Everything You Need To Know About Undergraduate Algebra

Nathanael Chwojko-Srawley

January 20, 2023

Contents

Preliminary Chapter **i**

0.1 Sets i

0.1.1 Propositions viii

0.1.2 Proving Propositions xi

0.1.3 Induction xi

0.1.4 *Pigeon-hole Principle xi

0.1.5 Equivalence Relations xi

0.1.6 Functions xiii

0.1.7 Adding Structures to Sets xvi

0.1.8 *Cardinality xvii

0.2 Number Theory xviii

0.2.1 GCD and LCM xviii

0.2.2 Modular Arithmetic's xxi

0.2.3 Counting Problems xxi

0.3 Graphs xxii

0.4 Categories xxiv

0.4.1 Defining Categories xxv

0.5 Matrices xxviii

I Groups **1**

1 Introduction to Groups **3**

1.1 Basic Properties and Definitions . . . 3
1.1.1 Consequences of Group Axioms . . . 11
1.2 Examples of Groups . . . 17
1.2.1 Cyclic Groups . . . 17
1.2.2 Dihedral Groups . . . 21
1.2.3 Symmetric Groups . . . 24
1.2.4 Quaternion Group . . . 31
1.2.5 *Matrix Groups . . . 33
1.3 Generators and Relations . . . 35
1.4 Homomorphisms and Isomorphisms . . . 38
1.4.1 Homomorphisms . . . 39
1.4.2 Isomorphisms . . . 48
1.5 Group Actions . . . 52
1.6 Cayley Graphs . . . 55

2 Category of Group 58
2.1 Basic properties of Morphisms . . . 64
2.2 Commutative diagrams . . . 68
2.3 More abstract examples . . . 69
2.4 Summary . . . 69

3 Sub Groups and Generating Groups 70
3.1 Basic Definition and Properties . . . 70
3.2 Examples of Subgroups . . . 72
3.2.1 Kernel and Image . . . 72
3.2.2 Center, Centralizers, and Normalizers . . . 74
3.3 Subgroups Generated by Subsets . . . 79
3.3.1 Cyclic Groups and Cyclic subgroups . . . 79
3.3.2 Generating Subgroups . . . 86
3.3.3 Free Groups . . . 89
3.3.4 Free Abelian Group . . . 97
3.3.5 *Free-Forgetful Adjoint . . . 98
3.4 Subgroup Lattices . . . 98
3.5 *Universal Property of kernel and coKernels . . . 102

4 Homomorphisms and Quotient Groups 105

4.1 Quotient Groups . . . 106
4.2 Presentations . . . 117
4.3 Application to Finite Groups . . . 120
4.4 Internal Direct Product . . . 125
4.5 Isomorphism Theorems . . . 127
4.6 Extensions . . . 134
4.7 Classifying all Finite Groups: Hölder Program . . . 137
4.8 Quotient Objects . . . 143

5 Group Actions 144
5.1 Defining Group Action . . . 144
5.2 Axiomatic representation . . . 147
5.2.1 A Word on Right-actions . . . 150
5.3 Orbit Stabalizer Theorem . . . 155
5.4 Application to Finite Groups . . . 156
5.4.1 Left Multiplication . . . 156
5.4.2 Conjugation . . . 161
5.4.3 Class Equation . . . 162
5.4.4 Sylow's Theorem . . . 165
5.4.5 Applications of Sylow's Theorem . . . 170
5.5 Actions on Groups . . . 171
5.5.1 A Word on Characteristic Groups . . . 174
5.6 Actions in general . . . 176

6 Composing Groups: Direct and Semi-Direct Products 177
6.1 Classical Definitions And Properties . . . 178
6.1.1 Fundamental theorem of Finitely Generated Abelian Groups . . . 180
6.2 Recognizing Direct Products . . . 181
6.2.1 A word on Perfect Groups . . . 186
6.3 Semi-Direct Products . . . 187
6.4 A Categorical approach of the Product . . . 194
6.4.1 The dual of the Product . . . 196
6.4.2 Universal Property of Semi-Direct Products? . . . 199

7 Further Topic in Group Theory 201
7.1 Generalization of Abelian Groups . . . 201

7.1.1 Solvable Groups . . . 203
7.1.2 Nilpotent Groups . . . 206
7.1.3 Polycyclic Groups . . . 210
7.2 Non-finitely generated abelian groups . . . 211
7.3 Linear Groups . . . 211

8 Generalization of Groups 212
8.1 Monoid . . . 212
8.2 Groupoid . . . 214

II Rings 216

9 Introduction To Rings 218
9.1 Basic Properties and Definitions . . . 219
9.1.1 Category of Ring . . . 232
9.1.2 Summary . . . 235
9.2 Examples of Rings . . . 236
9.2.1 *Formal Power Series . . . 241
9.3 Ring Homomorphisms and Quotient Rings . . . 246
9.4 properties and Further Structures of Ideals . . . 256
9.4.1 Operations and Generation on Ideals . . . 256
9.4.2 Ideals and Units . . . 262
9.4.3 Maximal Ideals . . . 263
9.4.4 Prime Ideals . . . 266
9.5 Ring of Fractions: Localization . . . 269
9.5.1 Constructing the Reals? . . . 273
9.6 Chinese Remainder Theorem . . . 274
9.7 *Monoidal Objects . . . 276

10 Exploring Integral Domains: Factorization and Valuation 277
10.1 Factorization . . . 279
10.1.1 Example of Atomic Domains: Noetherian Domains . . . 284
10.1.2 Unique Factorization Domains . . . 288
10.2 Euclidean Domains . . . 294
10.3 Factorization in the Gaussian Integers . . . 303

11 Exploring Polynomial Rings 306

11.1 Recap . . . 306
11.2 Polynomial Rings over Fields . . . 308
11.3 Polynomial Rings that are UFD's . . . 310
11.4 Irreducibility Criteria . . . 313
11.5 Polynomial Rings over Fields II . . . 317
11.6 Gröbner Basis . . . 318

12 Ordered Fields **322**

III Modules **323**

13 Introduction to Modules **326**
13.1 Basic Definitions and Examples . . . 327
13.1.1 Algebras . . . 333
13.1.2 Important Examples of Modules . . . 337
13.2 Homomorphism and Quotients of Modules . . . 339
13.2.1 Quotient Modules . . . 342
13.3 Generation of Modules . . . 345
13.3.1 Direct Sum of Modules . . . 346
13.3.2 Free Modules . . . 351
13.3.3 Presentations and Resolutions . . . 358
13.3.4 Finitely Generated Modules and Noetherian Modules . . . 360
13.3.5 Direct Sums and Homomorphisms . . . 363

14 Modules over Fields – Vector Spaces and Linear Algebra **365**
14.1 Vector Spaces and their Properties . . . 366
14.1.1 Quotient Spaces and Subspaces . . . 373
14.2 Matrix Representation . . . 376
14.2.1 Change of Basis . . . 381
14.2.2 Gaussian Elimination: Finding a Nice Basis . . . 383
14.2.3 Systems of Linear Equations Geometric Results . . . 385
14.3 Dual Spaces and Forms . . . 386
14.3.1 Dual Spaces . . . 388
14.3.2 Multilinear Forms . . . 393
14.3.3 Trace and Determinant . . . 395
14.3.4 Geometry and Inner Product . . . 398

15 Modules Over PID's **401**
15.1 Classification of Modules over PIDs . . . 402
15.2 Rational and Jordan Canonical Form . . . 407
15.2.1 Converting a square matrix to Ratioanl Canonical Form . . . 414
15.2.2 Jordan Cannonical Form . . . 416
15.2.3 Changing From one Canonical Form to Another . . . 417
15.2.4 Computation: Convert to JCF . . . 417

16 Tensor Products Of Modules **418**
16.1 What's a Tensor Product? . . . 418
16.1.1 Constructing the Tensor Product . . . 419
16.2 Properties of Tensor Products . . . 426
16.3 Functors Between Mod and Scalar Extension . . . 429
16.3.1 Universal property of $S \otimes_R N$ . . . 431
16.4 Tensor Algebras, Symmetric and Exterior Algebras . . . 434
16.4.1 Symmetric Algebras . . . 437
16.4.2 Exterior Algebras . . . 438
16.4.3 Homomorphisms of Tensor Algebras . . . 440
16.4.4 Symmetric and Alternating Tensors . . . 441
16.4.5 Interesting Applications . . . 442
16.5 Monoidal Categories . . . 442

17 Extension Problem – Prelude to Homological Algebra **443**
17.1 Categorical Preliminaries and Exact Sequences . . . 445
17.1.1 Exact Sequences . . . 452
17.1.2 Relating Short Exact Sequences . . . 454
17.1.3 Splitting Sequences . . . 457
17.2 Dual and Projective Modules . . . 459
17.3 Injective Modules . . . 467
17.3.1 *Duality Revisited . . . 473
17.4 Flat Modules . . . 474
17.5 Summary . . . 477

IV Fields **479**

18 Field Theory **481**

18.1 Properties of Fields . . . 481
18.2 Field Extensions . . . 487
18.2.1 Isomorphism of Fields Extensions . . . 498
18.3 Algebraic Extensions . . . 502
18.4 Historical Detour: Straightedge and Compass Construction . . . 510
18.5 Extensions that are Closures . . . 511
18.5.1 Algebraic Closure . . . 511
18.5.2 Splitting Field and Normal Extensions . . . 516
18.6 Separable Extensions . . . 524
18.6.1 Separability and Embeddings . . . 529
18.6.2 Separability and Simple Extensions . . . 532
18.7 Finite Fields and Cyclotomic Polynomials . . . 535
18.7.1 Cyclotomic Polynomials and Fields . . . 535
18.7.2 Finite Fields . . . 538

19 Galois Theory **542**
19.1 Galois Extension and Galois Group . . . 543
19.1.1 Galois Correspondence . . . 550
19.1.2 Composite Fields . . . 559
19.1.3 A bit of Character Theory . . . 561
19.2 Application of Galois Theory . . . 564
19.2.1 Fundamental Theorem of Algebra . . . 564
19.2.2 Construction of Regular n-gon . . . 565
19.2.3 Symmetric Functions . . . 565
19.2.4 Solvability By Radicals . . . 567
19.3 Galois Groups of Polynomials . . . 570
19.4 *Infinite Galois Extensions . . . 572

V Commutative Algebra **573**

20 More on Ring Theory and Extensions **575**
20.1 Ring and Integral Extension . . . 575
20.2 Localization . . . 583
20.2.1 Localization of Rings . . . 583
20.2.2 Localization of Modules . . . 587
20.2.3 Local Rings . . . 591

20.2.4 Nakayama's Lemma . . . 593
20.3 Completions . . . 598
20.3.1 Constructing Completions . . . 599
20.4 Discrete Valuation Domains . . . 601
20.5 Artinian Rings . . . 602

21 Relating Algebra to Geometry **607**
21.1 Reminder of Noetherian Rings . . . 607
21.2 Affine Algebraic Sets . . . 609
21.2.1 Zariski Topology . . . 616
21.2.2 Radical Ideals . . . 620
21.2.3 Algebraic Variety . . . 622
21.2.4 Primary Decomposition . . . 625
21.2.5 Computation of Algebraic Sets and k-algebras . . . 625
21.3 Hilbert's Nullstellensatz . . . 625
21.3.1 More on Zariski Topology . . . 629
21.4 Dimension Theory . . . 631
21.5 Spectrum of a Ring . . . 631

22 Relating Algebra to Number Theory **637**
22.1 Ring of Integers . . . 637
22.1.1 Trace and Norms . . . 640
22.1.2 Discriminant . . . 643
22.1.3 Structure of Ring of Integers . . . 648
22.2 Archimedean embeddings of Ring of Integers . . . 651
22.3 Further Structure of Ring of Integers . . . 653
22.4 Prime Decomposition in Ring of Integers . . . 654

VI Homological Algebra **655**

23 Abelian Categories **657**

VII Representation Theory **658**

24 Semisimple Rings and Modules **660**
24.1 Decomposing Modules . . . 661

24.2 Artin-Wederburn 669

25 Representation of Finite Groups **679**

25.1 Representations of Finite Groups 679

25.2 Character Theory 691

25.2.1 More Character Tables 707

25.2.2 Burnside's Theorem 709

25.3 Induced Representation 711

25.4 *Application to Harmonic Analysis 716

26 Quiver Theory – Lie Algebra Representation **717**

26.1 A Quick exploration of Lie Algebras 718

26.2 Quiver's and their Decomposition 718

26.2.1 Quiver Representation and Varieties 719

26.2.2 Decomposing Quivers 723

26.3 Root Systems 725

26.4 Gabriel's Theorem 730

26.5 Path Algebra of Quivers 736

VIII Category Theory **739**

27 An Introduction to Category Theory **741**

27.1 Basic Definitions and Examples 742

27.2 Duality 746

27.3 Functoriality 749

27.4 Natural Transformations 753

27.5 Universal Property 756

27.6 Adjoint 757

27.7 A Second Look at Category theory 758

27.8 Commutative Diagrams 759

IX Appendix **761**

A Axiom of Choice **763**

A.1 Build-up 764

A.2 Axiom of Choice Equivalences 766

A.3 Zorn's Lemma and Transfinite Induction 771

B Sub-Group of a Free Group is Free **772**

B.1 Using Cosets 772

B.2 Using Tree's 774

C Basic Topology and Topological Groups **775**

C.1 Definition and Properties 775

C.2 Construction Topologies 778

C.3 Compactness and Connectedness 780

C.4 Topological Groups 780

D Elements of Geometric Group Theory **785**

E Relating Algebra to Invariant Theory **786**

F Model Theory **787**

X Glossary and selected Answers **788**

G Glossary of Algebraic Structures **789**

G.1 Groups 789

G.2 Rings 789

G.3 Modules 790

G.4 Universal Properties 790

G.5 Character Tables 790

The title of the books 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 study, 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 proof, 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 learnt 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. 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 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.

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[1]. 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 is 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 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 572), 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,

[1] 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

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. A review of topology is given in appendix C.

Finally, 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*.

What is [Abstract] Algebra?

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. Abstract algebra is a generalization of this, it is as a way to study the properties of operations in a more general context, beyond the familiar numbers that we use in everyday life. For example, instead of just studying the properties of addition and multiplication for the natural numbers (1, 2, 3, etc.), we can also study operations for other sets of objects, such as polynomials, shapes, networks, and so forth. Some non-numerical examples of operations on objects that the reader may already be familiar with 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. squares can be rotated by multiple of $\pi/4$ degrees ($90°$) to get back a square that is indistinguishable from the original square (in thise case, the operation is rotation by $\pi/4$)

In all of these examples, we have a *algebraic structures*. An algebraic structure has two pieces of information: a set of objects, and a collection of operations that each can take in one more many objects in the set, and gives back another object in the set. The prototypical example of an algebraic structure is the set natural numbers and the elementary operations + which each takes in two numbers and returns another number. In algebra, we shall define new operations on sets of objects by defining what rules they must follow (the *axioms*) and study their properties. The advantage of this abstraction is we can study the property of the operations directly, giving us insights on all systems (be it in physics, economics, mathematics, etc.) in which we may define such an operation.

Given an algebraic structure, we may define *algebraic equations*, which combines objects from the algebraic structure using the operations provided with the algebraic structure. Elementary algebra

studies a special family of algebraic equations. The algebraic strucures is usually the numbers with operations $+$ and $\times$, and the algebraic equations are usually *polynomial*, that is equations of the form:

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$$

Particular polynomials of intesets were:

1. *Linear Equations* are equations of the form $ax + b = 0$ for numbers a, b and variable x.
2. *quadratic equations* are equations of the form $ax^2 + bx + c = 0$ for numbers a, b, c 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. For example:

- (little more advanced) *Differential Equations*: These are equations whose solutiosn are functions an operations are $+$, 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 a differentiable function

In abstract algebra, the set of objects that we use 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 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 we define the operation $+$ correctly in more general contexts, the solution to an equation $ax + b = 0$ need not be a number (for example, it may be a function, a matrix, a network, or a shape)!

So what types of properties of algebraic structures and algebraic equations do algebraist usually study? In short:

1. *Solving Equations*: Given an equation or system of equations, an algebraic wants to find solutions to the equation(s) (or determine that no solution exists). An algebraist will look for these solutions by purely "algebraic" means (for example, no calculus or tools ODEs would be used in solving an equation). When the objects of study are not numbers, these may get quite complicated.
2. *Finding properties of Algebraic Structures*: There is an incredible amount of different families of algebraic structures, each with varying properties. Some that shall be covered in this book are groups, rings, modules, fields, and algebras. 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) when is there infinitely many prime numbers?
   (c) when are objects "almost finite"

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 that the building blocks of an algebraic structure known as *finite groups* where classified (requiring over 10000 papers to be written)[2]. 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.

The general angle of attack in solving a problem algebraically is to find a way of inducing an algebraic structure given the mathematics the mathematician is working on. One of the main goals of abstract algebra is to understand the properties of algebraic structures and algebraic, and to use this understanding to solve problems.

(also add "models" systems, minus logic? Think of how you saw rings for a moment, and how you saw sheaves)

for Beginners

(incorporate this intuition of algebra: Algebra is about being able to break down your problem in pieces that we *know how to put it back together* through the use of several *functions*. Three examples that immediately come to mind is measure theory with measuring diagonal planes in $\mathbb{R}^n$, Fourier analysis and linear PDE's, and UFD's breaking down to irreducible numbers. The reason that "we know how to put it back together" is highlighted is because without that, we would just get the intuition behind a *topology*, and the use of functions tells us that we are working with *algebra*. For example, if it was through sequences, then we would be talking about *analysis* instead. This emphasis on the importance of how things combine also justifies later down the line why the union of subgroups/subrings/subspaces/etc is not a subgroup/subring/subspace/etc and why sums are the natural replacement)

(Also an eventual goal of establishing some confidence to try and attempt lemmas, theorems, or propositions before reading them. Near the end of the book, this will hopefully start manifesting itself in the readers.)

This book will be the study of algebra. The study of algebra is the study of algebraic structures. But what are algebraic structures? Algebraic structure are ...

Talk about "axioms"

Define the term "algebraic", and "structure"

Talk about proof types. necessary vs. sufficient. Existence problems. Classification problems. Relating examples/proofs/ideas.

some common terms like *a priori()*, *a fortiori*, *mutatis mutandis*, *clearly*, "The following are equivalent", *ad libitum*, etc.

Also talk about the further we go in the book, the more the motivation for the algebra will require more and more complex notion from other fields. They have interest in their own right in many ways, but they also "shine" the best when we know some more knowledge about these fields. I will try my

[2] this shall be explored further in section 4.7

best to make links between these fields and show the power that this is bringing, but sometimes, the only way to know the full power of the algebra is to know more of other fields. However, this will not be the case for most of the book. (perhaps some way of separating the book for this?)

So much explanation to not leave anything to chance/doubt. I want to make sure ppl understand.

(How a surprising amount of progress is made by a single result or illustrative example)

(perhaps a comment on how algebra might seem quite daunting to new-comers and a book of this magnitude is overwhelming. However, I think the fact that there is so much that can be said about a subject is simply fascinating, and once through the book the material will truly feel elementary and introductory.)

Advise on how to read the book

This book provides many proofs that are usually left out of an algebra textbook. This was done to have a source where the proofs are presented. However, it his *highly* encouraged that the reader not look at the proofs, but first try to work through them! For this reason, every proposition, lemma, and theorem, and corollary is will be marked with what I gauge as the difficulty of the proof. The reader then can at their discretion choose to work on it.

Any proof marked as easy are essentially mandatory exercise: they support the main definitions in such a way that if you don't know those proofs, you don't really know the full meaning of the definition.

Any proof marked as medium can be considered an exercise worth covering. They might not be as straight-forward as the easy proofs, but these will be the firs-time you encounter "non-trivial" problems in the discipline.

Any proof marked as hard is a proof that required either some creative leap, or decomposes the problem into many mini-proofs with some that require a creative leap. I still recommend you try these to get an idea; perhaps you'll get the first steps, and then reading the rest of the proof will make much more sense since you already saw how it started!

Preliminary Chapter

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 [3] – 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 called *naïve set theory*. We will be using these tools and question little why we are using these particular tools. The reason for naïve set theory is due to history. As set theory was in it's early stage of developments, these tools compromised all of set theory. However, there were a couple of important paradoxes which showed some inconsistencies in set theory [4]. From the existence of these paradoxes, a generation of mathematician's tried to free set theory from these problems [5]. 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

[3] With the exception of some very advanced logic, something that would be covered a logic class

[4] 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

[5] 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"

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: [6]

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

If naïve set theory is new to you, this section is going to go through a lot of terminology. If you're already comfortable with naïve set theory, you can look over this section if you want a refresher, or skip to ref:HERE for a different perspective than is usually presented for the material in section ref:HERE.

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. [7]. The one restriction we give to sets is that all the elements in a set must be different, in other 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 [8]) which are separated by a comma.

Example 0.1: 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:

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

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

[7] 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

[8] 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

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 our set with days of the week is the same if we re-arrange the days: In section ref:HERE, we study how to add an order to a set.

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

We will usually denote sets with capital letters (ex. $X = \{1, 2, 3\}$).

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 $\{1, 2, 3\} = \{3, 2, 1\}$ [9]. 2-tuples are sometimes called a *pair*. If it is clear from context, it is a common convention to drop the number infront 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)[10]. To write "a in A" in mathematical notation, we use the symbol $\in$. For example:

$$a \in A$$

[9] 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 patten follows for n-tuples

[10] 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

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".

Concept Check Is

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

answer:[a]

[a] flip this text? No, the set containing the single element$\{1, 2, 3\}$ is inside the above set

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[11]) means the two symbols represent the same thing. For example, the number one-half and two-quarters are equal:

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

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 [12]), we do the same convention as with $\in$ and put a slash across the symbol: $\frac{1}{2} \neq \frac{1}{3}$.

Example 0.2: 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

[11] 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

[12] 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)

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

(footnote about if we don't specify the set from which we're restricting we're restricting from "the set of all elements"? Is this true? How would that work?)

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

and is read "s in S such that [s satisfies the property P]".

Example 0.3: 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

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 0.1.1: 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 0.1.2: 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 (image of number line here?). 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 8.1

[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 $\mathbb{N}_{\frac{5}{2}+} = \left\{x \in \mathbb{N} \;\middle|\; x > \frac{5}{2}\right\}$ and $\mathbb{N}_{\frac{5}{2}-} = \left\{x \in \mathbb{N} \;\middle|\; x < \frac{5}{2}\right\}$, 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 $\mathbb{Q}_{\sqrt{2}^+} = \{x \in \mathbb{Q} \mid x > \sqrt{2}\}$ and $\mathbb{Q}_{\sqrt{2}^-} = \{x \in \mathbb{Q} \mid x < \sqrt{2}\}$ 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. [13]

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[14] . 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), 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 (usually a capital letter) 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. Propositions are most useful when we can pass in a variable in a proposition as though it's a function which evaluates to either true or false[15]

(don't know where to move these yet) (Maybe add somewhere how $a \in A$ can be thought of as the sentence $\exists x \in A$ such that $x = a$ is true?) (explain why it tries to get rid of ambiguity. For example, if we do not carefully define proposition, have the proposition P such that P is an interesting number. What does "interesting number" mean?)

(To avoid ambiguity, we will stick to statements which can be labeled as either True or False. The most explicit such sentence have value already in it, for example "2 is an even number" is true, while "2 is an odd number" is false. To expand our ability, we allow for variable, so we can write "x is an even number" and determine it's truth value given an [integer] value of x)

Example 0.4: 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", and then have $R(x) =$"$P(x)$, then $Q(x)$".

(Propositions of the form if $P(x)$ 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 $P(x) \Rightarrow Q(x)$,

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

[14] 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

[15] Say a quick word about how there are areas of math that try to generalize this that study *Topoi*

then $Q(x) \Rightarrow P(x)$. This need not be the case. For example, if we take $R^{-1}(x) = Q(x) \Rightarrow P(x)$ (better example here..)

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).

(proposition can also take in multiple variables, i.e. $P(x, y)$)

Example 0.5: Propositions 2

1. here. Give implications, if and only if, other such things

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 (about rainy weather? Show how you *cannot* give any conclusions. The famous "king of France is not wearing pants" example. Say there are multiple interpretations, one of them being vacuous truth. This is the one mathematicians took, and explain why.)

(thus, (False $\Rightarrow$ True) is True, and (False $\Rightarrow$ False) is *also* True)

Logical Quantifiers

We will now explore how to create more complicated properties by adding to our tool belt 5 "words" from which we can construct mathematical sentences:

1. $\forall$: the forall quantifier. This will mean we
2. $\exists$: the *existence* quantifier
3. $\exists!$ the *unique existence quantifier*

(some examples and exercises)

(A word on $\vee$ $\wedge$, but we'll mainly avoid these symbols for ease of reading)

(convention with set-builder notation where exists is assumed)

(table for $\rightarrow$ and explanation of $\Leftrightarrow$. Convention in definition when saying *if* meaning *iff*)

(A word on $|A|$, i.e., the cardinality of A)

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

1. subset:
$$A \subseteq B \Leftrightarrow (x \in A \Rightarrow x \in B)$$
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. Intersecting two sets:
$$A \cap B := \{x \mid x \in A \text{ and } x \in B\}$$
3. Union of two sets:
$$A \cup B := \{x \mid x \in A \text{ or } x \in B\}$$
if $A \cap B = \emptyset$, we say that A and B are *disjoint*
4. difference between sets:
$$A \setminus B := \{x \mid x \in A \text{ and } x \notin B\}$$
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.
5. Let A be a set. Then $P(A)$ is the set of all possible subsets of A. Note that if A is finite, then$|P(A)| = 2^{|A|}$ [16]
6. (Introduce tuples at some point?)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). For example
$$\{1, 2\} \times \{a, b\} = \{(1, a), (2, a), (1, b), (2, b)\}$$
and
$$\{a, b\} \times \{1, 2\} = \{(a, 1), (b, 1), (a, 2), (b, 2)\}$$
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:
$$\prod_{i=1}^{n} X_i$$

(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)

[16] This is even true when A is infinite, but we'd have to be more careful with how we define $2^{|A|}$, see exercise ref:HERE

0.1.2 Proving Propositions

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 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)

:exercise

I don't know where to put this exercise; it's not at all algebra. It's simply a cool exercise showing that if a series is divergent, you have to be careful because you can run into contradiction like $1 = 0$

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

0.1.3 Induction

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

0.1.4 *Pigeon-hole Principle

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

0.1.5 Equivalence Relations

(What does it mean two things to be equal? Fun to think about philosophically, and we'll give an answer to it here. There are actually many ways, depending on what you want to be "equal". Give different real world examples, there must be many non-mathematical ones. Then move on to the idea of relations. Say that one of the main relation is the equivalence relation, which will for us capture a generalization of the notation of "equality") 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 and odd numbers (more words here)

(define what a relation is, in particular if X has a relation $\sim$ and $x \sim y$, then that means that $(x, y) \in \sim \subseteq X \times X$)

(take not we're using 2-tuples, i.e. pairs, not sets with 2 elements)

Definition 0.1.3: [Binary] Relation

Let X be a set. Then Any subset $R \subseteq X \times X$ 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. $n > 2$), and make the sets different (ex. $X_1 \times X_2 \times \cdots X_n$)

We shall almost always think of relations

Example 0.6: Relations
here

A common relation to put on objects give a sense of bigger and smaller object:

DEFINE ORDER

Example 0.7: Orders
Let $\mathbb{N}^2 = \mathbb{N} \times \mathbb{N}$

1. 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

Look at Aluffi, he has an interesting build-up for equivalence relations. Also want to add more about order because of the 4th isomorphism theorems, Galois relations (ex. Fund thm of Gal theory and Hilbert Nullstellensatz), ordered fields, the frequent use of Zariski's lemma, and maximal normal subgroups or maximal ideals.

Definition 0.1.4: 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 0.8: 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. lots of cool example from Wikipedia that are beyond math like "has the same birthday"

(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"

(partitioning property. Like creating the "new objects" where we ignore the "details" that is "consumed" by the equivalence relation)

(natural projection?? Give $\pi : G \to \{0\}$ a natural projection that always exists?)

(should these be included somewhere? lattices Or perhaps under graphs?)

0.1.6 Functions

(common abuse of notation: f^{-1} both means pre-image and inverse, i.e., it's a function of the form $P(Y) \to P(X)$ if looking for pre-image, and it is of the form $Y \to X$ if looking for inverse)

Most of the time, we work with a very special type of relation: Given two sets X and Y, we want to relate each element of X to a single element of y. We call this relation a *function*.

(some intuition on what a function is supposed to represent before introducing the formal definition)

Definition 0.1.5: Graph of a Function and Functions

Let X and Y be sets. Then a *graph of a function* is a set $\Gamma = \{(x, y) \mid (x, y) \in X \times Y\}$ such that

1. For every $x \in X$, there exists an element $(x, y) \in \Gamma$
2. 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 $y, y' \in Y$

A *function* is defined to be a notation representation of the graph of a function. If Γ_f is the graph of a function and $(x, y) \in \Gamma_f$, then we write

$$f(x) = y$$

Function are also sometime called maps or mappings. Sometimes, we want to define a function, but don't need to reference to it again. Mathematicians sometimes like to use the following notation:

$$x \mapsto y$$

instead of $f(x) = y$

Example 0.9: Example

1. Give example of injective surjective and bijective functions
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 $a, b, c \in \mathbb{R}$

Proposition 0.1.1: Composition of functions

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

1. Define $(g \circ f)$ to be the function where $(g \circ f)(x) = g(f(x))$. Then $(g \circ f)$ is called *function composition*. $(g \circ f)$ is a function
2. $h \circ (g \circ f) = (h \circ g) \circ f$

Proof **:**
here

1. Familiarity with functions like e^x, $\log(x)$, and polynomials.
2. Knowing some properties of functions is very useful (ex. function composition, associativity of function composition, injective, surjective, injective then exists left inverse, surjective then exists right inverse. Surjective feels like you can somehow "be ahead of time", dk if that intuition is worth remembering)
3. Function are special relations.

Definition 0.1.6: Injective

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

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

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 0.1.2: Injective And Left Inverse

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

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

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

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

Definition 0.1.7: Surjective

Let $f : X \to Y$ be a function. Then f is *surjective* or a *surjection* if for all $y \in Y$, there exists an x such that $f(x) = y$, i.e. $f(X) = Y$.

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, in a sense acting as "representatives" for y. 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 0.1.3: Surjective And Right Inverse

Let $f : X \to Y$ be a surjective function. Then there exists a $r : Y \to X$ 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 $y \in Y$ 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 be seeing that every surjection defines an equivalence class:

Proposition 0.1.4: 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:

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

***Proof* :**
The 3 axioms of an equivalence class should be checked.

Thus, every $y \in Y$ is associated with an equivalence class $[x] \in X/\sim_f$.

(In fact, functions are special equivalence relations)

(how to decompose a function using equivalence relations)

$$A \twoheadrightarrow (A/\sim) \to \mathrm{im}(f) \hookrightarrow f(G)$$

(this means that when thinking of a function, you can think of it defining an equivalence relation, and when defining an equivalence relation, we can think of a function. In this way, a function is well-defined if the equivalence relation holds, that is, $a = b \Rightarrow f(a) = f(b)$ - I think)

(universal property here? – mention f_* is injective)

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 [17]. 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

Let $f : X \to Y$ and take $A \subseteq X$, $B \subseteq Y$. Show that $f^{-1}(A) \subseteq B \;\Leftrightarrow\; A \subseteq f(B)$

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?

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?

0.1.7 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 0.10: 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.

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

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 [18], and only near the end of the book do we encounter topological structures (p. ref:HERE) where it is assumed the reader has taken a course covering point-set topology[19]. 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?)

0.1.8 *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?)

[18] see section 3.4

[19] 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 the beginning of mathematics. In this section, the basic properties of numbers, like the basic operations of arithmetic's $(+, -, \cdot, \div)$, prime factorization, the order of numbers, and so froth, will be explored and expanded to explain them in more technical language to more easily convert this knowledge into abstract algebra. 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). (History of how these were symbolized, arriving at $\mathbb{N}$, or $\mathbb{N} \cup \{0\}$. Debt probably made us want to have negative numbers, so $\mathbb{Z}$. Realizing we have propositions made us want to have $\mathbb{Q}$. For awhile, it wasn't known if there was anything beyond $\mathbb{Q}$. There was this other concept called the continuum and was sometimes labeled c and called the real numbers $\mathbb{R}$. Eventually $\sqrt{2}$ was proven not to be there, and so $\sqrt{2} \in \mathbb{R} \setminus \mathbb{Q}$. All of these share some similar properties. The $\mathbb{C}$ are there for more elaborate reasons and are really called numbers for historic reasons)

(Nowadays, we would think of the set $\mathbb{N}$ be some countable quantity labeled as $\{0, 1, 2, ...\}$ with a function that takes in two elements of $\mathbb{N}$ and spits out another. Such a function is called a *binary operation.*)

(, elaborate on adding $+$ and $\cdot$, elaborate on how we also usually think of numbers having an order $\leq$)

(perhaps footnote that $\mathbb{N}$ is constructed through the Peano Axioms? We will show in this book how to construct $\mathbb{Z}$ from $\mathbb{N}$, and $\mathbb{Q}$ from $\mathbb{Z}$. We will skip the construction of $\mathbb{R}$ from $\mathbb{Q}$ since there is no algebraic construction of $\mathbb{R}$ from $\mathbb{Q}$ (as we will prove in appendix ref:HERE (Löwenheim-Skolem's Theorem)). Near the end (chapter 18), we will show how to construct $\mathbb{C}$ from $\mathbb{R}$. So we will have this chain

$$\mathbb{N} \rightsquigarrow \mathbb{Z} \rightsquigarrow \mathbb{Q} \overset{!}{\rightsquigarrow} \mathbb{R} \rightsquigarrow \mathbb{C}$$

where we will skip $\overset{!}{\rightsquigarrow}$)

(Elaborate on $\mathbb{C}$. Intuitively, you can think of it as adding an element i to $\mathbb{R}$ such that $i^2 = -1$. Notice there is no element $x \in \mathbb{R}$ such that $x^2 = -1$, so really are adding a new element. Then we define $\mathbb{C}$ to have element of the form $a + bi$ where $a, b \in \mathbb{R}$ and bi is just a new element itself)

(Sum and Product rule? Premutation and Combinations? Binomial Coefficients?)

(More can be found books on number theory HERE)

0.2.1 GCD and LCM

(define $a|b$ as a divides b, or a is a divisor for b if there exists a k such that $ak = b$)

(some easy immediate exercises: if $x\big|n_1$ and $x\big|n_2$, then $x\big|an_1 + bn_2$ for any $a, b \in \mathbb{Z}$. Also, if $a\big|b$ and $b\big|c$ then $a\big|c$) In this section, we expand on some basic number theory. we start by recalling that any integer $x \in \mathbb{Z}$ can be broken don into its prime factorization, that is, if $n \in \mathbb{Z}$ ($n \notin \{0, 1\}$), then

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

here p_i are distinct primes numbers and $n_i > 0$. Can you prove this using induction (or strong induction)?

Since all numbers can be broken don into multiples of primes, we can define a ay to compare ho similar to numbers are:

Definition 0.2.1: 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:

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

Similarly, we have the dual concept:

Definition 0.2.2: 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

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

To be able to find these values, Euclid has invented a ay to do so:

Theorem 0.2.1: 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 **:**
here

Example 0.11
Let $a = 20$, $b = 13$. Then:

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

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

Theorem 0.2.2: Extended Euclidean Algorithm

The Euclidean Algorithm

Proof **:**
here

Theorem 0.2.3: Bezout Identity

The Euclidean Algorithm

Proof **:**
here

Bezout'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 Bezout 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

(Basic mod arithmetic's? Some mod identities?)

1. The basic properties of $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ (ex. $3 < 4$ or $\forall x \in \mathbb{Q}$, there exists a y such that $xy = yx = 1$)
   - Extra properties of $\mathbb{R}$ and $\mathbb{C}$ (like uncountability, or $\|z\|$) will be brought up simply to use in more advanced examples
   - The fact that every number can be decomposed into prime-numbers is very important. Knowing why this decomposition is unique is also useful
2. The notion of finite and countably infinite. An understanding of uncountability is not required[20]
3. Least common multiple and Greatest common divisor, which I'll label lcm and gcd, though sometimes, I'll do (a, b) for $gcd(a, b)$ when it's not ambiguous
4. Modular arithmetic's. I will use the $\mathbb{Z}/n\mathbb{Z}$ notation for the set of elements $a \mod (n)$
5. some basic number theory like Bezout identity and Euclidean algorithm?

0.2.3 Counting Problems

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

[20] Uncountability is more useful in domains that heavily rely on topology

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 0.12: 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 0.13: examples of Graph Isomorphisms
here

These will come back in a exercise ref:HERE

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 [21]. 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. [22]
>
> 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

[21] Semantics is the study of the meaning that sentences (or any discourse) carries
[22] 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 [23]. 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 hat 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 27.

(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)

[23] 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 0.4.1: 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 14.2. 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 behind thms is:[

Some words on system of equations? Why matrix multiplication is defined to "represent" linear transformation in the since that

$$\mathcal{F}(AB) = \mathcal{F}(A) \circ \mathcal{F}B$$

We will show this much later. For now the definition is left as a reference HERE)

(Some idea on invertible matrices too)

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.)

(Also say how we are focusing on *finite* groups. The realm of infinite group is in fact much harder, and we will cover some of it in appendix D,)

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: 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 1.2: 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 3.3.2)
4. If $(A, \cdot)$ and $(B, \bullet)$ 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:
$$(a, b) * (c, d) = (a \cdot c, b \bullet d)$$

We can denote this group and its binary operation $(A \times B, (\cdot, \bullet))$. 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 6.

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 (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. 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:

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

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 8.1

[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 26.1

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. 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 numbers 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 $H \preceq G$

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

[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

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 1.3: 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. 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 1.4: 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 1.2 were abelian groups. We haven't yet introduced groups which are not abelian. The first such group will be the Dihedral group (definition 1.2.2)

[a] This is an example where the binary operation is implied. The binary operations shown in example 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

[2] for reference, see definition 4.1.2, 3.2.3, and section 6.3, 7.1, and 7.1

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]. Some other mathematician's are starting to call the study of something called "linear operators" in functional analysis as *non-commutative geometry*. All in all, the importance of this axiom cannot be understated.

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 1.5: 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

$\times$	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.

Exercise 1.1.1

1. Let $G = \mathbb{Q}$, and the binary operation be average : $\mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$, where $\text{average}(x, y) = \frac{x+y}{2}$. 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] 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 27

3. (If you know some complex analysis) In example 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 $\mathrm{Bij}(X)$ be the set of all bijections from X to X. Prove that $\mathrm{Bij}(X)$ is a group. In section 1.2.3, we will relabel this as $\mathrm{Sym}(X)$.
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 1.1.1: 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, 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 1.1.1: 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 $a * a^{-1} = e$. The proof works *mutandis mudandis* if we start with $a^{-1} * a = e$. Then

$$a * a^{-1} = e = e * e = (a * a^{-1}) * (a * a^{-1})$$

Next, we take full advantage of associativity:

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

multiplying on the left by a^{-1}, we get:

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

Simplifying the Left hand side, we get $a^{-1} * (a * a^{-1}) = (a^{-1} * a) * a^{-1} = e * a^{-1} = a^{-1}$ now

multiplying on the right side by a, we get:

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

wait.. there's a mistake..

Notice that $(a * a^{-1}) * (a * a^{-1}) = a * ((a^{-1}) * (a * a^{-1})) = ((a * a^{-1}) * a) * a^{-1}$ (explain this by letting different elements be considered part of he group). Then we can drop the parenthesis and get

$$e = a * (a^{-1} * a) * a^{-1}$$

then, multiplying by a^{-1} on the left and a on the right, we get:

$$a^{-1}a = aa^{-1}$$

as we sought to show

Proposition 1.1.2: 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: [4]

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

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 $0a = 0$ [5]. 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 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:

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:

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

[5] If you have covered rings, this might look familiar to the first property in proposition 9.1.1. 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 9.1.1: $0a = (1-1)a = a - a = 0$

Proposition 1.1.3: Cancellation Property: Unique solution to linear equations

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 [6] 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

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

structure [7] or drop the condition of associativity[8]. We will briefly see an alternative to associativity for an algebraic structure we will cover in part VII, chapter 26, but structures without associativity are outside the scope of this book[9]. 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 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 groups[10] which are commutative in chapter 6.1.1

[7] 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 $a \to (b \to c)$

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

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

[10] Technically, we will list all possible groups which are *finitely generated*, which is a concept we will introduce in chapter 3.3. I will address what this means once we reach theorem 6.1.1

: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 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 Cyclic Groups

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. Notice that 12 is the identity: for any time on the clock-face, adding 12 hours will bring you back to the same clock-face. Similar results appear for associativity and inverse elements. One can intuitively check that a clock forms a group.

The clock face follows the operation of $\mathbb{Z}/12\mathbb{Z}$. More generally, modular arithmetic's as seen in section 0.2.2, all form groups:

(definition here)

The fact that _____forms a group all comes from the properties of modular arithmetic's explored in section 0.2.2. One great consequence of this fact is that we can always find a subgroup of a group G that is the same as (a modular arithmetic's group TBD)[11]:

Proposition 1.2.1: 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

[11] A precise notion of "the same" is formulated in section 1.4

***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.

This subgroup has a special name, it's called a *cyclic subgroup*. Given any element $x \in G$, we can form a cyclic 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 1.6: 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

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

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 $x^a \in \langle x \rangle$, $a \leq k$.

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

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:

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

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 1.2.1: 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 1.7: 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 shown in example 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

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

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:

$$(a_1, a_2, a_3, ...) \cdot (b_1, b_2, b_3, ...) = (a_1b_1, a_2b_2, a_3b_3, ...)$$

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 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. Each of these shapes can be rotated or reflected in such a way that we get back the same shape. For example, if we choose a square:

and decided to rotate 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 (put a nice rotation symbol between the squares, like a curved arrow which say 90 (symbol for degrees) under it)

if we were to rotate it an eighth of a turn, we would not get the exact shape back:

IMG HERE

Let's label quarter turns to the right r (r for ***r**otate*).

We know introduce our first non-abelian group: the dihedral groups. The Dihedral group

The first intuition about this groups is by thinking about n-gons (or symmetric geometric objects) on a 2d surface, and consider any rotation or reflection on that makes the n-gon land exactly on itself:

(Also give the hand example)

Definition 1.2.2: Dihedral Groups

Let D_n denote the set of symmetries of a regular n-gon, where a symmetry is any rigid motion [a] of the n-gon which can be effected by taking a copy of the n-gon, moving this copy in any fashion in 3-space and then placing the copy back on the original n-gon so it exactly covers it We can describe each symmetry as a permutation, where the set $\{1, 2, \cdots, n\}$ describes the vertices of the n-gon, and the permutation describes where they end up.
Now, To make D_n into a group, by defining the operation between $s, t \in D_n$ to be the symmetry obtained by first applying t then s to the n-gon.

[a] Any function which doesn't stretch or squish the shape

This definition might seem extremely unsatisfactory, so let' prove that it' indeed a group:

Proof **:**
here

Even with this definition being a group, is there a better way to define it? There is; ha proper definition will come when we cover group presentations.

A few important properties of Dihedral groups:

$$|D_n| = 2n$$

Some authors like to put the order of the dihedral group in the subscript, that is they write D_n. 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.

This proof was a heavy leaning on the intuition behind reflections and rotations, I.e. for any n-gon, there are n rotations, and n axis of symmetry.

Proposition 1.2.2: 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)

Solution very intuition heavy ▼

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.

Exercise 1.2.2

1. (If you know norms and inner products) In this exercise, we'll define D_4 using properties of norms and inner products. Recall that an *isometry* on $\mathbb{R}^n$ is defined as a function $f : \mathbb{R}^n \to \mathbb{R}^n$ such that
$$\|f(x) - f(y)\| = \|x - y\| \qquad \forall x, y \in \mathbb{R}^n$$
    1. show that f is injective
    2. Show that $g(x) := f(x) - f(0)$ is *linear*. Deduce that $f(x)$ is an *affine* function[14] and conclude that f is bijective (hint, show that $\langle g(x), g(y)\rangle = \langle x, y\rangle$, and use this fact to show $g(x+y) = g(x) + g(y)$ and $g(ax) = ag(x)$ for all $a \in \mathbb{R}$).
    3. Let $A \subseteq \mathbb{R}^n$ be a set. Show that the collection of all functions $f : \mathbb{R}^n \to \mathbb{R}^n$ such that $f(A) = A$ forms a group. This groups is called the *groups of symmetries*.
    4. If $A = \{(-1,-1), (1,-1), (-1,1), (1,1)\}$, show that the group of symmetries is D_8.

[14] an affine function is a linear if we translate the function appropriately so that 0 maps to 0

1.2.3 Symmetric Groups

(S_n useful because if $|S| = |T|$ then $\mathrm{Sym}(S) \cong \mathrm{Sym}(T)$, so that gives us that information. However, $\mathrm{Sym}(S)$ can be easier to remember that we're just working with bijections, so in that sense it's "lower level". On the other hand $\mathrm{Aut}_{\mathbf{Set}}(S)$ reminds us that this is the set of isomorphisms in the category of **Set**)

(Mention how this group was the "original" group. When group theory started, it was the study of symmetric groups. So, if you wonder where the axioms of a group come from, they came from a generalization of these conditions. We will see later on that these axioms [will make all groups subgroups of permutation groups when we cover group actions TBBD)

Symmetric groups are one of the more important groups to get a handle on – they will show up extensively when covering group-actions. At this stage, I think the best way to think about symmetric groups is in the framework of thinking about groups as capturing different types of symmetry. If you take a cube, for example, and consider it's rotational symmetries, you can think of 24 different rotational symmetries which will leave the cube looking like a cube. If you also allow for reflections, that number will double to 48 (in fact, you just described D_{48}). If you further allow all faces of the cube to freely rotate, then you will get 188743680 symmetries.

Notice that when thinking about the symmetries, we were inherently limited by the structure we were trying to be symmetrical on. What if we try to find the symmetries of the most generic type of structure? One's first guess might be a set, but there's only one symmetry on a set since $\{a, b\} = \{b, a\}$. Thus, we also make order matter, we get a tuple, since a set would just be points arbitrarily relation between points of interest with no additional structure.

In a sense, a cube can also be represented as a tuple, but there is some extra structure on it (certain vertices are in relation with vertices edges, faces, etc., limiting what tuples we can form). So, what the symmetric group *is* is the group that acts on these tuples, making it in a sense the most generic group. This intuition will in fact be a rigorous concept once we reach Cayley's theorem (5.4.2)

Let N be a set. For now, let it be a finite set: countable an uncountable sets will be covered another time. If we take a bijective function $\sigma : N \to N$, then since it's one-to-one and onto, every element from the domain must uniquely map to some element in the co-domain. If we label the elements of N:

$$(1, 2, 3, 4, 5, 6, 7, 8, 9, ...)$$

then we after the σ permutes, we might get

$$(\sigma(1), \sigma(2), \sigma(3)\sigma(4), \sigma(5), \sigma(6), \sigma(7), \sigma(8), \sigma(9), ...) = (2, 4, 1, 3, 7, 9, 8, 5, 6, ...)$$

Note that since we're with function composition, it is already associative. The identity would just be the function that maps elements to themselves ($\sigma(1) = 1$), and since it's bijective, there exists an inverse. Hence, it is a group:

Definition 1.2.3: Symmetric Group

Let N be a set. Let S_N the set containing all bijective functions between the set and itself. Then S_N is a group with $\circ$ (i.e. function composition) as it's operation

Note that we can relabel the elements of N to $\{1, 2, 3..., n\}$, so we usually write S_n, where n represents the number of elements.

The example given above was a particular permutation. From stats, you'd know if $|A| = n$, then there is $n!$ total different ways of permuting the variables. Note that if we have two permutations, combining two permutation gives another permutation. Since it takes in two elements, then $\circ$ representing composition of functions is in fact a binary operation. Function operation is also automatically associative. There is also a notion of an identity function, that is, the permutation which doesn't move any of the elements ($id(a) = a$ for all $a \in A$). Since the function is bijective, it has a two-sided inverse, meaning every element has an inverse. Therefore, the set of all the permutations on a set is a group! Since what really matters is the cardinality of the group, since the elements can be relabeled, then if $|A| = n$, we can write S_n be represent the set of all permutations of A.

It is cumbersome to write-out the permutation by saying what each value of $\sigma \in S_n$ maps to. Instead, we'll write it like so

Figure 1.1: shortening notation for elements of S_n

This is a shorthand for writing out some $\sigma \in S_n$, by algorithmically finding the "cycles". In this case, we didn't have a case where an element didn't move (ex. $4 \mapsto 4$). If that were to happen, we'd simply not write down that element anywhere since (4) would not add any more information than leaving it out (we'd get the same amount of information either way).

