)$ and $k \in \mathrm{Hom}(c, y)$. Then $g_* f_*(h) = gfh = h$ and $f_* g_*(k) = fgk = k$, meaning f_* has a two-sided inverse, making it bijective

Conversely, if we assume f_* is bijective, and want to show $f : x \to y$ is an isomorphism. We need to find an inverse homomrphism such that $fg = 1_Y$, $gf = 1_X$. That would be the same as saying that we need to find an element $g \in \mathrm{Hom}(y, x)$ such that for $f_* : \mathrm{Hom}(y, x) \to \mathrm{Hom}(y, y)$, $f_*(g) = 1_Y$. Since f_* for any c, take $c = y$. Then we already have that $f_*(g) = fg = 1_Y$. To show that $gf = 1_X$, take $c = x$, and so $f_* : \mathrm{Hom}(x, x) \to \mathrm{Hom}(x, y)$ is bijective, and so $f_*(g) = gf = 1_X$. Thus, f (and g) are isomrophisms.

To prove the third implication, notice that we proved this equivalence for *any* categories, which thus *includes* the category C^{op}, meaning $f^{op} : y \to x$ in C^{op} is an isomorphism if and only if

$$f_*^{op} : \mathrm{Hom}_{C^{op}}(c, y) \to \mathrm{Hom}_{C^{op}}(c, x) \text{ is an isomrophism for all } c \in C^{op}$$

Thus, we can take the opposite of C^{op}, get C, which will make all pre-composition into post-composition, and so we actually automatically get that this implies e as well!

This lemma is actually very enlightetning: It shows how characterizing a isomorphism in *any* category by the isomorphisms between hom-sets in **Set**! This is called being *representable*, and we will discuss it in ref:HERE.

The idea of duality extends too to other categorical definition. For example, injectivity and surjectivty are sort of "dual" to each other, they each fulfill similar, but opposing roles. We can make that intuition rigorous!

Definition 30.1.3: Monomorphism And Epimorphism

Let $f : x \to y$ be a morphism in any category Then

1. f is a **monomorphism** if for two morphisms $h, k : w \rightrightarrows x$, $fh = fk \Rightarrow h = k$. We sometimes call f *monic* or *mono*, and can be denoted $f : x \rightarrowtail y$
2. f is a **epimorphism** if for two morphisms $h, k : w \rightrightarrows x$, $hf = kf \Rightarrow h = k$. We sometimes call f *epic* or *epi*, and can be denoted $f : x \twoheadrightarrow y$

Note that the monomprhisms/epimorphisms of C are the epimorphisms/monomprhisms of C^{op}! Furthermore, we can apply 30.1.2 and get that $f : x \to y$ is a monomrophism if and only if $f_* : \mathrm{Hom}(c, x) \to \mathrm{Hom}(c, y)$ is injective for all $c \in C$, and f is an epimorphisms if and only if $f^* : \mathrm{Hom}(y, c) \to \mathrm{Hom}(x, c)$ defined an injection (in a locally small category). In general, monomrophisms and epimorphisms match with injective and surjective when in a small category, but in general they need not be related

Example 30.1.2: : monomrophism, epimorphisms $\not\Rightarrow$ injective, surjective
The $\varphi : \mathbb{Z} \to \mathbb{Q}$ being monic/epic is of particularly intersting, since this is a ring-homomorphism, but there does not exist a non-trivial $\varphi : \mathbb{Q} \to \mathbb{Z}$ ring-homomorphism, showing how being *epic* here did not imply surjectivity.

30.2 Functoriality

> "Every sufficiently good analogy is yearning to become a functor" - John Baez, *Quantum Quanderies: A Category-Theoretic Perspective*

Now that we are somewhat familiar with the idea of Categories, we now add the next basic tool that always needs to be introduced when introducing new mathematical objects: how do categories relate to each other? What are the morphisms between categories?

Definition 30.2.1: Functor

A **Functor** $F : C \to D$ between categories C and D consists of

- An object $Fc \in D$ for each object $c \in C$
- A morphism $Ff : Fc \to Fc' \quad \in D$ for each $f : c \to c' \quad \in C$, so that the domain and codomain of Ff are, respectively, equal to F applied to the domain or comdain of f

where

1. For any composable pair f, g in C, $Fg \cdot Ff = F(g \cdot f)$. This condition is called *functoriality*
2. For each object $c \in C$, $F1_c = 1_{Fc}$.

That is, Functors are mappings of objects and morphisms that prserves all of the structure of a category, namely the domains and codomains, composition, and identities!

Example 30.2.1: Functors

1. There is a functor $F : \mathbf{Set} \to \mathbf{Set}$ which sends every set A to the powerset PA, and a function $f : A \to B$ to the direct-iamge function $Pf : PA \to PB$ that sends $A' \subseteq A$ to $f(A') \subseteq B$. A functor with the same and co-domain is called an *endofunctor*.
2. For each Algebraic examples (**Group**, **Ring**, **Field**, etc. we'll stick to **Group**) there exists a functor $F : \mathbf{Group} \to \mathbf{Set}$ that sends every *group* to it's underlying set, and every *group-homomorphism* to it's underlying function! In general, if you're forgetting some structure, you are a *forgetful functor*. Anotehr example would be form $\mathbf{Mod}_R \to \mathbf{Ab}$ and $\mathbf{Ring} \to \mathbf{Ab}$, where we forget the ring-action and 2nd binary operation respectively. Note that these functions are not necessarily inclusions (an abelian group might be expanded into multiple different types of rings), but there do exist thsoe which are inclusive, like form $\mathbf{Ab} \hookrightarrow \mathbf{Group}$. Note that this action can be considered a *left action*. We will soon see how we can define *right actions* in terms of functors.
3. there is also the opposite of a forgetful functor, usually called a *free functor*. There is a functor $F : \mathbf{Set} \to \mathbf{Group}$ which sends a set X to the *free group* on X!
4. An example outisde of Algebra, if you have taken some algebraic topology, you have encountered the fundametnal group, and that it related topological spaces and continuous function with groups and homomorphisms. In fact, there is a functo $\pi_1 : \mathbf{Top}_* \to \mathbf{Group}$ which does exactly that!

5. (To check) There is a functor $F : \mathbf{Vect}_K \to \mathbf{Mat}_K$ sending every linear transformation to it's matrix representation, and sending every Vector-space to the number representin it's dimension.

6. The chain rule expresses the functoriality of the derivative! Let $\mathbf{Euclid}_*$ denote the category whose objects are pointed finite-dimensional Euclidean spaces $(\mathbb{R}^n, a)$ [a] and whose morphisms are pointed differentiable function. Thus, we can define

$$D : \mathbf{Euclid}_* \to \mathbf{Mat}_{\mathbb{R}} \qquad D\mathbb{R}^n = n \quad D(f) = \text{ appropriate Jacobian Matrix}$$

   Where D sends the euclidean space to it's dimension, and the diffedifferentiable maps to the appropriate matrix! Given a map in $\mathbf{Euclid}_*$ form $g : \mathbb{R}^m \to \mathbb{R}^k$ carrying $f(a) \in \mathbb{R}^m$ to $g \cdot f(a) \in \mathbb{R}^k$, the functoriality of D asserts that the product of the Jacobian of F at a with the Jacobian of g at $f(a)$ equals the jacobian of $gf(a)$, that is, $D(gf) = D(g) \cdot D(f)$, which is *exactly* the chain rule!

7. We saw how a group G can be seen as a category in itsefl. Let's call it $\mathbf{G}$. We can then define a functor $X : \mathbf{G} \to C$ where X is an objet in C and an endormophism $g_* : X \to X$ for each g such that

   (a) $h_* g_* = (hg)_*$ for all $g, h \in G$

   (b) $e_* = 1_X$ where $e \in G$ is the identity element

   Notice that we just defined an *actions* of G on an *arbitrary* Category C! If $C = \mathbf{Set}$, then we get group-actions and a G-**set**. When $C = \mathbf{Vect}_K$, we get a G-**representation**, and when $C = \mathbf{Top}$, we get a G-**space**. Each of these shows how we can generalize the notion of groups acting!

8. We can make the notion of a commutative diagram rigorous with functors (`https://www.youtube.com/watch?v=U3nzEUEnLKQ`)

[a] recall the derivative is evaluated at a point

Since we introduced the idea of duality, what would be the dual of a functor?

Definition 30.2.2: Contravariant Functor

A **contravariant functor** F form C to D is a functor on $F : C^{op} \to D$. More precicely, a contravariant functor will

- carry objects $Fc \in D$ for each object $c \in C$
- will ary morphisms $Ff : Fc' \to Fc \quad \in D$ for each morphism $f : c \to c' \quad \in C$, so that the domain and codomain ofFf are equa to F applied to the codomain and domain of f, respectively.

such that

1. For any cpomposable pair f, g in C, $Ff \cdot Fg = F(g \cdot f)$
2. For each $c \in C$, $F(1_c) = 1_{Fc}$

The key difference is that we know that $F(fg) = Fg \cdot Ff$, or it varies *contrary* to how a functor varies. For this reason, we also call functors *covariant functors*.

Example 30.2.2: Covariant Functor

1. The functor $(__)^* : \mathbf{Vect}_k^{\mathrm{op}} \to \mathbf{Vect}_k$ carries a vectors space V to its dual $\mathrm{Hom}_k(V, k)$ and $f : V \to W$ to $f^* : W^* \to V^*$, which is the map of pre-compositions

$$(\varphi : W \to k) \mapsto (f \circ \varphi : V \to W \to k)$$

2. The same way $X : \mathbf{G} \to C$ define a left-action, $X : \mathbf{G}^{\mathrm{op}} \to C$ defines a *right action*.

3. The contravariant functor $P : \mathbf{Set}^{\mathrm{op}} \to \mathbf{Set}$ Sends A to $P(A)$ (the powerset of A) and and a function $f : A \to B$ to $f^{-1} : P(B) \to P(A)$ mapping $B' \subseteq B$ to $f^{-1}(B') \subseteq A$.

4. As a special case of the lsat example, let X be a topological space and $\mathcal{O}(X)$ represent all open sets of X. Then The functor $\mathcal{O} : \mathrm{Top}^{\mathrm{op}} \to \mathbf{Poset}$ that takes X to $\mathcal{O}(X)$ and a continuous map $f : X \to Y$ to $f^{-1} : \mathcal{O}(Y) \to \mathcal{O}(X)$ that takes the open set $U \subseteq Y$ to it's pre-image. In fact, we may say that f is continuous if the functor $\mathcal{O}$ exists. Similarly, we can define a functor $\mathcal{C}$ for closed subsets.

5. The contravariant functor $\mathrm{Spec} : \mathbf{CRing} \to \mathbf{Top}$ sends a commutative ring R to $\mathrm{Spec}(R)$ and $\varphi : R \to S$ to $\varphi^{-1} : \mathrm{Spec}(S) \to \mathrm{Spec}(R)$ where $\varphi^{-1}(P)$ is the pre-iamge of P.

6. (Talk about presheaves?)

With the notion of functor properly defined, we present *the* property which will make them interesting as a tool for comparing Categories

Lemma 30.2.3: Functors Preserve Isomorphissm

Functors Preserves Isomorphisms

Proof **:**
Consider a functor $F : C \to D$ and an isomorphism $f : x \to y$ in C with inverse $g : y \to x$. Applying the functoriality axioms, we get that

$$F(g)F(f) = F(gf) = F(1_X) = 1_{Fx}$$

meaning $Fg : Fy \to Fx$ is the left inverse to $Ff : Fx \to Fy$, and similarly if we swap f and g [a], showing that Fg is also a right inverse, and hence Ff is an isomorphism

[a] which we can easily do if we argue by duality

It sould be checked that functors do not preserve monomrophisms and epimorphisms (think of Spec), but using an argument similar to the one above, we can show a functor preserve split mono/epimorphisms.

We next introduce a very important general functor that appears a lot, and is often used to "represent" objects. Remember that in Category theory, the *relation* between objects is the key thing that *defines*

what an object is. The following Functor makes that rigorous!

Definition 30.2.4: Functor Represented By c

If C is locally small, then for any object $c \in C$, we may define a pari of covariant and contravariant **functors represented by** c:

$$C \xrightarrow{\mathrm{Hom}(c,_)} \mathbf{Set} \qquad C^{op} \xrightarrow{\mathrm{Hom}(_,c)} \mathbf{Set}$$

$$\begin{tikzcd} x \arrow[r, mapsto] \arrow[d, "f"'] \arrow[dr, phantom, "\longmapsto"] & \mathrm{Hom}(c,x) \arrow[d, "f_*"] \\ y \arrow[r, mapsto] & \mathrm{Hom}(c,y) \end{tikzcd} \qquad \begin{tikzcd} x \arrow[r, mapsto] \arrow[d, "f"'] \arrow[dr, phantom, "\longmapsto"] & \mathrm{Hom}(x,c) \\ y \arrow[r, mapsto] & \mathrm{Hom}(y,c) \arrow[u, "f^*"'] \end{tikzcd}$$

where f_* and f^* is the same functions presented in lemma 30.1.2 (pre- and post-composition)

These functors will play a major role, since as we've seen, being an isomorphism was intimitely related to whether a the representation was bijective on hom-sets. We wil return to this observation in chapter ref:HERE (next chapter).

The two functors in definition 30.2.4 can be combined. First, if **C** and **D** are functor, then the *product* of functors is a new category (C, D) where objects are tulpes (c, d) with $c \in C$ and $d \in D$, and morhpisms are pairs $(f : c \to c', g : d \to d')$, $f \in C$ and $g \in C$ where composition and the identity is component wise.

Definition 30.2.5: Two-Sided Represented Functor

Let C be a locally small category. Then The *two-sided representation functor*:

$$\mathrm{Hom}(__, __) : C^{\mathrm{op}} \times C \to \mathbf{Set}$$

maps (c, c') to $\mathrm{Hom}(c, c')$. The pair of morhisms (f, h) maps to (f^*, h_*) where:

$$\mathrm{Hom}(c, d) \xrightarrow{(f^*, h_*)} \mathrm{Hom}(c', d') \qquad g \mapsto hgf$$

So, we can now turn our attention to trying to ask the question we did last section when we defined **Cat**: what is the isomorphism between Categories? If we take the definition 30.0.4, then we would require that for a pair of functors $F : C \to D$ and $G : D \to C$ that GF and FG equal the identity functor on C and D. Sometimes, this is quite desirable: For example, if E/F is a finite galois extension and $\mathbf{Feild}_F^E$ represents all feels $F \subseteq k \subseteq E$, and $\mathcal{O}_G$ represents all subgruops $H \subseteq G = \mathrm{Aut}_F(E)$ whose objects are G/H (for a subgroup H) and morphisms are G-equivarient maps $G/H \to G/K$ (functions that commute with the G-action)

However, in many cases, this relation is a little too stringent to get comparative results

Example 30.2.3: Almost Category Isomorphism

p.20-21 – category theory in context with $\mathbf{Set}^{\partial}$ and $\mathbf{Set}_*$. The problem is essentially we don't want to define FG and GF to be the identity, but that $1_X(X)$ and $FG(X)$ to be equal up to isomorphism (and similarly for GF)

To solve this, we will loosen our condition of isomorphism between categories, and define a slightly weaker but more useful definition of "isomorphism" for categories in the next section

30.3 Natural Transformations

As a motivation for the concept of natural transformations, we revisit a concrete example of this notion that was already presented in the theory of dual spaces presented in section 13.3:

Example 30.3.1: Dual Space reminder

Let V be a finite-dimensional vector space over $\mathbb{F}$. We know that

$$V^* = \mathrm{Hom}(V, \mathbb{F})$$

is the *linear dual* of V. Recall that V and V^* have the same dimension. If $\beta = \{e_1, ..., e_n\}$ is a basis for V, and $\beta^* = \{e_1^*, ..., e_n^*\}$ is a set form V^* such that

$$e_j^*(e_i) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Then β^* forms a basis for V^*, and morover, the map $e_i \mapsto e_i^*$ is bijective and linear, meaning V and V^* are isomorphic ($V \cong V^*$).

We can further consider $V^{**} = \mathrm{Hom}(\mathrm{Hom}(V, \mathbb{F}), \mathbb{F})$ being the the set of all linear maps mapping all elements form $\mathrm{Hom}(V, \mathbb{F})$ to $\mathbb{F}$. We can in similar fashion make an isomorphism $V^* \cong V^{**}$ letting us conclude that $V \cong V^{**}$.

This whole construction started with a choice of basis form V, which need not be unique, and so it's not a "property" of V. However, we can still define an isomorphism between V and V^{**} *without* any reference to a basis. For any $v \in V$, we can define a map

$$(f \mapsto f(v)) : V^* \to \mathbb{F} \qquad (f \mapsto f(v))(g) = g(v)$$

that is, every functional $g \in V^*$ evaluates the value and get's us $g(v)$. This map is sometimes called the evaluation function. Let's switch up the notation for simplicity:

$$ev_v : V^* \to \mathbb{F} \qquad ev_v(g) = g(v)$$

where ev_v stands for evalutate at v. Now, defining map $V \to \mathrm{Hom}(V^*, \mathbb{F})$, $v \mapsto ev_v$ is an *isomorphism*. Bijectivity is left to check, for linearity, let φ be the function in question, so that for $k \in \mathbb{F}$ and $v, w \in V$:

$$ev(k \cdot v) = ev(kv) = ev_{kv} = k \cdot ev_v = kev(v)$$

$$ev(v + w) = ev_{v+w} = ev_v + ev_w = ev(v) + ev(w)$$

where kv represents the vector we get when applying the ring action on v. Note that we essentially used the lineary of the object we are mapping to in order to prove linearity.

