Introduction to Abstract Algebra

Timothy J. Ford

Department of Mathematics, Florida Atlantic University, Boca Raton, FL 33431

Email address: ford@fau.edu

URL: https://tim4datfau.github.io/Timothy-Ford-at-FAU/

Contents

Preface

The purpose of this book is to provide an introduction to the theory of abstract algebra in an efficient concise no-nonsense manner. The goal is to lay a solid foundation for future study of algebraic topics. It is intended to be accessible to first year graduate students and advanced undergraduate students in mathematics. Chapters two, three, four, five and six provide a solid introduction to group theory, ring theory, linear algebra and fields. A typical two-semester sequence on Abstract Algebra at the introductory level would cover much of the material in these five core chapters. Chapter one, a background chapter, contains much of the conventions concerning notation and terminology as well as a review of the material from set theory, elementary number theory, calculus, and matrix theory necessary for reading the rest of the book. Chapter seven is an additional chapter that applies the main results of the five core chapters to a topic normally seen in a more advanced algebra book, namely the group of ideal classes of a Dedekind domain.

Algebra is one of the fundamental areas of mathematics. Like most of modern mathematics, it is no exaggeration to say that Algebra is very abstract. The many abstract structures and constructions that exist in Algebra can be difficult to grasp upon first encounter. For this reason, it is sometimes helpful to have a "handle" to lend support. In its essence, Algebra is the study of polynomial equations. While not intending to be an oversimplification of the matter, keeping this in mind can be of help to a student trying to make sense of the many abstract notions that arise. For instance, Number Theory can be considered as that subset of Algebra that is concerned with polynomial equations for which the coefficients involve only natural numbers. Likewise, the origins of Group Theory lie in the study of solutions to polynomial equations in one variable. It was Galois who stressed the importance of looking at the permutations of the set of roots of a polynomial in one indeterminate. This led to what is now called Galois Theory, as well as to the notion of a group acting on a set, hence to what is now called Group Theory. The set of solutions to a system of polynomials in several variables is called an algebraic variety. Algebraic Geometry arose as the study of algebraic varieties. Linear Algebra is the study of systems of linear equations. Arising out of this study are what we now call vector spaces, and more generally, modules. Matrices turn out to have both practical and theoretical importance in Linear Algebra. Ring Theory can be thought of as the natural abstraction of the addition and multiplication operations possessed by the set of square matrices. Commutative Algebra naturally developed out of the study of properties of the ring of polynomial functions on an algebraic variety.

Chapter 1 includes a review of much of the background material. It serves as a reference for the rest of the book. This includes material that is ordinarily covered in the standard undergraduate courses on Discrete Mathematics, Introductory Number Theory, and Introductory Linear Algebra.

Chapter 2 is an introduction to Group Theory. There are many examples of finite groups. The standard counting theorems for cosets and factor groups, including Lagrange's Theorem are proved. The Isomorphism Theorems which are proved here for groups are referenced by the later chapters when the counterparts for rings and modules arise. There is a section devoted to group actions, conjugacy classes, the Class Equation, and semidirect products. Arbitrary direct products, both external and internal are defined. The important properties of permutations on finite sets and the symmetric groups are derived. Included are proofs of Cauchy's

Theorem, the Sylow Theorems, the Basis Theorem for a Finite Abelian Group, an introduction to the classification problem for finite groups, and an introduction to solvable groups.

Chapter 3 is an introduction to Ring Theory. Topics include ideals, factor rings, the Isomorphism Theorems, direct sums and products, the Chinese Remainder Theorem, factorization in commutative rings, euclidean domains, principal ideal domains, unique factorization domains, the quotient field of an integral domain, polynomial rings over a commutative ring, and a proof that a polynomial ring over a unique factorization domain is a unique factorization domain. We prove that a principal ideal domain satisfies the ascending chain condition on ideals and is a unique factorization domain. These results are used in Section 4.6 when we study finitely generated modules over a principal ideal domain. A euclidean domain is a principal ideal domain, hence is a unique factorization domain. Over an arbitrary euclidean domain, the Extended Euclidean Algorithm exists for computing a greatest common divisor $d = \gcd(a, b)$ and for solving the Bézout Identity $d = ax + by$.

Chapter 4 is an introduction to modules, vector spaces, algebras, and matrices. We define a module over an arbitrary ring, and a vector space over a division ring. Whenever possible, theorems on modules are treated as extensions of their counterparts for groups. Submodules, quotient modules, and the Isomorphism Theorems are covered. Direct products and direct sums are defined for an arbitrary family of modules. The fundamental properties of free modules are proved, including the existence of a basis. This then leads to an introduction to properties of vector spaces. In this chapter we also define an algebra over a commutative ring. In general, algebras are noncommutative. In an algebra over a field, we show that an algebraic element has a unique minimal polynomial. Included is an introduction to Matrix Theory over an arbitrary ring. Given finitely generated free R-modules M and N and given bases for M and N, a homomorphism from M to N has a matrix representation. The set of all n-by-m matrices over a ring R is a free module. Matrix multiplication is consistent with composition of homomorphisms. The set of all n-by-n matrices is itself a ring. Changing the basis of M corresponds to a similarity transformation of the matrix associated to an endomorphism of M. In this chapter we prove the Invariant Factor Form and the Elementary Divisor Form of the Basis Theorem for Finitely Generated Modules. Rather than state and prove these theorems for finitely generated $\mathbb{Z}$-modules or for modules over a euclidean domain, we instead prove them over an arbitrary Principal Ideal Domain. This permits us in Chapter 7 to prove the finiteness of the integral closure of a principal ideal domain in a finite separable extension of its quotient field (Corollary 7.1.16). This leads into an existence theorem for Dedekind domains (Theorem 7.3.2) that applies to the integral closure of a principal ideal domain in a finite separable extension of its quotient field.

Chapter 5 is an introduction to fields. This includes algebraic extensions, transcendental extensions, and the questions of antiquity involving straightedge and compass constructions. The existence and uniqueness of a splitting field is proved. The Primitive Element Theorem is proved for a finite separable extension. Our presentation of the Fundamental Theorem of Galois Theory is very close to the traditional approach of [**2**]. The importance of separable extensions is emphasized. There are proofs of the Embedding Theorem, the existence of a separable closure, that normal and separable implies Galois, and that for a field to be perfect it is

necessary and sufficient that every algebraic extension be separable. For a Galois extension there are the norm and trace functions. Hilbert's Theorem 90 is proved for a cyclic extension. Many of the fundamental properties of finite fields, cyclotomic extensions and cyclic extensions are proved. We prove that the field of symmetric rational functions is equal to the field of rational functions in the elementary symmetric polynomials and a symmetric polynomial is a polynomial in the elementary symmetric polynomials. When the ground field has characteristic zero and contains sufficiently many roots of unity, we show that a polynomial is solvable by radicals if and only if the Galois group solvable. This chapter relies heavily on results from Chapters 2, 3 and 4.

Chapter 6 is the study of linear transformations on a finite dimensional vector space over a field. Given a field k, a finite dimensional k-vector space V, and a linear transformation $\phi : V \to V$, the eigenvalues and eigenvectors of ϕ are studied. The eigenvalues are related to the existence of a basis such that the matrix of ϕ is in triangular form. Associated to the linear transformation ϕ is a module structure on V over the principal ideal domain $k[x]$. The Basis Theorems for a Finitely Generated Module over a Principal Ideal Domain are applied to derive the rational canonical form and the Jordan canonical form of ϕ. The existence and uniqueness of a reduced row echelon form is proved. The determinant function is defined as a symmetric multilinear form and the characteristic polynomial is defined as a determinant. This allows us to derive their important properties. This chapter relies heavily on results from Chapters 2 – 5.

Chapter 7 includes an introduction to integral extensions and the integral closure of an integral domain. An introduction to the properties of fractional ideals is given. We define the ideal class group, but for sake of brevity as well as completeness, we restrict our definition to Dedekind domains. A Dedekind domain is an integral domain S with quotient field L such that $S \neq L$, S is integrally closed in L, S has the ascending chain condition on ideals, and every nonzero prime ideal of S is maximal. We show that for a Dedekind domain the set of all fractional ideals is an abelian group. Given a principal ideal domain R with quotient field K and a finite separable extension L/K, the integral closure of R in L is a Dedekind domain. This important existence theorem implies, for example, that the ring of integers in an algebraic number field is a Dedekind domain. The last section is devoted to some applications and computations for two specific examples. The examples we consider are affine algebraic curves of degree two and degree three. Modulo some results we need from algebraic geometry (specifically, the Jacobian Criterion for Regularity and the proof that a nonsingular curve is rational if and only if there are two distinct points that are linearly equivalent), the group law on the cubic is presented in terms of the ideal class group of the affine coordinate ring. This chapter relies heavily on results from Chapters 2 – 6.

This book originated in the class notes that I compiled when I taught the two-semester sequence on Abstract Algebra at Florida Atlantic University for the Fall 2019 – Spring 2020 academic year. At my university, the students who take this course are either advanced undergraduates or first year graduate students. Throughout the course, I personally typeset the lecture notes and made them available for my students. Supplemental exercises were added as well. By the end of the course, I had accumulated most of the material in this document. In the following three year period, the book has since been used as the primary source for the same

algebra course at my university. A number of new topics have been included. Many examples and exercises have been added.

The intention in writing this book is to produce an introductory level abstract algebra book. Compared to other popular books on the subject, the plan is for it to be content-wise on the level of Herstein's [**15**] and Clark's [**6**], and presented in their same no-nonsense style. It deliberately covers fewer topics and is shorter in length than [**4**], and [**8**], for example. It is meant to be more accessible than Hungerford's [**16**], Rotman's [**23**], [**24**], [**25**], or Lang's [**18**].

I take this opportunity to express my gratitude for the positive influence three algebraists have had on my appreciation, attraction and love for this subject. I am grateful to Frank DeMeyer for introducing me to the amazing subject of Abstract Algebra, in the style of [**16**], to James Brewer for suggesting [**15**] for an introductory level textbook on this subject, and to Fred Richman who steered me into the straightforward approach employed by [**6**].

CHAPTER 1

Preliminaries and Prerequisites

Chapter 1 is intended to be used as a reference by the subsequent chapters. We assume the reader is familiar with most of the material. This chapter is not intended to be a substitute for an undergraduate textbook on Discrete Mathematics. Conventions, notation and terminology are established. Without undermining the importance of the subject matter, the goal of Chapter 1 is to efficiently and concisely set the table for the rest of the book. Therefore, a practical, or utilitarian approach is taken.

1. Background Material from Set Theory

Sets are the basic building blocks of abstract mathematics. We begin with sets of numbers, sets of letters, sets of sets, or sets of variables. We combine them, operate on them, compare them. Functions, relations and binary operations are themselves defined as sets.

A rigorous definition of a set is not attempted. Rather, we adopt the naive approach that a set is an abstract collection of objects, or elements. It is important to emphasize that the key property or attribute a set is required to possess is that it is possible to distinguish in an unambiguous way those elements that are in the set from those not in the set.

1.1. Sets and Operations on Sets. A *set* is a collection of objects X with a membership rule such that given any object x it is possible to decide whether x belongs to the set X. If x belongs to X, we say x is an *element* of X and write $x \in X$. Suppose X and Y are sets. If every element of X is also an element of Y, then we say X is a *subset* of Y, or that X is *contained* in Y, and write $X \subseteq Y$. If X and Y are subsets of each other, then we say X and Y are *equal* and write $X = Y$. The set without an element is called the *empty set* and is denoted $\emptyset$. The set of all subsets of X is called the *power set* of X, and is denoted 2^X. Notice that $\emptyset$ and X are both elements of 2^X. The *union* of X and Y, denoted $X \cup Y$, is the set of all elements that are elements of X or Y. The *intersection* of X and Y, denoted $X \cap Y$, is the set of all elements that are elements of X and Y. The *complement* of X with respect to Y, denoted $Y - X$, is the set of all elements of Y that are not elements of X. The *product* of X and Y, denoted $X \times Y$, is the set of all ordered pairs of the form (x, y) where x is an element of X and Y is an element of Y.

Let I be a set and suppose for each $i \in I$ there is a set X_i. Then we say $\{X_i \mid i \in I\}$ is a *family of sets indexed by* I. The *union* of the family is denoted $\bigcup_{i \in I} X_i$ and is defined to be the set of all elements x such that $x \in X_i$ for some $i \in I$. The *intersection* of the family is denoted $\bigcap_{i \in I} X_i$ and is defined to be the set of all elements x such that $x \in X_i$ for all $i \in I$.

The set of *integers* is $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$. The set of *natural numbers* is $\mathbb{N} = \{1, 2, 3, \dots\}$. The set of nonnegative integers is $\mathbb{Z}_{\geq 0} = \{0, 1, 2, 3, 4, \dots\}$. The set of *rational numbers* is $\mathbb{Q} = \{n/d \mid n \in \mathbb{Z}, d \in \mathbb{N}\}$ where it is understood that $n/d = x/y$ if $ny = dx$. The set of *real numbers* is denoted $\mathbb{R}$, the set of *complex numbers* is denoted $\mathbb{C}$.

If $n \in \mathbb{N}$ and $\{X_1, \dots, X_n\}$ is a family of sets indexed by $\{1, 2, \dots, n\}$, then we sometimes write $X_1 \cup \cdots \cup X_n$ instead of $\bigcup_{i=1}^n X_i$, and $X_1 \cap \cdots \cap X_n$ instead of $\bigcap_{i=1}^n X_i$. The *product* of the family, written $X_1 \times \cdots \times X_n$ or $\prod_{i=1}^n X_i$, is the set $\{(x_1, \dots, x_n) \mid x_i \in X_i\}$.

1.2. Relations and Functions. Let X and Y be nonempty sets. A *relation* between X and Y is a nonempty subset R of the product $X \times Y$. Two relations are equal if they are equal as sets. The *domain* of R is the set of all first coordinates of the pairs in R. The *range* of R is the set of all second coordinates of the pairs in R.

A *function* (or *map*) from X to Y is a relation $f \subseteq X \times Y$ such that for each $x \in X$ there is a unique $y \in Y$ such that $(x, y) \in f$. In this case, we say y is the *image* of x under f, and write $y = f(x)$. The range of a function f is also called the *image* of f. The image of f is denoted $f(X)$, or $\mathrm{im}(f)$. The notation $f : X \to Y$ means f is a function from X to Y. If $T \subseteq Y$, the *preimage* of T under f, denoted $f^{-1}(T)$, is the set of all elements $x \in X$ such that $f(x) \in T$. If $y \in Y$, we usually write $f^{-1}(y)$ instead of $f^{-1}(\{y\})$. If $S \subseteq X$, the *restriction* of f to S is the function $f|_S : S \to Y$ defined by $f|_S(x) = f(x)$ for all $x \in S$. The *identity map* from X to X, $1_X : X \to X$, is defined by $1_X(x) = x$ for all $x \in X$. If $S \subseteq X$, the *inclusion map* from S to X is the restriction of the identity map 1_X to the subset S. If $f : X \to Y$ and $g : Y \to Z$, the *product* or *composition map* is $gf : X \to Z$ defined by $gf(x) = g(f(x))$. If $h : Z \to W$, the reader should verify that $h(gf) = (hg)f$ so the product of functions is associative. We say that $f : X \to Y$ is *one-to-one* (or *injective*) in case $f^{-1}(y)$ is a set with exactly one element for each $y \in f(X)$. We say that $f : X \to Y$ is *onto* or (*surjective*) in case the image of f is equal to Y. If $f : X \to Y$ is one-to-one and onto, then we say that f is a *one-to-one correspondence* (or f is *bijective*). The reader should verify that the identity map 1_X is a one-to-one correspondence. If $S \subseteq X$, the reader should verify that the inclusion map $S \to X$ is one-to-one.

PROPOSITION 1.1.1. *Let $f : X \to Y$.*

1. *f is one-to-one if and only if there exists $g : Y \to X$ such that $gf = 1_X$. In this case g is called a* left inverse *of f.*
2. *If f is a one-to-one correspondence, then the function g of Part (1) is unique and satisfies $fg = 1_Y$. In this case g is called the* inverse *of f and is denoted f^{-1}.*
3. *If there exists a function $g : Y \to X$ such that $gf = 1_X$ and $fg = 1_Y$, then f is a one-to-one correspondence and g is equal to f^{-1}.*

PROOF. (1): View f as a subset of $X \times Y$ and define g as a subset of $Y \times X$. Because f is not onto, our definition of g on $Y - f(X)$ is ad hoc. For this reason, let x_0 be any element of X. Define $g = \{(f(x), x) \mid x \in X\} \cup \{(y, x_0) \mid y \in Y - f(X)\}$. Then g has the desired properties. The rest is Exercise 1.1.8. □

For the counterpart of Proposition 1.1.1 (1) with 'onto' instead of 'one-to-one', see Exercise 1.3.8.

A *commutative diagram* is a finite family of sets $D_V = \{X_1, \ldots, X_v\}$ together with a finite collection of functions $D_E = \{f_1, \ldots, f_e\}$ satisfying the following properties.

(1) Each f in D_E is a function from one set in D_V to another set in D_V.
(2) Given two sets X, Y in D_V and any two paths

$$X = A_0 \xrightarrow{f_{a_1}} A_1 \xrightarrow{f_{a_2}} \cdots \to A_{r-1} \xrightarrow{f_{a_r}} A_r = Y$$

$$X = B_0 \xrightarrow{g_{b_1}} B_1 \xrightarrow{g_{b_2}} \cdots \to B_{s-1} \xrightarrow{g_{b_s}} B_s = Y$$

from X to Y consisting of functions $f_{a_1}, \ldots, f_{a_r}$, $g_{b_1}, \ldots, g_{b_s}$ in D_E, the composite functions $f_{a_r} \cdots f_{a_1}$ and $g_{b_s} \cdots g_{b_1}$ are equal.

1.3. Binary Relations. A *binary relation on* X is a subset of $X \times X$. Suppose $\sim$ is a binary relation on X. If (x, y) is an element of the relation, then we say x is *related* to y and write $x \sim y$. Otherwise we write $x \not\sim y$. If $x \sim x$ for every $x \in X$, then we say $\sim$ is *reflexive*. We say $\sim$ is *symmetric* in case $x \sim y$ whenever $y \sim x$. We say $\sim$ is *antisymmetric* in case $x \sim y$ and $y \sim x$ implies $x = y$. We say $\sim$ is *transitive* if $x \sim z$ whenever $x \sim y$ and $y \sim z$. If $\sim$ is reflexive, symmetric and transitive, then we say $\sim$ is an *equivalence relation* on X. If $\sim$ is an equivalence relation on X, and $x \in X$, then the *equivalence class* containing x is $[x] = \{y \in X \mid x \sim y\}$. By $X/\sim$ we denote the set of all equivalence classes. The function $\eta : X \to X/\sim$ defined by $\eta(x) = [x]$ is called the *natural map*.

PROPOSITION 1.1.2. *Let X be a nonempty set and $\sim$ an equivalence relation on X.*

(1) If $x \in X$, then $[x] \neq \emptyset$.
(2) $\bigcup_{x \in X} [x] = X = \bigcup_{[x] \in X/\sim} [x]$
(3) If $x, y \in X$, then $[x] = [y]$ or $[x] \cap [y] = \emptyset$.

PROOF. Is left to the reader. □

Let X be a nonempty set. A *partition* of X is a family $\mathcal{P}$ of nonempty subsets of X such that $X = \bigcup_{P \in \mathcal{P}} P$ and if $P, Q \in \mathcal{P}$, then either $P = Q$, or $P \cap Q = \emptyset$. If $\sim$ is an equivalence relation on X, then Proposition 1.1.2 shows that $X/\sim$ is a partition of X. Conversely, suppose $\mathcal{P}$ is a partition of X. There is an equivalence relation $\sim$ on X corresponding to $\mathcal{P}$ defined by $x \sim y$ if and only if x and y belong to the same element of $\mathcal{P}$.

PROPOSITION 1.1.3. *Let X be a nonempty set. There is a one-to-one correspondence between the set of all equivalence relations on X and the the set of all partitions of X. The assignment maps an equivalence relation $\sim$ to the partition $X/\sim$.*

PROOF. Is left to the reader. □

Let U be any set, which we assume contains $\mathbb{N}$ as a subset. Define a binary relation on the power set 2^U by the following rule. If X and Y are subsets of U, then we say X and Y are *equivalent* if there exists a one-to-one correspondence $\alpha : X \to Y$. The reader should verify that this is an equivalence relation on 2^U. If X and Y are equivalent sets, then we say X and Y have the same *cardinal number*. For $n \geq 1$ define $\mathbb{N}_n = \{1, \ldots, n\}$. If a set X is equivalent to $\mathbb{N}_n$, then we say X

has cardinal number n and write $|X| = n$. The cardinal number of the empty set is defined to be 0. We write $|\emptyset| = 0$. We say a set X is *finite* if X is equal to the empty set, or equivalent to $\mathbb{N}_n$ for some n. Otherwise, we say X is *infinite*.

Let X be a set and $\leq$ a binary relation on X which is reflexive, antisymmetric and transitive. Then we say $\leq$ is a *partial order* on X. We also say X is *partially ordered by* $\leq$. If $x, y \in X$, then we say x and y are *comparable* if $x \leq y$ or $y \leq x$. A *chain* is a partially ordered set with the property that any two elements are comparable. If $S \subseteq X$ is a nonempty subset, then S is partially ordered by the restriction of $\leq$ to $S \times S$. If the restriction of $\leq$ to S is a chain, then we say S is a *chain in* X.

Let X be partially ordered by $\leq$ and suppose S is a nonempty subset of X. Let $a \in S$. We say a is the *least* element of S if $a \leq x$ for all $x \in S$. If it exists, clearly the least element is unique. We say a is a *minimal* element of S in case $x \leq a$ implies $x = a$ for all $x \in S$. We say a is a *maximal* element of S in case $a \leq x$ implies $x = a$ for all $x \in S$. A *well ordered* set is a partially ordered set X such that every nonempty subset S has a least element. The reader should verify that a well ordered set is a chain. An element $u \in X$ is called an *upper bound* for S in case $x \leq u$ for all $x \in S$. An element $l \in X$ is called a *lower bound* for S in case $l \leq x$ for all $x \in S$. An element $U \in X$ is a *supremum*, or *least upper bound* for S, denoted $U = \sup(S)$, in case U is an upper bound for S and U is a lower bound for the set of all upper bounds for S. The reader should verify that the supremum is unique, if it exists. An element $L \in X$ is an *infimum*, or *greatest lower bound* for S, denoted $L = \inf(S)$, in case L is a lower bound for S and L is an upper bound for the set of all lower bounds for S. The reader should verify that the infimum is unique, if it exists. A *lattice* is a partially ordered set X such that $\sup\{x, y\}$ exists and $\inf\{x, y\}$ exists, for every pair of elements x, y in X.

Let X be partially ordered by $\leq$. We say that X satisfies the *minimum condition* if every nonempty subset of X contains a minimal element. We say that X satisfies the *maximum condition* if every nonempty subset of X contains a maximal element. We say that X satisfies the *descending chain condition* (DCC) if every chain in X of the form $\{\ldots, x_3 \leq x_2 \leq x_1 \leq x_0\}$ is eventually constant. That is, there is a subscript n such that $x_n = x_i$ for all $i \geq n$. We say that X satisfies the *ascending chain condition* (ACC) if every chain in X of the form $\{x_0 \leq x_1 \leq x_2 \leq x_3, \ldots\}$ is eventually constant.

1.4. Permutations and Combinations. Let $n \geq 1$ and $\mathbb{N}_n = \{1, 2, \ldots, n\}$. A bijection $\sigma : \mathbb{N}_n \to \mathbb{N}_n$ is also called a *permutation*. Let S_n denote the set of all permutations of $\mathbb{N}_n$. In Example 2.1.15 we will call S_n the symmetric group on n letters. If $\sigma \in S_n$, then we can view $\sigma = (x_1, \ldots, x_n)$ as an n-tuple in the product $\prod_{i=1}^{n} \mathbb{N}_n$. The fact that σ is a bijection is equivalent to the statement that the n-tuple $(x_1, \ldots, x_n)$ contains no repeated elements. Therefore,

$$S_n = \left\{ (x_1, \ldots, x_n) \in \prod_{i=1}^{n} \mathbb{N}_n \mid \text{if } i \neq j, \text{ then } x_i \neq x_j \right\}.$$

Because there are n ways to pick x_1, $n-1$ ways to pick x_2, and so forth, a straightforward induction proof shows that the number of elements in S_n is equal to $n!$. If $1 \leq k \leq n$, then a *k-permutation* of $\mathbb{N}_n$ is a one-to-one function $\sigma : \mathbb{N}_k \to \mathbb{N}_n$. The k-permutations of $\mathbb{N}_n$ correspond to k-tuples $(x_1, \ldots, x_k)$ where each $x_i \in \mathbb{N}_n$ and

if $i \neq j$, then $x_i \neq x_j$. Again, a straightforward induction proof shows that the number of k-permutations of $\mathbb{N}_n$ is equal to $n(n-1)\cdots(n-k+1) = n!/(n-k)!$.

If X is a finite set with cardinality $|X| = n$, then we say X is an *n-set*. If $S \subseteq X$ and $|S| = k$, then we say S is a *k-subset of X*. The number of k-subsets of an n-set X is denoted $\binom{n}{k}$. The symbol $\binom{n}{k}$ is called the *binomial coefficient* and is pronounced n choose k because it is the number of different ways to choose k objects from a set of n objects.

As we saw above, the number of different k-permutations of $\mathbb{N}_n$ is equal to $n!/(n-k)!$. But a k-permutation of $\mathbb{N}_n$ can be viewed as a two step process. The first step is choosing a k-subset, which can be done in $\binom{n}{k}$ different ways. Then the elements of the k-set are permuted, which can be done in $k!$ ways. Viewing the number of k-permutations of $\mathbb{N}_n$ in these two different ways, we see that $n!/(n-k)!$ is equal to $\binom{n}{k}(k!)$. This leads to Part (3) of the next lemma.

LEMMA 1.1.4. *The following are true.*

(1) If $k < 0$ or $n < 0$ or $k > n$, then $\binom{n}{k} = 0$.
(2) If $n \geq 0$, then $\binom{n}{0} = \binom{n}{n} = 1$.
(3) If $0 \leq k \leq n$, then $\dbinom{n}{k} = \dfrac{n!}{k!(n-k)!}$.
(4) (Pascal's Identity) If $0 < k < n$, then $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$.

PROOF. Parts (1) and (2) follow straight from the definition of binomial coefficient. Part (3) follows from the paragraph above. Part (4) follows directly from the formula in (3) and is left as an exercise for the reader. □

1.5. Binary Operations. Let X be a nonempty set. A *binary operation* on X is a function $X \times X \to X$. If $*$ is a binary operation on X, the image of an ordered pair (x, y) is denoted $x * y$. The binary operation is said to be *associative* if $(x * y) * z = x * (y * z)$ for all $x, y, z \in X$. If e is a special element in X such that $x * e = e * x = x$ for all $x \in X$, then we say e is an *identity element* for $*$. If $x*y = y*x$ for all $x, y \in X$, then we say $*$ is *commutative*. If $(x, y) \mapsto x \cdot y$ is another binary operation on X such that $x\cdot(y*z) = (x\cdot y)*(x\cdot z)$ and $(x*y)\cdot z = (x\cdot z)*(y\cdot z)$ for all $x, y, z \in X$, then we say $\cdot$ *distributes over* $*$.

EXAMPLE 1.1.5. Here are some common examples of binary operations on sets.

(1) Addition of numbers is a binary operation on the set of real numbers $\mathbb{R}$. Addition is associative, commutative, and 0 is the identity element. Multiplication of numbers is a binary operation on the set of real numbers $\mathbb{R}$. Multiplication is associative, commutative, and 1 is the identity element. Multiplication distributes over addition.
(2) Let U be a nonempty set and $X = 2^U$. If A and B are in X, then so are $A \cup B$, $A \cap B$, and $A - B$. In other words, union, intersection, and set difference all define binary operations on X. Union and intersection are both associative and commutative. The distributive laws for union and intersection are in Exercise 1.1.6.
(3) Let X be a nonempty set and $\mathrm{Map}(X)$ the set of all functions mapping X to X. If $f, g \in \mathrm{Map}(X)$, then so is the composite function fg. Composition of functions is a binary operation on $\mathrm{Map}(X)$ which is associative. That is, $(f\,(gh))\,(x) = ((fg)\,h)\,(x)$ for all $f, g, h \in \mathrm{Map}(X)$ and $x \in X$.

If $|X| > 1$, then composition of functions in $\mathrm{Map}(X)$ is noncommutative. The identity map 1_X is the identity element.

(4) Let $\mathbb{R}^3 = \{(x_1, x_2, x_3) \mid x_1, x_2, x_3 \in \mathbb{R}\}$ be the set of all ordered 3-tuples over $\mathbb{R}$. The *cross product* of the vector $\mathbf{x} = (x_1, x_2, x_3)$ and the vector $\mathbf{y} = (y_1, y_2, y_3)$ is the vector $\mathbf{x} \times \mathbf{y} = (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1)$. Therefore, cross product is a binary operation on $\mathbb{R}^3$. This binary operation is not associative and not commutative.

1.6. Exercises.

EXERCISE 1.1.6. (Distributive Laws for Intersection and Union) Let $\{X_i \mid i \in I\}$ be a family of sets indexed by I and let Y be any set. Prove:

(1) $Y \cap \left(\bigcup_{i\in I} X_i\right) = \bigcup_{i\in I}(Y \cap X_i)$
(2) $Y \cup \left(\bigcap_{i\in I} X_i\right) = \bigcap_{i\in I}(Y \cup X_i)$

EXERCISE 1.1.7. (DeMorgan's Laws) Let $\{X_i \mid i \in I\}$ be a family of sets indexed by I and suppose U is an arbitrary set. Prove:

(1) $U - \left(\bigcup_{i\in I} X_i\right) = \bigcap_{i\in I}(U - X_i)$
(2) $U - \left(\bigcap_{i\in I} X_i\right) = \bigcup_{i\in I}(U - X_i)$

EXERCISE 1.1.8. Finish the proof of Proposition 1.1.1.

EXERCISE 1.1.9. Let $f : X \to Y$ and $g : Y \to Z$. Prove:

(1) If gf is onto, then g is onto.
(2) If gf is one-to-one, then f is one-to-one.
(3) If f is onto and g is onto, then gf is onto.
(4) If f is one-to-one and g is one-to-one, then gf is one-to-one.