Definition 1.2.4: Cycle (Cycle-Representation)

Given $\sigma \in S_n$, a cycle is the "chain" of numbers that link together till you get back to the beginning.
If we have a single cycle, we call it an n-cycle, where n represents the length of the cycle

Example 1.8: Cycle Representation
As we've already shown, $(1\,2\,4\,3)(5\,7\,8)(6\,9)$ is the cycle representation of the function I've outlined in the image.

This cycle is composed of a 4-cycle, a 3-cycle, and a 2-cycle.

We will define one more thing, since the idea come up so often

Definition 1.2.5: Transposition

2-cycles are usually referred to as ***transpositions***

The reasoning should be intuitive: transpositions simply transpose two elements ($a \mapsto b$, $b \mapsto 1$)

Note that when taking the product of two permutations when looking at the cycle form of σ, then since the operation is function composition, you "combine" the cycles from right to left

$$(1\,4\,5\,2)(4\,1\,3) = (1\,3\,5\,2)$$

The trick I did here was to start at 4, then noticed it goes to 1, but it doesn't stop at 1, cause 1 then goes to 4, so you get $4 \mapsto 4$, so you ignore it since it maps to itself. Next, I took 1. 1 maps to 3, then 3 maps to 3 in the next cycle, so I wrote $(1\,3$ to start a new cycle. next I see where 3 goes. It first go to 4, but 4 goes to 5, so $3 \mapsto 5$ giving us $(1\,3\,5$. Next, 5 goes to 5 in the first cycles, then 5 goes to 2 in the next, so $5 \mapsto 2$ giving us $(1\,3\,5\,2$. Next, we get that 2 maps to 2, then 2 maps to 1, so $2 \mapsto 1$. Notice that 1 is at the beginning of our cycle, so we'll close our parenthesis to show the cycle has finished $(1\,3\,5\,2)$.

Note that if we have $(1\,2)(3\,4)$, then you can easily check that $(3\,4)(1\,2)$ represents the identical permutation, and thus, *disjoint cycles commute*.

This might still work in the future: Programs to take the power of cycles and find cycles:

`http://webspace.ship.edu/jehamb/flash/cycleNotation.html`

`http://bluetulip.org/2014/programs/permutation.html?inputboxy=%281%2C2%29&resultbox=`

Note that that is not the unique way of writing down the cycle. In this case, they were all disjoint. You *can* write it down in terms of non-disjoint cycle, though it would be messier. The disjoint cycles are usually called n-cycles, where n represents the length of the cycle. If an n-cycles is of length 2 (i.e. a 2-cycle), it is also called a *transposition*.

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 1.9: : Interesting Exercises on Commutativity

We earlier mentioned how being commutative is a nice property, and that groups which are commutative are quite nice. We also said 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

Solution Try composing a n-cycles and a 2-cycle which are not disjoint form the left and right and see what you get ▼

With those extra properties of cycles, we will name one more concept for ease of labelling cycles:

Definition 1.2.6: 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 1.10: 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.

Also, you can algorithmically decompose any n-cycle into the product of transpositions. Let's take

$$(1\,2\,4\,3)$$

as in the image. Then

$$(1\,3)(1\,4)(1\,2) = (1\,2\,4\,3)$$

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 1.2.1: 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 **:**
I'll put the proof here once my homework is graded.

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 We can thus name certain properties of cycles based of off this. The first important one is something called *parity*

Definition 1.2.7: 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
Sometimes, 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 category of groups based around parity:

Definition 1.2.8: 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.

Properties of Symmetric Groups

These are properties that come up a lot in proofs. Not all of these are yet provable, but as the document goes on, we will fill in some of the gaps as we get the right tools to analyze them. One can attempt to solve some of these now if they want some practice with computing symmetric groups, or perhaps figure out the key insights.

Cardinality: $|S_n| = n!$ (think of permutations to get this number)

Order of an Element

Proposition 1.2.3: Order of element is p if and only if 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 **:**
Homework 5 first question

Note that if p is not prime, the "if and only if" is not necessarily true.

Example 1.11: : 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 $\text{lcm}(\sigma, \tau) > n$.

Proposition 1.2.4: 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 1.2.3, and since we're taking $\langle \sigma \rangle$, we'll use theorem 3.3.2 (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 1.2.3, 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 (1.2.3), 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 1.2.5: 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 $\tau = (a\,c)$, $\sigma\tau\sigma^{-1} = (\sigma(a), \sigma(b))$).

You can finish this proof, but it is long. It is easier to prove once your introduced to group action and conjugacy classes (see 5.4.1).

Generators

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:
>
> $$(a_1\, a_2 \ldots a_k\, n)$$
>
> 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.

Sub-Groups

Here are a couple of tricks I picked up while practicing. Note that some of these require information from further sections, but I don't know where to put them yet, so they'll go here:

1. One way to find subgroups is by looking at what you get by conjugating elements, since that moves things around in cycles in a nicely predictable way.
2. Next, if you have some subset $A \subseteq S_n$, and your trying to figure out properties of $\langle A \rangle$, try first figuring out the order of elements, since by Lagrange (4.3.1) you'd now that if $|x| = 3$, then $|x| \big| \langle A \rangle$. so for example, if you have $\langle (1\,2)(1\,2\,3\,4) \rangle$, you can notice that you cannot create an element of order 3. since $|S_4| = 4! = 24$, then you know that a subgroup would have order $1, 2, 3, 4, 6, 8, 12$. Since there is an element of order 4, there is at least 4 elements, plus the element $(1\,3)$, and we get at least 5 elements. Then, if we eliminate 6 and 12 (By Sylow's theorem eventually). So we're left with only 8. Since we're left with 8, then the usual suspect to try is D_8. Again, just need to think about the generators to check isomorphism, so let $r = (1\,2\,3\,4)$ and $\sigma = (1\,3)$. You need to check that $r^4 = 1, \sigma^2 = 1, rs = sr^{-1}$
3. A subgroup that comes up all the time is A_n. Counting comes up here and there when dealing with A_n.

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:

$$i^2 = j^2 = k^2 = -1 \tag{1.1}$$
$$1 \cdot a = 1 \cdot a \qquad \forall a \in Q_8 \tag{1.2}$$
$$ij = k \qquad ji = -k \tag{1.3}$$
$$ik = j \qquad ki = -j \tag{1.4}$$
$$jk = i \qquad kj = i \tag{1.5}$$

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 1.12: 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 1.2: 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 1.3: Q_8 Multiplication Table

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 1.2.3

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 section 0.5 (however, some of these collection of objects goes beyond that section). Let's first recall the notation we use to describe the collection of matrices:

Definition 1.2.9: Matrix Group Under Addition

The set $M_{n\times m}(\mathbb{F})$ forms a group under matrix addition

However, this group is not actually that interesting: it's in a sense equivalent to $\mathbb{F} \times \mathbb{F} \cdots \mathbb{F}$ $n \cdot m$ times [15]. The More interesting binary operation is multiplication, since it combines elements in much more interesting ways which you might still have an intuition form MAT223/24/40/47. As you might recall, Matrix multiplication requires dimensions to line up properly, so for associativity to hold, we'll have the dimensions be the same [16]. 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. As you probably have guessed, this implies that $\det(A) \neq 0$ for every matrix [17]. Such matrices are given a special name:

[15] We can actually show this once we get to Isomorphisms (definition 1.4.4)

[16] For future me: I believe, though haven't yet checked, that we might be able to define Groupoids on $n \times m$ matrices

[17] If this feels strange, a quick review of linear algebra notes here would be a good idea, since this group is used often, especially when rings are introduced.

Definition 1.2.10: General Linear Group

$$GL_n(F) = (\{A|A \text{ is an } n \times n \text{ matrix with entries from } F \text{ and } \det(A) \neq 0)\}$$

That is, the set of all matrices with non-zero determinant. When dealing with this group, you can use your linear algebra knowledge! For example, if $\mathbb{F} = \mathbb{C}$, then every matrix is diagonalizable, so

$$\forall A \in GL_n(\mathbb{C}), \exists B \in GL_n(\mathbb{C}) \text{ such that } BAB^{-1} = C$$

where C is the matrix with the eigenvalues of A. More generally, you know *any* square matrix has a Jordan Canonical Form. These ideas will be important when discussing conjugacy classes (5.4.1).

We will also name 3 special subgroup of GL_n since it comes up often enough:

Definition 1.2.11: Special Linear Group

$$SL_n(\mathbb{F}) = (\{A|A \text{ is an } n \times n \text{ matrix with entries from } F \text{ and } \det(A) = 1)\}$$

Definition 1.2.12: Orthogonal Group

$$O_n(\mathbb{F}) = \big(\{A|A^tA = AA^t = I\big)\}$$

In this group, the transformations are *distance-preserving* (you can think of the *unitary* or *orthogonal* operators in MAT247)

Definition 1.2.13: Special Orthogonal Group

$$SO_n(\mathbb{F}) = \{A|A^tA = AA^t = I\,, \det(A) = 1\}$$

If the underlying field is clear, we sometimes write $SO(n)$.

In this group, not only is distance preserved, but so is scale (think about rigid motions, $\|f(x)-f(y)\| = \|x-y\|$). This group will be used to generalize our notation of rotations of polygons to rotations of any objects in any dimensions.

Of all of these, all but $O_n(\mathbb{F})$ will appear commonly in this document

Exercise 1.2.4

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 14.3.4)

Proposition 1.2.6: 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 were 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.

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 inversible 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 3.3.3), 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 1.13: 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[18]. 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).

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 1.14: 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

[18] With only one exception. When a set of generators has no relationships, we call the group generated by that set a *free group*

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 1.2.3 and 3.3.3, 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 1.15: Insoluble Presentation
The group representation by this monstrous presentation:

$$\begin{array}{llll}
\langle & a, b, c, d, e, p, q, r, t, k & | & \\
& p^{10}a = ap, & pacqr = rpcaq, & ra = ar, \\
& p^{10}b = bp, & p^2adq^2r = rp^2daq^2, & rb = br, \\
& p^{10}c = cp, & p^3bcq^3r = rp^3cbq^3, & rc = cr, \\
& p^{10}d = dp, & p^4bdq^4r = rp^4dbq^4, & rd = dr, \\
& p^{10}e = ep, & p^5ceq^5r = rp^5ecaq^5, & re = er, \\
& aq^{10} = qa, & p^6deq^6r = rp^6edbq^6, & pt = tp, \\
& bq^{10} = qb, & p^7cdcq^7r = rp^7cdceq^7, & qt = tq, \\
& cq^{10} = qc, & p^8ca^3q^8r = rp^8a^3q^8, & \\
& dq^{10} = qd, & p^9da^3q^9r = rp^9a^3q^9, & \\
& eq^{10} = qe, & a^{-3}ta^3k = ka^{-3}ta^3 & \rangle
\end{array}$$

Figure 1.4: 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).

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 1.12, 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 1.16: "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 1.4.1). 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 1.17: 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:

$$\psi(e+e) = \psi(e) = 1 = 1 \cdot 1 = \psi(e) \cdot \psi(e)$$
$$\psi(o+o) = \psi(o) = 1 = (-1) \cdot (-1) = \psi(o) \cdot \psi(o)$$
$$\psi(o+e) = \psi(e+o) = \psi(o) = -1 = -1 \cdot 1 = 1 \cdot -1 = \psi(e) \cdot \psi(o) = \psi(o) \cdot \psi(e)$$

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 1.2(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 $\varphi : G \to H$ defiend 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 [19]. 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

[19]Elaborated in proposition 1.4.2

possible functions! That's a ton to check. We can deduce properties of homomorphisms which will let us diminish that number:

Proposition 1.4.1: 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.

***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 1.1.3
$$\varphi(1_G)\varphi(1_G)^{-1} = \varphi(1_G)\varphi(1_G)\varphi(1_G)^{-1}$$
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 1.1.3
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:[20]
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?
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 5)
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?

   Let $\varphi : G \to H$ be a homomorphism and $\ker(\varphi) = \{e_G\}$. Prove that φ is injective.

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:

[20] 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.

Proposition 1.4.2: 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 1.1.3, 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 [21]. 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 4): 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 1.4.2: 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

[21] 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 2

and the nodes are groups:

$$\begin{tikzcd}
 & U \arrow[d, shift left] \arrow[d, shift right] \arrow[dr, bend left] \arrow[dr, bend right] & \\
H \arrow[r, bend left] & L \arrow[l] \arrow[r] \arrow[d] & M \\
G \arrow[u] \arrow[ur] \arrow[r] \arrow[urr] & K \arrow[ur] \arrow[ul] & \\
 & & \\
 & & \\
 &

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 1.4.3: Set of Endomorphisms

1. Given a group G, an endomorphism is a homomorphism from an object to itself: $\varphi : G \to G$.
2. Let $\mathrm{End}(G) := \{\varphi : G \to G \mid \forall a, b \in G,\ \varphi(ab) = \varphi(a)\varphi(b)\}$ represent the set of all endomorphisms on G

What $\mathrm{End}(G)$ 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 1.17), 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 $\mathrm{End}(\mathbb{Z})$, 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 $\mathrm{End}(G)$, we can start getting a good idea on what are all sub-structures in G. More excitingly, if $\mathrm{End}(G)$ 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 $\mathrm{End}(G)$ 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

$$f, g \in \mathrm{End}(G) \qquad f + g \equiv (f + g)(x) = f(x) + g(x)$$

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 $\mathrm{End}(G)$, 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 $\mathrm{End}(G)$ is indeed a group. It's even an abelian group! This is an example of how being commutativity changes how a group effects.

Remark HomGrpIfAbRmk

In fact, more is true. If the domain of the functions are arbitrary sets, then $\mathrm{Hom}(A, G)$ is *still* an abliean group! This fact will be very useful to us much further don the line, in part II, III, and VII.

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 1.18: 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 1.4.4: 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 1.19: 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 1.4.4: 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 $(\mathrm{End}(G), \circ)$ 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 $\varphi^{-1} \circ \varphi = \varphi \circ \varphi^{-1} = \mathrm{id}$ (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 1.4.5: Automorphism

1. if $\varphi : G \to G$ is an isomorphism from a group to itself, then φ is called an ***automorphism***
2. Let $\mathrm{Aut}(G) := \{\varphi : G \to G \mid \varphi \text{ is an automorphism}\}$ be the collection of all automorphisms of G

Proposition 1.4.5: $\mathrm{Aut}(G)$ is a Group

Let G be a group and $\mathrm{Aut}(G)$ be the collection of automorphisms. Then $(\mathrm{Aut}(G), \circ)$ forms a group

Proof **:**
Proving it's a group should be doable at this point

Example 1.20: 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)

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 1.2[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:

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 1.4.6: 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 1.21
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 1.22

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 1.4.7: 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 1.4.7 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 1.23

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

Solution 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

Solution 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.

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 1.24

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 5.

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: [22]

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 1.25: 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:

$$\cdots \qquad -3 \longleftarrow -2 \longleftarrow -1 \longleftarrow 0 \longrightarrow 1 \longrightarrow 2 \longrightarrow 3 \qquad \cdots$$

   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:

$$\begin{tikzcd}
1 \arrow[r] \arrow[d, bend left] & b \arrow[r] \arrow[d, bend left] & b^2 \arrow[d, bend left] \arrow[dl, bend left] \\
a \arrow[r] \arrow[u, bend left] & ab \arrow[r] \arrow[u, bend left] & ab^2 \arrow[u, bend left] \arrow[ull, bend left]
\end{tikzcd}$$

3. Let $G = \mathbb{Z}/4\mathbb{Z}$. Then the Cayley graph is:

$$\begin{tikzcd}
0 \arrow[r] & 1 \arrow[r] & 2 \arrow[r] & 3 \arrow[lll, bend right]
\end{tikzcd}$$

[22] These conventions are usually followed in *geometric group theory*

4. Let $G = \mathbb{Z} \times \mathbb{Z}$. Then the Cayley graph is:

$$
\begin{tikzcd}
 & & & \vdots & & & \\
 & (-2,2) & (-1,2) \arrow[l] & (0,2) \arrow[l] \arrow[r] & (1,2) \arrow[r] & (2,2) & \\
 & (-2,1) \arrow[u] & (-1,1) \arrow[l] \arrow[u] & (0,1) \arrow[l] \arrow[u] \arrow[r] & (1,1) \arrow[r] \arrow[u] & (2,1) \arrow[u] & \\
\cdots & (-2,0) \arrow[u] \arrow[d] & (-1,0) \arrow[l] \arrow[u] \arrow[d] & (0,0) \arrow[l] \arrow[u] \arrow[d] \arrow[r] & (1,0) \arrow[r] \arrow[u] \arrow[d] & (2,0) \arrow[u] \arrow[d] & \cdots \\
 & (-2,-1) \arrow[d] & (-1,-1) \arrow[l] \arrow[d] & (0,-1) \arrow[l] \arrow[d] \arrow[r] & (1,-1) \arrow[r] \arrow[d] & (2,-1) \arrow[d] & \\
 & (-2,-2) & (-1,-2) \arrow[l] & (0,-2) \arrow[l] \arrow[r] & (1,-2) \arrow[r] & (2,-2) & \\
 & & & \vdots & & &
\end{tikzcd}
$$

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[23].

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[24]. 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, see appendix D

[23] In chapter 6, 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

[24] For example, they will come back when working with free groups and semi-direct products

2

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 1.2 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 2.0.1: 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 1.4.1 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 2.0.2: 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 2.1: Examples of Categories

1. It was mentioned that groups along with homomorphisms from a category. Indeed, by proposition 1.4.1 the composition of two homomorphsims is a homomorphism, so $gf = g \circ f$. Homomorphsims are associative since function composition is associative (proposition 0.1.1), 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 8.2))
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 **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 3.3.3. 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 2.1[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 2.0.3: [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 2.2: 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 2.3: 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 2.2, we shall further explore the use of commutative diagram to understand properties of categories.

2.1 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 0.1.2 and 0.1.3: 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 2.1.1: 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 2.1.1: 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 0.1.2, 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 2.1.2: 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 2.1.2: 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 2.1.3: 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 1.4.3. 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 27 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 2.4: sub-objects

1. **subobjects in Grp**: Recall in proposition 1.4.3 where we showed that if $(H, \cdot)$ 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 2.1.4: Isomorphism In C

Let C be a category, and f be a morphism. Then f is said to be an *isomorphism* if HERE.

Example 2.5: isomorphism in different Categories

1. In **Grp**, HERE
2. In **Set**, HERE
3. In the category of ordered sets, HERE

Proposition 2.1.3: 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 27

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.

2.2 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 2.2.1: Commutative Diagram

Give a nice definition here

Example 2.6: 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:

$$\begin{tikzcd} (a,b) \arrow[r, dotted, no head] \arrow[d, mapsto] & {} \\ a \cdot b \arrow[r, mapsto] & \varphi(a \cdot b) \arrow[u, dotted, no head] \end{tikzcd} \qquad \begin{tikzcd} (a,b) \arrow[r, mapsto] \arrow[d, dotted, no head] & (\varphi(a), \varphi(b)) \arrow[d, mapsto] \\ {} \arrow[r, dotted, no head] & \varphi(a) \cdot \varphi(b) \end{tikzcd}$$

which in it's familiar form would be $\varphi(a \cdot b) = \varphi(a)\varphi(b)$

2. maybe bring up how 1.6 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

2.3 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.

2.4 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.

3

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.

3.1 Basic Definition and Properties

First, a reminder of what it means to take a smaller group from the same group:

Definition 3.1.1: 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 3.1

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 3.1.1: 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 1.1.2). 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 3.1.1

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 $K \leqslant G$

3.2 Examples of Subgroups

In this section, we will explore "natural" subgroups as well as subgroups that arise given any homomorphism

3.2.1 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 3.2.1: 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

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 $\varphi = \psi$? (Not sure myself atm, just popped into my mind)

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 3.2.2: 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 1.4.1. 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 4.

Exercise 3.2.1

1. Do exercises with group actions showing that the kernel of the conjugation action is the centralizer

3.2.2 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 3.3.3. Overall, $T(G) \nleqslant G$ for any group 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 3.2.3: 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.

trivia The letter Z is because the German word for center is *Zentrum*.

Example 3.2: 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 3.2.4: Centralizer

Let G be an arbitrary group, and $A \subseteq G$[a]. Define

$$C_G(A) := \left\{g \in G | gag^{-1} = a \text{ for all } a \in A\right\}$$

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 3.3: 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 3.4: 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 3.2.5: 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 3.5: 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 3.2.1: 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 4 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 4. For now, you can keep them in the back of your mind.

Exercise 3.2.2

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 5.4.2)
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 $Z(G) \leqslant C_G(A)$?

3.3 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 14).

3.3.1 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 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 1.2.1) was discussed. Cyclic groups they pop-up quite often due to their simplicity.

[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 3.3.2 how to be more precise with the notion of smallest

Definition 3.3.1: 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 1.2.1 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 3.6

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 3.3.1: $|C_n| = |x|$

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 3.3.1: 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 1.1.2[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 3.3.2: 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 3.3.2

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 3.3.2, $x^{gcd(a,an)} = 1$

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 3.3.1: 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 3.3.2. 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 $\varphi \circ \psi = \psi \circ \varphi = \text{id}$ (technically, $\text{id}_{\langle x \rangle}$ or $\text{id}_{\langle y \rangle}$ 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 $\varphi \circ \psi = \psi \circ \varphi = \text{id}$ (technically, $\text{id}_{\mathbb{Z}}$ or $\text{id}_{\langle x \rangle}$ 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 3.3.3: 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).

Solution Dummit p.57-58 ▼

Finally, we'll put together all our knowledge into one theorem for easy reference:

Theorem 3.3.2: 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 (4.3.1), this proof will become much easier.

Note that the finite group examples of theorem 3.3.2 will eventually easily generalize to perhaps the most use theorem of the course: Lagrange Theorem (theorem 4.3.1), 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 3.3.2, 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 3.3.3: 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 3.3.2), 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 3.3.1, 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 3.7: : $H \notin Z(G)$
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 3.2.3 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 (4)

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!

3.3.2 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. 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 3.3.4: 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 3.3.2: 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 3.3.5

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 6.1.1 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:

$$G = \langle a, b \mid a^2 = b^2 = 1\rangle \tag{3.1}$$

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 3.1 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

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 3.3.1

[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?)

3.3.3 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 3.8: 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 1.4.1. 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 1.14 and section 1.2.3 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 3.8, 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 3.8 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 2 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 8.1 (which will turn out to be isomorphic to the set of all polynomials!), and *free-modules* in section 13.3.2

[6] You'll see these everywhere, to name a few: theorem 4.1.2, theorem ref:HERE (product), theorem ref:HERE (coproduct), proposition 9.2.1, theorem 9.5.3, theorem 16.3.1, theorem **??**, 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 3.3.6: $\mathcal{F}^A$ 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 1.4.2

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 1.4.2). 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 3.3.3: Word

Let A be some set. Then let $W(A)$ denote all possible combinations of *words*. TBD

Example 3.9: Example

1. here
2. here

Define reduction, define reduced word,

Lemma 3.3.1: reduced word lemma

reduced word lemma

Proof **:**
here

Proposition 3.3.7: Word Group

$(W(A), R)$ is a group

Proof **:**

1. (associativity)
2. ($\exists$ identity)
3. ($\exists$ inverse)

Theorem 3.3.3: 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 3.3.4: 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 3.3.4: Freely Generated

Let S be a set. Then we say that S *freely generates* a group G if $G = F(S)$.

Example 3.10: 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 3.1: 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

3.3.4 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 16 in which there is an initial object which is not a lift)

3.3.5 *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.

3.4 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 3.4.1: 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 3.11: 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 3.4.1: 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 3.12: 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 3.4.1. 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.

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 4

Exercise 3.4.1

1. Let $G = \mathbb{Q}$. If $1 \leqslant H \leqslant G$, show that there always exists a group H' such that $1 \leqslant 1 \leqslant H' \leqslant H$.

3.5 *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 1.4.2).

Theorem 3.5.1: 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 3.5.2: 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 4?

Also, for cokernel, this useful exact sequence might be helpful:

$$0 \to \ker(f) \to V \xrightarrow{f} W \to \operatorname{coker}(f) \to 0$$

4

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 section 0.1.6 in the preliminary chapter there exists an equivalence relation $\sim_\varphi$ on G induced by 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. This exploration will give us a new way of thinking and representing groups, and will go on to influence how to think 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

[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

see that when working with finite groups, the order of the group will give us a lot of information 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.

4.1 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 0.1.3 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

$$x \sim y \Leftrightarrow xb \sim xb \tag{4.1}$$

since we would also want that

$$[b \cdot a] = [b] \bullet [a]$$

we must have that

$$x \sim y \Leftrightarrow bx \sim by \tag{4.2}$$

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 4.1.1: $(G/\sim, \bullet)$ 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 $\bullet$ 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 $(G/\sim, \bullet)$.

First, note that the operation $\bullet : G/\sim \times G/\sim \to G/\sim$ $[a] \bullet [b] = [ab]$ 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 $\bullet$ is well-defined. Next, we need to show that $(G/\sim, \bullet)$ 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 propety, we will drop it without worry that there be any confusion, that is well write

$$[a][b] = [ab]$$

with the understanding that the operation on the left is in $G/\sim$ with $\bullet$ and on the right is G with $\cdot$.

We call $G/\sim$ the *quotient group* by $\sim$.

(word on congruence relation?)

Our

The fact that equivalence relations that respect 4.2 and 4.2 are the only ones that are congruence relations is very powerful, since this greatly limits what type of equivalence relations we have. For the rest of this section, we find all possible equivalence relations on a group G that satisfy property 4.1, 4.2, or both

We start by characterizing the equivalences classes of $G/\sim$ satisfying 4.1.

We will determine if there is any information we can figure out about the equivalence classes:

Proposition 4.1.1: Equivalence Classes Using property 4.1

Let $G/\sim$ be a quotient set given $\sim$ with proprety 4.1. 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 4.1. Using property 4.1 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 4.2, 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 4.1 $a^{-1}b \sim e_G$, 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 4.2, 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 4.1, 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 4.1 (and similarly 4.2), it's equivalence classes will be of the form

$$aK \qquad Ka$$

where a is the reprasentative 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 the equivalence classes of $G/\sim$.

Corollary 4.1.1: 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 4.1 or 4.2 (or both). Then every quotient set is of the same size, that is

$$aK = bK \qquad \forall a, b \in G$$

This theorem generalize to infinite groups

Proof **:**
here (does the proof acccomodate arbitrary size groups?)

Since these cosets represent our equivalence relations, we will give a special name to these quotient sets:

Definition 4.1.1: Cosets

Let $H \leqslant G$. Then for every $g \in G$, we can define

$$gH = \{gh \big| h \in H\}$$

we call this a [left]-***coset***. We can also define the ***right***-coset:

$$Hg = \{hg \big| h \in H\}$$

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 $G \setminus H$.

The definition of coset allowed for H to be any subgroup. Is it true that if H is a subgroup, then we can define an equivalence relation that satifies 4.1.1? In other words, we're asking for a converse to proposition 4.1.1[1], which stated that if $\sim$ respected property 4.1 (or 4.2), then the equivalence class containing e_G is a subgroup. The answer is yes:

Proposition 4.1.2: 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 4.1

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 4.1, 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 4.1

This proposition shows that any subgroup of H will define an equivalence relation $\sim_L$ respecting property 4.1, and thus by proposition 4.1.1[2] partition G (even better, equally partition G due to corollary 4.1.1). Similarly, we can show that the relatoin $a \sim_R b \Leftrightarrow ab^{-1} \in H$ gives an equivalence relation which satisfies property 4.2. Note that since H partitioned G, only $eH = H$ is a subgroup of G since only it contains the identity.

Combining proposition 4.1.1 and 4.1.2 we get the following one to one correspondence:

$$\{H|H \leqslant G\} \leftrightarrow \left\{ \begin{array}{c} \text{equivalence relations } \sim_L \\ \text{on } G \text{ satisfying} \\ \text{property 4.1} \\ \\ \\ \end{array} \right\}$$

Similarly:

$$\{H|H \leqslant G\} \leftrightarrow \left\{ \begin{array}{c} \text{equivalence relations } \sim_R \\ \text{on } G \text{ satisfying} \\ \text{property 4.2} \\ \\ \\ \end{array} \right\}$$

Thus completing our classification of equivalence relations that satisfy either 4.1 or 4.2, 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 4.1, and similarly for $G \setminus H$.

Is it the case that if we have G/H with the equivalence relation $\sim_L$, then this will have the same cosets as G/H satisfying $\sim_R$? The answer is actually no:

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 4.2. 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 4.1.1 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 4.1.3: 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 4.1.2: 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 4.1: 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 4.1.4: 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 4.1.3 is very important. The equivalence relatoin $\sim$ corresponding to H has both property 4.1 and 4.2 by proposition 4.1.2 and so by proposition 4.1.1 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 4.1.3: 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 4.1.2, and $\bullet$ is defined by

$$aHbH := (ab)H$$

An easy way to remember that this operation is defiend when is H is normal is by remembering $a(Hb)H = abHH = abH$.

Notationally, we can write $aK = \bar{a} = [a]$ 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

$$\pi : G \to H \qquad \pi(g) = gH$$

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 4.1.5: 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 4.1.1) 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 4.1.1 that it must factor through G/H. THis result is important enough to be given a name:)

Theorem 4.1.2: 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 4.1.1

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 4.1.2), we get

$$a \sim b \Rightarrow \varphi(a) = \varphi(b)$$

which put's us exactly at the inition condition for theorem 4.1.1, which gives us that the diagram commutes, as we sought to show.

What theorem 4.1.2 (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 4.1.2: 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 4.1.3: 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 4.1.2.

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)

$$1 \xrightarrow{\iota_1} K \xhookrightarrow{\iota_K} G \xrightarrow{\varphi} H \xrightarrow{\pi} 1$$

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 4.1.3 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 4.1.1

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 17.3) [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 $\mathbb{Z}/n\mathbb{Z}$, $n > 1$]

4.2 Presentations

We have seen in section 3.3.3 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

[2] reverse text? This might be giving the answer? Consider $G = H = N = \mathbb{Z}$, $K = 2\mathbb{Z}$

the set G to the underlying set of $(G, *)$ (where the underlying set of $(G, *)$ is G), we can extend it to a surjective group homomorphism $\varphi : F(G) \to (G, *)$. 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 3.3, 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 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 4.2.1: 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 $N = \langle R \rangle^{\trianglelefteq}$

From this, we can formally define a presentation

Definition 4.2.2: 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 (5.4.2): In Cayley's theorem, every group is the subgroup of a symmetric group. Here, every group is the quotient of a free group.

Example 4.2

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 4.2.1: 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).

4.3 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 4.3.1: 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 4.3

1. As a quick reminder, we used this theorem in (*corollary* 3.3.3)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* 3.3.2) 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 4.3.1: 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 4.3.1: 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 4.3.2: 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.

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 for 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 4.4: : 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 (3.3.4). 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 contradiciton. Note that since $|H| = 6$, then $|A_4 : H| = 2$, so by corollary (4.3.2), $H \trianglelefteq A_4$, 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 4.3.2: 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 (5.4.5) 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 5.4.6)

Cauchy's theorem is super important. It pop's a lot when trying to find groups of smaller order:

Example 4.5
If G is a finite abelian group with no normal subroups (besides 0 and G)[a], then $G \cong Z_p$.

Solution 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 3.3.1 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 4.3.3: 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 4.1: 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 6! 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 4.6: 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 $\mathbb{Z}$!

4.4 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 6, 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 4.4.1: 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

$$HK = \{\mathrm{id}, (123), (132), (123)(34), (132)(34), (34)\}$$

and $(132)(34) = (3421)$, but $(3421)^2 \notin HK$, meaning HK is not closed!

Proposition 4.4.1: 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 4.4.2: 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 4.4.1: 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 6 once we have learned the appropriate tools in chapter 5

Foreshadowing Thinking about HK like in corollary 4.4.1 will become the defining feature when we discuss *semi-direct* products in chapter 6.

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 will turn out then that there is a special type of ring (called a Dedekind ring) which will allow us to form a group 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.

4.5 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 4.1.2, 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 6)

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 4.1.2, but given it its classical name:

Theorem 4.5.1: 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 [3]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 4.7

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

$$S_n/A_n \cong \epsilon(S_n) = \{\pm 1\}$$

[3]In particular, it's a *natural homomorphism*. The word natural here has a technical meaning which we explore in section 27.4. 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

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 4.5.1: properties as consequence

Let $\varphi : G \to H$ be a group homomorphism

1. $\varphi$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 4.4.1) form a different angle.

Theorem 4.5.2: 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:

$$\begin{tikzcd}
 & G \arrow[d] & \\
 & AB \arrow[dl] \arrow[dr, red, "\trianglelefteq"] & \\
A \arrow[dr] & & B \arrow[dl] \\
 & A \cap B \arrow[d] & \\
 & \{e\} &
\end{tikzcd}
\quad \Rightarrow \quad
\begin{tikzcd}
 & G \arrow[d] & \\
 & AB \arrow[dl] \arrow[dr] & \\
A \arrow[dr, red, "\trianglelefteq"] & & B \arrow[dl] \\
 & A \cap B \arrow[d] & \\
 & \{e\} &
\end{tikzcd}$$

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 4.4.2, $AB \leqslant G$. 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 4.1.3, $A/\ker(\varphi) \cong AB/B$. 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 4.5.3: 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 4.1.3

$$(G/H)/\ker(\varphi) \cong G/K$$

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 4.8: 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 4.5.4: 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 4.5.2: 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 4.9: 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:

$$\begin{tikzcd}
 & & D_4 \arrow[dl, Rightarrow] \arrow[d, Rightarrow] \arrow[dr, Rightarrow] & & \\
 & \langle s, r^2 \rangle \arrow[dl, dash] \arrow[d, dash] \arrow[dr, Rightarrow] & \langle r \rangle \arrow[d, Rightarrow] & \langle rs, r^2 \rangle \arrow[dl, Rightarrow] \arrow[d, dash] \arrow[dr, dash] & \\
\langle s \rangle \arrow[drr, dash] & \langle r^2 s \rangle \arrow[dr, dash] & \langle r^2 \rangle \arrow[d, dash] & \langle rs \rangle \arrow[dl, dash] & \langle r^3 s \rangle \arrow[dll, dash] \\
 & & 1 & &
\end{tikzcd}$$

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 4.5.1

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 4.7. HERE

4.6 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 3.4.1), 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 4.10: 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 4.1.1). 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 4.11: Relation Within Cosets

Take $G/Z(G)$ (as an asside, 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 4.10, 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 4.6.1: 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 4.6.2: 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 $\operatorname{im}(f) = \ker(g)$, 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 $\operatorname{im}(\iota) = \ker(f)$ since f is injective, $\operatorname{im}(g) = \ker(\pi)$ since g is surjective, and $\operatorname{im}(f) = \ker(g)$ 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* [4]. 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 17). The exact sequence in definition 4.6.2 is the shortest

[4] 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)

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 4.12: 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.

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 17, 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 17.2, we will classify which *abelian groups* in a short exact sequence of abelian groups split[5] (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 4.6.1

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]
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!

4.7 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

[5] 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 $\mathbb{Z}$

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 [6]). 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 7.1, 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.

Definition 4.7.1: 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 [7]. 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 4.13: 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 wich no non-trivial subgroups (and so no normal subgroups)

[6] 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)

[7] 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

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 5, 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[8] (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 4.7.2: 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 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 4.14: 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.

[8] meaningful here meaning there is a notion of combining groups using short exact sequence, unlike general groups can't be

Example 4.15: 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 4.7.3: 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 [9], 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.

Example 4.16: 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:

$$1 \trianglelefteq \langle s \rangle \trianglelefteq \langle s, r^2 \rangle \trianglelefteq D_8 \qquad \text{and} \qquad 1 \trianglelefteq \langle r^2 \rangle \trianglelefteq \langle r \rangle \trianglelefteq G$$

[9] note that maximal subgroups by definition are proper subgroups

Note that not all infinte groups admit a compositions series

Example 4.17: 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 4.7.1: 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 4.7.2: 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 4.7.3: 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 4.18
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

4.8 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 1.4.3. 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?)

5

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!)

5.1 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 cna 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 1.2.2. 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 5.1.1: 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.

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 5.1: 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

$$(gh)(x) = g(h(x)) \qquad (gh) \cdot x = g \cdot (h \cdot x) \tag{5.1}$$

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

$$e(x) = x \qquad e \cdot x = x \tag{5.2}$$

(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?)

5.2 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.

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 5.2.1: 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 5.2.1 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 5.2.1: 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 5.2.2: 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 5.2.1

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?)

5.2.1 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 5.3.3) 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 5.3.2) 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 5.2.2: 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 5.4.2).

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 5.2: : 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 $x_1 + x_2$

Solution 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 5.2.3: Transitive

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 5.2.4: 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 5.2.1: Burnside'S Lemma

$$|X/G| = \frac{1}{|G|} \sum_{g \in G} |X_g|$$

Proof. `https://en.wikipedia.org/wiki/Burnside%27s_lemma` □

5.3 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 5.3.1: 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 5.3.2: 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 5.3.1: 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 5.3.3: 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 (5.2.2)

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 5.3.4: Faithful Group Action

An action is ***faithful*** if its kernel is the identity

5.4 Application to Finite Groups

(Words here)

5.4.1 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 5.3: 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 5.4: 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 5.5
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 5.4.1: 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 5.4.2: 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)

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 $n = 100$,

$$|S_{100}| = 93'326'215'443'944'180'000'000'000'000'000'000'000'000'000$$
$$'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000$$
$$'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000'000$$

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 5.4.1: 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 (4.3.1)

Proof **:**
p.120 Dummit and Foote

Here are a couple of exercises I found insightful:

Example 5.6: 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).

Solution 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 5.7: : 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 5.4.1

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]

5.4.2 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 (3.2.4, 3.2.5) 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 5.4.1: 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 5.4.2

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 5.2.2 to find the order of the orbits. We first find the order of the stabilizers. Let $a \in G$. then:

$$G_a = \left\{g \in G \middle| g \cdot a = gag^{-1} = a\right\} = C_G(a)$$

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 5.4.2: 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 (5.2.2), we know that $|X_S| = |G : G_S|$. And by definition of stabilizer, we know that

$$G_S = \left\{g \in G \middle| g \cdot S = gSg^{-1} = S\right\} = N_G(S)$$

that is, the stabilizers are the normalizers of S in G, therefore, $|X_S| = |G : N_G(S)|$.

Corollary 5.4.2: 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,

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: $\left\{gxg^{-1} \middle| g \in G\right\}$, 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

$$(g, h) \mapsto g^{-1}hg$$

5.4.3 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 3.2.3). 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 5.4.3: 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 5.8
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 3.3.3), 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 5.4.4: 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 5.4.3: 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 5.4.3: 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 5.9
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: $V_4 \leqslant A_4$

5.4.4 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 5.4.4: 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 5.10
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 5.4.5: 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 $p \big| n$). 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 5.11
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 5.4.5: 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 5.5.1, 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 $n_p \equiv 1 \mod (p)$. 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) $n_p \equiv 1 \mod p$
   (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 5.4.1: 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 (4.3.2) 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 (4.3.1), 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 (4.3.2), which would just say there exists a group of order p).

Corollary 5.4.4: 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

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 $n_p \equiv 1 \mod p$ 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 5.12

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

   Solution 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

   Solution 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 3.3.2 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 5.4.4, 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 5.4.6: 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 4.3.2

***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 3.3.2 $|\langle x^{p^{\alpha-1}} \rangle| = \frac{p^\alpha}{(p^\alpha, p^\alpha - 1)} = p$, 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

5.4.5 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 5.13: : 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

$$n_p \cong 1 \mod (p) \qquad n_q \cong 1 \mod (q)$$

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 3.3.2. 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 5.4.4), 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 5.14: : Groups of Order 30
p.143

Example 5.15: : Groups of Order 12
p.144

Example 5.16: : Groups of Order p^2q, $p \neq q$
p.144

Example 5.17: : Groups of Order 60
p.145

Corollary 5.4.5: 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!

5.5 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 1.4.5). 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 5.2.1), 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 (4.1.2): $N_G(H)/C_G(H) \cong I \leqslant \mathrm{Aut}(H)$.

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 6).

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 5.5.1: 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 5.5.1: 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 5.18
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 5.5.2: 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 5.5.1: 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 $Z(G)$.

5.5.1 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 5.5.3: 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 $\sigma \in \mathrm{Aut}(G)$.

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 5.19

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 6.2 and 7.1.1. Show that $G' \blacktriangleleft G$

Proposition 5.5.2: 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 4.7.3), 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 $\varphi \in \mathrm{Aut}(G)$, then it's also true for every $\varphi \in \mathrm{Inn}(G)$, 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 $\varphi \in \mathrm{Inn}(G)$, $\varphi(H) = H$. 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 $\varphi \in \mathrm{Aut}(K)$, $\varphi(K) = K$. Since $K \trianglelefteq$, we know that for $\psi \in \mathrm{Inn}(G)$, $\varphi(K) = gKg^{-1} = K$ (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 5.5.1: 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 $\mathrm{Aut}(H)\text{``} \leqslant'' \mathrm{Aut}(G)$, 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 19.1.3)

Exercise 5.5.1

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!)

5.6 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 $L(G) \blacktriangleleft Z(G)$

6

Composing Groups: Direct and Semi-Direct Products

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 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

[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

6.1 Classical Definitions And Properties

You might recall example 1.2[4], we showed we can create a group-structure given two groups. We revisit this definition here in generality:

Definition 6.1.1: 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 4.4.1)

Example 6.1

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 6.1.1: 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 6.1.1

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 6.1.1, 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).

Example 6.2: 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 6.3: 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.

[2] Think about $G = \prod_{n=1}^{\infty} \mathbb{Z}$

6.1.1 Fundamental theorem of Finitely Generated Abelian Groups

This theorem is stated, but not proved till till much later (chapter 15, theorem 15.1.3). 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 6.1.2: Finitely Generated Group

A group is finitely generated if $\exists A \subseteq G$ such that $\langle A \rangle = G$.

Definition 6.1.3: 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 6.1.1: 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 15.1.3. 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`

[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

Definition 6.1.4: 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

6.2 Recognizing Direct Products

Let's say we have a (not necessarily abeliean) 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 6.1.1 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 6.4: 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

$$xy = (1\ 2)44(2\ 3)2 = (1\ 2)(2\ 3)442 = (1\ 2\ 3)2$$

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 6.2.1: Commutator

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 6.2.1: 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 $\sigma \in \operatorname{Aut}(G)$, 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 5.5.2).

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 defiend 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 6.5

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 6.2.1: 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 4.4.1

Theorem 6.2.1: 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 6.2.2: 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 $k, k' \in HK$, $kk' = k'k$. $\square$

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 4.4.1, at least one of H or K are normal. We need to show that both H and K are normal. We know by proposition 6.1.1 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 4.4.2, and so by the lemma done earlier, every element of H commutes with elements of K.

Now, by lemma 6.2.1, 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 4.4.1) and *external direct product* (definition 6.1.1) of two subgroups H and K of G are equivalent if the two conditions in theorem 6.2.1 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 6.1.1 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.

6.2.1 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 6.2.3: Perfect Group

Let G be a group and G' be the commutator subgroup. If $G = G'$, then G is called a *prefect group*

Example 6.6: 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.

6.3 Semi-Direct Products

As we mentioned earlier, not all groups can be broken down into [non-trivial] direct products:

Example 6.7: $G \ncong H \times K$

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 4.3.2). Since one of the subgroups are normal, by proposition 4.4.2 $\langle r \rangle \langle \sigma \rangle$ is a group, and by proposition 4.4.1, 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 4.1.3 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 $b = \left\{e, r, r^2\right\}$ 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 4.4.2), 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 4.4.2. 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 $h_3 = h_1(k_1h_2k_1^{-1})$ (since $H \trianglelefteq G$, $gh_1g^{-1} \in H$) and $k_3 = k_1k_3$.

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 6.3.1: 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:

$$(h_1, k_1)(h_2, k_2) = (h_1 k_1 \cdot h_2, k_1 k_2)$$

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 6.3.1: 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 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 6.3.2: 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 6.8: 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 6.3.1: 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 17).

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 6.3.2: 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 6.9: 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.

6.4 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)

$$\begin{tikzcd}
 & G \arrow[d, "f"] \arrow[ddl, bend right, "f_A"'] \arrow[ddr, bend left, "f_B"] & \\
 & C \arrow[dl, "\pi_A"] \arrow[dr, "\pi_B"'] & \\
A & & B
\end{tikzcd}$$

[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!

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 6.10: Morphissm In $C_{A,B}$
here. Examples and counter-examples

Proposition 6.4.1: $C_{A,B}$ 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 6.4.1: 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 6.4.2: 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 6.4.2. All the steps generalize without a problem.

Exercise 6.4.1

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 6.4)

6.4.1 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 6.11: 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 6.1.1[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 3.3.3). Something analogous is defined called the *free product*.

Definition 6.4.1: 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 6.12: 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 6.4.2

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 6.13: 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 6.4.2: 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 6.4.3: 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.

6.4.2 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.

7

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)

7.1 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 7.1.1: 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 7.1.1: 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 7.1: 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 7.1.2: 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.

7.1.1 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 7.1.2: 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 7.1.3: 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 7.2: 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 7.1.1: 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 6.2.1. Then normality comes from proposition 5.5.2.

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 7.1.1: 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 6.2.1 $G_{(1)} \leqslant A_1$. 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 7.1.3: 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 7.1.1). 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 7.1.1: 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 7.1.3(2), G is solvable.

There is actually one more incredibly powerful theorem for which groups are solvable:

Theorem 7.1.2: 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 7.1.3: 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.

7.1.2 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 7.1.4: 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 7.3: 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), $|Z(G)| \big| |G|$, 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 7.1.4: 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 7.4: Example
here

Such groups are caleld *Engel groups*, and come up in Lie Algebra.

Proposition 7.1.5: 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 7.5: 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 7.1.6: 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 5.4.4, $P \operatorname{char} N$. 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 7.1.1

1. Prove that any nilpotent group of class 3 or lower is metabelian
2. Show that if $\operatorname{Aut}(G)$ is abelian, then G is nilpotent of class 2

7.1.3 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 7.1.5: 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 7.6: 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 7.1.7: 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 7.7: 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)

7.2 Non-finitely generated abelian groups

(Prüfer's Theroems!)

7.3 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?

8

Generalization of Groups

In this chapter, I will cover common algebraic objects which are generalizations of groups.

8.1 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 1.1.3). 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 8.1.1: 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 8.1
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 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 8.1.1: 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 8.2
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 8.1.1

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!!)

8.2 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 8.2.1: 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?

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* [3] 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 is 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[4] 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[5] (which we will present near the end of section 9.1.1). In that way, all groups can be made into rings, and we technically study them purely as rings where we ignore the first binary operation. However, we usually care a great deal about the second binary operation and it's interaction with the first binary operation, and we usually we will find "nice" binary operations to add onto abelian groups that will make their study more fruitful.

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.

[3]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`

[4]All be it to a very limited degree

[5]Similarly to how all 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. Afterwards, 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.

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 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 6), it will not obtrude our study of the ring operation. This is a common technic.

Example 9.1: 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 14.2.2), 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 25) 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 9.8 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 $\mathbb{Z}^\times, \mathbb{Q}^\times, \ldots$ 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 8.1). 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 1.4.3

Corollary 9.1.1: 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 9.1.1: 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 9.1.4: 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 9.1.5: 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 9.2: Exercise
Prove that any subfield of $\mathbb{R}$ must contain $\mathbb{Q}$

Solution 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 1.1.3). 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 9.3
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 9.4: 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 9.5

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 9.1.2: 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 9.1.6: 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.

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 9.1.3: 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 9.6: 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

$$\forall x, y \in \mathbb{Q}(\sqrt{5}) \qquad a + b\sqrt{D} - c - d\sqrt{D} = (a - c) + (b - d)\sqrt{D}$$
$$\forall x, y \in \mathbb{Q}(\sqrt{5}) \qquad (a + b\sqrt{D})(c + d\sqrt{D}) = (ac + bdD) + (ad + bc)\sqrt{D}$$

showing it's indeed a [commuative] 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 9.1.4: 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 cna 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 9.1.2. 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 9.1.7: Integral Domain

A commutative ring with identity $1 \neq 0$ is called *integral domain* if it has a no zero divisors

Example 9.7: 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. $(GL_n(\mathbb{F}), +, \cdot))$ is an integral domain (even a field)
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 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 9.1.8: 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 9.1.5: 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 9.1.4 can be expanded upon:

Corollary 9.1.2: Finite Integral Domain is a Field

If R is a finite integral domain, then R is a field

Just like in proposition 9.1.4, 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 VII 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 9.8: 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 $xy = yx$.

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 9.1.6: 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[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 1.4.2, 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 9.1.7: $\mathbb{Z}$ 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 9.1.1

$$\varphi'(1) = 1\,, \varphi'(2) = 1 + 1\,, \cdots$$

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 1.4.3 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)

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 5.4.2)! 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.

[7] Research **Ab**-enriched categories for information about this process

: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 9.8!

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 6 in creating larger groups using products. Thus, extensions of fields (similar to how they were defined for groups in section 4.6) 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.