> Thus, we defined this isomorphism with *no* reference to a basis on V. We can't do the same trick for $V \to V^*$, since if we try to do $v \mapsto f(v)$, it is ambiguous which f we are talking about.

What is happening here functorially is that we have a category $\mathbf{Vect}_{\mathbb{F}}$ we have a functor ev form it to the category $\mathbf{Vect}_{\mathbb{F}}^{**}$. What's *key* is that if know get any linear transformation $\varphi : V \to W$, and we ask ourseleves if there is an equivalent functor between W and W^{**}, the answer will be that we *can* do an equivalent construction, *irrelevant* to where the φ might map us! In this way, the functor does not rely on *anything* exccept what is *naturally* present in every vector-space! To speak more generally, we properly define the idea of natural functor:

Definition 30.3.1: Natural Transformation

Given categoreis C and D and functors $F, G : C \to D$, a **natural transformation** $\alpha : F \Rightarrow G$ consists of an arrow $\alpha_c : Fc \to Gc$ in D for each object $c \in C$, the collection of which define the **components** of the natural transformation, so that, for any morphism $f : c \to c'$ in C, the following diagram

$$\begin{tikzcd} Fc \arrow[r, "\alpha_c"] \arrow[d, "Ff"] & Gc \arrow[d, "Gf"] \\ Fc' \arrow[r, "\alpha_{c'}"] & Gc' \end{tikzcd}$$

commutes: $\alpha_c \cdot Gf = Ff \cdot \alpha_{c'}$. Or more generally, If $\mathrm{Hom}(C, D)$ is the category with *functors* form C to D as objects, then the morphisms in this category are *natural transformations*
A **natural isomorphism** is a natural transformation $\alpha : F \Rightarrow G$ in which every component α_c is an isomorphism. In this case, the natural isomorphism may be depicted as $\alpha : F \cong G$

Since are working in the category of functors $\mathrm{Hom}(C, D)$, it might be useful to also get a way to "draw" the functors of this category, which is usually done by:

$$\begin{tikzcd} C \arrow[r, bend left, "F"] \arrow[r, bend right, "G"'] \arrow[r, phantom, "\Downarrow \alpha"] & D \end{tikzcd}$$

So, in the example of $V \cong V^{**}$, we have the category $\mathrm{Hom}(\mathbf{Vect}_{\mathbb{F}}, \mathbf{Vect}_{\mathbb{F}})$. The map $ev : V \to V^{**}$ will define a *component*, with the *identity endofunctor* $\mathbf{Id}$ on V ($\mathbf{Id}V$) and the *double dual functor* $(_)^{**}$ ($(_)^{**}(V)$). That is, the diagram

$$\begin{tikzcd} V \arrow[r, "ev"] \arrow[d, "\varphi"] & V^{**} \arrow[d, "\varphi^{**}"] \\ W \arrow[r, "ev"] & W^{**} \end{tikzcd}$$

Decompoes itself to

$$\begin{tikzcd}
\mathbf{Id}V \arrow[r, "ev"] \arrow[d, "\varphi"] & (_)^{**}V \arrow[d, "\varphi^{**}"] \\
\mathbf{Id}W \arrow[r, "ev"] & (_)^{**}W
\end{tikzcd}$$

Further, since ev is an isomorphism, this is a *natural isomorphism*! If we think of this as an isomorphism in the $\mathrm{Hom}(\mathbf{Vect}_{\mathbb{F}}, \mathbf{Vect}_{\mathbb{F}})$ category, we can write this as

$$\begin{tikzcd}
\mathbf{Vect}_{\mathbb{F}} \arrow[rr, bend left, "\mathbf{Id}"] \arrow[rr, bend right, "(_)^{**}"'] & \Downarrow \alpha & \mathbf{Vect}_{\mathbb{F}}
\end{tikzcd}$$

The key take away here is that in natural transformations, we do not make choices, they arise natrually given the definition (one author liked to say "god-given"). We sometimes also say *cannonical* transformation too. So for example, $A \times B \to A$ $(a, b) \mapsto a$ can be shown to represent a natural transformation, HOwever, a function defined as "choose an element $a_0 \in A$, and then send every thing to a_0" wouldn't have a cannonical, since we made a choice.

Example 30.3.2: Natural Tranformations

1. Recall that a group homomrophism is a funciton $\varphi : G \to H$ which preserves the group action. If we take G, and H represent the underlying set, and $(G, \cdot_G)$, $(H, \cdot_H)$ represent the set with the underlying operation, then we can what's the most *natural* requirement for the operation to be preserved? Note that if we drop the operation preserving part of φ, we can think of the inputs of the binary operation as ^{a}

$$(\varphi \times \varphi) : G \times G \to H \times H \qquad (\varphi \times \varphi)(a, b) = (\varphi(a), \varphi(b))$$

   Then, we can combine this with the operations on the group to get the following diagram:

$$\begin{tikzcd}
G \times G \arrow[r, "\varphi \times \varphi"] \arrow[d, "\cdot_G"] & H \times H \arrow[d, "\cdot_H"] \\
G \arrow[r, "\varphi"] & H
\end{tikzcd}$$

   Thus, from a categorical perspective, the best way to define an operation to be preserving is if this diagram *commutes*, meaning:

$$\begin{tikzcd}
(a, b) \arrow[d, mapsto] \arrow[dr, dotted, no head] & \\
a \cdot b \arrow[r, mapsto] & \varphi(a \cdot b)
\end{tikzcd}
\qquad
\begin{tikzcd}
(a, b) \arrow[r, mapsto] \arrow[d, dotted, no head] & (\varphi(a), \varphi(b)) \arrow[d, mapsto] \\
 & \varphi(a) \cdot \varphi(b)
\end{tikzcd}$$

   which in it's familiar form would be $\varphi(a \cdot b) = \varphi(a)\varphi(b)$

2. Riesz represenation, the discrete category, G-sets, and the equivalence of A^B and $A \times B$ are all interseting examples I'll go over at some point.

3. Recall when we defined the *natural projection* (ref:HERE in preliminary chapter, category theory)? The natural projection is a natural transformation! The category is $\mathrm{Hom}((G,N),\mathbf{Group})$. The (G,N) category is the category consisting of all pairs (G,N) of groups and normal subgroups, and whose morphisms are "commutative squres", that is, $(N_1,G_1)\to(N_2,G_2)$ such that there is a homomorphism $n:N_1\to N_2$, $g:G_1\to G_2$ such that $N_1\hookrightarrow G_1$ and $N_2\hookrightarrow G_2$ remain. The point of this is to make maps between normal subgroups well defined. Finally, we have that

$$\begin{tikzcd}
(G_1,N_1) \arrow[r, "\pi"] \arrow[d, "\varphi"] & G_1' \arrow[d, "\psi"] \\
(G_2,N_2) \arrow[r, "\pi"] & G_2'
\end{tikzcd}$$

[a]The reason for this construction, as we shall soon see, relies on the *universla proprty of produts*. For now, since it's something all readers would know at this point, we leave it as intuitive

30.4 Universal Property

Finally, we will present a central concept in Category theory, one which we applied when looking at the Isomorphism Theorems for Groups and Rings. The more mathematics you do, the more you will define things in terms of their universal property.

One thing that might seem totally unintuitive at first is that there are categories with *initial* and *final* objects! That is, objects from which every object has a homomorphism from, and to which every object has a homomorphism to. In particular, there will be *unique* homomorphisms from initial/final objects of categories to/from any other objects in a category respectively. Because of the uniqueness of these homomorphisms, and because they go to *every* object, then we can actually characterize all homomorphisms in a cateogry given the inital homomorphism! That is, if A is an initial object, and X is some other object in the category, then

$$A \xrightarrow{\alpha} X$$

will be a unique morphism from A to X. Let's label this morphism α. Then let's say that $X\xrightarrow{\varphi}Y$ is a morphism. Let's label if φ. Since we have that $A\xrightarrow{\psi}Y$ exists and is unique (let's call it ψ), then

$$\begin{tikzcd}
A \arrow[d, "\alpha"] \arrow[dr, "\psi"] & \\
X \arrow[r, "\varphi"] & Y
\end{tikzcd}$$

must commute, that is, the arrow from $A\to Y$ must be $\psi=\alpha\circ\varphi$. That is, we characterized a morphism from two arbitrary objects using *only* the morphisms from the initial object!

A lot of the time, the initial objects of category will be not so interesting (ex. that of groups). However, there will be a "neigbhouring" category which will pocess much more interesting initial objects. The construction of this category is actually very similar for most universal properties, and so is given a name: *comma category*. The objects in this category will have the objects of two categories and a morphism between them, and natural transformation as morphisms. Think of the projection was from objects (G_1,N_1) in the previous section. That was actually a comma category!

Notice also how π didn't change both time. We will show how (G, N) are the *inital* objects of this category, and with this universal property, we will be able to characterize *all* homomorphisms from groups using them!

Before we get to the technicalities of the definition, here are a few more examples:

Example 30.4.1: Universal Properties

1. Let $\mathbb{1}$ denote a set with one element. Then this set with 1 elemnt has the following property

$$\text{for all sets } X\text{, there exists a } \textit{unique} \text{ map } X \text{ to } 1$$

2. Let R be a ring with identity $1 \neq 0$. Then

$$\text{For all rings } R\text{, there exists a } \textit{unique}$$

homomorphism $\mathbb{Z} \to R$ The homomorphism, simply take

$$\varphi(n) = \begin{cases} 1 + \cdots + 1 & \text{if } n > 0 \\ 0 & \text{if } n = 0 \\ -\varphi(-n) & \text{if } n < 0 \end{cases}$$

To prove uniqueness, let ψ be another function. We know in definition 9.1.2 that $\psi(1) = 1$. Then for $n > 0$, we have:

$$\psi(n) = \psi(1 + \cdots + 1) = \psi(1) + \cdots + \psi(1) = 1 + \cdots 1 = \varphi(n)$$

for all $n > 0$. If $n = 0$, $\varphi(0) = 0 = \psi(0)$. Finally, for the negative, $\psi(n) = -\psi(-n) = -\varphi(-n) = \varphi(n)$ whenever $n < 0$. Therefore, $\psi = \varphi$.

Crucially here, if we have $\varphi : A \to R$, then $A \cong \mathbb{Z}$

Proof. here □

3. Let V be a vector space with a basis β.

4. A non-algebra example would be with linking topology and discrete topology.

Many times, there is a universal property from the category of sets to bring from them the most common notions into the other categores (eg. equalizesrs are used to define sub-groups, co-equalizers, or equivalence relation, to define quotient groups, or bringing the limit of tow disconnected objects into other algebraic objects (i.e., defining the product in other concrete categoreis))

Definition 30.4.1: Informal Definition Of Universal Property

1. Let C and D be categories, and F be a functor from C to D, and let X be an object in D. A *universal arrow* from X to F is a pair $(U(X), \iota))$ where $U(X)$ is an object in C, $\iota : X \to FU(X)$ is a morphism in D satisfying the following property: for any object A in C, if φ is any morphism from X to FA in D, then there exists a unique morphism $\Phi : U(X) \to A$ in C such that $F(\Phi)\iota = \varphi$, i.e., the following diagram commutes

$$\begin{tikzcd} X \arrow[r, "\iota"] \arrow[dr, "\varphi"'] & FU(X) \arrow[d, "F(\Phi)"] \\ & FA \end{tikzcd}$$

2. Let C be a category and F be a functor from C to the category **Set**. A *universal element* of the functor F is a pair (U, ι) where U is an object in C and ι is an element of the set FU satisfying the following property: for any ojbect A in C and any element g in the set FA there is a unique morphism $\varphi : U_> A$ in C such that $F(\varphi)(\iota) = g$

More words here

Example 30.4.2: More Universal Properties

1. For concerete categories (I think they all have a faithful functor to **Set** but if they don't, concrete concrete categories with this functor), there can exist a universal property of free objects
2. universal property of quotients
3. universal property of equalizer (On an intuitive level given by alg geo, equalizer is the idea of two graphs of functions intersecting)
4. universal property of lifts
5. universal properties of products/co-products

30.5 Adjoint

here

For free objects, adjoint with some appropriate forgetful functor.

Adjoint of Hom and $\otimes$.

linear algebra adjoints?

30.6 A Second Look at Category theory

The concept of Categorization.

As an example, let's "categorify" the concept of groups:

Definition 30.6.1: Group Category

Let C be a cateogry with ($finite$) product and with final object 1. A *group object* in C consits of an object G of C and of morphisms

$$m : G \times G \to G,\ e : 1 \to G,\ i : G \to G$$

in C such that the diagram

$$\begin{tikzcd}
(G \times G) \times G \arrow[r, "m \times \mathrm{id}_G"] \arrow[d, "\cong"] & G \times G \arrow[dr, "m"] & \\
 & & G \\
G \times (G \times G) \arrow[r, "\mathrm{id}_G \times m"] & G \times G \arrow[ur, "m"] &
\end{tikzcd}$$

$$\begin{tikzcd}
1 \times G \arrow[r, "e \times \mathrm{id}_G"] \arrow[dr, "\cong"] & G \times G \arrow[d, "m"] \\
 & G
\end{tikzcd}
\qquad
\begin{tikzcd}
G \times 1 \arrow[r, "\mathrm{id}_G \times e"] \arrow[dr, "\cong"] & G \times G \arrow[d, "m"] \\
 & G
\end{tikzcd}$$

$$\begin{tikzcd}
G \arrow[r, "\Delta"] \arrow[d] & G \times G \arrow[r, "\mathrm{id}_G \times i"] & G \times G \arrow[d, "m"] \\
1 \arrow[rr, "e"] & & G
\end{tikzcd}
\qquad
\begin{tikzcd}
G \arrow[r, "\Delta"] \arrow[d] & G \times G \arrow[r, "i \times \mathrm{id}_G"] & G \times G \arrow[d, "m"] \\
1 \arrow[rr, "e"] & & G
\end{tikzcd}$$

commute

The $\Delta = \mathrm{id}_G \times \mathrm{id}_G$ is the "diagonal" morphism $G \to G \times G$ induced by the univeral property for products and the identity maps $G \to G$.

(example in topology, differential geometry, and in schemes. In topology, leads to topological groups, in diferential geometry, leads to lee groups, and in scheme's, lead to generlization of Galois theory, where instead of galois extensions corresponding to galois groups, we get inseparable extensions corresponding to algebraic group schemes)

Some miscellaneous insights

I have yet a place to put there insights so I'll put theme here

1. A monoid is like a group but you generalize to the fact that you can't "inverse" actions necessarily. A category is like a monoid, except you now have a "partial" function, i.e., you can't always compose any object to any object, to compose, you need to make sure the domain and co-domain are right. In this way, a category is a generalization of a monoid!!

2. **Set** can be thought as a 0-cateogy: a thing with only objects. Then a category has obects

and morphisms, and the level above that 2-category is a thing with object, moprhisms, and morphisms between morphisms. In this way, **Cat** can be thought of as a 2-category

- Though instead of saying a set is a 0-category, it might be more accurate to say a *topoi* is a 0-category, being a natural generalization of sets.

30.7 Commutative Diagrams

If there are arrows between many objects, we tend to draw them in diagrams:

$$\begin{tikzcd}
h' \arrow[drr] & & \\
& h \arrow[r, "\beta"] & l \\
f \arrow[uu] \arrow[ur, "\alpha"] \arrow[r, "\alpha'"] \arrow[drr] & g \arrow[ur, "\beta'"] & \\
& & g' \arrow[uu]
\end{tikzcd}$$

where each arrow reprents a morphism, and following the path in a diagram represents composition of morphisms. Diagrams like this usually have an extra property: any path with the same start and end point will lead to the same result. For example, in the above diagram, $\beta\alpha = \beta'\alpha'$. When a has this property, it is called a *commutative diagram*. Examples of commutative diagrams show up all the time (consider all the universal properties we've encountered). The name "commutative" most likely comes from the following diagram where $\cdot y$ and $\cdot x$ represent multiplying the group element 1 by some other element[1]

$$\begin{tikzcd}
1 \arrow[r, "\cdot x"] \arrow[d, "\cdot y"] & x \arrow[d, "\cdot y"] \\
y \arrow[r, "\cdot y"] & xy = yx
\end{tikzcd}$$

In general case where there is a square diagram, we would say that $f \circ \varphi = \varphi \circ g$, so φ "commuted". Even though f and g are different, they will be thought of as essentially the same, meaning commutative here would be meaningful. Though this is just one diagram, it is of such important (and proably most used) that the name commutative diagram stuck. Note too that though we have been dealing with squares, it need not be limited to squares: if we have longer length paths that have the same start and end point, the composition of those morphisms must be equal.