EXERCISE 1.1.10. Recall that the set of natural numbers is $\mathbb{N} = \{1, 2, \dots\}$ and if $n \in \mathbb{N}$, then $\mathbb{N}_n = \{1, 2, \dots, n\}$. Prove:

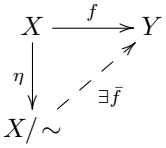
(1) If $f : \mathbb{N}_n \to \mathbb{N}_n$ is one-to-one, then f is onto.
(2) If $f : \mathbb{N}_n \to \mathbb{N}_n$ is onto, then f is one-to-one.

EXERCISE 1.1.11. (The Pigeonhole Principle) Let $f : \mathbb{N}_m \to \mathbb{N}_n$. Prove:

(1) If $m > n$, then f is not one-to-one.
(2) If $m < n$, then f is not onto.

EXERCISE 1.1.12. Let X and Y be finite sets. Show that $|X \times Y| = |X||Y|$.

EXERCISE 1.1.13. (Universal Mapping Property) Let $f : X \longrightarrow Y$ be a function. Let $\sim$ be an equivalence relation on X, and $\eta : X \to X/\sim$ the natural map. Show that if f has the property that $a \sim b$ implies $f(a) = f(b)$ for all $a, b \in X$, then there exists a function $\bar{f} : X/\sim \to Y$ such that $f = \bar{f}\eta$. Hence the diagram



commutes. This shows that if f is constant on equivalence classes, then f factors through the natural map η.

EXERCISE 1.1.14. Let $f : X \longrightarrow Y$ be a function. Define a relation $\approx$ on X by the rule: $x \approx y$ if and only if $f(x) = f(y)$. Prove:

(1) $\approx$ is an equivalence relation on X. If $x \in X$, then the equivalence class of x is $f^{-1}(f(x))$. We say the equivalence classes are the fibers of f.
(2) There exists a function $\bar{f} : X/\approx \to Y$ such that f factors through the natural map $\eta : X \to X/\approx$. That is, $f = \bar{f}\eta$.
(3) $\bar{f}$ is one-to-one.
(4) $\bar{f}$ is a one-to-one correspondence if and only if f is onto.

EXERCISE 1.1.15. Let X be an infinite set. Prove that X contains a subset that is equivalent to $\mathbb{N}$.

EXERCISE 1.1.16. Let X be a set. Prove that X is infinite if and only if there exists a one-to-one function $f : X \to X$ which is not onto.

EXERCISE 1.1.17. If $x \in \mathbb{R}$, the *floor* of x, written $\lfloor x \rfloor$, is the maximum of the set $\{k \in \mathbb{Z} \mid k \leq x\}$. The *ceiling* of x, written $\lceil x \rceil$, is the minimum of the set $\{k \in \mathbb{Z} \mid k \geq x\}$. Let $f : \mathbb{N}_m \to \mathbb{N}_n$. Prove:

(1) There exists $a \in \mathbb{N}_n$ such that the cardinality of the set $f^{-1}(a)$ is greater than or equal to $\lceil m/n \rceil$.
(2) There exists $b \in \mathbb{N}_n$ such that the cardinality of the set $f^{-1}(b)$ is less than or equal to $\lfloor m/n \rfloor$.

EXERCISE 1.1.18. Prove the Binomial Theorem:

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

where x and y are indeterminates and $n \geq 0$.

EXERCISE 1.1.19. Let X be a finite set. Use the Binomial Theorem to prove that $|2^X| = 2^{|X|}$.

EXERCISE 1.1.20. (Correspondence Theorem) Let X be a set and $\sim$ an equivalence relation on X. Let $\approx$ be an equivalence relation on $X/\sim$. Show that there exists an equivalence relation $\equiv$ on X and a one-to-one correspondence $\varphi : (X/\equiv) \to (X/\sim)/\approx$ such that the diagram

$$\begin{array}{ccc} X & \xrightarrow{\quad\eta_\sim\quad} & X/\sim \\ {\scriptstyle\eta_\equiv}\downarrow & & \downarrow{\scriptstyle\eta_\approx} \\ X/\equiv & \xrightarrow{\quad\varphi\quad} & (X/\sim)/\approx \end{array}$$

commutes where $\eta_\sim$, $\eta_\approx$, $\eta_\equiv$ are the natural maps.

2. Background Material from Number Theory

The basic results from Elementary Number Theory that will be required later in the text are included here. The set of integers is $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. We assume the reader is familiar with its partial ordering, and the binary operations of addition and multiplication. No attempt is made to construct the integers from from first principles. The set of natural numbers is $\mathbb{N} = \{1, 2, 3, \ldots\}$. The Well Ordering Principle is assumed as an axiom.

AXIOM 1.2.1. *(The Well Ordering Principle) If S is a nonempty subset of $\mathbb{Z}$ and S has a lower bound, then S contains a least element.*

PROPOSITION 1.2.2. *(Mathematical Induction) Let S be a subset of $\mathbb{N}$ such that $1 \in S$. Assume S satisfies one of the following.*

(1) For each $n \in \mathbb{N}$, if $n \in S$, then $n+1 \in S$.
(2) For each $n \in \mathbb{N}$, if $\{1, \ldots, n\} \subseteq S$, then $n+1 \in S$.

Then $S = \mathbb{N}$.

PROOF. Assume $S \subseteq \mathbb{N}$, $1 \in S$, and S satisfies (1) or (2). Let $C = \mathbb{N} - S$. For contradiction's sake assume $C \neq \emptyset$. By Axiom 1.2.1, C has a least element, say ℓ. Since $1 \in S$, we know $\ell > 1$. Therefore, $\ell - 1 \in S$ and $\ell \notin S$, which contradicts (1). Since ℓ is the least element of C, $\{1, \ldots, \ell - 1\} \subseteq S$ and $\ell \notin S$, which contradicts (2). We conclude that $C = \emptyset$, hence $S = \mathbb{N}$. □

PROPOSITION 1.2.3. *(The Division Algorithm) If $a, b \in \mathbb{Z}$ and $a \neq 0$, then there exist unique integers $q, r \in \mathbb{Z}$ such that $0 \leq r < |a|$ and $b = aq + r$.*

PROOF. First we prove the existence claim. The idea is to apply the Well Ordering Principle to the set $S = \{b - ax \mid x \in \mathbb{Z} \text{ and } b - ax \geq 0\}$. If $x > |b|$, then it follows that $b + |a|x \geq 0$. Therefore, either $b + ax$ or $b - ax$ is in S. By Axiom 1.2.1, S has a least element, say $r = b - aq$, for some $q \in \mathbb{Z}$. For contradiction's sake, assume $r \geq |a|$. Then $0 \leq r - |a| = b - aq - |a| = b - a(q \pm 1)$. This implies $r - |a| \in S$, contradicting the minimal choice of r.

To prove the uniqueness claim, suppose $b = aq + r = aq_1 + r_1$ and $0 \leq r \leq r_1 < |a|$. Then $|r_1 - r| = |a||q - q_1|$. Since $0 \leq r_1 - r < |a|$, this implies $q - q_1 = 0$. Hence $r_1 - r = 0$. □

Let $a, b \in \mathbb{Z}$. We say a *divides* b, and write $a \mid b$, in case there exists $q \in \mathbb{Z}$ such that $b = aq$. In this case, a is called a *divisor* of b, and b is called a *multiple* of a.

PROPOSITION 1.2.4. *Let $\{a_1, \ldots, a_n\}$ be a set of integers and assume at least one of the a_i is nonzero. There exists a unique positive integer d such that*

(1) $d \mid a_i$ for all $1 \leq i \leq n$, and
(2) if $e \mid a_i$ for all $1 \leq i \leq n$, then $e \mid d$.

We call d the greatest common divisor *of the set, and write $d = \gcd(a_1, \ldots, a_n)$.*

PROOF. Let S be the set of all positive linear combinations of the a_i

$$S = \{x_1a_1 + \cdots + x_na_n \mid x_1, \ldots, x_n \in \mathbb{Z},\ x_1a_1 + \cdots + x_na_n > 0\}.$$

The reader should verify that $S \neq \emptyset$. By Axiom 1.2.1, there exists a least element of S which we can write as $d = k_1a_1 + \cdots + k_na_n$ for some integers $k_1, \ldots, k_n$. Fix one i and apply the Division Algorithm to write $a_i = dq + r$ where $0 \leq r < d$. Solve $a_i = (k_1a_1 + \cdots + k_na_n)q + r$ for r to see that

$$r = a_i - (k_1a_1 + \cdots + k_na_n)q$$

is a linear combination of $a_1, \ldots, a_n$. Because $r < d$, we conclude that r is not in S. Therefore $r = 0$. This proves Part (1). The reader should verify Part (2) and the claim that d is unique. □

If $\gcd(a_1, \ldots, a_n) = 1$, then the set of integers $\{a_1, \ldots, a_n\}$ is said to be *relatively prime*. An integer $\pi \in \mathbb{Z}$ is called a *prime* in case $\pi > 1$ and the only divisors of π are $-\pi, -1, 1, \pi$.

LEMMA 1.2.5. *Let a, b and c be integers. Assume $a \neq 0$ or $b \neq 0$.*

(1) *(Bézout's Identity) If* $d = \gcd(a, b)$*, then there exist integers* u *and* v *such that* $d = au + bv$.
(2) *(Euclid's Lemma) If* $\gcd(a, b) = 1$ *and* $a \mid bc$*, then* $a \mid c$.
(3) *If there exist integers* u *and* v *such that* $1 = au + bv$*, then* $\gcd(a, b) = 1$.

PROOF. (1): This is immediate from the proof of Proposition 1.2.4.

(2): Assume $\gcd(a, b) = 1$. By Part (1) there exist integers u and v such that $1 = au + bv$. Then $c = acu + bcv$. Since a divides the right hand side, a divides c.

(3): This is immediate from the proof of Proposition 1.2.4. □

LEMMA 1.2.6. *Let* π *be a prime number. Let* a *and* $a_1, \ldots, a_n$ *be integers.*
(1) *If* $\pi \mid a$*, then* $\gcd(\pi, a) = \pi$*, otherwise* $\gcd(\pi, a) = 1$.
(2) *If* $\pi \mid a_1 a_2 \cdots a_n$*, then* $\pi \mid a_i$ *for some* i.

PROOF. (1): The proof is an exercise for the reader.

(2): For sake of contradiction, assume the statement is false. Let π and $a_1, \ldots, a_n$ be a counterexample such that n is minimal. Then π divides the product $a_1 \cdots a_n$ and by (1) $\gcd(\pi, a_i) = 1$ for each i. Again by (1), $n > 1$. By Lemma 1.2.5 applied to $a_1(a_2 \cdots a_n)$, $\pi \mid a_2 \cdots a_n$. By the minimal choice of n, π divides one of $a_2, \ldots, a_n$. This is a contradiction. □

PROPOSITION 1.2.7. *(The Fundamental Theorem of Arithmetic) Let* n *be a positive integer which is greater than* 1*. There exist unique positive integers* k*,* $e_1, \ldots, e_k$ *and unique prime numbers* $p_1, \ldots, p_k$ *such that* $n = p_1^{e_1} \cdots p_k^{e_k}$.

PROOF. First we prove the existence claim. If n is a prime, then set $k = 1$, $p_1 = n$, $e_1 = 1$, and we are done. In particular, the result is true for $n = 2$. The proof is by induction on n. Assume that every number in the set $\{2, 3, \ldots, n - 1\}$ has a representation as a product of primes. Assume $n = xy$ is composite and that $2 \leq x \leq y \leq n - 1$. By the induction hypothesis, both x and y have representations as products of primes. Then $n = xy$ also has such a representation. By Proposition 1.2.2, we are done.

For the uniqueness claim, assume

$$n = p_1^{e_1} \cdots p_k^{e_k} = q_1^{f_1} \cdots q_\ell^{f_\ell} \tag{2.1}$$

are two representations of n as products of primes. Let $M = \sum_{i=1}^{k} e_i$ and $N = \sum_{i=1}^{\ell} f_i$. Without loss of generality, assume $M \leq N$. The proof is by induction on M. If $M = 1$, then $n = p_1$ is prime. This implies $\ell = 1 = f_1$ and $q_1 = p_1$. Assume inductively that $M > 1$ and that the uniqueness claim is true for any product involving $M - 1$ factors. Using Lemma 1.2.6 we see that p_1 divides one of the q_i. Since q_i is prime, this implies p_1 is equal to q_i. Canceling p_1 and q_i from both sides of Eq.(2.1) results in a product of primes with $M - 1$ factors. By the induction hypothesis, we conclude that $k = \ell$ and the sets $\{p_1^{e_1}, \ldots, p_k^{e_k}\}$ and $\{q_1^{f_1}, \ldots, q_k^{f_k}\}$ are equal. □

DEFINITION 1.2.8. Let m be a positive integer. Define a binary relation on $\mathbb{Z}$ by the following rule. Given $x, y \in \mathbb{Z}$, we say x *is congruent to* y *modulo* m, and write $x \equiv y \pmod{m}$, in case $m \mid (x - y)$. By Proposition 1.2.9 this defines an equivalence relation on $\mathbb{Z}$. The set of all equivalence classes of integers modulo m is denoted $\mathbb{Z}/(m)$. The congruence class of x is denoted $[x]$.

PROPOSITION 1.2.9. *Let* m *be a positive integer.*

(1) *Congruence modulo m is an equivalence relation on $\mathbb{Z}$.*
(2) *$\{0, 1, \ldots, m-1\}$ is a full set of representatives for the equivalence classes. In other words, every integer is congruent to one of $0, 1, \ldots, m-1$ and no two distinct elements of $\{0, 1, \ldots, m-1\}$ are congruent to each other.*
(3) *If $u \equiv v \pmod{m}$ and $x \equiv y \pmod{m}$, then $u + x \equiv v + y \pmod{m}$ and $ux \equiv vy \pmod{m}$.*
(4) *If $\gcd(a, m) = 1$ and $ax \equiv ay \pmod{m}$, then $x \equiv y \pmod{m}$.*

PROOF. (1): Since $m \mid 0$, $x \equiv x \pmod{m}$ for every $x \in \mathbb{Z}$. If $x - y = mq$, then $y - x = m(-q)$. Therefore, $x \equiv y \pmod{m}$ implies $y \equiv x \pmod{m}$. If $x - y = mq$ and $y - z = mr$, then adding yields $x - z = m(q + r)$. Therefore, $x \equiv y \pmod{m}$ and $y \equiv z \pmod{m}$ implies $x \equiv z \pmod{m}$.

(2): By Proposition 1.2.3, if $x \in \mathbb{Z}$, then there exist unique integers q and r such that $x = mq + r$ and $0 \leq r < m$. This implies $x \equiv r \pmod{m}$, and $r \in \{0, 1, \ldots, m-1\}$. From the uniqueness part of the Division Algorithm, no two distinct elements of $\{0, 1, \ldots, m-1\}$ are congruent to each other.

(3): Write $u - v = mq$ and $x - y = mr$ for integers q, r. Adding, we get $u - v + x - y = (u + x) - (v + y) = m(q + r)$, hence $u + x \equiv v + y \pmod{m}$. Multiplying the first equation by x and the second by v we have $ux - vx = mxq$ and $xv - yv = mvr$. Adding, we get $ux - vx + xv - yv = ux - yv = m(xq + vr)$, hence $ux \equiv vy \pmod{m}$.

(4): By Lemma 1.2.5 we write $1 = au + mv$ for integers u, v. We are given that $a(x - y) = mq$ for some integer q. Multiply by u to get $au(x - y) = muq$. Substitute $au = 1 - mv$ and rearrange to get $x - y = mv(x - y) + muq$. Hence $x \equiv y \pmod{m}$. □

If $a, b \in \mathbb{Z} - \{0\}$, then $|ab|$ is a common multiple of both a and b. Therefore, the set $S = \{x \in \mathbb{N} \mid a \mid x,\ b \mid x\}$ is nonempty. By Axiom 1.2.1, S has a least element, which is called the *least common multiple* of a and b, and is denoted $\operatorname{lcm}(a, b)$.

PROPOSITION 1.2.10. *Suppose $a > 0$ and $b > 0$. Then the following are true.*
(1) *If $c \in \mathbb{Z}$ and $a \mid c$ and $b \mid c$, then $\operatorname{lcm}(a, b) \mid c$.*
(2) *$\gcd(a, b) \operatorname{lcm}(a, b) = ab$.*

PROOF. (1): Let $\operatorname{lcm}(a, b) = L$. By Proposition 1.2.3, $c = Lq + r$ where $0 \leq r < L$. Since $a \mid c$ and $a \mid L$, we see that a divides $r = c - Lq$. Likewise, $b \mid c$ and $b \mid L$ implies that b divides r. So r is a common multiple of a and b and $r < L$. By the definition of L, we conclude that $r = 0$.

(2): Write $d = \gcd(a, b)$. Then $(ab)/d = a(b/d) = (a/d)b$ is a common multiple of a and b. By (1), $L \mid (ab)/d$, or equivalently, $dL \mid ab$. By Lemma 1.2.5, $d = ax + by$ for some integers x, y. Multiply by L to get $dL = aLx + bLy$. Since L is a common multiple of a and b we see that ab divides $aLx + bLy = dL$. We have shown that $dL \mid ab$ and $ab \mid dL$. Both numbers are positive, so we have equality. □

THEOREM 1.2.11. *(Chinese Remainder Theorem) Let m and n be relatively prime positive integers. Then the function*

$$\mathbb{Z}/mn \xrightarrow{\psi} \mathbb{Z}/m \times \mathbb{Z}/n$$

defined by $\psi([x]) = ([x], [x])$ is a one-to-one correspondence.

PROOF. We know that ψ is well defined, by Exercise 1.2.19. By Exercise 1.1.12 and Proposition 1.2.9, $|\mathbb{Z}/m \times \mathbb{Z}/n| = |\mathbb{Z}/mn| = mn$. By Exercise 1.1.10, it is

enough to show ψ is one-to-one. Suppose $\psi([x]) = \psi([y])$. Then $x \equiv y \pmod{m}$ and $x \cong y \pmod{n}$, which implies $x - y$ is a common multiple of m and n. By Proposition 1.2.10, $x - y$ is divisible by $\operatorname{lcm}(m, n)$. But $\operatorname{lcm}(m, n) = mn$ since $\gcd(a, b) = 1$. This implies $x \equiv y \pmod{mn}$, and we have shown that ψ is one-to-one. □

For a generalization of Theorem 1.2.11, see Theorem 2.5.2.

Let $n \geq 2$. By Exercise 1.2.20, if $x \equiv y \pmod{n}$, then $\gcd(x, n) = \gcd(y, n)$. This says the function $\mathbb{Z} \to \mathbb{Z}$ defined by $x \mapsto \gcd(x, n)$ is constant on congruence classes. The set $U_n = \{[k] \in \mathbb{Z}/n \mid \gcd(k, n) = 1\}$ is called the set of *units modulo* n. The Euler ϕ-function, named for Leonhard Euler, is defined to be the number of units modulo n. That is, $\phi(n) = |U_n|$. In the terminology of Definition 2.1.1, Lemma 1.2.12 shows that U_n is an abelian group of order $\phi(n)$.

LEMMA 1.2.12. *Let $n \geq 2$.*

(1) If $[a] \in U_n$, then there exists $[b] \in U_n$ such that $[a][b] = [1]$.
(2) If $a, b \in \mathbb{Z}$ and $ab \equiv 1 \pmod{n}$, then $[a] \in U_n$ and $[b] \in U_n$.

PROOF. (1): If $[a] \in U_n$, then $\gcd(a, n) = 1$. By Lemma 1.2.5, there exist integers b, c such that $ab + nv = 1$. Therefore, $ab \equiv 1 \pmod{n}$.

(2): If $ab \equiv 1 \pmod{n}$, then $ab = nq + 1$ for some integer q. By Lemma 1.2.5, $\gcd(a, n) = 1$ and $\gcd(b, n) = 1$. □

PROPOSITION 1.2.13. *If p is a prime and $k \geq 1$, then $\phi(p^k) = p^k - p^{k-1} = p^k(1 - 1/p)$.*

PROOF. The multiples of p in the set $\{1, 2, \ldots, p^k\}$ are $p, 2p, \ldots, p^{k-1}p$. Since there are p^{k-1} multiples of p, there are $p^k - p^{k-1}$ numbers that are relatively prime to p. □

PROPOSITION 1.2.14. *Let $m \geq 2$, $n \geq 2$ and assume $\gcd(m, n) = 1$. Then $\phi(mn) = \phi(m)\phi(n)$.*

PROOF. By Theorem 1.2.11, the function $\psi : \mathbb{Z}/mn \to \mathbb{Z}/m \times \mathbb{Z}/n$ defined by $\psi([x]) = ([x], [x])$ is a one-to-one correspondence. We show that the restriction of ψ to U_{mn} induces a one-to-one correspondence $\rho : U_{mn} \to U_m \times U_n$.

If $\gcd(x, mn) = 1$, then by Lemma 1.2.5 there exist integers u, v such that $1 = xu + mnv$, hence $\gcd(x, m) = 1$ and $\gcd(x, n) = 1$. This proves that ρ is well defined. Since ψ is one-to-one, so is ρ. To finish the proof we show that ρ is onto. Let $([a], [b]) \in U_m \times U_n$. By Lemma 1.2.12 there exists $([x], [y]) \in U_m \times U_n$ such that $ax \equiv 1 \pmod{m}$ and $by \equiv 1 \pmod{n}$. Since ψ is onto, there exists $[k] \in \mathbb{Z}/mn$ such that $k \equiv a \pmod{m}$ and $k \equiv b \pmod{n}$. Likewise, there exists $[\ell] \in \mathbb{Z}/mn$ such that $\ell \equiv x \pmod{m}$ and $\ell \equiv y \pmod{n}$. By Proposition 1.2.9, $k\ell \equiv ax \equiv 1 \pmod{m}$ and $k\ell \equiv by \equiv 1 \pmod{n}$. Since ψ is one-to-one, $k\ell \equiv 1 \pmod{mn}$. By Lemma 1.2.12 this implies $[k] \in U_{mn}$, which proves ρ is onto. □

DEFINITION 1.2.15. Let $n \geq 1$ be an integer. The notation $\sum_{d|n}$ or $\prod_{d|n}$ denotes the sum or product over the set of all positive numbers d such that $d \mid n$. An integer n is said to be *square free* if for every prime p, n is not a multiple of p^2.

The Möbius function is defined by

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n \text{ is not square free,} \\ (-1)^r & \text{if } n \text{ factors into } r \text{ distinct primes.} \end{cases}$$

THEOREM 1.2.16. *(Möbius Inversion Formula) Let f be a function defined on $\mathbb{N}$ and define another function on $\mathbb{N}$ by*

$$F(n) = \sum_{d|n} f(d).$$

Then

$$f(n) = \sum_{d|n} \mu(d) F\left(\frac{n}{d}\right).$$

PROOF. The proof can be found in any elementary number theory book, and is left to the reader. □

2.1. Exercises.

EXERCISE 1.2.17. Let a and b be integers that are not both zero and let d be the greatest common divisor of a and b. Consider the linear diophantine equation: $d = ax + by$. Bézout's Identity says that there exist integers u and v such that $d = au + bv$.

(1) Show that the matrix $\begin{pmatrix} u & v \\ -b/d & a/d \end{pmatrix}$ is invertible over $\mathbb{Z}$. Find its inverse.
(2) If c is an integer, show that the linear diophantine equation $c = ax + by$ has a solution if and only if $d \mid c$.
(3) Assume $d \mid c$. Prove that the general solution to the linear diophantine equation $c = ax + by$ is $x = x_0 - tb/d$, $y = y_0 + ta/d$, where $t \in \mathbb{Z}$ and (x_0, y_0) is any particular solution.

EXERCISE 1.2.18. This exercise is based on Problem 1.3 of Adrian Wadsworth's book [**29**]. Let a and b be relatively prime positive integers and consider the set

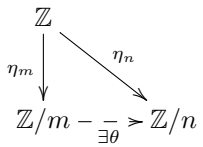
$$L = \{ax + by \mid x \text{ and } y \text{ are nonnegative integers}\}.$$

The problem is to find the integer ℓ satisfying these two properties: (1) $\ell - 1 \notin L$ and (2) if n is an integer and $n \geq \ell$, then $n \in L$.

You are encouraged to solve this interesting problem yourself. Alternatively, you may follow the six steps below which outline a solution.

(1) Prove that if $a = 1$ or $b = 1$, then L contains the set of all nonnegative integers.
(2) Prove that the integers a, b, ab, $(a-1)(b-1)$ are in L.
(3) Prove that $ab - a - b = (a-1)(b-1) - 1$ is not in L.
(4) Prove that if $n \geq ab$, then n is in L.
(5) Assume $a > 1$, $b > 1$, and let n be an integer satisfying $ab - a - b < n < ab$. Prove that n is in L.
(6) Let $\ell = (a-1)(b-1)$. Prove that $\ell - 1 \notin L$ and if $\ell \leq n$, then n is in L.

EXERCISE 1.2.19. Let $m, n \in \mathbb{N}$. Consider the diagram



where η_m and η_n are the natural maps. Show that there exists a function θ making the diagram commute if and only if n divides m. See Lemma 2.3.29 for an application of this result.

EXERCISE 1.2.20. Let $n \geq 1$. Show that the function $\mathbb{Z} \to \mathbb{Z}$ defined by $x \mapsto \gcd(x, n)$ is constant on congruence classes. In other words, show that $x \equiv y \pmod{n}$ implies $\gcd(x, n) = \gcd(y, n)$.

EXERCISE 1.2.21. Let p be a prime.

(1) If $1 \leq k \leq p-1$, show that p divides $\binom{p}{k}$.
(2) Show that $(a+b)^p \equiv a^p + b^p \pmod{p}$ for any integers a and b.
(3) Use (2) and Proposition 1.2.2 to prove that $(a+b)^{p^n} \equiv a^{p^n} + b^{p^n} \pmod{p}$ for any integers a and b and for all $n \geq 0$.

See Exercise 3.6.35 for a generalization of this result.

EXERCISE 1.2.22. Show that the Möbius function μ is multiplicative in the sense that if $\gcd(m, n) = 1$, then $\mu(mn) = \mu(m)\mu(n)$.

EXERCISE 1.2.23. Let $n \geq 0$ and $X = \prod_{i=1}^{n} \mathbb{Z}_{\geq 0} = \{(x_1, \ldots, x_n) \mid x_i \in \mathbb{Z}_{\geq 0}\}$, where $\mathbb{Z}_{\geq 0} = \{x \in \mathbb{Z} \mid x \geq 0\}$ is the set of nonnegative integers. The *lexicographical ordering* (also called alphabetical or dictionary ordering) on X is defined recursively on n. For $n = 1$, the usual ordering on $\mathbb{Z}$ is applied. If $n > 1$, then

$(v_1, v_2, \ldots, v_n) < (w_1, w_2, \ldots, w_n)$ if and only if

$$\begin{cases} (v_1, v_2, \ldots, v_{n-1}) < (w_1, w_2, \ldots, w_{n-1}), \text{ or} \\ (v_1, v_2, \ldots, v_{n-1}) = (w_1, w_2, \ldots, w_{n-1}) \text{ and } v_n < w_n. \end{cases}$$

If $\alpha, \beta \in X$, then we write $\alpha \leq \beta$ in case $\alpha < \beta$ or $\alpha = \beta$.

(1) Show that $\leq$ is a partial order on X. Show that X is a chain.
(2) If $\alpha \in X$, the *segment of X determined by α*, written $(-\infty, \alpha)$, is $\{x \in X \mid x < \alpha\}$. For which $\alpha \in X$ is
   (a) $(-\infty, \alpha) = \emptyset$?
   (b) $(-\infty, \alpha)$ finite?
   (c) $(-\infty, \alpha)$ infinite?
(3) Show that X with the lexicographical ordering $\leq$ is a well ordered set. That is, show that if $S \subseteq X$ and $S \neq \emptyset$, then S has a least element.

EXERCISE 1.2.24. Let $X = \{x_0, x_1, \ldots, x_{n-1}\}$ be a finite set and $\mathbb{Z}_{\geq 0}$ the set of nonnegative integers. If $U \subseteq X$, the so-called *indicator function* on U, denoted $\chi_U : U \to \{0, 1\}$, is defined by

$$\chi_U(x) = \begin{cases} 1 & \text{if } x \in U \\ 0 & \text{if } x \notin U. \end{cases}$$

Define $f : 2^X \to \mathbb{Z}_{\geq 0}$ by $f(U) = \sum_{i=0}^{n-1} \chi_U(x_i) 2^i$. Prove:

(1) f is a one-to-one correspondence between 2^X and $\{0, 1, \ldots, 2^n - 1\}$.
(2) $|2^X| = 2^{|X|}$.
(3) The ordering on 2^X induced by the function f makes 2^X into a well ordered set.