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.

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 ref:HERE). 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 ref:HERE).

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.

Polynomial Rings

Polynomial Rings are one of the first examples we'll give of an 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:

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)

$$\left(\sum_{i=0}^{\deg(p)} p_i x^i\right) \cdot \left(\sum_{i=0}^{\deg(q)} q_j x^j\right) = \sum_{k=0}^{\deg(p)+\deg(q)} \sum_{l=0}^{k} p_l q_{k-l} x^k$$

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.

[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

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 $r \mapsto r$.

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.

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

$$R[x, y] = (R[x])[y]$$

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. element 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.

Usually, when we maniuplate 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 aglebra and geometry in the 20th century.

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 9.2.1: 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

[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 27 needs to be read to reach section 27.4

[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 *direct limit*

***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.

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

Theorem 9.2.1: 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

$$\begin{tikzcd} R \arrow[r, "\varphi"] \arrow[d, "f"] & S \\ R[x] \arrow[ur, "\psi"'] & \end{tikzcd}$$

For those who are categorically inclined, try figuring out what is the category involved behind the scences (this is similar to the construction for free groups in section 3.3.3)

***Proof* :**
The proof becomes easy to see once you realise a polynomial:

$$a_n x^n + a_{n-1}x^{n-1} + \cdots a_1 x + a_0$$

becomes

$$\varphi(a_n)s^n + \varphi(a_{n-1})s^{n-1} + \cdots \varphi(a_1)s + \varphi(a_0)$$

As mentioned earlier, the polynomial rings $R[x]$ can be seen as an extension of R (in the sense that R is a subring of $R[x]$). There is a useful way of characterizing how to think about this extension: (mention commutativity)

A particular case of this is when $\varphi : R \to R$ is an identity map:

Definition 9.2.2: 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 whic 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 9.9
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.

:exercise

(Is $R[x]$ isomorphic to $\oplus_{i=1}^{\infty} R$?

The same way $\mathbb{Z}$ is a the initial ring, there is a notion of there being an initial polynomial ring (see Aluffi p. 130).

(Also a word on how $R[x]/(p(x))$ is in a sense forcing a solution onto the polynomial $R[x]$. Acutally, make this an exercise!! It would spoil too cool of a point. The idea being that we can let a be the "solution" in $R[x]/(p(x))$ as in $p(a) = 0$)

9.2.1 *Formal Power Series

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 20.3. For now, we may consider $f(x)$ in the above equation as "actually" being an infinite tulpe[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 9.2.3: 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 topology on our ring, something we shall not cover. We shall remark that it is very commmon in

[12] Is $\prod_{i=0}^{\infty}$ defined?

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 9.2.2: 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

(perhaps mention some behavior that differs from $R[x]$. For example, we have some more units, like $\frac{1}{1-x} \in R[[x]]$, among others)

Lemma 9.2.1: Existence Of More Units

Let $k[[x]]$ be the formal power series over the field k. Then $(1-x)$ has an inverse in $k[[x]]$

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 9.2.3: 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 9.2.1 to get that $1 - u(x)$ has an inverse. But then we may conclude that $f(x)$ has an inverse, thus completing the proof.

Definition 9.2.4: 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 9.2.4: 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 9.2.5: 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 9.2.6: 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

$$f \circ g = g \circ f = I$$

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 defiend 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 9.10: 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. $\log(x) =$

Exercise 9.2.1

1. Is $R[[x]]$ isomorphic to $\prod_{i=1}^{\infty} R$?)

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 seperate algebraic objects, they all follow the same rules as

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 3.3.5, **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" TBD. 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 consructions are called *free*.

Definition 9.2.5: Group Ring

Let R be a commutative ring and let G be any finite group, $G = \{g_0, g_1, ..., g_n\}$ where $g_0 = e$ is the identity, with the operation written multiplicity. Then define the *group ring* $R[G]$ (or RG) or G to be the set of all formal sums:

$$a_0g_0 + a_1g_1 + \cdots + a_ng_n \qquad a_i \in R \; g_i \in G$$

where addition is done pointwise:

$$\begin{aligned} &a_0g_0 + a_1g_1 + \cdots + a_ng_n + b_0g_0 + b_1g_1 + \cdots + b_ng_n \\ &= (a_0 + b_0)g_0 + \cdots + (a_n + b_n)g_n \end{aligned}$$

and the product is defined per monomial as $(ag_i)(bg_j) = (ab)g_k$ where ab is performed in R and $g_ig_j = g_k$, and then extended to:

$$\begin{aligned} &(a_0g_0 + a_1g_1 + \cdots + a_ng_n)(b_0g_0 + b_1g_1 + \cdots + b_ng_n) \\ &= \sum_{g_ig_j = g_k} a_ib_j \end{aligned}$$

Since $g_0 = e$, we often write $a_0g_0 = a_0$. 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 $a_i = \delta_{ij}$ where δ_{ij} is the Kronecker delta, we can identify $G \subseteq R[G]$, and G as a subgroup of $R[G]$ has it's original group oeration. 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

[13] We will cover free rings later

if and only if G is commutative. Another simple exercise to check is that R (identified 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 $R[G]$

Example 9.11: 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:

$$a + b \qquad ab$$

More to it than meets the eye! `https://en.wikipedia.org/wiki/Group_ring`

Note that group rings cannot be division rings (not every element has a multiplicative inverse), because a group rings always has zero divisors. Let $|G| > 1$, and let $g \in G$ such that $m = |g| > 1$, then

$$(1 - g)(1 + ... + g^{m-1}) = 1 - g^m = 1 - 1 = 0$$

(A word on monoidal rings? How they are bassically a free-monoid adjoined to R? Maybe have this as an exercise, i.e. $R[\mathbb{N}]$ related to $\mathbb{R}[x]$)

(also a free-forgetful adjoint here?)

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 9.12

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 9.2.3, 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 9.3.1: 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[14]. Kernels will soon be defined as *ideals* (definition 9.3.2) since they have more properties than ideals.

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 9.1.1. 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.

[14]The fact that the kernel is not a ring means that the category of **Ring** does not have kernel objects

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 4.6 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, 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:

$$(r + I) \cdot (s + I) = (rs) + I \qquad \text{(in equivalence class notation: } [x \cdot y] = [x] \cdot [y]\text{)}$$

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* [15]. 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 9.13: : 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

$$\left(\frac{a}{n} + \mathbb{Z}\right) = (a + \mathbb{Z})\left(\frac{1}{n} = \mathbb{Z}\right) = (0 + \mathbb{Z})\left(\frac{1}{n} + \mathbb{Z}\right) = [0]$$

[15] as before we'll delay defining the quotient object till we find the sufficient conditions on I independent of a ring homomorphism

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 $\frac{1}{n}$ and $\frac{1}{m}$. Then

$$[0] = \left[0 \cdot \frac{1}{m}\right][0]\left[\frac{1}{m}\right] = [1]\left[\frac{1}{m}\right] = \left[n \cdot \frac{1}{n}\right]\left[\frac{1}{m}\right] = [n]\left[\frac{1}{n}\frac{1}{m}\right] = [n]\left[\frac{1}{nm}\right] = \left[\frac{n}{nm}\right] = \left[\frac{1}{m}\right]$$

Which means that for every coset $(\frac{a}{b} + \mathbb{Z}) = (0 + \mathbb{Z})$, meaning that

$$(r + I)(s + I) \neq (rs + I)$$

unless $(r + I) = [0]$ or $(s + I) = 0$, meaning there is no salvagable 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

$$(i + I)(j + I) \in I \Rightarrow ij \in I$$

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 9.1.1, 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 9.3.2: 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 9.14

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))

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.

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 9.3.3: 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 9.3.2: 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 9.3.1: 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 4.1.2), 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 4.1.2.

As with groups, the bar notation $\overline{r}$ or $r \mod I$ notation is used to represent the equivalence class $[r]$.

Example 9.15

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 $\mathbb{F}[x,y]$, then $\mathbb{F}[x,y]/(x) \cong \mathbb{F}[y]$

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 9.16

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 9.3.2: 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))

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 24.2.2).

[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 $R \times S$?

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, since ideals are not subrings, ring theory is not the best context in which to introduce the concept (This reasoning sounds unsatisfying to me..). So, the proof that any ring is the quotient of a free ring is left for module theory.

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). (Perhaps on something how it's useful for polynomial rings?? How that goes to geometry?? Then this leads us to want ot understand ideals more, and hence why we look into generations of ideals in this section?)

(Why would we define these concept for ideals instead of rings? Because of it's connection to number theory I'd suppose)

(maybe because of the use of ideals in the isomorphism theorems?)

(Perhaps talk about ho certain number theoyr concepts like euclidean algorithm and Bezout's Identity translates into number theory (ex. the ideal (a, b) in $\mathbb{Z}$ is another ay of phrasing $xa + yb$))

(Let's say R is commutative. 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 $r \in R$)

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

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 defiend inductively by defining $I^1 = I$, and $I^n = II^{n-1}$ for $n = 2, 3, ...$ By convention $I^0 = R$.

5. 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?

Example 9.17

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.

$$\{r_1a_1 + \cdots + r_ka_k | r_i \in R\,, a_i \in A\}$$

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 $RA = AR = RAR = (A)$.

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)

Notice that we have defined the generation of ideals and not of rings. This is because 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 $1 \notin \overline{\{1+x\}}$[16], and it is

[16]This is actually a bit difficult to prove, that no polynomial $p(y)$ exists such that $p(1+x) = 0$, and so no polynomial

not an ideal since $(x+2)(x+1) \notin \overline{\{x+1\}}$[17]. 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.

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 9.18

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

   Solution 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 Bezouts 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

$q = p + 1$for which $q(x+1) = 1$. The full answer will be given in part IV when discussing transcendental extensions.

[17] 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

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 *Bezout 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 9.2.1, 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 9.4.2, $(a(x)) = \mathbb{Z}[x]$. 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 $R = \mathbb{F}[x]$ and take $I = \{p(x) \in F[x] | p(q) = 0\}$ 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 since being Noetherian is stable under taking the quotient, and is also stable under the "converse" of taking the quotient (if you've covred section 4.6, the Noetherian condition is stable under extensions). For the next proposition, note that a finitely generaed ideal is defined identically to a finitely generated ring (that it is generated by finitely many elements):

Proposition 9.4.1

let R be a Noetherian ring and I be any ideal of R. Then R/I is finitely generaed. 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 ring are those for which every ieal 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 9.4.4: 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*.

(example: show that $\mathbb{Z}$ is a pid?)

(Once Euc. Domains are introduced, and that $\mathbb{Z}$ is a Euc Domain, then this shows a much easier way of proving that $\mathbb{Z}$ is a pid, hence a ufd).

Proof **:**
$\mathbb{Z}$ is a Euclidean Domain, and hence is a Principal Ideal Domain, and hence is a Unique Factorization Domain.

(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 15 will heavily rely on this type of ring). (The point is to introduce them later in the chapter on factorization)

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.

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 9.4.2: 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 9.19: 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 9.4.1: 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 9.4.5: 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 VII

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 9.4.6: 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 9.20: Ring with maximal ideal
Take $\mathbb{Z}$ and $6\mathbb{Z}$. Since $6 = 2(x)$, $6\mathbb{Z} \subseteq 2\mathbb{Z}$. 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 9.21: 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 9.4.3: 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 9.4.4: 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 9.4.2. 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 9.4.2, 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 9.4.4 can be visualized with such a diagram:

$$\begin{array}{ccc} R & \qquad & R/M \\ | & & | \\ M & & 0 \end{array}$$

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 9.4.3 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[18]. 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 VII. 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 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

[18] Indeed, we can modify example 9.19 to see that $M_2(k)/(0)$ is not a field, and hence a counter-example

Example 9.22

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 hen we try to talk about divisibility over more general rings?

(TBD – ho did we get to defining prime as it is?)

If we take $(a) \subseteq R$ to be a principal ideal, then if $xy \in (a)$, then for some $z \in R$, $xy = za$ (TBD??)

If we take a ring R, If 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 9.4.7: 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.

Example 9.23: 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[19].

(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 $p|r$, 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 9.24: 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

[19] for example, proposition 10.1.7 and ref:HERE

This last example is important enough to be referenced many times, and so we make it a proposition:

Proposition 9.4.5: 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 9.4.2: 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 9.4.4 R/M is a field. A field is an integral domain, so the corollary folllows from proposition 9.4.5

Example 9.25

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 20.2. 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 $\left\{\frac{a}{1} | a \in \mathbb{Z}\right\}$ [20], showing how $\mathbb{Z}$ is a sub-ring of $\mathbb{Q}$.

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

[20] meaning construct a isomorphism

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

$$Q = \big\{[(a, b)] \big| a \in R, d \in D\big\}$$

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

$$\begin{aligned} [(a, b)] + [(c, d)] = [(ad + bc, bd)] \qquad &\Rightarrow \qquad \frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \\ [(a, b)] \cdot [(c, d)] = [(ac, bd)] \qquad &\Rightarrow \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd} \end{aligned}$$

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)

   Solution 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 9.5.1: 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 $\mathbb{R}^{\times}$.

Theorem 9.5.2: Smallest Field Containing R Is Q

Let R be an integral domain and let Q be the field of fractions of R. If a field $\mathbb{F}$ contains a subring $R' \cong R$, then the subfield of F generated by R' is isomorphic to Q. That is, the smalelst field containing an integral domain R is its field of fractions

Proof **:**
Use the uniqueness of the extension to show that $\Phi(Q)$ will be in any sub-field.

Here are some examples of ring of fractions

Example 9.26

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 $D = \left\{19, 19^2, ...\right\}$. 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 **:**
This was

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 20.2.

(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:

$$\frac{\mathbb{Z}[x_{1/2}, x_{1/3}, ..., x_{1/p}, ...]}{(x_{1/2} - 2, x_{1/3} - 3, ...,)}$$

Though this way of looking at it seems more complicated, it fits beter with the idea that we are addings 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)

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* [21]). 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

The Chinese remainder theorem is to me one of the clearest ay of seeing ho rings can be thought of as the generalization of integers. It's name sake comes from a book called the _____. It is said that the _____.

Interesting folklore origin: `https://en.wikipedia.org/wiki/Sunzi_Suanjing`

or

here

Before we get into the actual theorem, e'll first explore a direct application of chinese remainder theorem:

Example 9.27
simple form

$$\text{if } gcd(m,n) = 1 \text{ and } 0 \leq a < m\,, 0\ leb < n$$

then

$$\exists! k 0 \leq k < mn \text{ such that } k \equiv a \mod m\,, k \equiv b \mod n$$

Let's say: $m = 6$, $n = 5$, $a = 4$, $b = 2$. Then $k = 22$.

[21] See appendix ref:HERE

Solution e have

$$k \equiv 4 \mod 6, \ k \equiv 2 \mod 5$$

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:

$$k \equiv 5 \cdot 1 \mod 30, \ k \equiv 6 \cdot 1 \mod 30$$

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 [22]

Definition 9.6.1: coprime or 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 Bezout's identity, there exists numbers x, y such that $xa + yb = 1$, meaning $1 \in (a) + (b)$, meaning $(a) + (b) = \mathbb{Z}$ by propositon 9.4.2. 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 [23]

Proposition 9.6.1: Properties 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 **:**
Micahel Atiyah p. 7

Theorem 9.6.1: Chinese Remainder Theorem

Let $A_1, A_2, ..., A_k$ be ideals in R. The map

$$R \to R/A_1 \times \cdots \times R/A_k \qquad \text{defined by} \qquad r \mapsto (r + A_1, \cdots, r + A_k)$$

is a ring homomrophism with kernel $A_1 \cap \cdots \cap A_k$. 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 $A_1 \cap A_2 \cap \cdots \cap A_k = A_1 A_2 \cdots A_k$, 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 **:**
p.265-266

[22] why not co-prime ideals? We'll show how prime and maximal ideals will match in special rings in chapter 10 here we ould usually apply such a result. Hoever, we stay general since we can prove it in more general contexts

[23] For example, we'll use this theorem when looking at polynomial rings over fields

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)

(eucliean domains let you find the divisibility relations between *any* two elements!!!)

(Perhaps write how it is difficult to write all this content linearly, so a narrative of factorization was chosen (this narrative being the "trunk" from which many the other paths of exploration branch out of), and references are put here and there that shows where in the following "story" or "exploration" the reader can branch out and see what happens if a different path of exploration was chosen. Then, we will switch gears for a moment and talk about a different concept called valuation, or in particualr discrete valuation, which will allow us to bring in some notion of "size" or "linear ordering" to elements. We'll show how some rings have a "euclidean algorithm" that gives us an idea on how to divide elements when no units are present. In the end, we will show how our discussion on factorization is intimitely linked to rings with euclidean valuation)

The following diagram sums up the future results of this chapter:

Figure 10.1: Image form Alufi's textbook

> **Note**: Some results in this chapter apply beyond integral domains. It will be made extra-clear when that is.

(start section here?)

In fields, every element has an inverse. This is not so for general rings, as we have seen when studying units and zero-divisors. There is a weaker idea of inverse that is introduced when learning some basic number theory: *factorization*. In this way, we can somewhat "reverse" an element back at least some way; perhaps not all the way to 1, but but perhaps to some 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:

$$x = p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}$$

Is there a generalizable 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 like for integers), while other rings have a weaker condition (like the factorization not being unique). The focus of this chapter is to explore these ideas.

The overarching idea in this chapter 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 $2|8$, then there exists some $x \in \mathbb{Z}$ such that $2x = 8$, that is, 8 got subdivided into other elements. This idea will be at the heart of our following exploration.

The result will be 4 new classes of 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.

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. *Euclidean Domains* will have the property of the euclidean division algorithm applying to them to quickly find the greatest common divisor.

Since factorization is heavily inspired by the properties of $\mathbb{Z}$, all rings in this chapter will be *integral domains*.

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 $5 \not\big| 6$ and $5 \not\big| -6$. 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 checked 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: 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 10.1.1: 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, 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 10.1.3: 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 10.2: 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 units. If there existed a a, b such that $2 = ab$, then by our knowledge that 2 is a prime number, 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 9.23, 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 9.2.1:

$$\deg(x^2) = \deg(p(x)) + \deg(q(x))$$

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, 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, the ability to "work backwards in ideals" is a strong enough condition that it guarantee's irreducibility:

Proposition 10.1.2: 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 9.1.5), 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 10.3: 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 part ref:HERE (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 TBD – make sure this is actually doable, if it is then reference a theorem later in the book)

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 10.1.4: 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 10.1.5: 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

The following corollary is stated for the sake of reference:

Corollary 10.1.1: 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 been multiplied by a unit ($uq_1 q_2 \cdots q_n$), however since it is a unit the result is an associate to $u^{-1}r$ it is harmless to drop the unit in the definition.

Example 10.4: 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 were 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 11.5, $F_{0x} \subseteq \mathbb{F}[x]$ 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 9.4.5)
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 9.18[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 10.1.6: 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[1]:

[1] more on the generalization of it??

Proposition 10.1.3: 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 9.4.4). 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 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

This proof will become a special case of the equivalent proof for modules in section 13.3.4. 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 10.1.4: 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.

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.

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[2], 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[3] (TBD)) to a factorization condition (a condition for how to decompose elements into simpler elements).

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 10.1.2: Noetherian Domain $\Rightarrow$ Atomic Domain

Let R be a Noetherian Domain. Then it is an Atomic Domain

Proof **:**
By proposition 10.1.3, 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 10.1.4 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

[2] P.M. Cohn, Bezout rings and their subring; Proc. Camb. Phil.Soc. 64 (1968) 251–264
[3] assuming the Axiom of Choice

Note that the converse is not true. Unfortunately, there isn't a simple counter-example I know of until we prove the famous Hilbert's Basis Theorem in section 11.6. For the reader who wants to know the counter-example, they can either skip to Hilbert's basis theorem on page 319 (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 10.5: 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 9.18, 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 10.1.4. 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).

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 10.6: Atomic Domain $\not\Rightarrow$ Unique Factorization Domain

Take the ring F_{0x} from example 10.2. 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 10.2, $x^6 = x^2x^2x^2 = x^3x^3$ 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 10.7: 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 9.4.1, $\frac{\mathbb{C}[x,y,z,w]}{(xw-yz)} = R$ is a noetherian ring. It is also an integral domain, since $(xw - yz)$ is a prime ideal: if $ab \in (xw - yz)$, then $ab = p(x, y, z, w) \cdot (xw - yz)$ (TBD ON SHOWING IT'S INT DOMAIN). 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 associatse. 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 ref:HERE, 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 10.5 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 11.3.1 on page 311). 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 A guide on how either $\mathbb{Z}[\sqrt{-5}]$ is not a UFD but they are Noetherian Domains.

(random idea) Show that if $M \subseteq R$ is a maximal ideal then M is Noetherian. In fact, is the prime ideals are finitely generated, then R is Noetherian.

Exercise 10.1.1

1. Let R be a finitely generated ring. Does that imply R is Noetheria

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 10.1.7: 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 $q_1, q_2, ..., q_n$ 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

(I believe the set of irreducible elements form a free abelian monoid)

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.

Lemma 10.1.1: 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

As a useful reminder from the beginning of this chapter:

1. $b|a$ if and only if
2. $a = bb'$ if and only if
3. $a \in (b)$ if and only if
4. $(a) \subseteq (b)$

Proof **:**
These should all come rather quickly and are left for exercise ref:HERE

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. If you have covered the section on monoids (section 8.1), then the set of irreducible elemets are free generators to a free abelian monoid. If you have not covered in detail, this fact can be skipped without loosing anything in the process; the key take away is that every element is associted to a multiset. As an example of this, Unique Factorization Domains make it very easy to define the concept of *greatest common divisor*.

Definition 10.1.8: 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 divior* 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*. To remember this fact, it might be useful to remember the definition of gcd in terms of ideals:

For a, b in a commutative ring R, the $\gcd(a, b)$ is an element $d \in R$ such that $(a, b) \subseteq (d)$ and (d) is the smallest ideal that contains (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 10.2, 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 10.1.5: 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:

$$\gcd(a, b) = p_1^{\min(n_1,m_1)} p_2^{\min(n_2,m_2)} \cdots p_k^{\min(n_k,m_k)}$$

***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 10.1.1, 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 10.1.1 $c|d$, as we sought to show

As a final word on gcd's for now, 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. As such, gcd domains do not play an important role when it comes to understanding factorization, and so will be relegated to a couple of exercises. We will revisit the concept of gcd soon (not gcd domain), but from a computational prospective. In particular, recall that in $\mathbb{Z}$, the gcd of two numbers can be rather quickly identified by using the Euclidean Algorithm (theorem 0.2.1). If you know some computer science, the Euclidean algorithm runs in approximately $O(\ln(n))$ time, read "Big-O of log n" (or at least the similar *binary* gcd runs in that time), which is about as fast as it gets. This turns out to be much faster than "manually" decomposing two elements into it's factorization and finding the common elements (especially for "large" elements with many factors[4]). We will return to this shortly. For now, gcd can be thought of as being well-defined in UFDs.

Properties of UFD's

As we have shown in example 10.2 and 10.7 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 you have to be able to "work backwards' from anywhere in it's ideal (TBD). However, if you are a Unique Factorization Domain, then that is a sufficient condition for every irreducible to also be prime!

[4]the concept of "large" elements will in fact be shortly defined when we study Euclidean Domains

Proposition 10.1.6: 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 10.1.2, 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 10.1.1, 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 in only b, then $b \in (a)$. If in only 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.

This is an incredibly powerful property of Unique Factorization Domains. In fact, if we also add the ascending chain condition for principal ideals for every element (just like for an Atomic Domain), then these two conditions precisely define a UFD:

Theorem 10.1.1: 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 10.1.6 every irreducible element is a prime element, and by corrollary 10.1.1 R is an atomic domain, and so by proposition 10.1.4 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 10.1.4, 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 $q_2 q_3 \cdots q_n \dot{q}_1 = p_1 p_2 \cdots p_n$, 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 $uq_1q_2 \cdot q_n = q_1q_2 \cdots q_n$. 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 ho 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.

PID's imply UFD's

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 10.1.4. By proposition 10.1.1, 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; take a moment to think about it 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

As the hint suggests, we first prove the following important intermediary result:

Proposition 10.1.7: 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.

Proof **:**
Let R be a Principal Ideal Domain. If an ideal of R is maximal, by proposition 9.4.2 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.

This theorem in it of itself is already gives us some interesting results. For example, given some information about $R[x]$, we can deduce some properties of R:

Corollary 10.1.3: $R[x]$ Pid Then R is a Field

If R is any commutative ring such that the polynomial ring $R[x]$ is a Principal Ideal Domain (or euclidean domain), then R is a field

***Proof* :**
This is just juggling the proposition we already covered. Since R is a subring of $R[x]$, and $R[x]$ is a Principal Ideal Domain, then R is an integral domain (if you're not convinced, prove this)[a]. Now consider $R[x]/(x) \cong R$. Since R is an integral domain, then by proposition 9.4.5, (x) is a prime ideal. By proposition 10.1.7, (x) is maximal. Then, by proposition 9.4.4, R is a field, as we sought to show.

[a] If you use the definition of a ring that doesn't require a multiplicative identity, don't forget to prove that R must also have a multiplicative identity

Example 10.8
Taking the contra-positive, if R is not a field, then $R[x]$ is not a Principal Ideal Domain! Since $\mathbb{Z}$ is not a field, then $\mathbb{Z}[x]$ is not a Principal Ideal Domain! In example 9.18[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 11.2.

Going back to trying to prove Principal Ideal Domains are Unique Factorization Domains, the last step to this proof is to show that irreducible elements are prime:

Proposition 10.1.8: 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 10.1.2 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 10.1.7, 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 lemma 10.1.7, (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 10.1.1, 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 10.1.4: $\mathbb{Z}$ 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 10.1.7 is a Unique Factorization Domain.

Note that this is another condition to check an integral domain is a Principal Ideal Domain by taking the contra-positive: If an element is reducible, but not a prime, then the ring cannot be a Principal Ideal Domain (ex. in $\mathbb{Z}[\sqrt{-5}]$, 3 is irreducible, but not a prime).

However, 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 9.18, the ideal $(2, x)$ from the ring $\mathbb{Z}[x]$ is not a principal ideal[5]. $\mathbb{Z}[x]$ is however a Unique Factorization Domain. This proof will be relegated to after theorem 11.3.1 in section 11.3.

(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?)

10.2 Euclidean Domains

(word on them here: bassically, in a field, we always have for any a, b that $a = bs$ and $b = at$ for some s, t both non-zero. The best we can do is divide with a remainder, so we would get $a = qb + r$ or $b = pa + s$ for some p, q that might be zero, and an r and s that is somehow "smalle" then b or a. As an example (DO INTEGERS).)

(So far, we have been exploring divisibility in terms of being able to directly factor numbers. However, Euclid in approximately the year _____discovered that there is a simple algorithm that allows you to find the greatest common divisor of any two integers. The existence of this algorithm cannot be underestimaed: 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? We are fortunate to have an intuition when it comes to $\mathbb{Z}$ in thinking about how numbers divide[6]: We know that if we multiply two numbers, the result will be bigger than the other. (Similarly with $F[x]$). However, not all Unique Factorization Domain's has this notion of "size" associated with each element.

In this section, we will be showing how certain Unique Factorization Domain, in particular Principal Ideal Domain's, can have an notion of "size" assigned to every element. It is often easier to define such a notion

[5] Contrasted to how $\mathbb{Q}[x]$ is a Principal Ideal Domain. Notice that just like corollary 10.1.3 mentioned, this implies $\mathbb{Q}$ is a field

[6] perhaps this intuition is deeper than I think? It's not just forunate??

rather than showing that we can factor or that every ideal is principal. Using this new definition of "size", proposition 10.2.1 can be complted using the fact that all Principal Ideal Domain's are Unique Factorization Domain's.

With these results, it seems that PID's, are quite similar to Euclidean domains, and they almost are. The way we'll show this is by defining a new type of norm which is weaker than the norm we defined (the norm we defined will be a special case of this one). We'll show that all PID's have this norm:

Definition 10.2.1: Dedikend-Hasse Norm

Define N to be the *Dedekind-Hasse norm* 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 10.2.1: 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 10.2.1: Defning Dedekind-Hasse Norm

dedekindHassNorm let R be a Principal Ideal Domain. Then there exists a multiplicative Dedekin-Hasse norm.

Proof **:**
p.289

Example 10.9
Showing that $\mathbb{Z}[\sqrt{-19}]$ is a Principal Ideal Domain but not a Euclidean Domain.

(either an exercice or a proposition: Use the Chinese Remainder theorem to decompsoe $R/(a)$ for nonzero $a \in R$). :exercise

Let $R = \mathbb{Z}[\sqrt{-5}]$, and pick $3 \in R$. (p.284). 3 is irreducible in R (as shown previously). However, 3 is not prime, since $(2 - \sqrt{5})(2 + \sqrt{5}) = 3^2$. So, $3|3^2$, However, 3 does not divides $2 \pm \sqrt{5}$, since it divides non of it's coefficients.

REVISED UP TO HERE

When we study rings, integral domains are a nice class, since we have no zero-divisors ($\forall a \in R \not\exists b \in R$ such that $ab = 0$). However, this does not imply that the elements are units ($\forall a \in R \exists b \in R$ such that $ab = 1$). This would be the nicest property to have, for it means that we have a full multiplicative inverse. However, this does not always happen (think of $\mathbb{Z}$). However, there is still a notion of some numbers being "divisible" to a more limited degree in these rings like $\mathbb{Z}$. For example, in $\mathbb{Z}$, and so forth! None of these are units, so there isn't a x^{-1}, but we still have some notion of being able to divide in a limited capacity

As we've shown in section 9.1.1, $\mathbb{Z}$ and rings (in particular integral domains) are intimately related. So the first question we might ask is "is there anything I already know which gives an idea of divisibility in $\mathbb{Z}$". One approach can be to think of the the euclidean algorithm. If we do so, we must get an idea of how this algorithm would operate, how it would terminate (since we a greatest common divisor exists once we find the remainder to be zero). So to start, we must find a way to say we are getting "closer" to our result. For this, we'll define a way to "measure" the size of elements:

Definition 10.2.2: Norm

Any function $N : R \to \mathbb{Z}^+ \cup \{0\}$ with $N(0) = 0$ is called a *norm* on the integral domain R. If $N(a) > 0$ for $a \neq 0$, define N to be a *positive norm*

Note: Notice it says $\mathbb{Z}^+$!! this means the norm is *never negative*. It is possible that non-negative values are zero, but all values must non-negative.

Example 10.10
here

1. If you have done some analysis, you might ask what's the relation between this definition of norm and our usual definition of the norm:

   Definition 10.2.3: Norm From Analysis

   Let V be a vector-space over $\mathbb{F}$. Then $N : V \to \mathbb{F}$ is a norm if:

   (a) $N(u + v) \leq N(u) + N(v)\ \forall u, v \in V$

   (b) $N(av) = |a|N(v)\ \forall v \in V,\ a \in \mathbb{F}$

   (c) $N(v) = 0$ if and only if $v = 0$.

   Usually, N is denoted $\|\cdot\|$ and $N(v) \cong \|v\|$

2. Another norm we might already know for $z \in \mathbb{C}$, $N(z) = z\overline{z}$. In fact, this idea will be the basis for many norms we'll be seing, as lots of examples will be derived from sub-rings of $\mathbb{C}$. In particular, recall in example ref:HERE, where $\mathbb{Q}(\sqrt{D}) \subseteq \mathbb{C}$. we'll have a family of examples $\mathbb{Z}[\sqrt{D}] \subseteq \mathbb{Q}(\sqrt{D})$ which will inherit this norm from $\mathbb{C}$. We'll present some members of this family as examples of Euclidean domains

Definition 10.2.4: Euclidean Domain

The integral domain R is said to be a **Euclidean Domain** (or posess a *division algorithm*) if there is a norm N on R such that 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

The main thing we can do in euclidean domains is run the division algortihm to keep "dividing" our numbers

[7]

$$a = q_0 b + r_0 \tag{10.1}$$
$$b = q_1 r_0 + r_1 \tag{10.2}$$
$$r_0 = 2_0 r_1 + r_2 \tag{10.3}$$
$$\vdots \tag{10.4}$$
$$r_{n-1} = q_{n+1} r_n \tag{10.5}$$
$$\tag{10.6}$$

Such a sequence will terminate since $N(b) > N(r_0) > \cdots > N(r_n)$, and it will eventually hit 0. Note that the solution is *not* necessarily unique!

Example 10.11

1. Every field with the euclidean norm is a euclidean domain! In fact, any norm 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 euclidena domains must be domains (meaning no zero-divisors)! So oddities with multiplying thinking about the "size" of a number will not happen (it would break the triangle inequality in the clasical definition of the norm).
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, so it's construction in that sense is limited.
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 norm: 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 11.2.
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 norm on $\mathbb{C}$. That is because $\mathbb{Z}[i] \subseteq \mathbb{C}$ ($\mathbb{Z}[i] \subseteq \mathbb{Q}(i) \subseteq \mathbb{C}$). We'll show that Gaussian integers ($\mathbb{Z}[i]$) is a Euclidean Domain.

[7] This would be *actual* division in the ring of fractions

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 norms 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 norm (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.

Notice how central defining a norm 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!

Over the next few examples, we will encounter a common construction. This construction is called the *intergral closure* of the rationals. I'm not too sure about their number-theory origins, Dummit and Foote say it comes in chapter 15. For our sake, it will be a source of constructions of examples and counter-examples:

Example 10.12

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 $D \equiv 1 \mod 4$, 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

$$\omega = \begin{cases} \sqrt{D} & \text{If } D \equiv 2, 3 \mod 4 \\ \frac{1+\sqrt{D}}{2} & \text{If } D \equiv 1 \mod 4 \end{cases}$$

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 norm 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

$$N(a + b\omega) = (a + b\omega)(a - b\overline{\omega}) = \begin{cases} a^2 - b^2 D & \text{if } D \cong 2, 3 \mod 4 \\ a^2 + ab + \frac{1-\sqrt{D}}{2} & \text{if } D \cong 1 \mod 4 \end{cases}$$

where

$$\overline{\omega} = \begin{cases} -\sqrt{D} & \text{if } D \equiv 2, 3 \mod 4 \\ \frac{1-\sqrt{D}}{2} & \text{if } D \equiv 1 \mod 4 \end{cases}$$

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

One important consequence of the existence of this algorithm is that every Eucildean domain is a PID

Proposition 10.2.2: 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

The biggest use of this proposition is that we can show R is *not* an Euclidean Domain if it is *not* a PID. The big thing I think having a PID gives is giving us an idea of what should be a norm on your ring. We will show that not every PID is a euclidean domain soon, but most examples of PID's you'll encounter will be.

Example 10.13

1. Since $\mathbb{Z}$ is a Euclidean domain, thus A PID. 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 10.2.5: Side Divisor

Let $\overline{R} = R^{\times} \cup \{0\}$ be the colletion 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 existnece 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 10.2.3: 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 10.14: : 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)

Greatest common divisor

With some notion of divisibility, we can expand on some more notions from 9.1.1 about the euclidean algortihm. We can ask if we do this procedure for a, b and find that we terminate with d, then can anything else divide a and b? In $\mathbb{Z}$, this is certainly the case. We will explore this question for commutative rings in general before exploring it in Euclidean Domains:

Definition 10.2.6: Greatest Common Divisors In Commutative Rings

Let R be a commutative ring and let $a, b \in R$ with $b \neq 0$. Then

1. a is said to be a *multiple* of b if there exists an element of $x \in R$ with $a = bx$. In this case b is said to *divide* a or be a *divisor* of a, and is written $b|a$
2. A *greatest common divior* of a and b is a nonzero element d such that
   (a) $d|a$ and $d|b$ and
   (b) if $d'|a$ and $d'|b$ then $d'|d$

   A greatest common divisor of a and b will be denoted $gcd(a, b)$ or (abusing notation) (a, b)

Thus, if $d|b$ and $d|a$ then $(a) \subseteq (d)$, $(b) \subseteq (d)$. We can translate our intuition on divisors into inclusions of

ideals!! Notice that the GCD condition means that (d) is the *smallest* ideal that contains both of those values, for example:

$$gcd(12, 20) = 4,\ (12) \subseteq (4)\,(20)\,(4)$$

note that both 2 divides 12 and 20, but and so $(4) \subseteq (2)$, which is true.

Thus, we can re-state the aforementioned definition like so:

Definition 10.2.7: Great Comon Divisors In Terms Of Ideals

If I is the ideal of R generated by a and b, then d is a greatest common divisor of a and b if

1. I is contained in the principal ideal (d) $((a, b) \subseteq (d))$
2. If (d') is any principal ideal such that $(a, b) \subseteq (d')$, then $(d) \subseteq (d')$

In other wods, the greatest common divisor of a and b (if such exists) is a generator for the [unique] smallest principal ideal containing a and b.

Notice that this definition can be broadened beyond Euclidean Domains. For example, $\mathbb{Z}[x]$ is not a euclidean domain since $\mathbb{Z}[x]$ is not a PID. However, we can still say $(2, x) \subseteq (1)$, and since $(1) = \mathbb{Z}[x]$, it is tautologically the smallest ideal which contains $(2, x)$ and so 1 is the greatest common divisor of 2 and x.

The following sums up the note as a *sufficient* condition

Proposition 10.2.4: 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 *Bezout Domain* (more generally, if every *finitely generated* ideal is principal, then we have a Bezout Domain. For Principal Ideal Domains, every ideal is principal).

Note that it's not a *necessary* condition.

We'll quickly show that the greatest common divisor is unique up to multiplication by a unit:

Proposition 10.2.5: Uniqueness Of Greatest Common Divisor (In Integral Domain)

Let R be an integral domain. If two elements d and d' of R generate the same principal ideal $((d) = (d')$, then $d' = ud$ for some unit $u \in R$. In particular, if d and d' are both greatest common divisors of a and b, then $d' = ud$ for some unit u

Proof **:**
If either d or d' are zero then the assertion implies both are zero and the rest becomes trivial, so let them both be nonzero. Let $(d) = (d')$. This means that $d = xd'$ for some $x \in R$ and $yd = d'$ for some $y \in R$.

Then we get

$$\begin{aligned} d &= xd' \\ yd &= yxd' \\ d' &= yxd' && \text{because } yd = d' \end{aligned}$$

From here, we can proceed in 2 ways:

1. Notice that $d' - yxd' = 0 \Rightarrow d'(1 - yx) = 0$. since $d' \neq 0$ that means that $1 - yx = 0$, i.e., $xy = 1$. But then that implies x, y are both $1, -1$, or a unit. In all cases, they are units
2. Since $d' = (yx)d'$, and the multiplicative identity is unique, this forces $xy = 1$, and the argument follows the same way

If d and d' are greatest common divisors, then they must each divide each other (condition 2 of definition 10.1.8), and so we'll get $(d') \subseteq (d)$ and $(d') \supseteq (d)$, meaning $(d') = (d)$, which brings us back to were we started, and so the same reasoning applies!

Finally, we return to Euclidean domains with this question. The great thing about euclidean domains is that we have a way of computing the greatest common divisor:

Theorem 10.2.1: Greatest Common Divisors In Euclidean Domains

Let R be a Euclidean Domain and let a and b be nonzero elements of R. Let $d = r_n$ be the last nonzero remainder in the euclidean algorithm for a and b describved at the begining of this chapter. Then

1. d is a greatest common divisor of a and b and
2. The principal ideal (d) is the ideal generated by a and b. In particular, d can be written as an R-linear combination of a and b, i.e., there are elements x and y in R such that

$$d = ax + by$$

Proof **:**
p.275

10.3 Factorization in the Gaussian Integers

(introduction, getting grip on Unique Factorization Domain, introduction to some number-theory)

(Richerd Borched and Aluffi both has a cool geometric perspective on this! There is even a relation to complex analysis! There is even a generalization for the sum of 4 square numbers and the quaternions!)

(Also how this was done to generalize the notion of prime numbers by Gauss)

Recall $\mathcal{O}$ represents a ring of quadratic integers. We introduced the field-norm on it, where we showed a number is a unit if and only if $N(\alpha) = \pm 1$. 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:

$$p = N(\pi) = N(\alpha)N(\beta)$$

meaninig either $N(\alpha)$ or $N(\beta)$ is a unit, meaning it's value is ± 1. Thus, we have that

If $N(\alpha)$ is $\pm a$ for a prime (in $\mathbb{Z}$), then α is irreducible in $\mathcal{O}$

We can further explore what are the prime elements in $\mathcal{O}$. Let $\pi \in \mathcal{O}$ be a prime element, that is,(π) is a prime ideal generaed in $\mathcal{O}$. Then $(\pi) \cap \mathbb{Z}$ is also a prime ideal. If $a, b \in \mathbb{Z}$, and $ab \in (\pi)$ (an automatically $ab \in \mathbb{Z}$) then either a or b is an element of (π), so a or b is in $(\pi) \cap \mathbb{Z}$. Note that $\mathcal{O}$ is an extension of the integers, so $p = N(\pi) \in (\pi)$. Therefore, $N(\pi) \cap \mathbb{Z} = p\mathbb{Z}$. THus, since $p \in (\pi)$, then $\pi | p$ in $\mathcal{O}$, and so the prime elements in $\mathcal{O}$ can be determined by how they factor with prime elements in $\mathbb{Z}$! Let's say $\pi | p$. Then $p = \pi\pi'$, and so

$$p^2 = N(p) = N(\pi)N(\pi')$$

and since $p^2 = pp$, this implies that either $N(\pi) = \pm p^2$ or $N(\pi) = \pm p$. If the former, then $N(\pi') = \pm 1$, making $p = \pi$ (up to associates)! If $N(\pi) = \pm p$, then π' is also irreducible (since $p = N(\pi')$), and so $p = \pi\pi'$ decomposes into two irreducibles. Therefore, prime elements in $\mathcal{O}$ are the prime elements of $\mathbb{Z}$, or prime elements of $\mathbb{Z}$ decompose into two irreducible elements (note that irreducible does not imply prime, as we've shown with $\mathbb{Z}[\sqrt{-5}]$)

If we let $D = -1$, we get $\mathbb{Z}[\sqrt{-1}] = \mathbb{Z}[i]$ (the Gaussian integers). As we've seen in example 10.11, $\mathbb{Z}[i]$ is a Euclidean Domain (hence a Principal Ideal Domain, hence a Unique Factorization Domain), and so irreducible elements are prime elements by proposition 10.1.6.

Let's take $\alpha = a + bi \in \mathbb{Z}[i]$. Then $N(\alpha) = \alpha\overline{\alpha} = a^2 + b^2$. Thus, by what we've just shown, p will factor in $\mathbb{Z}[i]$ into precisely two irreducibels if and only if $p = a^2 + b^2$, that is, p is the sum of two integes squares. If p is not the sum of two integer squares, it will remain irreducible! We've just found under what condition elements in $\mathbb{Z}[i]$ are prime elements!

$2 = 1^2 + 1^2$, therefore 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 conjugte irreducibles $a + bi$ and $a - bi$ can be associate. Any larger element will have $N(\alpha) > 1$.

Any number in $\mathbb{Z}$ is congruent to $0, 1, 2, 3 \mod (4)$. 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 squres must be congruetn to 1 mod 4. Therefore, if $p \equiv 3 \mod 4$ in $\mathbb{Z}$, then p is *not* the sum of squares, and remains irreducible (and thus prime) in $\mathbb{Z}[i]$.

Now suppose that $p \equiv 1 \mod 4$. Then we will show that p will *not* be a prime in $\mathbb{Z}[i]$! That is, p will factor as $p = (a + bi)(a - bi)$ with two irreducible, $a + bi$, $a - bi \in \mathbb{Z}[i]$, or equivalently, p is the sum of two squres $p = a^2 + b^2$. This proof will be Fermat's Theorem on sums of squres

We first prove a lemma giving us some divisbility informabout of p in $\mathbb{Z}[i]$:

Lemma 10.3.1: divisility Of p in $\mathbb{Z}[i]$

The prime number $p \in \mathbb{Z}$ divides an integer of the form $n^2 + 1$ if and only if p is etiher 2 or is an odd prime congruent to $1 \mod 4$

Proof **:**
p.290

Theorem 10.3.1: Fermat's Theorem On Sums Of Squares

The prime p is the sum of two integer squres, $p = a^2 = b^2$, $a, b \in \mathbb{Z}$ if and only if $p = 2$ or $p \equiv 1 \mod 4$. Except for interchanging a and b or changing the sign of a and b, the represetnation of p as a sum of two squres is unique.
Furthermore, the irreducible elements in the Gausisan integers $\mathbb{Z}[i]$ are as follows:

1. $1 + i$ (which has norm 2)
2. the primes $p \in \mathbb{Z}$ with $p \equiv 3 \mod 4$ (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 $p \equiv 1 \mod 4$ (both of which have norm p).

Proof **:**
p.291

Corollary 10.3.1: Condition Sums of Squares in $\mathbb{Z}$

Let n be a positive integer and write

$$n = 2^k p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k} q_1^{b_1} q_2^{b_2} \cdots q_l^{b_l}$$

where $p_1, ..., p_k$ are distinct primes congruetn to $1 \mod 4$ and $q_1, ..., q_l$ are distinct primes congruent to $3 \mod 4$. Then n can be written as a sum of two squres in $\mathbb{Z}$, i.e. $n = A^2 + B^2$ with $A, B \in \mathbb{Z}$, if and only if each b_i is even. Further, if this condition on n is satisfied, then the number of represetnations of n as a sum of two squres if $4(a_1 + 1)(a_2 + 1) \cdots (a_k + 1)$

Proof **:**
p.292

11

Exploring Polynomial Rings

We will also show waht happens if we weaken the condition on $R[x]$ (ex. UFD, Integral domain, commutative ring?) intuition on polynomial rings:

1. here

11.1 Recap

Definition 11.1.1: Polynomial Rings

here

Proposition 11.1.1: Basic Properties

Let R be an integral domain. Then

1. here
2. here
3. here

Proposition 11.1.2: Ideals of R and $R[x]$

Let I be an ideal of the ring R and let $(I) = I[x]$ denote the ideal of $[R]$ benerated 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 $R[x]$.

***Proof* :**
To show the isomorphism, use $\varphi : R[x] \to (R/I)[x]$ with the natural projection and the first isomorphism theorem for rings.

For the 2nd part of the proof, note that since I is prime, R/I is an integral domain. Then by proposition 11.1.1, $(R/I)[x]$ is an integral domain, and by what we've just shown this implies (I) is an integral domain.

Example 11.1
Take $R = \mathbb{Z}$, $I = n\mathbb{Z}$. Then

$$\mathbb{Z}[x]/(n\mathbb{Z}[x]) \cong (\mathbb{Z}/n\mathbb{Z})[x]$$

where the isomorphism is the natural projection map $r \mapsto \overline{r}$ (or in $\mod$ notation, $r \mapsto r \mod n$)

Definition 11.1.2: Polynomial Ring Over Severable Variables

The polynomial ring in the variables $x_1, ..., x_n$ with coefficients in R, denoted $R[x_1, ..., x_n]$ is defined inductively by

$$R[x_1, x_2, ..., x_{n-1}, x_n] = R[x_1, x_2, ..., x_{n-1}][x_n]$$

If $x_1^{d_1} x_2^{d_2} \cdots x_k^{d_k}$ is a term, then the degree of the term is

$$d = d_1 + d_2 + \cdots + d_k$$

and the tuple $(d_1, d_2, ..., d_k)$ is the multidegree of the term.

Example 11.2
You can look at $\mathbb{Z}[x, y]$ (p. 298)

Finally, we have proved this fact about polynomial rings which are Principal Ideal Domain force R to be a field:

Corollary 11.1.1: $R[x]$ Pid Then R is a Field

If R is any commutative ring such that the polynomial ring $R[x]$ is a Principal Ideal Domain (or euclidean domain), then R is necessarily a field

***Proof* :**
see proof of corrollary 10.1.3

11.2 Polynomial Rings over Fields

We'll now explore polynomiral rings with coeficients 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 divsion 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 11.2.1: 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 $b(x)$ nonzeor, 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 caes:

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 produt of polynomials 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 11.2.1: 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 11.2.1), we know $\mathbb{F}[x]$ is a euclidean domain. By 10.2.2, it is a Principal Ideal Domain, and since PID's are UFD's, it is also a Unique Factorization Domain

Example 11.3

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).

Also, a quick word on field extensions: If you extend from a field F to E, then by the uniqueness of the euclidean algorithm, everything will still factor the same way as before: it cannot inroduce new solutions.

One final fact we'll introduce as a result of what we talked about:

Corollary 11.2.2: 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 11.4: generalization of Famous infinite primes proof
Let F be a finite field. Then there are infinitely many prime elements in $F[x]$

Solution 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. ▼

Exercise 11.2.1

1. Let $R[x]$ be a Principal Ideal Domain. Show that $R[x]$ is a Euclidean Domain.

11.3 Polynomial Rings that are 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 $F[x]$ intead 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 $p(x) \in F[x]$, and $F[x]$ is a Unique Factorization Domain, then $p(x)$ factors in $F[x]$. 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! The way we'll prove this is first by uniquely factoring in $F[x]$, then "clear denominators" to obtain unique factorization in $R[x]$!