(I just had another intuition on why this particular commutative diagram is very natural: We know in lin alg that ABA^{-1} for two invertible matrices represents change of basis, so a different way of representing the same thing, and that ABC for appropriate A and C make B equivalent to it (no intuition for this yet). In diff geo, $\varphi \circ f \circ \psi^{-1} = g$ is a coordinate representation of a map. In a sense, we are saying that these types of maps le us represent our original map in a different way. However,

[1] We will soon see how groups very naturally appear in category theory, so this with morphisms being multiplying is not too much of an abuse of notation

in more general contexts not all maps are invertible, so we instead write $\varphi \circ f = g \circ \psi$. This also has a new intuition added to it: since these maps are equal, we in a sense have tow "identical" ways of getting to the same co-domain. Many times, f and g wil be in one category, φ and ψ in another. It would be obvious if $\varphi \circ f = \psi \circ f$ since then $\varphi = \psi$, but what is cooler is if we can say "do f, then φ" and ask if the categor captures the structure to the point where we can say "actually, convert through some ψ", then find a g that still get's us to the same result. In a sense, g "represents" f through the maps φ and ψ (think of a bundle homomorphism where we require $\pi \circ F = f \circ \pi'$))

Part IX

Appendix

In the appendix, certain section will take liberties on how much the reader know's. In particular, the reader is expected to be comfortable with (some analysis or topology in certain appendicies, like category theory and geometric group theory).

A

Axiom of Choice

Look into this explanation of truth in mathematics, looks good!:

In mathematics, we are constantly making choices. We can make algorithmic choices (for example, if you're given the number n, choose the number $n+1$), or we make arbitrary choices (choose some natrual number n). The fact that we can make an arbitrary choice is uncontroversial in the axioms mathematicians used, that is ZF (short for Zermelo-Frankel axioms). So too is making a finite number of choices. These two "abilities" were also uncontroversial in other foundational formulations of mathematics, not just ZF. It might seem very natural then to be able to make an infinite, or more generally an arbitrary (including uncountably infinite or other cardinalities) amount of choices. However, making an arbitrary amount of choices turns out to have some very strange consequences. The German mathematician Ernst Zermelo axiomatically used the fact that you can make an arbitrary number of choices (which we will soon define what this means mathematically) to prove a very controversial conjecture called the *well-ordering Theorem* (which we shall cover later in this appendix), sparking a debate on whether mathematics should allow for arbitrary choice. One famous "counter-example" developed to "prove" that we shouldn't allow arbitray choice is a proof by Banach and Tarski that you can take a sphere, partition it into 6 subsets (which uses the axiom of choice), rotate them in certain ways, and recombine them to form 2 identical spheres! Importantly, we did *not* add any more points, yet we have somehow "doubled" the surface area of the points we've chosen [1]! This is called the *Banach-Tarski Paradox.*

Though this result is strange, as time went on, more and more mathematicians found that making an arbitrary number of choices produces some "natural" results. For example, the fact that every vector-space has a basis comes down to the ability of making an arbitrary number of choices. Another example (HERE: from that stackexchagne, I think it's the one saying "if $|A| = |B|$, then we need the AC to say there is an injective function from A to B"). In 1963, a mathematician named Paul Cohen proved that the making an arbitrary number of choices is independent of the standard system

[1] For those who have taken measure theory, the chosen subsets turn out to be *unmeasurable* sets, and so the concept of surface area does not apply to these sets which allows us to produce this strange result

of axioms ZF, meaning if we assume the we can do an arbitrary number of choices, it will not cause any more contradiction and ZF would already cause [2], and if we assume we can't make an arbitrary number of choices, it will *also* not cause any more contradiction than ZF would already cause. Through this result, most mathematicians now accept the axiom of choice as part of the standard model we use, which is represented as ZFC (ZF + the axiom of **C**hoice).

In Algebra, the axiom of choice most often appears in a special form called *Zorn's Lemma*. The aim of this chapter is to be able show that Zorn's lemma is equivalent to the axiom of choice.

A.1 Build-up

Definition A.1.1: Cartesian Product and Choice Function

Let $\{A_i\}_{i\in I}$ be any family of sets. Then the *Cartesian product* of this family, written $\prod_{i\in I} A_i$, is the set of all function $x : I \to \bigcup_{i\in I} A_i$, where $x(i) \in A_i$ for all $i \in I$ The function x is called a *choice function*.

Though we've defined what the cartesian product is, it does not mean that the function x exist! Naturally, if $I = \emptyset$ or if $A_i = \emptyset$ for some $i \in I$ (since then $\prod_{i\in I} A_i = \emptyset$), then this function exists. If $I \neq \emptyset$ or $A_i \neq \emptyset$, then it turns out the existence of this function is independent of the ZF axioms! Thus, the axiom of choice was formulated:

Axiom A.1.2: Axiom of Choice

The cartesian product of any non-empty family of non-empty sets is a non-empty set. In particular, if $\{A_i\}_{i\in I}$ is a family of sets such that $I \neq \emptyset$ and $A_i \neq \emptyset$ for each $i \in I$, then there exists at least one choice function for $\{A_i\}_{i\in I}$

Since the cartesian product exists (axiomatically), we will make some shorthand notation for it similar to that defined in chapter 5. If $\{A_i\}_{i\in I}$ is a family of sets, let $A^I = \prod_{i\in I} A_i$ where $A_i = A$ for each $i \in I$. If $|I|$ is finite, and so $|I| = n$ for some $n \in \mathbb{N}$, then we write A^n instead. If $|I|$ is countable, we sometimes use the A^∞ or $A^\mathbb{N}$ notation. the elements of A^n can be represented as n-tuples, that is

$$A^n = \{(a_1, a_2, ..., a_n \mid a_i \in A\}$$

and the elements of $A^\mathbb{N}$ can be represents as sequences

$$A^\mathbb{N} = \{(a_1, a_2, ..., a_n, ...) \mid a_i \in A\}$$

Example A.1.1: Cartesian Products

1. An example we've already encountered is $\mathbb{R}^n$ and $\mathbb{R}^\mathbb{N}$. This caretsian product is called *Euclidan Space*
2. Let $A = \{0, 1\}$. Then $A^\mathbb{N}$ is the set of all sequences with either 0 or 1 in each component. If we change notation a bit and let $A = \mathbb{Z}/2\mathbb{Z}$, and define addition point-wise, then this would

[2] The careful wording here is because it has *not* been proven that the axioms of ZF don't contradict each other – this remains an open problem in mathematics

be a group

(words here)

Definition A.1.3: Partially Ordered Sets

1. Let P be a non-mepty set. A *partial order* on P is a relation $\leq$ ($\leq \subseteq P \times P$) on P such that

   (a) **(reflective)** $x \leq x$ for all $x \in A$

   (b) **(antisymmetric)** if $x \leq y$ and $y \leq x$, then $x = y$

   (c) **(transitive)** if $x \leq y$ and $y \leq z$ then $x \leq z$

   Where we usually say P is a partiallly ordered set under the orering $\leq$ or that A is partially ordered by $\leq$.

   If $\leq$ also satisfies

   (d) **(trichotomy)** If $x, y \in P$, then $x \leq y$ or $y \leq x$,

   then $\leq$ is called a *total* (or linear, simple, complete) ordering on P. If further

   (e) **(Well-Ordering)**For all $\emptyset \neq A \subseteq P$, there exists an element $a \in A$ such that $a \leq x$ for all $x \in A$

   then we say that $\leq$ is a *well-ordering* of P, or that P is well-ordered by $\leq$

2. Let $(P, \leq)$ be a partially ordered set and let $A \subseteq P$. An element $u \in P$ is called an *upper bound* for A if for all $x \in A$, $x \leq u$. Similarly, $l \in P$ is called a *lower bound* if for all $x \in A$, $l \leq x$.

3. Let $(P, \leq)$ be a partially oredered set and let $A \subseteq P$. An element $m \in P$ is called a *maximal element* of P if whenever $m \leq x$ with $x \in P$, then $m = x$. Similarly, an element $m \in P$ is called a *minimal element* if whenever $x \leq m$ for $x \in P$, then $x = m$.

4. Let $(P, \leq)$ be a partially ordered set. Then a subset $C \subseteq P$ is called a *chain* in P if C is totally ordered under the order relation of $\leq$, that is, if $x, y \in C$, then either $x \leq y$ or $y \leq x$.

Note that it's called a *partial* order because becaue not all elements *necessarily* need to to have a relation between them: if $x, y \in P$ and $\leq$ is a parital order on P, then it need not be the case that $x \leq y$ or $y \leq x$. If this were the case, we would call $\leq$ a *total order*.

In the following examples we'll show the difference between being maximal and being an upper bound. Importantly, it is possible that *not all* $x \in P$ have the relation $x \leq m$ (i.e. m is not necessarily the "*largest*" or "*maximum*" element)

Example A.1.2: Partially Ordered Sets

1. Let S be an arbitrary set, and $P(S)$ be it's powerset. Then $(P(S), \subseteq)$ forms a partially ordered set. This set is *only* a partial order (it is not a total order) if $|S| > 1$: for two distinct

elements $x, y \in S$ $\{x\} \not\subseteq \{y\}$ and $\{y\} \not\subseteq \{x\}$. A chain in $(P(S), \subseteq)$ would look like:

$$S_1 \subseteq S_2 \subseteq S_3 \subseteq S_4 \subseteq \cdots$$

If $T \subseteq S$, then then T has an upperbound U:

$$U = \bigcup_{x \in T} \{x\}$$

Furthermore, S is also the maximal element of $P(S)$.

2. Let $S = \{a, b, b', c'\}$ be the set, and $\leq$ be the relation visualized by the graph:

$$\begin{tikzcd}
 & & c' \\
b & & b' \arrow[u] \\
 & a \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

Then both b and c' are maximal elements. S has no upper bound, but it has a as a lower bound.

3. Let S be the collection of *proper subsets* of $\mathbb{N}$ ordered by $\subseteq$. Notice that the chain

$$\{1\} \subseteq \{1, 2\} \{1, 2, 3\} \subseteq \cdots$$

does not have an upper bound. However, S *does* have maximal elements. $(\mathbb{N} - \{n\}) \in S$ for any $n \in \mathbb{N}$ is a maximal element of S

4. Let $S = \mathbb{R}$ with the usual order $\leq$. Notice that $\leq$ is a *total order* in this case. some sets of $\mathbb{R}$ have upper bound (ex. $[0, 1]$ or $\{0, 1, 2, 3\}$), while other's don't $(\mathbb{Z}, (0, \infty), (0, 1))$. $\mathbb{R}$ does not have maximal elements in this order.

When we say we give a partial order to an arbitrary set S, we will assume that this is the partial order on the power set given by $\subseteq$.

Next, we will build-up some facts to bootstrap the result

Definition A.1.4: Finite Character

Let $\mathcal{F}$ be a family of sets. Then $\mathcal{F}$ is said to be a *family of finite character* if for any set S, $S \in \mathcal{F}$ if and only if each finite subset of S is in $\mathcal{F}$.

Lemma A.1.5: Finite Character and Chains

Let $\mathcal{F}$ be a family of finite character and let $\mathcal{B}$ be a chain in $\mathcal{F}$. Then $\cup \mathcal{B} \in \mathcal{F}$

Proof **:**
Since $\mathcal{F}$ is of finite character, if we show that every finite subset of $\mathcal{B}$ is an element of $\mathcal{F}$. Let $\{x_1, x_2, ..., x_n\} = F \subseteq \cup\mathcal{B}$. Then there must exists sts $\mathcal{B}_1, ...\mathcal{B}_n \in \mathcal{B}$ such that $x_i \in \mathcal{B}_i$. Since $\mathcal{B}$ is a chain, there exists a $1 \leq j \leq n$ such that $B_i \subseteq B_j$, for all $i \in \{1, 2, ...n\}$. But then, $F \subseteq B_j \in \mathcal{F}$. Since $\mathcal{F}$ is of finite character, and $\mathcal{B}_j \in \mathcal{F}$, $F \in \mathcal{F}$. Since F was arbitrary, then this holds true for any finite subset of $\cup\mathcal{B}$, and so $\cup\mathcal{B} \in \mathcal{F}$, as we sought to show.

A.2 Axiom of Choice Equivalences

Note that terms like "lemma" or "theorem" for the following names are due to historical precedents rather than actually representing a lemma or theorem, since all of the following will be equivalent to the axiom of choice.

Theorem A.2.1: Turkey's Lemma

Every nomepty family of finite character has a maximal member

Theorem A.2.2: Hausdorff Maximality Principle

Every nonempty partially ordered set contains a maximal chain, i.e., every totally ordered subset is contained in a maximally totaly ordered subset

Maximally totally ordered here means that if it is enlraged in any way, it does not remained totally ordered

Theorem A.2.3: Zorn's Lemma

Every nonemtpy partially ordered set in which each chain has an upper bound has a maximal element

Theorem A.2.4: Well-Ordering Theorem

Every set can be well-orderd, i.e., if S is a set, then there exists some well-ordering $\leq$ on S

A well-ordering on a set S is a total order (i.e. a chain) on S with the additional property that every non-empty subset of S has a lower-bound. For example, given the usual orderings for the following examples, $\mathbb{N}$ is well-ordered (every subset has a lower-bound), while $\mathbb{Z}$ is not (namely, $\mathbb{Z}$ itself doens't have a lower-bound). A trickier example would be $\mathbb{R}$. The set (a, b) for $a, b \in \mathbb{R}$, $a \neq b$, does *not* have a minimal element, and so $\mathbb{R}$ is not well-ordered with the usual ordering.

Theorem A.2.5: Axiom of Choice Equivalences

The following are equivalent

1. The axiom of choice
2. Turkey's Lemma
3. The Hausdroff maximality principle
4. Zorn's Lemma
5. Well-ordering Theorem

Credit to the proof to

Proof **:**
Step (1) $\Rightarrow$ (2) will be the longest step requiring many techincal details. the other steps will come much more naturally. The proof will also be long since we will not be using heavy machinery from a course in set-theory (for ex. we will not use transfinite induction [directly?? are we secretly using it?])

(1) $\Rightarrow$ (2) For the sake of contradiction, let's say Turkey's lemma was false – there exists a non-empty family of sets of finite characetr that has no maximal element. let's say $\mathcal{F}$ was such a set. That means for every $A \in \mathcal{F}$, there exists a B such that $A \subsetneq B$. Let $\mathcal{O}_F = \{E \in \mathcal{F} \mid F \subsetneq E\}$. Then $\{\mathcal{O}_F \mid F \in \mathcal{F}\}$ is a set of sets indexed by $\mathcal{F}$. Thus, by the axiom of choice, there exists a choice function f defined on $\mathcal{F}$ such that $f(F) \in \mathcal{O}_F$ (i.e. $F \subsetneq f(F)$ for all F).

Using f, we'll define a new concept: a subfamily $\mathscr{F} \subseteq \mathcal{F}$ is f-inductive if:

(a) $\emptyset \in \mathscr{F}$
(b) if $A \in \mathscr{F}$ then $f(A) \in \mathscr{F}$
(c) if $\mathcal{C} \subseteq \mathscr{F}$ is a chain, then $\cup\mathcal{C} \in \mathscr{F}$

Notice that $\mathcal{F}$ is itself an f-inductive set since: $\mathcal{F}$ is non-empty (by assumption), so some $A \in \mathcal{F}$ exists, and $\emptyset \subseteq A$ is a finite set, and so $\emptyset \in \mathcal{F}$ since $\mathcal{F}$ has finite character, fulfilling the first condition. The second conditions holds true since the image of f is in $\mathcal{F}$, and the third condition holds by lemma A.1.5. Notice too that if the family $\mathscr{F}_i \subseteq \mathcal{F}$ is f-inductive (for some indexing set I), then so too is $\bigcap_i \mathscr{F}_i$. Thus, define the smallest f-inductive set of $\mathcal{F}$:

$$\mathscr{F}_0 = \{A \in \mathcal{F} \mid A \in \mathscr{F} \text{ for every } f\text{-inductive subset of } \mathcal{F}\}$$

The fact that $\mathscr{F}_0$ is the smallest f-inductive set will allow us to force certain properties on to it. Our goal is to show that $\mathscr{F}_0$ *must* be a chain, which we will use to arrive at a contradiction.