EXERCISE 1.2.25. (Partial Fractions) Let m and n be natural numbers such that $m < n$. Follow the following outline to show that a partial fraction decomposition exists for the rational number m/n. Prove:

(1) Let b and c be natural numbers such that $\gcd(b, c) = 1$. Let d be an integer.
   (a) There exist integers e, f such that $\frac{d}{bc} = \frac{e}{b} + \frac{f}{c}$.
   (b) If e' and f' are another pair of integers such that $\frac{d}{bc} = \frac{e'}{b} + \frac{f'}{c}$, then $e \equiv e' \pmod{b}$ and $f \equiv f' \pmod{c}$.
(2) Let a, b and c be natural numbers such that $\gcd(b, c) = 1$ and $a < bc$. Then there exist unique integers e, f such that $0 \leq e < b$, $-c < f < c$ and
$$\frac{a}{bc} = \frac{e}{b} + \frac{f}{c}.$$
(3) Let $p_1, \ldots, p_n$ be distinct prime numbers. For $i = 1, \ldots, n$, let $e_i \geq 1$ be a positive integer. Let $b = p_1^{e_1} \cdots p_n^{e_n}$. If a is a positive integer such that $a < b$, then there exist integers $a_1, \ldots, a_n$ such that
$$\frac{a}{b} = \frac{a}{p_1^{e_1} \cdots p_n^{e_n}} = \frac{a_1}{p_1^{e_1}} + \cdots + \frac{a_n}{p_n^{e_n}}$$
and $|a_i| < p_i^{e_i}$ for each i. The integers $a_1, \ldots, a_n$ are unique in the sense that if $a_1', \ldots, a_n'$ also satisfy the equation
$$\frac{a}{b} = \frac{a}{p_1^{e_1} \cdots p_n^{e_n}} = \frac{a_1'}{p_1^{e_1}} + \cdots + \frac{a_n'}{p_n^{e_n}},$$
then $a_i \equiv a_i' \pmod{p_i^{e_i}}$ for each i.
(4) (Base b Representation) Let b, n and a be natural numbers. Assume $b > 1$ and $a < b^n$. Then there exist unique integers $a_0, a_1, \ldots, a_{n-1}$ such that $0 \leq a_i < b$ for each i and
$$a = a_0 + a_1 b + \cdots + a_{n-1} b^{n-1}.$$
(5) Let b, n and a be natural numbers. Assume $b > 1$ and $a < b^n$. Then there exist unique integers $a_0, a_1, \ldots, a_{n-1}$ such that $0 \leq a_i < b$ for each i and
$$\frac{a}{b^n} = \frac{a_0}{b^n} + \frac{a_1}{b^{n-1}} + \cdots + \frac{a_{n-1}}{b}.$$

3. The Well Ordering Principle and Some of Its Equivalents

Most readers will prefer to make a quick scan of this section on first reading. It is included for completeness' sake as well as a tribute to the influence of [**17**, Chapter 0, Theorem 25] on the author's fondness for the subject. In this book, the only direct application of Zorn's Lemma, Proposition 1.3.3, is in the proof that a commutative ring contains a maximal ideal (see Proposition 3.2.27). As a historical note, Zorn's Lemma, which is equivalent to the Well Ordering Principle, has been called Zorn's Lemma since M. Zorn first used it to prove that a commutative ring contains a maximal ideal [**33**]. The Axiom of Choice, Proposition 1.3.5, guarantees that a product of nonempty sets is nonempty, but throughout this book we limit

our applications either to products of at most a countably infinite number of sets, or to a product of algebraic structures like groups for instance. Such a product always contains an identity element. In addition to the application to show the existence of maximal ideals, the other applications of the Well Ordering Principle or one of its equivalents appear in Section 4.2.3 and in several exercises that are inserted to challenge the reader.

Although we do not prove it here, the Well Ordering Principle, the Principle of Transfinite Induction, Zorn's Lemma, and the Axiom of Choice are logically equivalent to each other.

AXIOM 1.3.1. *(The Well Ordering Principle) If X is a nonempty set, then there exists a partial order $\leq$ on X such that X is a well ordered set. That is, every nonempty subset of X has a least element.*

Let X be a set and $\leq$ a partial order on X. If $x, y \in X$, then we write $x < y$ in case $x \leq y$ and $x \neq y$. Suppose $C \subseteq X$ is a chain in X and $\alpha \in C$. The *segment of C determined by α*, written $(-\infty, \alpha)$, is the set of all elements $x \in C$ such that $x < \alpha$. A subset $W \subseteq C$ is called an *inductive subset* of C provided that for any $\alpha \in C$, if $(-\infty, \alpha) \subseteq W$, then $\alpha \in W$.

PROPOSITION 1.3.2. *(The Transfinite Induction Principle) Suppose X is a well ordered set and W is an inductive subset of X. Then $W = X$.*

PROOF. Suppose $X - W$ is nonempty. Let α be the least element of $X - W$. Then W contains the segment $(-\infty, \alpha)$. Since W is inductive, it follows that $\alpha \in W$, which is a contradiction. □

PROPOSITION 1.3.3. *(Zorn's Lemma) Let X be a partially ordered set. If every chain in X has an upper bound, then X contains a maximal element.*

PROOF. By Axiom 1.3.1, there exists a well ordered set W and a one-to-one correspondence $\omega : W \to X$. Using Proposition 1.3.2, define a sequence $\{C(w) \mid w \in W\}$ of well ordered subsets of X. If w_0 is the least element of W, define $C(w_0) = \{\omega(w_0)\}$. Inductively assume $\alpha \in W - \{w_0\}$ and that for all $w < \alpha$, $C(w)$ is defined and the following are satisfied

(1) if $w_0 \leq w_1 \leq w_2 < \alpha$, then $C(w_1) \subseteq C(w_2)$,
(2) $C(w)$ is a well ordered chain in X, and
(3) $C(w) \subseteq \{\omega(i) \mid w_0 \leq i \leq w\}$.

Let $x = \omega(\alpha)$ and

$$F = \bigcup_{w<\alpha} C(w).$$

The reader should verify that F is a well ordered chain in X and $F \subseteq \{\omega(i) \mid w_0 \leq i < \alpha\}$. Define $C(\alpha)$ by the rule

$$C(\alpha) = \begin{cases} F \cup \{x\} & \text{if } x \text{ is an upper bound for } F \\ F & \text{otherwise.} \end{cases}$$

The reader should verify that $C(\alpha)$ satisfies

(4) if $w_0 \leq w_1 \leq w_2 \leq \alpha$, then $C(w_1) \subseteq C(w_2)$,
(5) $C(\alpha)$ is a well ordered chain in X, and
(6) $C(\alpha) \subseteq \{\omega(i) \mid w_0 \leq i \leq \alpha\}$.

By Proposition 1.3.2, the sequence $\{C(w) \mid w \in W\}$ is defined and the properties (4), (5) and (6) are satisfied for all $\alpha \in W$. Now set

$$G = \bigcup_{w<\alpha} C(w).$$

The reader should verify that G is a well ordered chain in X. By hypothesis, G has an upper bound, say u. We show that u is a maximal element of X. For contradiction's sake, assume X has no maximal element. Then we can choose the upper bound u to be an element of $X - G$. For some $w_1 \in W$ we have $u = \omega(w_1)$. For all $w < w_1$, u is an upper bound for $C(w)$. By the definition of $C(w_1)$, we have $u \in C(w_1)$. This is a contradiction, because $C(w_1) \subseteq G$. □

DEFINITION 1.3.4. Let I be a set and $\{X_i \mid i \in I\}$ a family of sets indexed by I. The *product* is

$$\prod_{i\in I} X_i = \{f : I \to \bigcup X_i \mid f(i) \in X_i\}.$$

An element f of the product is called a choice function, because f chooses one element from each member of the family of sets. Sometimes the product is called the *cartesion product*.

PROPOSITION 1.3.5. *(The Axiom of Choice) Let I be a set and $\{X_i \mid i \in I\}$ a family of nonempty sets indexed by I. Then the product $\prod_{i\in I} X_i$ is nonempty. That is, there exists a function f on I such that $f(i) \in X_i$ for each $i \in I$.*

PROOF. By Axiom 1.3.1, we can assume $\bigcup_{i\in I} X_i$ is well ordered. We can view X_i as a subset of $\bigcup_{i\in I} X_i$. For each $i \in I$, let x_i be the least element of X_i. The set of ordered pairs (i, x_i) defines the choice function. □

3.1. Exercises.

EXERCISE 1.3.6. Let I be a set and $\{X_i \mid i \in I\}$ a family of nonempty sets indexed by I. For each $k \in I$ define $\pi_k : \prod_{i\in I} X_i \to X_k$ by the rule $\pi_k(f) = f(k)$. We call π_k the *projection onto coordinate k*. Show that π_k is onto.

EXERCISE 1.3.7. Let X be a set that is partially ordered by $\leq$.

(1) Prove that X satisfies the descending chain condition (DCC) if and only if X satisfies the minimum condition.
(2) Prove that X satisfies the ascending chain condition (ACC) if and only if X satisfies the maximum condition.

EXERCISE 1.3.8. Use the Axiom of Choice to prove: A function $f : X \to Y$ is onto if and only if there exists a function $g : Y \to X$ such that $fg = 1_Y$. In this case g is called a *right inverse* of f.

4. Background Material from Calculus

As in Section 1.1.1, the set of real numbers is denoted $\mathbb{R}$.

THEOREM 1.4.1. *If a is a positive real number, then there exists a real number x such that $x^2 = a$. In other words, a positive real number has a square root.*

PROOF. See, for instance, [**27**, Theorem 7.8, p. 124]. □

THEOREM 1.4.2. *If n is a positive odd integer and $a_0, a_1, \ldots, a_{n-1}$ are real numbers, then there exists a real number x such that $x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$. In other words, a polynomial over $\mathbb{R}$ of odd degree has a root.*

PROOF. See, for instance, [**27**, Theorem 7.9, p. 125]. □

For properties of the complex numbers, the reader is referred, for example, to [**27**, Chapter 25]. The set of complex numbers, denoted $\mathbb{C}$, is identified with $\mathbb{R}^2$ and is a two-dimensional real vector space. A complex number is an ordered pair (a, b). A basis for $\mathbb{C}$ is $(1, 0)$, also denoted 1, and $(0, 1)$, also denoted i. In terms of this basis, the complex number (a, b) has representation $a + bi$. Addition of complex numbers is coordinate-wise: $(a + bi) + (c + di) = (a + c) + (b + d)i$. The additive identity is $0 = (0, 0)$ and the additive inverse of $a + bi$ is $-a - bi$. Multiplication distributes over addition, and $i^2 = -1$, hence $(a + bi)(c + di) = ac + (ad + bc)i + bdi^2 = (ac - bd) + (ad + bc)i$. The multiplicative identity is $1 = (1, 0) = 1 + 0i$. If $z = a + bi$, then the *absolute value* of z is $|z| = \sqrt{a^2 + b^2}$, which is equal to the length of the vector (a, b). Let $r = |a + bi|$. If θ is the angle determined by the vectors $z = a + bi$ and $1 = (1, 0)$, then the representation of z in polar coordinates is $z = a + bi = r\cos\theta + ir\sin\theta$. The complex conjugate of $z = a + bi$ is $\chi(z) = a - bi$. Then $z\chi(z) = a^2 + b^2 = |z|^2$ is a nonnegative real number. This implies if $z \neq 0$, then z is invertible and

$$z^{-1} = \frac{a - bi}{a^2 + b^2}.$$

The power series for the functions e^x, $\cos x$, and $\sin x$ are

$$\begin{aligned} e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \frac{x^6}{6!} + \frac{x^7}{7!} + \frac{x^8}{8!} + \ldots \\ \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} + \ldots \\ \sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots. \end{aligned}$$

These power series converge for every real number x. We define e^{ix} to be the substitution of ix into the power series. Using the identities $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, and $i^5 = i$, we have

$$\begin{aligned} e^{ix} &= 1 + ix + \frac{i^2x^2}{2!} + \frac{i^3x^3}{3!} + \frac{i^4x^4}{4!} + \frac{i^5x^5}{5!} + \frac{i^6x^6}{6!} + \frac{i^7x^7}{7!} + \frac{i^8x^8}{8!} + \ldots \\ &= 1 + ix - \frac{x^2}{2!} - \frac{ix^3}{3!} + \frac{x^4}{4!} + \frac{ix^5}{5!} - \frac{x^6}{6!} - \frac{ix^7}{7!} + \frac{x^8}{8!} + \ldots \\ &= \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} + \ldots\right) + i\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots\right) \\ &= \cos x + i\sin x. \end{aligned}$$

Therefore, if $z = a + bi$ has polar representation $r\cos\theta + ir\sin\theta$, then the representation for z in exponential form is $a + bi = re^{i\theta}$.

PROPOSITION 1.4.3. *In exponential notation, arithmetic in $\mathbb{C}$ satisfies the following formulas.*

(1) (Additive inverse) $-re^{i\theta} = re^{i(\theta+\pi)}$.
(2) (Multiplication) $re^{i\theta}se^{i\phi} = (rs)e^{i(\theta+\phi)}$.

(3) *(Complex conjugation)* $\chi\left(re^{i\theta}\right) = re^{-i\theta}$.
(4) *(Multiplicative inverse)* $\left(re^{i\theta}\right)^{-1} = r^{-1}e^{-i\theta}$.
(5) *(Square root) If* $r \geq 0$, *then* $z^{1/2} = \sqrt{re^{i\theta}} = \sqrt{r}e^{i\theta/2}$.
(6) *(nth root) If* $r \geq 0$, *then* $z^{1/n} = \left(re^{i\theta}\right)^{1/n} = r^{1/n}e^{i\theta/n}$.

PROOF. The proof is left to the reader. □

5. Background Material from Matrix Theory

The general theorems on matrices are derived and presented in Section 6.5. In that section, the entries of the matrices are usually from an arbitrary ring R. Nevertheless, matrices comprise an important source for examples. Therefore, we will be introducing many examples of groups, rings and modules of matrices throughout Chapters 2, 3, and 4 before the general theorems are proved. For example, the group of invertible two-by-two matrices is introduced in Example 2.1.21. The center of the ring of n-by-n matrices is computed in Example 3.1.13. The Cayley-Hamilton Theorem for two-by-two matrices is proved in Proposition 4.4.16.

In this short section we define the addition and multiplication formulas for matrices and derive some of their common properties. The goal is to be brief, with the understanding that most readers have already been exposed to this material. Throughout the remainder of this section we denote by R an arbitrary ring. In particular, R can be the ring of integers, $\mathbb{Z}$, or the ring of integers modulo n, $\mathbb{Z}/(n)$, or one of the fields $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$.

Let m and n be positive integers. An m-by-n *matrix* is an array consisting of m rows and n columns of elements from R. By $M_{mn}(R)$ we denote the set of all m-by-n matrices over R. If $A \in M_{mn}(R)$, we write $A = (a_{ij})$, where a_{ij} refers to the entry in row i and column j. If $m = n$, then we simply write $M_n(R)$ instead of $M_{nn}(R)$. Addition of matrices is coordinate-wise. If $B = (b_{ij}) \in M_{mn}(R)$, then $A + B$ is the matrix whose entry in position i, j is $a_{ij} + b_{ij}$. That is, $(a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$. We can multiply an element r of R (r is sometimes called a *scalar*) and a matrix $A = (a_{ij})$ by the rule $r(a_{ij}) = (ra_{ij})$. The m-by-n zero matrix, denoted 0, is the matrix whose entry in position i, j is 0. The matrix $(-a_{ij})$ is denoted $-A$.

PROPOSITION 1.5.1. *Let* A, B, C *be matrices in* $M_{mn}(R)$ *and* r, s *elements of* R. *Addition of matrices and multiplication by elements of* R *satisfy the following properties.*

(1) *Addition is associative:* $(A + B) + C = A + (B + C)$.
(2) *Addition is commutative:* $A + B = B + A$.
(3) *Additive inverses exist:* $-A + A = 0$.
(4) 0 *is the additive identity:* $0 + A = A$.
(5) *Scalar multiplication distributes over addition:* $r(A + B) = rA + rB$.
(6) *Another associative law:* $(rs)A = r(sA)$.
(7) *Another distributive law:* $(r + s)A = rA + sA$.

PROOF. The proof is left to the reader. □

If $A = (a_{ij}) \in M_{mn}(R)$ and $B = (b_{jk}) \in M_{np}(R)$, then the product AB is the matrix $C = (c_{ik})$ in $M_{mp}(R)$ whose entry in position i, k is $c_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk}$. The n-by-n *identity matrix* is the matrix I_n in $M_n(R)$ whose entry in position i, j is 1 if $i = j$ and 0 otherwise. If $r \in R$, then the matrix rI_n is called a scalar matrix.

PROPOSITION 1.5.2. *Addition and multiplication of matrices satisfy the following properties. In the following, assume $A \in M_{mn}(R)$, $B, B_1 \in M_{np}(R)$, and $C \in M_{pq}(R)$.*

(1) Multiplication is associative: $(AB)C = A(BC)$.
(2) I is the multiplicative identity: $I_m A = A = AI_n$.
(3) Multiplication distributes over addition from both the left: $A(B + B_1) = AB + AB_1$, and from the right: $(B + B_1)C = BC + B_1C$.

PROOF. The first property is called the associative law for matrix multiplication. We prove it and leave the rest to the reader. Write $A = (a_{ij})$, $B = (b_{jk})$, and $C = (c_{ks})$. Then AB is the m-by-p matrix whose entry in position i, k is

$$\sum_{j=1}^{n} a_{ij} b_{jk}.$$

The product $(AB)C$ is the m-by-q matrix whose entry in position i, s is

$$\sum_{k=1}^{p} \left(\sum_{j=1}^{n} a_{ij} b_{jk} \right) c_{ks}. \tag{5.1}$$

The product BC is the n-by-q matrix whose entry in position j, s is

$$\sum_{k=1}^{p} b_{jk} c_{ks}.$$

The product $A(BC)$ is the m-by-q matrix whose entry in position i, s is

$$\sum_{j=1}^{n} a_{ij} \left(\sum_{k=1}^{p} b_{jk} c_{ks} \right). \tag{5.2}$$

The fact that (5.1) and (5.2) are equal proves the associative law for multiplication. □

If e_{ij} is the m-by-n matrix with 1 in position (i, j) and 0 elsewhere, then e_{ij} is called an *elementary matrix*. Let $A = (a_{ij})$ be a matrix. Then A is an *upper triangular matrix* if $a_{ij} = 0$ whenever $i > j$. Likewise, A is a *lower triangular matrix* if $a_{ij} = 0$ whenever $i < j$. If $a_{ij} = 0$ whenever $i \neq j$, then we say A is a *diagonal matrix*. The notation $\operatorname{diag}(a_1, \ldots, a_n)$ represents the diagonal matrix $a_1 e_{11} + \cdots + a_n e_{nn}$.

As in Section 1.1.4, S_n denotes the symmetric group on n letters, where $n \geq 2$. Let σ be a permutation in S_n. The n-by-n matrix $P_\sigma = \sum_{i=1}^{n} e_{\sigma(i),i}$ is called the *permutation matrix* associated to σ. Notice that the n-by-n identity matrix I_n is equal to the permutation matrix $e_{11} + \cdots + e_{nn}$ associated to the identity permutation.

PROPOSITION 1.5.3. *Let σ be a permutation in S_n and P_σ the permutation matrix associated to σ. If A is an m-by-n matrix in $M_{mn}(R)$, and B is an n-by-p matrix in $M_{np}(R)$, then*

(1) AP_σ is the matrix obtained by permuting the columns of A by σ. That is, column i of AP_σ is column $\sigma(i)$ of A.
(2) P_σ is the matrix obtained by permuting the columns of I_n by the permutation σ. That is, column i of P_σ is column $\sigma(i)$ of I_n.

(3) $P_\sigma B$ *is the matrix obtained by permuting the rows of* B *by* σ^{-1}*. That is, row* i *of* $P_\sigma B$ *is row* $\sigma^{-1}(i)$ *of* B*.*

(4) P_σ *is the matrix obtained by permuting the rows of* I_n *by the permutation* σ^{-1}*. That is, row* i *of* P_σ *is row* $\sigma^{-1}(i)$ *of* I_n*.*

PROOF. (1): Let $1 \leq i \leq n$, and consider the product $Ae_{\sigma(i),i}$. If we denote column i of A by A_i, then $Ae_{\sigma(i),i}$ is the matrix $(0, ..., 0, A_{\sigma(i)}, 0, \ldots, 0)$ with column i equal to $A_{\sigma(i)}$, and all other columns equal to 0. Then AP_σ is the matrix $(A_{\sigma(1)}, A_{\sigma(2)}, \ldots, A_{\sigma(n)})$ obtained by permuting the columns of A by σ.

(3): Follows from an argument similar to that used to prove (1).

(2) and (4): Follow from (1) and (3) respectively. □

CHAPTER 2

Groups

Groups arise in all areas of Mathematics. All of the other algebraic structures that arise are also based on groups. A module is an abelian group, a ring is an additive abelian group and the set of invertible elements of a ring is a multiplicative group. For this reason the theorems of this chapter are fundamental.

In a concrete sense, a group is a set of permutations of a set. E. Galois first emphasized the importance of studying permutations of the roots of polynomials. Group Theory can be viewed as an axiomatic abstraction of permutation groups.

1. First Properties of Groups

The notion of a binary operation on a set was introduced in Section 1.1.5. The main ideas remain the same, but we recast them in light of the present context. Let G be a nonempty set with a binary operation $G \times G \to G$. Usually the binary operation on a group will be written either multiplicatively or additively. In the multiplicative notation, the *product* of two elements a, b in G is denoted by ab, an identity element will usually be denoted e or 1 and the inverse of an element a will be written a^{-1}. If additive notation is used, the *sum* of a and b is denoted by $a+b$, an identity is usually denoted 0 and $-a$ denotes the inverse of a.

1.1. Definitions, Terminology, and First Properties. After the formal definition of a group is stated, some of the first properties of groups are proved. For instance, we establish the associative law for arbitrary finite products, the notion of a general power, the uniqueness of an identity element, the solvability and cancellation theorems. Some of the first examples of groups are given. These include the group of integers modulo n under addition, the group of units modulo n under multiplication, the group of permutations of a set, the symmetric group on n letters, the group of symmetries of a regular n-gon, the quaternion 8-group, and the general linear group of 2-by-2 matrices with entries in a field.

Definition 2.1.1. Let G be a nonempty set with a multiplicative binary operation. If $a(bc) = (ab)c$ for all $a, b, c \in G$, then the binary operation is said to be associative. In this case, G is called a *semigroup*. If G is a semigroup and G contains an element e satisfying $ae = ea = a$ for all $a \in G$, then e is said to be an identity element and G is called a *monoid*. Let G be a monoid with identity element e. An element $a \in G$ is said to be *invertible* if there exists $a^{-1} \in G$ such that $aa^{-1} = a^{-1}a = e$. The element a^{-1} is called the *inverse of* a. A monoid in which every element is invertible is called a *group*. In other words, a group is a nonempty set G together with an associative binary operation such that an identity element e exists in G, and every element of G is invertible. If $xy = yx$ for all $x, y \in G$, then the binary operation is said to be commutative. A commutative group is called an *abelian group*.

If G has an additive binary operation, then the associative law is $(a+b)+c=a+(b+c)$ for all $a,b,c\in G$. The element $0\in G$ is an identity element if $a+0=0+a=a$ for all $a\in G$. The element a is invertible if there exists an inverse element $-a\in G$ such that $a+(-a)=(-a)+a=0$. The commutative law is $a+b=b+a$ for all $a,b\in G$. As a rule, additive notation is not used for nonabelian groups.

EXAMPLE 2.1.2. Let X be a nonempty set. A one-to-one correspondence $\sigma: X\to X$ is also called a permutation of X. The set of all permutations of X is denoted $\mathrm{Perm}(X)$. We are in the context of Example 1.1.5 (3) and $\mathrm{Perm}(X)$ is a subset of $\mathrm{Map}(X)$. Composition of functions is a binary operation on $\mathrm{Map}(X)$ which is associative and 1_X is the identity element. Therefore, $\mathrm{Map}(X)$ is a monoid. If σ and τ are permutations of X, then so is the composite function $\sigma\tau$, by Proposition 1.1.1. Therefore, $\mathrm{Perm}(X)$ is a group with identity element 1_X. See Example 2.1.15 for the important special case where $|X|$ is finite. We will see later (for example, Example 2.1.15) that if $|X|>2$, then $\mathrm{Perm}(X)$ is nonabelian.

EXAMPLE 2.1.3. Here are some examples of abelian groups.

(1) Under addition, $\mathbb{Z}$ is an abelian group with identity 0. The inverse of x is written $-x$.
(2) Let $n\in\mathbb{N}$. Proposition 1.2.9 shows that under addition, $\mathbb{Z}/(n)$ is an abelian group with identity $[0]$. The inverse of $[x]$ is $[-x]$. We have $|\mathbb{Z}/(n)|=n$.
(3) Let $n\in\mathbb{N}$. Lemma 1.2.12 shows that the set of units modulo n, U_n, is a multiplicative abelian group. The identity element is $[1]$ and $|U_n|=\phi(n)$.

Let G be a multiplicative semigroup. The associative law on G says that $(ab)c=a(bc)$. In other words, a product of length three has a unique value regardless of how we associate the multiplications into binary operations using parentheses. When writing a product abc it is not necessary to use parentheses. The next lemma extends this result to products of arbitrary finite length.

LEMMA 2.1.4. *(General Associative Law) Let G be a semigroup, $n\geq 1$, and $x_1x_2\cdots x_n$ a product involving n elements of G. Then the product has a unique value regardless of how we associate the multiplications into binary operations using parentheses.*

PROOF. First we define a standard value for $x_1x_2\cdots x_n$ by the recursive formula:

$$x_1x_2\cdots x_n=\begin{cases}x_1 & \text{if } n=1\\ (x_1x_2\cdots x_{n-1})x_n & \text{if } n>1.\end{cases}$$

Now we show that any association of $x_1x_2\cdots x_n$ will result in the value defined above. The proof is by induction on n. If $n\leq 3$, then this is true by the associative law on G. Inductively assume $n>3$ and that the result holds for any product of length less than n. Let $x_1x_2\cdots x_n$ be a product involving n elements. Assume the product is associated into binary operations using parentheses. Then the last binary operation can be written as

$$(x_1x_2\cdots x_m)(x_{m+1}\cdots x_n)$$

and by the induction hypothesis, the two products $x_1x_2\cdots x_m$ and $x_{m+1}\cdots x_m$ have unique values regardless of how they are associated. If $m=n-1$, then we are

done, by the induction hypothesis. Assume $1 \leq m < n-1$. Using the associative law on G and the induction hypothesis, we get

$$\begin{aligned}(x_1x_2\cdots x_m)(x_{m+1}\cdots x_n) &= (x_1x_2\cdots x_m)((x_{m+1}\cdots x_{n-1})x_n)\\ &= ((x_1x_2\cdots x_m)(x_{m+1}\cdots x_{n-1})x_n)\\ &= (x_1x_2\cdots x_{n-1})x_n\\ &= x_1x_2\cdots x_n\end{aligned}$$

which completes the proof. □

DEFINITION 2.1.5. Let G be a group, $a \in G$, and n a nonnegative integer.

(1) If G is a multiplicative group, then the *nth power of a* is defined recursively by the formula:

$$a^n = \begin{cases} e & \text{if } n = 0 \\ aa^{n-1} & \text{if } n > 0. \end{cases}$$

We define a^{-n} to be $(a^{-1})^n$. Using induction, the reader should verify that $(a^{-1})^n$ is equal to $(a^n)^{-1}$.

(2) For an additive group G, the counterpart of the nth power of a is *left multiplication of a by n*, which is defined recursively by:

$$na = \begin{cases} 0 & \text{if } n = 0 \\ a + (n-1)a & \text{if } n > 0. \end{cases}$$

We define $(-n)a$ to be $n(-a)$, which is equal to $-(na)$.

PROPOSITION 2.1.6. *Let G be a group and a, b, c elements of G.*

(1) There exists a unique x in G such that $ax = b$.
(2) There exists a unique y in G such that $ya = b$.
(3) We have $ab = ac$ if and only if $b = c$.
(4) We have $ab = cb$ if and only if $a = c$.

Parts (1) and (2) are called the solvability *properties, Parts (3) and (4) are called the* cancellation *properties.*

PROOF. (3): Assume we have $ab = ac$. Multiply both sides on the left by a^{-1} to get $a^{-1}ab = a^{-1}ac$. Since $a^{-1}ab = eb = b$ and $a^{-1}ac = ec = c$, we get $b = c$. Conversely, multiplying both sides of $b = c$ from the left with a yields $ab = ac$.

(1): Let $x = a^{-1}b$. Multiply by a on the left to get $ax = aa^{-1}b = eb = b$. If x' is another solution, then $ax = ax'$ and by Part (3) we have $x = x'$.

Parts (4) and (2) are proved in a similar manner. □

LEMMA 2.1.7. *If G is a group and x, y are elements of G, then the following are true.*

(1) If $x^2 = x$, then $x = e$. We say that a group has exactly one idempotent.
(2) If $xy = e$, then $y = x^{-1}$.
(3) $(x^{-1})^{-1} = x$.
(4) $(xy)^{-1} = y^{-1}x^{-1}$.

PROOF. (1): Applying Proposition 2.1.6 (3) to $x^2 = x = xe$, we have $x = e$.
(2): Applying Proposition 2.1.6 (1) to $xy = e$, we have $y = x^{-1}e = x^{-1}$.
(3): Applying Proposition 2.1.6 (1) to $x^{-1}x = e$, we have $x = (x^{-1})^{-1}$.
(4): Applying (2) to $(xy)(y^{-1}x^{-1}) = e$, we have $y^{-1}x^{-1} = (xy)^{-1}$. □

EXAMPLE 2.1.8. Let G be a group. Let $a \in G$ be a fixed element. Then "left multiplication by a" defines a function $\lambda_a : G \to G$, where $\lambda_a(x) = ax$. Part (1) of Proposition 2.1.6 says that λ_a is onto and Part (3) says that λ_a is one-to-one. Therefore, λ_a is a one-to-one correspondence. Likewise, "right multiplication by a" defines a one-to-one correspondence $\rho_a : G \to G$ where $\rho_a(x) = xa$.

DEFINITION 2.1.9. If G is a group, then the *order of* G is the cardinality of the underlying set. The order of G is denoted $[G : e]$ or $|G|$ or $o(G)$.

DEFINITION 2.1.10. Let G be a group and $a \in G$. The *order of* a, written $|a|$, is the least positive integer m such that $a^m = e$. If no such integer exists, then we say a has infinite order.

DEFINITION 2.1.11. Let G and G' be groups. A function $\theta : G \to G'$ is called an *isomorphism of groups*, if θ is a one-to-one correspondence and $\theta(xy) = \theta(x)\theta(y)$ for all $x, y \in G$. In this case, we say G and G' are *isomorphic* and write $G \cong G'$. From an abstract algebraic point of view, isomorphic groups are indistinguishable.

1.2. Examples of Groups.

EXAMPLE 2.1.12. In this example we show that there is up to isomorphism only one group of order two. By Example 2.1.3, a group of order two exists, namely the additive group $\mathbb{Z}/2$. Let $G = \{e, a\}$ be an arbitrary group of order two, where e is the identity element. By Example 2.1.8, left multiplication by a is a permutation of G. Since $ae = a$, this implies $aa = e$. In other words, there is only one binary operation that makes $\{e, a\}$ into a group. If $G' = \{e, b\}$ is a group, then the function that maps $e \mapsto e$, $a \mapsto b$ is an isomorphism.