The first thing to do is make precise the noti noof comparing factorization of a polynomial in $F[x]$ and a factorization in $R[x]$

Proposition 11.3.1: 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 $F[x]$ then $p(x)$ is reducible in $R[x]$. More precisely, if $p(x) = A(x)B(x)$ for some nonconstatn polynomials $A(x), B(x) \in F[x]$, then there are nonzero elements $r, s \in F$ such that $rA(x) = a(x)$ and $sB(x) = b(x)$ both lie in $R[x]$ and $p(x) = a(x)b(x)$ is a factorization in $R[x]$

I find it helpful to sometimes think of as if $p(x), q(x) \in \mathbb{Q}[x]$, then $rp(x), sq(x) \in \mathbb{Z}[x]$ for appropriates $r, s \in \mathbb{Q}$.

Proof **:**
Let's say $p(x) \in F[x]$ 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:

$$d = p_1 \cdots p_n$$

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 (proposition 10.1.6), and so $R[x]/p_1R[x]$ is an integral domain (proposition 9.4.5) and $R[x]/p_1R[x] \cong (R/p_1R)[x]$ (proposition 11.1.2).

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. 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 F is a field, the elements of R becomes units in $F[x]$ (the units in $F[x]$ 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 $F[x]$

Corollary 11.3.1: Irreducibility In $R[x]$ And $F[x]$

Let R be a UFD, let F 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 $F[x]$. In particular, if $p(x)$ is a monic polynomial that is irreducible in $R[x]$, then $p(x)$ is irreducible in $F[x]$.

***Proof* :**
We'll proceed by the contrapositive: $p(x)$ is reducible in $R[x]$ if and only if $p(x)$ is reducible $F[x]$. The ($\Leftarrow$) direction is Gauss's Lemma, so let's go ($\Rightarrow$). Let's say $p(x)$ is reduble 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 $F[x]$ are not units, meaning $p(x) = a(x)b(x)$ is also a factorization in $F[x]$, completing the proof!

With this, we prove the main theorem of this section:

Theorem 11.3.1: 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 11.1.1), 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 F be the field of fractions of R. Then since $F[x]$ is a Unique Factorization Domain (proposition 11.2.1),

$p(x)$ can be factored uniquely into irreducibles in $F[x]$. By Gauss's Lemma, such a factorization implies a factorization in $R[x]$ whose factors are F-multiples of the factors in $F[x]$. 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 11.3.1, 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 $F[x]$. 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 11.3.1, each $q_i(x)$ and $q_j'(x)$ is an irreducible in $F[x]$. By uniqueness of factorization on $F[x]$, $n = m$, 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 $u \in F[x]$. It remains to show they are associates in $R[x]$. By proposition 11.1.1, the untis of $F[x]$ 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 11.3.2: 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 11.3.1, 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 11.5

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 11.6: 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 $F[x] : x^2 = (x-i)(x+1)$. so in particualr $p(x)$ *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 11.3.1, $\mathbb{Z}[2i]$ is *not* a Unique Factorization Domain! If it were, then it should've been reducible!
2. (give counter-example for $R[[x]]$)

11.4 Irreducibility Criteria

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, 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 irreducability criteria.

We'll be focusing on polynomials with one variable for now. By Gauss' Lemma, it suffices to consider the factorizations in $F[x]$ where F is the field of fraction of R. Ocassionally, we'll let R be an integral domain to generalize some results.

Let's start with finding factors of one (linear factors):

Proposition 11.4.1: Finding Linear Factors

Let F be a field and let $p(x) \in F[x]$. 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 $\alpha \in F$ such that $p(\alpha) = 0$

Proof **:**
Let $p(x)$ be a polynomial and $q(x)$ be a factor. If $q(x)$ is a factor of degree 1, then we can assume this factor is of the form $(x - \alpha)$ (up to unit) for some $\alpha \in F$, implying $p(\alpha) = 0$.

Conversely, if $p(\alpha) = 0$ for some $\alpha \in F$, then by the division algorithm in $F[x]$ we may write

$$p(x) = q(x)(x - a) + r$$

Since $p(\alpha) = 0$, then r msut be 0, and hence $p(x)$ has $(x - \alpha)$ as a factor

Proposition 11.4.2: 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

The reason for this is that if we have a degree higher than 4, then we can reduce to two irreducible polynomials of degre 2, like $(x^2 + 1)(x^2 + 1)$ in $\mathbb{Q}[x]$.

This next proposition let's us limit the number's we need to check when polynomials have constant coefficients. This can be generalized to coefficients to a Unique Factorization Domain, but we don't loose any details doing it with $\mathbb{Z}[x]$ since $\mathbb{Z}$ is initial and $\mathbb{Z}[x]$ is uniquely defined using the universal property of polynomials:

Proposition 11.4.3: Rational Roots Test

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 integer coefficients. If $r/s \in \mathbb{Q}$ 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 $p(x)$: $r \big| a_0$ and $s \big| a_n$. 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 $\mathbb{Q}(x)$.

***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, $s \big| a_n r^n$. Since $gcd(r, s) = 1$, 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 11.7: 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 corrollary 11.3.1, 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 11.4.2, 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).

IMAGE HERE USING TIKZ

We use another trick for these type of polynomials: We can use 11.1.2 ($R[x]/I[x] \cong (R/I)[x]$) to generalize the rational root test:

Proposition 11.4.4: Reduction Test

Let I be a proper ideal in the integral domain R and let $p(x)$ be a nonconstant *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 smalelr 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 11.1.2, 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 $\mathbb{Z}/n\mathbb{Z}$!) 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 11.8: Limitation to reduction method
Take $x^4 + 1 \in \mathbb{Z}[x]$. This polynomial is irreducible in $\mathbb{Z}[x]$, but is reducible for every prime. We will not prove this (chapter 14 in dummit), but you can for now verify emperically. TBD

So, if you get that a polynomial is reducible in $(R/I)[x]$, that unfortunately does not tell you it's reducible in $R[x]$.

Further examples are

1. p.309 – good exampels

A special case of proposition 11.4.4 that is quite useful is called Eisenstein's Criterion (or Eisenstein-Schönenmann Criterion, since Schönenmann discovered it first)

Proposition 11.4.5: Eisenstein's Criterion

Let P be a prime ideal of the integral domain 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 11.4.4 (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!

Corollary 11.4.1: Eisenstein's Criterion For $\mathbb{Z}[x]$

Let p be a prime in $\mathbb{Z}$ and let $f(x) = x^n + a_{n-1}x^{n-1} + \cdots a_1x + a_0$. 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 11.4.5 and using corrollary 11.3.1.

Here are some examples using Eisenstein's criterion, and some more tricks when finding irreducible polynomials:

Example 11.9: 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. on cyclotomic polynomials?
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.

11.5 Polynomial Rings over Fields II

With more knowledge of Unique Factorization Domain polynomial rings, we return to examining $F[x]$ where F is a field.

First, recall the following result:

Proposition 11.5.1: Recall: In PID, Prime ideal implies Maximal Ideal

The maximal ideals in $F[x]$ are the ideals $(f(x))$ genearted by irreducible polynomials $f(x)$. In particualr ,$F[x]/(f(x))$ is a field if and only if $f(x)$ is irreducible

Proof **:**
This follows from proposition 10.1.7 when applied to the Principal Ideal Domain $F[x]$.

Proposition 11.5.2: Chinese Remainder On Polynomials

Let $g(x)$ be a nonconstant monic element of $F[x]$ 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:

$$F[x]/(g(x)) \cong F[x]/(f_1(x)^{n_1}) \times F[x]/(f_2(x)^{n_2}) \times \cdots \times F[x]/(f_k(x)^{n_k})$$

Proof **:**
This follows from the Chinese Remainder Theorem (theorem 9.6.1), 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 $F[x]$, 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 $(f_i^{n_i}) + (f_j^{n_j}) = \mathbb{F}[x]$ by propositin 9.4.2).

Next, we concern ourselves with the number of roots of a polynomial over a field F.

Proposition 11.5.3: Finding Roots Of Polynomial

If the polynomial $f(x)$ has roots $\alpha_1, \alpha_2, ..., \alpha_n$ in F (not necessarily distinct) then $f(x)$ has $(x - \alpha_1) \cdots (x - \alpha_k)$ as factors. In particular, a polynomial of degree n in one variable over a field F has at most n roots in F, even counted with multiplicity.

Proof **:**
The first statement follows by induction on proposition 11.4.1). Since linear factors are irreducible, the second statement follows since $F[x]$ is a Unique Factorization Domain.

This last result has a fascinating consequence

Theorem 11.5.1: 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 fundametnal theorem for finitely generated abliean groups!

Since $F^\times$ is an abliean group, then:

$$\mathbb{Z}/n_1\mathbb{Z} \times \mathbb{Z}/n_2\mathbb{Z} \times \cdots \times \mathbb{Z}/n_k\mathbb{Z}$$

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 divindg 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 thatn n_k rots of the polynomial $x^{n_k} - 1$ in the field F, contradiction proposition 11.5.3!! Hence, $k = 1$, and the group is cyclic!

Corollary 11.5.1: $(\mathbb{Z}/p\mathbb{Z})^\times$ 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 11.5.1

Corollary 11.5.2: Isomorphisms Of Multiplicative Groups

Let $n \geq 2$ be an integer wit hfactorization $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n}$ 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 an da 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 **:**
p.314

11.6 Gröbner Basis

We now consider polynomials over several variables.

From theorem 11.2.1 that if F is a field, then $F[x]$ is a Euclidean Domain. By corollary 11.3.2, 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 11.5 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.

These rings will be greatly explored in chapter ref:HERE for HERE. For our purposes, they give us a way to characterise polynomial rings over several variables:

Theorem 11.6.1: Hilbert's Basis Theorem

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 11.6.1: 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 10.5 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 11.6, 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 11.6.2: 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 11.6.3: 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.

Exercise 11.6.1

1. Alternative HBT: (the one uses Diksons lemma)

12

Ordered Fields

I'm tempted to put a quick chapter here on ordered fields due to it's ties to valuation's and to how in many classes the real numbers are introduced axiomatically, one of the axioms being the order axiom, and then the completeness axiom is introduced on top of that.

Part III

Modules

explanation here

integarte:

https://mathoverflow.net/questions/61678/
what-is-the-general-geometric-interpretation-of-modules-in-algebraic-geometry
https://math.stackexchange.com/questions/1502957/intuition-for-when-modules-are-flat

intuitions

1. ring actions
2. better than rings
3. Groups seem to be well-thought intuitively as symmetries, while rings seem to be well-thought of as numbers. So, it's like having numbers act on geometric objects with symmetrices. However, limits to this intuition exist: A ton of geometric objects actually derive their structure from rings: that's practically the entire ethos of algebraic geometry, taking a ring and finding inherent geometry in it [1]

Also, we asked earlier if the set of ideals forms a ring. The answer is no, many axioms fail (additive inverse not well-defined). We'll instead analyse the structure of the ideals of rings in a different manner: we'll see that the ideals of the ring greatly affect how rings act on different groups! However, this is in a sense cheating: This will not be an algebraic structure on rings, but a way of analyzing rings through actions. To find the answer of what type of operation works on the set of ideals, it will be a Quantale. See also

https://en.wikipedia.org/wiki/Modular_lattice

When we talked about groups, we saw that the kernel of groups were groups, 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 begining 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 exlude 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 (ex $i : \mathbb{Z} \to \mathbb{Q}$) (TBD). 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. this

Modules behave better: they are more similar to abliean groups than they are to rings! If you have any sub-module, then the quotient will always be well-defined. (TBD).

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 we asked "how can we generalize this", 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)

[1] I'm talking about $\mathbf{Spec}(R)$

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

(I think there should be a chapter on some categorical constructions. There is a chapter on the tensor product, perhaps there should also be a chapter on the limit and colimit. You can also perhaps work with the chapter on prelude to homological algebra)

13

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).

13.1 Basic Definitions and Examples

Definition 13.1.1: [left]-Module

Let R be a ring (more generally a rng). A *left R-module* or a *let 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 suh 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 $rm = mr$.

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 13.1). 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 13.1, 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.

[1] That is, there must exist an $m \in M$ st $r \cdot m = s \cdot m$ implies $r = s$

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 13.1.2: 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, $\circ$, 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 13.1.3: 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 13.1.4: 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 13.1.5: 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 13.1.1: 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 13.1: 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 13.3.2. 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 $\mathrm{Ann}_R(M)$, 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 5.2.2)

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 24 and 27[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 13.3.1.

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 3.3.3

[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 13.2: 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 16, section 16.3.1

Vector-spaces

If the ring is a field, we give a special name to such a module:

Definition 13.1.6: 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 14, 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 14). 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 13.3: 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! (reference to how this can be useful! TBD)

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.

If $R = \mathbb{F}$ is a field, we give a special name to sub-modules:

Definition 13.1.7: Subspace

If $W \subseteq V$ is a submodule of a vector space, then we call it a *subspace*

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 13.1.1

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 3.3.4).
    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 3.3.5)
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)?

13.1.1 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

[2] We'll explore module generation in section 13.3

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 13.1.8: 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 13.1.8 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 26.

[3]Though a good basis in differential geometry would go a long way for intuition

From now on, when we say algebra or R-algebra, we will mean an associative algebra as defined in definition 13.1.8. Algebra homomorphisms are defined similarly to how they've been defined before (can you figure it out without looking at the following definition?)

Definition 13.1.9: 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 16.4.

Example 13.4: 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 $s \cdot r = r$.
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.

[4] natural meaning it satisfies the appropriate universal property

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 13.1.1: 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 9.2.1), 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
$$R[x_1, x_2, ..., x_n] \twoheadrightarrow S$$

13.1.2 Important Examples of 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 two other important classes of R-modules, where one will show how we can use modules to study abelian groups, while the other will lay the foundation on how we will generalize the Fundamental Theorem for Finitely Generated Abelian Groups and (if you've done linear algebra) Jordan's Decomposition of Matrices.

Example: $\mathbb{Z}$-module

Let the ring $R = \mathbb{Z}$, 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 13.1.2: 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.

Example: $\mathbb{F}[x]$-module

This example will be best motivated if you have done linear algebra before, in which case $\mathbb{F}[x]$-modules will in fact be the collection of linear transformation. If you have not studied linear algebra, then this section will be a prelude to the study of decomposition theorems that we will cover in chapter 15.

For this example, we'll have a lot of components. Let $\mathbb{F}$ be a field, let x be an indeterminate and let R be the polynomial ring $\mathbb{F}[x]$. Let V be an $\mathbb{F}$-module (i.e. a vector space over $\mathbb{F}$). Notice that we're using the same field for V and for R. Finally, let T be a linear operator on V, that is a linear transformation from V to V. This example will show you how to extend the $\mathbb{F}$-module V (i.e. a vector-space) into an $\mathbb{F}[x]$-module. To accomplish this, we'll take advantage of nice properties of linear transformations.

First, define

$$T^0 = I, \qquad T^n = T \circ T \circ \cdots \circ T \qquad (n \text{ times })$$

where I is the identity map from V to V and $\circ$ is function composition (which makes sense since T maps to V). Here's the crucial part in why we use a linear function: the composition and scalar multiplication of linear functions are linear functions

We'll now define the ring action of $\mathbb{F}[x]$-module (i.e. how polynomials act on elements in V). Let $p(x) \in \mathbb{F}[x]$ be a polynomial. Then define an action of the ring element $p(x)$ on the module element b by:

$$\begin{aligned} p(x)v &= (a_n T^n + \cdots + a_1 T + a0)(v) \\ &= a_n T^n(v) + \cdots + a_1 T(v) + a_0 v \end{aligned}$$

(i.e. $p(x)$ acts by substituting the linear transformation T for x in $p(x)$ and applying the resulting linear transformation to v). One should check that these rules satisfy the module axioms that make V into an $\mathbb{F}[x]$-module.

1. distributive via scalars and module elements
2. associativity comes from properties of linear transformations
3. the module is unital, since the constant polynomial of 1 will send v back to v.

Here's the key part; note that $\mathbb{F}$ is a sub-ring of $\mathbb{F}[x]$. As you'll see in homework, this means that V is also a $\mathbb{F}$-module, which is a fact we already knew, since we knew V was a vector-space. Thus, our new module with constant polynomials if simply our vector space, and the definition of $\mathbb{F}[x]$-module *extends* the action of $\mathbb{F}$ to an action of the larger ring $\mathbb{F}[x]$.

The interesting result from this comes from choosing different linear operators T! In general, there can by many different $\mathbb{F}[x]$-module structures for the same V.

Example 13.5

1. If $T = 0$, 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 $\mathbb{F}[x]$-module structure is just like the $\mathbb{F}$-module structure.
2. Another simple example is $T = I$, 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 $\mathbb{F}^n$ and let T be the "shift operator"
$$T(x_1, ..., x_n) = (x_2, x_3, ..., x_n, 0)$$
Since V could be a vector-space, let e_i be the usual basis vectors. Then applying this linear transformation to only the basis, we get:
$$T^k(e_i) = \begin{cases} e_{i-k} & \text{if } i > k \\ 0 & \text{if } i \leq k \end{cases}$$
So that if $m < n$
$$(a_m x^m + ... + a_0)e_n$$
]

So for example, if $m < n$,
$$a_m x^m + a_{m-1}x^{m-1} + \cdots + a_0)e_n = (0, 0, ..., 0, a_m, a_{m-1}, ..., a_0)$$

With this generalisation, the $\mathbb{F}[x]$-module will in fact describe *all* linear transformation of T on a vector space V over $\mathbb{F}$. We can now what do $\mathbb{F}[x]$-submodules look like. Let W be such a submodule. Since $\mathbb{F}$ is a sub-ring of $\mathbb{F}[x]$, W is also an $\mathbb{F}$-submodule, which since $\mathbb{F}$ is a field means that W is a sub-space. Next, W must still be closed under under the ring action, that is, $T(W) \subseteq W$. If one recalls MAT247, such a function is called *T-stable* or *T-invariant*. Thus, W is a $\mathbb{F}[x]$-submodule of V where W is *T-stable*.

(for more p.341)

Eventually, the theory behind these will reflect with *rational cannonical forms* and its *Jordan canonical forrms*.

Exercise 13.1.2

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?

13.2 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 13.2.1: 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 $\varphi \circ \varphi^{-1} = \varphi^{-1} \circ \varphi = \mathrm{id}$ 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) $\mathrm{im}(\varphi) = \varphi(M) = \{n \in N | \varphi(m) = n\}$
3. Define $\mathrm{Hom}_R(M, N)$ be the set of all R-module homomorphisms from M into N, $\mathrm{End}_R(M)$ the set of all R-module homomorphsism from M to itself, and $\mathrm{Aut}_R(M)$ the set of all all R-module isomorhpisms from M to itself. A function $\varphi \in \mathrm{End}_R(M)$ is called an *R-module enormophism*, and a function $\varphi \in \mathrm{Aut}_R(M)$ 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 13.2.1: 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 13.2.2: 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 13.2.3: 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 13.2.4: 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 13.6: 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 13.2.1: 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]

Exercise 13.2.1

1. Let M be an R-module. Prove that $\mathrm{Hom}_R(R, M) \cong M$
2. Let R^{op} be the opposite ring of R (see exercise ref:HERE), and let $\mathrm{End}_R(R)$ be all R-module homomorphisms from R to itself. Considering $(\mathrm{End}_R(R), +, \circ)$ as a ring show that $R^{\mathrm{op}} \cong_{\mathbf{Ring}} \mathrm{End}_R(R)$ [Hint ; think of Yoneda's lemma]

13.2.1 Quotient Modules

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

[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 $\circ$

an R-module structure on the abelian group M/N. It turns out that like abelian groups, every sub-module is "normal"!

Proposition 13.2.2: 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:

$$\begin{aligned}
\pi(rm) &= rm + N \\
&= r(m+N) && \text{(by definition of the action of } R \text{ on } M/N) \\
&= r\pi(m)
\end{aligned}$$

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 13.2.2: 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 13.2.1: 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.

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 4.7, we showed that mathematicians managed to find the building blocks of extensions, known as simple groups, and classify all finite simple groups, and in section 6.4, 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. Unlike groups, no large-scale classification project exists for R-modules for any particular ring R. Instead, the direct product of simple R-modules are studied for their simplicity, culminating in some very important results for representation theory. This is further explored in part VII, chapter 24. 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 17, and explored in full length in part VI.

The vocabulary for a sequence of R-module homomorphisms is quite important, and so we define the term here for future use:

Definition 13.2.3: 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 directio nand $\ker(d_i) = \operatorname{im}(d_{i+1})$ 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)

13.3 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 3.3.2: we take a subset of an R-module, and permit all possible 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. 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 3.3.4, 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 13.2.1, $\operatorname{Hom}_R(M, N)$ 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 $\operatorname{Hom}_R(M, N)$ have as an R-module, or even better that $\operatorname{End}_R(M)$ 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 14 and chapter 15, the reader can skip to part VII and read chapter 24 (only requiring knowledge from section 18.5 for the last 2 propositions of that chapter). A particular example of the above is when $N = R$ so that we consider $\operatorname{Hom}_R(M, R)$. 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[6].

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

[6] dual modules are central to Functional analysis and Harmonic Analysis

the fundamental theorem of finitely generated abelian groups (theorem 6.1.1) 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 15), 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":

$$r_1a_1 + r_2a_2 + s_1a_1 = (r_1 + s_1)a_1 + a_2$$

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 14, 15, and 25 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.

13.3.1 Direct Sum of Modules

We first update the terminology from the generation of groups and rings for the generation of modules:

Definition 13.3.1: 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 13.3.1 are indeed modules. If R is a *rng*, then it is not necessarily the case that $A \subseteq RA$.

Example 13.7: 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 (wich 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 6.1.2).
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. Let $\varphi : M \to N$ be an R-module homomorphism. Here is a neat way of writing down it's co-domain. Let
$$\varphi(M) + N = \{\varphi(m) + n | \varphi(m) \in \varphi(M),\ n \in N\}$$
Then this set represents all possible combinations of the image of M with any element of n. Moding by N, we get:
$$\frac{\varphi(M) + N}{N}$$
which is allowed since $N \subseteq \varphi(M) + N$, and we can mod by any submodule. Thus, we can write the image of φ in terms of sums of modules! Even better, if $A = M/B$, then $M = A + B$, since any element in M is some element of A (i.e. a coset representative hen taking $b = 0$), and an element of B.

Naturally, we would want to find when direct sums of modules behave like direct products, or more specifically external direct sums

Definition 13.3.2: 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 $\bigoplus_{i \in I} M_i \subsetneq \prod_{i \in I} M_i$.

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:

$$\mathbb{Z} \cong 2\mathbb{Z} \qquad \mathbb{Z} \neq 2\mathbb{Z}$$

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)[7]. 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 13.8: 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.

[7] 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 3.3

Proposition 13.3.1: 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
$$(n_i)_{i \in I} \mapsto ☼_i n_i$$
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
$$\begin{aligned} \pi(r \cdot (a_1, a_2, ..., a_k) &= \pi(ra_1, ra_2, ..., ra_k) && (13.1)\\ &= ra_1 + ra_2 + \cdots + ra_k && (13.2)\\ &= r \cdot (a_1 + a_2 + \cdots + a_k) && (13.3)\\ &= r\pi(a_1, a_2, ..., a_k) && (13.4) \end{aligned}$$
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 13.3.3: 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 13.3.1. 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 24, we will further explore how to find when a module M can be decomposed into direct products.

13.3.2 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[8]. 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 3.3.3 and 4.2.

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:

$$x = \sum_{a\in B\subseteq A} r_a \cdot a \quad \text{for some subset } B \subseteq A$$

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 13.3.1, 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

[8]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 $\mathbb{Z} \hookrightarrow \mathbb{Q}$

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 3.3.3 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.

OnRecovingABox 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 14

[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 13.3.1: 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:

$$(f + a)(a) = f(a) + g(a)$$
$$(rf)(a) = r(f(a))$$

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 13.3.4.

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 13.3.1

$$F^R(A) = Ra_1 \oplus Ra_2 \oplus \cdots \oplus Ra_n$$

Since $R \cong Ra_i$ (under $r \mapsto ra_i$) by proposition 13.3.1 again, the direct sum is isomorphic to R^n.

Definition 13.3.4: 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 $x = r_1 a_1 + \cdots r_n a_n$, for some $n \in \mathbb{Z}^+$, which by proposition 13.3.1 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 infintie 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[9]). Furthermore, the notion of rank is not necessarily well-defined if R is not commutative. We shall explore more of this in chapter 14.

Example 13.9: 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.
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 13.1, 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 13.2.1. 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
$$\psi(x) = \psi(\sum_{i=1}^{3} r_i s_i) = \sum_{i=1}^{3} r_i \varphi(s_i)$$

[9] Inspired by the link here

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 14, 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 3.3.4, 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 13.3.1: 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 13.3.1.

Proof **:**
exercise

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 14 and 17.

Measuring degree of non-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 13.3.5: Rank

For any integral domain R, the *rank* of an R-module M is the maximum number of R-linearly independent elements of M.

The notion of rank is similar to free rank, and it will in fact match exactly the idea of free rank for modules over PID's (and hence *a fortiori* over fields), but for modules over weaker rings, rank and generating elements need not match.

Example 13.10: Different Rank and Generation
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.

Using the notion of rank, we can see how a module might fail to be free. In the "worst case scenario", we can have single elements fail to be linearly independent:

Definition 13.3.6: 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 elemetns of M is denoted $\mathrm{Tor}_R(M)$ or $\mathrm{Tor}(M)$ is the context is clear.

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 13.3.7: 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 13.1(4) when discussing modules where the annihilator is trivial; such a module is given a name:

Definition 13.3.8: 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

Similarly, we have an equivalent notion for trivial torsion submodules

Definition 13.3.9: 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 trosion free is much more flexible than being free:

Proposition 13.3.2: 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 and R is free, then M is 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.

Exercise 13.3.1

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

13.3.3 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 13.3.3: 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 13.3.10: 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 we have not written

$$0 \to R^n \to R^m \to M \to 0 \tag{13.5}$$

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 (13.5) 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 as the above chain. We will prove this result in a moment, we will assume it for a bit to show some cool facts. Note that $\mathbb{Z}$ is a Principal Ideal Domain, and so for any [finitely generated? TBD] 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 (13.5) 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 (13.5) 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 13.3.11: 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[10].

[10] In chapter 17, we will introduce projective, injective, and flat modules, and then in part VI, we will introduce projective, injective, and flat resolutions

Example 13.11: Free Resolution

Let G be a finitely generated abelian group. Then by the fundamental theorem for such groups (theorem 6.1.1),

$$G \cong \mathbb{Z}^n \oplus \mathbb{Z}/n_1 Z \oplus \mathbb{Z}/n_2\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/n_k\mathbb{Z}$$

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 VI.

We now proceed to classify all R-modules M with free resolutions length 1. (finish here TBD)

13.3.4 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 13.3.12: 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 13.3.13: 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 13.3.2: 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 noethierna, an the uni prop of free modules).[11]. In proposition 9.4.3, 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)[12]. 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 13.3.2

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 the following chain of subrings:

$$\begin{gathered}\mathbb{R}[x] \\ \subseteq \\ \text{Analytic functions on } \mathbb{R} \\ \subseteq \\ \text{Analytic functions on } [0,1] \\ \subseteq \\ \text{Analytic functions on } (0,1) \\ \subseteq \\ \text{Analytic functions at } 0 \\ \subseteq \\ \text{Smooth functions near } 0 \\ \subseteq \\ \text{Former Power series}\end{gathered}$$

[11]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"?

[12]See Noetherian topologies for an interesting example

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

13.3.5 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 13.3.4: 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 $\varphi \in \mathrm{Hom}_R\left(\bigoplus_{i\in I} M_i, N\right)$

$$\varphi((\alpha_i)_{i\in I}) = \varphi(\alpha_1) + \cdots + \varphi(\alpha_n)$$

Now, it is immediate that each function on the right hand side is naturally isomorphic to a function $\varphi_i \,\mathrm{Hom}_R(M_i, N)$. 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:

$$\mathrm{Hom}_R\left(\bigoplus_i M_i, \bigoplus_j N_j\right) \cong \prod_{i,j} \mathrm{Hom}_R(M_i, N_j)$$

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 $\mathrm{Hom}_R(\oplus_i^m M_i, \oplus_j^n N_j)$ and matrix rings, especially if $M_i = N_j$. More on that in chapter 25, section 24.

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 13.12: 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 13.3.3

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).

14

Modules over Fields – Vector Spaces and Linear Algebra

This chapter focuses on the study of modules where the ring that is acting on the space is a field (i.e. a [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 13.1.6, 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 for modules over fields that is now considered standard:

[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., $M = \text{Span}(A)$

14.1 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 14.1: Vector Spaces

We already gave 4 examples in example 13.3. 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 $V = \text{Hom}_{\mathbf{Set}}(X, k)$ (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:

$$(f + g)(x) = f(x) + g(x) \qquad \forall x \in X$$

[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).

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. 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 $\mathrm{Hom}_{\mathbf{Set}}(X, k)$.

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 $c\dot{p}(x_1, x_2, ..., x_n, ...) = cp(x_1, x_2, ..., x_n, ...)$.
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

$$V := \left\{ f \in C^\infty(\mathbb{R}) \mid f'' + f \equiv 0 \right\}$$

Then using elementary ODE theory, we find that:

$$V = \mathrm{span}_{\mathbb{R}}(\{e^t, e^{-t}\})$$

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 usedis *Fourier Analysis*.

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 for some point $p \in M$, denoted $\mathrm{der}_p(M) \subseteq \mathrm{Hom}_{\mathbf{Smooth}}(C^\infty(M), \mathbb{R})$ forms a vector-space of dimension n, and is a common definition of the *tangent space* of the manifold M at the point p.

[a] See a textbook covering Fourier Analysis, ex. Pugh or Folland

Now, we shall develop a more sophisitaced notion of generating set:

Definition 14.1.1: 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
$$a_1m_1 + a_2m_2 + ... + a_nm_n = 0$$
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:
$$\begin{tikzcd} F^R(I) \arrow[r, "\varphi"] & M \\ I \arrow[u, "j"] \arrow[ur, "\iota"'] & \end{tikzcd} \tag{14.1}$$
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 (14.1) 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 (14.1) 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 emphesize the fact that it's ordered, a basis is called an *ordered basis*. In terms of indexing sets, the ordered eleemnts are generalized to elements having different indexes. Notice how being linearly independent is different from proposition 13.3.1(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 14.1.1: 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 naturla 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 14.2: 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 14.1.2: 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.

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!

Proposition 14.1.1: Vector Space, Maximally Linearly Independent Then 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, $v \in \text{span}(B)$, showing that B generated V. By lemma 14.1.1, 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 14.1.2: 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 14.1.1: 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 proposition 14.1.1, 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

[3]Though it is true for free groups, see appendix B

Example 14.3: Non Free Submodule

Take $\mathbb{Z}_6$ over itself, andconsider 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 proposition 14.1.1 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:

Theorem 14.1.1: 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}:

$$b_{k+1} = \beta_1 b_1 + \cdots + \beta_k b_k + \alpha_{k+1} a_{k+1} + \cdots + \alpha_n a_n \tag{14.2}$$

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:

$$\beta_1 b_1 + \cdots + \beta_{k+1} b_{k+1} + \alpha_{k+2} a_{k+2} + \cdots + \alpha_n a_n = 0 \tag{14.3}$$

By equation (14.2), we can make equation (14.3) a linear combination of elements of $(b_1, b_2..., b_k, a_{k+1}, ..., a_n)$:

$$(\beta_1 + \beta_{k+1}\beta_1')b_1 + \cdots + (\beta_k + \beta_{k+1}\beta_k')b_k + (\beta_{k+1}\alpha_{k+1}')a_{k+1} + \cdots + (\alpha_n + \beta_{k+1}\alpha_n')a_n = 0$$

[4] see theorem 15.1.1

[5] And will be used in chapter 25, section 24

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

$$\begin{aligned} 0 &= \beta_1 b_1 + \cdots + 0b_{k+1} + \alpha_{k+2}a_{k+2} + \cdots + \alpha_n a_n \\ &= \beta_1 b_1 + \cdots + \beta_k b_k + \alpha_{k+2}a_{k+2} + \cdots + \alpha_n a_n \\ &= \beta_1 b_1 + \cdots \beta_k b_k + 0a_{k+1} + \alpha_{k+2}a_{k+2} + \cdots + \alpha_n a_n \end{aligned}$$

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 14.1.2: 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 14.1.2: 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 **??**, 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.

Corollary 14.1.3: 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 $v = \sum_{\beta_i \in \mathcal{B}} b_i \beta_i \in W$, 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 14.2, and that vector spaces may have many different bases (of the same cardinality by the replacement theorem):

Example 14.4: Different Basis

1. Any uncountable subset of $\mathbb{C}[x]$ is linearly dependent, since there is a basis $\{1, x, x^2, ...\}$. Let $V = \mathbb{Q}^n$. Can you think of others? If you know about inner product spaces, can you think of an orthonormal [hamel][a] basis?
2. As we've defined in example 13.7the 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?

[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?

14.1.1 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 14.1.3: 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 14.1.4: 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

Definition 14.1.3: Rank And Nullity

If $\varphi : V \to W$ is a linear transformation of vector sapces 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 14.1.5: Properties of 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 between free

[6] In particular, if $T_p f$ is the tangent map at p, then $T_p f$ would drop rank

$\mathbb{Z}$-modules, but is certainly not surjective! 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 (see exercise ref:HERE). 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 transfomration

Example 14.5: Example
cool example with $|GL_n(V)|$

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 14.1.4: 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 **??**).

Proof **:**
exercise

Corollary 14.1.6: 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 14.1.4.

We shall further explore spaces whith this property in section 17.2.

Exercise 14.1.1

[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 17.2 for a generalized notion of splitting and section 6.4.2 for identifying groups that split

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 13.1]
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.

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:

14.2 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 14.2.1: 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

$$\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{14.4}$$

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 (14.4): $[\varphi(v)]_{\beta_W} = M_{\beta_V}^{\beta_W}[v]_{\beta_V}$.

Example 14.6: 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 14.2.1: 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 (14.4). 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 14.2.1: 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 $M_{\beta_U}^{\beta} V(\varphi)$ and $M_{\beta_V}^{\beta_W}(\psi)$. In fact, the coefficents of $M_{\beta_U}^{\beta_W}(\psi \circ \varphi)$ are exactly what you get by multiplying the other two matrices! This is summed up in the following theorem:

Theorem 14.2.2: 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

$$M_{\beta_U}^{\beta_W}(\psi \circ \varphi) = M_{\beta_V}^{\beta_W}(\psi) M_{\beta_U}^{\beta_V}(\varphi)$$

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 14.2.2: 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 14.2.2, (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 14.2.3: 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 14.2.2, 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 14.2.1. Theorem 14.2.2 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 14.1.5 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 14.2.2: 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 14.1.3)

Proposition 14.2.1: 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 14.1.5, φ is an isomorphism. Thus, there exists a linear transfomration φ^{-1} such that $\varphi^{-1} \circ \varphi = \mathrm{id}$. By using theorem 14.2.1, let B be the representation of φ^{-1} with respect to γ and β. Using theorem 14.2.2 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 14.2.1

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:

$$[f] = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

14.2.1 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:

$$F^{\oplus \beta_A} \xrightarrow{\psi^{-1} \circ \varphi} F^{\oplus \beta_B}$$

Then by theorem 14.2.1, 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 14.2.3: 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 14.2.4: 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

[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 14.3.4

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:

$$\varphi_2^{-1} \circ \alpha \circ \varphi_1 : F^{\oplus\beta_A} \to F^{\oplus\beta_C} \tag{14.5}$$

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:

$$\psi_2^{-1} \circ \alpha \circ \psi_1 : F^{\oplus\beta_B} \to F^{\oplus\beta_D} \tag{14.6}$$

now with some manipulation, we can transform equation 14.6 into equation 14.5 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 14.2.2: 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 14.2.4: 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. a choice of bases) that will make it "easier" to see what the linear transformation is doing? For the rest of this section, this question will be explored to great success, in particular, all linear transformation can be represented in "reduced form" and which can be algorithmically found.

14.2.2 Gaussian Elimination: Finding a Nice Basis

Let $\varphi : V \to W$ be a linear transofmration, 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 14.2.2, 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 14.2.3: 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 14.2.5: 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 14.2.3 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 14.2.6: 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 1.2.10. 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 14.2.3 becomes and "iff"

Proposition 14.2.4: 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* :**
p. 322 of Aluffi. Idea: take advantage tha 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*. The method is common enough to be given a name: *Gaussian Elimination*. [10]. Since Gaussian elimination usually reffers to operations on the left, we may think of this as a form of "row"-equivalence. TBD/

Corollary 14.2.5: Equivalence Of Matrices

Let $R = K$ be a field and $P \in M_{m,n}(R)$ be any amtrix. 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 entrices 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 transforamtion's betewen two finite dimensional vector-spaces. This also show's that if we look at a linear transformation in the right ay (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 14.1.4 to see how this intuition and the existence of compliments of subspaces work together.

Gaussian Elimination over Euclidean Domain

(Aluffi. p. 323, cool stuff, in particular get to smith normal form)

(a or on ho can't do it ith Principal Ideal Domain, but since PID's have some versio nof a gcd, ew ill still be able to produce a smith normal form though we ill require something more than elementary matrices)

14.2.3 Systems of Linear Equations Geometric Results

(How we alrady know how to solve $ax + b = c$ if $a, b, c \in R$ is a ring or $a, b, c \in k$ is a field.)

(Another geometric result: We have just classified all the hype-planes in terms of equations! If there is an n dimensional space, then if we have k independent equation, we have an $n - k$ hyperplane. Each hyperplane is a translation os the plane passing through the origin. Thus, it suffices to classify all subspaces to classify all hyper-planes!!)

(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 $\mathrm{GL}_n(R)$.)

[10] 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

Example 14.7: Exercices
Something with transpose. Perhaps even with conjugate? Get started with $\langle A, A*\rangle$ stuff?

Exercise 14.2.2

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 hte 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 $\mathrm{Hom}_F(V, V)$.)

14.3 Dual Spaces and Forms

(A good chunk of this section was inspired by Terrence Tao's blog: here)

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[11]). 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 injective map
$$\psi : G \to S_A$$

Notice how the first one defiend us what's happening btween elements in a sort of "local" way, while in the 2nd example we defiend it by using properties of the larger group S_A and how elements of G embedd 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

[11] If you know Fourier Analysis, think of phase space

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 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, 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 why 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[12] Hausdorff spaces B^*! Thus, the "dual" of Boolean Algebra is all continuous maps from B^* to $\mathbb{R}$! More generally, locally compact[13] topological spaces (which we can consider as a geometric object) are the dual of a type of algebra known as a C^*-algebras[14] (which we consider as algebraic objects): this is known as *Gelfand Duality*[15]. 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.

[12] A topological space in which the only connected sets are singletons

[13] a topological space in which for any point x, there exists a compact neighbourhood

[14] A Banach Algebra over $\mathbb{C}$ with a map known as an involution satisfying the properties of an adjoins

[15] This duality requires a good deal more topology and analysis than is part of the scope of this book. However, in part V, chapter 21.5, we will see the Zariski Duality which studies a geometric-algebraic duality without the need for much analysis or topology

14.3.1 Dual Spaces

Definition 14.3.1: 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 13.2.1, $\mathrm{Hom}_k(V, k)$ 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 14.3.2: Linear Functionals

Let V^* be the dual space of V. Then $f \in V^*$ is called a *linear functional* or *linear form*.

Example 14.8: 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 will assume $\int_{\mathbb{R}} |f| dx < \infty$. 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.

[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 dimensionae will in fact greatly complicate the structure of dual spaces[16]. 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[17]. 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 14.3.3: 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 14.3.1: Dual Basis

Let $\beta*$ be defined as in definition 14.3.3. Then $\beta*$ is a basis for V^*

Proof **:**
exercise

As a conseuqence, V^* has the same dimension as V

[16]This is one of the analytic properties that was discussed in the footnote in page 366

[17]Give references to books HERE, ex. Folland

Proposition 14.3.2: Reflectivity Of Finite Dimensional Vector Spaces

Let V be a finite dimensional vector space. Then $V \cong V^*$.

Proof **:**
By proposition 14.3.1, since V and V^* are of the same dimension, $V \cong V^*$.

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 ot recover the structure of V? The answer is almost: the following proposition shows how much information we can recover:

Proposition 14.3.3: 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. Remember to take advantage of the properties of linearity and that this must be true for *all* functionals.

The reason we can't completely recover V is because there can be many vector-spaces for which the $\mathrm{Hom}_k(V, k)$ is it's dual.

Example 14.9: Example
here

Since we can map any V to V^*, if we can map any $f : V \to W$ to an appropriate "dual" function f^* where composition is respected, then we would get that $(___)^*$ is a functor from Vect_k to itself. We indeed have such a concept:

Definition 14.3.4: Dual map (Transpose)

Let $f : V \to W$ be a linear transformation. then the *transpose* of f, denoted f^*, is defined as:

$$f^* : W^* \to V^* \qquad f^*(\varphi) = \varphi \circ f$$

Note that the tranpose of a function insures that maps that used to "out of" the space V will now "go into" the space V^* (and similalry for maps that go into V). Thus, this definition matches with our goal for V^* to be the dual of V. The name transpose comes from the matrix representation of f^*

Proposition 14.3.4: 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 corrollary, 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 iamge of f, we get the dual equivalent for f^*, given an appropriate "dual" for the kernel:

Definition 14.3.5: 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 ref:HERE)

Lemma 14.3.1: 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 14.3.5: 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 **:**
exercise

The first point in proposition 14.3.5 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 14.3.6: 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 *naturaly isomorphic* to V. To explain the idea of a natural isomorphism, we will show how V is *not* naturally isomorphic to V^*

Example 14.10: 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^* whre:

$$e_j^*(e_i) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

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 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.

Remark In 1955, mathematician NAME HERE discovered that the above map in infinite dimensions is always *injective*, leading to a natural embedding of V into V^{**}. If the map is also surjective, then V is called a *reflexive space*. If V is finite dimensional, V is always a reflexive space.

Exercise 14.3.1

1. Some things with infinite dimensions

2. Let $U \subseteq V$ be a subspace. Shwo that $U^{00} \cong U$. If $\widehat{(___)}$ is the natural embedding of V into V^{**}, then $\hat{U} = U^{00}$

3. Show that:

   1. $(U_1 + U_2)^0 = U_1^0 \cap U_2^0$
   2. $(U_1 \cap U_2)^0 = U_1^0 + U_2^0$

14.3.2 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[18])

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:

Definition 14.3.7: Bilinear and Multilinear Maps

Let V and W be vector spaces over k. The na 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 $v_i, v_j \in V$:

$$\varphi(v_1, ..., v_i, ..., v_j, ..., v_n) = -\varphi(v_1, ..., v_j, ..., v_i, ..., v_n)$$

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 16, we will show how we can study multilinear maps via linear algebra.

[18] There are also many examples in differential geometry, such as covariant/contravariant/mixed tensor fields on a manifold

Example 14.11: 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 16.4.
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 infintie 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 14.3.8: 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 14.3.2: 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 14.3.9: 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 14.3.2

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.

14.3.3 Trace and Determinant

In this section, we will go over to common functionals on $M_n(k)$ 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.

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 14.3.10: 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 14.3.6: 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 14.3.11: 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 $\mathrm{tr}(f)$ to be

$$\mathrm{tr}(f) := \mathrm{tr}([f]_\beta^\gamma)$$

Corollary 14.3.1: Trace Of Dual

Let $f : V \to W$ and $f^* : W^* \to V^*$ be the dual map of f. Then:

$$\mathrm{tr}(f) = \mathrm{tr}(f^*)$$

Proof **:**
exercise

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 it will provide by taking in n-copies of the vector-space, we will be able to detect when a matrix is invertible, and will give a very clear geometric picture of how a linear transformation effects a 1-cube in $\mathbb{R}^n$:

Definition 14.3.12: 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 $\det : (M_n(R))^n \to R$:

$$\det(A) := \sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^n a_{i\sigma(i)} \in R$$

Proposition 14.3.7: 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 14.3.8: 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 14.3.9: Invertability And Determinants

Let R be a commuative 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[19].

Proposition 14.3.10: 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)

Exercise 14.3.3

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 ($\overline{x + iy} = x - iy$)

[19] In particular, proposition 20.1.1

2. Let $A \in M_n(k)$. Define the *adjucate* of A to be $\operatorname{Ajd}(A) = C^T$ where C is the cofactor expansion matrix of A. show that if $\det(A) \neq 0$, then $A^{-1} = \det(A)^{-1} \operatorname{Adj}(A)$.

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: $(\det)'(I) = \operatorname{tr}$. 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 $(D_A \det)(X) = \operatorname{tr}(\operatorname{Adj}(A^t)X)$.

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 14.3.1: 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 14.3.13: Resultant

The determinant in theorem 14.3.1 $\operatorname{Res}(f,g)$ is called the *resultant*.

14.3.4 Geometry and Inner Product

A multilienar form of particular importance is known as an *inner product*, which is a symetric bilinear map with the extra property that $\varphi(v,v) \geq 0$ for any $v \in V$. On a high-level, defining an inner product defines

many geometric notions, namely *distance*, *size*, and *angle*, and is important in defining the notion of *length*. In this section, we will show how to define these notions

(should I somehow say that inner products will be in a sense greatly generlized in the tensor product section? Or not? hat it really does is allo a notion of "angles" to be generlized.)

(perhaps already introduce symmetric forms since the inner product is almost a symmetric form?)

Definition 14.3.14: Inner Product

Let V be a vector-space over k where k is either $\mathbb{R}$ or $\mathbb{C}$. Then define $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ to be

1. sesquilinear[a]: $\langle a + b, c \rangle = \langle a, c \rangle + \langle b, c \rangle$ and conjuate linear on the second term, and scalar's pass normally in the 1st, and conjugation through the second.
2. $\langle v, v \rangle \geq 0$ and $= 0$ if and only if $v = 0$
3. $\langle a, b \rangle = \overline{\langle b, a \rangle}$

[a] The term "sesqui" means $1\frac{1}{2}$, making it an apt choice of terminology

(also found this about generlizing to other fileds (commetned))

This section develops further on finding nice bases on a vector-space. It is not necessary for any other part of hte book (I think).

In many areas of mathematics we commonly upgrade vector-spaces with other structures add different ideas (ex. completeness, distance, measure, norm, algebra). We will do just that by adding inner prodcuts, and explain why that is interesting.

Riesz Representation?

Example 14.12: Inner Product

1. Let V be any vector-space over a field with characteristic zero 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:
$$\langle v, w \rangle := \sum_{\beta \in \mathcal{B}} v_\beta w_\beta$$
It should be checked that this is indeed a vector-space. This says that *every* vector-space has an inner product, and justifies our intuition view of vector-spaces in cartisian 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, 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 funcitonal anslysis (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.

:exercise

(do some exercises showing that inner product's really let you talk about geometric concepts. For example, given 3 points in $\mathbb{R}^2$ that form a triangle, find the area of that triangle. Or build-up to the concept of parrallelapiped and cross product; tease it's importance in defining inner product spaces in cool places in functional analysis).

Theorem 14.3.2: 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:

$$b_1 = \frac{v_1}{\|v_1\|}$$
$$b_i = \frac{v_i - \sum_{j=1}^{i-1} \langle v_i, b_j \rangle b_j}{\left\| v_i - \sum_{j=1}^{i-1} \langle v_i, b_j \rangle b_j \right\|} \qquad 2 \leq i \leq n$$

Since $v_1 \neq 0$ and $v_i \notin \text{span}(b_1, b_2, ..., b_{j-1})$ for eac h$j \geq 2$, the denominator are all nonzero. By construction, these vectors satisfy the condition in the theorem, and they are certainly orthonormal, completing the proof.

15

Modules Over PID's

The title of this chapter should be really *Modules Over Finitely Generated PID's.* but (justification here: The jist is finite, most the time we'll be dealing with this).

(I think I should)

In this chapter, we cover how finitely generated modules over PID's make module structure very nice. Just like how finitely generated free R-modules have a simple strutre (ex. free module M over R of rank n)

$$M \cong R^n$$

so too finitely generated modules over PID's have a similar relation:

$$M \cong R^k \oplus R/(a_1) \oplus \cdots \oplus R/(a_m)$$

We'll define the terms a_m and how to find k soon. Essentially, we'll split M 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 [nice] quotient rings of R, meaning the entire structure of M can be viewed through the structure of R!

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 6.1.1) Since $\mathbb{Z}$ is a finitely generated Principal Ideal Domain, if M is a $\mathbb{Z}$-module (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) (A word on why this is useful, perhaps both computationally, and why that is important, and theoretically)

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 [1]

15.1 Classification of Modules over PIDs

This entire next section is a build-up to be used in one part of Theorem 15.1.1, in particular, Dummit trying to show PID's are Noetherian Modules so he can pick a maximal submodule.