To show that $\mathscr{F}_0$ is a chain, we will have to go through some technical steps. First, let $\mathcal{H} = \{A \in \mathscr{F}_0 \mid \text{if } B \in \mathscr{F}_0 \text{ and } B \subsetneq A \text{ then } f(B) \subseteq A\}$. The set $\mathcal{H}$ is the set of all elements $A \in \mathscr{F}_0$ such that B element is an element of $\mathscr{F}_0$ and is properly contained in A, f of that elements is contained in or equal to A. However, some oddities occurs with this set too: note that $\mathcal{H}$ is non-empty since vacusously,

$\emptyset \in \mathcal{H}$ (since (False $\Rightarrow$ True) and (False $\Rightarrow$ False) are both *true* statements!!) This oddity is necessary, since we are guaranteed that $\mathcal{H}$ is non-empty. Now, we'll show that if $A \in \mathcal{H}$ and $C \in \mathscr{F}_0$, then one of two relations *must* hold between A and C: either $C \subseteq A$ or $f(A) \subseteq C$. Notice we are forcing a subset relation between all elements of $\mathscr{F}_0$ and $\mathcal{H}$ – wel'l use this to create a total order later. First, let's prove the last assertion:

Proof. Define $\mathcal{G}_A = \{C \in \mathscr{F}_0 \mid C \subseteq A \text{ or } f(A) \subseteq C\}$. By definition, $\mathcal{G}_A \subseteq \mathscr{F}_0$. If we show that $\mathcal{G}_A$ is f-inductive, then $\mathscr{F}_0 \subseteq \mathcal{G}_A$, and since $\mathscr{F}_0$ is the smallest f-inductive subset, it must be that $\mathcal{G}_A = \mathscr{F}_0$ which will tell us that any element $C \in \mathscr{F}_0$ must be either contained in A or contain $f(A)$. For notational convinience, let's label the condition $C \subseteq A$ as [1] and condition $F(A) \subseteq C$ as [2]

(a) Since $\emptyset \in \mathscr{F}_0$, $\emptyset \subseteq A$, and so by [1] $\emptyset \in \mathcal{G}_A$

(b) Let $C \in \mathcal{G}_A$. We must show that $f(C) \in \mathcal{G}_A$. Since $C \in \mathcal{G}_A$, then we have 3 possibilites (we're splitting [1] into 2 seperate cases:

   i. $C \subsetneq A$: Since $C \subsetneq A$ and $A \in \mathcal{H}$, then $f(C) \subseteq A$, so by [1] $f(C) \in \mathcal{G}_A$

   ii. $C = A$: Since $C = A$ then $f(C) = f(A)$. Which is equivalent to $f(C) \subseteq f(A)$ and $f(A) \subseteq f(C)$. For our purposes, we care that $f(A) \subseteq f(C)$. By condition [2], $f(C) \in \mathcal{G}_A$

   iii. $f(A) \subseteq C$, then since $C \subsetneq f(C)$ by definition of f, $f(A) \subseteq f(C)$, and so by condition [2], $f(C) \in \mathcal{G}_A$

   in all cases, $f(C) \in \mathcal{G}_A$

(c) Let $\mathcal{C} \subseteq \mathcal{G}_A$ be a chain in $\mathcal{G}_A$. We must show that $\cup\mathcal{C} \in \mathcal{G}_A$. Then, either for all $C \in \mathcal{C}$, $C \subseteq A$, in which case $\cup\mathcal{C} \subseteq A$, and so by [1] $\cup\mathcal{C} \in \mathcal{G}_A$, or there is some $C \in \mathcal{C}$ $f(A) \subseteq C \subseteq \cup\mathcal{C}$, in which case by [2] $\cup\mathcal{C} \in \mathcal{G}_A$. In both cases, the chain is in $\mathcal{G}_A$

Thus, $\mathcal{G}_A$ is indeed f-inductive for any A, and so $\mathscr{F}_0 = \mathcal{G}_A$

□

Using the previous result, we will force one more important result: $\mathcal{H} = \mathscr{F}_0$.

Proof. As before, it is already given that $\mathcal{H} \subseteq \mathscr{F}_0$. If we show $\mathcal{H}$ is f-inductive, we'll have that $\mathcal{H} = \mathscr{F}_0$:

(a) As mentioned before, $\emptyset$ has no proper subset, so $(F \Rightarrow T)$ is vacuously true, meaning $\emptyset \in \mathcal{H}$

(b) Let $A \in \mathcal{H}$. We must show $f(A) \in \mathcal{H}$, that is if $B \in \mathscr{F}_0$ and $B \subsetneq f(A)$ then $f(B) \subseteq f(A)$. Since $B \in \mathscr{F}_0 = \mathcal{G}_A$, then either $B \subseteq A$ or $f(A) \subseteq B$. Since $B \subsetneq f(A)$, it cannot be the latter, thus we must have that $B \subseteq A$. From here, we have two possibilities:

   i. $B \subsetneq A$: If so, then since $A \in \mathcal{H}$, by definition $f(B) \subseteq A \subseteq f(A)$, where the last $\subseteq$ comes from the definition of f. Thus, $f(B) \subseteq f(A)$

   ii. $B = A$: Then $f(A) = f(B)$, in particular $f(B) \subseteq f(A)$

   in either case, we get $f(B) \subseteq f(A)$

(c) Let $\mathcal{C} \subseteq \mathcal{H}$ be a chain. To show $\cup\mathcal{C} \in \mathcal{H}$, it must be that for any $B \in \mathscr{F}_0$ where $B \subsetneq \cup\mathcal{C}$, then $f(B) \subseteq \cup\mathcal{C}$. We will do a sneaky trick for this proof using the law of excluded middle [a]: For each $C \in \mathcal{C}$, either "$C \subseteq A$ for some

A", or "$A \subsetneq C$ for all A" (which is it's direct negation). If the latter was true then

$$B \subsetneq \cup\mathcal{C} \subseteq \cup\{f(A) \mid A \in \mathcal{C}\} \subseteq B$$

implying $B \subsetneq B$ – a contradiction. Thus, it must be that for some $A \in \mathcal{C}$, $B \subseteq A$. There are again two cases

i. $B \subsetneq A$: Since $A \in \mathcal{H}$, then $f(B) \subseteq A \subseteq \cup\mathcal{C}$, so $f(B) \subseteq \cup\mathcal{C}$

ii. $B = A$: Then we have $B \in \mathcal{H}$, so $\mathcal{G}_B$ is defined and $\mathscr{F}_0 = \mathcal{G}_B$. Since $\mathcal{G}_B$ is f-inductive, $\cup\mathcal{C} \in \mathcal{G}_B = \mathscr{F}_0$. Thus, either $\cup\mathcal{C} \subseteq B$ or $f(B) \subseteq \cup\mathcal{C}$, and since $\cup\mathcal{C} \subseteq B$ is not possible since $B \subsetneq \cup\mathcal{C}$, it must be that $f(B) \subseteq \cup\mathcal{C}$.

In either case, $f(B) \subseteq \cup\mathcal{C}$, showing that $\cup\mathcal{C} \in \mathcal{H}$, showing that $\mathcal{H}$ is f-inducitve

Thus, $\mathscr{F}_0 = \mathcal{H}$ as we sought to show □

Finally, we get to combine all these facts. If $A \in \mathscr{F}_0 = \mathcal{H}$ and $B \in \mathscr{F}_0 = \mathcal{G}_A$, then either $B \subseteq A$ or $A \subseteq f(A) \subseteq B$. We couldn't do this earlier since we didn't know that $\mathscr{F}_0 = \mathcal{H}$ and $\mathscr{F}_0 = \mathcal{G}_F$ for any $F \in \mathscr{F}_0$. Thus, we have that $A \subseteq B$ or $B \subseteq A$, showing that $\mathscr{F}_0$ is a chain. We can now define $M = \cup\mathscr{F}_0$. Since $\mathscr{F}_0$ is f-inductive, then by the 3rd condition $M \in \mathscr{F}_0$. By the 2nd condition:

$$\cup\mathscr{F}_0 = M \subsetneq f(M) \in \mathscr{F}_0$$

But $f(M) \subseteq \cup\mathscr{F}_0 = M$, so $M \subsetneq f(M) \subseteq M$ implying $M \subsetneq M$ – a contradiction. The source of the contradiction was assuming an f existed which let's as choose an ever bigger value, and so no such function can exist, meaning for some A, there does not exist a B such that $A \subsetneq B$, but then A is maximal. And so, we have that (1) $\Rightarrow$ (2)

(2) $\Rightarrow$ (3) Let $(P, \leq)$ be a non-empty partialy orderd set. We want to show that $(P, \leq)$ contains a maximal chain (i.e. a chain C that cannot be enlarged further without loosing its total order). Let $\mathcal{F}$ be the set of all chains in P, orderd by inclusion (so if $P = A \leq B$ is a chain and $Q = A \leq B \leq C$, then $P \leq Q$). This set is non-empty, since $\emptyset \in \mathcal{F}$ and $\{x\} \in \mathcal{F}$ for all $x \in P$. Furthermore, Since any finite subchain of any infinite chain in $\mathcal{F}$ would be in $\mathcal{F}$, $\mathcal{F}$ has finite character. Thus, by Turkey's Lemma, $\mathcal{F}$ has a maximal element, meaning P has a maximal chain.

(3) $\Rightarrow$ (4) Let $(P, \leq)$ be a non-empty partially ordered set where each chain has an upper bound. Then by the Hausdroff Maximality Principle, there exists a maximal chain $M \subseteq P$. Since M is a chain, it has an upper-bound – let's say m. Then m is must be a maximal element of P: If there exists some $x \in P$ such that $m \leq x$ and $m \neq x$, then $M \cup \{x\}$ is a chain that is properly included in M, but M is maximal – a contradiction. Thus, m is maximal in P

(4) $\Rightarrow$ (5) Let S be any set, and let $\mathcal{L}$ be the family of all well-ordered set $(W, \leq)$ such that $W \subseteq S$. This set is non-empty, since for each $x \in S$, $\{\{x\}, \{x, x\}\} \in \mathcal{L}$. In the upcoming proof, we will give an ordering to $\mathcal{L}$ and use Zorns Lemma to show that the maximal element will be $(S, \leq)$ with appropriate $\leq$

Define $(W_1, \leq_1) \preceq (W_2, \leq_2)$ if either

(a) $W_1 = W_2$ and $\leq_1 = \leq_2$ or

(b) there exists a $a \in W_2$ such that $W_1 = \{x \in W_2 \mid x \leq_2 a,\ x \neq a\}$, and $\leq_2$ agrees with $\leq_1$ on W_1 ($\leq_1 \subseteq \leq_2$ [b]). In this case, W_2 is called a continuition of $(W_1, \leq_1)$

It should be relatively clear that $(\mathcal{L}, \preceq)$ is a partial ordering on $\mathcal{L}$. Note that it is not a total ordering: If W_i and W_j do not have the same minimal element, then it $W_i \not\preceq W_j$ and $W_j \not\preceq W_i$. This means that any chain of elements will end up with the same minimal element. Now, let $\ell = \{(W_i, \leq_i)\}_{i \in I}$ be any non-empty chain in $\mathcal{L}$ (relative to $\preceq$). We want to show it must have an upper-bound. Set $W = \bigcup_{i \in I} W_i$ and $\leq = \bigcup_{i \in I} \leq_i$ (see definition 0.1.6). Then $(W, \leq)$ is a total order (see exercise ref:HERE). We'll show it is a well-ordering. Let A be a non-empty subset of W. Since it's non-empty, there exists some $i \in I$ such that $A \cap W_i \neq \emptyset$. Since $(W_i, \leq_i)$ is well-ordered, there is an element $a \in A \cap W_i$ such that $a \leq_i x$ for all $x \in A \cap W_i$. Suppose now that there is an element $b \in A$ such that $b \leq a$. Then by definition of $\preceq$, it must be that $b \in W_i$, so $b \leq_i a$, so $b = a$. Thus, A has a smallest element a in $(W, \leq)$, meaning $(W, \leq) \in \mathcal{L}$. Furthermore, it should be clear that $(W, \leq)$ is an upper-bound for ℓ

Thus, by Zorn's Lemma, $\mathcal{L}$ has a maximal element $(W_0, \leq_0)$. If $W_0 = S$, we're done – $\leq_0$ is our well-ordering of S. If $W_0 \neq S$, let $z \in S \cap W_0$. Define $(W_0 \cup \{z\}, \leq)$, where

$$\leq = \leq_0 \cup \{(x, z \mid x \in W_0 \cup \{z\}\}$$

that is, place z after everything in W_0. Then we can see that $(W_0 \cup \{z\}, \leq) \in \mathcal{L}$ and $(W_0, \leq_0) \preceq (W_0 \cup \{z\}, \leq)$ and $(W_0 \cup \{z\}, \leq) \not\preceq (W_0, \leq_0)$. However, this contradicts the maximality of $(W_0, \leq_0)$, and so it must be that $W_0 = S$, showing that S does indeed have a well-ordering.

(5) $\Rightarrow$ (1) Let $\{A_i\}_{i \in I}$ be any non-empty family of non-empty sets, and let $S = \cup_{i \in I} A_i$. By the well-ordering principle, there exists an $\leq$ that well-orders S. Now, for each $i \in I$, let $f(i)$ be the smallest element of A_i (relative to $\leq$). Then f is a chocie function on the family $\{A_i\}_{i \in I}$

[a] For every proposition P, either P is true or $\neg P$ is true. This is implied by ZF

[b] If these seems strange, review the technical definition of a relation in section 0.1.4

This textbook contains many cases here zorn's lemma must be used, for example:

1. All proper ideals in a ring [with identity] is contained in a maximal ideal
2. All vector-spaces have a basis

(The proof of Schröder-Bernstein Theorem requires Zorn's lemma, so I can show it here, it's in Follands Real Analysis preliminary chapter)

(maybe also show Shcröder-Bernstein's lemma in the finite case after introducing cardinality?

A.3 Zorn's Lemma and Transfinite Induction

For the purposes of this textbook, the most important consequence of the Axiom of Choice is Zorn's lemma. I think a section on Zorn's lemma is important to be able to reference it and explain how it

is a sort of "super induction".

(Include transfinite induction, used a bit when showing IBN for vector spaces)

(there is a surprising amount of progress made in fields because of a single illustrative observation: In representation theorem,the fact that any $N \subseteq M$ has a compliment implies M is the direct sum of simple modules gives rise to a ton of rep theory. The fact that any R-module over a vector-space is free gives rise to liner algebra. The fact k-homomorphism from from k to F can always be extended to a k-homomorphism to algebraically closed field over k $\Phi : F \to L$ leads to results about normal extensions that are abused like crazy. I'm sure there is something more with Baer's criterion. I just realized that what I actually just rambled about is how Zorn's lemma opened up many new avenues of thinking...)

B

Sub-Group of a Free Group is Free

In this book, the fact that a subgroup of a free-group is free doesn't come into play. It is interesting tangentially as an example of a property which is closed under taking subgroups, contrasting to how being finitely generated is is not a property closed under taking subgroups, or that a sub-module of a free module is not always free, but we don't use this fact anywhere else. However, some might be curious as to how we would prove this fact. The goal of this section is to do just that using only group-theoretical knowledge.

B.1 Using Cosets

There are in fact many proofs of this fact, including a famous one by Serre topological properties of trees (connected graphs with no cycles). Such a proof would be covered in geometric group theory.

Lemma B.1.1: Finite Index Subgroup Of Finitely Generated Subgroup

Let G be a finitely generated subgroup, and let $H \leqslant G$ be a subgroup of finite index (i.e. $|G : H| = s$ for some $s \in \mathbb{N}$). Then H is finitely generated

Proof **:**
Let $G = \langle x_1, x_2, ..., x_n \rangle$ be generated by the elements $X = \{x_1, x_2, ..., x_n\}$, $H \leqslant G$ so that $|G : H| = s$, and let $A = \{a_1 = 1, ..., a_s\}$ be a set of representative sfor the right cosets of H (also known as a transversal for H in G).

Let $x \in H$. Since $H \leqslant G$, then $x \in G$. Since G is finitely generated we have that

$$x = y_1 y_2 \cdots y_n$$

where y_i is either a generator or an inverse of a generator. Since $y_1, y_2, ..., y_n$ are not necessarily in H, this is not a finite representation of x. Instead, we'll show that from these we can create a representation with elements in H. Take the representative a_{i_1} which represents the coset Hy_1, so $a_{i_1} \in Hy_1$, meaning $a_{i_1} = h_1 y_1$ for some $h_1 \in H$. Then

$$x = y_1 a_{i_1}^{-1} a_{i_1} y_2 y_3 \cdots y_n$$

Next, let a_{i_2} be the represetnative for $Ha_{i_1} y_2$ so that $a_{i_2} = h_2 a_{i_1} y_2$. In particular, $h_2^{-1} = a_{i_1} y_2 a_{i_2}^{-1}$ so that this conjugation is part of H. Now, re-write

$$x = y_1 a_{i_1}^{-1} a_{i_1} y_2 a_{i_2}^{-1} a_{i_2} y_3 \cdots y_n$$

Similarly for a_{i_3}, picking a representative from $Ha_{i_2} y_3$. Going on in this fashion, we'll end up with an equation

$$x = y_1 a_{i_1}^{-1} a_{i_1} y_2 a_{i_2}^{-1} a_{i_2} y_3 a_{i_3}^{-1} a_{i_3} \cdots y_n a_{i_n}^{-1} a_{i_n} \tag{B.1}$$

Thus, every element $x \in H$ can be in written out as in equation B.1. Now, as we noticed earlier: $a_j y_k a_l^{-1} \in H$. If $a_{i_n} \in H$. Thus, every triple $a_j y_k a_l^{-1}$ in equation B.1 is in H. The only element that might not be in H is a_{i_n} (if a_{i_n} is a representative for a coset zH for $z \neq 1$). If $a_{i_n} \in H$, we'd have a finite representation, so suppose $a_{i_n} \notin H$. Then we can "squish" everything down to

$$x = ha_{i_n}$$

where $h = y_1 a_{i_1}^{-1} a_{i_1} y_2 a_{i_2}^{-1} a_{i_2} y_3 a_{i_3}^{-1} a_{i_3} \cdots y_n a_{i_n}^{-1} \in H$. But then, this implies that $x \in Ha_{i_n}$, and since cosets partition groups (ref:HERE (preliminary chapter on equivalence relation)), we have $x \notin H$ – a contradiction to our original assumption. Therefore, $a_{i_n} \in H$, and so x is finitely represented by the set of triplet products $a_j y_k a_l^{-1}$, wherre $a_j, a_l \in A$ an $y_k \in X \cup X^{-1}$. Since x was arbitrary, this is true for any element of H, and so H is finitely generated.