EXAMPLE 2.1.13. We know from Example 2.1.3 that the additive group $\mathbb{Z}/3$ is an abelian group of order three. In this example we show that up to isomorphism there is only one group of order three. Let $G = \{e, a, b\}$ be an arbitrary group of order three, where e is the identity element. By Example 2.1.8, λ_a and ρ_a are permutations of G. By cancellation, $ab = b$ leads to the contradiction $a = e$. Since $ae = a$, we conclude that $ab = e$ and $aa = b$. Similarly, $ba = b$ is impossible, hence we conclude that $ba = e$. We have shown that $G = \{e, a, a^2\}$ and a has order 3. Suppose $G' = \{e, c, c^2\}$ is another group of order 3. Then the assignments $a^i \mapsto c^i$ for $i = 0, 1, 2$ define an isomorphism.

EXAMPLE 2.1.14. If $X = \{x_1, \ldots, x_n\}$ is a finite set, then a binary operation on X can be represented as an n-by-n matrix with entries from X. Sometimes we call the matrix the "multiplication table" or "addition table". If the binary operation is $*$, then the entry in row i and column j of the associated matrix is the product $x_i * x_j$. For instance, the multiplication and addition tables for $\mathbb{Z}/6$ are:

$*$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

$+$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

If the binary operation $*$ on X is commutative, then the matrix is symmetric with respect to the main diagonal. If $X, *$ is a group, then by Example 2.1.8, each row

of the multiplication table is a permutation of the top row and each column is a permutation of the leftmost column. See Exercise 2.1.28 for more examples.

EXAMPLE 2.1.15. Let $n \geq 1$ and $\mathbb{N}_n = \{1, 2, \ldots, n\}$. The set of all permutations of $\mathbb{N}_n$ is called the *symmetric group on n letters* and is denoted S_n. In Example 2.1.2 we saw that composition of functions makes $S_n = \text{Perm}(\mathbb{N}_n)$ into a group. As shown in Section 1.1.4, the group S_n has order $n!$. A permutation can be specified using an array of two rows. For example,

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & \ldots & n \\ a_1 & a_2 & a_3 & \ldots & a_n \end{bmatrix}$$

represents the permutation $\sigma(i) = a_i$. Occasionally we simply write the same permutation as the ordered list of length n: $\sigma = (a_1, a_2, a_3, \ldots, a_n)$. The so-called cycle notation is a very convenient way to represent elements of S_n. Let $\{a_1, \ldots, a_k\} \subseteq \mathbb{N}_n$. The *k-cycle* $\sigma = (a_1 a_2 \ldots a_k)$ is the permutation of $\mathbb{N}_n$ defined by:

$$\sigma(x) = \begin{cases} x & \text{if } x \notin \{a_1, \ldots, a_k\} \\ a_1 & \text{if } x = a_k \\ a_{i+1} & \text{if } x = a_i \text{ and } 1 \leq i < k. \end{cases}$$

Notice that a k-cycle has order k in the group S_n. The inverse of $\sigma = (a_1 a_2 \ldots a_k)$ is $\sigma^{-1} = (a_1 a_k a_{k-1} \ldots a_2)$ which is also a k-cycle. It is important to realize that the notation for k-cycles is not unique. The k-cycles $(a_1 a_2 \ldots a_k)$, $(a_2 \ldots a_k a_1)$, $\ldots$, $(a_k a_1 \ldots a_{k-1})$, all denote the same permutation. A 2-cycle is also called a *transposition*. The identity element of S_n is usually denoted e. For example, $(abc)(ab) = (ac)$ and $(ab)(abc) = (bc)$. Therefore, S_n is nonabelian if $n > 2$. The group table for $S_3 = \{e, (abc), (acb), (ab), (ac), (bc)\}$ is:

$*$	e	(abc)	(acb)	(ab)	(ac)	(bc)
e	e	(abc)	(acb)	(ab)	(ac)	(bc)
(abc)	(abc)	(acb)	(e)	(ac)	(bc)	(ab)
(acb)	(acb)	(e)	(abc)	(bc)	(ab)	(ac)
(ab)	(ab)	(bc)	(ac)	(e)	(acb)	(abc)
(ac)	(ac)	(ab)	(bc)	(abc)	(e)	(acb)
(bc)	(bc)	(ac)	(ab)	(acb)	(abc)	(e)

EXAMPLE 2.1.16. Let T be a regular triangle with vertices labeled $1, 2, 3$. A *symmetry* of T is any one-to-one correspondence $\sigma : T \to T$ that preserves distances and maps adjacent vertices to adjacent vertices. Therefore, σ is a permutation of the three vertices. Conversely, a permutation of $\{1, 2, 3\}$ uniquely determines a symmetry of T. The group of symmetries of T is therefore equal to S_3.

EXAMPLE 2.1.17. Now let $n > 2$ and let T_n be a regular n-gon with vertices labeled $1, 2, \ldots, n$ consecutively. A symmetry of T_n is any one-to-one correspondence $\sigma : T_n \to T_n$ that preserves distances and maps adjacent vertices to adjacent vertices Therefore, σ is a permutation of the n vertices. If $n > 3$, a permutation of $\{1, 2, \ldots, n\}$ does not necessarily determine a symmetry of T_n. When $n > 3$, the group of symmetries of T_n is therefore a proper subgroup of S_n. The group of all symmetries of T_n is called the *dihedral group* D_n. A rotation of T_n through an

angle of $2\pi/n$ corresponds to the permutation

$$R = \begin{bmatrix} 1 & 2 & 3 & \dots & n-1 & n \\ 2 & 3 & 4 & \dots & n & 1 \end{bmatrix}$$

which in cycle notation is the n-cycle $R = (12 \dots n)$. Therefore, R^k is a rotation of T_n through an angle of $2\pi k/n$, hence R has order n. A top to bottom flip of T_n across the line of symmetry containing vertex 1 corresponds to the permutation defined by

$$H = \begin{cases} \begin{bmatrix} 1 & 2 & 3 & \dots & k & k+1 & \dots & n-1 & n \\ 1 & n & n-1 & \dots & k+2 & k+1 & \dots & 3 & 2 \end{bmatrix} & \text{if } n = 2k \text{ is even,} \\ \begin{bmatrix} 1 & 2 & 3 & \dots & k & k+1 & \dots & n-1 & n \\ 1 & n & n-1 & \dots & k+1 & k & \dots & 3 & 2 \end{bmatrix} & \text{if } n = 2k-1 \text{ is odd.} \end{cases}$$

In cycle notation, H can be represented as

$$H = \begin{cases} (2,n)(3,n-1)\cdots(k,k+2) & \text{if } n = 2k \text{ is even,} \\ (2,n)(3,n-1)\cdots(k,k+1) & \text{if } n = 2k-1 \text{ is odd.} \end{cases}$$

Then $HH = e$, hence H has order 2. The reader should verify that $HRH = R^{-1}$. Any symmetry of T_n is either a rotation or a rotation followed by a flip. Therefore we see that $D_n = \{H^i R^j \mid 0 \le i \le 1, 0 \le j < n\}$ is a nonabelian group of order $2n$.

EXAMPLE 2.1.18. Let R_4 be a nonsquare rectangle with vertices labeled consecutively $1, 2, 3, 4$. The group of symmetries of R_4 can be viewed as a subgroup of S_4 as well as a subgroup of D_4. In the notation of Example 2.1.17, the group of symmetries of R_4 is $\{H^i R^j \mid 0 \le i \le 1, 0 \le j \le 1\}$, which is a group of order four. In cycle notation, this group is $\{e, (14)(23), (12)(34), (13)(24)\}$. Note that the group is abelian and every element satisfies the identity $x^2 = e$.

EXAMPLE 2.1.19. The *quaternion 8-group* is $Q_8 = \{1, -1, i, -i, j, -j, k, -k\}$ with identity element 1. The multiplication rules are: $(-1)^2 = 1$, $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$. This is an example of a nonabelian group of order eight. The group table for Q_8 is:

	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

For a continuation of this example, see Exercise 2.4.23. The group Q_8 is also called the group of quaternion units because it is a subgroup of the group of units of the ring of real quaternions $\mathbb{H}_{\mathbb{R}}$, which was discovered by W. R. Hamilton. For the definition of the ring of quaternions over an arbitrary field, see Example 3.1.14.

EXAMPLE 2.1.20. Let F be a field. If α is a nonzero element of F, then α has a multiplicative inverse, denoted α^{-1}. The set of all nonzero elements of F is a multiplicative group. This group is denoted F^* and is called the *group of units of* F.

EXAMPLE 2.1.21. Let F be a field. The set of all n-by-n matrices over F is denoted $M_n(F)$. In this example, we assume the reader is familiar with the basic properties for multiplication of matrices (see Section 1.5). In this example our goal is to show that the set of 2-by-2 matrices over F with nonzero determinant is a group. For $n = 2$, the determinant function $\det : M_2(F) \to F$ is defined by

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

To show that the determinant function is multiplicative, start with the product of two arbitrary 2-by-2 matrices:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}.$$

The determinant formula applied on the left hand side yields: $(ad-bc)(eh-fg) = adeh - adfg - bceh + bcfg$. The reader should verify that this is equal to the determinant of the right hand side: $(ae + bg)(cf + dh) - (ce + dg)(af + bh)$. A matrix α is invertible if there is a matrix β such that $\alpha\beta = \beta\alpha = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Taking determinants, this implies $\det\alpha \det\beta = 1$. In other words, if α is invertible, then $\det\alpha \neq 0$. Notice that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} = (ad - bc) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

If $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc \neq 0$, then the matrix is invertible and the inverse is given by the formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = (ad - bc)^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

The set

$$\mathrm{GL}_2(F) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(F) \mid ad - bc \neq 0 \right\}$$

is the set of all invertible 2-by-2 matrices over F and is called the *general linear group of* 2*-by-*2 *matrices over* F. For a continuation of this example when $F = \mathbb{Z}/2$ is the field of order 2, see Exercise 2.1.26.

EXAMPLE 2.1.22. The Klein Viergruppe, or Klein 4-group, is $V = \{e, a, b, c\}$ with multiplication rules: $a^2 = b^2 = c^2 = e$, $ab = ba = c$. Notice that V is isomorphic to the group of symmetries of a nonsquare rectangle presented in Example 2.1.18 by the mapping: $a \mapsto (14)(23)$, $b \mapsto (12)(34)$, $c \mapsto (13)(24)$.

1.3. Exercises.

EXERCISE 2.1.23. Let G be a monoid with identity element e.

(1) Show that G has exactly one identity element. In other words, show that if $e' \in G$ has the property that $ae' = e'a = a$, then $e = e'$.

(2) Show that an invertible element of G has a unique inverse. In other words, if $aa^{-1} = a^{-1}a = e$ and $aa' = a'a = e$, then $a^{-1} = a'$.
(3) Suppose $a, r, \ell \in G$ satisfy the identities: $ar = e$ and $\ell a = e$. Show that $r = \ell$ and a is invertible.
(4) Suppose every element of G has a left inverse. In other words, assume for every $a \in G$ there exists $a_l \in G$ such that $a_l a = e$. Show that G is a group.
(5) If $a \in G$ is invertible, then a^{-1} is invertible and $(a^{-1})^{-1} = a$.
(6) If a and b are invertible elements of G, then ab is invertible and $(ab)^{-1} = b^{-1}a^{-1}$.

EXERCISE 2.1.24. Let G be a group. The *opposite group of* G is denoted G^o. As a set, G^o is equal to G. The binary operation on G^o is reversed from that of R. Writing the multiplication of R by juxtaposition and multiplication of R^o with the asterisk symbol, we have $x * y = yx$. Show that G^o is a group. Show that G is isomorphic to G^o.

EXERCISE 2.1.25. Let G be a group. Prove the following:

(1) If $x^2 = e$ for all $x \in G$, then G is abelian.
(2) If $|G| = 2n$ for some $n \in \mathbb{N}$, then there exists $x \in G$ such that $a \neq e$ and $a^2 = e$.

EXERCISE 2.1.26. In this example, we assume the reader is familiar with the basic properties for multiplication of matrices (see Proposition 1.5.2). In particular, multiplication of matrices is associative and the product of a two-by-two matrix times a two-by-one column vector is defined by:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} au + bv \\ cu + dv \end{pmatrix}.$$

Let $G = \mathrm{GL}_2(\mathbb{Z}/2)$ be the group of two-by-two invertible matrices over the field $\mathbb{Z}/2$ (see Example 2.1.21). List the elements of G and construct the group table (see Example 2.1.14). Show that G has two elements of order three and three elements of order two. Let

$$a = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, c = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and consider the set of column vectors $\{a, b, c\}$ over $\mathbb{F}_2$. For every matrix α in G, show that left multiplication by the matrix α defines a permutation of the set $\{a, b, c\}$. Comparing the group table for G with the group table given in Example 2.1.15 for S_3, the symmetric group on 3 letters, show that $\mathrm{GL}_2(\mathbb{Z}/2)$ is isomorphic to S_3.

EXERCISE 2.1.27. Let K and H be groups. Define a binary operation on $K \times H$ by $(x_1, y_1)(x_2, y_2) = (x_1x_2, y_1y_2)$. Show that this makes $K \times H$ into a group with identity element (e, e), and the inverse of (x, y) is (x^{-1}, y^{-1}). Show that $K \times H$ is abelian if and only if K and H are both abelian.

EXERCISE 2.1.28. For various values of n, each of the following matrices is an n-by-n multiplication table representing a binary operation $*$ on the set $I_n = \{0, 1, \ldots, n-1\}$. In each case, determine whether the binary operation (a) is commutative, (b) is associative, (c) has an identity element, and (d) is a group.

(1)

$*$	0	1	2	3
0	0	0	0	0
1	0	1	1	3
2	0	2	3	0
3	0	3	1	2

(2)

$*$	0	1	2	3	4	5	6	7
0	4	2	6	0	7	1	5	3
1	5	4	0	1	6	7	3	2
2	1	7	4	2	5	3	0	6
3	0	1	2	3	4	5	6	7
4	7	6	5	4	3	2	1	0
5	6	0	3	5	2	4	7	1
6	2	3	7	6	1	0	4	5
7	3	5	1	7	0	6	2	4

(3)

$*$	0	1	2	3	4	5	6	7
0	4	5	3	2	0	1	7	6
1	7	4	5	6	1	2	3	0
2	3	7	4	0	2	6	5	1
3	2	6	0	4	3	7	1	5
4	0	1	2	3	4	5	6	7
5	6	0	1	7	5	3	2	4
6	5	3	7	1	6	0	4	2
7	1	2	6	5	7	4	0	3

(4)

$*$	0	1	2	3	4	5	6	7
0	7	2	1	4	3	6	5	0
1	2	7	0	5	6	3	4	1
2	1	0	7	6	5	4	3	2
3	4	5	6	7	0	1	2	3
4	3	6	5	0	7	2	1	4
5	6	3	4	1	2	7	0	5
6	5	4	3	2	1	0	7	6
7	0	1	2	3	4	5	6	7

(5)

$*$	0	1	2
0	2	0	1
1	0	1	2
2	1	2	0

(6)

$*$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	0	4	5	3
2	2	0	1	5	3	4
3	3	5	4	0	2	1
4	4	3	5	1	0	2
5	5	4	3	2	1	0

EXERCISE 2.1.29. In the following, $\mathbb{Z}/n$ denotes the group of integers modulo n (see Example 2.1.3), D_n is the dihedral group (see Example 2.1.17), V is the Klein 4-group (see Example 2.1.22), Q_8 is the quaternion 8-group (see Example 2.1.19). Prove:

(1) $\mathbb{Z}/4$ is not isomorphic to V.
(2) $\mathbb{Z}/(2n)$ is not isomorphic to D_n.
(3) No two of $\mathbb{Z}/8$, Q_8, or D_4 are isomorphic to each other.

2. Subgroups and Cosets

A subgroup of a group G is a subset that is itself a group under the binary operation on G. One way we study groups is in terms of their subgroups. A subset X of a subgroup H is said to be a generating set for H if H is the smallest subgroup of G that contains X. One way to study subgroups is in terms of their generators. Associated to a subgroup H is an equivalence relation on G called left congruence modulo H. Specifically, two elements x and y of G are left congruent modulo H if there is an element z in H such that $y = xz$. The equivalence class of x is the set $xH = \{xz \mid z \in H\}$, which is called the left coset of x modulo H. The set of all left cosets of H in G is denoted G/H and there is a natural map $\eta : G \to G/H$. If G is a finite group, then all left cosets of H have the same cardinality. Lagrange's Theorem says the order of G is divisible by the order of H and the quotient is equal to the number of cosets of H. If H and K are two subgroups of G, then

the number of elements in the set HK is equal to the order of H times the order of K divided by the order of the intersection $H \cap K$. This important formula is called a counting theorem. The last part of this section contains results on cyclic subgroups. A subgroup is cyclic if it contains a generating set consisting of a single element. In particular, we show that an infinite group always contains an infinite number of cyclic subgroups. One reason cyclic subgroups are important is that any group is the union of its cyclic subgroups.

2.1. First Properties of Subgroups. First we state the formal definition of a subgroup. Then in Lemma 2.2.2 we give two more equivalent conditions.

DEFINITION 2.2.1. If G is a group and H is a nonempty subset of G that is a group under the binary operation on G, then we say H is a *subgroup of* G and write $H \leq G$.

If H is a nonempty subset of a group G, then Lemma 2.2.2 provides two useful tests for whether H is a subgroup of G or not.

LEMMA 2.2.2. *Let G be a group and H a nonempty subset of G. The following are equivalent.*
(1) H is a subgroup of G.
(2) For all a, b in H we have $ab \in H$ and $a^{-1} \in H$.
(3) For all a, b in H we have $ab^{-1} \in H$.

PROOF. (2) implies (1): Let $a \in H$. Then $e = aa^{-1} \in H$. The associative law applies on G, hence on H. The other group properties are included in (2).

(1) implies (3): Let a and b be elements of H. If H is a group, then $b^{-1} \in H$ and $ab^{-1} \in H$.

(3) implies (2): Let a and b be elements of H. By (3) we have $aa^{-1} = e \in H$, $ea^{-1} = a^{-1} \in H$, and $a(b^{-1})^{-1} = ab \in H$. □

EXAMPLE 2.2.3. Let G be a group. Then $\{e\}$ and G are both subgroups of G. We call these the *trivial subgroups of* G. A nontrivial subgroup is also called a *proper subgroup*.

When G is a finite group, Proposition 2.2.4 is a subgroup test that is a simplified form of Lemma 2.2.2.

PROPOSITION 2.2.4. *Let G be a group and H a finite subset of G. If for all $a, b \in H$ we have $ab \in H$, then H is a subgroup of G.*

PROOF. Assume $a, b \in H$ implies $ab \in H$. By Lemma 2.2.2, to show H is a subgroup it suffices to show that $a \in H$ implies $a^{-1} \in H$. Let $|H| = n$. Define $f : \mathbb{N}_{n+1} \to H$ be defined by $f(i) = a^i$. Since $a \in H$, we see from Definition 2.1.5 that f is well defined. The Pigeonhole Principle (Exercise 1.1.11) implies that there exists a pair $0 < i < j \leq n + 1$ such that $a^i = a^j$. Then $j - i > 0$, so $e = a^{j-i}$ is in H. If $j - i = 1$, then $a = e$, which implies $a^{-1} = e \in H$. If $j - i > 1$, then $e = a^{j-i} = aa^{j-i-1}$, which implies $a^{-1} = a^{j-i-1} \in H$. □

LEMMA 2.2.5. *Let G be a group and $X \subseteq G$. Let $\mathcal{S} = \{H \leq G \mid X \subseteq H\}$, and let*

$$\langle X \rangle = \bigcap_{H \in \mathcal{S}} H$$

be the intersection of all subgroups of G containing X. Then the following are true.

(1) $\langle X\rangle$ *is the smallest subgroup of* G *containing* X*.*
(2) $\langle X\rangle$ *is the trivial subgroup* $\{e\}$ *if* $X = \emptyset$*, otherwise*

$$\langle X\rangle = \{x_1^{e_1}\cdots x_n^{e_n} \mid n \geq 1, e_i \in \mathbb{Z}, x_i \in X\}.$$

PROOF. (1): We know $\mathcal{S}$ is nonempty because $G \in \mathcal{S}$. Therefore, (1) follows straight from Exercise 2.2.23.

(2): If $X = \emptyset$, then $\{e\} \in \mathcal{S}$, so $\langle X\rangle = \{e\}$. Assume $X \neq \emptyset$. By Lemma 2.2.2, the set $S = \{x_1^{e_1}\cdots x_n^{e_n} \mid n \geq 1, e_i \in \mathbb{Z}, x_i \in X\}$ is a subgroup of G. Since $X \subseteq S$, we have $\langle X\rangle \subseteq S$. Let $x_1^{e_1}\cdots x_n^{e_n}$ be a typical element of S. For each i, $x_i \in X$ implies x_i is in the group $\langle X\rangle$. By Definition 2.1.5, the power $x_i^{e_i}$ is in $\langle X\rangle$. Therefore, the product $x_1^{e_1}\cdots x_n^{e_n}$ is in $\langle X\rangle$. This proves $S \subseteq X$. □

DEFINITION 2.2.6. In the context of Lemma 2.2.5, the set $\langle X\rangle$ is called the *subgroup of* G *generated by* X. If $X = \{x_1, \ldots, x_n\}$ is a finite subset of G, then we sometimes write $\langle X\rangle$ in the form $\langle x_1, \ldots, x_n\rangle$. A subgroup $H \leq G$ is said to be *finitely generated* if there exists a finite subset $\{x_1, \ldots, x_n\} \subseteq H$ such that $H = \langle x_1, \ldots, x_n\rangle$. We say H is *cyclic* if $H = \langle x\rangle$ for some $x \in H$.

PROPOSITION 2.2.7. *The set of all subgroups of* G*, ordered by set inclusion, is a lattice. We call this partially ordered set the* subgroup lattice of G*.*

PROOF. Let A and B be subgroups of G. The subgroup generated by the set $A \cup B$ is the least subgroup of G that contains both A and B. By Exercise 2.2.23, $A \cap B$ is the largest subgroup of G that is contained in both A and B. □

DEFINITION 2.2.8. Let G be a group and H a subgroup of G. If x and y are elements of G, then we say x *is congruent to* y *modulo* H if $x^{-1}y \in H$. In this case we write $x \equiv y \pmod{H}$.

LEMMA 2.2.9. *Let* G *be a group and* H *a subgroup. Then congruence modulo* H *is an equivalence relation on* G*.*

PROOF. If $x \in G$, then $x^{-1}x = e \in H$, so $x \equiv x \pmod{H}$. Assume $x \equiv y \pmod{H}$. Then $x^{-1}y \in H$, which implies $y^{-1}x = (x^{-1}y)^{-1} \in H$, hence $y \equiv x \pmod{H}$. Assume $x \equiv y \pmod{H}$ and $y \equiv z \pmod{H}$. Then $x^{-1}yy^{-1}z = x^{-1}z \in H$, which implies $x \equiv z \pmod{H}$. □

The following notation will be used frequently in our study of groups.

DEFINITION 2.2.10. Let G be a group and assume A and B are nonempty subsets of G. If G is a multiplicative group, then

$$AB = \{xy \mid x \in A, y \in B\}.$$

In case A or B is a singleton set, we write xB or Ay instead of $\{x\}B$ or $A\{y\}$. If G is an additive group, then

$$A + B = \{x + y \mid x \in A, y \in B\}.$$

In case A or B is a singleton set, we write $x + B$ or $A + y$ instead of $\{x\} + B$ or $A + \{y\}$.

LEMMA 2.2.11. *Let* G *be a group,* H *a subgroup, and* $x, y \in G$*. The following are equivalent.*
(1) $x \equiv y \pmod{H}$*.*
(2) $y = xh$ *for some* $h \in H$*.*

(3) $xH = yH$.

Under congruence modulo H*, the equivalence class of* x *is* xH.

PROOF. (1) is equivalent to (2): We have $x \equiv y \pmod{H}$ if and only if $x^{-1}y \in H$ which is true if and only if $x^{-1}y = h$ for some $h \in H$ which is equivalent to $y = xh$ for some $h \in H$.

(3) implies (2): We have $y = ye \in yH = xH$. Therefore, $y = xh$ for some $h \in H$.

(2) implies (3): Suppose $y = xh$, for some $h \in H$. For every $z \in H$, $yz = x(hz) \in xH$. Hence $yH \subseteq xH$. Also, $x = yh^{-1}$ implies $xz = y(h^{-1}z) \in yH$, which implies $xH \subseteq yH$.

By (1), (2) and (3) above, the equivalence class of x modulo H is $\{y \in G \mid y \equiv x \pmod{H}\} = \{y \in G \mid y = xh \text{ for some } h \in H\} = xH$, which proves the last statement. □

2.2. Cosets and Lagrange's Theorem. Let G be a group and H a subgroup. By Lemma 2.2.9, congruence modulo H is an equivalence relation on G. Therefore G is partitioned into equivalence classes. If $x \in G$, then by Lemma 2.2.11, the equivalence class of x is xH. The set xH is called *the left coset of* x *modulo* H. The set of all left cosets of G modulo H is $G/H = \{xH \mid x \in G\}$. By Proposition 1.1.2 two cosets are either disjoint or equal as sets. The *index of* H *in* G is the cardinality of the set G/H and is denoted $[G : H]$.

There is a right hand version of the above, which we will briefly describe here. We say x *is right congruent to* y *modulo* H if $yx^{-1} \in H$. This defines an equivalence relation on G. The equivalence class of x is the set Hx which is called the *right coset of* x *modulo* H. The set of all right cosets is denoted $H\backslash G$. In general, the partitions G/H and $H\backslash G$ are not equal. That is, a left coset is not necessarily a right coset (see Lemma 2.3.5). In Exercise 2.2.25 the reader is asked to show that there is a one-to-one correspondence between G/H and $H\backslash G$. That is, G/H and $H\backslash G$ both have cardinality equal to $[G : H]$.

LEMMA 2.2.12. *Let* G *be a group and* $H \leq G$*. Given* $x, y \in G$ *there is a one-to-one correspondence* $\phi : xH \to yH$ *defined by* $\phi(z) = (yx^{-1})z$*. If* $|H|$ *is finite, then all left cosets of* H *have the same number of elements.*

PROOF. For any $h \in H$, $yx^{-1}xh = yh \in yH$. We see that ϕ is a well defined function. The function $\psi(w) = xy^{-1}w$ is the inverse to ϕ. Applying Proposition 1.1.1, it follows that ϕ is a one-to-one correspondence. □

If H is a subgroup of G, then *a complete set of left coset representatives for* H *in* G is a subset $\{a_i \mid i \in I\}$ of G where we have exactly one element from each left coset. The index set I can be taken to be G/H. If $\{a_i \mid i \in I\}$ is a complete set of left coset representatives, then $G = \bigcup_{i \in I} a_i H$ is a partition of G. For example, if $m \geq 1$, then Proposition 1.2.9 (2) shows that $\{0, 1, \ldots, m-1\}$ is a complete set of left coset representatives for $\langle m \rangle$ in $\mathbb{Z}$.

THEOREM 2.2.13. *If* $K \leq H \leq G$*, then* $[G : K] = [G : H][H : K]$*. If two of the three indices are finite, then so is the third.*

PROOF. Let $\{a_i \mid i \in I\}$ be a complete set of left coset representatives for H in G and Let $\{b_j \mid j \in J\}$ be a complete set of left coset representatives for K in H.

Then $G = \bigcup_{i\in I} a_i H$ is a partition of G and $H = \bigcup_{j\in J} b_j K$ is a partition of H. So

$$\begin{aligned} G &= \bigcup_{i\in I} a_i H \\ &= \bigcup_{i\in I} a_i \left(\bigcup_{j\in J} b_j K \right) \\ &= \bigcup_{i\in I} \left(\bigcup_{j\in J} a_i b_j K \right). \end{aligned}$$

To finish the proof, we show that $\{a_i b_j \mid (i,j) \in I \times J\}$ is a complete set of left coset representatives for K in G. It suffices to show the cosets $a_i b_j K$ are pairwise disjoint. Assume $a_i b_j K = a_s b_t K$. Then $a_i b_j = a_s b_t k$ for some $k \in K$. Recall that b_j, b_t, k are in H. Then we have $a_i = a_s h$, for some $h \in H$. Hence $a_i H = a_s H$, which implies $i = s$. Canceling, we get $b_j = b_t k$, or $b_j K = b_t K$, which implies $j = t$. This proves $[G : K] = [G : H][H : K]$. The index $[G : K]$ is infinite if and only if $[G : H]$ is infinite or $[H : K]$ is infinite. This proves the theorem. □

COROLLARY 2.2.14. *(Lagrange's Theorem) If G is a group and $H \leq G$, then $|G| = [G : H]|H|$.*

PROOF. Apply Theorem 2.2.13 with $K = \langle e \rangle$. □

2.3. A Counting Theorem.

LEMMA 2.2.15. *Let G be a group containing subgroups H and K. Then HK is a subgroup of G if and only if $HK = KH$.*

PROOF. See Definition 2.2.10 for the definition of the set HK. First assume $HK = KH$. To show HK is a subgroup we show that the criteria of Lemma 2.2.2 (1) are satisfied. In the following, h, h_1, h_2, h_3 denote elements of H and k, k_1, k_2, k_3 denote elements of K. Let $h_1 k_1$ and $h_2 k_2$ be arbitrary elements of HK. Since $HK = KH$, there exist h_3, k_3 such that $k_1 h_2 = h_3 k_3$. Now $(h_1 k_1)(h_2 k_2) = h_1(k_1 h_2)k_2 = h_1(h_3 k_3)k_2 = (h_1 h_3)(k_3 k_2)$ is an element of HK. By Lemma 2.1.7, $(hk)^{-1} = k^{-1}h^{-1}$ is is an element of $KH = HK$. This proves HK is a subgroup.