Returning to our problem of slitting linearly indepdendent and dependent elements, if R is [an intergral domain] of rank n, then if

$$r_1 y_1 + r_2 y_2 + \cdots r_n y_n = 0$$

with not all zero r_1,then the elements $y_1, y_2, ..., y_n$ are linearly dependent in some way. Since we're in a module, we can combine these together to form some element r' and y' such that

$$r' y' = 0$$

where $r' \neq 0$. In this way, we can think of y' and 0 being "linearly dependent". We give a name to such "not linearly independent" elements. Let:

$$\mathrm{Tor}(M) = \{m \in M | r \cdot m = 0 \text{ for some nonzero } r \in R\}$$

This is a submodule of M called *the* torsion submodule of M, and if N is any submodule of $\mathrm{Tor}(M)$, N is called *a* torsion submodule of M (so the torsion submodule of M is the union of all torsion submodules of M, i.e., is the maximal torsion submodule of M). If M has no torsion elements, then $\mathrm{Tor}(M) = 0$, so the torsion module always exists, so let $\mathrm{T}(M)$ represent the maximal torsion module. If $\mathrm{T}(M) = 0$, the module of M is said to be *torsion free*.

We can in fact find the element of R that create torsion modules. If we take the elements $r, r' \in R$ such that $r \cdot m = 0$ and $r' \cdot m = 0$, then $(r + r') \cdot m = r \cdot m + r' \cdot m = 0$ and $(rr') \cdot m = r \cdot (r' \cdot m) = r \cdot 0 = 0$. Furthermore, if $s \in R$ is any element of R, then $(sr) \cdot m = s \cdot (r \cdot m) = s \cdot 0 = 0$, so this is in fact an ideal of R.

For any submodule N of M, the *annihilator* of N is the ideal of R defined by

$$\mathrm{Ann}(N) = \{r \in R | rn = 0 \forall n \in N\}$$

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 15.1: 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

[1] wikipedia: Structure theorem for finitely generated modules over a principal ideal domain, Non-finitely generated modules

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

Fortunately, since our modules are finitely generated, the annihilator will always be non-empty, since if we annihilate the generators, we annihilate the entire module.

More words here justifying why we care about the next theorem

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, then we wouldn't get $x + y \in (x, y)$. if it were of rank 2, then $yx - xy = 0$ (i.e., not linearly independent). Any choice will fail: take a look at

`https://math.stackexchange.com/questions/74940/`
`are-submodules-of-finitely-generated-free-modules-always-free`

This next result let's us determine on what ring can we guarantee that submodules of free modules are free:

Theorem 15.1.1: M Free Over Pid of finite rank, Then Submodules is Free

Let R be a Principal Ideal Domain, let M be a free R-module of finite rank n, and let N be a submodule of M. Then

1. N is free of rank m, $m \leq n$ and
2. there exists a basis $y_1, y_2, ..., y_n$ of M so that $a_1y_1, a_2y_2, ..., a_my_m$ is a basis of N where $a_1, a_2, ..., a_n$ are nonzero elements of R with the divisibility relations

$$a_1|a_2|\cdots|a_m$$

Proof **:**
p. 460-462 – worth going over most again.

With this powerful way of finding finding a linearly independent set of a module over a Principal Ideal Domain, we can use it to find a very natural decomposition of the ring:

We will decompose a module M into the "free" part and "the rest". The structure of the rest will be the direct sum of cyclic modules (meaning the structure is actually very simple, since cyclic modules one of the simplest module structures!)

Recall that if R is a [unital] ring, then R over itself is a cyclic R-module, since $R \cdot 1 = R$. Then if (a) is an idea of R, it doesn't contain 1, so $R/(a)$ is still cyclic. since we're working in a Principal Ideal Domain, all the invariant factors will be of this form, meaning every $R/(a)$ is a cyclic R-module.

In general, if C is a cyclic module ($C = Rx$, where R is not necessarily commutative or unital), then

$$\pi : R \to C \qquad \pi(r) = rx$$

is surjective by assumption of $C = Rx$. Then by the first isomorphsim theroem

$$R/\ker(\pi) \cong C$$

and if R is a PID, $\ker(\pi) = (a)$ so

$$R/(a) \cong C$$

where $(a) = \mathrm{Ann}(C)$.

TBD

Theorem 15.1.2: Fundamental Theorem Of Existence: Invariant Factor Form

Let R be a Principal Ideal Domain and let M be a finitely generated R-module.

1. Then M is isomorphic to the direct sum of finitely many cyclic modules. More precicely,

$$M \cong_R R^r \oplus R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_m)$$

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 torsione 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 particualr M is a tortion module if and only if $r = 0$ and in this case the annihilator of M is the ideal (a_m).

***Proof* :**

M is finitely generated by assumption over the Principal Ideal Domain R. Since M is finitely generated, let $x_1, x_2, ..., x_n$ be the set of generator of M of minimal cardinality. Let R^n be the free R-module of rank n with basis $b_1, b_2, ..., b_n$. (perhaps word on previous theorem?? Take

$$\pi : R^n \to M \qquad \pi(b_i) = x_i$$

By the fist isomorphism theorem for modules,

$$R^n / \ker(\pi) \cong_R M$$

Now, by theorem 15.1.1, we can choose another basis $y_1, y_2, ..., y_n$ of r^n so that $a_1y_1, a_2y_2, ..., a_my_m$ is a basis of $\ker(\pi)$ for some elements $a_1, a_2, ..., a_m$ of R with $a_1|a_2|\cdots|a_m$. Thus,

$$M \cong_R R^n / \ker(\pi) = (Ry_1 \oplus Ry_2 \oplus \cdots \oplus Ry_n)/(Ra_1y_1 \oplus Ra_1y_2 \oplus \cdots \oplus Ra_my_m)$$

We'll re-write this equation by showing that the map

$$Ry_1 \oplus Ry_2 \oplus \cdots \oplus Ry_n \to R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_m) \oplus R^{n-m}$$

has the same kernel (dummit made this exercie 7 TBD)

Thus, we equivalently get that:

$$M \cong_R R^{n-m} \oplus R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_m)$$

if any a_i is a unit, then $R/(a_i) = R/R = 0$, so we can dispose of any summand in the right hand side with units generating the ideals. Letting $r = n - m$ proves the 1st assertion

For the 2nd, Since $R/(a)$ is a torsion R-module (the ideal (a) annihilates $R/(a)$), we get that M is corsion free if and only if $M \cong R^r$.

For part 3, similarly

Definition 15.1.1: Free Rank or Betti Number

The integer r in theorem 15.1.2 is called the *free rank* or the *betit number* of M and the elements $a_1, a_2, ..., a_m \in nR$ (defined up to multiplication by units in R) are caleld the *invariatn factors* of M.

Note we are yet to prove the invariant factors are unique, so we can't yet say they are *the* invariant factors, and similarly for the (a) free rank. We'll prove the decomposition is unique along-side another decomposition we're about to present refining the invariant factor form.

If we use the Chinese Remainder Theoem (theorem 9.6.1) on the cyclic modules of theroem 15.1.2 so that M is isomorphic to the direct sum of cyclic modules whose annihilators are as simple as possible, namely either (0) or generated by powers of primes in R, we'll get a 2nd representation.

In particular, if we let a be a nonzero element of a Principal Ideal Domain R, then since it's also a Unique Factorization Domain:

$$a = up_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n}$$

where each p_i are distinct priems in R and u is a unit.

more words here

Theorem 15.1.3: Fundamental Theorem Of Existence: Elementary Divisor Form

Let R be a PID and let M be a finitely generated R-module. Then M is the direct sum of a finite number of cylcic modules whose annihilators are either (0) or generated by powers of prime sin R, i.e.,

$$M \cong R^r \oplus R/(p_1^{\alpha_1}) \oplus R/(p_2^{\alpha_2}) \oplus \cdots \oplus R/(p_n^{\alpha_n})$$

where $r \geq 0$ is an integer and $p_1^{\alpha_1}, ..., p_n$ are positive powers of (not necessarily distinct) primes of R

Proof **:**
either here or before theorem? p.464

Definition 15.1.2: Elementary Divisors

Let R be a Principal Ideal Domain and let M be a finitely generated R-module as in theroem 15.1.3. The primes powers $p_1^{\alpha_1}, ..., p_n^{\alpha_n}$ (defined up to multiplication by units in R) are called the *elementary divisors* of M

words here

Theorem 15.1.4: 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 int odistinct 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 generaed, then each N_i is the direct sum of finitely many cycli modules whose annihilators are divisros of $p_i^{\alpha_i}$

Proof **:**
We proved this in the finite case in theorem _____. For the general case, here (p.465)

We give a name to the N_i modules

Definition 15.1.3: P_I-Primary Components

The submodules N_i in the previous theroem are called *p_i-primary components* of M.

At this point, we are able to show a decomposition exists, but we can't show it is unique. If it were unique, then once we find a decomposition, we know we found *the* decomposition. We first do a build-up

Lemma 15.1.1: PID Modules quotient by Prime

Let R be a PId and let p be a prime in R. Let F dentoe the field $R/(p)$.

1. Let $M = R^r$. Then $M/pM \cong F^r$
2. Let $M = R/(a)$ 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. Let $M = R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_k)$ where each a_i is divisible by p. Then $M/pM \cong F^k$

Proof **:**
short. p. 466

Theorem 15.1.5: Fundamental Theorem Of Uniqueness Of Factors

Let R be a Principal Ideal Domain. Then

1. Two finitely generated R-modules M_1 and M_2 are isomorphic if and only if they have the same free rank and the same list of invariant factors.
2. Two finitely genreated R-modules M_1 and M_2 are isomorphic if and only if they have the same free rank and the same list of elementary divisors.

Proof **:**
long p. 466-7

Corollary 15.1.1: Computation Shortcut For Invariant Factors

Let R be a Principal Ideal Domain and let M be a finitely generaetd R-module. Then:

1. The elementary divisors of M are the prime power factors of the invariant factors of M.
2. The alrgest invariant factor of M is the product of the larget 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 **:**
Follow's by the uniqueness given from 15.1.5

Corollary 15.1.2: The Fundamental Theorem of Finitely Generated Abelian Groups

See theorem 6.1.1.

Proof **:**
Take $R = \mathbb{Z}$ in theorem 15.1.2, 15.1.3 and _____. Note that the invariatn factors are listed in reverse order in Chapter 6 for ease of computation.

15.2 Rational and Jordan Canonical Form

(Some here needs to be re-writen)

With the theory build-up, we can now apply these results to simplify many ring structures.

We'll bring back the work we've done in section 13.1.2. Let $\mathbb{F}$ be an arbitrary field, let $R = \mathbb{F}[x]$, let V be a finite dimensional vectorspace over $\mathbb{F}$ (an $\mathbb{F}$-module), and let T be a fixed linear transformation on V ($T : V \to V$). We can consider V as an $\mathbb{F}[x]$-module where the element x acts on V as a linear transformation T. Since V is a finite dimensional vector space over $\mathbb{F}$, it is *a fortiori* a finitely generated vector-space over $\mathbb{F}[x]$ (since $\mathbb{F}$ is still part of it, it already has all the necessray generators). Thus, the theroems from the previous section will apply.

Note that any free $F[x]$-module would be isomrophic to the direct sum of copies of F, and so would be an infinite dimensional vector-space over F. So, if V is finite dimensional over F, then V would have to be

a torsion $F[x]$-module (i.e., have a free-rank of 0). Thus, by the fundamental theroem of invariant factors (theorem 15.1.2), V is isomorphic as an $F[x]$-module to the direct sum of cyclic, torsion $F[x]$-modules!

A key use of this is that we can now tell when two matrices represent the same linear transformation, i.e., when two given $n \times n$ matrices are similar.

(A word on group rep)

Here are a few reminder from chapter 14

Definition 15.2.1: Eigenvalues, Eigenvectors, Eigenspaces

1. here
2. here
3. here

define determinant of lin transform

how to find eigenvetor

Definition 15.2.2: Characteristic Polynomial

Let x be an indeterminate over F. The polynomial $\det(xI - T)$ is called the *characteristic polynomial* of T and will be denoted $c_T(x)$. If A is an $n \times n$ matrix with coefficeints in F, $\det(xI - A)$ is called the *characteristic polynomial of* A and will be denoted $c_A(x)$

Now, since V is a finitely generated torsion module over $F[x]$, then there exists an annihilator. Since we're working over a Principal Ideal Domain, the annihilator would be generated by a single polynomial! We give sucha polynomial a name:

Definition 15.2.3: Minimal polynomial

The unique monic polynomial which generates the ideal $\mathrm{Ann}(V)$ in $F[x]$ is aclled the *minimal polynomial* of T and will be dneotd $m_T(x)$. The uniqe monic polynomial of smallest degree which when evaluated at the matrix A is the zero amtrix is caleld the *minimal polynomial* of A and will be denotd $m_A(x)$.

We can see that the degree of the minimal polynomial is at most n^2, since the vector space of all linear transformations on an n dimensional vector-space is n^2. We'll that the minimal polynomial must divide the charactersitic polynomial, limiting this dimension to n.

Now, by theorem 15.1.2, we have the follwoing $F[x]$-module isomorphism:

$$V \cong F[x]/(a_1(x)) \oplus F[x]/(a_2(x)) \oplus \cdots \oplus F[x]/(a_m(x))$$

where $a_1(x), a_2(x), ..., a_m(x)$ are polynomial in $F[x]$ of degree at leat 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 $(a_m(x)$ (theroem 15.1.2), we get:

$$a_1(x)|a_2(x)|\cdots|m_A(x)$$

we will make a proposition of this for easy reference:

Proposition 15.2.1: 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 15.1.2)

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 $a(x) = x^k + b_{k-1}x^{k-1} + \cdots + b_1 x + b_0$ is any monic polynomial in $F[x]$ and $\overline{x} = x \mod (a(x))$.

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 15.2: 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 15.2.4: 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 hthe 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 $F[x]$) sicne 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 [2] 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 15.1.5. Thus, we can give this matrix a name:

[2] showing this eventually using the universal property?

Definition 15.2.5: 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 leat 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:

Theorem 15.2.1: Rational Canonical Form For Linear Transformations

Let V be a finite dimensional vector sapce over the field F and let T be a linear transfomration of V

1. There is a basis for V wit hrespect to whic 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 lesat one with $a_1(x)|a_2(x)|\cdots|a_m(x)$
2. The rational canonical form of T is unique

Proof **:**
some of what was said before? p.476 Dummit

remark about similary

Theorem 15.2.2: Similarty And RCF

Let S and T be linear transfomration of V. Then the following are equivalent

1. S and T are similar linear transfomrations
2. The $F[x]$-modules obtained from V via S and via$T$ are isomorphic $F[x]$-modules
3. S and T have the same rational canonical form

Proof **:**
p.476

Theorem (mabye should be corrollary) saying ibid for matrices

with that, we can give these a name:

Definition 15.2.6: Invariant Factor

The *invariant factors* of an $n \times n$ matrix over a field F are the invariant factors of its rational canonical form.

thereom (maybe should be corrollary) stating similarty with matrices

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 15.2.1: 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 somtimes 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 15.2.1, 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.

Similarly for the characteristic polynimal, since it's entries only rely on A(TBD?)

(2) here

With the theory established, we'll move on to trick for finding the invariant factors.

build-up to Caylay Hamilton

Lemma 15.2.1: Characteristic Polynomial For Block Matrix

Let $a(x) \in F[x]$ be any monic polynomial

1. The characteristic polynomial of the companion matrix of $a(x)$ is $a(x)$
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 matrice s$A_1, A_2, ..., A_k$, 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.

Proposition 15.2.2: Cayley-Hamilton Theorem

Let A be an $n \times n$ matrice sover the field F

1. The characteristic polynomial of A is the product of all the invariant factors of A
2. (The Cayley-Hamilton Theroem) The minimal polynomial of A divides the characteristic polynomial of A
3. The characteristic polynomial of A divides some power of the minmal polynomial of A. In particular, these polynomials have the same roots, not counting multiplicities.

Similarly if we replace the matrix A with it's respective linear trnsfomration T of an n-dimensional vector spac eover F.

***Proof* :**
p.479

Another way of phrasing the Cayley Hamilton theorem is that for a matrix A, there exists a polynomial $p(x) \in F[x]$ such that $p(A) = 0$. (justification of this interpretation). 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)

Build-up now on compution RCF

Theorem 15.2.3: Computation Of Invariant Factors

Let A be an $n \times n$ matrix over the field F. Using the three elementary row and column operations above, the $n \times n$ matrix $xI - A$ with entries from $F[x]$ can be put into the diagonal form (called the *Smith Normal Form* of A)

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & a_1(x) & & & \\ & & & & a_2(x) & & \\ & & & & & \ddots & \\ & & & & & & a_m(x) \end{pmatrix}$$

with monic nonzero elements $a_1(x), a_2(x), ..., a_m(x)$ of $F[x]$ with degrees at least one satisfying $a_1(x)|a_2(x)|cdots|a_m(x)$. The elements $a_1(x), a_2(x), ..., a_m(x)$ ar ethe invariant factors of A

***Proof* :**
Dummit left it as an exercice

15.2.1 Converting a square matrix to Ratioanl 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 15.3: 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 transformation (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:

$$\begin{pmatrix} -7 \\ 7 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} A \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -4 \\ 3 \\ 0 \end{pmatrix}$$

which gives the same P as before.

15.2.2 Jordan Cannonical Form

Lots of words to convert.

Let's assume the eigenvalues are in the field so that the characteristic polynomials splits into linear factors

Like before, we'll choose a basis for $F[x]/((x-\lambda)^k)$. We'll purposefully be choosing our basis so that our representation is particularly simple:

$$(\overline{x}-\lambda)^{k-1}, (\overline{x}-\lambda)^{k-2}, ..., \overline{x}-\lambda, 1$$

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 transformation T to x, and get

$$this$$

which in matrix form will be represtned 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 15.2.7: 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

up to permutation of blocks

define JCF for matrix and transformation

theroem stating exists a basis for V that makes T into JCF (and re-caps re-ordering property)

same theorme for matrices

corrollary saying if A is sim to D then D is A's JCF, also saying to diag matrices sim if and only if smae diag entries up to permutation

corrollary saying if F contains all eigenvalues, then F sim to a diag matrix over F if and only if minimal polynomial has no repeated roots

15.2.3 Changing From one Canonical Form to Another

15.2.4 Computation: Convert to JCF

16

Tensor Products Of Modules

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.

16.1 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 14, section 14.3, 14.3.3, and 14.3.4 (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! 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!, they let you define a way to "formally mulitply" two modules in such a way that the two modules structures combine to create a new single module. We already encountered something that sounds like we multiplied two modules, i.e. 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,

$$\begin{tikzcd}
M \arrow[dr, "\iota_M"] \arrow[ddr, bend right, "f"'] & & N \arrow[dl, "\iota_N"'] \arrow[ddl, bend left, "g"] \\
& M \times N \arrow[d, "f+g"] & \\
& L &
\end{tikzcd}$$

Thus, the map $f + g : M \times N \to L$ is already a linear map. However, we are interested in the case where $M \times N \to L$ is *not* linear. 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 mneumonic for remembering a tensor product is that tensor products let's us "multiply" elements of modules. The result of this multiplication will be inside a new module $\mathbb{R} \otimes \mathbb{R}$. Though this mneumonic 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 14.11). 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 16.4, 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 $c \cdot n, c \cdot m$. 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 13.1.1, 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.

16.1.1 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:

$$(m_1 + m_2, n) = (m_1, n) + (m_2, n) \tag{16.1}$$
$$(m, n_1 + n_2) = (m, n_1) + (m, n_2) \tag{16.2}$$
$$\tag{16.3}$$

Furthermore, we want scalars to commute:

$$(rm, n) = (m, rn) \tag{16.4}$$

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:

$$(m_1 + m_2, n) - (m_1, n) - (m_2, n) \tag{16.5}$$
$$(m, n_1 + n_2) - (m, n_1) - (m, n_2) \tag{16.6}$$
$$(rm, n) - (m, rn) \tag{16.7}$$

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 (16.1), (16.4), and (**??**) 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 (**??**), $[(rm, n)] = [(m, rn)]$. This will be key in ensuring our map will be *linear*. The R-module $F(M \times N)/K$ is given a name:

Definition 16.1.1: 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 16.1.1: 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:

$$(rm, n) = (m, rn) \tag{16.8}$$

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:

$$(mr, n) = (m, rn) \tag{16.9}$$

Furthermore, it is no longer clear what the R-module structure may be[1]. However, it is always an abelian group. Thus, inspired by equation (16.9), we define the following generlization of a R-bilinear map:

[1]Why is not as simple as defining $r \cdot (m, n) = (mr, n)$?

Definition 16.1.2: 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 (16.1), (16.4), and (16.9) to be the generators of a group K' and define:

$$F^{\mathbf{Ab}}(M \times N)/K' = \mathbb{Z}^{\oplus M \times N}/K'$$

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 16.1.1. 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 16.1.2: 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 16.1.1 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 16.1.3: (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 16.1: 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 16.1.4: 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 16.2: 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 16.3: 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 16.1.3: 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 16.1.1: 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 16.1.5: 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 16.1.4: 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 16.3

Theorem 16.1.5: 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 16.1.1

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.

16.2 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 16.2.1: 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 16.2.1: 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 16.2.2: Tensor of Vector-spaces

Let V, W be vector-spaces over F. Then (look at exercice 12 proof).

***Proof* :**

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 $\psi : \oplus_i V \to V \otimes_F V$, $\psi(V_i) = \sum V_i \otimes e_i$.
Now define $\varphi : V \times V \to \oplus_e V$, $\varphi\,(v \sum a_i e_i) = (va_i)$
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

$$\begin{aligned}(\phi \circ \psi)(V_i) &= \phi\left(\sum V_i \otimes e_i\right)\\ &= (V_i)\end{aligned}$$

$$\begin{aligned}\Big(\psi \circ \phi)(V \otimes \left(\sum a_i e_i\right)\Big) &= \psi(Va_i)\\ &= \sum Va_i \otimes e_i\\ &= V \otimes \sum a_i e_i\end{aligned}$$

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 16.1: type up!

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 16.2.3: 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 16.2.4: 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 16.2.1: 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 16.2.2: 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 16.4: Tensoring Algebras

1. the example elaborated was $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$

16.3 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:

$$(s_1 + s_2)n = s_1 n + s_2 n \tag{16.10}$$
$$s(n_1 + n_2) = sn_1 + sn_2 \tag{16.11}$$
$$s_1(s_2 n) = (s_1 s_2)n \tag{16.12}$$

and in particular for that last relation:

$$s_1(rn) = (s_1 r)n \tag{16.13}$$

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

$$(s_1 + s_2, n) - (s_1, n) - (s_2, n) \tag{16.14}$$
$$(s, n_1 + n_2) - (s, n_1) - (s, n_2) \tag{16.15}$$
$$(sr, n) - (s, rn) \tag{16.16}$$

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:

$$(s_1 + s_2) \otimes n = s_1 \otimes n + s_2 \otimes n \tag{16.17}$$
$$s \otimes (n_1 + n_2) = s \otimes n_1 + s \otimes n_2 \tag{16.18}$$
$$sr \otimes n = s \otimes rn \tag{16.19}$$

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:

$$\begin{array}{ccc}
R & \subseteq & S \\
N & \xrightarrow{\varphi_R} & S \otimes_R N
\end{array}$$

16.3.1 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. [2]. 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:

$$\begin{tikzcd} S \otimes_R L \arrow[r, "\psi"] & L \\ N \arrow[u, "\iota"] \arrow[ur, "\varphi"'] & \end{tikzcd}$$

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 16.3.1: 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 13.3.1), 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 $s\varphi(n)$: $\psi((s, n)) = s\varphi(n)$

[2] Notice these requirements and the definition of initial/final objects in 1.4.2, and of the proof that $\mathbb{Z}$ is initial in **Ring** (proposition 9.1.7) are similar? This is no coincidence: $S \otimes_R N$ is the initial object in the appropriate category. More on this in ref:HERE

$$\begin{tikzcd} F(S \times N) \arrow[r, "\psi"] & L \\ S \times N \arrow[u, "\iota"] \arrow[ur, "f"] & \end{tikzcd}$$

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 16.3.1: 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 16.5

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 16.6: 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.

16.4 Tensor Algebras, Symmetric and Exterior 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 we briefly said in section 4 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 of interst, this must be "closed" under iteration, that is $m_1 \otimes m_2 \otimes m_3$, $m_1 \otimes m_2 \otimes m_3 \otimes m_4$, 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

$$\mathcal{T}^k(M) = M \otimes_R M \otimes_R \cdots \otimes_R M \quad k \text{ factors}$$

and set $\mathcal{T}^0(M) = R$. The elements of $T^k(M)$ are called *k-tensors* (defined in section 4). 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 16.4.1: $\mathcal{T}(M)$ 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*) 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 $\Phi|_M = \varphi$

Note: For addition, it is just like with tensor products. that is, if two simple tensor are not in the same residue class, then
$$a \otimes b + c \otimes d$$
is the result of the sum of the two. Same with tensor algebras, so
$$a \otimes b + a \otimes b \otimes c$$
is the answer to the sum of those two tensors. Furthermore, 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.

***Proof* :**
The proof for 1 follows easily through computation, and the proof for (2) follows from the universal property of algebras with some additional computationsl checks.

Definition 16.4.1: 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 16.4.1: $\mathcal{T}^k(V)$ basis and dimension

If we take $M = V$ a finite dimensional vector space over a field k with basis $\mathcal{B} = \{v_1, ..., v_n\}$, Then the k-tensors
$$v_{i_1} \otimes v_{i_2} \otimes \cdots \otimes v_{i_k} \qquad \text{with } v_{i_j} \in \mathcal{B}$$
are a vector space basis of $\mathcal{T}^k(M)$ over k (with the understanding that the basis vector is the element $1 \in k$ when $k = 0$). 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* k in the (noncommuting) variables $v_1, ..., v_n$. The analogous result also holds for finitely generated free modules over any commutative ring (i.e. $R^s \otimes_R R^t \cong R^{st}$)

Example 16.7: 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
$$(r, \overline{p})(s, \overline{q}) = (r, \overline{p}, s, \overline{q}) = (rs, \overline{p}, 0, \overline{q})$$
TBD on how multiplication is defined. The ring $R/(x)$ 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
$$\mathcal{T}(\mathbb{Z}/n\mathbb{Z}) = \mathbb{Z} \oplus (\mathbb{Z}/n\mathbb{Z}) \oplus (\mathbb{Z}/n\mathbb{Z}) \oplus \cdots$$
It follows 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?

With some basic examples under our belt, let's start looking at the algebraic properties of tensor algebras. for one, multiplying two elements in $\mathcal{T}(M)$ gives us an element where its "order" is that of adding the "orders" of the multiplyers. This is called a "grading" or "degree" structure (reminicent of a polynomial ring)

Definition 16.4.2: 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 addititve 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 caled the *homogeneous component of* R *of degree* k.
2. An ideal I of a graded ring R is called a *graded ideal* if $I = \bigoplus_{k=0}^{infty} (I \cap R_k)$.
3. A ring homomorphism $\varphi : R \to S$ between two graded rings is called a *homomorphism of graded rings* if it respects the grading structure on R and S, i.e., if $\varphi(R_k) \subseteq S_k$ for $k \in \mathbb{N}$.

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 neccessarily 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 16.8: 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 generallym $R[x_1, x_2, ..., x_n]$ is graded since:
$$R[x_1, x_2, ..., x_n] \cong R \oplus \langle x_1, x_2, ..., x_n \rangle \oplus \langle x_1^2, x_1 x_2, ..., x_n^2 \rangle \oplus \cdots$$
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].

3. See D&F p.444

[a] Rees Algebras correspond to the notion of *Blow ups* in Algebraic Geometry

Definition 16.4.3: Graded Homomorphism

Let R, S be graded rings. Then a ring homomorphism $\varphi : R \to S$ is called a *graded homomorphism* if

$$\varphi(S_i) \subseteq T_i$$

kernel of graded homomorphisms are called *graded ideals* or *homogeneous*

The ideal I is graded if whenever a sum $i_{k_1} + \cdots + i_{k_n}$ of homogeneous elements with distinct degrees $k_1, ..., k_n$ is in I, then each of the individual summands $i_{k_1}, ..., i_{k_n}$ is itself in I!

Proposition 16.4.2: 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

Solution 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. (as you can check) ▼

The next two sections construct tensor algebras with more structure using quotients! After the next two chapters, we'll construct the same algebras, but by showing that they're also sub-algebras of a tensor algebra! For me, the quotient construction gives us intuition on what symmetric/exterior algebras are, while the sub-group construction gives us an intuition on the mereology of tensor algebras, and since it's sub-structure, it makes many theories from tensor algebras easily apply to these sub-algebras.

16.4.1 Symmetric Algebras

Symmetric algebras are the natural "abelianized" version of tensor algebras (similar to the construction of the commutator quotient G/G' from group theory). What does it mean to be the *natural* way of doing this? In our case, it means the construction of a commutative quotient ring of $\mathcal{T}(M)$ through which R-module homomorphisms from M to any *commuttaive* R-algebra must factor (will be made clear in an upcomming theorem)

Definition 16.4.4: 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 $\mathcal{S}(M)$.

Note: There are elements of higher degree than 2 in $\mathcal{C}(M)$ since the *ideal* $\mathcal{C}(M)$ of the *ring* $\mathcal{T}(M)$ must be closed under multiplication, multiplication is defined by squishing together the components of the tensors

The elements of $R = \mathcal{T}^0(M)$ an $M = \mathcal{T}^1(M)$ commutve in the quotient ring $\mathcal{S}(M)$ trivially. By contruction, it naturally follows that $\mathcal{S}(M)$ is a commutative ring. It is also easy to show that $\mathcal{S}(M)$ is a graded ideal by our construction, which by our previous proposition means its homogeneous components of degree k follow: $\mathcal{S}^k(M) = \mathcal{T}^k(M)/\mathcal{C}^k(M)$. (MORE INFO HERE p.445)

We'll show some basic properties of $\mathcal{S}(M)$ to not need to go through the construction of $\mathcal{C}(M)$ and compute quotients.

Theorem 16.4.2: 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. here
2. here
3. here

Proof. p.446 □

Corollary 16.4.1: Dim of $\mathcal{S}(V)$

$$\dim_F(\mathcal{S}^k(V)) = \binom{k+n-1}{n-1}$$

Proof. here □

16.4.2 Exterior Algebras

Exterior algebras are extremely immportant: they are the natural algebraic structure to represent *orientable area.*

Definition 16.4.5: 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 $m \in M$. The exteriro algebra $\mathcal{T}(M)/\mathcal{A}(M)$ is dentoed by $\bigwedge(M)$ and the image of $m_1 \otimes \cdots \otimes m_k$ in $\bigwedge(M)$ is denoted by $m_1 \wedge \cdots \wedge m_k$.

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 identityf 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 caleld the *wedge (or exterior) product.* By definition of the quotient, this multiplication is alternating in the sense that

the product of $m_1 \wedge \cdot \wedge m_k$ is 0 in $\bigwedge(M)$ if $m_i = m_{i+1}$ for any $1 \leq i < k$. Then

$$\begin{aligned} 0 &= (m + m') \wedge (m + m') \\ &= (m \wedge m) + (m \wedge m') + (m' \wedge m) + (m' \wedge m') \\ &= (m \wedge m') + (m' \wedge m) \end{aligned}$$

Show that multiplication is also anticommutative on the simple tensor

$$m \wedge m' = -m' \wedge m$$

for every $m, m' \in M$. Not ethat this is not true in general for $a, b \in \bigwedge(M)$, that is

$$ab = -ba$$

is not always true

SHOW HERE

Proposition 16.4.3: properties of exterior algebras

Let M be an R-module over the commutative ring R and let $\bigwedge(M)$ be its exterior algebra.

1. Ths 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 $i \neq j$

   (a) (universal property for algernating 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 defiend by

$$\iota(m_1, ..., m_k) = m_1 \wedge \cdots \wedge m_k$$

Solution here ▼

pertinent examples

1. If V is a 1 dimensional vector space over $\mathbb{F}$ 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

$$\begin{aligned} &\bigwedge^0(V) = F \\ &\bigwedge^1(V) = V \\ &\bigwedge^2(V) = V \qquad\qquad i \geq 2 \end{aligned}$$

and 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 $\mathbb{F}$. 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 we'd get some at least one of v, v' appearing twice in any simple product. (Another way of finding dimension in book)

Thus, it follows that

$$\bigwedge(V) = F|oplusV \oplus F(v \wedge v') \oplus 0 \oplus \cdots$$

Notice by our previous examples that exterior algebras are in fact always finite dimensional (over a field). THis fact will persist when increasing the dimension of V!

Corollary 16.4.2: dimension of $\bigwedge(V)$

$\dim_F(\bigwedge^k(V)) = \binom{n}{k}$.

Proof. intuitive □

example here (and show some similarity to determinant)

16.4.3 Homomorphisms of Tensor Algebras

If $\varphi : M \to N$ is any R-module homomorphism, then there is an inducd map on the kth tensor power

$$T^k(\varphi(: m_1 \otimes \cdots \otimes m_k \mapsto \varphi(m_1) \otimes \cdots \otimes \varphi(m_k)$$

IT follows that this maps sends generators of each of the homogeneous components of $\mathcal{C}(M)$ and $\mathcal{A}(M)$ to tehmselves. 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)

An interesting example is when $M = V$ is an n-dimensional vector space over a field $\mathbb{F}$ and $\varphi : V \to V$ is an endomorphism. in this case by _____, $\bigwedge^n(\varphi)$ maps the 1-dimensional space $\bigwedge^n(V)$ to itself. Let $v_1, ..., v_n$ be a basis for V, so that $v_1 \wedge \cdots \wedge v_n$ is a basis for $\bigwedge^n(V)$. Then:

$$\bigwedge^n(\varphi(v_1 \wedge \cdot \wedge v_n) = \varphi(v_1) \wedge \cdots \wedge \varphi(v_n) = D(\varphi)v_1 \wedge \cdots \wedge v_n$$

for some scalar $D(\varphi) \in F$ which is dependent on the function φ. In particular, by what we know from linear algebra, we know that any endormorphism (and more generally linear tranformation), we can associate it with an $A \in M_{n\times n}(F)$. Further, the can define a map $D : M_{n\times n}(F) \to F$ where $D(A) = D(\varphi)$. The use of the symbol D here is no coincidence, since D is in fact the determinant map!

Proposition 16.4.4: endormoprhisms of exterior tensors and the determinant

If φ is an endomorphism on a n-dimensional vector space V, then $\bigwedge^n(\varphi)(w) = \det(\varphi)w$ for all $w \in \bigwedge^n(V)$.

Solution The proof is quite intuitive. ▼

There's also a very natural generalisaiton of this:

Lemma 16.4.1: $\bigwedge^k(V)$ and l-tensors $l < k$

From Munkres Analysis.

Proof. Check munkres. Very intuitive. □

I don't get the example though

The inclusion map $\varphi : I \hookrightarrow R$ of the ideal (x, y) into the ring $R = \mathbb{Z}[x, y]$, both considedered 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

I don't see why the first equality is 0. There is one way I see it so, but then the other should also be zero.

16.4.4 Symmetric and Alternating Tensors

We've defined the symmetric and exteior algebras by quotienting away the values that should be considered "the same". 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 subalgebras. 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 very difficult, as we'll see in an example.

For any R-module M there is a natural left group action of the symmetric group S_k on $M \times M \times \cdots M$ (k factors) given by permutting the factors:

$$\sigma(m_1, ..., m_k) = (m_{\sigma^{-1}(1)}, ..., , m_{\sigma^{-1}(k)})$$

for each HERE

Definition 16.4.6: Symmetric and alternating k-tensors

here

more here

Example

Exercise 16.4.1

1. Let $v_1, ..., v^n \in \bigwedge^1(V)$ be linearly dependent vectors. Show that $v_1 \wedge \cdots \wedge v_n = 0$.

16.4.5 Interesting Applications

here

Application to Physics

Clifford algebra stuff?

Application to Differential Geometry

Weyl Algebra and Differential Forms?

Application in Representation Theory

group algebra, Lie algebra, univresal envelopping algebra?

Applicatoin in Analysis

stotastic integration, Banach Algebra?

16.5 Monoidal Categories

Monoidal Categories use the tensor notation and are interesting; should I add this here?
see: this

17

Extension Problem – Prelude to Homological Algebra

In this chapter, we take a closer look at the extension of R-modules. Recall that in section 4.6, we saw that a group G is an extension of a group A by B[1] if

$$G \cong_{\text{Grp}} B/A$$

Similarly, an R-module M is an extension of a module A by B if

$$M \cong_R B/A$$

In section 13.3, 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. As we saw in section **??**

These questions are not necessary for part IV and V, and so this chapter can be comfortably skipped until part VI. However, these concepts do show up in a couple of places as easter eggs for those who might want to see them.

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 departement is defined by what they do: You can have a department of finance, a departement of Public Relations, a department for cashiers, departement of logistics and so on. Given a certian departement, it might need to talk to some departement or it might not. For example, Let's say the departement of Public Relations needs to talk to the departement of cashiers (for example, to deal with customers issues) but it does not need to talk to the departement 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 departement and each edge represents whether the departement needs to be in touch with another departement.

[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

> So far, we have defined departements by what they do "internally" (ex. work with cashiers or logistical workers) and have given that they can interact with other departements. Let's now say that it is your job to create a departement which must be able to communicate with all departements. Notice that the defining feature of this departement is not something "internal", but by the fact that it relates to every other government – something "relational".

If we replace the term departement for R-module and "need for interaction" with homomorphism, then we have a similar idea of how to define an R-module *homologically*: 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: free-modules have the property that for any module M, there exists a surjective homomorphism $F(M) \to M$.

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.

Note that we can generalize this, and instead of using R-modules and R-module homomorphisms, we can use general objects and categories. We will in fact do this in part VI once we have gained some footing in the category of R-**Mod**.

Before going into how we will gain our footing, it must be mentioned that we imagined a directed graph in the analogy, but we will not be using directed graphs. However, there is another way of defining objects which *do* use direct graphs. We will explore these ideas in chapter 26 (Quiver theory).

The way we will gain our footing is by studying a particular problem which we have referenced way back in section ref:HERE. Given a group A and C, what are the possible group G such that

$$1 \to A \to G \to C \to 1$$

is a short exact sequence. Since we are working with R-modules, we can rephrase this as asking us to find an R-modules M given A and C such that

$$0 \to A \to M \to C \to 0$$

is a short exact sequence. This was called the *extensions problem* for groups which we have mentioned in section 4.7 (The 2nd part of Jordan Hölder's program). The way we will approach this problem is by studying how much information we can deduce about M given information about homomorphisms from some module, say D to A and C. That is, given we have $f : D \to A$ and $h : D \to C$ that commutes with the short exact sequence, what can we deduce any properties for some $g : D \to M$, and then use this information to deduce properties about M (since homomorphisms forces a lot of structure on the codomain). Conversely, if we have a homomrophism $f' : A \to D$ and $h' : C \to D$ that commutes with the short exact sequence, what can we deduce information about some homomorphism $g' : M \to D$, and then use this information to deduce properties of M? In terms of a commutative diagram:

$$\begin{tikzcd}
 & & D \arrow[dl, "f"'] \arrow[d, dashed, "g"] \arrow[dr, "h"] & & \\
0 \arrow[r] & A \arrow[r] & M \arrow[r] & C \arrow[r] & 0
\end{tikzcd}$$

and

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r] \arrow[dr, "f'"'] & M \arrow[r] \arrow[d, dashed, "g'"] & C \arrow[r] \arrow[dl, "h'"] & 0 \\
 & & D & &
\end{tikzcd}$$

As we will see, certain R-module M will be able to be completely defined by the homomorphisms of the R-modules on the two sides of the short exact sequence. Those who cannot be completely defined by these homomorphsisms will be "off" by a quantifiable degree. (TBD on building up to Ext group?)

After having explored this question, we will turn out attention to a second way of analyzing short exact sequence homologically using tensors. Recall in ref:HERE that the tensor product is left-adjoint to Hom. It is thus naturally related to the sort of questions we were asking earlier (TBD).

(Generalize into going into some basic homology?)

(In the final part of this chapter, we will venture into some quick applications beyond algebra? Like grad$\circ$ curl$= 0$? or DeRham cohomology? Finding holes in spaces? Put that in homological algebra proper?)

(A word somewhere that since ideals don't have 1's they are not rings, and so don't define short exact sequences)

(A word on how these concepts are introduced early since proj/inj/flat modules come up here and there?)

17.1 Categorical Preliminaries and Exact Sequences

In this section, it will be taken that chapter 2 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:

$$\begin{tikzcd}
F(b) \arrow[r] & F(c) \arrow[r] & F(f) \\
F(a) \arrow[u] \arrow[r] & F(d) \arrow[u] \arrow[ur] \arrow[r] & F(e) \arrow[u] \arrow[ul]
\end{tikzcd}$$

and if F is contravariant, the comutative diagram gets sent to:

$$\begin{tikzcd}
F(b) \arrow[d] & F(c) \arrow[l] \arrow[d] & F(f) \arrow[l] \arrow[d] \arrow[dl] \\
F(a) \arrow[ur] & F(d) \arrow[l] \arrow[ur] & F(e) \arrow[l] \arrow[ul]
\end{tikzcd}$$

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 17.1: 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 $C \to \mathbf{Set}$ 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 2, 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*[2]. 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 14.3 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:

[2] There will be examples of isomorphism of categories, famously the Galois Correspondence can be made into a category isomorphism between the right categories

Definition 17.1.1: 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

(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 17.1.2: 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 17.1.1: 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 17.2: Equivalence Of Categories
Let $\mathbf{Mat}_k$ be the categories whose objects are the natural numbesr $\mathbb{N}$ 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:

$$\begin{tikzcd}
a \arrow[r] \arrow[dr] \arrow[ddr, bend right] & b \arrow[r] \arrow[d] \arrow[dd, bend left] & c \arrow[dl] \arrow[ddl, bend left] \\
& n \arrow[d] & \\
& x &
\end{tikzcd}$$

For example, in the above diagram the colimit is $\mathbb{Q}$. To generlize this notion, the notion of commutative diagram must be abstracted:

Definition 17.1.3: 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 17.1.4: [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 17.1.5: [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 17.3: Limits And Colimits

1. products
2. equaliers and kernels
3. All universal properties are examples of limits and colimits.
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:

$$\cdots \xrightarrow{\varphi_{45}} A_4 \xrightarrow{\varphi_{34}} A_3 \xrightarrow{\varphi_{23}} A_2 \xrightarrow{\varphi_{12}} A_1 \tag{17.1}$$

Then $\varprojlim F$ is an object Aendowed with morphisms $\varphi_i : A \to A_i$ such that the following diagram commutes:

$$\begin{tikzcd}
 & & & A \arrow[dl, dashed, "\cdots"'] \arrow[dl, "\varphi_4"] \arrow[d, "\varphi_3"] \arrow[dr, "\varphi_2"] \arrow[drr, "\varphi_1"] & & \\
\cdots \arrow[r, "\varphi_{45}"] & A_4 \arrow[r, "\varphi_{34}"] & A_3 \arrow[r, "\varphi_{23}"] & A_2 \arrow[r, "\varphi_{12}"] & A_1
\end{tikzcd}$$

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**Mod**, we have that they always exist. Indeed, say we are given a diagram as in equation (17.1). 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 $A =_i A_i$.

5. a very similar construction occurs for colimits, in particular with the indexing category:

$$1 \to 2 \to 3 \to \cdots$$

(put the rest of the construction 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 17.1.6: 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 17.4: 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 $G : S$-**Mod** $\to R$-**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: $G^{\mathrm{ab}} = G/[G, G]$)

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 17.1.1: 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.

17.1.1 Exact Sequences

We generlize now the results of section 4.6:

Definition 17.1.7: 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.

Proposition 17.1.2: 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 17.5: Exact Sequences

Recall the automorphism group $\text{Aut}(G)$. We mentioned in section 5.5.1 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 17.1.2, 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 17.1.1: 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 4.6.

Definition 17.1.8: 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".

grpExtensionRemark 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 17.6: 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.

17.1.2 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 17.1.9: 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 17.7

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 17.1.3: 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:

$$\text{By commutativity} \Rightarrow \psi'(\alpha(a)) = \beta(\psi(a)) = \beta(b) = 0 \tag{17.2}$$

$$\text{Crucially, since } \psi', \alpha \text{ inj} \Rightarrow a = 0 \tag{17.3}$$

$$\text{Therefore} \quad b = \psi(a) = \psi(0) = 0 \tag{17.4}$$

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 17.1.4: 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 17.1.3.

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)

17.1.3 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 17.1.10: 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 exetnsion may equivalently be phrased in the language of homomorphisms

Proposition 17.1.5: 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

Solution 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 17.1.11: 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 17.1.6: 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.

Solution 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?

17.2 Dual and Projective Modules

(I think I will start by motivating projective module with dual modules, and then work on generalizing)

(I also got rid of Dummit's intuition of using the information from homomorphism from D to L and M to find information about M:

$$\begin{tikzcd} & & D \arrow[dl] \arrow[dr] & & \\ 0 \arrow[r] & L \arrow[r] & M \arrow[r] & N \arrow[r] & 0 \end{tikzcd}$$

this should somehow be re-integrated)

(explain too why we like functors: If $A \xrightarrow{f} B$ is a morphism in some category C and $\mathcal{F} : C \to D$ is a functor, then $\mathcal{F}(A) \xrightarrow{\mathcal{F}(f)} \mathcal{F}(B)$ exists in D, that is, they "preserve" the fact that there is a relation between A andd B. When we find a functor, we can rejoice in knowing some of the relations have "survived". An example where we cannot feel this joy is with Aut(_). THis is not a functor (see comment in latex document, link is causing not to compile)

(in particular, $\mathrm{Hom}(N, ___)$ and $\mathrm{Hom}(___, M)$ are natural functors. Let $R-$Mod be the category of R-modules. Then define the following functor from R-Mod $\to$ **Set**:

$$\mathrm{Hom}(M, ___) : D \mapsto \mathrm{Hom}_C(M, D) \qquad \mathrm{Hom}(___, N) : D \mapsto \mathrm{Hom}_C(D, N)$$

then $\mathrm{Hom}(M, ___)$ is a covariant functor and $\mathrm{Hom}(___, N)$ is a contravariant functor. If $\mathrm{Hom}_C(A, B)$ is also an abelian group and the maps are also homomorphisms, then we call a such a functor an *additive functor* and we have a map from $C \to$ **Ab**.

Next, 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 – we will find out how to transfer the rather complicated work of computing short exact sequences of modules to suitable short exact sequence's with the Hom functor.

Suppose we have a short exact sequence of R-modules $0 \to L \to M \to N \to 0$ – we shall constantly reference this short exact sequence. We know that for any R-module D, we can define the functor $\mathrm{Hom}(D, ___)$ and

$\mathrm{Hom}(___, D)$. The first thing we will ask is when are $\mathrm{Hom}(D, ___)$ and $\mathrm{Hom}(___, D)$ *exact functors*, that is, Given a short exact sequence like the one earlier, when is

$$0 \to \mathrm{Hom}_R(D, L) \to \mathrm{Hom}_R(D, M) \to \mathrm{Hom}_R(D, N)$$

exact, or when is

$$0 \to \mathrm{Hom}_R(N, D) \to \mathrm{Hom}_R(M, D) \to \mathrm{Hom}_R(L, D)$$

exact (notice that for $\mathrm{Hom}(___, D)$, the order is flipped since $\mathrm{Hom}(___, D)$ is a contravariant functor). We will start by studying $\mathrm{Hom}(D, ___)$.

We will start with the easier scenario: 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 an 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 17.8: Counter-Example

Take the $\mathbb{Z}$-module homomorphism $\pi : \mathbb{Z} \to \mathbb{Z}2\mathbb{Z}$ where $\pi(z) = z \mod 2$ 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 nothing mapping to $\mathrm{id}_{\mathbb{Z}/2\mathbb{Z}}$ (TBD – be more precise here)

More concisely, if $\varphi : A \to B$ is surjective, that is $A \to B \to 0$ is exact, then it need not be the case that $\varphi' : \mathrm{Hom}_R(D, A) \to \mathrm{Hom}_R(D, B)$ is surjective, that is $\mathrm{Hom}_R(D, A) \to \mathrm{Hom}_R(D, B) \to 9$ need not be exact.

We flesh out these observation in the following proposition

Proposition 17.2.1: 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 $F \in \mathrm{im}(\psi')$

Thus, $\ker(\psi') = \mathrm{im}(\varphi')$, showing that the sequence is exact at $\mathrm{Hom}_R(D, M)$ as well, as we sought to show.

We can make the previous "if" condition into and "if and only if" if we limit ourselves to $0 \to L \to M \to N$, that is

$$0 \to L \to M \to N \text{ is exact if and only if } 0 \to \mathrm{Hom}_R(D, L) \to \mathrm{Hom}_R(D, M) - > \mathrm{Hom}_R(D, N) \text{ is exact } \forall D$$

The "if" is shown by using the same proof as the previous proposition (notice we did not use surjectivity of φ in it). To show the "only if", recall that $\mathrm{Hom}_R(R, D) \cong D$ (see exercise re:HERE (in Generation of Modules). Then since R is always an R-module, simply choose $D = R$.