This theorem is interesting in it's own right as a partial result to when a subgroup of a finitely generated subgroup is finitely generated [1]. It also has application in geometric group theory where we are interested in finding waht is the growth type of the function $\mathfrak{g}(G)(n)$ (ex. polynomial growth, exponential growth) which represents the number of words which require n elements from the generators to represent it. In this case, what we found was if $m(H)$ represents the minimum number of generators for the group H, then $m(H) \leq 2m(G)|G : H|$. For more, see cite:HERE

The part that is more intriguing for us is the way we have constructed the elements of H. Though we've limited to $|H : S|$ being finite, we can still represent the elments of $H \leqslant G$ with the same trick! Using this fact, we will show that the elements of a subgroup of a free-group will have to be relationless:

Theorem B.1.2: Nielsen-Schreier's Theorem

Let G be a free group and $H \leqslant G$. Then H is a free gorup

[1] recall that $F_2' \leqslant F_2$, the commutator subgroup of the free group on 2 elements, is not finitely genrated (see ref:HERE)

***Proof* :**
Let F be any free group freely generated [a] by the set of elements X. Note that X need not be finite, or even countable. By the well-ordering principle (ref:HERE), fix some ordering of the generators and their inverses. Then, order the elements of F, first by lengh of their words, and for words with the same length, order them lexicographically (i.e. the same way you order a dictionary). This is now a well-ordering of F.

Let $H \leqslant F$. Choose as representatives for the cosets of H the element which is first in the ordering. Letting A be the set of these representatives, then we saw in B.1.1 H is generated by elements of the form $a_j x_k a_l^{-1}$, $a_j, a_l \in A$, $x_k \in X$. Let's label the element $a_j x_k a_l^{-1}$ as $y_{j,k}$, and let Y represent the set of all these elements. We will show that Y is the set of free generators of H by showing: (1) the generators cannot be reduced, (2) the words the generators form cannot be reducted

1. The elements $a_j x_k a_l^{-1}$ are in reduced form.

   Suppose not. Then some terms in the word must cancel, which would only be possible if x_k is cancelled out somehow. Let's say it was cancelled out on the left. Then $a_j = u x_k^{-1}$ and u $u \in H a_j$. But then, $u = a_j x_k$, meaning that u is the representative of $H a_j x_k$, that is $u = a_l$. Thus, we get
   $$a_j x_k a_l^{-1} = u x_k^{-1} x_k a_l^{-1} = a_l x_k^{-1} x_k a_l^{-1} = 1$$
   However, no such relation exists between these elements in F – a contradiction. The same argument applies if x_k is cancelled from the right, or take the inverse of $a_j x_k a_l^{-1}$

2. If x is any reduced word by non-identity elements in Y, then when writing x as a word from X, the elements x_k (or their inerses) in the middle of $y_{j,k}$ are not cancelled

   Let's say some x_k is cancelled by some y on the right, say in the product $y_{j,k} y_{r,s}$. Then the a_l^{-1} occuring in $y_{j,k}$ must be cancelled first, then x_k would be cancelled. One way for this to happen is for a_r to start with $al x_k^{-1}$. Then like before, x_k^{-1} is also a representative of its coset. BUt looking at $y_{j,k}^{-1} = al x_k^{-1} a_j^{-1}$, we see that a_j represents the coset of $a_l x_k^{-1}$. That means that $a_j = a_l x_k^{-1}$, and thus $y_{j,k} = 1$.

   If x_k is cancelled by x_s in $y_{r,s}^{-1}$ to its right, then $y_{r,s}^{-1}$ msut start with a_l and $x_s = x_k$, but then $y_{r,s} = y_{j,k}$, and so the word x is not reducced in Y, contradiction our assumption.

   Finally, it might be that x_k is cancelled a letter in the a_s^{-1} that ends $y_{r,s}$. Then inverting x, we get a wrod in whic $y_{r,s}^{-1} y_{j,k}^{-1}$ occurs, and the x_s^{-1} in $y_{r,s}$ is cancelled in the same manner as in the the first case for x_k, and just like before this implies that $y_{r,s} = 1$.

   Similarly if x_k is cancelled from the left, showing that x cannot be reduced using elements from Y

From this, we see that Y freely generates H, and so H is free, as we sought to show.

[a] as in the elements of X can be treated as letters

B.2 Using Tree's

A perhaps more intuitive proof of this theorem comes from some interesting results on groups acting on trees. HERE (Clara Löh's proof is really concise and I like it!)

C

Basic Topology and Topological Groups

This appendix is by no means supposed to be a subsitute for a proper introduction of Topology. Rather, it should be treated as either a revision or a test of knowledge to see if the prerequisite Topology knowledge for chapter ref:HERE has been aquired. To that end, this will be a speed-run of definitions and basic concepts with some exercises sprinkled in to give the reader the opportunity to check their knowledge. Good books to get a proper grasp of topology would be Munkre's Topology or _____.

C.1 Definition and Properties

Definition C.1.1: Topology

Let X be a set and $\mathcal{T}_X \subseteq P(X)$ be a subset of the powerset of X such that

1. $\emptyset$ and X are in $\mathcal{T}_X$
2. For any arbitrary collection of elements in $\mathcal{T}_X$, their union is in $\mathcal{T}_X$
3. for any finite collection of elements in $\mathcal{T}_X$, their intersection is in $\mathcal{T}_X$

Then $(X, \mathcal{T}_X)$ is call a *topological space* and $\mathcal{T}_X$ is called a *topology on* X. Elements of $\mathcal{T}_X$ are called *open sets*.

Usually we omit the $\mathcal{T}_X$ and simply call X the topology, thus we say that $Y \subseteq X$ is open if and only if $Y \in \mathcal{T}_X$. If $x \in X$, then a *neighbhorhood* of X will be an *open set* $U \in \mathcal{T}$ such that $x \in U$. In

some book, a neighbhorhood is simply some set containing x and need not be open; sometimes to emphasize that the neighbhorhood is open, it is called an *open neighbhorhood*. The *neighbhorhood system* of an element $x \in X$ is the collection of all open neighbhorhoods containing x. Topologies can also be defined in terms of closed sets:

Proposition C.1.2

Let X be a set and $\mathcal{T}_X^C \subseteq P(X)$ be subset of the powerset such that

1. $\emptyset$ and X are in $\mathcal{T}_X^C$
2. For any finite collection of elements in $\mathcal{T}_X^C$, their union is in $\mathcal{T}_X^C$
3. for any arbitrary collection of elements in $\mathcal{T}_X^C$, their intersection is in $\mathcal{T}_X^C$

Then $\mathcal{T}_X^C$ forms a topology on X identical to the topology $\mathcal{T}_X$ if every set is replaced with its copmliment.

***Proof* :**
This should be done as an exercise

Many times, we define a topology through the use of a topological basis (not to be confused with a vector-space basis)

Definition C.1.3: Topological Basis

Let X be a set. Then a subset of the powerset $\mathcal{B} \subseteq P(X)$ is called a *topological basis* (or just *basis* if unambiguous) where:

1. for every $x \in X$, there is some $U \in \mathbb{C}\mathrm{P}$ such that $x \in U$
2. For every $U, V \in \mathcal{B}$ and every $x \in U \cap V$, there exists a $B \in \mathcal{B}$ such that $x \in B \subseteq U \cap V$

If only the first condition is met, then $\mathcal{B}$ is called a *topological subbasis* (or *subbasis* for short). The topology $\mathcal{T}_\mathcal{B}$ on X is the topology where $U \in \mathcal{T}_\mathcal{B}$ if and only if U is the union of some subfamily of $\mathcal{B}$. If $\mathcal{B}$ is a subbasis, then U is some union and intersection of a subfamily of $\mathcal{B}$.

Two topologies that can be naturally defined on any set X are the *trivial topology* of $\mathcal{T}(X) = \{\emptyset, X\}$ and the *discrete topology* of $\mathcal{T}(X) = P(X)$. The Euclidean topology is a common topology defined on $\mathbb{R}^n$ for $n \geq 1$. Another common topology is defined on real or complex vector-spaces using a function called a *norm* (very similar to the euclidean norm used in when defining euclidean domains): a norm is a function $\|\cdot\| : V \to \mathbb{R}$ where

1. $\|x + y\| \leq \|x\| + \|y\|$
2. $\|c \cdot x\| = |c| \cdot \|x\|$
3. $\|x\| = 0$ if and only if $x = 0$

Then then the collection of all $B_r(x) = \{y \in X \mid \|x - y\| \leq r\}$ forms a topological basis for a topology. Another common way of defining a topological space is by using a metric: if (X, d) is a metric space,

then the collection of all $B_r(x) = \{y \in X \mid d(x, y) < r\}$ forms a basis for X.

We relate two topologies by taking *continuous functions* between them:

Definition C.1.4: Continuous Function

Let X, Y be topological spaces and $f : X \to Y$ be a function. Then f is called a continuous function is the pre-iamge of every open set of Y is in X:

$$f^{-1}(U) \in \mathcal{T}(X)$$

Common examples of continuous functions are the polynomials functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ given the euclidean topology in both cases. If a function is also bijective, then we call such a function a bijective continuous function. Unlike homomorphisms (be it groups, rings, etc.) a bijective continuous function does *not* imply the inverse function is continuous, or equivalently, that the image of open sets are open (can you think of an easy example from $\mathbb{R}^2$ to $\mathbb{R}$?), and so we give a special name to a bicontinuous bijective function:

Definition C.1.5: Homeomorphim

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be two topological spaces. Then a function $f : X \to Y$ is called a *homeomorphism* if f is bijective and

$$f^{-1}(\mathcal{T}_Y) = \mathcal{T}_X$$

examples of homeomorphisms are any bijective function if the codomain has the indiscrete topology or the domain has the discrete topology, the identity function or $f(x) = x^3$ on $\mathbb{R}$ in the discrete, indiscrete, and euclidean topology. A functions that is bijective continuous but are not homeomorphisms is the identity function from X to X where the domain has the discrete topology and the codomain has the trivial topology. Note that for any topological space X, the identity function is continuous, the composition of continuous functions is continuous, and associativty of continuous functions follows from the associativity of functions. Hence, the collection of topological spaces and continuous maps from a category

Definition C.1.6: Category Of Top

The collection of all topologies on any set and the collection of continuous maps between topologies forms a category, labelled as **Top**, with objects being topologies and morphisms begin continuous maps. Homeomorphisms are the isomorphisms of **Top**.

If the codomain of a continuous function is an algebra, (usually $\mathbb{R}$ or $\mathbb{C}$), we can perform arithmetics on continuous functions:

Proposition C.1.7: Arithmetics of Continuous Functions

Let $f : X \to A$ and $g : X \to A$ be two continuous functions with A being an R-algebra (usually $\mathbb{R}$ or $\mathbb{C}$). Then $f + g$, fg and cf for $c \in A$ are continuous functions.

This fact is used in a couple of examples in the book, and so proving this proposition is an important exercise. If the fact that the codomain is an R-algebra makes the proof seem complicated, simply let

$A = \mathbb{R}$ or $\mathbb{C}$ with the euclidean topology (it is all that is really necessary to prove).

Given an arbitrary set $U \subseteq X$, we may consider the set of all closed subsets containing U. Since the arbitrary intersection of closed sets is closed, there is a notion of smallest closed subsets *containing* U. Similarly, we may consider all open subsets *contained* in U, and take their union, giving the dual

Definition C.1.8: Closure And Interior

Let $U \subseteq X$ be an arbitrary set. Then

1. The *closure* of U, denoted $\overline{U}$ is the intersection of all closed sets containing U
2. The *interior* of U, denoted $\text{int}(U)$, is the union of all open sets contained in U.

Given a topological space, injective and surjective maps give rise to natural spaces, and the product of sets has a natural topology:

Definition C.1.9: Subspace, Quotient Space, and Product Space

Let X be a topological space. Then

1. the subset $Y \subseteq X$ is a *subsapce* of X if it is endowed with the topology given by the sets $\{U \cap Y \mid U \subseteq X \text{ open}\}$
2. a set of equivalence space $Y = X/\sim$ is a *quotient topology* if it is endowed with the topology where $U \subseteq Y$ is open if and only if $\pi^{-1}(U) \subseteq X$ is open (where $\pi : X \to Y$ is the natural projection)
3. Given a collection of topological spaces $\{X_i\}_{i \in I}$ for an indexing set I, the product topology is the topology endowed on products of the sets $Y = \prod_{i \in I} X_i$ where $\pi_i^{-1}(U)$ forms a sub-basis for Y.

The reader may verify these satisfy the necessary universal properties: the universal property of subspaces, the universal property of quotient spaces (which is the dual of the former universal property), and the universal property of products.

Recall that a topological space has varying notion of separation defined on it. The most prominent of which is the Hausdorff axiom (also known as T_1):

Definition C.1.10: Hausdorff space (T_2)

Let X be a topological space, and let $x, y \in X$ be arbitrary different points. Then X is called *Hausdorff* if there always exists open sets U, V such that $x \in U$, $y \in V$, and $U \cap V = \emptyset$.

There are two weaker conditions that do show up in this book:

1. Féchet (T_1) means that for each $x, y \in X$, there exists open neigbhorhoods U, V such that $x \in U$, $y \in V$, but $y \notin V$ and $x \notin U$. Note that U and V need not be disjoint. Key feature of T_1 is that all singletons are closed.

2. Even weaker is Kolmogorov (T_0) (Zariski topologies are T_0 but not T_1)

Proposition C.1.11: Hausdorff And Diagonals

Let X be a topological space. Then X is Hausdorff if the diagonal $\Delta \subseteq X \times X$ with the product topology is closed.

Proof **:**
exercise

(define it, show that if $A \subseteq X$ is closed, then A is compact)

A final tool that is very common in topology is that of the *sequences*[1]

Definition C.1.12: Sequence

Let X be any set. Then a *sequence*, is a function $f : \mathbb{N} \to X$. Usually, the element $f(n)$ is labeled as x_n for some dummy variable x, and the function f is represented as $(x_n)_{n=1}^{\infty}$ or simply (x_n) for short.

Sequences become interesting once they are in a topological space, and there is a notion of them "approaching" a point. In a metric space, we can define this notion using epsilon-balls. However, there is a more general notion of convergence for topological spaces that are non-metrizable

Definition C.1.13: Convergent Sequence

Let X be a topological space, (x_n) a sequence in X, and $x \in X$ an element of X. Then (x_n) is said to *converge to* x if for every neigbhorhoods U of x, there exists a natural number $S(U) \in \mathbb{N}$ such that

$$x_n \in U \qquad \forall n \geq S(U)$$

Proposition C.1.14: Convergent Sequence In Hausdroff Spaces

Let X be a topological space and (x_n) a convergent sequence. Then (x_n) converges to *at most* 1 point. The resulting point is called the *limit point*.

Proof **:**
exercise

For most cases that are considered in algebra and analysis, every point $x \in X$ has a "countable neighborhood basis", which essentially translates to every open set is "locally" a countable union of open sets. This property is called *first-countable*. It is not important for this book, but it is important to be able to state the following result:

[1] More generally, ultra-filters or nets may be more appropriate for the topological setting, however they are not necessary for this review of topology

Proposition C.1.15: Closed Sets In Hausdroff Spaces w.r.t. Sequences

Let $U \subseteq X$ be any subset of a first-countable Hausdorff topological space X. Then if $\lim U$ represents the collection of all limit points of all possible convergent sequences $(x_n) \subseteq X$, then:

$$\overline{U} = \lim U$$

Proof **:**
exercise

Most spaces covered in a topological class, and all topological spaces covered in this book, have this property and hence closedness can be characterized with sequences. If $\overline{U} = X$ (i.e. the closure gives the entire space), then we give such a set a name

Definition C.1.16: Dense Set

Let X be a topological space. Then if $U \subseteq X$ is a subset of X such that $\overline{U} = X$, then U is called a *dense set*.

These are very important in real and functional anslysis, and they show up in the discussion of completion in section 20.4.

If X is a metric space, we can define the notion of a cauchy sequence:

Definition C.1.17: Cauchy Sequence

Let X be a metric space with metric d. Then (x_n) is a *cauchy sequence* if for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n, m > N$,

$$d(x_n, x_m) < \epsilon$$

Given any metric space, we can "complete" it by adding all of the cauchy sequence that do not converge in the following way: for any two cauchy sequences, they are called *equivalent* if $x_n - y_n \to 0$ in G. Then the collection of all equivalent cauchy sequence is called the *completion of the metric space*. It should be checked that this is equivalence to the following definition:

Definition C.1.18: Complete Metric Space

Let X be a metric space. Then X is compelte if all cauchy sequences are convergent sequences

Example C.1.1: Complete Spaces
$\mathbb{R}$ is a complete space, while $\mathbb{Q}$ is not

The notion of completeness given in section 20.4 is in fact a generalization of the above notion, as we'll show in section C.4. If is often the case that when working with complete spaces, we want to have

C.2 Construction Topologies

The subspace, product, and quotient topology are all special examples of the following two constructions:

Definition C.2.1: Weak [initial] Topology

Let $F = \{f : X \to Y_\alpha\}_{\alpha \in A}$ be a collection of maps from a set A to topological spaces Y_α indexed by $\alpha \in A$. Then the coarsest topology on X which makes each f continuous is called the *weak topology*. In particular, a basis for the weak topology generated by F consists of $f_\alpha(U)$ for all open sets $U \subseteq Y_\alpha$ for every $\alpha \in A$.