Conversely, suppose HK is a subgroup. Consider the function $i : G \to G$ defined by $i(x) = x^{-1}$. By Lemma 2.1.7, i^2 is the identity function. Thus i is a one-to-one correspondence. Since HK is a group, the restriction of i to HK is a one-to-one correspondence. That is, $i(HK) = HK$. If $hk \in HK$, then $i(hk) = (hk)^{-1} = k^{-1}h^{-1}$ is in KH, which shows $HK = i(HK) \subseteq KH$. Consider $kh \in KH$. Then $i(kh) = (kh)^{-1} = h^{-1}k^{-1}$ is in HK. Therefore, kh is the inverse of an element in the subgroup HK. By Lemma 2.2.2, $kh \in HK$, which implies $KH \subseteq HK$. □

THEOREM 2.2.16. *Let G be a group. If H and K are finite subgroups of G, then*

$$|HK| = \frac{|H||K|}{|H \cap K|}.$$

PROOF. We do not assume HK is a group. Let $C = H \cap K$. Then C is a subgroup of H. Let $\{h_1, \ldots, h_n\}$ be a full set of left coset representatives of C in

H, where $n = [H : C]$. Then $H = \bigcup_{i=1}^{n} h_i C$ is a disjoint union. Since $C \subseteq K$, by Proposition 2.1.6 we have $CK = K$, hence

$$HK = \bigcup_{i=1}^{n} h_i CK = \bigcup_{i=1}^{n} h_i K.$$

The last union is a disjoint union. To see this, suppose $h_i K = h_j K$. Then $h_j^{-1} h_i \in H \cap K = C$, which implies $i = j$. By Lemma 2.2.12 we can now count the cardinality of HK:

$$|HK| = \sum_{i=1}^{n} |K| = n|K| = [H : H \cap K]|K|.$$

By Corollary 2.2.14, we are done. □

2.4. Cyclic Subgroups. In the next theorem we show that the additive group $\mathbb{Z}$ is cyclic and every subgroup is of the form $\langle n \rangle$ for some $n \geq 0$. Moreover, the equivalence relation of Definition 2.2.8 defined in terms of the subgroup $\langle n \rangle$ is equal to the equivalence relation of Definition 1.2.8 defined in terms of divisibility by n.

THEOREM 2.2.17. *Let $\mathbb{Z}$ be the additive group of integers.*

(1) Every subgroup of $\mathbb{Z}$ is cyclic. The trivial subgroups of $\mathbb{Z}$ are: $\langle 0 \rangle$ and $\mathbb{Z} = \langle 1 \rangle$. If H is a nontrivial subgroup, then there is a unique $n > 1$ such that $H = \langle n \rangle = n\mathbb{Z} = \{nk \mid k \in \mathbb{Z}\}$.

(2) If $n \geq 1$ and $H = \langle n \rangle$, then $x \equiv y \pmod{H}$ if and only if $x \equiv y \pmod{n}$. That is, the coset $x + \langle n \rangle$ in $\mathbb{Z}/\langle n \rangle$ is equal to the congruence class $[x]$ in $\mathbb{Z}/n$.

PROOF. (1): Let $H \leq \mathbb{Z}$ and assume $H \neq \langle 0 \rangle$. If $x \in H - \langle 0 \rangle$, then so is $-x$. By the Well Ordering Principle (Axiom 1.2.1) there is a least positive integer in H, say n. We prove that $H = n\mathbb{Z}$. Let $x \in H$. By the Division Algorithm (Proposition 1.2.3) we can write $x = nq + r$ where $0 \leq r < n$. By Definition 2.1.5, $nq \in H$. Therefore, $r = x - nq$ is in H. By the choice of n, this implies $r = 0$. Hence $x \in n\mathbb{Z}$.

(2): This follows from the fact that $x - y \in H$ if and only if n divides $x - y$. □

Let G be a group and a an element of finite order in G. Recall (Definition 2.1.10) that the order of a, written $|a|$, is the least positive integer m such that $a^m = e$.

LEMMA 2.2.18. *Let G be a group, $a \in G$, and assume $|a| = m$ is finite. Then the following are true.*

(1) $|a| = |\langle a \rangle|$.

(2) $\langle a \rangle = \{e, a, a^2, \ldots, a^{m-1}\}$.

(3) For each $n \in \mathbb{Z}$, $a^n = e$ if and only if m divides n.

(4) For each $n \in \mathbb{Z}$, $|a^n| = m/\gcd(m, n)$.

(5) Let $b \in G$. Assume $|b| = n$ is finite, $ab = ba$, and $\langle a \rangle \cap \langle b \rangle = \langle e \rangle$. Then $|ab| = \operatorname{lcm}(m, n)$.

PROOF. (1) and (2): Let $m = |a|$. Then $m > 0$, $a^m = e$, and if $m > 1$, then $a^{m-1} \neq e$. Let $n \in \mathbb{Z}$. Applying Proposition 1.2.3, there exist unique integers q and r such that $n = mq + r$ and $0 \leq r < m$. Then $a^n = (a^m)qa^r = a^r$. Therefore, $\langle a \rangle = \{e, a, a^2, \ldots, a^{m-1}\}$. It follows that $|\langle a \rangle| = m$.

(3): First assume $n = mq$. Then we have $a^{mq} = (a^m)^q = e^q = e$. Conversely assume $a^n = e$. By Parts (1) and (2), if $n = mq + r$ and $0 \leq r < m$, then $a^r = e$, which implies $r = 0$.

(4) and (5): This part of the proof is Exercise 2.2.29. □

COROLLARY 2.2.19. *If $|G|$ is finite, and $a \in G$, then the following are true.*

(1) $|a|$ is finite.
(2) $|a|$ divides $|G|$.
(3) $a^{|G|} = e$.

PROOF. (1): Proposition 2.2.4 shows that $|a|$ is finite.
(2) and (3): These follow immediately from Lemma 2.2.18 and Corollary 2.2.14. □

COROLLARY 2.2.20. *Let $a \in \mathbb{Z}$. Then the following are true.*

(1) (Euler) If $m \in \mathbb{N}$ and $\gcd(a, m) = 1$, then $a^{\phi(m)} \equiv 1 \pmod{m}$.
(2) (Fermat) If p is prime and $\gcd(a, p) = 1$, then $a^{p-1} \equiv 1 \pmod{p}$.

PROOF. As noted in Example 2.1.3, U_n, the group of units modulo n, has order $\phi(n)$. If p is prime, then $\phi(p) = p - 1$. □

COROLLARY 2.2.21. *Let G be a group satisfying $|G| > 1$. If G has no proper subgroup, then $|G|$ is finite, $|G|$ is prime, and G is cyclic.*

PROOF. Let $a \in G - \langle e \rangle$. Since G has no proper subgroup and $\langle e \rangle \neq \langle a \rangle$ is a subgroup of G, we have $\langle a \rangle = G$. Look at the set $S = \{e, a, a^2, \dots\}$. If there is a relation of the form $a^k = a^m$, where $k < m$, then $|a|$ is finite, hence G is finite. Conversely, if G is finite, then Proposition 2.2.4 shows that there is a relation $a^k = a^m$, where $k < m$. Assume for contradiction's sake that G is infinite. Then $a \neq a^n$, for all $n > 1$. Thus, $\langle a^2 \rangle$ is a proper subgroup of G, a contradiction. We conclude that $G = \langle a \rangle = \{e, a, \dots, a^{n-1}\}$ is a finite cyclic group of order n, for some n. Assume for contradiction's sake that $n = xy$ where $1 < x \leq y < n$. By Lemma 2.2.18 (4), $\langle a^x \rangle = \{e, a^x, a^{2x}, \dots, a^{(y-1)x}\}$ has order y, hence G has a proper subgroup, which is a contradiction. This proves n is prime. □

COROLLARY 2.2.22. *Let G be a group. If G has only a finite number of subgroups, then G is finite.*

PROOF. Suppose G is an infinite group. We prove that G has infinitely many subgroups. Let $x_1 \in G$ and set $X_1 = \langle x_1 \rangle$. By Theorem 2.2.17, the additive group of integers $\mathbb{Z}$ has infinitely many distinct subgroups, namely $\{\langle n \rangle \mid n \geq 0\}$. If X_1 is infinite, then the same proof shows that X_1 has infinitely many distinct subgroups, namely $\{\langle x_1^n \rangle \mid n \geq 0\}$. From now on assume every element of G has finite order. Then $G - \langle x_1 \rangle$ is infinite. Pick $x_2 \in G - \langle x_1 \rangle$. Then $\langle x_1 \rangle \neq \langle x_2 \rangle$. Assume inductively that $n \geq 1$ and $x_1, x_2, \dots, x_n$ are in G such that $X_1 = \langle x_1 \rangle, \dots, X_n = \langle x_n \rangle$ are n distinct subgroups. Then $\bigcup_{i=1}^{n} X_i$ is finite. Pick $x_{n+1} \in G - X_1 - X_2 - \cdots - X_n$ and set $X_{n+1} = \langle x_{n+1} \rangle$. Then by induction there exists an infinite collection $\{X_i \mid i \geq 1\}$ of distinct subgroups of G. □

2.5. Exercises.

EXERCISE 2.2.23. (An intersection of subgroups is a subgroup.) Let G be a group, I a nonempty set, and $\{H_i \mid i \in I\}$ a family of subgroups of G indexed by I. Show that

$$\bigcap_{i \in I} H_i$$

is a subgroup of G.

EXERCISE 2.2.24. Let G be a group and X, Y, Z subgroups of G. Prove that if $Y \subseteq X$, then $X \cap YZ = Y(X \cap Z)$.

EXERCISE 2.2.25. Let G be a group and H a subgroup of G. We denote by G/H the set of all left cosets of H in G, and by $H\backslash G$ the set of all right cosets of H in G. Show that the assignment $xH \mapsto Hx^{-1}$ defines a one-to-one correspondence between G/H and $H\backslash G$.

EXERCISE 2.2.26. Let G be a group containing finite subgroups H and K. If $|H|$ and $|K|$ are relatively prime, show that $H \cap K = \langle e \rangle$.

EXERCISE 2.2.27. This exercise is a continuation of Exercise 2.1.27. Let K and H be groups and $K \times H$ the product group. Show that $K \times \langle e \rangle = \{(x, e) \mid x \in K\}$ and $\langle e \rangle \times H = \{(e, y) \mid y \in H\}$ are normal subgroups of $K \times H$.

EXERCISE 2.2.28. Consider the symmetric group S_3 of order 6. Show that S_3 has 4 proper subgroups. Let H be the subgroup of order 2 generated by the transposition (12). Compute the three left cosets of H and the three right cosets of H.

EXERCISE 2.2.29. Prove Parts (4) and (5) of Lemma 2.2.18.

EXERCISE 2.2.30. Let p be a prime number and G a finite group of order p. Prove:

(1) G has no proper subgroup.
(2) There exists $a \in G$ such that $G = \langle a \rangle$.
(3) G is abelian.

EXERCISE 2.2.31. Let $(\mathbb{R}, +)$ denote the additive group on $\mathbb{R}$. Then $(\mathbb{Q}, +)$ is a subgroup of $(\mathbb{R}, +)$ and $(\mathbb{Z}, +)$ is a cyclic subgroup of both $(\mathbb{Q}, +)$ and $(\mathbb{R}, +)$. Show that the set $\{x \in \mathbb{R} \mid 0 \leq x < 1\}$ is a complete set of left coset representatives for $\mathbb{Z}$ in $\mathbb{R}$. Show that the set $\{x \in \mathbb{Q} \mid 0 \leq x < 1\}$ is a complete set of left coset representatives for $\mathbb{Z}$ in $\mathbb{Q}$. See Exercise 2.3.23 for a continuation of this exercise.

EXERCISE 2.2.32. Let G be a finite group of order m and $a \in G$. Let $n \in \mathbb{Z}$ and assume m and n are relatively prime. Show that there exists $b \in G$ such that $a = b^n$.

EXERCISE 2.2.33. Let G be a finite group and a, b elements of G such that $|a| = 2$, $|b| = 2$, and $ab \neq ba$. Show that the subgroup of G generated by a and b is isomorphic to a dihedral group D_n for some $n > 2$ (see Example 2.1.17).

3. Homomorphisms and Normal Subgroups

A homomorphism of groups is a function $\phi : G \to G'$ from a group G to a group G' which preserves the binary operations on G and G'. In other words, the binary operation on the image of ϕ agrees with the binary operation on G. Important properties of the groups G and G' are studied in terms of the homomorphism ϕ. The function ϕ induces a binary relation on G and the fibers of the map are the equivalence classes. If e' is the identity element of G', then the fiber $\phi^{-1}(e')$ is called the kernel of ϕ. The kernel of ϕ is denoted $\ker(\phi)$ and is a subgroup of G. The image of ϕ is a subgroup of G'. Given a group G and subgroup H, left congruence modulo H is an equivalence relation on G (Lemma 2.2.9). Equivalence classes

are called left cosets and the set of left cosets is denoted G/H. There is a natural surjection $\eta : G \to G/H$. If the binary operation on G induces a group structure on the set of cosets modulo H, then H is called a normal subgroup. The kernel of a homomorphism ϕ is a normal subgroup. Theorem 2.3.12 is of fundamental importance and says that ϕ factors in a natural way into the onto homomorphism $\eta : G \to G/\ker(\phi)$ followed by a one-to-one homomorphism $G/\ker(\phi) \to G'$. Therefore, the homomorphic images of G correspond to the quotient groups G/H where H is a normal subgroup of G.

3.1. Definition and First Properties of Normal Subgroups. A function from one group to another that preserves the binary operations is called a homomorphism. After stating the formal definition of a homomorphism, some of the first properties are established. For instance, the homomorphic image of the identity element is the identity element, the inverse of a homomorphic image of an element is the homomorphic image of the inverse, the homomorphic image of a subgroup is a subgroup.

If H is a subgroup of G, then H is a normal subgroup if and only if the binary operation on G turns the set of left cosets G/H into a group and in this case the natural map $G \to G/H$ is a homomorphism of groups (Lemma 2.3.5).

DEFINITION 2.3.1. A *homomorphism of groups* is a function $\phi : G \to G'$ from a group G to a group G' such that $\phi(xy) = \phi(x)\phi(y)$ for all $x, y \in G$. If ϕ is onto, we say ϕ is an *epimorphism*. If ϕ is one-to-one, we say ϕ is a *monomorphism*. If ϕ is one-to-one and onto, then as in Definition 2.1.11 we say ϕ is an *isomorphism*. A homomorphism from G to G is called an *endomorphism of* G. An isomorphism from G to G is called an *automorphism of* G.

DEFINITION 2.3.2. Let $\phi : G \to G'$ be a homomorphism of groups. The *kernel of* ϕ is $\ker(\phi) = \{x \in G \mid \phi(x) = e\}$. In Lemma 2.3.3 (4) we prove that $\ker(\phi)$ is a subgroup of G.

LEMMA 2.3.3. *If $f : G \to G'$ is a homomorphism of groups, then the following are true.*

(1) $f(e) = e$.
(2) For each $x \in G$, $f(x^{-1}) = f(x)^{-1}$.
(3) If H is a subgroup of G, then $f(H)$ is a subgroup of G'. If there is a containment relation $H_1 \subseteq H_2$, then $f(H_1) \subseteq f(H_2)$.
(4) If H' is a subgroup of G', then $f^{-1}(H')$ is a subgroup of G. If there is a containment relation $H_1' \subseteq H_2'$, then $f^{-1}(H_1') \subseteq f^{-1}(H_2')$. In particular, $\ker f$ *is a subgroup of $f^{-1}(H')$.*

PROOF. (1): $f(e) = f(ee) = f(e)f(e)$. By Lemma 2.1.7 (1), $f(e) = e$.

(2): By (1), we have $e = f(e) = f(xx^{-1}) = f(x)f(x^{-1})$. By Lemma 2.1.7 (2), $f(x^{-1}) = f(x)^{-1}$.

(3): Let x and y be arbitrary elements of H. By (2), we have $f(xy^{-1}) = f(x)f(y)^{-1}$. By Lemma 2.2.2, this proves $f(H)$ is a subgroup of G'. The second statement is left to the reader.

(4): Let x and y be arbitrary elements of G such that $f(x) \in H'$ and $f(y) \in H'$. Then $f(xy^{-1}) = f(x)f(y)^{-1} \in H'$. By Lemma 2.2.2, this proves $f^{-1}(H')$ is a subgroup of G. The second statement is left to the reader. Since $\langle e \rangle$ is a subgroup of H', $\ker f = f^{-1}(e)$ is a subgroup of $f^{-1}(H')$. □

DEFINITION 2.3.4. Let G be a group. For every $a \in G$, let $\alpha_a : G \to G$ be defined by $\alpha_a(x) = a^{-1}xa$. If X is a nonempty subset of G, then $\alpha_a(X) = a^{-1}Xa$ is called the *conjugate of X by a*.

The next lemma lists the fundamental properties of normal subgroups. The definition follows the lemma.

LEMMA 2.3.5. *Let G be a group and H a subgroup of G. The following are equivalent.*

(1) For each $x \in G$, $x^{-1}Hx \subseteq H$.
(2) For each $x \in G$, $x^{-1}Hx = H$.
(3) For each $x \in G$ there exists $y \in G$ such that $xH = Hy$.
(4) For each $x \in G$, $xH = Hx$.
(5) For each $x \in G$ and $y \in G$, $xHyH = xyH$.
(6) There is a well defined binary operation $G/H \times G/H \to G/H$ on G/H defined by the rule $(xH, yH) \mapsto xyH$.
(7) There is a binary operation on G/H such that the natural map $\eta : G \to G/H$ is a homomorphism of groups.
(8) There exists a group G' and a homomorphism of groups $\theta : G \to G'$ such that $H = \ker\theta$.

PROOF. (1) implies (2): Let $x \in G$. First apply (1) to x, yielding $x^{-1}Hx \subseteq H$. Now conjugate by x^{-1} and apply (1) with x^{-1} to get $H = (xx^{-1})H(xx^{-1}) \subseteq xHx^{-1} \subseteq H$.

(2) implies (3): Let $x \in G$. Apply (2) to x^{-1} to get $xHx^{-1} = H$. This implies $xH = Hx$.

(3) implies (4): Given $x \in G$, there exists $y \in G$ such that $xH = Hy$. Since x is in $xH = Hy$, this implies $x = hy$ for some $h \in H$. Therefore $y = h^{-1}x$ and $Hy = Hh^{-1}x = Hx$.

(4) implies (5): Let $x \in G$ and $y \in G$. By (4) applied to y, $yH = Hy$. Therefore, $xHyH = x(Hy)H = x(yH)H = xyH$.

(5) implies (6): This is immediate.

(6) implies (7): By (6), $(xH, yH) \mapsto xyH$ defines a binary operation on G/H. The associative law on G implies the associative law also holds on G/H. The identity element is the coset eH and $(xH)^{-1} = x^{-1}H$. Therefore G/H is a group and it is now clear that the natural map $\eta : G \to G/H$ is a homomorphism.

(7) implies (8): The kernel of $\eta : G \to G/H$ is $\eta^{-1}(eH) = H$.

(8) implies (1): Let $\theta : G \to G'$ be a homomorphism of groups and assume $H = \ker\theta$. By Lemma 2.3.3 (4), $\ker(\theta) = \theta^{-1}(\langle e\rangle)$ is a subgroup of G. Given $x \in G$ and $h \in H$ we have $\theta(h) = e$. Hence $\theta(x^{-1}hx) = \theta(x)^{-1}\theta(h)\theta(x) = \theta(x)^{-1}\theta(x) = e$. Therefore, $x^{-1}Hx \subseteq \ker\theta = H$. □

DEFINITION 2.3.6. If G is a group and H is a subgroup of G satisfying any of the equivalent conditions of Lemma 2.3.5, then we say H is a *normal subgroup of G*. The group G/H is called the *quotient group*, or *factor group*. If N is a normal subgroup of G, we sometimes write $N \trianglelefteq G$.

EXAMPLE 2.3.7. Let G be a group.

(1) The trivial subgroups $\langle e\rangle$ and G are normal in G.
(2) If G is abelian and H is a subgroup of G, then for every $x \in G$, $xH = Hx$, hence H is normal. The quotient group G/H is abelian because G is abelian.

3.2. The Isomorphism Theorems. Theorem 2.3.12, says that any homomorphism of groups $\theta : A \to B$ factors in a natural way into a surjection $A \to A/\ker(\theta)$ followed by an injection $A/\ker(\theta) \to B$. This provides us with a valuable tool for defining a homomorphism on a quotient group A/N. As applications, we prove the Isomorphism Theorems (Theorem 2.3.14) and the Correspondence Theorem (Theorem 2.3.15).

LEMMA 2.3.8. *Let $\phi : G \to G'$ and $\phi_1 : G' \to G''$ be homomorphisms of groups. Then the following are true.*

(1) The composite $\phi_1\phi : G \to G''$ is a homomorphism of groups.
(2) The kernel of ϕ, $\ker(\phi)$, is a normal subgroup of G.
(3) The function ϕ is one-to-one if and only if $\ker(\phi) = \langle e \rangle$.

PROOF. (1): This follows straight from:

$$\phi_1\phi(xy) = \phi_1(\phi(x)\phi(y)) = \phi_1\phi(x)\phi_1\phi(y).$$

(2): By Lemma 2.3.5 (8), $\ker(\phi)$ is a normal subgroup of G.

(3): If ϕ is one-to-one, then $\ker(\phi) = \phi^{-1}(\langle e \rangle) = \langle e \rangle$. If $\ker(\phi) = \langle e \rangle$ and $\phi(x) = \phi(y)$, then $\phi(xy^{-1}) = \phi(x)\phi(y)^{-1} = e$, so $xy^{-1} \in \ker(\phi)$. Therefore, $x = y$ and ϕ is one-to-one. □

EXAMPLE 2.3.9. If $\phi : G \to G'$ is an isomorphism of groups, then as in Definition 2.1.11 we say G is isomorphic to G', and write $G \cong G'$. If $\phi_1 : G' \to G''$ is another isomorphism of groups, then by Lemma 2.3.8 and Exercise 1.1.9, the composite $\phi_1\phi$ is an isomorphism. The reader should verify that isomorphism defines an equivalence relation on the set of all groups.

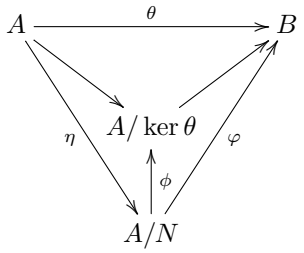
EXAMPLE 2.3.10. Let G be a group. The set of all automorphisms of G is denoted $\mathrm{Aut}(G)$. By Lemma 2.3.8 the composition of automorphisms is an automorphism. In the notation of Example 2.1.2, $\mathrm{Aut}(G)$ is a subgroup of $\mathrm{Perm}(G)$.

DEFINITION 2.3.11. Let G be a group and $a \in G$. As in Definition 2.3.4, conjugation by a defines the function $\alpha_a : G \to G$, where $\alpha_a(x) = a^{-1}xa$. In Exercise 2.3.21 the reader is asked to prove that α_a is an automorphism of G. We call α_a the *inner automorphism of G defined by a*. The set of all inner automorphisms is a subgroup of $\mathrm{Aut}(G)$.

THEOREM 2.3.12. *Let $\theta : A \to B$ be a homomorphism of groups. Let N be a normal subgroup of A contained in $\ker\theta$. There exists a homomorphism $\varphi : A/N \to B$ satisfying the following.*

(1) $\varphi(aN) = \theta(a)$, or in other words $\theta = \varphi\eta$.
(2) φ is the unique homomorphism from $A/N \to B$ such that $\theta = \varphi\eta$.
(3) $\mathrm{im}\,\theta = \mathrm{im}\,\varphi$.
(4) $\ker\varphi = \eta(\ker\theta) = \ker(\theta)/N$.
(5) φ is one-to-one if and only if $N = \ker\theta$.
(6) φ is onto if and only if θ is onto.

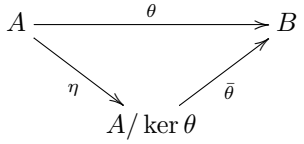
(7) There is a unique epimorphism $\phi : A/N \to A/\ker\theta$ such that the diagram



commutes.

PROOF. The map φ exists by Exercise 1.1.13. The proofs of (1) – (6) are left as an exercise for the reader. Part (7) results from an application of Parts (1) – (6) to the natural map $A \to A/\ker\theta$. □

COROLLARY 2.3.13. *If $\theta : A \to B$ is a homomorphism of groups, then there exists a unique monomorphism $\bar{\theta}$ such that $\theta = \bar{\theta}\eta$. Hence θ factors into an epimorphism η followed by a monomorphism $\bar{\theta}$ and the diagram*



commutes.

PROOF. This is Theorem 2.3.12 (5). □

An important application of Theorem 2.3.12 is the definition of a homomorphism on a factor group. It is worth stressing this subtle but useful tool. Suppose A and B are groups, N is a normal subgroup of G, and one wishes to define a homomorphism $\varphi : A/N \to B$. By Theorem 2.3.12, to do this it suffices to define a homomorphism $\theta : A \to B$ and show that $N \subseteq \ker\theta$. To reiterate, the function is defined on the top group A instead of on the factor group A/N. Typically this is the preferred approach because it immediately implies that the function φ is well defined on A/N which consists of cosets of N, not elements of A. This is the method of proof used below in the proof of Theorem 2.3.14.

THEOREM 2.3.14. *(The Isomorphism Theorems) Let G be a group.*

(1) If $\theta : G \to G'$ is a homomorphism of groups, then the map $\varphi : G/\ker\theta \to \operatorname{im}\theta$ sending the coset $x\ker\theta$ to $\theta(x)$ is an isomorphism of groups.

(2) If A and B are subgroups of G and B is normal, then the natural map

$$\frac{A}{A \cap B} \to \frac{AB}{B}$$

sending the coset $x(A \cap B)$ to the coset xB is an isomorphism of groups.

(3) If A and B are normal subgroups of G and $A \subseteq B$, then B/A is a normal subgroup of G/A and the natural map

$$\frac{G/A}{B/A} \to G/B$$

sending the coset containing xA to the coset xB is an isomorphism of groups.

PROOF. (1): By Lemma 2.3.3 (3), the image of G is a subgroup of G'. This is Parts (e) and (f) of Theorem 2.3.12.

(2): By Exercise 2.3.20, AB is a group, B is normal in AB, and $A \cap B$ is normal in A. Let $f : A \to (AB)/B$ be the set containment map $A \to AB$ followed by the natural map $AB \to (AB)/B$. If $a \in A$ and $b \in B$, then $abB = aB$, hence f is onto. Let $a \in A$. Then $aB = B$ if and only if $a \in B$. Therefore the kernel of f is $A \cap B$. Part (2) follows from Part (1) applied to the homomorphism f.

(3): By Theorem 2.3.12 (7) applied to the natural map $G \to G/B$, there is a natural epimorphism $\phi : G/A \to G/B$ defined by $\phi(xA) = xB$. The kernel of ϕ consists of those cosets xA such that $x \in B$. That is, $\ker \phi = B/A$. Part (3) follows from Part (1) applied to the homomorphism ϕ. □

THEOREM 2.3.15. *(The Correspondence Theorem) Let G be a group and A a normal subgroup of G. There is a one-to-one order-preserving correspondence between the subgroups B such that $A \subseteq B \subseteq G$ and the subgroups of G/A given by $B \mapsto B/A$. Moreover, B is a normal subgroup of G if and only if B/A is a normal subgroup of G/A.*

PROOF. Let $\eta : G \to G/A$ be the natural homomorphism. By Lemma 2.3.3, if B is a subgroup of G, then $\eta(B)$ is a subgroup of G/A, and if H is a subgroup of G/A, then $\eta^{-1}(H)$ is a subgroup of G containing A. If $B_1 \subseteq B_2$, then $\eta(B_1) \subseteq \eta(B_2)$. Likewise, if $H_1 \subseteq H_2$, then $\eta^{-1}(H_1) \subseteq \eta^{-1}(H_2)$. Since η is onto, $\eta\eta^{-1}(H) = H$. By Exercise 2.3.17, if B is a subgroup of G containing A, then $B = \eta^{-1}\eta(B)$. This proves the first claim.

For the last claim, let B be a subgroup of G containing A. If B is normal, then by Theorem 2.3.14 (3), $\eta(B)$ is normal in G/A. Conversely assume $\eta(B)$ is normal in G/A. Then B is equal to the kernel of the composite map $G \to G/A \to (G/A)/\eta(B)$, hence is normal in G. □

EXAMPLE 2.3.16. Let $(\mathbb{R}, +)$ be the additive abelian group of real numbers and $(\mathbb{R}_{>0}, \cdot)$ the multiplicative abelian group of positive real numbers. Define $\phi : (\mathbb{R}, +) \to (\mathbb{R}_{>0}, \cdot)$ by $\phi(x) = e^x$. Then $\phi(x + y) = e^{x+y} = e^x e^y = \phi(x)\phi(y)$, so ϕ is a homomorphism. Define $\psi : (\mathbb{R}_{>0}, \cdot) \to (\mathbb{R}, +)$ by $\psi(x) = \ln x$. Then $\psi(xy) = \ln xy = \ln x + \ln y = \psi(x) + \psi(y)$, so ψ is a homomorphism. Since ϕ and ψ are inverses of each other, they are isomorphisms. Hence $(\mathbb{R}, +)$ and $(\mathbb{R}_{>0}, \cdot)$ are isomorphic groups.

3.3. Exercises.

EXERCISE 2.3.17. Let $f : G \to G'$ be a homomorphism of groups. Prove:

(1) If H is a subgroup of G and $\ker f \subseteq H$, then $f^{-1}f(H) = H$.
(2) If G is abelian, then $\mathrm{im}(f)$ is abelian.

EXERCISE 2.3.18. Let $G, +$ be an additive abelian group. Let $n \in \mathbb{Z}$ and $x \in G$. If $n > 0$, then $nx = \sum_{i=1}^{n} x = x + \cdots + x$ is the sum of n copies of x. If $n < 0$, then $nx = |n|(-x) = \sum_{i=1}^{|n|}(-x)$, and $0x = 0$.

(1) Show that "left multiplication by n" defines a function $\lambda_n : G \to G$ by the rule $\lambda_n(x) = nx$. Show that λ_n is an endomorphism of G.
(2) Show that the kernel of λ_n is $G(n) = \{x \in G \mid |x| \mid n\}$, hence $G(n)$ is a subgroup of G.

(3) Show that the image of λ_n is $nG = \{nx \mid x \in G\}$, hence nG is a subgroup of G.

When the group operation is written multiplicatively, the counterpart of λ_n is the "nth power map" which is denoted $\pi^n : G \to G$ and is defined by $\pi^n(x) = x^n$. In this case, $\mathrm{im}(\pi^n)$ is denoted G^n.

EXERCISE 2.3.19. Let G be a group and H a subgroup. Prove that if $[G : H] = 2$, then H is a normal subgroup.

EXERCISE 2.3.20. Let G be a group containing subgroups H, K, and N. Prove the following:

(1) If N is a normal subgroup of G, then NK is a subgroup of G. Moreover, K is a subgroup of NK, and N is a normal subgroup of NK.
(2) If N is normal, then $N \cap H$ is a normal subgroup of H.
(3) If H and K are both normal, then HK is a normal subgroup of G.