As we've seen, it need not be the case that $\mathrm{Hom}(D, ___)$ maps a short exact sequence to a short exact sequence (see example 17.8). Since there might not be a homomorphism from D to M that maps to a homomorphism from D to N. If a homomorphism $f : D \to N$ is mapped to by a homomorphism $F : D \to M$, we say that F is a *lift* of f (or *extensions* of f). In our persuit of trying to understand extensions, it would be useful to have a type of R-module, let's say P, such that $\mathrm{Hom}(P, ___)$ is exact. Such a P would be special module, and so we will give it a name:

Definition 17.2.1: Projective module

An R-module P is called *projective* $\mathrm{Hom}(P, ___)$ is an exact additive functor.

This is a good theoretical definition, but it doesn't lend itself well to computationally. The theorem following the proposition will give a nice characterization of all projective modules (explain why this propositin is here? Perhaps have it earlier in the book? It interupts the flow to just put it here, it would be more convinient to have it after the big thoerem, but I want examples of projective modules after the big theorem. Maybe have some, then introduce this prop, then give more?)

Proposition 17.2.2: Hom Commutes with Direct Sums

Let D, L, and N be R-modules. Then:

1. oplus transfers on left one
2. oplus transfers on right one

and this result can be extended inductively.In other words, Hom *commutes* with finite direct sumes.

Proof. homework □

(word on finite and infinite direct sums maybe?)

Theorem 17.2.1: properties of projective modules criterion

Let P be an R-module. Then the following are qeuivalent:

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 liefts 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, and two implies one comes from HERE.

$(2 \Rightarrow 3)$ 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:

$$\begin{tikzcd} & P \arrow[d, "f"] \arrow[dl, dashed, "F"'] & \\ M \arrow[r, "\varphi"] & N \arrow[r] & 0 \end{tikzcd}$$

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" by defining the map $P \to M$ where $M \twoheadrightarrow N$. In this way, we can see where the terminology "projective" comes from, since we can see that there is a sense in which we can go backwards and still have a homomorphism, and since M then has a surjective homomorphism onto N, we can imagine this as a projection. 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.

Recall that a free modules doesn't need to be free over free ring, we *can* make free modules over, say, finite rings:

$$M = \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \ [= R]$$

which is a module over itself. This module is projective since it is itself free, so it is it's own summand of a free module (i.e. "$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \cong \mathbb{Z}/4\mathbb{Z}$? is the single element of a summand to a free module). Taking $\mathbb{Z}/2\mathbb{Z}$, we see that it's *not* a free $\mathbb{Z}/4\mathbb{Z}$-modules. However, it *is* a summand of a free module, so it's projective! In other words

$$\text{free} \Rightarrow \text{projective}, \ \text{projective} \not\Rightarrow \text{free}$$

Summing this up in a corrollary:

Corollary 17.2.1: 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 17.2.1. 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:

Example 17.9: Projective Modules

Use lemma that comutes $\oplus$ and Hom to construct some more cool projective modules?

1. If F is a field, then F-modules (i.e. vector spaces over F) are free (theorem **??**), and hence projective
2. More generally, any ring R over itself is a free R-module or rank 1, and so is projective. We 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 17.2.1, 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 17.2.1 , 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 17.2.1. 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

$$0 \to \mathbb{Q} \to \mathbb{Z}^{\oplus \mathbb{Q}} \to \mathbb{Q} \to 0$$

does not split. (also found: Take $A = \mathbb{Z}[x_1, ...]$, $B = Q$. There is a surjection from A to B given by mapping x_i to 1_i. So, we define g to be this map, f to be the identity. But there is no possible $h : \mathbb{Q} \to A$ because the only units are ± 1)

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. (show or remind? that a finitely generated module over a Principal Ideal Domain is projective if and only if it's free?)
9. Given two projective modules P_1 and P_2, then $P_1 \oplus P_2$ is also projective.
10. (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 17.2.3: Projective Modules Over Pid'S

Let M be a projective module over a Principal Ideal Domain R. Then R is a free-module.

***Proof* :**
here. THis? proof online

Also found this: proj mod over PID's

(A word on how we will eventually measure how close an object is to being "exact", defining the Ext group in the process?)

17.3 Injective Modules

We know explore the dual question by taking $\mathrm{Hom}(___, D)$ instead of $\mathrm{Hom}(D, ___)$. This 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 17.10: 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.

Theorem 17.3.1: $\mathrm{Hom}(___, D)$ 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 theorem 17.2.1, 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)

(In dummit and foote, they also look at the $M \to N \to 0$ exact if and only if $\mathrm{Hom}_R(M, D) \to \mathrm{Hom}_R(N \to D \to 0$ exact scenario, and do a long proof. Not sure yet if I shoul include this)

Thus, similarly to how $\mathrm{Hom}(D, ___)$ is always right-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 17.3.1: 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 17.3.2: 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 **:**

$(1 \Rightarrow 2)$ here

$(2 \Rightarrow 3)$ 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

$(3 \Rightarrow 2)$ here

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. A criterion still exists, provided by Reinhold Baer who introduced the idea of injective modules in the 1940s:

Theorem 17.3.3: Baer's Criteron

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. 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 17.3.2, 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 $f(m) = f(m'n) \stackrel{\text{def}}{=} m' \cdot f(n) = m' \cdot x$ 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 generlize 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 17.11: 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 we saw in chapter 15, (TBD) no finitely generated abelian group is divisible, hence no finitely generated $\mathbb{Z}$-module is injective
6. Let $R = F$ be a field. Then (TBD) every F-module is injective! (split exact sequence if and only if basis can be extended, and we proved basis can be extended (does this prove it is injective? projective?))
7. (final example in dummit p. 397)

(modify this section to make it more compatible with talking about injective envelops)

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. There is a dual result for injective modules! We will first proof a build-up lemma:

Lemma 17.3.1: $\mathbb{Z}$-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 17.3.3. Then

$$M = F(A)/K \subseteq Q/K$$

showing that M is a submodule of an injective module, finishing the proof.

Now generalizing to general injective R-modules:

Lemma 17.3.2: R-module Submodules of Injective Modules

Let R be a ring with 1 and let M be an R-module. Then M is contained in an injective R-module.

Proof **:**
Dummit made this an exercise (p. 398)

Finally, we can prove a proper dual to the universal property of free modules:

Theorem 17.3.4: Universal Property Of Injective Hulls

HERE

***Proof* :**
Dummit recommends a book on p. 398 – I can check it out, or look at another place

Something here relates to representation theory.

17.3.1 *Duality Revisited

Thinking of that Richard Borcherds Video. (Rings 12 Duality and injective modules)

In section 14.3, 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 17.12: 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 $\left((\mathbb{Z}/8\mathbb{Z})^\times\right)^*$ 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 25, section 25.2. In character theory, following elements of the dual are also called *Dirichelt Characters*

(The rest is in this video: commented)

17.4 Flat Modules

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 17.4.1: 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 17.4.2: $D \otimes_R$ ___ 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 ___, 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 17.4.1: 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)

Solution already outlined earlier ▼

Definition 17.4.1: 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 17.4.1: 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 17.13: 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 17.4.2: tensor product of projective modules

If R is commutative then the tensor product of two projective R-modules is projective

Proof. p.402 □

17.5 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[3])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!)

[3]It was a geometric proof involving fitting squares of different sizes together

18

Field Theory

A reminder that a field F is a unital commutative division ring (definition 9.1.5). 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 9.4.2) 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].

18.1 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 9.4.1)

Proposition 18.1.1: 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.

[1] A famous book explaining these notions is *Althor's Complex Analysis*, in which the first chapter explores complex numbers

Proof **:**
This is proposition 9.4.2 (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 18.1.1: 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*.

Example 18.1: 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 extensiosn 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 9.6, 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 considerd an overload of notation, but since the kernel of a ring homomorphism between fields must *always* be trival, it is safe to see that K/k would not represent a quotient of the field K by k.

Another thing we can obeserve 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 $f\text{—} : \mathbf{Fld} \to \mathbf{Ring}$.

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 18.1.2: 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 18.1.3: 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 18.1, $\mathbb{C}/\mathbb{R}$ 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 example **??**, 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 18.1.4: 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 14.2, 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 18.5.4).

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 19.1.4, 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 9.1.7):

$$\mathbb{Z} \to k,\ n \mapsto \begin{cases} \underbrace{1+1\cdots+1}_{n \text{ times}} & n > 0 \\ -(\underbrace{1+1\cdots+1}_{n \text{ times}}) & n < 0 \\ 0 & n = 0 \end{cases}$$

We will use multiplication notation $n \cdot 1$ as a shorthand [2]. Recall $\operatorname{im}(\mathbb{Z})$ 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:

Definition 18.1.5: 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 defiend 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 18.1.6: 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.

[2] Note that we can also have defined it as a $\mathbb{Z}$-module on k

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 $\mathrm{ch}(k) = n$, $n \geq 0$, and $\varphi : k \to K$ is a homomorphism, then $\mathrm{ch}(K) = n$ (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

Roots of Unity

As we have shown in theorem 11.5.1, $F^\times$ is locally cyclic. We recall the theorem here for convinience:

Theorem 18.1.1: $F^\times$ 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 18.1.7: 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 18.2

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 18.1.8: 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]

[3] Remember by the fundamental theory of finitely generated abelian groups that any finitely generated abelian

Exercise 18.1.1

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 3.3.2: Take all subfield of F containing k and A, and intersect them. Similarly to section 3.3.2, 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 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 18.1.9: 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 18.1.10: 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 $F[S] = F[\alpha_1, \alpha_2, ..., \alpha_n]$.

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?

[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

Proposition 18.1.2: 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 3.3.2 with appropriate modifications for the added operation of multiplication and inversing.

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 18.1.11: 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 3.3.2, we can describe K_1K_2 as the intersection of all subfields contianing K_1 and K_2 (or any number of subfields).

Example 18.3: 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 $\sqrt[6]{2}$

18.2 Field Extensions

We now continue our exploration of field extensions. As we saw in proposition 18.1.2, 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 18.2.3, 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!!

Taking a moment

Theorem 18.2.1: 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.*

(A word on how Cauchy bassically discovered it first, and Kronecker generlized it? In the book on History of Number Daniel sent)

***Proof* :**
Consider

$$K = k[x]/(p(x))$$

Since $k[x]$ is a Principal Ideal Domain (corrollary 11.2.1) and since $p(x)$ is irreducible (so prime by proposition 10.1.8), then $p(x)$ is maximal (proposition 10.1.7). Then $k[x]/(p(x))$ is a field by proposition 9.4.4. 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 18.1.1, 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

$$\begin{aligned} p(\overline{x}) &= \overline{p(x)} && \text{since } \pi \text{ is a homomorphism} \\ &= p(x) \mod (p(x)) && \text{in } k[x]/(p(x)) \\ &= 0 \end{aligned}$$

[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 9.2.1), $g(x) \mapsto g(\alpha)$ 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 extensiosn (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 18.4: 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 18.2.4.

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

[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 18.5.1

some polynomial equation), and isomorphically identify $K \xrightarrow{\sim} k(\alpha)$ (which we will show in more detail in theroem 18.2.3).

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 18.5.2, 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 18.5: 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 11.4.2). 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 19.1.1. 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 19.2.4, 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!

[7] Though, this is not much of an exception since $\varphi = \frac{1+\sqrt{5}}{2}$

Theorem 18.2.2: 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 $\theta = x \mod (p(x)) \in K$. 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

$$b_0 + b_1x + \cdots + b_{n-1}x^{n-1} \equiv 0 \mod (p(x))$$

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 18.2.4).

As a reminder of arithmetics in $F[x]/(p(x))$:

Corollary 18.2.1: Shortcut On Arithmetics In K

Let K be as in theorem 18.2.1, and let $a(\theta), b(\theta) \in K$ be two polynomials of degree $< N$ in θ. Then addition in K is defiend simply bu usual polynomial addition and multiplicatin in K is defiend 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 18.2.1 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 18.6: 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 18.2.2, $K/\mathbb{R}$ 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.
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:

[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

Theorem 18.2.3: 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 18.2.1 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 9.5.2), 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 $1[= \alpha^0], \alpha, \alpha^2, \ldots$ are all linearly independent over F in $F(\alpha)$ (since $1[x^0], x, x^2, \ldots$ 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 18.2.2, $[F(\alpha) : F] = \deg m_{\alpha,k}(x)$, and so the extension is finite, completing the proof.

This theorem motivates the following vocabulary:

Definition 18.2.1: 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 18.2.2)

Definition 18.2.2: 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 18.2.2: 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 18.2.3

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 18.2.1 to study study irreducible polynomials that aquire a root.

Going back to the minimal polynomial that we found in the proof,

Definition 18.2.3: Minimal Polynomial

Let $k(\alpha)$ be a finite simple extension as in theorem 18.2.3, 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 18.2.1

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 $L[x]$.

[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

***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):

$$m_{\alpha,F}(x) = m_{\alpha,L}(x)g(x) + r(x)$$

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 18.2.2: 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 11.6)

Example 18.7: 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 18.2.3, 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 18.2.1, we get to easily represent $F(\alpha)$

Corollary 18.2.3: Representing $F(\alpha)$

Suppose in theorem 18.2.3 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 18.2.3: 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 18.8: 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 18.8 need not be isomorphic (as fields) if the roots solve different polynomials, meaning different irreducible polynomials will produce different extensions:

Example 18.9: $\mathbb{Q}(\sqrt{2}) \ncong \mathbb{Q}(\sqrt{5})$

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 extensiosn is to find extensions in which for any irreducible (or more generally any polynomial) will contain *all* of the roots.

18.2.1 Isomorphism of Fields Extensions

As we have seen in theorem 18.2.1, 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 18.4, $\mathbb{Q}(\sqrt{2})$ 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 18.2.4: 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:

$$f^{\varphi} : k_1[x] \to k_2[x]$$

defiend by applying φ to the coefficeitns 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 18.2.3, 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 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:

$$\frac{1}{1-\alpha} \qquad 1 + \frac{1}{\alpha}$$

[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

[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

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 18.2.4: 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 18.2.4: 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.

***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 18.2.4, 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)
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 18.10: 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 18.2.1

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

$$N = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \;\middle|\; a, b \in \mathbb{R} \right\}$$

[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 18.6

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.

18.3 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 18.3.1: 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 18.11: 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 18.15

[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 18.3.1: 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 18.2.2, 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 18.3.1: 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 18.3.1, 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 18.3.2: 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 18.3.1 is not true: If K is algebraic over F, this does not mean that K/F has finite degree

Example 18.12: Algebraic and infinite degree
We require corrollary 18.3.5 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 18.3.1: 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 18.3.3: 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 18.13: 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 18.5. 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 18.3.1: 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 18.2.1, we can replace F with F_i. Then these can all be characterized in terms of simple extensions, meaning we can use proposition 18.3.1. 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:

$$[K : F] = [F_k : F_{k-1}][F_{k-1} : F_{k-2}] \cdots [F_1 : F_0] \leq [F(\alpha_1) : F][F(\alpha_2) : F] \cdots [F(\alpha_k) : F] \tag{18.1}$$

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 18.3.2: 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 18.12 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 18.3.1, 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 (18.1), we have that K/F is finite.

Using what we've found, it is easier to compute the degree of non-simple extensions:

Example 18.14: 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 dgree 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 11.4.2, 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 18.3.4: 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 18.3.2, and so are algebraic by corrollary 18.3.1 (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 corrollary 18.3.1 every element will be a root of a polynomial over F).

Corollary 18.3.5: 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 (18.3.4), we get that the collection is closed under $+$ and $\times$. From here, it is easy to verify any extra axiom we want.

Example 18.15: 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 18.12.
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 18.3.2 a bit further to algebraic extensions (finite or infinite)

Theorem 18.3.3: 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 18.3.2

$$k(e_1, ..., e_{n-1}) \subseteq k(e_1, ..., e_{n-1}, \alpha)$$

is a finite extension. Furthermore, by theorem 18.3.2 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 18.3.2: 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 18.3.6: Degree of Composite Field and Equality

Suppose that $[K_1 : F] = n$, $[K_2 : F] = m$ in proposition 18.3.2, 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 18.3.2, $[K_1K_2 : F] \leq nm$, we get $[K_1K_2 : F] = nm$

Exercise 18.3.1

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.

18.4 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.

[15] and an additional instrument to measure the angle between the person and the horizon, and the north star

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:

Theorem 18.4.1: 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

18.5 Extensions that are Closures

Through theorem 18.2.1, 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 18.3.5, 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 (18.15) (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 18.2.4, 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.

18.5.1 Algebraic Closure

We start off this seciton by remding the reader what an algebraically closed field is:

Definition 18.5.1: 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 algebaric closure:

Lemma 18.5.1: 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 (18.15)(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 (18.15)(2) how π and e are not algebraic).

Definition 18.5.2: 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 18.5.2: 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 18.2.1 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 9.4.3). 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 18.5.3

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.

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:

[16]This is an example of a *direct limit*

Theorem 18.5.1: 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 (18.15)(2) assuming $\mathbb{C}$ is algebraically closed.

Proof **:**
First, by corrollary 18.3.5, $\overline{k}$ 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 18.3.3), thus α is algebraic over k. But then by definition of $\overline{k}$, $\alpha \in \overline{k}$. Thus, by lemma 18.5.1$\overline{k}$ is algebraically closed, completing the proof

Next, we strive for uniqueness. As we establisehd in theorem 18.2.1, 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 18.5.4: Finding Homomorphisms Over Algebraic Extensions

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 17.3.3). 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 18.2 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 18.2.4: 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 18.5.2: 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 18.5.4, 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 18.5.1 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.

18.5.2 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

Definition 18.5.3: 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"

[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

Theorem 18.5.3: 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 18.5.4 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 18.16: 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 $x^4 + 1$, $x^4 - 1$, and $x^4 + 2$). 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. To undersatnd this phenomenon better, we give a name to fields with this property

Definition 18.5.4: 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 follwoing proofs. It is no harm on making F/k possibly not algebraic, since we work with finite extensions in the majority of cases (maybe it's better to say F/k is algebraic for the infinite case? maybe not? TBD)

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 18.2.4 apply to normal extensions[19]

[19] In fact, normal subgroups got their name from normal extensions! The connection will be made in the chapter 19 when discussing Galois Theory in theorem 19.1.3

Theorem 18.5.4: 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 by corrollary 18.2.4, $\varphi(\alpha)$ 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 18.2.4, 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 18.5.3, F must be finite.

Since $\mathrm{id} : k \to k$ is a field homomorphism, by theorem 18.2.4, 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 18.5.3), 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.

Remark Notice how condition (3) of theorem 18.5.4 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 18.17: 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 18.16, $\mathbb{Q}(\sqrt[4]{2}, i)$ 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 18.5.1: 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 18.18: Low Tolerance for Normal Extension
Consider $L/K/k$ where:

$$\begin{tikzcd}
\mathbb{Q}(\sqrt[3]{2}, \zeta_3) \arrow[d, dash] \\
\mathbb{Q}(\sqrt[3]{2}) \arrow[d, dash] \\
\mathbb{Q}
\end{tikzcd}$$

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:

$$\begin{tikzcd}
\mathbb{Q}(\sqrt[4]{2}) \arrow[d, dash] \\
\mathbb{Q}(\sqrt{2}) \arrow[d, dash] \\
\mathbb{Q}
\end{tikzcd}$$

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 18.5.5: 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 18.5.2: 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 18.5.1

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.

18.6 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 18.19: 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 11.3.1. 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 18.6.1: 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 "seperated" from each other if the polynomial is *inseparable*. 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 perserves roots (theorem 18.2.4). 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 $\overline{k}$.

Definition 18.6.2: 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.

RemarkSeparableEx Note that it is note 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 18.6.3: 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

$$D_x\left(\sum_i^n a_i x^i\right) = \sum_i^n i a_i x^{i-1} \tag{18.2}$$

[20](the correspondence Galois group is trivial otherwise)

Notice that this derivative acts exactly like the derivative seen in a regular calculus course [21]. 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 (18.2) 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.

Proposition 18.6.1: 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 11.3.1 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:

[21] A good resources to learn this would be Spivak's Analysis

Proposition 18.6.2: 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 18.6.1, $f(x)$ 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 18.6.1: 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 18.6.2

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 18.6.4: 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 18.6.5: 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 18.6.3: 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 18.6.1, any irreducible polynomial is already seperable, so let's say $\mathrm{ch}(k) = p$ for some prime p.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 irreducability of $f(x)$, showing us that no such polynomial can exist. 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 18.19, 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 18.6.2: 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 18.6.3 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.

18.6.1 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 18.5.2 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 18.6.6: 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 18.2.4.

($\Leftrightarrow$) moment to ponder

Proposition 18.6.4: 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 18.2.4. Consider the collection of k-homomorpihism $\{\iota_r : k(\alpha) \to \overline{k}\}$. By corrollary 18.2.4, 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 theorem 18.5.4, $k(\beta) \subseteq \overline{k}$ exists, and by theorem 18.2.4, 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 18.6.3: 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 18.5.4, 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 18.6.5: 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 18.6.4 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 18.6.4, corollary 18.6.3, and theorem 18.3.1:

$$\begin{aligned}
[F : k]_s &= [k(\alpha_1, ..., \alpha_{n-1})(\alpha_n) : k(\alpha_1, ..., \alpha_{n-1})]_s \cdots [k(\alpha_1) : k]_s \\
&\leq [k(\alpha_1, ..., \alpha_{n-1})(\alpha_n) : k(\alpha_1, ..., \alpha_{n-1})] \cdots [k(\alpha_1) : k] \\
&= [F : k]
\end{aligned}$$

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 18.6.3

($3 \Rightarrow 2$) 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 18.6.3,

$$[F : k(\alpha)]_s[k(\alpha) : k]_s = [F : k]_s = [F : k] = [F : k(\alpha)][k(\alpha) : k]$$

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 18.6.4 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 **??**: 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 18.6.6: 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 18.6.5. 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 18.6.5, $k(\alpha_1, ..., \alpha_n)/k$ 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.

18.6.2 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 18.6.7: 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 18.7.4. 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 18.6.1: 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 18.7.4.

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:

$$f(x) = \prod_{\iota \neq \iota'} \big((\iota(\alpha)x + \iota(\beta)) - (\iota'(\alpha)x + \iota'(\beta))\big) \in \overline{k}[x]$$

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 18.2.4 from simple to finite extensions:

Corollary 18.6.4: Order of $\mathrm{Aut}_k(F)$ Of Finite Extension

Let F/k be a finite seprable extension. Then:

$$|\mathrm{Aut}_k(F)| \leq [F : k]$$

with equlaity if and only if F/k is a finite separable normal extension.

***Proof* :**
By theorem 18.6.1, if K is finite and separable, it is simple, and so the inequality immediately holds by corrollary 18.2.4. Equality holds if the minimal polynomial splits, which it does over it's splitting field, which is a normal extension by theorem 18.5.4. 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 18.20: 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 18.6.7

Exercise 18.6.1

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)

18.7 Finite Fields and Cyclotomic Polynomials

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 18.6.1 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[22]

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.

18.7.1 Cyclotomic Polynomials and Fields

We start off with some reminders:

Definition 18.7.1: 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 $\zeta_4 = i$. 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. $\zeta_5 = \frac{\sqrt{5}-1}{4} + \frac{\sqrt{10+2\sqrt{5}}}{4}$

[22] 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.

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 18.4. 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 requier 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 18.7.2: 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 (TBD), there are $\varphi(n)$ primitive roots of 1, where φ is Euler's totient function (in Aluffi, its exercsie II.6.14). 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 18.21: Cyclotomic Polynomials

1. Let $n = p$ be a prime. Then every non-identity eleent 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 $\binom{n}{k}$).

For ease of reference, we record what happens when n is not prime:

Lemma 18.7.1: Cyclotomic Polynomial Computation

For all $n > 0$,

$$x^n - 1 = \prod_{1 \leq d \mid n} \Phi_d(x)$$

Proof **:**
(Justification here by Aluffi p. 446)

$$x^n - 1 = \prod_{\zeta \in \mu_n} (x - \zeta) = \prod_{1 \leq d \mid n} \left(\prod_{\substack{\zeta \text{ primitive} \\ d\text{th root of } 1}} (x - \zeta) \right) = \prod_{1 \leq d \mid n} \Phi_d(x)$$

Corollary 18.7.1: Integer Coefficients

The cyclotomic polynomials $\Phi_n(x)$ have integer coefficients

Proof **:**
p. 447

Example 18.22: 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 18.7.1: 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 18.7.3: Cyclotomic Field

The splitting field $\mathbb{Q}(\zeta_n)$ for the polynomial $x^n - 1$ over $\mathbb{Q}$ is the nth *cyclotomic field*

Proposition 18.7.2: Cyclotomic Field Automorphism

$$\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta_n)) \cong (\mathbb{Z}/n\mathbb{Z})^\times$$

Proof **:**
p. 448 Aluffi. 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. The homomorphism sends $j([m]_n)$ to the [unique] automorhpism $\zeta_n \mapsto \zeta_n^m$. Make sure it's representative independent and injective (i.e.

well-defined).

18.7.2 Finite Fields

Let F be a finite field of characteristic p. Then as we remarked under definition 18.1.6,

$$F/\mathbb{F}_p$$

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 18.7.1: 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 elemetns. Then F is a splitting field for $x^q - x$ over $\mathbb{F}_p$

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 18.7.2: 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 18.7.1 and the uniquness of the splitting field

Definition 18.7.4: 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 18.23: Galois Fields
Good example by aluffi involving the reducibility of $x^4 + 1$ over *any* field $\mathbb{F}_p$

Corollary 18.7.3: 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 18.7.1, 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 18.1.1 ($F^\times$ 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 18.7.4: 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 18.6.7. If k is a finite field, then so is F, and the extensions are automatically simple by corollary 18.7.3.

Using these results, we can be more gain a lot of information about polynomials over finite fields:

Corollary 18.7.5: 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 fintie 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 18.7.6: Factorizing $x^{q^n} - x$

here p. 443 aluffi

Proof **:**

Example 18.24: Factorizing Polynomials over Finite Fields
p. 443 Aluffi

Another result that finiteness brings is that the structure of $\operatorname{Aut}_{\mathbb{F}_p}(\mathbb{F}_{p^d})$ is very simple. We already know that any finite field is perfect, hence all algebraic extensions are separable, and so

$$|\operatorname{Aut}_{\mathbb{F}_p}(\mathbb{F}_{p^d})| = d$$

However, we can say even more about the structure of the group

Proposition 18.7.3: 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 **??**)

$$x^p \mod x \mod p$$

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 fintie fields seems to be close to another known category)

Exercise 18.7.1

1. Let $f(x)$ be an irreducible polynomial of degree n over $\mathbb{F}_p$. Sho that any field extension of $f(x)$ splits and is isomorphic to $x^{p^n} - x$

19

Galois Theory

(I think it's important to emphasize here why groups are called "symetries" and how what we talked about when saying that Aut capture this idea of symmetry. In this case, mathematicians where looking for relations between elements of a polynomial for a long time. Some of these were found with sucess, for example if $x^2 + bx + c$ has roots α and β, then $\alpha + \beta = -b$ and $\alpha\beta = q$[1]. However, the effort to understand these brought limited results to pressing questions like the solvability of the quintic. One other approach that was considerd was to see what permutations of roots of polynomials produces the same results. For example taking the polynomial:

$$x^4 - 5x^2 + 6 = (x^2 - 2)(x^2 - 3)$$

it reduces to 2 irreducible polynomials over $\mathbb{Q}$. Adjoining $\sqrt{2}$ and $\sqrt{3}$ to $\mathbb{Q}$, we can factor further and get:

$$= (x - \sqrt{2})(x + \sqrt{2})(x - \sqrt{3})(x + \sqrt{3})$$

In this case, the roots are: $\sqrt{2}$, $-\sqrt{2}$, $\sqrt{3}$, and $-\sqrt{3}$. Now, in the original formula, there are only 2 spots in which we can plug in roots. The possibilities are:

$$((\sqrt{2})^2-2)((\sqrt{3})^2-3) \qquad ((-\sqrt{2})^2-2)((\sqrt{3})^2-3) \qquad ((\sqrt{2})^2-2)((-\sqrt{3})^2-3) \qquad ((-\sqrt{2})^2-2)((-\sqrt{3})^2-3)$$

The realization that at first seems a little trivial is that we can "permute" between these possibilities by permutting the roots: if we map

$$\begin{cases} \sqrt{2} \mapsto -\sqrt{2} \\ -\sqrt{2} \mapsto \sqrt{2} \\ \sqrt{3} \mapsto -\sqrt{3} \\ -\sqrt{3} \mapsto \sqrt{3} \end{cases}$$

[1] See section 19.2.3 for much details about these relations

then we get one of those 4 possibilities. However, if we premute:

$$\begin{cases} \sqrt{2} \mapsto \sqrt{3} \\ -\sqrt{2} \mapsto -\sqrt{3} \\ \sqrt{3} \mapsto \sqrt{2} \\ -\sqrt{3} \mapsto -\sqrt{2} \end{cases}$$

Then the equation no longer equals 0 ("has a solution"), since:

$$((\pm\sqrt{3}^2 - 2)((\pm\sqrt{2})^2 - 3) = -1$$

Thus, the 4 permutations of the roots that let the polynomial remain with solutions permute $\sqrt{2} \mapsto \pm\sqrt{2}$ and $\sqrt{3} \mapsto \pm\sqrt{3}$.

Another example is the following: take $x^4 - 2$. This is irrreducible by Eisenstens. (should I put this here or later? basically, I wanna say that we can reduce it by adding a root $\zeta_4^k\sqrt{2}$, but also the product of two of these elements produce $\pm\sqrt{2}$ giving us an "intermediate step" of irreducibles with their own minimal polynomial)

) Word on Galois investigating permutation of roots which leads to current result? Word on where Galois theory is today?

19.1 Galois Extension and Galois Group

First, let's review the recall the definition of an automorphism and define a new term:

Definition 19.1.1: 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 19.1: 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

$$\mathrm{Aut}_P(F) = \mathrm{Aut}(F)$$

Lemma 19.1.1: Automorphism Group

Let F be any field. Then $\mathrm{Aut}(F)$ is a group under composition and $\mathrm{Aut}_k(F)$ is a subgroup

***Proof* :**
This all comes down to the properties of isomorphissm (ex. invertability, associativity of functions etc.) and the fact that composing two automorphisms σ and τ, $\sigma\tau$ is possible since the co-domain and domain match, and it produce another automorphism.

Similarly, if two automorphissm σ, τ fix F, then the composition $\sigma\tau^{-1}$ will also fix F, meaning $\mathrm{Aut}(K/F)$ is a subgroup by the subgroup criterion (proposition 3.1.1).

Next, we recall a very useful criterion for finding automorphisms of algebraic extensions K/F

Proposition 19.1.1: 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

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

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

$$\sigma(\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0) = \sigma(0)$$
$$\sigma(\alpha)^n + a_{n-1}\sigma(\alpha)^{n-1} + \cdots + a_1\sigma(\alpha) + \sigma(a_0) = 0 \qquad \sigma \text{ fixes } a_i \text{ for each } i$$

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 19.2: 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 18.2.2, 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 18.14). 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 (18.16), 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[2]

Lemma 19.1.2: 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

[2]Note that $\mathrm{Aut}_{\mathbf{Grp}}(-)$ is *not* a functor; there are counter-examples, see exercise ref:HERE

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 19.1.2: 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.

Definition 19.1.2: 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 19.3: Fixed Field

1. The fixed field of $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\sqrt{2}))$ is $\mathbb{Q}$. As we have seen in example 19.2.
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 19.1.3: 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 19.1.2.

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 19.4: 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 19.1.3: 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 18.6.1, 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 18.5.4, 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:

$$q_\alpha(x) = x^n - s_1 x^{n-1} + s_2 x^{n-1} - \cdots + (-1)^{n-1} s_{n-1} x + (-1)^n s_n$$

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 19.1.4: $H = \mathrm{Aut}_{F^H}(F)$

Let F/k be a *finite* extension, and let H be a subgroup of $\mathrm{Aut}_k(F)$. Then $|H| = [F : F^H]$ and

$$H = \mathrm{Aut}_{F^H}(F)$$

In particular, the Galois correspondence is surejctive for a finite field (for any subgroup $H \leqslant \mathrm{Aut}_k(F)$, there exists an intermediate field (namely F^H) such that $H = \mathrm{Aut}_{F^H}(F)$)

Intuitively, we want to show that $\mathrm{Aut}_{F^H}(F)$ *only* fixes F^H, and so $H = \mathrm{Aut}_{F^H}(F)$.

***Proof* :**
Let $H \leqslant \mathrm{Aut}_k(F)$. By proposition 19.1.3, we have that $H \leqslant \mathrm{Aut}_{F^H}(F)$, so $|H| \leq |\mathrm{Aut}_{F^H}(F)|$. Since $\mathrm{Aut}_{F^H}(F)$ is finite, verifying that $\geq$ holds will complete the proof.

To see this, recall that we've shown in lemma 19.1.3 that $F = F^H(\alpha)$ for some $\alpha \in F$. Thus, by corollary 18.2.4, $|\mathrm{Aut}_{F^H}(F)|$ 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 19.1.3. By construction, the number of roots of $q_\alpha(x)$ must be less than or equal $|H|$, giving us:

$$|\mathrm{Aut}_{F^H}(F)| \leq |H|$$

And since $|\mathrm{Aut}_{F^H}(F)| \geq |H|$, we get $|\mathrm{Aut}_{F^H}(F)| = |H|$ and since $H \subseteq \mathrm{Aut}_{F^H}(F)$, then $H = \mathrm{Aut}_{F^H}(F)$, giving the desired equality.

Finally, to show that $[F : F^H] = H$, since F/F^H is finite, separable, and normal, by corollary 18.6.4, $[F : F^H] = |\operatorname{Aut}_{F^H}(F)|$, 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?)

19.1.1 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 19.1.1: 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 18.5.4
- $2 \Leftrightarrow 3$ in corrollary 18.6.4
- $3 \Leftrightarrow 4$ follows from proposition 19.1.4, 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 18.5.4 along with the fact that finite separable extensions are simple (theorem 18.6.1)

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 19.1.4, $\operatorname{Aut}_{F^{\operatorname{Aut}_k(F)}}(F) = \operatorname{Aut}_k(F)$, 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 19.1.4, 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 $E = F^{\mathrm{Aut}_E}(F)$, 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 19.1.3: 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 19.1.1.

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 19.1.1: 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 18.6.1, any algebraic extension of F is separable, and hence by theorem 19.1.1, it is Galois.

Definition 19.1.4: 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 19.1.5: 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 19.1.6: 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:

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 19.1.1(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 19.1.3, 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 Galoi 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 19.5: 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 19.2), 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 (18.14) 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 19.1.3 to find the associated fixed field given the subgroups of $\operatorname{Gal}(K/\mathbb{Q})$. 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 5.4.5) 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 19.6.

6. As we've seen with cyclotomic fields, by proposition 18.7.2,

$$\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta_n)) \cong (\mathbb{Z}/n\mathbb{Z})^{\times}$$

giving us a Galois group for every $n \in \mathbb{N}_{>0}$.

7. By theorem 18.7.1, 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 18.7.3, 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 19.1.1 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 19.1.2: 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 19.1.1, 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 19.1.3), 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 19.6: Lattice Correspondence

1. As we've shown in example (19.5)(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 (19.5)(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 18.5.1 and example 18.18). 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 19.1.3: 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 19.1.6) 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 18.5.4 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 18.6.1, 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: $\operatorname{Aut}_k(F)$ acts on I: for $g \in \operatorname{Aut}_k(F)$ 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 19.1.1). 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 $\mathrm{Aut}_k(F)$

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 19.1.3

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 19.1.1, 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 19.1.1). Define the homomorphsim 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 homomrophsim is surjective (since the G-action acts transitively on I), and $\ker(\rho)$ is all automorphissm that restrict to the identiy on E, that is, $\mathrm{Aut}_E(F)$. But then, we know that the kernel of a group homomorphsim is normal, and by the 1st isomorphism theorme 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 extensiosn, 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 19.1.2: 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 19.7: Application To Lattices
(Draw two latties, and show the normal extensions)

Foreshadowing Can give the connection between Galois group/extensions to Algebraic geometry with regular covers and deck transformations

19.1.2 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 chapter, 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. (why does F/k need to be finite in the following theorem?)

Proposition 19.1.5: Galois Composite Fields

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:

$$\begin{tikzcd}
 & KF \arrow[dl, dash] \arrow[dr, dash, "G"] & \\
F \arrow[dr, dash, "G"] & & K \arrow[dl, dash] \\
 & F \cap K \arrow[d] & \\
 & k &
\end{tikzcd}$$

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 19.1.4, $H = \mathrm{Aut}_{E^H}(E)$. 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 (19.1.1)(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 19.8: Composite fields determining Abelian Extensions

1. As we have seen in section 18.7.1, 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 19.1.5, 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 then show that all finite abelian group show up as a field extension of $\mathbb{Q}$ in section ref:HERE.

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 $n \in \mathbb{N}$

Corollary 19.1.3: 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 (for example, $k = \mathbb{Q}$, $E = \mathbb{Q}(\sqrt[3]{2})$, and $F = \mathbb{Q}(\zeta_3 \sqrt[3]{2})$), since $EF = \mathbb{Q}(\rho\sqrt[3]{2})$ (with minimal polynomial $x^3 - 2$, and hence degree 3 extension) and:

$$3 = [\mathbb{Q}(\rho\sqrt[3]{2})] \neq \frac{[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}][\mathbb{Q}(\zeta_3\sqrt[3]{2}) : \mathbb{Q}]}{[\mathbb{Q} : \mathbb{Q}]} = 9$$

19.1.3 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 19.1.7: 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 19.9: 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 19.1.8: 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 19.1.4: 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 $g \in G$:

$$a_1\chi_1(g) + a_2\chi_2(g) + \cdots + a_n\chi(g) = 0 \tag{19.1}$$

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 (19.1) holds for every element of G, we have that:

$$a_1\chi_1(g_0g) + a_2\chi_2(g_0g) + \cdots + a_n\chi_n(g_0g) = 0 \tag{19.2}$$

i.e.

$$a_1\chi_1(g_0)\chi_1(g) + a_2\chi_2(g_0)\chi_2(g) + \cdots + a_n\chi_n(g_0)\chi_n(g) = 0 \tag{19.3}$$

Now, multiplying equation (19.1) by $\chi_n(g_0)$ and subtracting by equation (19.3), 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 19.1.4: 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 19.1.4

Using some character theory, we can

Proposition 19.1.6: 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 $\varphi(\delta)$, whic 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 19.1.4, 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 19.1.1

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 $p \equiv 1 \mod n$)

19.2 Application of Galois Theory

19.2.1 Fundamental Theorem of Algebra

(Some build-up with properties of $\mathbb{C}$ that are given a exercsie 5.23)

Theorem 19.2.1: 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 19.2.1: 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 18.5.4, 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[3].

Theorem 19.2.2: 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 16.4))

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 VI, section ref:HERE) of many classical groups

19.2.2 Construction of Regular n-gon

19.2.3 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:

$$P_n(x) = x^n - s_1(t_1, ..., t_n)x^{n-1} + \cdots + (-1)^n s_n(t_1, ..., t_n)$$

[3]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

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 19.2.3: 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 19.2.1: 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 Galoi group S_n

Corollary 19.2.2

Let G be a finite group. Then there exists a Galois extension F/k such that $\operatorname{Aut}_k(F) \cong G$

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 19.2.3

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.

19.2.4 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.

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 19.2.1: 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 extensiosn that is *Galois*:

Lemma 19.2.2: 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 18.6.1), 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 19.2.3: 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 19.2.2, 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 19.2.2: 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 19.2.4: 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 19.1.6.

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 18.5.1), and each extension:

$$k(\delta_1, ..., \delta_{i-1}) \subseteq k(\delta_1, ..., \delta_i)$$

satisfies proposition 19.1.6, and hence each G_i is cyclic. By theorem 19.1.3, 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 19.1.6 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 19.2.1: 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 19.1.5, 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 19.2.4, 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 19.2.4, $L(\zeta)/k(\zeta)$ 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 19.2.4: 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 19.2.3 and proposition 19.2.1.

Theorem 19.2.4: Abel's Theorem

Let $f(x)$ be an polynomial of degree 5 or higher. Then it admits no solutions by radicals

Proof **:**
By lemma 19.2.1, 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.

19.3 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 19.3.1: 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 18.2.4, 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 19.3.1: 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 19.3.2: 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 19.3.2: 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 19.10: 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?)

19.4 *Infinite Galois Extensions

here, talk about the krull topology and the limit construction.

Part V

Commutative Algebra

This is the part of the book were we are not studying a new object as the central focus (commutative rings and modules are at the heart of commutative algebra). Rather, it is the basic language that deals with *algebraic number theory* and *algebraic geometry*.

In this chapter, some basic topological notions ill be used. The geometric intuition presented in this chapter ill be lost on those who have not taken or learnt some basic topology (up to the separation axioms, in particular T_1 and T_2, and compactedness/connectedness). Appendix C serves as a review of what topological concepts will be used in this chapter.

(Perhaps mention how though it is not a necessity to know topology, a lot of geometrical concepts are described in the language of topology, and so it is useful. However, to some degree, the geometric nature is simply evident since basic geometric shape's like lines, planes, circles, and so forth will appear, and no topology is needed to see the sets representing those shapes are geometric. However, a lot of the study of [non-classical] geometry has it's basis in topology)

(A word on how algebraic topology is a very different field of study that is not algebraic geometry)

(A word on how commutative algebra is often also associted with the phrase "commutative geometry". There is a new field of math developping called "non-commutative geometry" (maybe also non-commutative algebra?) which I don't yet know about because I did not at MAT436/437)

20

More on Ring Theory and Extensions

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.

(Any field extension is a k-algebra! If in the Kroneker construction (theorem 18.2.1) we use a *reducible* polynomials, then the resulting object is no longer a field (it contains zero-divisors), but remains to be a k-aglebra! Hence k-algebras are the natural object to study more quotient's of polynomial rings.)

20.1 Ring 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 20.1.1: 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 injecive, 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 20.1.2: 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 originated from the following requirement in number theory: Let K be a finite extension of $\mathbb{Q}$. The field K is called an *algebraic number field*. Each algebraic number field has an associated *ring of integers* $\mathcal{O}_K = \overline{\mathbb{Z}}^K$. These are fundamental in algebraic number theory. We want to have a correspondence between algebraic extensions of $\mathbb{Q}$ and integral extensions of $\mathbb{Z}$. To that end, we would want $\mathcal{O}_{\mathbb{Q}} = \mathbb{Z}$. Since the polynomials are monic, it must be that the only polynomials that are satisfied with coefficents from $\mathbb{Q}$ are:
$$x - t = 0 \qquad t \in \mathbb{Z}$$
If we allowed for non-monic coefficients, then $\mathcal{O}_{\mathbb{Q}} = \mathbb{Q}$, breaking the correspondence. For a more general 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 coefficeint 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.

Integral elements share many of the same properties as algebraic elements:

Proposition 20.1.1: 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 20.1.1: 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 20.1.1 $R[s]$ 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 20.1.2: 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 infintie 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 18.3.2). 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 20.1.2: 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 20.1: 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 $F = \mathrm{Frac}(R)$. 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:

$$a^n = b(-r_{n-1}a^{n-1} - \cdots - r_1 ab^{n-2} - r_0 b^{n-1})$$

so $b \big| a^n$, and hence $b \big| a$. 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 $a/b \in \mathrm{Frac}(R)$ 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 11.3.1), 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:
$$\mathbb{Z}\left[\frac{1+\sqrt{3}}{2}\right]$$

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 20.1.3: 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 $J \cap R$. It should be checked that

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 $\mathbb{Z} \hookrightarrow \mathbb{Q}$)
6. If J is maximal, J^c need not be maximal, and similarly if I is maximal then I^e need not be maximal.

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: Show the following 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. On the otherhand, the correpsondence of $\varphi(R)$ and S is much harder.

Example 20.2: 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 **??**), it can be deduced that:

1. $(2)^e = (1+i)^2$
2. If $p \equiv 1 \mod 4$, then p^e is the product of two distinct prime ideals (ex. $(5)^e = (2+i)(2-i)$)
3. if $p \equiv 3 \mod 4$, 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:

Theorem 20.1.1: Integral Extension Prime Correspondence (Cohen–Seidenberg Theorem)

Let $R \subseteq S$ be a subring and S be an integral extension of R

1. (**Lying-over Property**) Let $p \subseteq R$ be a prime ideal. Then there is a prime ideal $q \subseteq S$ such that $P = Q \cap R$. Furthermore, p is maximal if and only if q 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**) Let S be an integarl domains, $p_1 \subseteq p_2 \subseteq \cdots \subseteq p_n$ 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.
4. (**Incomparability property**) Let q_1 and q_2 lie over p. Then $q_1 \not\subseteq q_2$ and $q_2 \not\subseteq q_1$

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 clsoed 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.

Theorem 20.1.2: Universal Property of R-algebras

Let S be an R-algebra generated by n elements (not necessarily finite). Let $\{x_i\}_{i \in I}$ be an indexing set of indeterminate variables where $|I| = n$. Then there exists a surjective homomorphism making the following diagram commute:

$$\begin{tikzcd} R[(x_i)_{i \in I}] \arrow[r] & S \\ A \arrow[u] \arrow[ur] & \end{tikzcd}$$

Proof **:**
This is theorem 13.1.1.

It is useful to show that if $\varphi : A \to B$ is an R-algebra homomorphism, then φ is constant on R in the

polynomial representation of the R-algebras.

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.

20.2 Localization

(re-write this motivation: I moved localization out of algebraic geometry for now since it feels it will allow me to talk about Completion and it will be useful for number theory too)

20.2.1 Localization of Rings

The detail of the motivation will come here later. For now, 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 20.2.1: 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 20.2.1: 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 20.2.1: 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: 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:
$$\mathbb{Z}_{(p)} = \left\{ \frac{a}{b} \in \mathbb{Q} \;\middle|\; p \nmid b \right\}$$
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?
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 9.4.4, 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 20.2.1: 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 $J \subseteq D^{-1}R$,

$$J = {}^{\iota}({}^{\pi}J)$$

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:

$${}^{\pi}(\lim I) = \{r \in R \mid dr \in I \text{ for some } d \in D\}$$

As a consequence, ${}^{\iota}I = D^{-1}R$ 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 $D^{-1}R$:

$$\left\{\begin{matrix}\text{prime ideals of } R \text{ with} E \\ P \cap D = \emptyset\end{matrix}\right\} \underset{\pi}{\overset{\iota}{\rightleftarrows}} \left\{\text{prime ideals of } D^{-1}R\right\}$$

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 20.2.1 is given a name:

Definition 20.2.2: 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

Exercise 20.2.1

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. Show that if R is a localization at P for prime P, then R_P is a local ring.

20.2.2 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 20.2.1, 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 20.2.2: 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.

***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. This is the case:

Proposition 20.2.3: 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 $\mathrm{im}(D^{-1}f) \subseteq \ker(D^{-1}g)$. 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 $sm \in \ker(g) = \mathrm{im}(f)$, implying for some $m' \in M_3$, $f(m') = tm$. Putting it all together, we get that in $M_2[D^{-1}]$:

$$\frac{m}{d} = \frac{f(m')}{ds} = (D^{-1}f)\left(\frac{m'}{ds}\right) \in \mathrm{im}(D^{-1}f)$$

showing that $\ker(D^{-1}g) \subseteq \mathrm{im}(D^{-1}f)$, which along with the fact that $\mathrm{im}(D^{-1}f) \subseteq \ker(D^{-1}g)$ show's that

$$\ker(D^{-1}g) = \mathrm{im}(D^{-1}f)$$

completing the proof.

It follows that if $\varphi : M_1 \to M_2$ is injective, so is $D^{-1}\varphi : D^{-1}M_1 \to D^{-1}M_2$ (what about surjectivity?). If we use section 17.4, then we can use the following observation to arrive at this result more quickly:

Proposition 20.2.4: 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:

$$(r/d) \times m \mapsto \frac{am}{d}$$

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 20.2.2: 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 20.2.4 and the appropriate isomorphism theorem for modules.

(Some more cool results from Micheal's p. 43)

Local Properties

Localization allows us to diminsh 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 (Why would we do that? TBD). This

motivates the following concept:

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 localized ring (resp. module). Here are some examples of local properties:

Proposition 20.2.5: 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 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 20.2.6: Injectivity/Surjectivity Is A Local Property

Let $\varphi : M \to N$ be an R-module homomorphism. Then the following are equivalent:

1. φ is injective
2. $\varphi_p : M_p \to N_p$ is injective for all prime ideals $p \subseteq R$
3. $\varphi_m : M_m \to N_m$ is injective for all maximal ideals $m \subseteq R$

Similarly for surjective.