Note that it is not interesting to ask what is the finnest topology on X that makes all the maps continuous; that would simply be the discrete topology on X. Note too that since the weak topology $\mathcal{T}$ is the coarest topology on X that makes all the maps continuous, if $\mathcal{T}'$ is another topology on X that makes each map in F continuous, then $\mathcal{T}' \subseteq \mathcal{T}$. Note too that we showed in the definition how to actually create the weak topology, however for the skeptical reader it should be verified that the weak topology does indeed exist

Example C.2.1: Weak Topology

1. Let $\{X_\alpha\}_{\alpha \in A}$ be some collection of topologies and $X = \prod_{\alpha \in A} X_\alpha$. Let $\Pi = \{\pi_\alpha : X \to X_\alpha\}$ be the collection of functions. Then the weak topology on X generated by F is *precisely* the product topology of X This is be

2. Let X be any topological space and $A \subseteq X$ any subset. Let $\iota : A \to X$ be the inclusion map. Then the weak topology on A generated by ι is the *subspace* topology.

One of the great advantages of of defining weak topologies is that we can make the collection of f_α and spaces X_α anything we want. This flexibility will the definition's of many interesting topologies on funcitonal spaces (ex. collection of smooth functions on compact spaces, collection of bounded continuous functions, or simply collection of bounded functions, and so forth). All such spaces are of key importance in funtional analysis. Weak topologies are also a special case of a much more general construction, which we'll explore at the end of this chapter.

Finally, there is also the dual of weak topologies where we take maps that map *into* X.

Definition C.2.2: Final Topologies

Let Y be a set, $\{X_\alpha\}_{\alpha \in A}$ be a collection of topological spaces, and $F = \{f_\alpha : X_\alpha \to Y\}$ be a collection of maps into Y. Then the the finnest topology on Y that makes each f_α continuous is called the *final topology*. In particular, $U \subseteq X$ is open if and only if $f_\alpha(U)^{-1}$ is open for each $\alpha \in A$

Again, it should be verified that this is indeed a topology. While the weak (or initial) topology seeks for a domain with the coarsest topology, final topologies seek a codomain with the finnest topology. This is because the coarsest topology on the co-domain will always create continuous maps. Unlike the initial topology, the final topology is not given the name "strong" to contrast the name "weak" for initial topologies. There is in fact a topology known as the strong topolgoy on a space X: If $\|\cdot\|$

is a norm on X[2], the strong topology on X is the topology generated by all the balls $B_r(x)$ for all $x \in X$ and $r > 0$. The contrast between strong and weak topologies will make more sense when studying functional analysis, where there we would much rather a space have a strong topology, but it will sometimes be unfeasible (ex. the space of all smooth functions, or infinite dimensional vector spaces won't have a good norm defined on them), and so we will "weaken" the topology that will create topology that will not be induced by something as strong as a norm (usually, it the weak topology will be induced by a countable collection of seminorms, see Folland chapter 5 for more information).

Example C.2.2: Final Topologies
Let X be an arbitrary topological space an $A \subseteq X$ a subspace. Let $\pi : X \to A$ be the projection map. Then the final topology on A is the *quotient topology*.

Both the initial and final topologies can be generalized even further. If you have read VIII, both the weak and strong topologies are simple examples of *limits* and *colimits* in **Top**. These two also go by the name of *projective limit* and *injective limit*, or *colimit* and *limit*.

Exercise C.2.1

1. There are some cool exercises Hossain gave that I'd like to include for those who have covered the section on commutative algebra and would like to know some more topological properties of $\mathrm{Spec}(R)$.
2. Let X be a topological space. Show that X is Hausdorff if and only if the diagonal $\triangle = \{(x, y) \in X \times X \mid x = y\}$ is closed.

C.3 Compactness and Connectedness

Definition C.3.1: Compact Space

Let X be a topological space. If $\{U_i\}_{i \in I}$ is a collection of open sets such that $\bigcup_{i \in I} U_I = X$, then $\{U_i\}_{i \in I}$ is called an *open cover*. An *open subcover*, or *subcover* for short, is a subst of an open cover. If any open cover of X adminits a finite subcover, then X is *compact*.

Proposition C.3.2: Closed Subset Of Compact Spaces

Let $U \subseteq X$ be a closed subset of a compact space. Then U is compact is a subspace.

Proof **:**
exercise

[2]See definition ref:HERE

Definition C.3.3: Connected Space

Let X be a topological space. Then if there exists no non-trivial proper clopen subsets, X is connected. Equivalently, X is connected if there does not exist two non-empty open sets $U_1, U_2 \subseteq X$ such that $X = U_1 \cup U_2$.

C.4 Topological Groups

In this book, topological groups show up when talking about the completion of rings in section 20.4 and when talking about the representation of non-finite groups in chapter 24. The topological considerations in both of those cases were purposefully scarce to not have any topological requirements for learning those materials. However, having a topological perspective brings out a deeper perspective, and so they are further explored in this section.

Definition C.4.1: Topological Group

Let G be a topological space. Then if there exists a binary operation $+ : G \times G \to G$ satisfying the axioms of a group that is also *continuous* (where $G \times G$ has the product topology), and a $(_)^{-1} : G \to G$ mapping $g \mapsto g^{-1}$ that is *continuous*, then G is a *topological group*

Though we have defined topological groups, the same definition can be done for any other algebraic object we've worked with, notably *topological rings*, *topological R-modules*, and *topological R-algebras*. Topological groups (resp. rings, modules, algebras) is a special example of a *group object* in the category of **Top** (see ref:HERE, currently at the end of Cat theory part).

Example C.4.1: Topological Group

1. For any group G, if G is equipped with the discrete topology it is a topological group, and all homomorphisms are continuous between any other topological group. In this way, the theory of topological groups can be thought as subsuming group theory. Any group with the indiscrete topology (also known as trivial topology) is also a topological group.

2. If $H \leqslant G$ is also a topological space, then H is a topological group with the subspace topology.

3. The real numbers $\mathbb{R}$ and $\mathbb{R}^\times$ under multiplication is a topological group (can you prove continuity of $+$, $\cdot$, and $(_)^{-1}$?). More generally, $\mathbb{R}^n$ is a topological group, or any *topological vector-space* (a vector-space where the action is continuous as well) is a topological group

4. The circle S^1, and the n-torus $(S^1)^n$ are topological groups. Show that $\mathbb{R}/\mathbb{Z}$ is both homeomorphic and isomorphic to S^1. If we take S^1 as a subset of $\mathbb{C}$, then we $S^1 = U(1) = \{x \in \mathbb{C} \mid |x| = 1\}$.

5. The general linear group over the reals $\mathrm{GL}_n(\mathbb{R})$ is a topological group. Similarly, the special linear group,$\mathrm{SL}_n(\mathbb{R}) = \{A \in \mathrm{GL}_n \mid \det(A) = 1\}$, is a topological group.

6. Take $O(n) \subseteq \mathrm{GL}(n)$, where $O(n) = \{A \in \mathrm{GL}_n(\mathbb{R}) \mid AA^t = I\}$. This is a topological group known as the *orthogonal group*, since it consists of all orthonormal basis of $\mathbb{R}^n$. Similarly, we

can define $U(n) = \left\{ A \in \mathrm{GL}_n(\mathbb{R}) \;\middle|\; A\overline{A}^t = I \right\}$ is a topological group known as the *unitary group*. Similarly to $\mathrm{SL}_n(\mathbb{R})$, we have:

$$\mathrm{SO}(n) = \mathrm{SL}_n(\mathbb{R}) \cap O(n) \qquad \mathrm{SU}(n) = \mathrm{SL}_n(\mathbb{R}) \cap U(n)$$

Note that all these groups are *compact*.

7. (If you know some Differential Geometry) Any Lie group is naturally a topological group. An example of a topological group that is not a Lie group is $\mathbb{Q}$ with the subspace topology induced by $\mathbb{R}$.
8. If you have covered chapter 20, section 20.4, then the p-adic integers, $\mathbb{Z}_p$ has the induced inverse-limit topology (in this case, the p-adic topology). This group is compact, homeomorphic to the Cantor set, but is totally disconnected (hence cannot be a Lie Group)
9. Similarly to $\mathbb{Z}_p$, giving $\mathbb{Q}$ a different p-adic metric, we can take a different completion to form $\mathbb{Q}_p$, which is also a Lie Group.

Proposition C.4.2: Topological Criterion

Let G be a group and a topological space. Then G is a topological group if and only if the map $\cdot : H \times H \to H$, $(x, y) \mapsto xy^{-1}$ is continuous

Proof **:**
exercise

Definition C.4.3: Topological Homomorphism/Isomorphism

Let $f : G \to H$ be a function between topological groups. Then if f is a continuous homomorphism (resp. a homeomorphic isomorphism) then f is called a *topological homomorphism* (resp. *topological isomorphism*)

Example C.4.2: Topological Isomorphism
Naturally, the identity map is a topological isomorphism. If the topology is discrete, every homomorphism/isomorphism is a topological homomorphism/isomorphism.

It is possible that a homomorphism/continuous map is not a continuous map/homomorphism. As an example of a continuous, even homeomorphic map, that is not a homomorphism is the following: let $g \in G$ and $\varphi_g : G \to G$ be the map sending $x \mapsto gx$. Show that φ_g is a homeomorphism. This example is particularly interesting since it shows that every $g \in G$ induces a homeomorphism! A space in which the maps φ_g are homeomorphisms is called a *homogeneous space*. As a consequence of this, if U is a neighborhood of 1, then gU is a neighborhood of g, meaning that the topology of G is determined by the neighborhoods of 1!

As an exercise, it is worth seeing how if $H \leqslant G$, then G/H has the quotient topology, is a homogeneous space, and if $H \trianglelefteq G$, then G/H is a topological group. Furthermore, There exists continuous homomorphisms $f : G \to H$ where $G/\ker(f) \ncong f(H)$ (i.e. there exists continuous bijective

homomorphisms that are not isomorphisms in the category of topological groups). However, the 3rd isomorphism theorem is essentially the same, as can be done as an exercise.

Some authors make being Hausdorff (or at least T_1) part of the definition of the group. This limitation is not imposed in this definition, since in algebraic geometry, there will be non-Hausdorff spaces (ex. the a-adic topology on certain topological rings are non-Hausdorff). For Topological groups, there is a way to always find a quotient topology that is Hausdorff:

Proposition C.4.4

Let G be a topological group, $e \in g$ the identity, and $H = \bigcap_{\substack{U \subseteq G \text{ open} \\ e \in U}} U$ Then:

1. H is a subgroup
2. H is the closure of e, i.e. $\overline{e} = H$
3. G/H is Hausdorff
4. G is Hausdorff if and only if $H = \{e\}$

***Proof* :**
exercise

Topological groups are explored in a great many areas; we will list a few places topological groups show up without going into any detail. Perhaps the most studied topological groups are *locally compact topological groups* (topological groups where for any point $x \in G$, there exists a compact neighborhood). These are very useful in analysis: one can always define a measure on them (called the *Haar measure*), and the representation of topological groups leads to the *Fourier transform*, which is the decomposition into irreducible representations of the continuous group action of $\mathbb{R}$ on the Hilbert space $L^2(\mathbb{R})$. (Something with homotopy theory. Don't enough yet to understand). Finally, topological groups show up in *Hilbert's fifth problem*. First, it was proven that if $f : G \to H$ is a continuous homomorphism between lie groups, then f is in fact *smooth*(!). Thus, if G is a topological group, if there can exist a smooth structure on it, making it a lie group, then that structure must be unique. Hilbert's fifth problem asks if any topological group is also a topological manifold (i.e. locally homeomorphic to $\mathbb{R}^n$), then is it a lie group (i.e. the operation is smooth, and the group is locally diffeomorphic to $\mathbb{R}^n$). The answer was shown in 1955 to in fact be yes(!) by Andrew Gleason, Deane Montgomery, and Leo Zippin. Even better, G will always be *real analytic*.

C.5 Completion in Topology

The last subject of topological groups we will explore here is the *completion* of topological groups, since it generalizes some of the results of section 20.4 in chapter 20. This definition of Cauchy generalizes that of the definition of a Cauchy sequence in a metric space by realizing that we only need to have a notion of "continuous subtraction" to define a Cauchy sequence:

Definition C.5.1: Cauchy Sequence In Topological Group

Let G be a topological group and $(x_n) \subseteq G$ be a sequence. Then (x_n) is a *Cauchy sequence* if for every neighborhood U around 0, there exists a natural number $s(U) \in \mathbb{N}$ such that

$$x_n - x_m \in U \qquad \forall n, m > S(U)$$

Just like for metric spaces, the collection of all equivalent Cauchy sequence is the completion of the group G, which we'll denote as $\hat{G}$ (notice that if x_n, y_n are Cauchy sequence, so is $(x_n - y_n)$, and so $\hat{G}$ depends only on the equivalence classes). For any $g \in G$, we can define the constant sequence (g), which is an element of $\hat{G}$. We thus identify G in $\hat{G}$ by the map $\iota : G \to \hat{G}$ via $g \mapsto [(g)]$. Note that in general that ι is *not* in general injective. In fact, we have that:

$$\ker \iota = \bigcap_{\substack{0 \in U \\ U \text{ open}}} U$$

and so ι is injective if and only if G is Hausdorff. Completion is functorial: Since continuous homomorphisms preserve Cauchy sequences, $f : G \to H$ maps to $\hat{f} : \hat{G} \to \hat{H}$, and $g \hat{\circ} f = \hat{g} \circ \hat{f}$.

This construction is quite general: we can allow for $G = \mathbb{Q}$ with the euclidean topology, then $\hat{G} = \mathbb{R}$. We can also allow for $G = \mathbb{Q}$, and define a different metric (known as the p-metric), which will give us $\hat{G} = \mathbb{Q}_p$. We will focus on a particular case that useful for commutative algebra and show how it relates to completion.

Define a topology on a commutative group G by taking a descending sequence of subgroups[3]:

$$G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_n \supseteq \cdots$$

This is called a *filtration* of the group G. Let the collection of (G_n) define the *local neighborhood* of 0, and $x + G_n$ be an open neighborhood of x. Then this generates a topology, which we'll call the *subgroup topology*: $U \subseteq G$ is a neighborhood of 0 if and only if U contains some G_n. An example of such a topology would be:

$$\mathbb{Z} \supseteq p\mathbb{Z} \supseteq p^2\mathbb{Z} \supseteq \cdots \supseteq p^n\mathbb{Z} \supseteq \cdots$$

The subgroup topology is interesting in that every subgroup G_n is open and closed (that is, clopen and hence connected). First, for any $g \in G_n$, $g + G_n \subseteq G_n$, and so G_n is open. Thus, for any $h \notin G_n$, $h + G_n$ is open (since translation is a homeomorphism), and so:

$$H = \bigcup_{h \notin G_n} h + G_n$$

is open and disjoint from G_n. In fact, $H^c = G_n$, showing that G_n is *closed*.

Now, since G with the subgroup topology is a topological group, we can define $\hat{G}$ in the exact same way. Since the open sets are derived from a descending chain of subgroups G, we see can observe that cauchy sequences will be induce a special type of sequence called a *coherent sequence* which shall be equivalent to Cauchy sequence in the subgroup topology. Let (x_n) be a Cauchy sequence in G with the subgroup topology generated by (G_n). Then in G/G_n, the image of(x_n) must be eventually constant; say it is ζ_n. When considering G_{n+1}, we may get different constant, say ζ_{n+1}.

[3] If G were not commutative, we would take a descending sequence of normal subgroups

When considering the relation between these two, it is natural to see that the natural projection map ζ_{n+1} to ζ_n, that is the map:

$$\pi_{n+1} : G/G_{n+1} \to G_n$$

Now, for all n, we will get some ζ_n. The collection of these ζ_n's is called a *coherent sequence*. We chose the maps to be projection maps, however we may imagine a similar such sequence with any such functions, as long as they map to the right elements. Thus, being a little more general, a coherence sequence is a collection (ζ_n) such that for a collection of functions f_n, we have:

$$f_{n+1}(\zeta_{n+1}) = \zeta_n \qquad \forall n \in \mathbb{N}$$

It is left as an exercise that equivalent Cauchy sequences define the same coherent sequence, and that a coherent sequence defines a Cauchy sequence, and hence these two concepts capture the same idea. For this reason, the collection of all coherent sequences is also the completion of G, and is also denoted $\hat{G}$.

This now brings us to the same definition as given in section 20.4 in the case of Abelian groups: an *inverse system* is a collection of subgroups $\{G_n\}$ and functions f_n such that

$$f_{n+1} : G_{n+1} \to G_n$$

Then the group of all coherent sequences, $\hat{G}$, is called the *limit*, and is denoted $\varprojlim G_n$, where the homomorphism f_n are usually implicitly understood. With this notation, we have that:

$$\hat{G} \cong \varprojlim G/G_n$$

The most common subgroup topology and inverse system is when $G = R$ and $G_n = I^n$ is an ideal. In this case, the subgroup topology of R is called the *I-adic topology*, or simply the *I-topology*. Recall that we have defined a norm on a vector-space to define a topology. We can similarly define a norm on any topological group defined via the I-adic topology which will generate the I-adic topology: Define

$$\|x\| = C^{-n} \qquad \text{if} \qquad x \in I^n,\ x \notin I^{n+1}$$

for some $C \in \mathbb{N}$. If $x \in I^n$ for all $n \in \mathbb{N}$, then we will say that $\|x\| = 0$. Then through this norm, we can define a metric:

$$d(x, y) = \|x - y\|$$

and so we can define a completion, $\hat{G}$ as defined in definition C.1.18, which the reader should verify is the same as $\varprojlim G/G_n$. This new metric we have defined is also more powerful than the usual euclidean metric, namely:

$$\|x + y\| \leq \max(\|x\|, \|y\|) \qquad d(x, y) \leq \max(d(x, z), d(z, y))$$

This metric condition is *much* stronger and is called an *ultra-meric*. The reader is invited to prove the following incredible properties of topologies with ultra-metrics:

1. Given a collection of any three points $x, y, z \in X$ one of the following three equalities must hold:

   $$d(x, y) = d(y, z) \qquad d(x, z) = d(y, z) \qquad d(x, y) = d(z, x)$$

   Hence, every triple *must* form an isosceles triangle.