EXERCISE 2.3.21. Let G be a group. For every $a \in G$, let $\alpha_a : G \to G$ be defined by $\alpha_a(x) = a^{-1}xa$. In the terminology of Definition 2.3.4, $\alpha_a(x)$ is the conjugate of x by a. Prove that α_a is an automorphism of G.

EXERCISE 2.3.22. (The conjugate of a subgroup is a subgroup.) Let G be a group, S a nonempty subset of G, and $a \in G$. By Definition 2.3.4, the conjugate of S by a is defined to be $\alpha_a(S) = a^{-1}Sa$. Prove that S is a subgroup of G if and only if S^a is a subgroup of G.

EXERCISE 2.3.23. Let $S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$. Then $S^1 = \{e^{2\pi i\theta} \mid 0 \leq \theta < 1\}$ is the unit circle in the complex plane (see Section 1.4).

(1) Show that multiplication in $\mathbb{C}$ makes S^1 into a group.
(2) Let $(\mathbb{R}, +)$ denote the additive group on $\mathbb{R}$. Show that the function $f : (\mathbb{R}, +) \to S^1$ defined by $f(\theta) = e^{2\pi i\theta}$ is an onto homomorphism. Compute the kernel of f. Show that f induces an isomorphism $\mathbb{R}/\mathbb{Z} \cong S^1$ (see Exercise 2.2.31).
(3) If $n \in \mathbb{N}$, then the nth power map $z \mapsto z^n$ is an endomorphism of S^1 (see Exercise 2.3.18). Let μ_n denote the kernel of the nth power map. Show that $\mu_n = \{e^{2\pi ik/n} \mid k \in \mathbb{Z}\}$ is the set of all nth roots of unity in $\mathbb{C}$.
(4) Show that the function $\phi : \mathbb{Z} \to \mu_n$ defined by $\phi(k) = e^{2\pi ik/n}$ is an epimorphism. Compute the kernel of ϕ. Show that ϕ induces an isomorphism $\mathbb{Z}/n \cong \mu_n$.
(5) Let $\mu = \bigcup_{n \geq 1} \mu_n$. Show that μ is a group. Define $h : \mathbb{Q} \to \mu$ by $h(r) = e^{2\pi ir}$. Show that h is an epimorphism. Compute the kernel of h. Show that h induces an isomorphism $\mathbb{Q}/\mathbb{Z} \cong \mu$ (see Exercise 2.2.31).

EXERCISE 2.3.24. Let G be a finite group of order $n = [G : e]$. Let p be a prime number such that $p \mid n$ and $p^2 > n$. Assume G contains a subgroup H of order p. (This is always true, by Cauchy's Theorem, Theorem 2.7.3.) Prove:

(1) H is the unique subgroup of G of order p.
(2) H is a normal subgroup of G.

EXERCISE 2.3.25. A group G is said to be *simple* if the only normal subgroups of G are $\langle e \rangle$ and G. Prove that a group G is simple if and only if for every nontrivial homomorphism of groups $f : G \to G'$, f is a monomorphism.

EXERCISE 2.3.26. This exercise is a continuation of Exercise 2.2.27. Let K and H be groups and $K \times H$ the product group. Define four functions

(1) $\iota_1 : K \to K \times H$, $\iota_1(x) = (x, e)$
(2) $\iota_2 : H \to K \times H$, $\iota_2(y) = (e, y)$
(3) $\pi_1 : K \times H \to K$, $\pi_1(x, y) = x$
(4) $\pi_2 : K \times H \to H$, $\pi_2(x, y) = y$

Show that ι_1 and ι_2 are monomorphisms. Show that π_1 and π_2 are epimorphisms. Show that $\operatorname{im} \iota_1 = \ker \pi_2 = K \times \{e\}$ and $\operatorname{im} \iota_2 = \ker \pi_1 = \{e\} \times H$.

3.4. More on Cyclic Groups. One way we study a group is in terms of its subgroups. From this point of view, cyclic subgroups are the basic building blocks of a group. A group is the union of its cyclic subgroups.

A cyclic group $A = \langle a \rangle$ is generated by a single element. Theorem 2.3.27 shows that if A is infinite, then A is isomorphic to the additive group $\mathbb{Z}$. In this case A has two generators, namely a, and a^{-1}. If A is finite of order n, then A is isomorphic to $\mathbb{Z}/n$ and A has $\phi(n)$ generators, namely $\{a^i \mid 1 \leq i \leq n-1, \gcd(i, n) = 1\}$. Lemma 2.3.29 shows that any homomorphism $A \to G$ of groups defined on A is completely determined by the image of a generator. Necessary and sufficient conditions for the existence of a homomorphism $A \to G$ are derived. In Theorem 2.3.30 we show that the group of all automorphisms of a cyclic group of order n is isomorphic to the group of units modulo n. The group of automorphisms of an infinite cyclic group is a group of order two. As an application of these theorems on cyclic groups, we exhibit the classic proof by mathematical induction that a finite abelian group of order n contains an element of order p if p is a prime divisor of n (Theorem 2.3.32).

THEOREM 2.3.27. *Let $A = \langle a \rangle$ be a cyclic group. Then the following are true.*

(1) A is abelian.
(2) Every subgroup of A is cyclic.
(3) Every homomorphic image of A is cyclic.
(4) There is a unique $n \geq 0$ such that A is isomorphic to $\mathbb{Z}/\langle n \rangle$.
(5) If $n = 0$, then
  (a) A is infinite and
  (b) A is isomorphic to $\mathbb{Z}$.
(6) If $n > 0$, then
  (a) A isomorphic to $\mathbb{Z}/n$, hence A is finite of order n,
  (b) if H is a subgroup of A, then $|H|$ divides n,
  (c) for every positive divisor d of n, A has a unique subgroup of order d, namely $\langle a^{n/d} \rangle$,
  (d) if d is a positive divisor of n, then A has $\phi(d)$ elements of order d, where ϕ is the Euler function.

PROOF. (4): Let $\theta : \mathbb{Z} \to A$ be the function defined by $\theta(i) = a^i$. Since A is generated by a, by Lemma 2.2.5, a typical element of A is a^i for some $i \in \mathbb{Z}$. Therefore, θ is onto. Since $\theta(i + j) = a^{i+j} = a^i a^j = \theta(i)\theta(j)$, θ is an epimorphism. By Theorem 2.2.17 there is a unique $n \geq 0$ such that $\ker(\theta) = \langle n \rangle$. By Theorem 2.3.14 (1), θ induces an isomorphism $\bar{\theta} : \mathbb{Z}/\langle n \rangle \to A$.

(1): This follows from (4) and Exercise 2.3.17 (2).

(3): If $\phi : A \to G$ is a homomorphism, then $\phi(a^i) = \phi(a)^i$. Therefore we have $\operatorname{im}(\phi) = \langle \phi(a) \rangle$.

(2): By (4), $\bar{\theta} : \mathbb{Z}/\langle n\rangle \to A$ is an isomorphism. By (3) and Theorem 2.2.17 every subgroup of $\mathbb{Z}/\langle n\rangle$ is cyclic. Theorem 2.3.15 now applies.

(5): Follows from (4).

(6): Assume $n > 0$ and d is a positive divisor of n. By Lemma 2.2.18, $|a^{n/d}| = d$. Thus, $\langle a^{n/d}\rangle$, is a subgroup of order d. Now suppose $|a^x| = d$. By Lemma 2.2.18, $\gcd(x, n) = n/d$. By Bézout's Identity, Lemma 1.2.5, we can write $n/d = xu + nv$, for some $u, v \in \mathbb{Z}$. Since $a^{n/d} = (a^x)^u(a^n)^v = (a^x)^u$ we see that $\langle a^{n/d}\rangle \subseteq |a^x| = d$. Both groups have order d, hence they are equal. By Lemma 2.2.18, the number of elements of order n in A is equal to the cardinality of the set $\{x \in \mathbb{Z} \mid 1 \leq x \leq n \text{ and } \gcd(x, n) = 1\}$, which is equal to $\phi(n)$. Therefore, the number of elements of order d in a cyclic group of order d is $\phi(d)$. □

Example 2.3.28. In this example we show that up to isomorphism there are exactly two groups of order four. The group $\mathbb{Z}/4$ is a cyclic abelian group of order four, generated by either [1], or [3]. As in Example 2.1.22, the Klein Viergruppe is denoted V. Then $V = \{e, a, b, c\}$ has multiplication rules: $a^2 = b^2 = c^2 = e$, $ab = ba = c$ and is a noncyclic abelian group of order four. Let G be an arbitrary group of order four. We will show that G is either isomorphic to $\mathbb{Z}/4$, or V. Let $a \in G - \{e\}$. By Corollary 2.2.19, the order of a is either 2 or 4. If $|a| = 4$, then by Lemma 2.2.18, $G = \langle a\rangle$ is cyclic. By Theorem 2.3.27 (4), G is isomorphic to $\mathbb{Z}/4$. If G has no element of order 4, then every element a of G satisfies: $a^2 = e$. By Exercise 2.1.25, G is abelian. Let $a \in G - \{e\}$. Let $b \in G - \{e, a\}$. Let $c = ab$. Then $G = \{e, a, b, c\}$, $a^2 = b^2 = c^2 = e$, $ab = ba = c$, and it is clear that G is isomorphic to V.

Lemma 2.3.29. *Let $A = \langle a\rangle$ be a cyclic group and G any group.*

(1) Let $\phi : A \to G$ be a homomorphism of groups. Then ϕ is completely determined by the value $\phi(a)$.

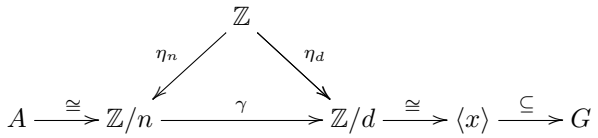
(2) Let $x \in G$.

(a) If the order of A is infinite, then there is a homomorphism $\theta : A \to G$ defined by $\theta(a) = x$.

(b) If A has finite order $|A| = n$, then there is a homomorphism $\theta : A \to G$ defined by $\theta(a) = x$ if and only if x has finite order $|x| = d$ and $d \mid n$.

Proof. (1): We have $\phi(a^i) = \phi(a)^i$.

(2): Part (a) was proved in the proof of Part (4) of Theorem 2.3.27. We prove Part (b). Assume A is finite and $|A| = n$. If there is a homomorphism $\theta : A \to G$, then by Exercise 2.3.44 the order of $\theta(a)$ is a divisor of n. Conversely, assume $|x| = d$ is finite and $d \mid n$. By Theorem 2.3.27 there is an isomorphism $A \cong \mathbb{Z}/n$ defined by $a^i \mapsto [i]$. Likewise, there is an isomorphism $\mathbb{Z}/d \cong \langle x\rangle$ defined by $[1] \mapsto x$. If η_n and η_d are the natural maps, then by Exercise 1.2.19, there exists a homomorphism γ such that the diagram



commutes. Define the homomorphism θ to be the composition of the four homomorphisms in the bottom row. As required, we have $\theta(a) = x$. □

As in Example 2.3.10, if G is a group, then $\mathrm{Aut}(G)$ denotes the group of all automorphisms of G. In Theorem 2.3.30 we show that the group of automorphisms of a cyclic group is an abelian group.

THEOREM 2.3.30. *Let $n \in \mathbb{N}$ be a positive integer. The group of automorphisms of the cyclic group of order n is isomorphic to the group of units modulo n. That is,*

$$\mathrm{Aut}(\mathbb{Z}/n) \cong U_n$$

which is a group of order $\phi(n)$. The group of automorphisms of the infinite cyclic group $\mathbb{Z}$ is isomorphic to the group of order two. That is,

$$\mathrm{Aut}(\mathbb{Z}) \cong \{1, -1\}.$$

PROOF. We utilize Theorem 2.3.27, Lemma 2.3.29, and Exercise 2.3.18. Let $A = \langle a \rangle$ be an arbitrary cyclic group. Given $r \in \mathbb{Z}$, the rth power map on A is denoted $\pi^r : A \to A$ and is defined by $\pi^r(a) = a^r$. If $\alpha : A \to A$ is an endomorphism of A, then $\alpha(a) = a^s$ for some integer s. Since

$$\alpha(a^t) = \alpha(a)^t = (a^s)^t = a^{st} \tag{3.1}$$

we see that $\alpha = \pi^s$. That is, every endomorphism of A is π^r for some $r \in \mathbb{Z}$. This also shows that the composite function $\pi^s\pi^t$ is equal to π^{st}. The image of $\pi^r : A \to A$ is the subgroup $\langle a^r \rangle$.

Case 1: Assume A is finite of order n. Then $a^r = a^s$ if and only if $r \equiv s \pmod{n}$. This proves that there are n distinct endomorphisms of A, namely $\{\pi^0, \pi^1, \ldots, \pi^{n-1}\}$. The generators of A are $\{a^r \mid \gcd(r, n) = 1\}$, which is a set of order $\phi(n)$. Since π^r is one-to-one and onto if and only if a^r is a generator of A, this proves that there are $\phi(n)$ automorphisms of A, namely $\{\pi^r \mid 1 \leq r \leq n-1, \gcd(r, n) = 1\}$. By Example 2.1.3, the group of units modulo n is an abelian group of of order $\phi(n)$. Define $\theta : \mathrm{Aut}(\mathbb{Z}/n) \to U_n$ by $\theta(\pi^r) = r$. Then we have shown that θ is an isomorphism of groups.

Case 2: Assume A is infinite. Then $a^r = a^s$ if and only if $r = s$. By Theorem 2.2.17, the two generators of A are $\{a, a^{-1}\}$. Therefore, the two automorphisms of A are π^1 and π^{-1}. □

REMARK 2.3.31. To effectively apply Theorem 2.3.30 it is necessary to have a complete description of the groups U_n in terms of the factorization $n = p_1^{e_1} \cdots p_k^{e_k}$ of n into prime numbers. We remark here that this computation of the group U_n will come later and depends on several theorems. For instance, by the Chinese Remainder Theorem (Corollary 2.5.3) $U_n \cong U_{p_1^{e_1}} \times \cdots \times U_{p_k^{e_k}}$. This allows us to compute the group U_n in terms of groups of the form U_{p^a}, where p is prime. To classify the groups U_{p^a} we will employ the Basis Theorem for Finite Abelian Groups (Theorem 2.8.7). In case $n = 2^a$, Proposition 2.8.8 gives a description of the group of units U_{2^a}. When p is an odd prime, see Proposition 3.6.19 for a description of the group of units U_{p^a}.

In general, if G is a finite group and p is a prime divisor of $|G|$, then G has an element of order p. This is known as Cauchy's Theorem and we will eventually present two proofs in Corollary 2.4.15 and Theorem 2.7.3. As an application of Theorem 2.3.27, an abelian version of Cauchy's Theorem is stated and proved in Theorem 2.3.32 below. The proof is an example of an important divide and conquer technique in Group Theory. The goal is to show that a group G has a

certain property. The strategy is to find a normal subgroup N such that both N and the quotient group G/N have that property, and from there proceed to show that the group G has it as well. The proof below is by induction on the order of G. The induction step uses Lagrange's Theorem (Corollary 2.2.14) and the fact that if N is a subgroup of G, then G/N is an abelian group (Example 2.3.7). The key step in the induction argument is that an element of order p in the quotient group G/N "lifts" to an element in G whose order is a multiple of p.

THEOREM 2.3.32. *(Cauchy's Theorem for Abelian Groups) Let G be a finite abelian group and p a prime number. If p divides $|G|$, then G contains an element of order p.*

PROOF. The proof is by induction on the order of G. Let $n = |G|$. Since p divides n, we know $n > 1$. If $p = |G|$, then by Exercise 2.2.30, there exists $a \in G$ such that $G = \langle a \rangle$, hence $|a| = p$. Inductively assume n is composite and that the result holds for all abelian groups of order less than n. By Corollary 2.2.21, we know G has a proper subgroup, call it N. If p divides $|N|$, then by our induction hypothesis, N has an element of order p. Therefore, assume p does not divide $|N|$. Since G is abelian, by Example 2.3.7, N is a normal subgroup and G/N is abelian. By Corollary 2.2.14, p divides $|N|[G : N]$. Since p does not divide $|N|$, we have p divides $[G : N]$. By our induction hypothesis, G/N has an element of order p. Suppose $b \in G$ and bN has order p in G/N. Since G is finite, b has finite order. By Exercise 2.3.44, p divides the order of b. By Theorem 2.3.27, $\langle b \rangle$ contains an element of order p. □

EXAMPLE 2.3.33. In this example we show that up to isomorphism there are exactly two groups of order six. By Example 2.1.3, we know that $\mathbb{Z}/6$ is an abelian group of order six. We know from Example 2.1.15 that the symmetric group on 3 letters, S_3, is a nonabelian group of order 6. Let G be a group of order six. Let $a \in G$ and set $A = \langle a \rangle$. By Corollary 2.2.19, $|a| \in \{1, 2, 3, 6\}$. If G has an element of order 6, then by Theorem 2.3.27, G is isomorphic to $\mathbb{Z}/6$. Assume from now on that G has no element of order 6. For contradiction's sake, suppose G has no element of order 3. Then every element of G satisfies $x^2 = e$. By Exercise 2.1.25, G is abelian and there exists $a \in G$ such that $|a| = 2$. Then $A = \langle a \rangle$ is normal and G/A has order three. By Exercise 2.3.44, if the generator of G/A is bA, then b has order 3 or 6, a contradiction. We have shown that G has an element a of order 3. If $A = \langle a \rangle$, then by Exercise 2.3.24, A is the unique subgroup of order 3. Then $G - A$ consists of elements of order 2. Let $b \in G - A$. The coset decomposition of G is $A \cup bA = \{e, a, a^2\} \cup \{b, ba, ba^2\}$. Since $[G : A] = 2$, by Exercise 2.3.19 A is normal. By Lemma 2.3.5, $bA = Ab$. Therefore, $ab \in \{b, ab, a^2b\}$. We know $ab \neq b$ since $a \neq e$. If $ba = ab$, then by Lemma 2.2.18, $|ab| = 6$, a contradiction. Therefore, $ab = ba^2$. We have proved that $G = \{e, a, a^2, b, ba, ba^2\}$ where $a^3 = b^2 = e$ and $ab = ba^2$. The reader should verify that the assignments $a \mapsto (123)$, $a^2 \mapsto (132)$, $b \mapsto (12)$, $ba \mapsto (23)$, and $ba^2 \mapsto (13)$ define an isomorphism $G \cong S_3$.

3.5. The Center of a Group. The center of a group is defined and as an exercise the reader is asked to prove that the center is a normal subgroup. As examples, we compute the center of the quaternion 8-group, the dihedral groups, the symmetric groups, and the general linear group of 2-by-2 matrices over a field.

DEFINITION 2.3.34. Let G be a group. The *center of* G, denoted $Z(G)$, is defined to be $\{x \in G \mid xa = ax \text{ for all } a \in G\}$. In Exercise 2.3.42 the reader is asked to prove that $Z(G)$ is a normal subgroup of G.

EXAMPLE 2.3.35. Let Q_8 be the quaternion 8-group of Example 2.1.19. In Exercise 2.4.23 the reader is asked to prove that the center of Q_8 is the unique subgroup of order two.

EXAMPLE 2.3.36. Let $n \geq 3$ and let D_n be the dihedral group (see Example 2.1.17). Then D_n is the group of symmetries of a regular n-gon. If H is the horizontal flip and R the rotation, then $D_n = \{H^i R^j \mid 0 \leq i \leq 1, 0 \leq j < n\}$ is a nonabelian group of order $2n$. The relations $H^2 = R^n = e$ and $HRH = R^{-1}$ hold. Hence the conjugate of R by H is $HRH = R^{-1}$ and the conjugate of H by R is $R^{-1}HR = HR^2$. Conjugation by R is an automorphism, so if $0 \leq i < n$, then $R^{-1}(HR^i)R = R^{-1}HRR^i = HR^2R^i$. This proves HR^i is not in $Z(D_n)$. It also shows that $Z(D_n)$ is a subgroup of $\langle R \rangle$. A typical element of $\langle R \rangle$ is R^i. We show that R^i is in $Z(D_n)$ if and only if $n \mid 2i$. Since $e \in Z(D_n)$, assume $0 < i < n$. Conjugation by H is an automorphism, so $HR^iH = R^{-i}$. We see that R^i is in $Z(D_n)$ if and only if $R^i = R^{-i}$, which is true if and only if $n = 2k$ is even and $i = k$. So if $n = 2k$ is even, $Z(D_n)$ is the subgroup $\langle R^k \rangle$ which has order 2. If n is odd, then it follows that the center of $D_n = \langle e \rangle$. In summary, we have shown that

$$Z(D_n) = \begin{cases} \langle R^{n/2} \rangle & \text{if } n \text{ is even} \\ \langle e \rangle & \text{if } n \text{ is odd.} \end{cases}$$

EXAMPLE 2.3.37. Let $n \geq 3$ and let S_n be the symmetric group on n letters (see Example 2.1.15). We show that $Z(S_n) = \langle e \rangle$. Let $\pi \in S_n$ and assume $\pi \neq e$. First assume $\pi(a) = b$ and $\pi(b) = c$, where a, b, c are distinct. Let τ be the 2-cycle (ab). Then $\pi\tau(a) = \pi(b) = c$ and $\tau\pi(a) = \tau(b) = a$, which shows π is not central. Now suppose $\pi(a) = b$ and $\pi(b) = a$. Let σ be the 2-cycle (bc), where a, b, c are distinct. Then $\pi\sigma(a) = \pi(a) = b$ and $\sigma\pi(a) = \sigma(b) = c$, which shows π is not central. If $\pi \neq e$, then π falls into one of these two cases. This shows $Z(S_n) = \langle e \rangle$.

EXAMPLE 2.3.38. Let F be a field and $\mathrm{GL}_n(F)$ the general linear group of invertible n-by-n matrices over F. For instance, if $n = 1$, then $\mathrm{GL}_1(F)$ is simply the set $F - \{0\}$ of invertible elements in F, which we denote F^*. If $n = 2$, then we saw in Example 2.1.21 that

$$\mathrm{GL}_2(F) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid ad - bc \neq 0 \right\}.$$

To compute the center, assume $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is a central matrix. Then

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} d & c \\ b & a \end{pmatrix}$$

shows that $a = d$ and $b = c$. Now

$$\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a-b & 0 \\ b & a+b \end{pmatrix}$$

shows that $b = 0$. Therefore, a central matrix is diagonal. It is routine to show that a diagonal matrix $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$ is central. This computation shows that $Z(\mathrm{GL}_2(F))$ is

equal to $\left\{ \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \mid a \in F^* \right\}$. If we define $\delta : F^* \to \mathrm{GL}_2(F)$ to be the diagonal map, $\delta(x) = \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}$, then δ is a monomorphism and $\mathrm{im}(\delta) = Z(\mathrm{GL}_2(F))$. The quotient, $\mathrm{GL}_2(F)/F^*$, is denoted $\mathrm{PGL}_2(F)$ and is called the *projective general linear group of* 2*-by-*2 *matrices over* F.

EXAMPLE 2.3.39. Let F be a field. Let $\det : \mathrm{GL}_2(F) \to F^*$ be the determinant function, where $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$. In Example 2.1.21 we showed that det is an epimorphism on multiplicative groups. This is proved in Lemma 6.3.5 below for all n. The kernel, $\ker(\det)$, which is the set of all matrices with determinant equal to 1, is denoted $\mathrm{SL}_2(F)$ and is called the *special linear group of* 2*-by-*2 *matrices over* F. By Theorem 2.3.14 (1) there is an isomorphism of groups

$$\mathrm{GL}_2(F)/\mathrm{SL}_2(F) \cong F^*.$$

See Exercise 2.5.18 for a computation of $\mathrm{SL}_2(\mathbb{Z}/3)$.

EXAMPLE 2.3.40. As in Example 2.1.15, the group of permutations of the set $\{1, 2, 3\}$ is

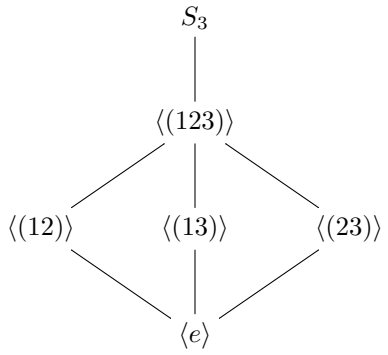
$$S_3 = \{e, (123), (132), (12), (13), (23)\}$$

and is called the *symmetric group on* 3 *elements*. The group S_3 is isomorphic to D_3, the group of symmetries of an equilateral triangle (see Example 2.1.16). Also, S_3 is isomorphic to $\mathrm{GL}_2(\mathbb{Z}/2)$, the group of invertible 2-by-2 matrices over the field of order 2 (see Exercise 2.1.26). The group table for S_3 is listed in Example 2.1.15. The cyclic subgroups of S_3 are:

$$\begin{aligned} \langle e \rangle &= \{e\} \\ \langle (123) \rangle = \langle (132) \rangle &= \{e, (123), (132)\} \\ \langle (12) \rangle &= \{e, (12)\} \\ \langle (13) \rangle &= \{e, (13)\} \\ \langle (23) \rangle &= \{e, (23)\} \end{aligned}$$

Since S_3 is a subgroup of itself, there are exactly 6 subgroups. The center of S_3 is the trivial subgroup $\langle e \rangle$, by Example 2.3.37. The commutator subgroup (see Exercise 2.3.46) of S_3 is the cyclic subgroup $\langle (123) \rangle$, by Exercise 2.3.47. There is one subgroup of order 6, one subgroup of order 3, three subgroups of order 2, and one subgroup of order 1. The three elements of order 2 are not central, hence the subgroups of order 2 are not normal. The commutator subgroup and the trivial subgroups are normal. By Proposition 2.2.7, the set of all subgroups of a group is

a lattice. The subgroup lattice of S_3 is



EXAMPLE 2.3.41. In Example 2.1.17 we defined the dihedral group D_n as the group of symmetries of a regular n-gon. For instance, if $n = 4$, the dihedral group

$$D_4 = \{e, (1234), (13)(24), (1432), (13), (24), (12)(34), (14)(23)\}$$

is a group of order 8 and is the group of symmetries of a square. In this example we use cycle notation, so $R = (1234)$ represents a rotation of the square through an angle of 90 degrees. The horizontal flip that fixes vertex 1 is $H = (24)$. The multiplicative powers of each element of D_4 are given in the rows of the following table. The order of the element is listed in the last column.

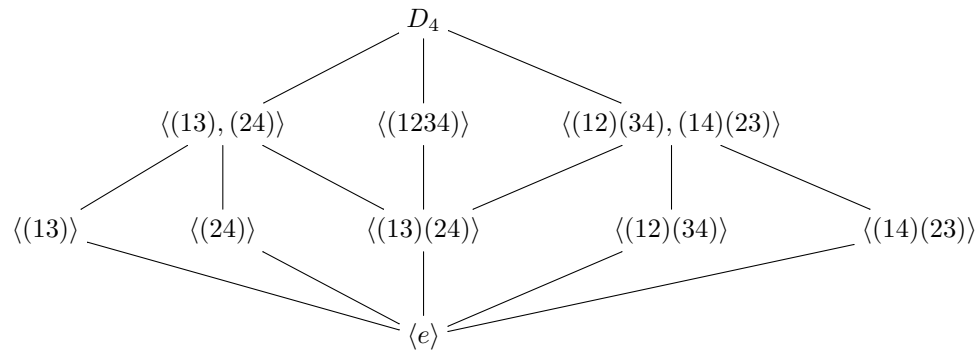
x	x^2	x^3	x^4	$\lvert x\rvert$
e				1
(1234)	$(13)(24)$	(1432)	e	4
$(13)(24)$	e			2
(1432)	$(13)(24)$	(1234)	e	4
(13)	e			2
(24)	e			2
$(12)(34)$	e			2
$(14)(23)$	e			2

There are 2 elements of order 4, 5 elements of order 2, and 1 element of order 1. Each element of order 2 generates a cyclic subgroup of order 2. The elements of order 4 are inverses of each other and generate the only cyclic subgroup of order 4 in D_4. There are two more subgroups of order 4 that are not cyclic:

$$\langle(13), (24)\rangle = \{e, (13), (13)(24), (24)\}$$
$$\langle(12)(34), (14)(23)\rangle = \{e, (12)(34), (13)(24), (14)(23)\}.$$

The trivial subgroups $\langle e\rangle$ and D_4 are normal. The three subgroups of order 4 are normal, by Exercise 2.3.19. The center of D_4 is the cyclic subgroup $\langle(13)(24)\rangle$ and is normal, by Exercise 2.3.42. The commutator subgroup of D_4 is the cyclic subgroup $\langle(13)(24)\rangle$, by Exercise 2.3.47. The only subgroups of D_4 that are not normal are the four cyclic subgroups of order 2 that are not central. The subgroup

lattice of D_4 is



where a line indicates set containment.

3.6. Exercises.

EXERCISE 2.3.42. Let G be a group. As in Definition 2.3.34, the center of G is the set $Z(G) = \{x \in G \mid xy = yx \text{ for every } y \in G\}$. Prove the following:

(1) $Z(G)$ is an abelian group.
(2) $Z(G)$ is a normal subgroup of G.
(3) If H and K are groups, then $Z(H \times K) = Z(H) \times Z(K)$.
(4) If $G/Z(G)$ is a cyclic group, then G is abelian.

EXERCISE 2.3.43. Let G be a group and $\mathrm{Aut}(G)$ the group of all automorphisms of G. As in Exercise 2.3.21, for every $a \in G$, let $\alpha_a : G \to G$ be defined by $\alpha_a(x) = a^{-1}xa$. Define $\theta : G \to \mathrm{Aut}(G)$ by $\theta(a) = \alpha_{a^{-1}}$. Show that θ is a homomorphism of groups. The image of θ is called the group of inner automorphisms of G and is denoted $\mathrm{Inn}(G)$. Show that $\ker(\theta)$ is equal to $Z(G)$, the center of G. Conclude that $\mathrm{Inn}(G)$ is isomorphic to $G/Z(G)$.

EXERCISE 2.3.44. Let $\theta : G \to G'$ be a homomorphism of groups and $x \in G$ an element of finite order. Show that $|\theta(x)|$ divides $|x|$.

EXERCISE 2.3.45. Let n be a positive integer. Prove that $\sum_{d|n} \phi(d) = n$. See Definition 1.2.15 for the notation $\sum_{d|n}$.