Proof **:**

(1) $\Rightarrow$ (2) Since $0 \to M \xrightarrow{\varphi} N$ is exact, by proposition 20.2.3, $0 \to M_p \xrightarrow{\varphi_p} N_p$ 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 20.2.3 $0 \to K_m \to M_m \to N_m$ is exact, therefore $K'_m \cong \ker(\varphi_m) = 0$ since φ_m is injective. Thus, by proposition 20.2.5 φ is injective.

to show surjectivity, simply reverse the arrows.

A final example that will be useful is that faltness is a local property:

Proposition 20.2.7: 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:

$$\begin{aligned} N \to P \text{ is injective } &\Rightarrow N_m \to P_m \text{ is injective} && \text{proposition 20.2.6} \\ &\Rightarrow N_m \otimes_{R_m} M_m \to P_m \otimes_{R_m} M_m \text{ is injective} && \text{flatness property} \\ &\Rightarrow (N \otimes_R M)_m \to (P \otimes_R M)_m \text{ is injective} && \text{proposition ??} \\ &\Rightarrow N \otimes_R M \to P \otimes_R M \text{ is injective} && \text{proposition 20.2.6} \end{aligned}$$

Hence, M is flat.

Exercise 20.2.2

1. Show that M is projective if and only if each M_p is free.

20.2.3 Local Rings

Recall that the localization of $\mathbb{Z}$ at any prime ideal (p), $\mathbb{Z}_{(p)}$, left $\mathbb{Z}_{(p)}$ with a unique maximal ideal. It was left as an excercise 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 associatd 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 20.2.3: 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 20.2.8: 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 20.4: 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 20.2.9: 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 20.2.4: Residue Field

Let R be a commutative local ring with maximal ideal m. Then R/m is called the *residue field*.

20.2.4 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 20.2.5: 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:

$$\mathrm{Jac}(R) = \bigcap_{\substack{M \subseteq R \\ \text{maximal}}} M$$

(analogous to Fratini subgroup)

Proposition 20.2.10: 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 $1 - rx = y \in M'$, shwoing 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 20.2.1: 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 20.2.2: 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 20.2.1, $M/F = \{0\}$, 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 20.2.3: 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 20.2.2, $M = F$, 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 15, 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 20.2.2: 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 20.2.11: 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 $f(M \mod \mathfrak{m}M) = N \mod \mathfrak{m}N$. Since N is finitely generated, so is any quotient of it, and so $N \mod \mathfrak{m}N$ 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, $N = \mathrm{im}(f) + \mathfrak{m}N$. Thus, by Nakayama's lemma, $N = \mathrm{im}(f)$.
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 20.2.2, $x_1, ..., x_n$ forms a basis for M. Suppose now that $f(a_1 x_1 + \cdots + a_n x_n) = 0$. 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. $\langle f_{(\mathfrak{m})}(\overline{x}_1), ..., f_{(\mathfrak{m})}(\overline{x}_n)\rangle = \mathrm{im}(f_{(\mathfrak{m})})$ 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:

$$a_1 f(x_1) + \cdots + a_n f(x_n) = f(a_1x_1 + \cdots + a_nx_n) = 0$$

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 20.2.12

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 9.1.2). Naturally, this is not necessarily the case for infinite rings ($\mathbb{Z}$ being the prototypical example). We may thus take $Q = \mathrm{Frac}(R)$, and consdier the R-module Q

Corollary 20.2.3: Infinite Generation Of Ring Of Fractions

Let $Q = \mathrm{Frac}(R)$ 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 infintie 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 20.2.4: 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!

20.3 Completions

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, we will see that the exact same way the rational numbers complete the reals, the completion of a ring is the equivalence class of all Cauchy-sequences which converge to the same point. We will use completions to construct rings like the ring of *formal power series* $R[[x_1, x_2, ..., x_n]]$, which is the ring with infinite polynomials, the ring of *infinite sequences*, the *p-adic integers* $\mathbb{Z}_p$, and the *profinite integers* $\hat{\mathbb{Z}}$. Unlike completion of $\mathbb{Q}$ to $\mathbb{R}$, there are multiple ways of completing a ring depending on which prime ideal we choose. For example, $\mathbb{Z}$ may be completed with respect to any prime ideal $p \subseteq \mathbb{Z}$.

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 have 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 $y^2 - x^2(x+1) = 0$

(Useful in both algebraic geometry and number theory, so should be introduced now)

(In this section, topological notion's will be introduced, and so an aquientance of the material in appendix C is recommended. It should be noted that the topologies that will be discussed will be very different than that of euclidean topologies: HERE)

20.3.1 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:

Definition 20.3.1: Completion Of A Ring

Let R be a ring and $I \subseteq R$ be an ideal, and:

$$\cdots \xrightarrow{f_3} R/(I^3) \xrightarrow{f_2} R/(I^2) \xrightarrow{f_1} R/I$$

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 dentoed as:

$$\varprojlim R/I^n := \hat{R}$$

and is called the *inverse limit*. The collection $\{R/I^n\}$ is calledhte *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 $\{R/a^nR\}$ is an inverse sytem with ideal $(a) \subseteq R$, and any $x \in R$, define a neighbourhood basis (or local basis) $\{x + a^nR\}_{n=1}^{\infty}$. Then this forms a basis for a topology. This topology is called the *a-addic topology*, where a represents the ideal (and so I-addic topology in general). U is open in this topology if and only if for each $x \in U$, there exists a positive integer n such that
$$x + a^nR \subseteq U$$
Then R is Hausdorff if and only if $\cap a^nR = 0$. If this is the case, define the metric $d(x, y) = 2^{-n}$ whenever $x \neq y$ where n is an integer such that $x - y \in a^nR - a^{n-1}R$.

   This exact same process can be repeated on an arbitrary R-module, where the inverse system would be $\{M/a^nM\}$, and for a given ideal $I \subseteq R$, we would define the I-addic topology on M.

2. this is a special case of a more general construction of an *inverse limit* in the category of **Ring**

3. By how we've defined it, any inverse system will be surjective. Such systems are somtimes *surjective systems*.

Example 20.5: Completion Of Rings

1. Let $R = k[x]$ and $I = (x)$. Then:
$$\varprojlim(k[x]/(x)^n) = k[[x]]$$
(explanation here)

2. Let $R = \mathbb{Z}$ and $I = (10)$. Then $\varprojlim \mathbb{Z}/(10)^n$ is represented by all infinitely long numbers (show how).

Completions are a way to construct rings. It would be nice if exactness interactd nicesely 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 20.3.1: Inverse Limits And Exactness

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

***Proof* :**
p. 104 Michael Atiyah

(more corrollaries here that I'll consider thinking about later)

Filtration

It is worth noting that there is antoher way of giving the topology talked about above using *filtrations* (Michael Atiyah p. 105).

20.4 Discrete Valuation Domains

(Can we define a norm on a vector-space over a finit field? no, homogeneity fails. But we still have *some* notion of size. Defining a norm was super-useful for studying infinite vector-spaces. We will osensibly be doing the same here!!)

(cool intuition here)

Definition 20.4.1: 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 $x = 0$

Definition 20.4.2: Discrete Valuation Induces Ring

Let ν be a discrete valuation. Then

$$\mathcal{O}_K = \{x \in K \mid \nu(x) \geq 0\}$$

forms a ring, called the *discrete Valuation Ring* (DVR)

(NOTE every DVR has a natural topology defined on it! $|x - y| = 2^{\nu(x-y)}$)

Proposition 20.4.1: DVR Equivalences

Let R be a discrete Valuation Ring. Then the following are equivalent:

1. There is some discrete valuation ν on the field of fractions K of R such that $R = \{0\} \cup \{x \in K : (x)0\}$
2. R is a local principal ideal domain, and not a field.
3. R is a valuation ring with a value group isomorphic to the integers under addition.
4. R is a local Dedekind domain and not a field.
5. R is a Noetherian local domain whose maximal ideal is principal, and not a field.
6. R is an integrally closed Noetherian local ring with Krull dimension one.
7. R is a principal ideal domain with a unique non-zero prime ideal.
8. R is a principal ideal domain with a unique irreducible element (up to multiplication by units).
9. R is a unique factorization domain with a unique irreducible element (up to multiplication by units).
10. R is Noetherian, not a field, and every nonzero fractional ideal of R is irreducible in the sense that it cannot be written as a finite intersection of fractional ideals properly containing it.

Proof **:**
Do these as an exercise!

20.5 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 interseting results, namely the *descending chain condition*. In this section, we explore thep propertie 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 produt of matrix ring over divison rings)

Definition 20.5.1: 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 20.5.1: 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 refersed

Just like noetherian modules, if $N \subseteq M$ and M/N are artinian, so is M.

Lemma 20.5.1: Artinian Module Closed Under Quotient

Let M be an artinian module and $N \subseteq M$. Then M/N is also Artinian

Proof **:**
exercise

However, unlike Noetherian rings, artinian rings need to satisfy stronger properties on their ideals:

Proposition 20.5.2: 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 differ's 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 20.5.1. 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 20.5.2: 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 20.6: 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 20.5.1: Structure Theorem On Artinian Rings

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

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 20.5.1: 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 21.2.4, 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 20.5.1, R/m^k is artinian, and so it follows that R is artinian.

Corollary 20.5.2: 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 20.5.2: 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: 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

Exercise 20.5.1

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 $M = \ker(\varphi^n) \oplus \operatorname{im}(\varphi^n)$.

21

Relating Algebra to Geometry

In this section we will explore how to establish a link between geometrical concepts with algebric concepts. (TBD: move intuition from section 21.2 to here)

(This chapter builds up the dictionary between geometry and algebra, and ends on briefly outlining a prelude to modern algebraic geometry and how this dictionary can be generalized to general rings).

21.1 Reminder of Noetherian Rings

Definition 21.1.1: 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 21.1.1: 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 21.1.2: 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 21.1: 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*.

Theorem 21.1.1: Hilbert's Basis Theorem

Let R be a Noetherian ring. Then $R[x]$ is a Noetherian ring. More generally, $R[x_1, x_2, ..., x_n]$ (a polynomial ring over severable variables) is Noetherian.

Proof **:**
this is Theorem 11.6.1 and corollary 11.6.1.

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 20.5, 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 21.1.1

1. 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?
2. Let S be a finitely generated R-algebra, where R is Noetherian. Prove that S is Noetherian.
3. 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?
4. 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 (move this to field extension...)
4. $R := k + xk[x, y]$ as a subring of $k[x, y]$, k a field.

5. 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?)

21.2 Affine Algebraic Sets

In this whole section, 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 however a natural way of thinking about any such polynomial as a function (which is probably how most first encounter polynomials): if $f \in k[x_1, x_2, ..., x_n]$, where , 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

$$k[x_1, ..., x_n] \to \mathrm{Hom}_{\mathbf{Set}}(k^n, k) \qquad f \mapsto ((a_1, ...a_n) \mapsto f(a_1, ..., a_n)) \tag{21.1}$$

Note that the codomain of this function 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, $\mathrm{Hom}_{\mathbf{Set}}(k^n, k)$ has a natural k-algebra structure given since the codomain k is a k-algebra. The image of the above function is given a name:

Definition 21.2.1: Regular Function

Let F represent the function in equation (21.1). Then the collection $\mathrm{im}(F)$ is called the *regular functions*, and is usually denoted $k[\mathbb{A}^n]$ or $\mathrm{Hom}_{\mathbf{Reg}}(k^n, k)$.

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 to work with. In Algebra, we can think of polynomial functions as the "natural" types of functions in a similar sense to how continuous 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 a polynomial:

Example 21.2: Zeros' of Polynomials

Let $R = \mathbb{R}$ be the real numbers. Then:

Give a circle, and perhaps some fancy shape, but also say there are limits to what geometric shapes this method can capture, as we will soon see.

Therefore, 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!

Before continuing, there is an important notational change that is important to clarify. The set of regular function was written as $k[\mathbb{A}^n]$, and not $k[k^n]$, that is, instead of writing the domain of a polynomial as k^n, we we wrote it as $\mathbb{A}^n$ (or in some cases $\mathbb{A}^n_k$ if the field not clear from context). The set $\mathbb{A}^n$ is called it the *affine n-space*. For example, if $p \in \mathbb{R}[x, y]$, then $p : \mathbb{A}^2 \to \mathbb{A}^1$ will be our function. As a set, $\mathbb{A}^2$ ($\mathbb{A}^2_k$) is identical to k^2. The reason for the notational change is that k^n is commonly associated with the vector-space over k of dimension n. A vector-space has a "special" element: the identify. In $\mathbb{A}^n$, every element is treated equally: the $(0, 0, ..., 0)$ element will play no special role. Thus, we can think of $k[\mathbb{A}^n]$ to mean we are not giving any special role to the points of the domain (similarly to how in Classical Euclidean Geometry, there is no "origin")

With the notation established, we'll give a name to the zero's of a set:

Definition 21.2.2: 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* (or just algebraic set for short). V is called the *locus of S in $\mathbb{A}^n$*.

The word "affine" in the definition is because we are taking the zero's of an affine space. There is a different construction (which I don't understand yet) which starts with the dual of vector-spaces over algebraically closed fields, which is naturally a graded symmetric algebra and so also acts like polynomial rings, and yet another starts with $\mathbb{C}\mathrm{P}^n$. The advantage of this I'm not aware of yet.

From now on, when I use the notation $A \subseteq \mathbb{A}^n$ for some variable A (not necessary the letter A), 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 21.3: 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 more roots is $0(x)$, which has all of k as its zero. Therefore, in $\mathbb{A}^1$, the affine sets are

$$\emptyset \qquad \text{finite sets} \qquad k$$

   This also shows that not all sets are algbraic 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

$$\mathbb{Z}(f) = \left\{(x, y) \in \mathbb{A}^2 \mid x^2 + y^2 = 1\right\}$$

   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?)

4. Given any linear transformation

5. Show that $V(z^2 - 1) \subseteq \mathbb{C}$ is no an algebraic set.

6. 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)]

7. Let k be a finite field and $S \subseteq \mathbb{A}^n(k)$. Show that S is always an algebraic set.

8. 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 $\mathbb{R}^3$.

9. 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 21.2.1: 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 inclusino)
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:
$$x \in \mathcal{Z}(\cup_{i \in I} S_i) = \{y \in \mathbb{A}^n \mid f(y) = 0,\ \forall f \in \cup_{i \in I} S_i\}$$
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 $1 \in I$), 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 let $\mathcal{Z}$ be a function from the ideals of $k[\mathbb{A}^n]$ to the algebraic sets in $\mathbb{A}^n$:

$$\mathcal{Z} : \{\text{ideals in } k[\mathbb{A}^n]\} \to \{\text{affine algebraic sets in } \mathbb{A}^n\}$$

Furthermore, $\mathcal{Z}$ is a surjection from the ideals of $k[x_1, x_2, ..., x_n]$ and the algebraic sets, meaning we can without any loss work with the ideals of $k[\mathbb{A}^n]$. Next, since $k[x_1, ..., x_n]$ is Noetherian, every ideal is finitely generated: $I = (f_1, ..., f_m) = (f_1) \cup (f_2) \cup \cdots \cup (f_m)$. By the 3rd point, we see that:

$$\mathcal{Z}(I) = \mathcal{Z}(f_1) \cap \mathcal{Z}(f_2) \cap \cdots \cap (f_m)$$

meaning we can work on seeing if there is a decomposition theorem for algebraic sets! This will indeed be our endeavour in section ref:HERE. Next, the finiteness condition on the product of ideals is a necessary condition for closure of algebraic varities: see if you can come up with a countable collection of algebraic sets whose union is not an algebraic set (such a set exists). Next, since $\mathcal{Z}(A) = \mathcal{Z}((A))$, notice that $\cap_\alpha \mathcal{Z}(I_\alpha) = \mathcal{Z}(\sum_\alpha I_\alpha)$; this should be proven[1]! Next, we took the product of ideal $\prod_i^n I_i$ instead of the inersection. Since it always the case that $IJ \subseteq I \cap J$ then in the cases where $IJ \subsetneq I \cap J$, $\mathcal{Z}(I \cap J) \subsetneq \mathcal{Z}(I) \cup \mathcal{Z}(J)$[2]. 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 obselete (see exercise ref:HERE- 2nd one as of time of writing this document)

Perhaps most importantly for this chapter, notice that these properties are the properties of a topology (see appendix C):

[1] Hint: prove that $I + J = (I \cup J)$, i.e. $I + J$ is the smallest ideal containing I and J

[2] The condition that $IJ = I \cap J$ is surprisingly technical, requiring that $\mathrm{Tor}^1(R/I, R/J) = 0$, see ref:HERE

Definition 21.2.3: 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[3]

Towards the goal of finding an inverse, 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 21.2.4: 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 *vanising ideal* associated with V

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

$$\mathcal{I} : \{V \subseteq \mathbb{A}^n\} \to \{\text{ideals of } k[\mathbb{A}^n]\}$$

Example 21.4: 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. Ho if k is infinite , $\mathcal{I}(\mathbb{A}^n) = 0$. Ho not otherise
3. here

Just like $\mathcal{Z}$, $\mathcal{I}$ shares similar properties:

[3]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

Proposition 21.2.2: 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 21.2.1.

Notice that we have ommited showing what $\mathcal{I}(A \cap B)$ is. The main 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 $(\mathcal{I}(A) \cup \mathcal{I}(B)) = \mathcal{I}(A) + \mathcal{I}(B)$, that

$$\mathcal{I}(A \cap B) = \mathcal{I}(A) + \mathcal{I}(B)$$

However, this is not generally the case. 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 $\mathcal{I}(A \cap B) \supseteq \mathcal{I}(A) + \mathcal{I}(B)$. 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 21.2.3: 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, $\mathcal{I}(\mathcal{Z}(\mathcal{I}(V))) = \mathcal{I}(V)$, $\mathcal{Z}(\mathcal{I}(\mathcal{Z}(V))) = \mathcal{Z}(V)$

***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 21.2.3 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 we had was that

$$\mathcal{Z}(I) = \mathcal{Z}(f_1) \cap \mathcal{Z}(f_2) \cap \cdots \cap (f_m)$$

what does this imply about the decomposition of algebraic sets to simpler algebraic sets? Furthermore, we have been shown theoretically what properties algebriac sets have, but how would we go about actually computing one? These questions will recieve answers in the following section, and will help us in gaining an intuition for how the *Nullstellensatz* is proved.

Exercise 21.2.1

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 $\mathcal{Z}(I \cap J) = \mathcal{Z}(IJ)$.

21.2.1 Zariski Topology

In this section, we will be assuming basic topological knowledge from the reader.

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 are the empty set, finite sets, and A^1 set, these are the closed sets, 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 seperate 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 21.3), an and so it is always a Fréchet space (or T_1). (A word on finite fields?).

This result generalizes for any k^n

Proposition 21.2.4: Zariski Topology Always T_1

Let k be an infinite field and k^n have the Zariski topology. Then k^n is T_1, i.e. every singleton is always closed, and 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 Hasudorff, we already showed it in the case of $n =$ in example 21.3, so we'll do k^n with $n > 1$. Take $k \times \{0\}^{n-1}$ with the sub-space topology. Then any such set 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 comlpiment has co-finitely many ponits. It is now easy to verify that the space is not Hasudorff.

Two topological concepts which should be immediately introduced to be explored as the chapter develops are the following:

Definition 21.2.5: 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 $\overline{A}$.

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 21.2.3:

Corollary 21.2.1: Zariski Closure

Let $A \subseteq \mathbb{A}^n$. Then:

$$\overline{A} = \mathcal{Z}(\mathcal{I}(A))$$

Proof **:**
p. 678 dummit if you're stuck: bassically, $\mathcal{I}(A)$ 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.

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 21.2.5: Basis For Zariski Topology

Let $A \subseteq k[\mathbb{A}^n]$. Define

$$D(A) = k^n \setminus \mathcal{Z}(A) = \{a \in k^n \mid f(a) \neq 0,\ \forall f \in A\}$$

Then the collcetion $\{D(f)\}_{f \in k[A^n]}$ forms a [topological] basis for the Zariski topology on $\mathbb{A}^n$.

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 $U = k^n \setminus \mathcal{Z}(I)$. 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.

(Next, want to find the contuous functions. They are in fact the *regular functions* described earlier. Make it inuitive on how for this reason, the Zariski topology is the "coarest" topology in which polynomials are continuous)

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 stabalize:

Proposition 21.2.6: Zariski Topology Is Noetherian

Let k^n have the zariski topology. Then k^n is Notherian.

***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 k^n is in fact always *compact*, but we require the Nullstellensatz to prove this result [4].

The next topological notion we wish to bring in is the idea of *subspaces* and subspace topologies. In particular, if we have any subset $A \subseteq k^n$, we can define the Zariski topology on A by taking the intersection of all open sets of A. Note that all of the previous properties we have still apply to A since all singleton's are closed (proposition 21.2.4). We will always limit ourselves to taking A to be an algebraic set, since algebaic set are the object of interest[5]. The first thing we wish to establish is for 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?

taking a moment

Definition 21.2.6: Affine Coordinate Ring

Let $k[X]$ be a polynomial ring over n variables, V be an algebraic set, and $I(V)$ be the coresponding coordinate ideal. Then:

$$k[X]/I(V)$$

is called the *coordinate ring*.

Notice that this ring can be quite naurally defined: Since $\mathrm{Hom}_{\mathbf{Set}}(V, k)$ has a natural ring structure (namely $f + g(a) = f(a) + g(a)$ and $fg(a) = f(a)g(a)$), take the ring homomorphism:

$$k[x_1, ..., x_n] \to \hom_{\mathbf{Set}}(V, k) \qquad f \mapsto (a \mapsto f(a))$$

Then we see that the kernel of this map is by definition $I(V)$, hence:

$$k[V] = k[x_1, ..., x_n]/I(V) \hookrightarrow \hom_{\mathbf{Set}}(V, k)$$

Notice how this notation is the same as for regular functions: $k[\mathbb{A}^n]$. This matches, since $I(\mathbb{A}^n) = 0$ (when the field is infinite).

With this, we may define continuous functions between subspaces:

[4] the Nullstellensatz is theorem 21.3.4 on page 627 and then the Zariski topology being compact is proposition 21.3.1 on page 629

[5] Once we show that k^n is compact in the Zariski topology $A \subseteq k^n$ will also be compact, being a closed set of a compact space

Definition 21.2.7: Regular Functions on Subspaces

Let $V, W \subseteq \mathbb{A}^n$ be algebaic sets. Then a map $\varphi : V \to W$ is called a *regular function*, *regular map*, or *morphism*, if there exists polynomials $\varphi_1, \varphi_2, ..., \varphi_n \subseteq k[x_1, x_2, ..., x_n]$ such that

$$\varphi(a_1, a_2, ..., a_n) = (\varphi_1(a_1, a_2, ..., a_n), \varphi_2(a_1, a_2, ..., a_n)..., \varphi_n(a_1, a_2, ...a_n))$$

Example 21.5: Coordinate Ring

1. Let $V = \mathbb{A}^n$ so that $\mathcal{I}(V) = 0$. Then $k[V] = k[\mathbb{A}^n]$, meaning all the functions of $\mathbb{R}^n$
2. Let $V = \mathcal{Z}(xy - 1)$, which represents the hyperbola $y = 1/x$ in $\mathbb{R}^2$. Then the coordiante 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 $\overline{f}$.

(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)

If $k[\mathbb{A}^n]$ is the algebraic representation of $\mathbb{A}^n$, then $k[V]$ is the algebraic representation of V. (define "same function on $k[V]$")

(build up to this)

Theorem 21.2.1: 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

$$\{\text{regular functions from } V \text{ to } W\} \leftrightarrow \{k\text{-algebra homomorphisms from } k[W] \text{ to } k[V]\}$$

$$\left\{\begin{matrix}\text{regular functions} \\ \varphi : V \to W\end{matrix}\right\} \underset{\Phi}{\overset{f \circ \varphi}{\rightleftarrows}} \left\{\begin{matrix}k\text{-algebra homomorphism} \\ \tilde{\varphi} : k[W] \to k[V]\end{matrix}\right\}$$

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

(a ay to define a morphism without any reference to the ambient space)

Proposition 21.2.7: Regular Function Characterization

Let $\varphi : V \to W$ be a regular function with $\tilde{\varphi} : k[W] \to k[V]$ the associated k-algebra homomorphism. Then:

1. The kernel of $\tilde{\varphi}$ is $\mathcal{I}(\varphi(V))$
2. The Zariski closure of $\varphi(V)$ is the zero set in W of $\ker(\tilde{\varphi})$. As a consequence, the homomorhpism $\tilde{\varphi}$ is injective if and only if $\varphi(V)$ is Zariski dense in W

Proof **:**
p. 678 if you're lost

(allos using Grobner basis)

(example with implicitization and projection of alg sets)

21.2.2 Radical Ideals

As we've seen in proposition 21.2.1, given some polynomial f or collection of polynomials F, the set of zero's of these polynomials does not change if we take higher powers of these (ex. $f, f^2, f^3, \ldots$ 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 21.2.8: 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 hich some poer of them lie in I, that is:

$$\sqrt{I} = \operatorname{rad}(I) := \left\{ a \in R \;\middle|\; a^k \in I \text{ for some } k \in \mathbb{N} \right\}$$

   here $\operatorname{rad}(I)$ stands for the radical of I. It is often written as $\sqrt{I}$.
2. The radical of the zero ideal ($\operatorname{rad}((0))$) is calle the *nilradical* of R
3. An ideal I is called a *radical ideal* if $I = \operatorname{rad}(I)$ ($I = \sqrt{I}$)
4. The ideal $\sqrt{(0)} + N$ is called the *nilradical*.

Note that the nilradical by definition contains all the nilpotent elements of R, and that $\operatorname{rad}(\operatorname{rad}(I)) = \operatorname{rad}(I)$.

Proposition 21.2.8: $\operatorname{rad}(I)$ Is An Ideal

Let $I \subseteq R$ be an ideal of a commutative ring. Then $\operatorname{rad}(I)$ is indeed an ideal (containing I). Furthermore, $\operatorname{rad}(I)/I$ is the nilradical of R/I.

The last statement shows us that R/I has no nilpotent elements if and only if $I = \operatorname{rad}(I)$.

Proof **:**
p. 673 dummit if you're stuck

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. Similar to in Galois Theory, as the number of prime's get larger, the radical get's smaller:

Proposition 21.2.9: Radical Through Prime Ideals

Let $I \subsetneq R$ be a proper ideal of a commutative ring and $\operatorname{rad}(I)$ the radical ideal of I. Then $\operatorname{rad}(I)$ 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 cae 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 21.2.2: Prime Or Maximal Then Radical

Let $I \subseteq R$ be a prime or maximal ideal. Then I is a radical ideal

***Proof* :**
Let P be a prime or maximal ideal. Since maximal ideals are prime, by proposition 21.2.9,

$$P = \text{rad}(P)$$

completing the proof.

Example 21.6: Radical Ideals

1. Easy for a UFD
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 21.2.10: 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 prove one direction of this:

Proposition 21.2.11: 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 also get that $\sqrt{I} = \mathcal{I}(\mathcal{Z}(I))$, 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 $\mathcal{Z}(I) = \emptyset$, so $\mathcal{I}(\mathcal{Z}(I)) = \mathbb{R}[x]$ But $I \supsetneq \mathbb{R}[x]$, 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 propertie of $\mathcal{I}$ we have hat $\mathcal{I}(\mathcal{Z}(I)) = ((x - i)(x + i)) = (x^2 + 1) = I$, 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 21.3.4 (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 Algebraic Variety

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 posess

Definition 21.2.9: Reducible and Irreducible

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*

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 21.2.12: 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 21.2.3: 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 21.2.2: 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 21.2.10: 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, $1 \leq i \leq n$

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 21.2.2. It then follows easily that any closed subset of X has this property too.

(this lends to the important ability to generate a field out of $k[V]$, giving us the poer of field theory)

Definition 21.2.11: Rational Functions On V

Let $V \subseteq \mathbb{A}^n$ be an algebraic variety with assocatied (integral domain) coorinate ring $k[V]$. Then the field of fractions $\text{Frac}(k[V])$ is called the field of *rational functions* on V, and is denoted $k(V)$.

The *dimension* of V is the transcendence degree of $k(V)$ over k, and is denoted $\text{trdeg}(V)$, or $\dim(V)$ if it is clear that V is an algebraic variety.

(many examples)

21.2.4 Primary Decomposition

(ring parrallel)

(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)

Corollary 21.2.4: 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 **:**

21.2.5 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 eventualy understand.

Restrictions on the ru

21.3 Hilbert's Nullstellensatz

(Noether's normalization theorem backbone of many results, including zariski (which goes on to prove Nullstellensatz), and defining dimension of varieties (through means of trdeg i.e. $\dim(X) = \text{trdeg}(k(X)/k)$)). 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 a the genelization of the idea that any field extension can be split into a transcendental extension, folloed by an algebraic extensio:

Theorem 21.3.1: 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:

$$g = cx_n^N + \sum_{i=1}^{N-1} h_i(X_1, ..., X_{m-1})x_n^i$$

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 thm is the basis for a lot. One really important and rather crazy result 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 genered by *just* allowing addition!!)

Theorem 21.3.2: 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 20.1.2, 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 21.3.3: 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 18.5.1), 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 21.3.4: Hilbert's Nullstellensatz

Let k be an algebraically closed field. Then for any ideal $I \subseteq k[x_1, x_2, ..., x_n]$,

$$\mathcal{I}(\mathcal{Z}(I)) = \text{rad}(I)$$

Furthermore, the maps $\mathcal{I}$ and $\mathcal{Z}$ are order-reversing bijections:

$$\left\{\text{affine algebraic sets}\right\} \underset{\mathcal{Z}}{\overset{\mathcal{I}}{\rightleftarrows}} \left\{\text{Radical Ideals}\right\}$$

***Proof* :**
First, we've already shown that $\text{rad}(I) \subseteq \mathcal{I}(\mathcal{Z}(I))$ (proposition 21.2.11), so it remains to show hat $\text{rad}(I) \supseteq \mathcal{I}(\mathcal{Z}(I))$. 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, $g \in \text{rad}(I)$. Since g was arbitrary, we see that $\mathcal{I}(\mathcal{Z}(I)) \subseteq \text{rad}(I)$, and so we get $\mathcal{I}(\mathcal{Z}(I)) = \text{rad}(I)$.

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.

(corrollary vairant + the Remark Aluffi did, words on further generaliation to Jacobain Radicals?)

(note that this theorem can be generlized for non-algebraically closed fields as long as e take $\mathbb{A}^n$ to be over $\overline{k}$)

(this now shows motivation for the correspondence between primary ideals and prime ieals, as well as the fact that the intersection of radical ideals is radical, and that any radical ideal is the intersection of all prime

ideals containing it)

(Also, we can finally complete the following:

$$\mathcal{I}(A \cap B) = \sqrt{\mathcal{I}(A) + \mathcal{I}(B)}$$

)

Overall, we get the following dictionary between the langauge of geometry and the language of algebra:

Geometry	Algebra
Affine algebraic sets V	coordinate rings $k[V]$
points of V	maximal ideals of $k[V]$
affine algebraic subsets of V	radical ideals of $k[V]$
subvarieties of V	prime ideals of $k[V]$
regular function: $\varphi : V \to W$	k-algebra homomorphism $\tilde{\varphi} : k[W] \to k[V]$

21.3.1 More on Zariski Topology

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 21.3.1: Zariski Topology is Compact

Let k^n have the Zariski topology. Then k^n is compact

Proof **:**
Let $\mathcal{O}$ index an open cover for k^n

$$k^n = \bigcup_{\alpha \in \mathcal{O}} U_\alpha$$

By proposition 21.2.5, any open set is the union of sets of the form $D(f)$, so we may re-write this as:

$$k^n = \bigcup_{\alpha \in \mathcal{O}'} D(f_\alpha)$$

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 coefficeints $g_i \in k$, 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

$$\emptyset = \bigcap_{i=1}^{n} \mathcal{Z}(f_i) \quad \Leftrightarrow \quad k^n = \bigcup_{i=1}^{n} D(f_i)$$

Recall that for any algebraic subset $V \subseteq k^n$, the corresponding coordinate ring $k[V]$ is used to define the subspace topology. Using the Nullstellensatz, e may deduce that it cannot have nilpotent elements:

Corollary 21.3.1: Coordinate Ring Always Reduced

Let $V \subseteq k^n$. 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.

Definition 21.3.1: Reduced Ring

Let R be a ring. Then R is said to be *reduced* if $x^n = 0$ implies $x = 0$

Corollary 21.3.2: Nullstellensatz For Subspaces

Let $V \subseteq k^n$ 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

Thus, we can imagine a a change of subsets with the following property

$$\begin{tikzcd}
\{\text{radical ideals of } k[V]\} & \{\text{closed subsets of } V\} \\
\{\text{Prime ideals of } k[V]\} \arrow[u, "\subset"] & \{\text{irreducible subsets of } V\} \arrow[u, "\subset"] \\
\{\text{Maximal ideals of } k[V]\} \arrow[u, "\subset"] & \{\text{points of } V\} \arrow[u, "\subset"]
\end{tikzcd}$$

(How this links to Localization here)

So far, we have estalished the link between ideals of a commutative ring R and the Zariski topology on R. Since we are working with a topology, we are often interested about the behavior near a point p (for example, when defining continuity or differentiability at p, we don't just care what happens at p, but near p). Let's say $p \in X$ where $X \subseteq \mathbb{A}^n$ is an algebraic set. Then the Zariski open sets that contain p are sets of the form $X - Y$ where $Y \subseteq X$ is an algebraic set and $p \notin Y$. However, Y here is in the subset topology given by X, and so in general is not an algebraic set of our Zariski topology (there is exericse 3.5 in Harthorne about a plane and a point, will checkout later). We will mitigate this by picking small neighbourhoods of p which correspond to a large algebraic set Y which can be associated to the vanishing of a single function f which does not vanish at p. In this case, it will turn out that $X - Y$ is an algebraic set in $\mathbb{A}^{r+1}$.

Exercise 21.3.1

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)

21.4 Dimension Theory

Put here the last chapter of Michael's commutative algebra book; I think it's really cool and insightful.

21.5 Spectrum of a Ring

We have now created a dictionary translating properties of algebraic sets to coordinate rings. In particular, we have that every algebraic set correponds to a radical ideal (and an algebraic variety corresponds to a prime ideal), points of algebraic sets correspond to maximal ideals, and regular functions correspond to k-algebra homomorphisms between coordiante rings. All of this relied on the intuition of treating elements of $k[x]$ as functions that we may evaluate. What if we can generalize this notion to arbitrary rings R? How can we define elements $f \in R$ as "functions" that can "evaluate" at a point? To what space would the evaluation result in? So far, it was taken for granted that for $k[x_1, x_2, ..., x_n]$, the evalutation of the elements results in an element in k^n. To generlize this dictionary, we would have to create a new space to evaluate to, as well as a new notion of "function" for arbitrary elements of rings.

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, Given $V = \{x \in k \mid x = 0\}$, there is no way of recovering that V was generated by polynomial of degree 2). This will lead to the notion of a *scheme*. In this section, we build up the definitions and key ideas for I believe these are incredibly interesting concepts. However, further exploration are far outside the scope of this book and is the beginning of algebraic goemetry. The notions themselves are quite interesting and will be an interesting starting point for an introductory text to algebraic geometry.

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-variantely 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 $\varphi^{-1}(P)$ is prime. This inspires the following definition:

Definition 21.5.1: 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 $\mathrm{Spec}(R)$:

$$\mathrm{Spec}(R) := \{P \subseteq R \mid P \text{ is a prime ideal of } R\}$$

Similarly, the collection of all maximal ideals is called the *maximal spectrum* of R and is denoted $\mathrm{mSpec}(R)$.

Example 21.7: 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 21.5.1: $\mathrm{Spec}(-)$ 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 $\mathrm{Spec}(R)$ 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 $\mathrm{Spec}(R)$ (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" $P \in \mathrm{Spec}(R)$?

We generlize this to the following: let's first consider when $f \in k[x]$. If $\mathrm{Spec}(k[x])$ is the spectrum of $k[x]$, then the *value* of f at $p \in \mathrm{Spec}(R)$ 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:

$$R \to \mathrm{Hom}_{\mathbf{Set}}\left(\mathrm{Spec}(R), \bigsqcup_{P \in \mathrm{Spec}(R)} R/P\right) \qquad f \mapsto (P \to f \mod P)$$

Definition 21.5.2: Value Of Element

Let $f \in R$. then the "value" of f at the "point" $P \in \mathrm{Spec}(R)$ is the element:

$$f(P) := \overline{f} \in R/P$$

Next, we see that we can define the maps $\mathcal{Z}$ and $\mathcal{I}$ as follows. For $\mathcal{Z}$, we get:

$$\mathcal{Z}(A) = \{P \in \mathrm{Spec}(R) \mid A \subseteq P\} \subseteq \mathrm{Spec}(R)$$

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 $I \subseteq P$, so $f \in P$ for every $f \in I$. If we view $f \in R$ as a "function" on $\operatorname{Spec}(R)$, we see that $P \in \mathcal{Z}(I)$ if and only if

$$f(P) = f \mod P = 0 \in R/P$$

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:

$$\mathcal{I}(Y) = \bigcap_{P \in Y} P$$

Proposition 21.5.2: 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 P contains $\operatorname{rad}(I)$, hence $\mathcal{Z}(I) = \mathcal{Z}(\operatorname{rad}(I))$. Since $\operatorname{rad}(I)$ is the inersection of all prime ideals containing I, the definition of $\mathcal{I}(I)$ gives $\mathcal{Z}(\operatorname{rad}(I)) = \mathcal{Z}(\mathcal{I}(I))$. Similarly:
$$\mathcal{I}(\mathcal{Z}(I)) = \mathcal{I} \bigcap_{P \in \mathcal{Z}(I)} P = \bigcap_{I \subseteq P} P = \operatorname{rad}(I)$$
2. essentially immediate
3. essentially immediate

This shows that we are back in the condition's for the Nullstellensatz: every $\mathcal{Z}(I)$ in $\operatorname{Spec}(R)$ occurs as a radical ideal, which is unique since $\mathcal{I}(\mathcal{Z}(I)) = \operatorname{rad}(I)$. Furthermore, the 2 other statements shows that we have also kept the same topology on $\operatorname{Spec}(R)$ 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 21.5.3: Zariski Topology

Let R be a ring and $\operatorname{Spec}(R)$ the corresponding spectrum. Then the collection of $\mathcal{Z}(I)$ defines the closed sets of the *zariski topology* on $\operatorname{Spec}(R)$.

Definition 21.5.4: Closed Points

Let $m \subseteq R$ be a maximal ideal of R and $\operatorname{Spec}(R)$ be the specturm of R. Then the maximal ideals of R in $\operatorname{Spec}(R)$ 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 $\operatorname{Spec}(R)$ is still T_1, and again rarely Hausdroff.

Definition 21.5.5: 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 21.5.3: Nullstellensatz On Spectrum

Let R be a ring and $\text{Spec}(R)$ the spectrum with the Zariski topology. Then the maps $\mathcal{Z}$ and $\mathcal{I}$ define order-reversing bijections:

$$\left\{\begin{matrix}\text{Zariski closed subset of} \\ \text{Spec}(R)\end{matrix}\right\} \underset{\mathcal{Z}}{\overset{\mathcal{I}}{\rightleftarrows}} \{\text{Radical Ideals of } R\}$$

Proof **:**
Use proposition 21.5.2.

Example 21.8: 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 $\text{Spec}(R)$.

The 3rd example given can be generlized:

Definition 21.5.6: 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 $\text{Spec}(R)$ have prime ideals. The following generlizes the notion of regular function)

Proposition 21.5.4: Spectrum Map

Let $\varphi : R \to S$ be a ring homomorphism. Then there is an indcued map $\varphi^* : \text{Spec}(S) \to \text{Spec}(R)$ mapping

$$\varphi^*(P) = \varphi^{-1}(P)$$

This map is continuous with respect to to the Zariski Topology

***Proof* :**
exercise

(many examples)

Proposition 21.5.5: 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 21.5.7

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 21.5.8: 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 21.5.9: Stalk

here

Definition 21.5.10: 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 $\varphi^{\vee}: W^{\vee} \to V^{\vee}$ that pre-composes a linear functional $W \to \Bbbk$ with φ to obtain a linear functional $V \xrightarrow{\phi} W \xrightarrow{\omega} \Bbbk$.
>
> (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 $C\colon \mathsf{Top}^{\mathrm{op}} \to \mathsf{Poset}$ 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

Algebraic number theory can be thought of as looking for solutions to equations. Perhaps the most famous number theory problem 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 Problem*. This is trivial if $n = 1$ (ex. $1+2=3$). It is more interseting 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 *pythagorian triple*.

(I wouldn't mind going over here the development of Fermat's last theorem, how it is porved in $n = 2$, how it is generlized for regular primes, and how there are stumbling blocks for irregular primes. For now, I will assume I developed all of this.)

As we see in the above, the ring $\mathbb{Z}[\omega]$ is not always a Unique Factorization Domain, however there is a generalization by Dedekind and Kummer allowing us to consider the more general notion of factorizing the ideals of a ring to prime ideals. One of the key facts is to work with a ring other than $\mathbb{Z}[\omega]$ that is Unique Factorization Domain.

> **Warning** Some elementary analysis will be assumed in this section, including some properties of the real numbers.

22.1 Ring of Integers

We start by defining our terms.

Definition 22.1.1: Algebraic Number Fields

Let k be a field. If $\mathbb{Q} \subseteq k$ is a algebraic extension of $\mathbb{Q}$, then k is called an *algebraic number field*.

Example 22.1: 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 fields always have degree 2. notice that we can consider any $m \in \mathbb{Z}$, however we get many equality of fields, for exmaple $\mathbb{Q}(\sqrt{12}) = \mathbb{Q}(3)$. The field $\mathbb{Q}(i)$ is both a quadratic field and a cyclotomic field since $i = \zeta_4$. As an exercise, show that $\mathbb{Q}(-3)$ is a cyclotomic field.

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 ireducible. An important ring of integer that is worth isolating is the following:

Definition 22.1.3: Algebraic Integers

The collection of elements

$$\mathcal{O}_{\mathbb{C}} = \overline{\mathbb{Z}}^{\mathbb{C}}$$

are called the *algebraic integers*.

Example 22.2: Ring Of Integers

1. Let $k = \mathbb{Q}(\sqrt{m})$. Then:
$$\mathcal{O}_k = \mathbb{Z}[\sqrt{m}] \qquad m \equiv 2, 3 \mod 4$$
$$\mathcal{O}_k = \mathbb{Z}\left[\frac{1+\sqrt{m}}{2}\right] \qquad m \equiv 1 \mod 4$$
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 inegers. The rest is now computation.
2. $\mathcal{O}_{\mathbb{Q}} = \mathbb{Z}$. For this reason, the integers in algebraic number theory are sometimes called the rational integers
3. $\mathcal{O}_{\mathbb{Q}(i)} = \mathbb{Z}[i]$. 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 in the equation:
$$\alpha^2 - 2a\alpha + (a^2 - b^2) = 0$$
$-2a$ and $a^2 - b^2$ are integers. From the second equation we get $(a-b)(a+b) = z$ for some $z \in \mathbb{Z}$. Doing some manipulations, we get that $a, b \in \mathbb{Z}$, and so $\alpha = a + bi$, $a, b \in \mathbb{Z}$, telling us that $\mathcal{O}_{\mathbb{Q}(i)} = \mathbb{Z}[i]$

4. 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 $\omega \notin \mathbb{Z}[\sqrt{-3}]$. Thus, we will need more tools to find $\mathcal{O}_F$.

We should think of ring of integers as being partially ordered 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 for any field k), 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 $a_i \in \mathbb{Z}$ and
$$\varphi(a_i) = \varphi(1 + \cdots + 1) = \varphi(1) + \cdots + \varphi(1) = 1 + +1 = a_i$$
Hence $\varphi(\alpha) \in \mathcal{O}_L$. From this, we may define $\mathcal{O}_- : \mathbf{Fld}^0 \to \mathbf{Ring}$ where $\mathbf{Fld}^0$ are fields of characteristic 0. Since we have an order on the $\mathcal{O}_k$ that mirrors the order of field inclusions, we would like to know how to move between the orders, and furthermore we may wonder what the adjoint to this functor is. We shall first show how to go from rings of integers back to their original fields

Proposition 22.1.1: Field Of Fraction Of Ring Of Integers

Let R be an integral domain, $K = K(R)$ its field of fraction, and $B = \overline{R}^L$ where L/K is an algebraic extension. Then $L = K(B)$. Equivalently, $L = \langle B \rangle_{K(R)}$.

In number theory, we would have $R = \mathbb{Z}$ and $K(R) = \mathbb{Q}$ so that $B = \overline{\mathbb{Z}}^L = \mathcal{O}_L$. Then $L = K(\mathcal{O}_L)$.

Proof **:**
Let R, K, B and L be defined as in the question. We shall first show that if $\alpha \in L$ is algearic 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 multiplce 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 $L = \langle B \rangle_{K(B)}$.

With that, to see that $L = K(B)$, notice that every $\alpha \in L$ can be written as $\alpha = b/d$ for $d \in R$, hence $L \subseteq K(B)$. Conversely, if $a/b \in K(B)$, where $a, b \in B$ and each satisfy an R-polynomial, then they are certainly in B, giving equality.

Corollary 22.1.1: 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 22.1.1.

Proposition 22.1.2: Relating Rings Of Integers

Let k be a field and $\mathcal{O}_k$ the ring of integers of k. Then:

1. $\mathcal{O}_k \cap \mathbb{Q} = \mathbb{Z}$. More generally, if L/k, then $\mathcal{O}_L \cap k = \mathcal{O}_k$
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.

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 20.2.4 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 how to find elements in $\mathcal{O}_k$.The strategy in example 22.2 to find the structure of $\mathcal{O}_k$ gets much harder for more complicated roots of irreducibles, even once we have degree 3, and especially beyond degree 5. 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 $\mathcal{O}_K$. Before developing those tools, there are some simple remraks we may already make about $\mathcal{O}_k$:

22.1.1 Trace and Norms

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 of eleements 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 independnet calculations, it is meaningul to ponder what the trace/determinant of this transformation and what information it may provide us:

Definition 22.1.4: 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 22.3: 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 tarce 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 compelx number (up to a factor).

2. Let $M = k(\alpha)$ finite, 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:

$$\operatorname{Nm}_{F/k}(a_1 a_2) = \operatorname{Nm}_{F/k}(a_1)\operatorname{Nm}_{F/k}(a_2) \quad \operatorname{Nm}_{F/k}(a) = a^n$$

Most of the properties of the trace and norm come from our observation in example 22.3, namely that the trace and norm of α is the first/last coefficient of its last coefficient, which we know information about from section 19.2.3. This is important enough to be given its own box:

Proposition 22.1.3: Trace And Norm via Galois Conjugates

Let M/k be a finite extension where where $[M : k] = n$ and let $\alpha \in M$ be algebraic over k. Then:

$$\operatorname{Trace}(\alpha) = [M : k(\alpha)] \cdot \sum_{\substack{r \\ \text{root of } m_{\alpha,k}}} r \qquad \operatorname{Norm}(\alpha) = \left(\prod_{\substack{r \\ \text{root of } m_{\alpha,k}}} r \right)^{[M:k(\alpha)]}$$

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 embedings 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 concretly computing the trace, we get the $n-1$th coefficeint 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 19.2.3, we see that

$$\alpha_{n-1} = \sum_i \sigma_i(\alpha)$$

that is, it is the sum of the roots. Similarly, we get that $\operatorname{Nm}_{F/k}(\alpha) = a_0 = \prod_i \sigma_i(\alpha)$, 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 22.1.2: 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 $\alpha \in L$:

$$\mathrm{tr}_{L/k}(\alpha) = \sum_i^n \sigma_i(\alpha) \qquad \mathrm{Nm}_{L/k}(\alpha) = \prod_i^n \sigma_i(\supset$$

Proof **:**
this is just a re-phrasing of the previous result.

This immediately shows us that $\mathrm{tr}_{M/k}(\alpha), \mathrm{Nm}_{M/k}(\alpha) \in k$, since $\pm a_{n-1}, a_0 \in k$. To relate this back to your goal of undersatnding $\mathcal{O}_M$, 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:

Corollary 22.1.3: 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 fration of R. Then if $\alpha \in F$ is integral over R,

$$\mathrm{tr}_{F/K(R)}(\alpha), \mathrm{Nm}_{F/K(R)}(\alpha) \in R$$

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 22.1.2.

If $R = \mathbb{Z}$ and $F/\mathbb{Q}$ is a finite extension, then this means that we may identify algebraic elements by noticing that $\mathrm{tr}_{F/\mathbb{Q}}(\alpha), \mathrm{Nm}_{F/\mathbb{Q}}(F/\mathbb{Q}) \in \mathbb{Z}$. Another little less trivial result is how to deal with towers of extensions $M/L/k$:

Proposition 22.1.4: Trace And Norm w.r.t. Towers

Let $M/L/k$ be algebraic 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 the embeddings of L in $\mathbb{C}$ which fix K pointwise and let $\tau_1, \tau_2, ..., \tau_m$ be the embddings in $\mathbb{C}$ fixing L pointwise. We will want to compsoe σ 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; rebael our extensiosn 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 embddings 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).