2. If $B_r(x)$ is an open ball then every point inside the ball is it's center: if $d(x, y) < r$:

$$B_r(x) = B_r(y)$$

Every intersection must contain each other:

$$B_r(x) \cap B_s(y) \Rightarrow B_r(x) \subseteq B_s(y) \qquad \text{or} \qquad B_y(s) \subseteq B_r(x)$$

If $r > 0$, then $B_r(x)$ is clopen.

What is notable is that completion is the beginning of real analysis (and hence also PDEs), and given the strong algebraic structure introduced we also have variants such completions coming up in complex analysis. In CITE:CFT, we shall use different completion of $\mathbb{Z}$ with p-adic metrics to construct fields in which we are "keeping track' of a single prime that is splitting or ramifying.

Theorem C.5.2: Universal Property Of Completion

Here

Proof **:**
here

C.5.1 Classifying Complete Archimidean Fields

(Should I show that $\mathbb{R}$ and $\mathbb{C}$ are the only complete archimidean fields upto isomorphism?

Exercise C.5.1

1. (Relate this to DVR topologies?)

D

Lattice Theory and Universal Algebras

(here: Two important classifying objects but not really central to any current big field of study. Lattice theory going to be presented as build-up to Stone Duality, while Universal algebra presented as a start to Model Theory)

D.1 Lattice Theory

We first generalize the notion of gcd and lcm for partial orders:

Definition D.1.1: GCD and LCM for Partially ordered sets

Let $(S, \leq)$ be a partially ordered set. Then

1. a *lowerbound* of a subset $T \subseteq S$ is an element $l \in S$ such that $l \leq t$ for all $s \in T$. A *greatest lower bound* (gcd) is a lowerbound g such that for any other lower-bound l, $l \leq g$. In lattice theory, the greatest lower bound is also called the *meet* or *infimum*

2. an *upperbound* of a subset $T \subseteq S$ is an element $u \in S$ such that $t \leq u$ for all $s \in T$. A *least upper bound* (gcd) is a upper-bound ℓ such that for any other upper-bounds u, $\ell \leq u$. In lattice theory, the greatest lower bound is also called the *join* or *supremum*

The reader should check that if a meet/join exists, then it is unique. If the meet/join of a subset

exists for a subset $T \subseteq S$, then it is denoted:

$$\bigwedge_{t \in T} t \qquad \text{and} \qquad \bigvee_{t \in T} t$$

Definition D.1.2: Lattice

Let $(S, \leq)$ be a partially ordered set. Then

1. If for any two $a, b \in S$, $a \wedge b$ exists, then S is called a *lower semi-lattice*
2. If for any two $a, b \in S$, $a \vee b$ exists, then S is called a *upper semi-lattice*
3. If S is both a lower and upper semi-lattice, S is called a *lattice*.

Proposition D.1.3: Properties Of Lattice Binary Operation

Let S be a semi-lattice. Then $\vee$ and $\wedge$ as binary operations are

idempotent, commutative, associative, and order preserving.

Furthermore, by using induction, every finite subset has a meet and join

Proof **:**
Exercise. Note that it is enough to prove it for $\wedge$ as $(S, \leq)^{\text{op}} = (S, \leq^{op})$ turns any meet into a join.

Example D.1.1: Lattices

1. $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and real fields (see section 19.1) with the usual $\leq$ all form lattices; what is $a \wedge b$ and $a \vee b$ for any a, b in these sets?
2. Let S be a set with powerset $P(S)$. Then with $\cup$ and $\cap$, $P(S)$ is a lattice
3. Let S be the collection of all (subgroups/ideals/submodules) of (group/ring/module). Then S forms a lattice with $A \cap B \in S$ and $\langle A \cup B \rangle \in S$. In the ideal/submodule case, $\langle A \cup B \rangle = A + B$.
4. Any partially ordered set can be represented by a graph where an edge represents where $x \leq y$ if $x \to y$. Then the following are examples of non-lattices:

5. Let X be a topological space. Then the set of open sets forms a lattice, for the union of two open/closed sets is an open/closed set.

Unlike most algebraic structures encountered with two binary operations, lattices are *not* distributive in general (can the reader find a counter-example?). They have the following alternative:

Proposition D.1.4: Absorption Law

Let L be a lattice. Then:

$$x \wedge (x \vee y) = x \qquad x \vee (x \wedge y) = x$$

Proof **:**
exercise

Using the above result, we can formulate lattices purely algebraically:

Proposition D.1.5: Algebraic Definition Of A Lattice

Let L be a set with binary operations $\vee$ and $\wedge$ that are associative and commutative. Then if these satisfy the absorption laws, then L forms a lattice

Proof **:**
exercise

Naturally, a lattice homomorphism if a order-homomorphism such that $\varphi(x \vee y) = \varphi(x) \vee \varphi(y)$ and $\varphi(x \wedge v) = \varphi(x) \wedge \varphi(y)$. Note that not all order-homomorphisms are lattice homomorphisms (as the reader should come up with a counter-example). However an order-isomorphism is a lattice-isomorphism (see exercise ref:HERE).

Lattices as they are can be considered as two commutative *semigroups* with a compatible structure (see section 7); the absorption being the unique feature of lattices. If furthermore the lattice is bounded, then it shall have a monoid structure.

Definition D.1.6: Bounded Lattice

Let L be a lattice. Then L is said to be *bounded* if there exists an minimal element (denoted 0 or $\perp$) and a maximal element (denoted 1 or $\top$) such that for all $x \in L$:

$$\perp \leq x \leq \top \qquad \text{equiv.} \qquad 0 \leq x \leq 1$$

As an algebraic structures, we can think of a bounded lattice as a set $(L, \vee, \wedge, 0, 1)$ such that $(L, \vee, \wedge)$ is a lattice, 0 is the identity element for the join operation $\vee$, and 1 is the identity element for the meet operation $\wedge$:

$$a \vee 0 = a \qquad a \wedge 1 = a$$

If we require the stronger condition that every subset $S \subseteq L$ of a lattice has a supremum and infimum, we get the following stronger notion:

Definition D.1.7: Complete Lattice

Let L be a lattice. Then if every subset $S \subseteq L$ has a gcd and lcm, then L is said to be a *complete lattice*

This makes L a bounded lattice, for $L \subseteq L$ must have a gcd and lcm. Every finite set can be made into a complete lattice, however $\mathbb{N}, \mathbb{Q}, \mathbb{R}$ are incomplete. Every closed set of $\mathbb{R}$ is complete. More notably, $P(X)$ for any set X, as well as the collection of (subgroups/ideals/submodules) of (groups/rings/modules) forms a complete lattice.

Given any lattice, it can be embedded in a complete lattice. Rather interestingly, this shall link to Galois connections and the closure map:

Definition D.1.8: Closure Map

Let $(S, \leq)$ be a partially ordered set. Then the *closure map* is an order-homomorphism $\Gamma S \to S$ that is idempotent, $(\Gamma \circ \Gamma) = \Gamma$, and expanding, $((\Gamma(x) \geq x)$, for all $x \in X)$. An element $x \in S$ is *closed relative to* Γ if $\Gamma(x) = x$.

The closure map is so named since the closure map $A \mapsto \overline{A}$ is the quintessential closure map. As we shall see when discussing pointless topology, lattices and topologies shall be intimately related, elucidating better this example.

Lemma D.1.9: Closed Element Infimum and Supremum

Let L be a complete lattice and Γ a closure map. Then the set of closed elements relative to Γ is closed under infimums and supremums.

Proof **:**
exercise

Next, we define the Galois connection more generally:

Definition D.1.10: Galois Connection

Let X, Y be two partially ordered sets. Then a *galois connection* between X, Y is a pair of order-reversing homomorphism, say $\alpha : X \to Y$ and $\beta : Y \to X$ such that for all $x \in X$ and $y \in Y$:

$$(\beta \circ \alpha)(x) \geq x \qquad (\alpha \circ \beta)(y) \geq y$$

Naturally, the Galois correspondence is a Galois connection, and the generalization to closed groups shows the correspondence in the infinite case.

Lemma D.1.11: Galois Connection And Lattices

Let X, Y be partially ordered sets and (α, β) a Galois connection. Then

1. α, β, induce order-reversing mutually inverse bijections between $\mathrm{im}(\alpha)$ and $\mathrm{im}(\beta)$
2. $\alpha \circ \beta$ and $\beta \circ \alpha$ are closure maps
3. If X, Y are complete lattices, $\mathrm{im}(\alpha)$ and $\mathrm{im}(\beta)$ are complete lattices

Proof **:**
exercise

Theorem D.1.12: Lattice Embedding Theorem

Let $(L, \leq)$ be a partially ordered set. Then L can be embedded in a complete lattice $\hat{L}$ in such a way that all infimums and supremums of L are preserved.

Proof **:**
Grillet (p.544, rather standard argument)

Examples of Lattices

Lattices are rather general objects. As it turns out, when working the set of sub-objects of an R-module with $(+, \cap)$, we get some more interesting behavior:

Definition D.1.13: Modular Lattice

Let L be a lattice. Then L is called *modular* if whenever $x \leq z$,

$$x \vee (y \wedge z) = (x \vee y) \wedge z$$

Note that for any lattice, if $x \leq z$ then x and $y \wedge z$ are lower bounds of $\{x \vee y, z\}$ and hence:

$$x \vee (y \wedge z) \leq (x \vee y) \wedge z$$

Thus, a lattice is modular if and only if $x \leq z$ implies $x \vee (y \wedge z) \geq (x \vee y) \wedge z$. Note that the opposite of a modular lattice is modular, hence the theory is self-dual. As mentioned, the name modular lattice comes straight form the lattice of submodules:

Example D.1.2: Modular Lattices
Letting $(\vee, \wedge) = (\cap, +)$, and taking the lattice of submodules of an R-module. Then if $A \subseteq C$, then taking $x \in (A + B) \cap C$ we have $x = a + b \in C$ for some appropriate $a \in A \subseteq C$. Then $b = x - a \in C$, and $x = a + b \in A + (B \cap C)$

Detecting when a lattice fails to be modular is similar to how to detect a complete bipartite graph:

Proposition D.1.14: Detecting Modular Lattice

Let L be a lattice. Then L is modular if and only if it does not contain a sub-lattice isomorphic to N_5:

$$\begin{tikzcd}
 & 1 & \\
a \arrow[ur] & & c \arrow[ul] \\
 & & b \arrow[u] \\
 & 0 \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

Proof **:**
Certainly this lattice is not modular (as the reader should check). For the converse, use the fact the lattice is not modular to find elements that do not satisfy the modularity property and construct N_5

Modular lattices help generalizes results such as theorem 3.7.5(2) (i.e. that all maximal normal series have unique maximal length).

Proposition D.1.15: Modular Lattices And Finite Maximal Chains

Let L be a modular lattice. Then any two finite maximal chains have the same length

Proof **:**
exercise (or p.547 of Grillet)

A little more restrictive than modular lattices are *distributaive* lattices:

Definition D.1.16: Distributive Latice

Let $(L, \vee, \wedge)$ be a lattice. Then L is a *distributive lattice* if one of the two equivalent conditions hold for all $x, y, z \in L$:

1. $x \vee (y \wedge z) = (x \vee y) \wedge (y \wedge z)$
2. $x \wedge (y \vee z) = (x \vee y) \wedge (y \vee z)$

Proposition D.1.17: Detecting Distributive Lattices

Let L be a lattice. Then L is a distributive lattice if and only if L does not have a sublattice isomorphic to N_5 or M_5, where M_5 is the lattice:

$$\begin{tikzcd}
 & 1 & \\
a \arrow[ur] & b \arrow[u] & c \arrow[ul] \\
 & 0 \arrow[ul] \arrow[u] \arrow[ur] &
\end{tikzcd}$$

Proof **:**
exercise

The power-set $P(X)$ is the canonical distributive lattice with $(\cup, \cap)$. In fact a lattice is distributive if and only if it is isomorphic to a sublattice of $P(X)$ for some appropriate X. To see this, we require the following build-up

Definition D.1.18: Lattice Irreducible Elements

Let L be a lattice. Then an element $i \in L$ is called *sup-irreducible* (or just *irreducible*) if i is not a minimal element and:

$$x \vee y = i \quad \Rightarrow \quad x = i \text{ or } y = i$$

The collection of irreducible elements is denoted $\mathrm{Irr}(L)$.

In $P(X)$, singletons are irreducible.

Proposition D.1.19: Lattice And Irreducible

Let L be a lattice satisfying the descending chain condition. Then every elemetn $x \in L$ is the join of the set of irreducible elements of L

Proof **:**
exercise (see Grillet p.551)

Definition D.1.20: Order Ideal

Let S be a partially ordered set. Then a [left] order ideal $I \subseteq S$ is a subset such that if $y \in I$ and:

$$x \leq y \in I \quad \Rightarrow \quad x \in I$$

Order ideals naturally create a distributive lattice given any partially ordered set:

Proposition D.1.21: Order Ideals And Partially Ordered Sets

Let S be a partially ordered set. Then the order ideals, ordered by inclusion, produce a distributive lattice, denoted $\text{Ideals}(S)$

Proof **:**
exercise.

Theorem D.1.22: Finite Distributive Lattices And Ideals

Let L be a finite lattice. Then L is a distributive lattice if and only if

$$L \cong \text{Ideals}(\text{Irr}(L))$$

Proof **:**
exercise

hello world!

This result can be generalized to all ideals, see Grillet for further explanation.

A final important lattice naturally arises in mathematical logic, known as a *boolean lattice*. In a Boolean lattice, the meet, join, and "complement" correspond to conjunction, disjunction, and negation in mathematical logic:

Definition D.1.23: Boolean Lattice

Let L be a bounded lattice (hence containing a 0 and 1). Then a *complement* of an element $a \in L$ is an element a' such that $a \wedge a' = 0$ and $a \vee a' = 1$. Then a lattice is a *Boolean lattice* if every element has as compliment

The reader should verify that compliments are unique, and that the operation of taking the compliment, $(-)'$ is a lattice-homomorphism.

Example D.1.3: Boolean Lattice
For an easy example of a Boolean lattice take the power-set $P(X)$ with compliment $Y' = X \setminus Y$. For a more interesting example, let R be a Boolean ring (i.e. a ring such that for $x \in R$, $x^2 = x$). Then define a partial order on R via:

$$x \leq y \qquad \text{if and only if} \qquad xy = x$$

Then we get a boolean lattice via

$$x \wedge y = xy \qquad x \vee y = x + y + xy \qquad x' = 1 - x$$

Interestingly, the all Boolean lattices can be converted into Boolean rings, for a Boolean lattice $(L, \vee, \wedge, (-)', 0, 1)$ define:

$$x = y = (x' \wedge y) \vee (x \wedge y') \qquad xy = x \wedge y$$

denote this ring $R(L)$. Then

$$\mathrm{L}(\mathrm{R}(L)) = L \qquad \mathrm{R}(\mathrm{L}(R)) = R$$

Since power-sets are naturally boolean lattices, they offer themselves to the following classification theorem:

Theorem D.1.24: Finite Boolean Lattice Classification

Let L be a finite lattice. Then L is a boolean lattice if and only if there exists a set X such that

$$L \cong P(X)$$

Proof **:**
build off of the same result for distributive lattices.

To generalize to general boolean lattices, we require the following lemmas:

Lemma D.1.25: Stone Space Basis

Let L be a Boolean lattice and let X be the set of all prime ideals. Then the set

$$V(a) = \{P \in X \mid a \notin P\}$$

is a basis for the topology on X. Furthermore, $V(0) = \emptyset$, $V(1) = X$, and $V(a') = V(a)'$

This topology shall be given a name soon

Proof **:**
exercise

Theorem D.1.26: Boolean Lattice Classification

Let L be a lattice. Then L is a Boolean lattice if and only if there exists a set X such that L is isomorphic to a subset of $P(X)$.

Proof **:**
build off of the same result for distributive lattices.

As mentioned, the above topology has a special name:

Definition D.1.27: Stone Space

Let L be a Boolean lattice and X the collection of prime ideals of L. Then with the above basis, X is called the *stone space*.

Proposition D.1.28: Stone Space Properties

Let X be a stone space. Then X is compact, Hausdorff, and totally disconnected

Proof **:**
For compactness, it is recommended the reader proves it by showing every ultra-filer converges.

Given a space with these properties, note that every finite union/intersection of closed/open sets is closed/open, and similarly for compliment, hence it can be shown that it forms a Boolean lattice $\mathrm{L}(X)$, and so a stone space can be defined with the above properties.

D.2 Pointless Topology

Using some lattice theory, we may find the closest result to an "algebraic dual" to topological spaces (mirroring the stone duality and the duality presented in section 21.7). The main idea is that we can focus

(The overall goal is to get the adjoint pair (Ω, pt) between Top and Loc, and that it restricts to *Stone Duality*)

Heyting Algebras

(they deserve a mention due to their importance in logic)

D.3 Universal Algebras

(generalizes many theorem that have the same pattern, including isomorphism theorems, free constructions, and that there will exist sub-algebras, products, and quotient algebras).