EXERCISE 2.3.46. Let G be a group. The *commutator subgroup of* G is the subgroup of G generated by the set $\{x^{-1}y^{-1}xy \mid x, y \in G\}$ and is denoted G'. Prove:

(1) G' is a normal subgroup of G.
(2) G/G' is abelian.
(3) If N is a normal subgroup of G such that G/N is abelian, then $G' \subseteq N$.
(4) If H is a subgroup of G and $G' \subseteq H$, then H is normal in G.

EXERCISE 2.3.47. Let $G = D_n$ be the dihedral group of order $2n$. Compute the commutator subgroup G' (see Exercise 2.3.46).

EXERCISE 2.3.48. Let

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 6 & 1 & 5 & 3 & 7 & 2 \end{bmatrix}, \quad \tau = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 2 & 4 & 3 & 6 & 1 & 7 \end{bmatrix}$$

be permutations in S_7. Compute $\tau\sigma\tau^{-1}$. Write σ, τ, $\tau\sigma\tau^{-1}$ using cycle notation. Show that σ factors into a 4-cycle times a 3-cycle. Show that $\tau\sigma\tau^{-1}$ factors into a 4-cycle times a 3-cycle. This is a special case of Lemma 2.6.9.

EXERCISE 2.3.49. Let G be a group and $X \subseteq G$. Let $\mathcal{S}$ be the set of all normal subgroups H in G such that $X \subseteq H$. Prove that $N = \bigcap_{H\in\mathcal{S}} H$ is a subgroup of G satisfying:

(1) N is a the smallest normal subgroup of G containing X.
(2) N is equal to the subgroup of G generated by the set $\bigcup_{g\in G} gXg^{-1}$.

We call N the *normal subgroup of G generated by X*.

EXERCISE 2.3.50. Let F be a field and $G = \mathrm{GL}_2(F)$ the general linear group of 2-by-2 matrices over F. Show that the commutator subgroup G' (see Exercise 2.3.46) is a subgroup of the special linear group $\mathrm{SL}_2(F)$ (see Example 2.3.39). For a continuation of this example, see Exercise 2.3.54.

EXERCISE 2.3.51. Let $\mathrm{GL}_2(F)$ be the general linear group of invertible 2-by-2 matrices over the field F and $\det : \mathrm{GL}_2(F) \to F^*$ the determinant function (see Example 2.1.21). Consider the following sets consisting of upper triangular matrices in $\mathrm{GL}_2(F)$:

$$U = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in M_2(F) \mid ad \neq 0 \right\},$$
$$D = \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \in M_2(F) \mid b \in F \right\}.$$

(1) Show that U is a subgroup of $\mathrm{GL}_2(F)$.
(2) Show that $\det : U \to F^*$ is an epimorphism of groups and describe the kernel as a set of matrices.
(3) Show that D is isomorphic to $(F, +)$, the additive group of the field F.
(4) Show that D is a normal subgroup of U and $U/D \cong F^* \times F^*$.
(5) Show that D is equal to the commutator subgroup of U. For the definition of commutator subgroup see Exercise 2.3.46.

For a continuation of this example, see Exercise 2.3.52.

EXERCISE 2.3.52. As in Exercise 2.3.51, let F be a field, $\mathrm{GL}_2(F)$ the general linear group of 2-by-2 matrices over F, and U the subgroup of $\mathrm{GL}_2(F)$ consisting of all upper triangular invertible matrices.

(1) Define $\theta : U \to F^*$ by $\theta \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = d$. Show that θ is a group epimorphism. Let $T = \ker\theta$. Describe T as a set of matrices.
(2) Show that
$$W = \left\{ \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \in M_2(F) \mid a \in F^* \right\}$$
is a subgroup of U. Assume $F \neq \mathbb{Z}/2$. In other words, assume F contains at least three elements. Show:
   (a) W is not a normal subgroup of U.
   (b) The normal subgroup of U generated by W (for this terminology, see Exercise 2.3.49) is the group T of Part (1).

For a continuation of this example, see Exercise 2.5.24.

EXERCISE 2.3.53. Let $\mathbb{C}^*$ be the group of all nonzero complex numbers under multiplication and $S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$ the subgroup of all complex numbers of absolute value 1 (see Exercise 2.3.23). Show that the quotient group $\mathbb{C}^*/S^1$ is isomorphic to $(\mathbb{R}_{>0}, \cdot)$, the multiplicative abelian group of positive real numbers. For a continuation of this exercise see Exercise 2.5.27.

EXERCISE 2.3.54. This exercise is a continuation of Exercise 2.3.50. Let F be a field and assume $F \neq \mathbb{Z}/2$. In other words, assume F is a field that has at least three elements. Show that the commutator subgroup of $\mathrm{GL}_2(F)$, the general linear group of 2-by-2 matrices over F, is equal to $\mathrm{SL}_2(F)$, the special linear group. (Although the proof is relatively long and tedious, it is elementary and involves only material already covered in this book.)

EXERCISE 2.3.55. Let Q_8 be the quaternion 8-group of Example 2.1.19 and D_4 the dihedral group of Example 2.1.17. Let C_4 be a cyclic group of order 4. For each of the following statements, either exhibit an example to substantiate the claim, or prove that the claim is false.

(1) There exists a monomorphism of groups $C_4 \to Q_8$.
(2) There exists an epimorphism of groups $Q_8 \to C_4$.
(3) There exists a monomorphism of groups $C_4 \to D_4$.
(4) There exists an epimorphism of groups $D_4 \to C_4$.

See Exercise 2.4.42 for a continuation of this exercise.

EXERCISE 2.3.56. Let G be a group and $\mathrm{Aut}(G)$ the group of all automorphisms of G. Let $\mathrm{Inn}(G)$ denote the group of inner automorphisms of G (see Exercise 2.3.43). Show that $\mathrm{Inn}(G)$ is a normal subgroup of $\mathrm{Aut}(G)$.

EXERCISE 2.3.57. Let $\theta : A \to B$ be an isomorphism of groups. Prove:

(1) θ maps the center of A isomorphically onto the center of B. That is, $\theta(Z(A)) = Z(B)$.
(2) θ maps the commutator subgroup of A isomorphically onto the commutator subgroup of B. That is, $\theta(A') = B'$.

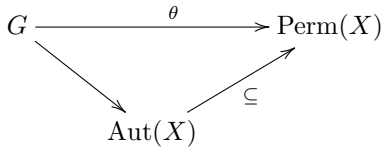
EXERCISE 2.3.58. Let G be a group containing subgroups A and B such that $A \subseteq B \subseteq G$. In this context, each of the following statements is true or false. If true, present a proof. If false, exhibit a counterexample.

(1) If A is normal in G, then A is normal in B.
(2) If A is normal in B and B is normal in G, then A is normal in G.
(3) If B is normal in G and $A = B'$ is the commutator subgroup of B, then A is normal in G.
(4) If B is normal in G and $A = Z(B)$ is the center of B, then A is normal in G.
(5) If A is normal in G, then B is normal in G.

4. Group Actions

The topic of the present chapter is Group Theory. As we have emphasized before, Group Theory arises as the axiomatic abstraction of permutation groups. In this section the connection between the abstract notion of a group G and the concrete notion of the group of all permutations of a set X is formalized. If X is any nonempty set with group of permutations $\mathrm{Perm}(X)$, then a homomorphism of

groups $\theta : G \to \mathrm{Perm}(X)$ associates to an element g of G a permutation $\theta(g)$ on X. We call such a representation of G an "action by G on X". We also say "G acts on X as a group of permutations". Cayley's Theorem, named after A. Cayley, says that a group G always has a representation as a group of permutations of the set $\{g \mid g \in G\}$ of elements of G. There is an important special case that arises when X is not only a set, but also a group. In case X is a group, $\mathrm{Aut}(X)$ is a subgroup of $\mathrm{Perm}(X)$. Whenever G acts as a group of permutations of X and $\theta : G \to \mathrm{Perm}(X)$ factors through $\mathrm{Aut}(X)$, so that



commutes, then we say "G acts as a group of automorphisms of X". A group action by G on X gives rise to an equivalence relation on X. In this case, the equivalence class of an element $x \in X$ is the set of all images $\theta(g)(x)$ where g is parametrized by G. This set is called the orbit of x under G. The set of all orbits is called the orbit space. Associated to x is the subgroup of G associated to all g such that the permutation $\theta(g)$ fixes x. Another important subset of X is the set of all points x that are fixed by every $\theta(g)$. Conjugation is an important action by G on itself. The orbits under this action are called the conjugacy classes. The subset of G fixed by this action is the center of G. When G is finite, the so-called Class Equation is an important counting theorem. We end this section with an important construction that allows us to construct a new group, called the semidirect product of H and K, whenever K acts as a group of automorphisms of H. Many of the results and notions from the present section will play an important role when we study Galois Theory in Section 5.3.

4.1. Group Actions, Orbits and Stabilizers.

LEMMA 2.4.1. *Let G be a group and S a nonempty set. The following are equivalent.*

(1) There is a homomorphism of groups $\theta : G \to \mathrm{Perm}(S)$.
(2) There is a function $G \times S \to S$, where the image of the ordered pair (g, x) is denoted $g * x$, and the properties
  (a) (associative law) $(g_1 g_2) * x = g_1 * (g_2 * x)$ for all $g_1, g_2 \in G$, $x \in S$ and
  (b) ($e \in G$ acts as the identity function) $e * x = x$, for all $x \in S$
  are satisfied.

PROOF. (1) implies (2): Instead of $\theta(g)(x)$ we will write $g * x$. The assignment $(g, x) \mapsto g * x$ defines a function $G \times S \to S$. Then

$$\begin{aligned}(g_1 g_2) * x &= \theta(g_1 g_2)(x) \\ &= \theta(g_1)(\theta(g_2)(x)) \\ &= g_1 * (g_2 * x)\end{aligned}$$

and $e * x = \theta(e)(x) = 1_S(x) = x$.

(2) implies (1): For each $g \in G$, define $\lambda_g : S \to S$ to be the "left multiplication by g" function defined by $\lambda_g(x) = g * x$. Since $g * g^{-1} = g^{-1} * g = e$, λ_g is a

permutation of S. Define $\theta : G \to \mathrm{Perm}(S)$ by $\theta(g) = \lambda_g$. The associative law implies $\theta(g_1 g_2) = \theta(g_1)\theta(g_2)$, so θ is a homomorphism. □

In light of Lemma 2.4.1 we make the following definition.

DEFINITION 2.4.2. Let G be a group and S a nonempty set. We say G *acts on* S *as a group of permutations*, if there is a homomorphism of groups $\theta : G \to \mathrm{Perm}(S)$. If $g \in G$ and $x \in S$, instead of $\theta(g)(x)$ we usually write $g * x$. If θ is one-to-one, then the group action is said to be *faithful*. If the image of θ is $\langle 1_X \rangle$, then the group action is said to be *trivial*. In Lemma 2.4.4 below, the kernel of θ is denoted G_0. Hence G_0 is a normal subgroup of G.

EXAMPLE 2.4.3. Let G be a group. As in Example 2.1.8, if $a \in G$, then $\lambda_a : G \to G$ is the "left multiplication by a" function and λ_a is a permutation of the set G. Since $\lambda_{ab} = \lambda_a \lambda_b$, the assignment $a \mapsto \lambda_a$ defines a homomorphism of groups $\lambda : G \to \mathrm{Perm}(G)$. Proposition 2.1.6 shows that λ is one-to-one.

LEMMA 2.4.4. *Let G be a group acting on a set X. Then*

$$G_0 = \{g \in G \mid g * x = x \text{ for all } x \in X\}$$

is a normal subgroup of G.

PROOF. As in Lemma 2.4.1, there is a homomorphism of groups $\theta : G \to \mathrm{Perm}(X)$ and G_0 is equal to the kernel of θ. □

THEOREM 2.4.5. *(Cayley's Theorem) A finite group of order n is isomorphic to a subgroup of the symmetric group S_n.*

PROOF. Let $G = \{g_1, \ldots, g_n\}$ be a fixed enumeration of the elements of G. Then we can identify $\mathrm{Perm}(G)$ with the symmetric group S_n. By Example 2.4.3, G is isomorphic to a subgroup of S_n. □

EXAMPLE 2.4.6. Let G be a group and H a subgroup. Let a, x, y be elements of G. Then $xH = yH$ if and only if $axH = ayH$ because $(ax)^{-1}ay = x^{-1}y$. If $a \in G$ and $xH \in G/H$, then $a * xH = (ax)H$ defines an action by G on the set G/H by left multiplication. The reader should verify that the criteria of Lemma 2.4.1 (2) are satisfied.

LEMMA 2.4.7. *Let H and K be groups. The following are equivalent.*

(1) There is a homomorphism of groups $\theta : K \to \mathrm{Aut}(H)$.
(2) There is a function $K \times H \to H$, where the image of the ordered pair (k, x) is denoted $k * x$, and the properties
  (a) (associative law) $(k_1 k_2) * x = k_1 * (k_2 * x)$ for all $k_1, k_2 \in k$, $x \in H$ and
  (b) ($e \in K$ acts as the identity function) $e * x = x$, for all $x \in H$
  (c) (distributive law) $k * (xy) = (k * x)(k * y)$ for all $k \in K$, $x, y \in H$.
  are satisfied.

PROOF. (1) implies (2): We identify $\mathrm{Aut}(H)$ with a subgroup of $\mathrm{Perm}(H)$. Then by Lemma 2.4.1, K acts on H as a group of permutations. The action by K on H is defined by $k * x = \theta(k)(x)$ and properties (a) and (b) are satisfied. The distributive law follows from the fact that $\theta(k)$ is a homomorphism if $k \in K$.

(2) implies (1): By Lemma 2.4.1, $K \to \mathrm{Perm}(H)$ is a homomorphism of groups, where $k \mapsto \lambda_k$. For $k \in K$, λ_k is a permutation of H. The distributive law implies λ_k is a homomorphism. □

In light of Lemma 2.4.7 we make the following definition.

DEFINITION 2.4.8. Let H and K be groups. We say K *acts on* H *as a group of automorphisms*, if there is a homomorphism of groups $\theta : K \to \mathrm{Aut}(H)$.

EXAMPLE 2.4.9. Let G be a group and $g \in G$. Conjugation of X by g is the automorphism $\alpha_g : G \to G$ defined by $\alpha_g(x) = g^{-1}xg$. (Exercise 2.3.21). By Exercise 2.3.43, there is a homomorphism of groups $\theta : G \to \mathrm{Aut}(G)$ defined by $g \mapsto \alpha_{g^{-1}}$. By Definition 2.4.8, this implies conjugation defines an action by G on itself as a group of automorphisms. The kernel of θ is $Z(G)$, the center of G. More generally, if N is a normal subgroup of G, and $g \in G$, then α_g restricts to an automorphism of N. Therefore there is a homomorphism $G \to \mathrm{Aut}(N)$ defined by $g \mapsto \alpha_{g^{-1}}$. Thus, conjugation defines an action of G on the normal subgroup N as a group of automorphisms. See Exercise 2.4.22 for a continuation of this example.

DEFINITION 2.4.10. Let G be a group acting as a group of permutations of a nonempty set X. Define a relation $\sim$ on X by the rule $x \sim y$ if $y = g * x$ for some $g \in G$. Then $x = e * x$ implies $x \sim x$, and if $y = g * x$, then $x = g^{-1} * y$. Moreover, if $y = g_1 * x$ and $z = g_2 * y$, then $z = g_2g_1 * x$. This proves that $\sim$ is an equivalence relation on X. The equivalence class of x is called the *orbit of* x. The orbit of x is equal to $G * x = \{g * x \mid g \in G\}$. The set of orbits is denoted X/G. If $x \in X$, then the *stabilizer of* x *in* G is $G_x = \{g \in G \mid g * x = x\}$. It is shown in Theorem 2.4.11 below that G_x is a subgroup of G, therefore, G_x is sometimes called the *subgroup fixing* x. If $G_x = G$, then we say x *is fixed by* G. The set $X_0 = \{x \in X \mid g * x = x \text{ for all } g \in G\}$ is the set of all x in X that are fixed by G.

THEOREM 2.4.11. *Let G be a group acting on a nonempty set X. If $x \in X$, then G_x, the stabilizer of x in G satisfies the following properties.*

(1) G_x is a subgroup of G.

(2) The length of the orbit $G * x$ is equal to the index $[G : G_x]$.

PROOF. (1): Since $e \in G_x$, we have $G_x \neq \emptyset$. If $a, b \in G_x$, then $ab * x = a * (b * x) = a * x = x$, hence $ab \in G_x$. If $a * x = x$, then $x = a^{-1} * x$. This proves G_x is a subgroup of G.

(2): We show that there is a one-to-one correspondence between the set of left cosets of G_x in G and the set $G*x$. Define a function $f : G \to G*x$ by $f(g) = g*x$. Then f is onto. Define a relation on G by the rule: $g \approx h$ if and only if $f(g) = f(h)$. By Exercise 1.1.14, $\approx$ is an equivalence relation. Notice that $g \approx h$ if and only if $g^{-1}h \in G_x$, which is equivalent to $g \equiv h \pmod{G_x}$. Therefore, $\bar{f} : G/G_x \to G * x$ is a one-to-one correspondence. □

4.2. Conjugates and the Class Equation.

EXAMPLE 2.4.12. Let G be a group and 2^G the power set of G. If X is a subset of G, and $a \in G$, then the conjugate of X by a is $\alpha_a(X) = a^{-1}Xa$ (Definition 2.3.4). Define a function $\theta : G \to \mathrm{Perm}(2^G)$ by $\theta(a) = \alpha_{a^{-1}}$. Since $\alpha_{(ab)^{-1}}(X) = (ab)X(ab)^{-1} = a(bXb^{-1})a^{-1} = \alpha_{a^{-1}}\alpha_{b^{-1}}(X)$, this implies $\theta(ab) = \theta(a)\theta(b)$. That is, θ is a homomorphism of groups. By Definition 2.4.2, conjugation defines an action by G as a group of permutations of the set 2^G. This action is therefore given by the formula $a * X = aXa^{-1}$. The stabilizer of X in G is usually called the *normalizer of* X *in* G and is denoted $N_G(X) = \{a \in G \mid aXa^{-1} = X\}$. The orbit of X under this action is the set $\{aXa^{-1} \mid a \in G\}$ of all distinct conjugates of X by elements of G.

PROPOSITION 2.4.13. *Let G be a group and X a subset of G. The normalizer of X in G satisfies the following properties.*

(1) $N_G(X)$ is a subgroup of G.
(2) If H is a subgroup of G, then $N_G(H)$ is the largest subgroup of G containing H as a normal subgroup.
(3) The number of distinct conjugates of X by elements in G is $[G : N_G(X)]$.

PROOF. (1) and (3): These follow from Theorem 2.4.11.

(2): Since H is a subgroup, $a^{-1}Ha = H$ for all $a \in H$. Therefore, $H \subseteq N_G(H)$. If $x \in N_G(H)$, then $x^{-1}Hx = H$. Therefore, H is normal in $N_G(H)$. Suppose $H \leq K \leq G$ and H is a normal subgroup of K. For all $x \in K$, $x^{-1}Hx = H$, hence $K \subseteq N_G(H)$. □

As in Example 2.4.9, let G be a group acting on itself by conjugation. Assume moreover that G is a finite group. If $x \in G$, the orbit of x is $G*x = \{a^{-1}xa \mid a \in G\}$ and is called the *conjugacy class of x*. The number of conjugates of x is the length of the orbit $G * x$. By Theorem 2.4.11, $|G * x| = [G : N_G(x)]$. If x is in $Z(G)$, the center of G, then $N_G(x) = G$ and $G * x = \{x\}$. Since $|G|$ is finite, there are a finite number of conjugacy classes. If $x_1, \ldots, x_n$ is a full set of representatives for the conjugacy classes that are not in $Z(G)$, then

$$\begin{aligned} G &= Z(G) \cup (G - Z(G)) \\ &= Z(G) \cup \left(\bigcup_{i=1}^{n} G * x_i \right) \\ &= Z(G) \cup G * x_1 \cup \cdots \cup G * x_n \end{aligned}$$

is a disjoint union. Taking cardinalities of both sides of this equation yields the next corollary.

COROLLARY 2.4.14. *(The Class Equation) Let G be a finite group and $x_1, \ldots, x_n$ a full set of representatives for the conjugacy classes that are not in $Z(G)$. Then*

$$|G| = |Z(G)| + \sum_{i=1}^{n} [G : N_G(x_i)].$$

As an application of Corollary 2.4.14, we prove Cauchy's Theorem. Recall that we already proved Theorem 2.3.32, which is the abelian version of this result. A second more concise proof of Cauchy's Theorem is given below in Theorem 2.7.3.

COROLLARY 2.4.15. *(Cauchy's Theorem) Let G be a finite group of order n and p a prime divisor of n. Then G contains an element of order p.*

PROOF. The proof is by induction on n. If G is abelian, then G has an element of order p, by Theorem 2.3.32. We see from Exercise 2.2.30 and Example 2.3.28 that any group of order five or less is abelian. Inductively assume $n \geq 6$, G is nonabelian, and that the result holds for any group of order less than n. Let $x_1, \ldots, x_m$ be a full set of representatives for the conjugacy classes that are not in $Z(G)$. By our induction hypothesis, $m \geq 1$. Solving the Class Equation of Corollary 2.4.14 for $|Z(G)|$, we have

$$|Z(G)| = |G| - \sum_{i=1}^{m} [G : N_G(x_i)]. \tag{4.1}$$

For each x_i, $N_G(x_i)$ is a proper subgroup of G. If p divides $|N_G(x_i)|$ for some i, then by our induction hypothesis, there is an element of order p in $N_G(x_i)$. Therefore, assume for every i that p does not divide $|N_G(x_i)|$. By Corollary 2.2.14, $|G| = |N_G(x_i)|[G : N_G(x_i)]$. Since p divides $|G|$ and p does not divide $|N_G(x_i)|$, we have p divides $[G : N_G(x_i)]$, for every i. Therefore, p divides the right hand side of (4.1). Hence p divides $|Z(G)|$. By Theorem 2.3.32, we know that $Z(G)$ has an element of order p. □

4.3. Semidirect Product. As in Definition 2.4.8, let H and K be groups and assume K acts on H as a group of automorphisms. In this context, we can build a new group which contains subgroups that are isomorphic to H and K. The underlying set for our bigger group is the cartesian product $H \times K$, but the new group is not the product group of Exercise 2.1.27 because a modified binary operation is used.

Define a binary operation on $H \times K$ by the rule:

$$(x_1, k_1)(x_2, k_2) = (x_1(k_1 * x_2), k_1k_2).$$

The binary operation defined above makes $H \times K$ into a group. The proof is Proposition 2.4.16 (1) and follows straight from the definitions and Lemma 2.4.7. The identity element is (e, e) and the inverse of (x, k) is $(k^{-1} * x^{-1}, k^{-1})$. This group is denoted $H \rtimes K$ and is called the *semidirect product* of H and K.

PROPOSITION 2.4.16. *Let K act on H as a group of automorphisms and let $H \rtimes K$ be the semidirect product of H and K. Then*

(1) $H \rtimes K$ is a group.
(2) $N = \{(x, e) \mid x \in H\}$ is a normal subgroup of $H \rtimes K$ and is isomorphic to H.
(3) The quotient $(H \rtimes K)/N$ is isomorphic to K.
(4) $C = \{(e, k) \mid k \in K\}$ is a subgroup of $H \rtimes K$ and K is isomorphic to C.

PROOF. (1): By the distributive and associative laws of Lemma 2.4.7, the last term of

$$\begin{aligned}((x_1, k_1)(x_2, k_2))\,(x_3, k_3) &= (x_1(k_1 * x_2), k_1k_2)(x_3, k_3)\\ &= (x_1(k_1 * x_2)(k_1k_2) * x_3, k_1k_2k_3)\end{aligned}$$

and the last term of

$$\begin{aligned}(x_1, k_1)\,((x_2, k_2)(x_3, k_3)) &= (x_1, k_1)(x_2(k_2 * x_3), k_2k_3)\\ &= (x_1k_1 * (x_2(k_2 * x_3)), k_1k_2k_3)\end{aligned}$$

are equal. Therefore, the binary operation on $H \rtimes K$ is associative. The proof that the identity element is (e, e) is left to the reader. Applying Lemma 2.4.7 and Exercise 2.4.21, an argument similar to the above proves that $(k^{-1} * x^{-1}, k^{-1})$ is the inverse of (x, k).

(2): Define $\iota_1 : H \to H \rtimes K$ by $\iota(x) = (x, e)$ and $\iota_2 : K \to H \rtimes K$ by $\iota(k) = (e, k)$. Define $f : H \rtimes K \to K$ by $f(x, k) = k$. The reader should verify that ι_1, ι_2 and f are homomorphisms of groups. As in Exercise 2.3.26, the reader should verify that f is onto, the kernel of f is the set N, ι_1 and ι_2 are one-to-one, and $\text{im}(\iota_1) = \ker(f)$. By Lemma 2.3.5, N is a normal subgroup of $H \rtimes K$. This proves (2).

Part (3) follows from (2) and Theorem 2.3.14 (1). Lastly, the image of ι_2 is the set C, which proves Part (4). □

When is a given group G isomorphic to a semidirect product? This question is the motivation for Corollary 2.4.17. It provides sufficient conditions for a group G to be a semidirect product of two subgroups N and K. If N is a normal subgroup of G and K is an arbitrary subgroup, then by Exercise 2.4.22, conjugation defines an action by K on N as a group of automorphisms. On elements the action is defined by the rule: $k * x = kxk^{-1}$, for all $k \in K$ and $x \in N$.

COROLLARY 2.4.17. *Let G be a group containing subgroups N and K satisfying:*

(1) $G = NK$,
(2) N is normal in G, and
(3) $N \cap K = \langle e \rangle$.

Let K act on N by conjugation. Then the semidirect product $N \rtimes K$ is isomorphic to G.

PROOF. Define $f : N \rtimes K \to G$ by $(x, k) \mapsto xk$. Then

$$\begin{aligned} f((x_1, k_1)(x_2, k_2)) &= f(x_1 k_1 * x_2, k_1 k_2) \\ &= f(x_1 k_1 x_2 k_1^{-1}, k_1 k_2) \\ &= x_1 k_1 x_2 k_1^{-1} k_1 k_2 \\ &= x_1 k_1 x_2 k_2 \\ &= f(x_1, k_1) f(x_2, k_2). \end{aligned}$$

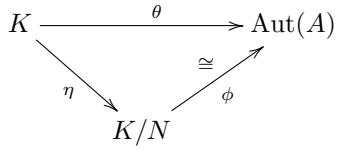
This shows f is a homomorphism. By (1) it follows that f is onto and by (3) it follows that f is one-to-one. □

EXAMPLE 2.4.18. In this example we show that for any $n \geq 1$ there exists a nonabelian group of order $6n$. The strategy is to construct a semidirect product $A \rtimes K$ where A has order 3, K has even order and K acts in a nontrivial way on A. From now on, let $A = \langle a \rangle$ be a cyclic group of order 3. By Theorem 2.3.30, $\mathrm{Aut}(A)$ is isomorphic to U_3 which is a group of order 2. By Proposition 2.4.16, to construct a semidirect product $A \rtimes K$, it suffices to find a group K and an onto homomorphism $\theta : K \to \mathrm{Aut}(A)$. By Theorem 2.3.14 (1), it suffices to find a group K which contains a normal subgroup of index 2. Any subgroup of index 2 is normal, by Exercise 2.3.19. Without being specific, let K be any group which contains a subgroup N of index 2. Before proceeding with our construction, we mention a few of the many choices for K.

(1) By Theorem 2.3.27, a cyclic group of order $2n$ has a subgroup of index 2. This shows that for any $n \geq 1$ there exists at least one abelian group K of order $2n$ which satisfies our criteria. We will see in Section 2.8 below that any finite abelian group of even order has a normal subgroup of index 2.
(2) If $n \geq 3$, a nonabelian choice for our group K would be the dihedral group D_n which has order $2n$ and contains a normal subgroup of order n (Example 2.1.17).
(3) Another nonabelian choice for K is the symmetric group S_n where $n \geq 3$ (Example 2.1.15). We will see in Corollary 2.6.15 below that S_n contains a unique subgroup of index 2 which is denoted A_n and is called the alternating group on n letters.
(4) There are many infinite groups for which our construction applies. For example, $K = \mathbb{Z}$ and $N = \langle 2 \rangle$.

(5) Another infinite choice for K is the group of nonzero real numbers $\mathbb{R}^*$ which contains the subgroup $\mathbb{R}_{>0}$ of positive real numbers. The quotient group $\mathbb{R}^*/\mathbb{R}_{>0}$ has order 2 (see Exercise 2.5.26).
(6) Notice that for a given group K, there may be multiple choices for N. For example, we saw in Example 2.3.41 that D_4 contains 3 subgroups of order 4.

Now we continue with our example. Assume from now on that A is a cyclic group of order 3, and K is a group containing a normal subgroup N such that $[K : N] = 2$. Since K/N and $\mathrm{Aut}(A)$ are cyclic groups of order 2, they are isomorphic to each other. Let $\theta = \phi\eta$ be the composition of the natural map η followed by the isomorphism ϕ. Then the diagram



commutes. Using θ we construct the semidirect product $A \rtimes K$. Since θ is onto, by Theorem 2.3.30, there exists some $k \in K$ such that $k * a = a^{-1}$. Therefore $(e, k)(a, e) = (a^{-1}, k)$ is not equal to $(a, e)(e, k) = (a, k)$. This shows elements of the subgroup $\{(x, e) \mid x \in A\}$ do not commute with elements of the subgroup $\{(e, k) \mid k \in K\}$. The group $A \rtimes K$ is nonabelian, even if K is abelian.

Next we present an example of a semidirect product of an arbitrary abelian group and a cyclic group of order two. The reader is asked to show in Exercise 2.4.25 below that the dihedral group D_n is the semidirect product of a cyclic group of order n and a group of order two. Therefore, the example in Proposition 2.4.19 is very general and includes as special cases the dihedral groups.