Example 22.4: Finding integral Elements

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, $n \neq m^2$ 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 $b = 1/2$. Simiarly, 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 $n \equiv 1 \mod 4$.

A useful result that we shall use later is the following:

Proposition 22.1.5: Norm And Units

Let $M/\mathbb{Q}$ and $\mathrm{Nm}_{M/\mathbb{Q}}$. Then the units in $\mathcal{O}_M$ are all elements where $\mathrm{Nm}_{M/\mathbb{Q}}(\alpha)$ map to a unit

Proof **:**

Let $\alpha \in \mathcal{O}_M$ so that $\mathrm{Nm}_{M/\mathbb{Q}}(\alpha) \in \mathbb{Z}$. If α is a unit in $\mathcal{O}_M$, then $1/\alpha \in \mathcal{O}_M$ exists. Then:

$$1 = \mathrm{Nm}_{M/k}(1) = \mathrm{Nm}_{M/k}(\alpha \cdot 1/\alpha) = \mathrm{Nm}_{M/k}(\alpha)\,\mathrm{Nm}_{M/k}(1/\alpha)$$

Since $\alpha, 1/\alpha$ is integral over $\mathbb{Z}$, $\mathrm{Nm}_{M/k}(\alpha) \in \mathbb{Z}$, hence it must be that both map to ± 1, which are the units of $\mathbb{Z}$, completing the proof.

22.1.2 Discriminant

Recall in section 14.3.2 that multi-linear forms have matrix representations and so the determinant of multi-linear maps is well-defined:

Definition 22.1.5: [General] Discriminant

Let $\varphi : V \times V \to k$ be a symmyetric k-bilinear map ($\varphi(x, y) = \varphi(y, x)$). Then given a basis $v_1, v_2, ..., v_n$, the *discriminant* of φ is:

$$\operatorname{disc}(\varphi) = \det \begin{pmatrix} \varphi(v_1, v_1) & \varphi(v_1, v_2) & \cdots & \varphi(v_1, v_n) \\ \varphi(v_2, v_1) & \varphi(v_2, v_2) & \cdots & \varphi(v_2, v_n) \\ \vdots & \ddots & \vdots & \\ \varphi(v_n, v_1) & \varphi(v_n, v_2) & \cdots & \varphi(v_n, v_n) \end{pmatrix}$$

and is denoted $\operatorname{disc}(\varphi)$.

Let $\{e_1, e_2, ..., e_n\}$ be a basis for V and $\{v_1, v_2, ..., v_n\}$ be some collection of elements so that

$$v_i = \sum_i a_i e_i$$

Then:

$$\varphi(v_k, v_l) = \sum_{i,j} \varphi(a_{ki} e_i, a_{lj} e_j) = a_{ki} \varphi(e_i, e_j) a_{kl}$$

hence, as matrices, we have the (component-wise) equality:

$$((v_k, v_l)) = A \cdot (\varphi(e_i, e_j)) A^{\mathrm{tr}}$$

where $A = (a_{ij})$. Then we have the determinant of the matrix:

$$\det((\varphi(f_i, f_j)) = \det(A)^2 \cdot \det((\varphi(e_i, e_j))) \tag{22.1}$$

Showing the discriminant is defined up to the square of a unit. Hence, if $\operatorname{disc}(\varphi) \neq 0$, $\operatorname{disc}(\varphi)$ is well-defined in $k^\times/(k^\times)^2$. Since we are working with bilinear forms, we need to upgrade the notion of non-generate form:

Definition 22.1.6: Non-Generate Form

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: $\{v \in V \mid \varphi(x, v) = 0, \ \forall x \in V\}$

if φ is non-degenerate, it always induces an isomorphism $V \mapsto V^\vee$, $v \mapsto (x \mapsto \varphi(v, x))$. The simplest such map is the following

Definition 22.1.7: Trace Pairing

Let F/k be a finite field extension. Consider:

$$\mathrm{tr}_{F/k}(\alpha\beta) : F \times F \to k \qquad (\alpha_1, \beta) \mapsto \mathrm{tr}_{F/k}(\alpha\beta)$$

Then $\mathrm{tr}_{F/k}(\alpha\beta)$ is called the *trace-pairing*.

Note that the trace pairing is non-degenerate. We see in definition **??** 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.

Definition 22.1.8: 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:

$$\mathrm{disc}_{L/k}(\alpha_1, \alpha_2, ..., \alpha_n) = \det((\sigma_i(\alpha_j))_{ij})^2$$

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 22.1, 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 proerly 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 14.3.9).

Proposition 22.1.6: Discriminant Via Trace-Pairing

$$\mathrm{disc}_{L/k}(\alpha_1, \alpha_2, ..., \alpha_n) = \det(\mathrm{tr}_{L/k}(\alpha_i\alpha_j))$$

Proof **:**
This follows immediately from the matrix equation. if $[\sigma_i(\alpha_j)]$ represents the entry of the matrix, then

$$[\sigma_j(\alpha_i)][\sigma_i(\alpha_j)] = [\sigma_1(\alpha_i\alpha_j) = \cdots + \sigma_n(\alpha_i\alpha_j)] = [\mathrm{tr}_{L/k}(\alpha_i\alpha_j)]$$

Using the fact that the determinant is multiplicative and invariant under transposes, we get our result.

The discriminant gives us lots of information, we shall list out a few notable ones (some pertinent for Galois groups were given in section 19.3).

Corollary 22.1.4: Detecting Algebraic Integers using Discriminant

Let K be a number field. Then $\mathrm{disc}_{K/\mathbb{Q}}(\alpha_1, ..., \alpha_n) \in \mathbb{Q}$, and if all $\alpha_1, \alpha_2, ..., \alpha_n$ are algebraic integers, then $\mathrm{disc}(\alpha_1, \alpha_2, ..., \alpha_n) \in \mathbb{Z}$

Proof **:**
we know that $\mathrm{tr}_{K/\mathbb{Q}}(\alpha_i\alpha_j) \in \mathbb{Q}$ by proposition 22.1.3 and are integers via corollary 22.1.3, hence $\mathrm{disc}_{K/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n) \in \mathbb{Q}$, and if they are all algebraic integers then $\mathrm{disc}_{K/\mathbb{Q}}(\alpha_1 n\alpha_2. \cdots .\alpha_n) \in \mathbb{Z}$.

Corollary 22.1.5: Discriminant And Linear Independence

$\mathrm{disc}_{K/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n) = 0$ 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 $\mathrm{disc}_{K/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n) = 0$. Conversely, assuming $\mathrm{disc}_{K/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n) = 0$, 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 froms 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 following corollary will be important when we prove that $\mathcal{O}_k$ is a finitely generated abelian group. Note that in this case we see that $\operatorname{disc}(\mathcal{O}_K/\mathbb{Z})$ is well-defined for any basis since if we change $\mathbb{Z}$-basis through A, then $\det(A) \in \{-1, 1\}$, and so $\det(A)^2 = 1$.

Corollary 22.1.6: Discriminant And Comparing Basis

Let $B/\mathbb{Z}$ be a finitely generated free abelian group of rank n. Then $N \subseteq B$ is a $\mathbb{Z}$-submodule of finite (i.e. $[B : N] < \infty$) if $\operatorname{disc}_{B/\mathbb{Z}}(\gamma_1, \gamma_2, ..., \gamma_n) \neq 0$. Then:

$$\operatorname{disc}(N) = [B : N]^2 \operatorname{disc}(B)$$

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 22.2), which complete the proof.

As a consequence of this result, if $\operatorname{disc}(N) = \operatorname{disc}(B)$, then a basis for $[B : N] = 1$, that is $B = N$! Hence, we may use discriminants to identify finitely generated free abelian groups!

Another important interpretation of the discriminant is the following:

Proposition 22.1.7: Discriminant via Difference of Products

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

$$\Delta = \prod_{1 \leq i < j \leq n} (\alpha_i - \alpha_j)$$

***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:

$$\begin{pmatrix} 1 & \mathrm{tr}(\alpha) & \mathrm{tr}(\alpha^2) & \cdots & \mathrm{tr}(\alpha^{n-1}) \\ \mathrm{tr}(\alpha) & \mathrm{tr}(\alpha^2) & \mathrm{tr}(\alpha^3) & \cdots & \mathrm{tr}(\alpha^n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}(\alpha^{n-1}) & \mathrm{tr}(\alpha^n) & \mathrm{tr}(\alpha^{n+1}) & \cdots & \mathrm{tr}(\alpha^{2n-2}) \end{pmatrix}$$

which 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:

$$\det(\mathrm{tr}_{B/A}(\alpha_i\alpha_j)) = \prod_{1 \leq i < j \leq n} (x_i - x_j)$$

as we sought to show.

Example 22.5: Discriminant On Simple Extensions

1. The term discriminant might be familiar from solving quadratic equation where we had $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$; the above proposition link the two. Namely if $f(x) = ax^2 + bx + c$, then its roots are

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

   Then the square of the difference of the roots gives $b^2 - 4ac$.
2. Marcus p.19-20

The discriminant gives us another way of detecting the separability of an extension

Proposition 22.1.8: Discriminant And Separability

Let F/k be a finite extension. Then $\mathrm{disc}_{F/k}$ is a nonzero function if and only if F is separable

***Proof* :**
By proposition 22.1.7, 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.

Another powerful use is to check whether two number fields are isomorphic

Proposition 22.1.9: 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

Note that it is possible that there are multiple fields with the same discriminant, however it is a result that we shall not show that there are only finitely many fields for each discriminant value.

Proof **:**
here

A counter-example to the converse would be $\mathbb{Q}(\alpha), \mathbb{Q}(\beta), \mathbb{Q}(\gamma)$ 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. One final useful application that will be given is how the discriminant can be used to find the ring of integers

Example 22.6: Discriminant And Ring Of Integers
see this link

22.1.3 Structure of Ring of Integers

We shall first show that $\mathcal{O}_K$ is a $\mathbb{Z}$-module,

Proposition 22.1.10: Finite Generation of $\mathcal{O}_k$

Let R be a normal integral domain, $(\overline{R}^{K(R)} = R)$,$K = K(R)$ the field of fraction of R, and let $B = \overline{R}^F$ where F/K is a separable extension of deree n. Then there exists free A-submodules M' of F such that

$$M \subseteq B \subseteq M'$$

Hence, B 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.

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). To see this result more concretly, it is worth stating the following immediate corollary:

Corollary 22.1.7: $\mathcal{O}_K$ Is Free $\mathbb{Z}$-Module

Let $F/\mathbb{Q}$ be a finite extension of degree n. Then $\mathcal{O}_F$ is a free $\mathbb{Z}$-module of degree n.

Proof **:**
Of proposition 22.1.10
Let $K = K(R)$ and let $\{\beta_1, \beta_2, ..., \beta_n\}$ be a basis for F over K. We shall use this basis to squish B between two finitely generated R-modules (note that B is an R-module since B/R is an interal extension)

For the first R-module, as we saw in the proof of proposition 22.1.1, there exists a nonzero $d \in R$ such

that $d\beta_i \in B$. Clearly $\{d\beta_1, d\beta_2, ..., d\beta_n\}$ is still a K-basis for F. Now, consider $\langle d\beta_1, d\beta_2, ..., d\beta_n\rangle_R = M$. Since B is an R-module and $d\beta_i \in B$, we have that $M \subseteq R$.

For the second R-module, define:

$$M^\vee = \left\{b \in F \;\middle|\; \forall \ell \in M, \mathrm{tr}_{F/K}(b, \ell) \in \mathbb{Z}\right\}$$

Since the trace-pairing is nondegernate,there exists a "dual" K-basis $\beta_1', ..., \beta_m'\}$ of M given by $\mathrm{tr}(\beta_i\beta_j) = \delta_{ij}$ (note that this is a basis of M, not $M^\vee$).

They key result we shall prove is:

$$M = R\beta_1 + R\beta_1 + \cdots + R\beta_n \subseteq B \subseteq R\beta_1' + R\beta_2' + \cdots + R\beta_m' = M^\vee \tag{22.2}$$

The first inclusion is certainly immediate since each $\beta_i \in B$ and $R \subseteq B$, hence we show the second inclusion. Let $b \in B$ be arbitrary. Then 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 $\beta_i, b \in B$, $b\beta_i \in B$, hence $\mathrm{tr}(b\beta_i) \in R$. But:

$$\mathrm{tr}(b\beta_i) = \mathrm{tr}(\sum_j b_j\beta_j'\beta_i) = \sum_j b_j\,\mathrm{tr}(\beta_j'\beta_i) = \sum_j b_{j\,ij} = b_i$$

Hence, $b_i \in R$, showing equation (22.2). If R is Noetherian, then M' is a Noetherian R-module, and hence B 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 B 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 equl to n, an so it has rank n as an R-module, completing the proof.

Note that separability is a necessary hypothesis to insure the non-degeneracy of the trace-pairing, hence when working over number fields this shall always be the case. Naturally, if R is not a Principal Ideal Domain, then $\mathcal{O}_F$ need not be a free $\mathcal{O}_K$-module. For now, we shall always let $R = \mathbb{Z}$ so that $K = \mathbb{Q}$ nd $F/\mathbb{Q}$ is a number field.

Definition 22.1.9: Integral Basis

Let K be a number field. Then a $\mathbb{Z}$-basis for $\mathcal{O}_K$ is called an *integral basis* for K.

Corollary 22.1.8: Maximality Of Ring Of Integers for Number Fields

The ring of integers of a number field L is the largest subringthat 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 alos a finitely generated $\mathbb{Z}$-submodule, then every eleemnt of B is integral over $\mathbb{Z}$ (proposition 20.1.1(3)). hence, $B \subseteq \mathcal{O}_L$.

Proposition 22.1.11: 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 $(d \big| m_i^2)$

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 22.1.1).

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 embeddigns 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, Hence, for any $\mathcal{O}_K$, if $\alpha_1, \alpha_2, ..., \alpha_n$ is a basis for $K/\mathbb{Q}$

$$\bigoplus \alpha_i\mathbb{Z} \subseteq \mathcal{O}_K \subseteq \bigoplus \frac{\alpha_i}{d}\mathbb{Z}$$

Example 22.7: 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:

$$\text{tr}(b \cdot 1) = 2r \qquad (b\sqrt{5}) = 10s$$

Hence, if $2r \in \mathbb{Z}$ and $10s \in \mathbb{Z}$, we get:

$$L^\vee = \langle \frac{1}{2}, \frac{\sqrt{5}}{10} \rangle_{\mathbb{Z}}$$

Note that $L^\vee$ is biger 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 $\mathcal{O}_K^\vee \neq \mathcal{O}_K$ (TBD)

The previous proposition also gives more motivation for squaring the determinant in the definition of the discriminant. By corollary 22.1.4, if $(\alpha_1, \alpha_2, ..., \alpha_n)$ is a $\mathbb{Z}$-basis for $\mathcal{O}_F$, then $\text{disc}_{F/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n) \in \mathbb{Z}$ 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, hence we can define:

Definition 22.1.10: Discriminant Of Ring Of Integers

Let K be a number field and $\mathcal{O}_K$ the associated ring of integers. Then for any $\mathbb{Z}$-basis $\alpha_1, \alpha_2, ..., \alpha_n$ of $\mathcal{O}_K$

$$\text{disc}(\mathcal{O}_K) = \text{disc}_{K/\mathbb{Q}}(\alpha_1, \alpha_2, ..., \alpha_n)$$

An important word of warning must be given about free $\mathbb{Z}$-modules. It is possible that A, B are two free $\mathbb{Z}$-module of rank 2, 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 $B \subseteq A$[1].

Example 22.8: Strange 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 $\langle a_1, a_1 \rangle_{\mathbb{Q}}$ 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 the next section.

22.2 Archimedean embeddings of Ring of Integers

This is build-up to define volume forms soon and to better understand what type of $\mathbb{Z}$-modules the rings of integers are. Let $n = [K : \mathbb{Q}]$ so that $\text{Hom}_{\mathbb{Q}}(K, \mathbb{C}) = n$. Let $n = r_1 + 2r_2$ 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 embeded into all $\mathbb{C}$ not strictly in $\mathbb{C}$. Note that each embedding into $\mathbb{C}$ comes in complex-conjugate pair, justifying the $2r_2$:

$$\tau_1, \tau_2, ..., \tau_{r_2}, \overline{\tau_1}, ..., \overline{\tau_{r_2}}$$

Definition 22.2.1: Archimedean Embedding

Let $K/\mathbb{Q}$ be a finite fieldextension. Then its *Archimedean embedding* is defined as

$$\begin{aligned} \sigma : & K \hookrightarrow \mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2} \\ & \alpha \mapsto (\sigma_1(\alpha), ..., \tau_{r_2}(\alpha)) \end{aligned}$$

For future purposes, we shall require one more definition:

[1] $b \in B$ can be 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

Definition 22.2.2: [Integral] Lattice

Let V be a finite dimensional vector space over $\mathbb{Q}$ or $\mathbb{R}$. 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
2. Λ is discrete in the topological space V (there is an open ball near the origin that intersects only the origina)
3. Λ is a free $\mathbb{Z}$-module on n generators

It requires a bit of work, but any two of these imply the third (see *A guide to Algebraic Number Theory* by Swinnerton-Dyer), p. 120).

Lemma 22.2.1: Equivalence Of Conditions

Any two of these conditions implies the third

Proof **:**
We shall prove this for the real case $\mathbb{R}$.

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, hecne 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

$$0 = \sum Na_i\ell_i = \sum [Na_i]\ell_i{}_{\in L} + \sum \{Na_i\}\ell_i{}_{\in B_\epsilon(0)}$$

where $\{Na_i\}$ denote the fractional part, completing the proof.

[a] see section 0.1.4

The discrete condition need only be checked at the origin due to linearity (think of topological groups). An example of a free $\mathbb{Z}$-module in $\mathbb{R}^2$ that is not discrete is

$$\langle (1,1), (\sqrt{2}, \sqrt{2}) \rangle \subseteq \mathbb{R}^2$$

In particular, we may get as close as we want to $(0,0)$ by adding these two elements.

Proposition 22.2.1: Closure Of Archimedean Embedding

Let $K/\mathbb{Q}$ be a finite field extension and σ the archemedian embedding. Then:

1. $\langle \sigma(K) \rangle_{\mathbb{R}} = \mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2}$
2. $\sigma(\mathcal{O}_k)$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 **:**

1.
2. here

22.3 Further Structure of Ring of Integers

As we saw, if $K = \mathbb{Q}(\alpha)$, it is not always the case that $\mathcal{O}_K = \mathbb{Z}[\alpha]$. Since any finite field extension can is a simple extension by the primitive extension theorem, then the same applies for $\mathbb{Z}[\alpha]$ (use Gauss's lemma if need be). We can be more precise with the structure via the following :

Theorem 22.3.1: 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 polynyomial $f_i \in \mathbb{Z}[x]$ of degree i for $1 \leq i \leq n-1$ such that

$$1, \frac{f_1(\alpha)}{d_1}, \cdots \frac{f_{n-1}(\alpha)}{d_{n-1}}$$

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

22.4 Prime Decomposition in Ring of Integers

Definition 22.4.1: Dedekind Domain

Let R be an integral domain. Then if R is:

1. Noetherean
2. every nonzero prime ideal is maximal
3. R is normal ($\overline{R}^{K(R)} = R$)

Theorem 22.4.1: Ring Of Integers Are Dedekind Domains

Let $\mathcal{O}_k$ be a ring of integers. Then it is a Dedekind domain.

Proof **:**

Part VI

Homological Algebra

In the section on Cohomology, it would be nice as an example to do the cohomology of a galois group. Explain how a cohomology is a "property" a group can have.

(foothold for when I have time to write down my readings on homological algebra)

23

Abelian Categories

(here)

(also more generally useful, for ex. comes up in sheaf theory and algebraic geometry)

(A quick mention on how it would probably be better to reserve the symbol Hom for additive categories while Mor(

Part VII

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[1]

(The reader may want to go back to section 13.3 to review about direct sums of modules)

(Also, this entire link is really cool: from wiki. Also has some cool examples of duality!!!)

[1] see Langland's program

24

Semisimple Rings and Modules

As mentioned in the introduction, representation theory is all about decomposing structures. In part I, we showed in chapter 4, section 4.7, 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 14 that if $R = k$ is a field k, then any k-module is free, and in chapter 15 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?

24.1 Decomposing Modules

Definition 24.1.1: 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 17.2). 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 over a Principal Ideal Domain is semi-indecomposable due to the Decomposition Theorem from chapter 15), 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 24.1.4)[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:

Definition 24.1.2: 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,

[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

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 24.1.1: 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-commuative 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.

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 24.1.1: 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.

[6] In fact, any element must generate it, can you see why?

***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 24.1.2: 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 9.4.4, 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 13.3.4)

Lemma 24.1.3: 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:

$$\operatorname{End}_R(M) \cong_{\mathbf{Ring}} \operatorname{End}_R(S_1^{\oplus n_1}) \oplus \cdots \oplus \operatorname{End}_R(S_k^{\oplus n_k})$$

Furthermore, for each $S_i^{\oplus n_i}$, we have that direct sums "pull out" to become matrices:

$$\operatorname{End}_R(S_i^{\oplus n_i}) \cong_{\mathbf{Ring}} M_{n_i}(\operatorname{End}_R(S)) = M_{n_i}(D_i)$$

where $D = \operatorname{End}_R(S_i)$ and $M_{n_i}(D_i)$ is the matrix ring over the division ring $\operatorname{End}_R(S_i)$.

***Proof* :**
We invoke proposition 13.3.4 to get the result in terms of groups, and then use Schur's lemma and check that the ring operation is compatible:

$$\begin{aligned}\mathrm{End}_R(M) &= \mathrm{Hom}_R\left(\bigoplus_i S_i^{\oplus n_i}, \bigoplus_j S_j^{\oplus n_j}\right)\\ &\cong_{\mathbf{Grp}} \bigoplus_{i,j} \mathrm{Hom}_R\left(\ S_i^{\oplus n_i}, S_j^{\oplus n_j}\right)\\ &\cong_{\mathbf{Grp}} \bigoplus_{i,j} \mathrm{Hom}_R(\ S_i, S_j)^{\oplus n_i n_j}\end{aligned}$$

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 14.2. 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 24.1.4: 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 13.3.1.

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 13.3.1. 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 13.3.1,

$$N = \sum_{i \in I} S_i \cong \bigoplus_{i \in I} S_i$$

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 internal-SumVsSumRmk, $\mathbb{Z}$ 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 fintie 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 24.1.2: 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 24.1.4. 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 24.1.4, we have that $M = N \oplus N'$, and so by proposition 13.3.1

$$M/N' \cong N$$

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 24.1.3: 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 24.1.2 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 24.1.5: 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.

Remark leftNoethArtRemark

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 necesary 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 24.1.1 more precise:

Corollary 24.1.4: 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 24.1.1

1. Let S be a simple module. Show that any nonzero submodule of the injective enveloppe $E(S)$ is indecomposable, but not simple.

[7] $k \times \{0\} \subseteq k \times k$ is an ideal and a field with identity $(1, 0)$

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)

24.2 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 24.1.5 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 9.5 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 24.2.1: 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 $k\langle x, y\rangle$).

Proposition 24.2.1: 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 24.2.1: Yonneda's Lemma for opposite Rings

Let R be a ring and consider the ring $\mathrm{End}_R(R)$ to be the set of all R-module homomorphisms from R (as an R-module) to itself. Then $\varphi : R^{\mathrm{op}} \to \mathrm{End}_R(R)$, $r \mapsto (x \mapsto xr)$ is a ring isomorphism:

$$R^{\mathrm{op}} \cong_{\mathbf{Ring}} \mathrm{End}_R(R)$$

Proof **:**
Let φ_r be the given map. Let's first check that $\varphi_r \in \mathrm{End}_R(R)$. 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 $\varphi_r \in \mathrm{End}_R(R)$. Next, we check that φ is ineed a ring-homomorphism. By the properties of modules, it is easy to see that:

1. $\varphi_1 = \mathrm{id}$, 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 $\varphi \in \mathrm{End}_R(R)$. 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 $R \cong R^{\mathrm{op}}$, showing that in this case $R \cong \mathrm{End}_R(R)$. Using this, we will use lemma 24.2.1 along with lemma 24.1.3 to upgrade lemma 24.1.5:

Theorem 24.2.1: 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 24.1.5, 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 13.3.1). 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 24.2.1 and lemma 24.1.3:

$$\begin{aligned} R^{\mathrm{op}} &\cong_{\mathbf{Ring}} \mathrm{End}_R(R) \\ &\cong_{\mathbf{Ring}} M_{n_1}(D_1) \times \cdots \times M_{n_k}(D_k) \end{aligned}$$

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].

Proposition 24.2.2: 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 $C = D^{\oplus n}$,

$$C = \bigoplus_i De_i$$

[8]This will be an example of R being both a simple R-module and a simple ring

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 24.1.3 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.

Remark divRingIsSimpleRmk

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 24.2.1: 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 24.2.2

$$M_n(D) \cong_{\mathbf{Ring}} D^{\oplus n} \oplus \cdots_{\mathbf{Ring}} \oplus D^{\oplus n}$$

And by corollary 24.1.4,

$$M \cong D^{\oplus n}$$

Corollary 24.2.2: 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 24.2.2: 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 24.1: 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 24.1, 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 24.2.3: 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 24.2.4: 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 24.2.3, $P \cong_{A \times B\text{-Mod}} M \oplus N$ 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 24.2.3: 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 24.2.3, 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 24.2.4, 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 24.2.4: 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 24.2.3 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.

Remark rightNoethArtRemark
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 24.2.2: 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 24.2.4, 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 $D_i^{\oplus n_i} \cong_R (D_i')^{\oplus n_i'}$, $\text{End}_{M_{n_i}(D_i)}(D_i^{\oplus n_i}) \cong_{\textbf{Ring}} \text{End}_{M_{n_i'}(D_i')}((D_i')^{\oplus n_i'})$. By corollary 24.2.2,

$$D_i^{\text{op}} \cong_{\textbf{Ring}} \text{End}_{M_{n_i}(D_i)}(D_i^{\oplus n_i}) \cong_{\textbf{Ring}} \text{End}_{M_{n_i'}(D_i')}((D_i')^{\oplus n_i'}) \cong_{\textbf{Ring}} (D_i')^{\text{op}}$$

which, taking op on both sides, we see that:

$$D_i \cong_{\textbf{Ring}} D_i'$$

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 14.1.1) applies to divions rings (notice that commutativity was not necessary), and so corollary 14.1.2 and theorem **??** give us that $n = m$, completing the proof.

This theorem also shows us that if R is a semi-simple ring, then R is indeed the direct product of simple rings (since $M_{n_i}(D_i)$ is a simple ring), but it will always be a *finite* product of simple rings. Notice too that this theorem just gave use the answe to the classification of all *finite simple rings*: namely they would have ot be of the form $M_n(\mathbb{F}_q)$ (why fields? TBD) 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!

Example 24.2: 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 24.2.5: 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

$$R \cong_{\mathbf{Ring}} M_{n_1}(D_1) \times \cdots \times M_{n_l}(D_l)$$

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 24.2.6: 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 24.2.7: 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.

Exercise 24.2.1

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 $m \in M$

25

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.

25.1 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 25.1.1: 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 $\varphi : G \to \mathrm{GL}_n(k)$

Usually, when we say representation we will mean *linear representation*. If a vector-space V has a represen-

[1] Not sure about where to address infinite groups

tation, 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 25.1.2: 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 24: 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 $a = ae = ag_1$.

3. There is a natural k action on $k[G]$ defined by

$$b \cdot \left(\sum_i a_i g_i\right) = \sum_i (b a_i) g_i$$

[2]See curryingRmk
[3]A word on this really being a natural way?

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 15.2).

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 25.1.3: 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 25.1.4: 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 hvae 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 25.1: 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 19.1.3 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 25.1.1: 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 $V \not\cong_G W$, 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 25.1.5: 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 25.1.3, page 688).

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 25.1.6: 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 24.1.1 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!

Proposition 25.1.1: 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 24.1.4[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 25.2: 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. Shwo that $k[C_p] \cong k[a]/(a^n - 1)$
2. Show that if $\mathrm{char}\, k \nmid |G|$, 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 $\mathrm{char}(k) = p$, 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 24.2.7) 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 25.1.2: Matrix Representation Consequences

Let $\rho : G \to \mathrm{GL}(V)$ be a representation over a field k where $\mathrm{char}(k) \nmid |G|$. 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 25.1.3: 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 25.1.2: 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 24.2.7.

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 25.1.4: 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 24.2.7, $\mathbb{C}[A]$ 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 25.3: Decomposition Of Modules

1. DnF p. 847, and a bit about $\dim(V) = 1$ example right under definition

Exercise 25.1.1

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?)

25.2 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 25.2.1: 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 25.4: 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 25.1. 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 25.2.1: 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 25.2.1: 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}

Remark We can prove this directly without using Artin-Wedderburn, which we will

Corollary 25.2.1: 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 25.2.2: 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 25.5: 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 25.2.3: Premuation Represenation

Let G act on a fintie 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 25.2.2: 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 25.2.4: 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 25.2.3: 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 17.4.1 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 25.2.4: 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 25.2.3 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 25.2.5: 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 25.1.3 if we let $\langle , \cdot, \cdot \rangle$ be the usual conjugate dot product on $\mathbb{C}^n$. (TBD)

Lemma 25.2.5: 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 25.2.2: 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 25.2.3: 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 25.2.6: 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 25.6: 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 25.2.1: 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 25.7: 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 25.2.2: 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 25.1.2:

$$e_i = \sum_{g \in \mathrm{conj}(g)} g \in Z(\mathbb{C}[G])$$

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 25.1.2, 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 20.1.1, $\mathbb{Z} \subseteq Z(\mathbb{Z}[G])$ 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 24.2.6 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:

$$\sum_i^r \frac{\mathrm{conj}(g_i)}{n} \chi_V(g_i) \overline{\chi_V(g_i)}$$

Then we've just shown that $\frac{\mathrm{conj}(g_i)}{n} \chi_V(g_i)$ 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, $\alpha \in \mathrm{Frac}(R)$, 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 25.2.1

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.

25.2.1 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 25.1 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 25.1, 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

25.2.2 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 25.2.6: 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 Bezout ($1 = am + b|\operatorname{conj}(g)|$)

$$\frac{1}{n}\chi_V(g) \in \overline{\mathbb{Z}}^{\mathbb{C}}$$

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 $G = \operatorname{Gal}(N/\mathbb{Q})$ 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 **??**).

Theorem 25.2.2: 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 25.2.6

$$\chi_i(h) = 0 \qquad \text{or} \qquad \chi_i(1)$$

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)

25.3 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 25.3.1: 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 25.2.3 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 25.8: 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 25.3.1: 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 25.3.1: 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 25.3.1: 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 25.3.2: Computing Induced Representation

Let $H \leqslant G$ of a fintie 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 $h = g_i^{-1}gg_j$, this fully describes $\Phi(g)$, as we sought to show.

Corollary 25.3.2: 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:

$$\mathrm{Ind}_H^G(\chi) = \sum_{i=1}^{m} \chi(g_i^{-1}gg_i)$$

where $\chi(g_i^{-1}gg_i) := 0$ if $g_i^{-1}gg_i \notin H$. 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 25.3.2: 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:

$$\begin{tikzcd} & W_1 \\ V \arrow[ur, dash] \arrow[dr, dash] & \\ & W_2 \end{tikzcd}$$

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:

$$\begin{tikzcd} V \arrow[dr, dash] & \\ & W \\ V' \arrow[ur, dash] & \end{tikzcd}$$

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.

25.4 *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)

26

Quiver Theory – Lie Algebra Representation

(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.

26.1 A Quick exploration of Lie Algebras

(I would like to prelude the importance of quiver theory within lie algebras)

26.2 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 26.2.1: 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 26.1: 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 26.2.2: 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 26.2.3: 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:
$$\begin{tikzcd} & \cdot \arrow[d] & \\ \cdot \arrow[r] \arrow[dr] & i & \cdot \arrow[l] \arrow[dl] \\ \cdot \arrow[ur] & \cdot \arrow[u] & \cdot \arrow[ul] \end{tikzcd}$$
3. We say a vertex $i \in V$ is a *source* if all edges containing i are outgoing. Intuitvely, this can be visualized as:
$$\begin{tikzcd} \cdot \arrow[dr] & \cdot \arrow[d] & \cdot \arrow[dl] \\ \cdot \arrow[r] & i \arrow[l] \arrow[r] & \cdot \\ \cdot \arrow[ur] & \cdot \arrow[u] & \cdot \arrow[ul] \end{tikzcd}$$
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.

26.2.1 Quiver Representation and Varieties

Definition 26.2.4: 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 26.2: 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
$$\begin{tikzcd} & 2 \arrow[loop above, "f"] \\ 1 \arrow[r, "e"] & \end{tikzcd}$$
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 26.2.5: 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 26.2.6: 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 26.2.7: 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 26.2.1: 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 26.3: 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:

$$\begin{tikzcd} V \arrow[d, "\varphi"] & \mathbb{F} \arrow[r, "1"] \arrow[d, "1"] & \mathbb{F} \arrow[d, "0"] \\ W & \mathbb{F} \arrow[r, "0"] & \{0\} \end{tikzcd}$$

   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 26.2.8: 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 $\operatorname{im}(\varphi)$ is the subrepresentation of W consisting of $\operatorname{im}(V(i)) \subseteq W_i$

Note that $\ker(\varphi)$ and $\operatorname{im}(\varphi)$ 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 26.2.9: Quiver Isomorphisms

Let $V, W \in \operatorname{Rep}(Q)$. 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 26.4: 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 26.2.10: Quotient Quiver

Let $V \in \operatorname{Rep}(Q)$ 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 26.2.2: 1st Isomorphism for Quivers

Let $V, W \in \operatorname{Rep}(Q)$ and $\varphi : V \to W$ be a quiver morphism. Then:

$$\operatorname{im}(\varphi) = V / \ker(\varphi)$$

Proof **:**
left as an exercise

Definition 26.2.11: Product Quiver

Let $V, W \in \operatorname{Rep}(Q)$. Define $V \oplus W \in \operatorname{Rep}(Q)$ 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 26.2.12: 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 26.2.3: 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 $\mathrm{Rep}(Q, \underline{d})$ 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 26.2.1: Isomorphism Equivalence

Let $\mathrm{Rep}(Q, \underline{d})$ be a quiver variety and $G(\underline{d})$ be the group as defined in proposition 26.2.3. Then two representations $V, W \in \mathrm{Rep}(Q, \underline{d})$ 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:

$$\{G(\underline{d})\text{-orbit of } \mathrm{Rep}(Q, \underline{d})\} \rightleftarrows \{\text{iso classes of rep with dim } \underline{d}\}$$

Proof **:**
As we've seen in example 26.4, 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.

26.2.2 Decomposing Quivers

Similarly to all other algebraic structures we've encountered, we have a notion of simple objects:

Definition 26.2.13: Simple Representation

Let Q be a quiver. Then $V \in \mathrm{Rep}(Q)$ 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 26.2.14: 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 26.5: 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 26.2.2: 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 26.6: 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 15). Using this knowledge, decompose any representation $T : V \to V$.

26.3 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 26.3.1: 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 26.7: 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 26.3.2: 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 26.8: Simple Roots

homework, exercise 8 in Jimails notes

Theorem 26.3.1: 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 26.3.3: 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 26.9: Cartan Matrix
exercise 10

(word to motivate the following visualization)

Definition 26.3.4: 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 26.10: 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 26.3.2: 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 26.3.5: 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 1.2.12)

Example 26.11: Weyl Group
p. 11 Jimail

Definition 26.3.6: 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 26.3.3: 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 26.3.7: 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 26.3.4: 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 26.3.8: 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 26.12: Adapted Dynkin Diagrams
p. 15 Jaimal

Theorem 26.3.5: 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

26.4 Gabriel's Theorem

In this section, we come back to Quiver's. Recall in example 26.6 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 26.4.1: 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 26.13: 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 26.6 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 26.4.1: 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:

$$\left(\overset{V(1)}{\bullet} \xrightarrow{T_e} \overset{V(2)}{\bullet} \right) \mapsto \left(\overset{\mathrm{coker}(T_e)}{\bullet} \xleftarrow{S_f} \overset{V(2)}{\bullet} \right)$$

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 26.4.2: 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 26.14: 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 26.4.1: 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 26.4.3: 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 26.4.2: Root Lattice Map

here exists a map $\dim_Q : \{\text{iso classes of Rep's}\} \to \mathbb{Z}_+^I$

Proof **:**
not proved

Definition 26.4.4: 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 26.15: 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 26.4.3: 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 26.4.4: 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

26.5 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 $I = \text{Ann}_R(M)$).

Let $Q = (V, E)$ be a quiver. For vertices i, j of Q, Define the set $\text{Path}(i, j) = \text{span}\{\text{paths } \gamma \text{ from } i \text{ to } j\}$ (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 26.5.1: 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

$$S_e(\varphi^v_a(\gamma)) = S_e S_\gamma(v) = S_{e\gamma}(v) = \varphi^v_b(e\gamma) = \varphi^v_b(T_e(\gamma))$$

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 26.5.1: Path Algebra

Let $Q = (V, E)$ be a quiver. Then the *path algebra* $P(Q)$ is defined as follows:

1. As a vector-space $P(Q) = \mathrm{span}\{\text{paths in } Q\}$, 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 VIII

Category Theory

(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!

27

An Introduction to Category Theory

In this chapter, we introduce the basic terminology as well as a plethora of examples, and then introduce some natural categorical concepts such as *duality*, *functoriality*, and *natural transformations*. Since many of these concepts were already introduced at many places in the books (for example, chapter 2, and section ref:HERE (preliminaries to hom-alg)), these will be covered at "high-speed".

27.1 Basic Definitions and Examples

Definition 27.1.1: 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 27.1: 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 27.2: 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 27.3: : 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 eample). 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 interseting in one more generelization where the morphisms of this category become the objects, and the morphism being a morphism between functor!

As noted in a footnote of the definion, 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:

Russell's Paradox Fix S to be the set of all sets (it conatins $\{\}$, $\mathbb{N}$, $\{\{\mathbb{R}\}\}$, 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 27.1.2: 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 27.1.3: 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 27.4: 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 27.1.4: 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 27.5: 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 27.1.5: Groupoid

A **groupoid** is a category in which every morphism is an isomrophism

Example 27.6: 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 1.1.2. 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 27.1.6: Subcategory

A subcategory D of a category C is defined by restricting to a subcollection of objects and subcollection of morphisms subject to the requrieemetns 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 27.7: 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 27.1.1: 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

27.2 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 27.2.1: 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 27.8: 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 27.2.1: 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 27.2.2: 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 27.2.1 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 27.9: : 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.

27.3 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 27.3.1: 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 27.10: 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-dimensinoal 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 27.3.2: 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 27.11: Covariant Functor

1. The functor $(___)^* : \mathbf{Vect}_k^{\text{op}} \to \mathbf{Vect}_k$ carries a vectors space V to its dual $\text{Hom}_k(V, k)$ 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, $X : \mathbf{G}^{\text{op}} \to C$ defines a *right action*.
3. The contravariant functor $P : \mathbf{Set}^{\text{op}} \to \mathbf{Set}$ 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 $\mathcal{O} : \text{Top}^{\text{op}} \to \mathbf{Poset}$ 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 $\text{Spec} : \mathbf{CRing} \to \mathbf{Top}$ sends a commutative ring R to $\text{Spec}(R)$ and $\varphi : R \to S$ to $\varphi^{-1} : \text{Spec}(S) \to \text{Spec}(R)$ 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 27.3.1: 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 27.3.3: 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 27.2.1 (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 27.3.3 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 27.3.4: 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 27.1.4, 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 27.12: 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

27.4 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 14.3:

Example 27.13: 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 27.4.1: 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:

$$\begin{tikzcd} C \arrow[rr, bend left, "F"] \arrow[rr, bend right, "G"'] & \Downarrow \alpha & D \end{tikzcd}$$

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

$$\begin{tikzcd} \mathbf{Vect}_{\mathbb{F}} \arrow[r, bend left, "\mathbf{Id}"] \arrow[r, bend right, "(_)^{**}"'] \arrow[r, phantom, "\Downarrow \alpha"] & \mathbf{Vect}_{\mathbb{F}} \end{tikzcd}$$

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 27.14: 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:

$$\begin{tikzcd} (a, b) \arrow[r, dotted, no head] \arrow[d, mapsto] & & (a, b) \arrow[r, mapsto] \arrow[d, dotted, no head] & (\varphi(a), \varphi(b)) \arrow[d, mapsto] \\ a \cdot b \arrow[r, mapsto] & \varphi(a \cdot b) & & \varphi(a) \cdot \varphi(b) \arrow[l, dotted, no head] \end{tikzcd}$$

   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

27.5 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 thier 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 wich 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 27.15: Universal Properties

1. Let $\mathbb{1}$ denote a set with one element. Then this set with 1 elemnt has the following property

   for all sets X, there exists a *unique* map X to 1

2. Let R be a ring with identity $1 \neq 0$. Then

   For all rings R, there exists a *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 27.5.1: 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 27.16: 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

27.6 Adjoint

here

For free objects, adjoint with some appropriate forgetful functor.

Adjoint of Hom and $\otimes$.

linear algebra adjoints?

27.7 A Second Look at Category theory

The concept of Categorization.

As an example, let's "categorify" the concept of groups:

Definition 27.7.1: 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 extensiosn 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.

27.8 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, 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

[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

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 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

[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

[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

than ZF would already cause. Though 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.1.1: 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 6. 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: 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 be a group

(words here)

Definition A.1.2: 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.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.1.3: 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.1.1: 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 defiend 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.1.1. 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 0.1.3). 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 0.1.5

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 appendix D (maybe a geometric group theory thing or Gromov's theorem appendix?) or check out _____'s *How Groups Grow*.

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.1.1: Nielsen-Schreier's Theorem

Let G be a free group and $H \leqslant G$. Then H is a free gorup

***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] recall that $F_2' \leqslant F_2$, the commutator subgroup of the free group on 2 elements, is not finitely genrated (see ref:HERE)

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 $a_j = ux_k^{-1}$ and u $u \in Ha_j$. But then, $u = a_j x_k$, meaning that u is the representative of $Ha_j x_k$, that is $u = a_l$. Thus, we get

   $$a_j x_k a_l^{-1} = ux_k^{-1} x_k a_l^{-1} = a_l x_k^{-1} x_k a_l^{-1} = 1$$

   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 alx_k^{-1}. Then like before, x_k^{-1} is also a representative of its coset. BUt looking at $y_{j,k}^{-1} = alx_k^{-1} a_j^{-1}$, we see that a_j represetns 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 ommit 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.1.1

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.1.2: 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 $U \in \mathbb{C}\text{P}$ 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 defiend 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.1.3: 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.1.4: 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.1.5: 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 beign 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.1.2: 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).

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.1.6: 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 the ook:

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)

(closure, dense sets, interior)

(define it, show that if $A \subseteq X$ is closed, then A is compact)

(might need to define sequence's if I will use it to give an intuition to completion of groups through through analogy to completing topological groups through means of Cauchy sequenced)

A final tool that is very common in topology is that of the *sequence*

Definition C.1.7: 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.

Sequence are become intersting 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.1.8: 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

$$x_n \in U \qquad \forall n \geq S(U)$$

If X is a metric space, we can define the notion of a cauchy sequence:

Definition C.1.9: 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.1.10: Complete Metric Space

Let X be a metric space. Then X is compelte if all cauchy sequences are convergent sequences

Example C.1: Complete Spaces
$\mathbb{R}$ is a complete space, while $\mathbb{Q}$ is not

The notion of completeness given in section 20.3 is in fact a generalization of the above notion, as we'll show in section C.4.

C.2 Construction Topologies

(product topology. How the product respects the universal property of the product. Is the quotient needed?)

The subspace, product, and quotient topology are all special examples of the following two consructions:

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: 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[1], 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).

[1] See definition ref:HERE

Example C.3: 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 generlized even further. If you have read VIII, both the weak and strong topologies are simple examples of *inverse limits* and *direct limits* 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

(Would it be fruitful to give that alternate characteriztion of compactness)

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 fintie subcover, then X is *compact*.

(Give definition for subsets of a space)

Proposition C.3.1: Equivalent Definition Of Compatness

1. If every continuous function $f : M \to \mathbb{R}$ is bounded, then M is compact
2. If every continuous bounded function $f : M \to \mathbb{R}$ achieves a maximum or minimum, thne M is compact
3. If every continuous function $f : M \to \mathbb{R}$ has compact rang $f(M)$ then M is compact
4. If every nested decreasing sequence of nonempty closed subsets of M has nonempty intersection then M is compact

Proof **:**
exercise

C.4 Topological Groups

In this book, topological groups show up when talking about the completion of rings in section 20.3 and when talking about the representation of non-finite groups in chapter 25. The topological considerations in both of those cases were purposefully ommited 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 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: 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 *toplogical 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 $O(n) = \left\{A \in \mathrm{GL}_n(\mathbb{R}) \;\middle|\; AA^t = I\right\}$. 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 Geomtery) 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 20.3, then the p-addic integers, $\mathbb{Z}_p$ has the induced inverse-limit topology (in this case, the p-addic 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-addic metric, we can take a different completion to form $\mathbb{Q}_p$, which is also a Lie Group.
10. In commutative algebra, it is common to define a topology on a group in the following way: Let G be a commutative group and take a descending sequence of subgroups[a]:
$$G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_n \supseteq \cdots$$
Then let the collection of (G_n) define the *local neighbhorhood* of 0, and $x + G_n$ be an open neighbhorhood of x. Then this generate's a topology, which we'll call the *subgroup topology*. This topology is particularly important when we talk about the *completion* of groups; more on that at the end of this section.

[a] If G were not commutative, we would take a descending sequence of normal subgroups

Proposition C.4.1: 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.4.2: 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.5: Topological Isomorphism
Naturally, the identity map is a topologica isomorphism. If the topology is discrete, every homomorphism/isomrophism is a topological homomorphism/isomorphism.

It is possible that a homomorphism/continuous map is not a continuous map/homomrophism. As an example of a contiuous, even homeomorphic map, that is not a homomrophism 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 intersting 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 neighbourhood of 1, then gU is a neighbourhood of g, meaning that the topology of G is determined by the neighbourhoods of 1!

As an exercise, it is worth seeing how if $H \leqslant G$, then G/H has the quotient topology, is a homogenous 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 leats 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-addic 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.4.2

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 Haudorff
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 neighbourhood). 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*.

The last subject of topological groups we will explore here is the *completion* of topological groups, since it generlizes some of the results of section 20.3 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.4.3: 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 neighbourhood 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 colleciton of all equivalent cauchy sequence is the completion of of the group G, which we'll dentoe 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 Hausdroff. Completion is functorial: Since continuous homomorphisms preserve cauchy sequences, $f : G \to H$ maps to $\hat{f} : \hat{G} \to \hat{H}$.

This construciton is quite general: we can allow for $G = \mathbb{Q}$ and then $\hat{G} = \mathbb{R}$. We can also allow for $G = \mathbb{Q}$, and define a different metrc (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 the topology at the end of example C.4 given by subgroups (i.e. the subgroup topology). In this topology, $U \subseteq G$ is a neighbourhood 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 based completely on the subgroups of G, we will define the equivalent of a cauchy sequence in completely algebraic terms, which will call a *coherent sequence*. 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}. When considering the relation between these two, it is natural to see that the natural projeciton map maps 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 projectino maps, howver we may imagine a simlar 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 let 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 20.3 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 *inverse limit*, and is denoted $\varprojlim G_n$, where the homomorphissm f_n are usually implicitely 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-addic 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 any topological group defined via the I-addic topology which will generate the I-addic 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.1.10, which we can see is the same as $\varprojlim G/G_n$. Overall, the inverse limits of rings in terms of ideal powers has a very convenient topology in which we can combine the notion of completion with the usual notion of completion.

Exercise C.4.1

1. (Relate this to DVR topologies?)

D

Elements of Geometric Group Theory

(Explain how the study of infinite groups starts "analytical" ideas, leading us to defining some new terms and studying groups from a different perspective (I'm thinking of Sofic groups))

1. Explain basics with using Cayley tree's (and other ay's of definign geometric groups theory)
2. Growth of free groups, then define other types of groth (mention unsolved problems)
3. Defining quasi-isometries
4. Gromov's Theorem?
5. Tits Alternative?
6. Application to other fields?
7. Burnside's group?
8. perhaps ends with paradoxical groups and their link to the Banach-Tarski paradox (in Clara löh's book under amenable groups!)

(also give proper textbooks which will continue this discussion in further detail. Do the same actually for every section in the appendix)

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).

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

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

etc.

free rings

G.3 Modules

free modules

G.4 Universal Properties

(list all universal properties covered in the book)

G.5 Character Tables