E

Relating Algebra to Invariant Theory

Invariant theory is not really studied anymore, so it doesn't feel like it should be part of the main material. However, it presents some very interesting results, and also covers as basic matrical the 3rd of the 3 big Hilbert Theorem (Basis, Nullstellensatz, and Syzygy).

F

Model Theory

1. I want to build up the basics of Model theory for it's interesting use in Algebraic Geometry. In particualr, I want to write here the build up and proof of the Ax-Grothendeick theorem.
2. I would like to have the Lefschetz Principle and Löwenheim-Skolem theorem (the Löwenheim-Skolem theorem is doubly interesting since it can be used to formally show there is no "algebraic"' construction of $\mathbb{R}$ from $\mathbb{Q}$)
3. perhaps as a mini bonus or as exercises, I can include some other infinite algebraic contructions that can be achieved using Model Theory (ex infinite torsion group).

F.1 Löwenheim-Skolem Theorem

Part X

Glossary and selected Answers

G

Glossary of Algebraic Structures

In this chapter, we sum up all groups we've talked about of all different orders to be easily referenced.

G.1 Groups

subgroup chain:

$$\text{normal} \supset \text{abelian} \supset \text{Characteristic}$$

The klein four-group has an abelian but non-characteristic subgroup. Groups of small order.:
Free groups
here. on nice it's not lagging.

G.2 Rings

here

$$\text{rngs} \supset \text{rings} \supset \text{commutative rings} \supset \text{integral domain} \supset \text{Factorization Domains} \supset \text{UFD}$$
$$\supset \text{PID} \supset \text{Euclidean Domain} \supset \text{Field} \supset \text{algebraically closed fields}$$

The $\mathbb{Z}[\sqrt{D}]$ rings.
The finite rings
integral domains
free rings

G.3 Modules

free modules see

G.4 Universal Properties

(list all universal properties covered in the book)

G.5 Character Tables

Index

Abel's Theorem, 679
Abelian Group, 11
Absolute Value, 694
Action
    Group, 139
    Ring, 346
Acyclic Functor, 1163
Additive Function, 860
Adjoint Functor, 712
Adjoint Operator, 462
Affine Cone, 827
Affine Coordinate Ring, 798
Affine Scheme, 881
Affine Space, 782
Affording Representation, 1041
Algebra, 354
Algebraic Closure, 610
Algebraic Element, 589
Algebraic Integer, 886
Algebraic Multiplicity, 447
Algebraic Number Field, 885
Algebraic Set
    Affine, 786
    Projective, 832
Algebraically Closed, 609
Alternating Group, 28
Alternating Tensor, 529
Annihilator
    Dual Map, 432
    R-module, 379
Archimedean Embedding, 913
Archimedean Field, 691
Artin-Rees Lemma, 750
Artin-Schreir Theorem, 693
Artin-Wedderburn Theorem, 1035
Artinian Module, 773
Artinian Ring, 773
Associated Graded Ring, 755
Associative Elements, 292
Atomic Domain, 296
Automorhpism
    Group, 52

Baer's Criterion, 553
Basis, 394
Bijective, xxi
Bilinear Map, 436
Bimodule, 502
Binary Operation, 5
Birational Map, 819
Boundary, 1120
Burnside's Theorem, 1074

Cardinality, xxii
Cartan Matrix, 1094
Category, 213, 1195
    Abelian, 1115
    Additive, 1114
    Complexes, 1121
    Homotopy, 1139
    Locally Small, 1199
    Opposite, 1201
    Small, 1198
Cauchy's Theorem, 115
Cauchy-Schwartz Inequality, 454

Cayley Graph, 57
Cayley Table, 12
Cayley's Theorem, 152
Center of Group, 66
Centralizer of Group, 67
Chain, 715
Chain Complex, 1120
Chain Homotopy, 1134
Change of Basis, 412
Character, 1053
Character Table, 1057
Characteristic Ideal, 480
Characteristic Polynomial, 443
    $\mathfrak{m}$-primary Ideal, 865
Characteristic Subgroup, 167
Chinese Remainder Theorem, 287
Class Equation Theorem, 156
Class Function, 1053
Class Group, 936
Class Number, 940
Closed Real Field, 692
Closure, 77
    Algebraic, 610
    Galois, 669
    Normal, Field, 622
    Normal, Group, 110
    Zariski, 793
Codimension, 859
Cohomology, 1123
Coinduced G-module, 1184
Colimit, 708
Comaximal Ideal, 286
Commutative Diagram, 221, 708
Commutator Subgroup, 176
Companion Matrix, 486
Complement
    Group, 179
Complete Ring, 747
Completion
    $\mathfrak{a}$-adic of Ring, 741
Composite Field, 582
Composition Series
    Group, 132
    R-module, 715
Conductor, 962
Conjugacy, 154
Coprime Numbers, xxix
Cosets, 102
Coxeter Element Lemma, 1103
Curves, 859
Cycle, 1120
Cycle Type, 27
Cyclic Group, 71
Cyclotomic Field, 639
Cyclotomic Polynomial, 637

Decomposition Field, 976
Decomposition Group, 976
Dedekind Domain, 924
Dedekind-Hasse Valuation, 317
Dedekind-Kummer Theorem, 962
Degree of Field Extension, 577
Derived Functor, 1153
Determinant, 440
Diagonalizable, 449
Dihedral Group, 31
Dimension
    Quiver, 1087
    Vector Space, 400
Dimension Theorem, 870
Direct Product
    Group, 172
    R-module, 368
Direct Sum
    Group, 193
    R-module, 371
Dirichlet Unit Theorem, 945
Discriminant
    Bilinear Map, 895
    Field, 897
    Polynomial, 680
    Relative, 957
    Ring of Integers, 904
Distinguished Open Sets, 794
Divisibility, xxvii
    Ring, 291
Divisible Group, 556
Divisible Subgroup, 558
Division Ring, 233
Domain, 239
Domain of Definition, 818
Double Complex, 1166
Double Dual, 433
Dual Basis, 430
Dual Map, 431
Dual Space, 428

Dynkin Diagram, 1094

Eigenspace, 447
Eigenvalue, 445
Eigenvector, 447
Elementary Divisor, 476
Elementary Matrix, 416
Endomorhpisms
    For Groups, 48
Enough Injectives, 1141
Enough Projectives, 1141
Epimorphism, 219
Equivalence
    Matrix, 413
Equivalence of Categories, 706
Equivalence Relation, xiv
Euclidean Algorithm, xxix
Euclidean Domain, 311
Euclidean Valuation, 311
Euler Form, 1102
Euler Totient, 20
Evaluation Map, 249
Exact Chain Complex, 364
Exact Couple, 1188
Exact Sequence, 533
Exact Triangle, 1175
Exterior Algebra, 522

Faithful Action, 149
Faithful Module, 380
Feit-Thompson's Theorem, 200
Fermat's Little Theorem, xxxiii
Fermat's Sum of Two Squares Theorem, 322
Field, 233
Field Characteristic, 578
Field Extension, 576
    Abelian, 656
    Algebraic, 599
    Cyclic, 656
    Finite, 577
    Galois, 656
    Kummer, 999
    Normal, 618
    Radical, 675
    Ramified, 965
    Separable, 623
    Simple, 581
    Solvable, 677
    Transcendental, 599
    Unramified, 965
Filtration, 747
Finite Map
    Affine Variety, 850
    Quasi-Projective Variety, 851
Finitely Generated
    Group, 174
    R-module, 384
Fixed Field, 651
Fixed Field Automorphism, 596
Fixed Homomorphism, 578
Flat Module, 567
form, 826
Formal Derivative, 253, 624
Formal Power Series, 251
Fractional Ideal, 929
Free
    Abelian Group, 174
    R-Module, 375
Free Action
    Group, 147
Free Product, 192
Free Rank, 378
Freely Generated, 87
Freyd-Mitchell Theorem, 1118
Frobenius Automorphism, 990
Frobenius Homomorphism, 626
Frobenius Reciprocity, 1077
Frobenius's Theorem, 673
Function, xvi
Functor, 215
Fundamental Theorem of Algebra, 672
Fundamental Theorem of Galois Theory
    Finite, 655
    Infinite, 685

G-Module, 1179
Gabriel's Theorem, 1098
Galois Conjugate, 680
Galois Field, 641
Galois Group, 656
GCD
    Ideal, 927
    Numbers, xxviii
    Ring Elements, 303
GL_n, 35
General Linear Matrix, 416

Generating
Ideal, 269
Generators
Group, 37
R-module, 366
Geometric Multiplicity, 447
Graded Algebra, 517
Graded ring, 517
Grand-Schmidt Theorem, 458
Graph, xxxv
Graph Isomorphism, xxxv
Group, 6
Group Character, 669
Group Extension, 127
Group Lattice, 91
Group Ring, 256
Groupoid, 209, 1200

Height of Prime, 857
Hermitian Matrix, 453
Hilbert Function
Graded Module, 863
Projective Variety, 866
Hilbert's 90th Theorem, 999
Hilbert's Basis Theorem, 716
Hilbert's Nullstellensatz, 812
Homogeneous Polynomial, 826
Homogenization, 828
Homology, 1123
Homomorhpism
R-module, 347
Homomorphism
R-Algebra, 355
Group, 41
Quiver, 1087
R-module, 359
Representation, 1040
Ring, 231
Short Exact Sequence, 536

Ideal, 262
Ideal Contraction, 763
Ideal Extension, 763
Ideal Group, 936
Idempotent Element, 1033
Image, 64
Independence
Algebraic, 589
Linear, 394
Index, 112
Induced G-module, 1184
Inertia Degree, 958
Inertia Field, 983
Inertia Group, 983
Injective, xx
Injective Module, 552
Injective Object, 1140
Inner Automorphism, 165
Inner Product, 451
Class Function, 1061
G-Invariant, 1045
Integral Basis, 904
Integral Domain, 238
Integral Element, 758
Integral Ideal, 929
Integral Lattice, 913
Integral Trace Dual, 897
Internal Direct Product, 118
Irreducible Component, 804
Irreducible Elements, 293
Irrelevant Ideal, 835
Isometry, 452
Isomorphism
Category, 220
Groups, 50
Ring, 258
Isomorphism of Varieties, 800
Isomorphism Theorem
Group
First, 106, 120
Fourth, 124
Second, 121
Third, 123
R-Modules, 363
Ring
2,3,4, 266
First, 264

Jacobson Radical, 734
Jordan Block, 493
Jordan Canonical Form, 494
Jordan-Hölder Program, 134

Kernel
Group, 64
Group Action, 149

Ring, 258
Kronecker's Theorem, 583
Kronecker-Weber Theorem, 999
Krull Dimension, 774
Krull Principal Ideal Theorem, 870
Krull Topology On Galois Group, 684
Krull's Theorem, 753

Lagrange Theorem, 112
Lasker-Noether Theorem, 808
LCM
Ideal, 927
Numbers, xxviii
Ring Elements, 303
Legendre Symbol, 970
length of Composition Series, 715
Limit, 708
Linear Functionals, 429
Local Property, 728
Local Ring, 731
Localization, 721
Lusztig's Theorem, 1097
Lying Over, 765

Mapping Cone, 1129
Maschkle's Theorem, 1047
Matrix Group, 35
Matrix Ring, 255
Maximal
Ideal, 274
Metabelian Group, 195
Minimal Polynomial, 590
Minkowski's Bound, 923
Modular Arithmetics, xxxi
Module, 346
Monogenic Number Ring, 907
Monoid, 207
Monomorhpism, 218
Multilinear Map, 436, 505
Multiset, xxiii

Natural Transformation, 705
Nielsen-Schreier's Theorem, 1234
Nilpotent Group, 200
Noether's Normalization Theorem, 810
Noetherian
R-Module, 384
Ring, 271, 714
Nonsingular Matrix, 410
Norm
Absolute, 919
Character, 1064
Field, 890
Relative, 955
Vector Space, 455
Normal Closure
Field, 622
Group, 110
Normal Operator, 466
Normal Subgroup, 104
Normalizer, 68
Null-Homotopic, 1133
Nullity, 402

Opposite Ring, 1027
Orbit, 148
Orbit-Stabilizer Theorem, 144
Order of an element, 22
Order on Set, xiii
Ordered Field, 689
Orghogonal, 438
Orthogonal, 457
Orthogonal Compliment, 459
Orthogonal Element, 1033
O_n, 35
Orthogonal Operator, 468
Orthonormal, 457
Outer Automorphism, 167

p-Group, 159
Parity of Cycle, 28
Partially Ordered Set, 1225
Partition, xv
Perfect Field, 626
Perfect Group, 181
Permutation Representation, 143
Character, 1058
Picard Group, 1012
Poincaré Series, 861
Polycyclic Group, 204
Polynomial Ring, 246
Presentation
Formal Definition, 37
Module, 382
Quotient Definition, 111
Primary Ideal, 807
Prime

Elements in Ring, 293
Prime Ideal, 278
Ramified, 958
Split, 958
Totally Ramified, 958
Unramified, 958
Prime Ideal Conjugate, 975
Prime Subfield, 579
Primitive Element Theorem, 633
Primitive Root of Unity, 580
PID, 272
Product of Projective Space, 845
Profinite Completion, 682
Profinite Group, 682
Projection Map, 404
Projective Module, 544
Projective Object, 1140
Projective Plane Curve, 855
Projective Space, 826
Pythagorean Theorem, 458

Quadratic Reciprocity, 970
Quadratic Residue, 969
Quasi-Isomorphism, 1125
Quiver, 1084
Quotient
Group, 100
Ring, 264

R-balanced Map, 501
R-module of Finite Length, 777
Radical Ideal, 801
Ramification Index, 958
Rank
Matrix, 417
Vector Space, 402
Rational Canonical Form, 487
Rational Closed Set, 819
Rational Function On quasi-projective Variety, 843
Rational Functions, 816
Rational Mapping, 818
Reduced Ring, 815
Reducible Component, 804
Reflection Functor, 1099
Regular Function, 782
Quasi-projective Variety, 839
Regular Local Ring, 872
Regular Map
Quasi-projective Variety, 840
Regular Point, 816
Regulator, 947
Relation, xiii
Replacement Theorem, 398
Representation
Induced, 1076
Matrix, 406
Quiver, 1086
Simple, 1090
Representation, Linear, 1039
Residue Field, 733
Resolution
Free, 382
Injective, 562, 1146
Projective, 1146
Resultant, 443
Riesz Representation Theorem, 460
Ring, 228
Ring Extension, 758
Ring of Fraction, 281
Ring of Fractions, 283
Ring of Integers, 886
Order, 1011
Root Lattic, 1101
Root of Unity, 580
Root System, 1092

Saturated Ideal, 725
Section Map, 540
Self-Adjoint Operator, 466
Semi-Direct Product, 183
Semi-Simple Ring, 1018
Separable Degree, 628
Separable Polynomial, 623
Set Operations, ix
Short Exact Sequence, 534
Side Divisor, 316
Similarity, 414
Simple
Group, 130
Ring, 274, 1018
Size
of vector, 454
Snake Lemma, 1127
Solvable Group, 197
SL_n, 35

SO_n, 36
Spectral Decomposition, 449
Spectral Sequence, 1189
Spectrum of Ring, 876
Split Exact Sequence, 539
Splitting Field, 614
Stabalizer, 148
Standard Resolution, 1182
Subgroup, 10, 61
Submodule, 348
Subnormal Series, 131
Subobject, 219
Subring, 231
Subspace, 353
Sum
    R-module, 363
Surface, 859
Surjective, xx
Sylow p-Group, 159
Sylow's Theorem, 160
Symmetric Algebra, 520
Symmetric Group, 25
Symmetric Tensor, 529
System of Linear Equations, 421

Tensor Algebra, 516
Tensor Product, 500, 503
Tensor-Hom Adjoint, 564
Terminal Object
    For Groups, 47
Tits Form, 1102
Torsion Element, 379
Torsion-Free Module, 379
Trace
    Field, 890
    Matrix, 439
Trace Pairing, 896
Transcendental Dimension, 858
Transcendental Element, 589
Transitive Action, 147
Transpose, 431
Transposition, 27

Unique Factorization Domain
    UFD, 296, 302
Unit, 235
Unitary Operator, 468
Universal Property
    Coproduct in Ab, 193
    Cokernel, 94
    Completions, 1250
    Derived Cat enough Proj, 1150
    Field of Fractions, 284
    Free Groups, 86
    Free R-algebra, 357
    Free R-Modules, 373
    Group Kernel, 93
    Injective Hull, 562
    Localization, 720
    Polynomial Rings, 249
    Product of Groups, 189
    Quotient Sets, xxxix
    Scalar Extensions, 513
    Tensor Product (Commutative Ring), 500
    Tensor Product (non-Commutative Ring), 502

Valuation
    Archimedean, 695
    Discrete, 769
    Real, 768
    Ring, 310
Valuation Ring, 769
Valuation Totally Ordered Abelian Group, 768
Vanishing Ideal, 790
Variety
    Affine, 804
    Projective, 834
    Quasi-projective, 839
Vector Space, 351
Volume
    Determinant, 442
    Lattice, 917

Weyl Group, 1095
Wild Inertia Group, 991

Zariski Dense, 793
Zariski Topology
    Affine Space, 790
    Projective Space, 834
    Spectrum, 878
Zariski's Lemma, 811
Zero-Divisor, 235