Let A be an abelian group, written multiplicatively. Let $\sigma : A \to A$ be the function defined by $\sigma(x) = x^{-1}$. In the notation of Exercise 2.3.18, σ is equal to the (-1)-power map and is denoted π^{-1}. Hence σ is an automorphism of A. Since $\sigma^2 = 1$, $\langle\sigma\rangle$ is a subgroup of $\mathrm{Aut}(A)$ of order 2 or less. Let $C = \langle c\rangle$ be a cyclic group of order 2. By Lemma 2.3.29, there exists a homomorphism $\theta : C \to \mathrm{Aut}(A)$ defined by $\theta(c) = \sigma$. By Lemma 2.4.7, the rule $c * x = x^{-1}$ for all $x \in A$ defines an action by C on A as a group of automorphisms.

PROPOSITION 2.4.19. *In the above context, let A be an abelian group and $C = \langle c\rangle$ a cyclic group of order 2 acting on A by the rule $c * x = x^{-1}$ for all $x \in A$. Let $G = A \rtimes C$ be the semidirect product.*

(1) The commutator subgroup of G is the set $G' = \{(x^2, e) \mid x \in A\}$.
(2) The group G is abelian if and only if $x^2 = e$ for all $x \in A$.
(3) If G is nonabelian, then $Z(G)$, the center of G, is equal to the set $\{(x, e) \mid x \in A$ and $x^2 = e\}$.

PROOF. For reference we list some useful multiplication identities for the group G which should be verified by the reader. In the following, x and y are arbitrary

elements of A.

$$(x,c)^{-1} = (x,c) \tag{4.2}$$

$$(x,e)^{-1} = (x^{-1},e) \tag{4.3}$$

$$(x,c)(y,c) = (xy^{-1},e) \tag{4.4}$$

$$(x,c)(y,e) = (xy^{-1},c) \tag{4.5}$$

$$(x,e)(y,c) = (xy,c) \tag{4.6}$$

$$(x,e)(y,e) = (xy,e) \tag{4.7}$$

The commutator subgroup of G is the subgroup generated by the set of all commutators $\{a^{-1}b^{-1}ab \mid a,b \in G\}$ (Exercise 2.3.46). Using (4.2) – (4.7), a general commutator in G is one of the following.

$$(x,c)(y,c)(x,c)(y,c) = (xy^{-1},e)(xy^{-1},e) = (x^2y^{-2},e) \tag{4.8}$$

$$(x,c)(y^{-1},e)(x,c)(y,e) = (xy,c)(xy^{-1},c) = (y^2,e) \tag{4.9}$$

$$(x^{-1},e)(y,c)(x,e)(y,c) = (x^{-1}y,c)(xy,c) = (x^{-2},e) \tag{4.10}$$

$$(x^{-1},e)(y^{-1},e)(x,e)(y,e) = (x^{-1}y^{-1}xy,e) = (e,e) \tag{4.11}$$

It is evident that the commutator subgroup G' is equal to $\{(x^2,e) \mid x \in A\}$. This proves (1). Part (2) follows from (1) since $G' = \langle (e,e) \rangle$ if and only if $x^2 = e$ for all $x \in A$. It follows from (4.5) – (4.7) that $(x,e) \in Z(G)$ for all $x \in A$ such that $x = x^{-1}$. Conversely, if $x \neq x^{-1}$, then $(x,e)(e,c) = (x,c)$ is not equal to $(e,c)(x,e) = (x^{-1},c)$ and $(x,c)(e,c) = (x,e)$ is not equal to $(e,c)(x,c) = (x^{-1},e)$. This proves (3). □

We end this section with the following application of Cauchy's Theorem (Corollary 2.4.15). In Proposition 2.4.20 we classify all groups of order pq, where p and q are distinct primes.

PROPOSITION 2.4.20. *Let p and q be distinct primes, and assume $p < q$. Let G be a group of order pq. Then G is isomorphism to one of two groups. Either G is abelian and isomorphic to $\mathbb{Z}/(pq)$, or G is nonabelian, $q \equiv 1 \pmod{p}$, and G is isomorphic to a semidirect product $\mathbb{Z}/q \rtimes \mathbb{Z}/p$.*

PROOF. The proof is divided into four steps.

Step 1. Assume G is an abelian group. By Theorem 2.3.32, let a be an element of order p and b an element of order q. By Lemma 2.2.18 (5), ab has order pq and G is cyclic. By Theorem 2.3.27, an abelian group of order pq is isomorphic to $\mathbb{Z}/(pq)$.

Step 2. Now we show that any group G of order pq is isomorphic to a semidirect product of a cyclic group of order q and a cyclic group of order p. By Corollary 2.4.15, let P be a subgroup of G of order p and Q a subgroup of G of order q. By Exercise 2.2.26, $P \cap Q = \langle e \rangle$. By Exercise 2.3.24, Q is normal in G. Then P acts on Q by conjugation (Exercise 2.4.22) and there is a homomorphism $\theta : P \to \mathrm{Aut}(Q)$. By Corollary 2.4.17, G is isomorphic to the semidirect product $Q \rtimes P$.

Step 3. Now we prove that there exists a nonabelian group of order pq if and only if $q \equiv 1 \pmod{p}$. Assume G is a group of order pq. By Step 2, there is a homomorphism $\theta : P \to \mathrm{Aut}(Q)$. By Theorem 2.3.30, $\mathrm{Aut}(Q) \cong U_q$ is an abelian group of order $\phi(q) = q - 1$. By Theorem 2.3.14 (1) and Corollary 2.2.14, the image of θ has order 1 or p. First we assume q is not congruent to 1 modulo p, and prove that G is abelian. If p does not divide $q - 1$, then p does not divide

the order of $\mathrm{Aut}(Q)$, so $\mathrm{im}(\theta) = \langle e \rangle$. By Exercise 2.4.41, G is isomorphic to the direct product of P and Q. By Exercise 2.1.27, G is an abelian group. Conversely, assume $q \equiv 1 \pmod{p}$ and show that there is a nonabelian group of order pq. In this case, p divides $q-1$ and by Corollary 2.4.15, there is a subgroup of order p in U_q. By Lemma 2.3.29, there exists a monomorphism $\theta : \mathbb{Z}/p \to \mathrm{Aut}(\mathbb{Z}/q)$. By Exercise 2.4.41, the semidirect product $\mathbb{Z}/q \rtimes \mathbb{Z}/p$ is a nonabelian group of order pq.

Step 4. We show that up to isomorphism a nonabelian group of order pq is unique. This part of the proof uses the fact that U_q, the group of units modulo q, is a cyclic group of order $q-1$. This result will be proved in Corollary 3.6.12 below. Since $\mathrm{Aut}(Q) \cong U_q$ is cyclic of order $q-1$, there is a unique subgroup of order p in $\mathrm{Aut}(Q)$ (Theorem 2.3.27 (6)). The monomorphism $\theta : \mathbb{Z}/p \to \mathrm{Aut}(\mathbb{Z}/q)$ is unique up to the choice of a generator for $\mathbb{Z}/p$. If $\theta_1 : \mathbb{Z}/p \to \mathrm{Aut}(\mathbb{Z}/q)$ is another monomorphism, then there exists $\sigma \in \mathrm{Aut}(\mathbb{Z}/p)$ such that $\theta = \theta_1\sigma$. So if $\mathbb{Z}/q \rtimes_\theta \mathbb{Z}/p$ is the semidirect product defined using θ, and $\mathbb{Z}/q \rtimes_{\theta_1} \mathbb{Z}/p$ is the semidirect product defined using θ_1, then the map defined by $(x, y) \mapsto (x, \sigma(y))$ is an isomorphism of groups. □

4.4. Exercises.

EXERCISE 2.4.21. Let H and K be groups. Recall (Definition 2.4.8) that we say K acts as a group of automorphisms of H if there is a homomorphism of groups $\theta : K \to \mathrm{Aut}(H)$. In this case, write $k * x$ instead of $\theta(k)(x)$. Prove the following:

(1) $k * e = e$ for all $k \in K$.
(2) $(k * x)^{-1} = k * x^{-1}$ for all $k \in K$, $x \in H$.

EXERCISE 2.4.22. Let G be a group containing a normal subgroup N. Let K be an arbitrary subgroup of G. Generalize Example 2.4.9 by showing that conjugation defines an action by K on N as a group of automorphisms. Specifically, show that if $k \in K$ and $x \in N$, then $k * x = kxk^{-1}$ defines an action by K on N as a group of automorphisms.

EXERCISE 2.4.23. Let $Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$ be the quaternion 8-group of Example 2.1.19. Show that every subgroup of Q_8 is normal. Let Z denote the center of Q_8. Show that Z is a group of order two and is contained in every nontrivial subgroup of Q_8. Show that Q_8 is not a semidirect product of two subgroups. Show that Z is equal to the commutator subgroup of Q_8.

EXERCISE 2.4.24. Let $m, n \in \mathbb{N}$ be positive integers. Show that there are $\gcd(m, n)$ distinct homomorphisms from $\mathbb{Z}/m$ to $\mathbb{Z}/n$. See Exercises 3.1.17 and 2.8.14 for a continuation of this exercise.

EXERCISE 2.4.25. If $n \geq 3$, show that the dihedral group D_n is isomorphic to the semidirect product of a cyclic subgroup of order n and a cyclic subgroup of order two.

EXERCISE 2.4.26. Let p be an odd prime. Let G be a group of order $2p$. Show that G has a unique subgroup of order p. Denote by P the subgroup of G of order p. Show that G is isomorphic to the semidirect product of P and a cyclic subgroup of order two that acts on P by conjugation. Show that G is isomorphic to either the cyclic group $\mathbb{Z}/2p$ or the dihedral group D_p.

EXERCISE 2.4.27. Show how to construct a nonabelian group of order $9 \cdot 37$ that contains a cyclic subgroup of order 9 and a cyclic subgroup of order 37.

EXERCISE 2.4.28. Let G be a group and H a subgroup of G. As in Example 2.4.6, G acts on G/H by left multiplication. By Lemma 2.4.1, there is a homomorphism of groups $\theta : G \to \mathrm{Perm}(G/H)$. As in Lemma 2.4.4, denote the kernel of θ by G_0. Prove:

(1) G_0 is a normal subgroup of G contained in H.
(2) If N is a normal subgroup of G contained in H, then $N \subseteq G_0$. Conclude that G_0 is the largest normal subgroup of G contained in H.
(3) If $[G : H] = n > 1$ and $[G : 1]$ does not divide $n!$, then H contains a nontrivial normal subgroup of G. Conclude that G is not a simple group.

EXERCISE 2.4.29. Let p be a prime and G be a group of order p^2. Apply Exercise 2.4.28 to show that every subgroup of G is normal. If G has order p^r, $r > 1$, show that every subgroup of order p^{r-1} is normal in G.

EXERCISE 2.4.30. Let p and q be primes such that $q \equiv 1 \pmod{p}$. Show how to construct a nonabelian group of order pq.

EXERCISE 2.4.31. Let $Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$ be the quaternion 8-group of Example 2.1.19. Show that $Q_8 = \{1\} \cup \{-1\} \cup \{\pm i\} \cup \{\pm j\} \cup \{\pm k\}$ is the decomposition of Q_8 into conjugacy classes.

EXERCISE 2.4.32. The group of symmetries of a square is

$$D_4 = \{e, (1234), (13)(24), (1432), (12)(34), (14)(23), (13), (24)\}.$$

Show that

$$D_4 = \{e\} \cup \{(13)(24)\} \cup \{(1234), (1432)\} \cup \{(24), (13)\} \cup \{(12)(34), (14)(23)\}$$

is the decomposition of D_4 into conjugacy classes.

EXERCISE 2.4.33. The group of symmetries of a regular pentagon is

$$\begin{aligned} D_5 = \{&e, (12345), (13524), (14253), (15432), \\ &(25)(34), (15)(24), (13)(45), (12)(35), (14)(23)\}. \end{aligned}$$

Show that

$$\begin{aligned} D_5 = \{e\} &\cup \{(12345), (15432)\} \cup \{(13524), (14253)\} \\ &\cup \{(25)(34), (15)(24), (13)(45), (12)(35), (14)(23)\} \end{aligned}$$

is the decomposition of D_5 into conjugacy classes.

EXERCISE 2.4.34. Show how to construct two nonisomorphic nonabelian groups of order 40 each of which is a semidirect product of two cyclic groups.

EXERCISE 2.4.35. Let G be a finite group and H a subgroup of G. Suppose the only normal subgroup of G contained in H is $\langle e \rangle$. Show that G is isomorphic to a subgroup of S_n, where $n = [G : H]$.

EXERCISE 2.4.36. For the following choices of p and q, show how to construct a nonabelian group of order pq which is a semidirect product of two cyclic groups.

(1) $p = 5$, $q = 11$.
(2) $p = 7$, $q = 29$.

EXERCISE 2.4.37. Let p be a prime number and n an integer such that $0 < n < p$. Prove that if G is a finite group of order pn and P is a subgroup of order p, then P is normal.

EXERCISE 2.4.38. Let G be a finite group and H a subgroup. Prove the following generalization of Exercise 2.3.19. If p is the smallest prime that divides $[G : 1]$ and $[G : H] = p$, then H is a normal subgroup.

EXERCISE 2.4.39. Let G be a finite group and suppose G has n conjugacy classes. Prove that $\sum_{x \in G} |N_G(x)| = n|G|$.

EXERCISE 2.4.40. Let G be a group and a, b elements of G. Prove:

(1) The cyclic subgroups $\langle ab \rangle$ and $\langle ba \rangle$ are conjugates of each other.
(2) The order of ab is equal to the order of ba.

EXERCISE 2.4.41. Let K act as a group of automorphisms on the group H. Let $\theta : K \to \mathrm{Aut}(H)$ be the associated homomorphism. Let $H \rtimes K$ be the semidirect product. Prove:

(1) If $\operatorname{im} \theta = \langle e \rangle$, then $H \rtimes K$ is isomorphic to $H \times K$.
(2) If $\operatorname{im} \theta \neq \langle e \rangle$, then elements of $N = \{(x, e) \mid x \in H\}$ do not necessarily commute with elements of $C = \{(e, k) \mid k \in K\}$. The group $H \rtimes K$ is nonabelian.

EXERCISE 2.4.42. This exercise is a continuation of Exercise 2.3.55. Let Q_8 be the quaternion 8-group of Example 2.1.19 and D_4 the dihedral group of Example 2.1.17. Let C_4 be a cyclic group of order 4.

(1) How many distinct homomorphisms $\theta : C_4 \to Q_8$ are there? How many are monomorphisms?
(2) How many distinct homomorphisms $\theta : Q_8 \to C_4$ are there? How many are epimorphisms?
(3) How many distinct homomorphisms $\theta : C_4 \to D_4$ are there? How many are monomorphisms?
(4) How many distinct homomorphisms $\theta : D_4 \to C_4$ are there? How many are epimorphisms?

EXERCISE 2.4.43. Let G be a group. Show that if N is a normal subgroup of G, then N is a disjoint union of conjugacy classes.

5. Direct Products

We have already defined the direct product $K \times H$ of two groups K and H. The binary operation on $K \times H$ is coordinate-wise. If $\{G_i \mid i \in I\}$ is a family of groups indexed by an arbitrary set I, then coordinate-wise multiplication is a binary operation on the product $\prod_{i \in I} G_i$ (Definition 1.3.4). The Chinese Remainder Theorem for the product of finite cyclic groups is proved in Theorem 2.5.2. When $N_1, \ldots, N_m$ are normal subgroups of a group G, then G is the internal direct product of $N_1, \ldots, N_m$ if the product map $\prod_{i=1}^{m} N_i \to G$ is an isomorphism of groups. The notion of a free group on a set X is introduced. This allows us to present a group G in terms of generators and relations.

5.1. External Direct Product. Given an arbitrary family of groups $\{G_i \mid i \in I\}$, a binary relation is defined on the product $G = \prod_{i\in I} G_i$ and the resulting group is called the direct product. In other words, we build the big group G and the building blocks are the groups G_i. Because the product G is constructed from the unrelated groups G_i, this is also called the external direct product. From this point of view, the groups G_i are not subgroups of G. Nevertheless, the canonical injection and projection maps give rise to homomorphisms $\iota_k : G_k \to G$ and $\pi_k : G \to G_k$.

DEFINITION 2.5.1. Let I be an index set and $\{G_i \mid i \in I\}$ a family of multiplicative groups indexed by I. Although the groups G_i in general are not equal as sets and have no common elements, we abuse notation and use the same symbol e to denote the identity element of each group G_i. As defined in Definition 1.3.4, the product is $\prod_{i\in I} G_i = \{f : I \to \bigcup_{i\in I} \mid f(i) \in G_i\}$. The product is a group if the binary operation is defined to be coordinate-wise multiplication: $(fg)(i) = f(i)g(i)$. The identity element is the constant function $e(i) = e$ and the inverse of f is defined by $f^{-1}(i) = (f(i))^{-1}$, the coordinate-wise inverse. The group $\prod_{i\in I} G_i$ is called the *direct product*. Sometimes $\prod_{i\in I} G_i$ is called the *external direct product* to distinguish it from the construction in Definition 2.5.4 below. For every $k \in I$ there is a *canonical injection map* $\iota_k : G_k \to \prod_{i\in I} G_i$ which maps $x \in G_k$ to $\iota_k(x)$, where

$$\iota_k(x)(i) = \begin{cases} x & \text{if } i = k \\ e & \text{otherwise.} \end{cases}$$

The *canonical projection map* is $\pi_k : \prod_{i\in I} G_i \to G_k$ where $\pi_k(f) = f(k)$. The reader should verify that ι_k is a monomorphism, π_k is an epimorphism and $\pi_k\iota_k = 1_{G_k}$. Since the product $\prod_{i\in I} G_i$ contains the constant function e, it is not necessary to apply the Axiom of Choice (Proposition 1.3.5).

When $I = \{1, \ldots, n\}$ is a finite set, the direct product is identified with the set of n-tuples $\{(x_1, \ldots, x_n) \mid x_i \in G_i\}$ and it is written $G_1 \times \cdots \times G_n$ or $\prod_{i=1}^n G_i$. Multiplication is defined coordinate-wise, hence $(x_1, \ldots, x_n)(y_1, \ldots, y_n) = (x_1y_1, \ldots, x_ny_n)$. The identity element is $(e, \ldots, e)$, and $(x_1, \ldots, x_n)^{-1}$ is $(x_1^{-1}, \ldots, x_n^{-1})$.

Theorem 2.5.2 is a generalization of Theorem 1.2.11 and Corollary 2.5.3 is a generalization of Proposition 1.2.14.

THEOREM 2.5.2. *(Chinese Remainder Theorem) Let m and n be positive integers and let*

$$\psi : \mathbb{Z} \to \mathbb{Z}/m \times \mathbb{Z}/n$$

be defined by $\psi(x) = (\eta_m(x), \eta_n(x))$, where $\eta_m : \mathbb{Z} \to \mathbb{Z}/m$ and $\eta_n : \mathbb{Z} \to \mathbb{Z}/n$ are the natural maps. Then the following are true:

(1) $\ker(\psi) = \langle M \rangle$, *where* $M = \operatorname{lcm}(m, n)$.
(2) ψ *is onto if and only if* $\gcd(m, n) = 1$.
(3) $\mathbb{Z}/m \times \mathbb{Z}/n$ *is cyclic if and only if* $\gcd(m, n) = 1$.

PROOF. (1): Since η_m and η_n are homomorphisms, it is routine to verify that ψ is a homomorphism. By Theorem 2.2.17, the kernel of η_m is $m\mathbb{Z}$ and the kernel of η_n is $n\mathbb{Z}$. We see that $\ker(\psi) = \ker(\eta_m) \cap \ker(\eta_n)$ is equal to $\{x \in \mathbb{Z} \mid m \mid x \text{ and } n \mid x\}$. By Theorem 2.2.17, $\ker(\psi)$ is generated by $M = \operatorname{lcm}(m, n)$.

(2): Let $d = \gcd(m, n)$. By Proposition 1.2.10, $Md = mn$. By Theorem 2.3.14, $\operatorname{im}(\psi)$ is isomorphic to $\mathbb{Z}/M$, which has order M. We see that ψ is onto if and only if $M = mn$, which is true if and only if $d = 1$.

(3): If $d = 1$, then the direct product $\mathbb{Z}/m \times \mathbb{Z}/n$ is cyclic by (2). Assume $d > 1$. To show the direct product is not cyclic, we show that it contains more than $\phi(d)$ elements of order d and apply Theorem 2.3.27 (6). Let $A = \{x \in \mathbb{Z}/m \mid |x| = d\}$. Then $|A| = \phi(d)$. If $x \in A$, then by an application of Lemma 2.2.18 (5) we see that $(x, 0)$ has order d in the direct product. Likewise, if $B = \{y \in \mathbb{Z}/n \mid |y| = d\}$, then $|B| = \phi(d)$ and $(0, y)$ has order d, for each $y \in B$. Therefore, the direct product contains at least $2\phi(d)$ elements of order d. This proves (3). □

COROLLARY 2.5.3. *Let m and n be relatively prime positive integers. Then*

$$\mathbb{Z}/mn \xrightarrow{\psi} \mathbb{Z}/m \times \mathbb{Z}/n$$

defined by $\psi([x]) = ([x], [x])$ is an isomorphism of additive groups. The restriction of the map ψ induces an isomorphism

$$U_{mn} \xrightarrow{\psi} U_m \times U_n$$

of multiplicative groups of units.

5.2. Internal Direct Product. The topics in this section are motivated by the question: When is a group G isomorphic to a direct product? The answer to this question is the basis for the definition of the internal direct product given below. If G is a group, and $N_1, N_2, \ldots, N_m$ a collection of normal subgroups of G, then the notion of internal direct product is defined (Definition 2.5.4). We prove in Lemma 2.5.5 (5) that algebraically, this notion is isomorphic to the external direct product $\prod_{i=1}^{m} N_i$ defined in Definition 2.5.1. In other words, G is the internal direct product of $N_1, N_2, \ldots, N_m$ if and only if the product map $(x_1, \ldots, x_m) \mapsto x_1 x_2 \cdots x_m$ is an isomorphism of groups $\prod_{i=1}^{m} N_i \to G$. If G is a group, then to show that G is a direct product it suffices to show that G is the internal direct product of a family of normal subgroups. Therefore, the difference between the external and internal direct product is the point of view.

DEFINITION 2.5.4. Let G be a group and $N_1, N_2, \ldots, N_m$ a collection of subgroups of G satisfying:

(1) N_i is a normal subgroup of G for each i,
(2) $G = N_1 N_2 \cdots N_m$, and
(3) if $x_i \in N_i$ for each i and $e = x_1 x_2 \cdots x_m$, then $x_i = e$ for each i.

Then we say G is the *internal direct product* of $N_1, \ldots, N_m$.

LEMMA 2.5.5. *Suppose G is the internal direct product of $N_1, N_2, \ldots, N_m$. Then the following are true.*

(1) If $i \neq j$, then $N_i \cap N_j = \langle e \rangle$.
(2) If $i \neq j$, $x_i \in N_i$, $x_j \in N_j$, then $x_i x_j = x_j x_i$.
(3) For each i let $x_i, y_i \in N_i$. If $x = x_1 x_2 \cdots x_m$, and $y = y_1 y_2 \cdots y_m$, then
  (a) $xy = (x_1 y_1)(x_2 y_2) \cdots (x_m y_m)$, and
  (b) $x^{-1} = x_1^{-1} x_2^{-1} \cdots x_m^{-1}$.
(4) If $x \in G$, then x has a unique representation as a product $x = x_1 x_2 \cdots x_m$, where $x_i \in N_i$ for each i.
(5) G is isomorphic to the (external) direct product $N_1 \times N_2 \times \cdots \times N_m$.

PROOF. (1): Let $x \in N_i \cap N_j$. Assume $1 \leq i < j \leq m$. In the product $N_1 \cdots N_i \cdots N_j \cdots N_m$ we have

$$e = e \cdots x \cdots x^{-1} \cdots e$$

where the ith factor is x, the jth factor is x^{-1}, and all other factors are the group identity e. By the uniqueness property of Definition 2.5.4, $x = e$.

(2): Because N_i and N_j are normal in G, we have $x_i y_j x_i^{-1} x_j^{-1}$ is in $N_i \cap N_j = \langle e \rangle$.

(3): The two identities follow immediately from Part (2).

(4): Assume $x = x_1 x_2 \cdots x_m$, where $x_i \in N_i$ for each i. Assume $x = y_1 y_2 \cdots y_m$, where $y_i \in N_i$ for each i is another such representation. Using Part (3), we get

$$e = x x^{-1} = (x_1 y_1^{-1}) \cdots (x_m y_m^{-1}).$$

By the uniqueness property of Definition 2.5.4, $x_i = y_i$ for each i.

(5): Let $\psi : N_1 \times N_2 \times \cdots \times N_m \to G$ be the function defined by multiplication in the group G: $\psi(x_1, x_2, \ldots, x_m) = x_1 x_2 \cdots x_m$. By Part (3), ψ is a homomorphism. By Definition 2.5.4, ψ is a one-to-one correspondence. □

PROPOSITION 2.5.6. *Let G be a group and $N_1, \ldots, N_m$ a collection of normal subgroups. Then the following are equivalent.*

(1) G is the internal direct product of $N_1, \ldots, N_m$.
(2) The function $\phi : N_1 \times \cdots \times N_m \to G$ defined by $\phi(x_1, \ldots, x_m) = x_1 \cdots x_m$ is an isomorphism of groups.
(3) $G = N_1 \cdots N_m$ and the intersection $N_k \cap (N_1 \cdots N_{k-1} N_{k+1} \cdots N_m) = \langle e \rangle$ is the trivial subgroup for each k.
(4) $G = N_1 \cdots N_m$, and $N_1 \cap N_2 \cdots N_m = N_2 \cap N_3 \cdots N_m = \cdots = N_{m-1} \cap N_m = \langle e \rangle$.

PROOF. (1) implies (2): This is Lemma 2.5.5 (5).

(2) implies (3): Since ϕ is onto we have $G = N_1 \cdots N_m$. Let x be an arbitrary element of $N_k \cap (N_1 \cdots N_{k-1} N_{k+1} \cdots N_m)$. We can write x in two ways: $x = x_k \in N_k$, and $x = x_1 \cdots x_{k-1} x_{k+1} \cdots x_m \in N_1 \cdots N_{k-1} N_{k+1} \cdots N_m$. Therefore $x = \phi(e, \ldots, e, x_k, e, \ldots, e) = \phi(x_1, \ldots, x_{k-1}, e, x_{k+1}, \ldots, x_m)$. Since ϕ is one-to-one, $x_1 = e$, ..., $x_m = e$. Hence $x = e$.

(3) implies (4): For each $k = 1, \ldots, m-1$ we have the set containment:

$$N_{k+1} \cdots N_m \subseteq N_1 \cdots N_{k-1} N_{k+1} \cdots N_m.$$

Therefore, $N_k \cap (N_{k+1} \cdots N_m) \subseteq N_k \cap (N_1 \cdots N_{k-1} N_{k+1} \cdots N_m) = \langle e \rangle$.

(4) implies (1): Let $e = x_1 x_2 \cdots x_m$ be a representation of e in $N_1 N_2 \cdots N_m$. Then $x_1^{-1} = x_2 \cdots x_m$ is in $N_1 \cap N_2 \cdots N_m = \langle e \rangle$. Therefore, $x_1 = e$ and $x_2 \cdots x_m = e$. Inductively, assume $1 < k < m$ and $x_k \cdots x_m = e$. Then $x_k^{-1} = x_{k+1} \cdots x_m$ is in $N_k \cap N_{k+1} \cdots N_m = \langle e \rangle$. Therefore, $x_k = e$ and $x_{k+1} \cdots x_m = e$. By induction, we are done. □

5.3. Free Groups. In this section the free group $F(X)$ on an arbitrary alphabet X is defined. As abstract groups we see that free groups are fundamental. For instance, $F(X)$ satisfies a universal mapping property and every group G is the homomorphic image of a free group. This allows us to introduce the notion of a group presentation in terms of generators and relations.

Let X be a set, which will be called the *alphabet*. A *word* on the alphabet X is a finite string of the form

$$w = a_1^{\epsilon_1} a_2^{\epsilon_2} \cdots a_n^{\epsilon_n}$$

where $n \geq 0$, each a_i is an element of X and $\epsilon_i \in \{-1, 1\}$. The *length* of the string is n. The only string of length 0 is called the *empty string* and is denoted e. A

string is *reduced* if it contains no substrings of the form xx^{-1} or $x^{-1}x$, for $x \in X$. Every word can be reduced in a unique way by recursively striking out all of the substrings of the form xx^{-1} or $x^{-1}x$.

LEMMA 2.5.7. *Let $v = a_1^{\epsilon_1}a_2^{\epsilon_2}\cdots a_n^{\epsilon_n}$ and $w = b_1^{\phi_1}b_2^{\phi_2}\cdots b_p^{\phi_p}$ be reduced words on the alphabet X. There exist factorizations of v and w into substrings $v = v_1v_2$, $w = w_1w_2$ such that v_2w_1 reduces to the empty word e and the reduction of vw is equal to v_1w_2. The factors v_1, v_2, w_1, w_2 are unique.*

PROOF. If v has length $n = 0$, then take $v_1 = v_2 = w_1 = e$ and $w_2 = w$. In this case, $vw = v_1w_2$ and we are done. Inductively assume $n > 0$ and that the result holds for any reduced word of length $n - 1$. If $a_n^{\epsilon} \neq b_1^{-\phi_1}$, then vw is reduced. In this case, take $v = v_1$,$v_2 = w_1 = e$, and $w_2 = w$. Otherwise, delete a_n^{ϵ} from the end of v and $b_1^{-\phi_1}$ from the front of w, and apply the induction hypothesis to obtain factorizations:

$$a_1^{\epsilon_1}a_2^{\epsilon_2}\cdots a_{n-1}^{\epsilon_{n-1}} = v_1v_3$$
$$b_2^{\phi_2}\cdots b_p^{\phi_p} = w_3w_2$$

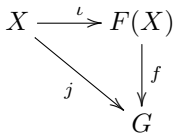
Setting $v_2 = v_3a_n^{\epsilon}$ and $w_1 = b_1^{\phi_1}w_3$, we have $v_2w_1 = v_3a_n^{\epsilon}b_1^{\phi_1}w_3$ reduces to v_3w_3 which reduces to the empty word e. Also, the reduction of vw is equal to the reduction of $v_1v_3w_3w_2$ which is equal to v_1w_2. This proves the existence of the factorization. The uniqueness of v_3 and w_3 implies the uniqueness of v_2 and w_1. □

LEMMA 2.5.8. *Let $F(X)$ be the set of all reduced words on X. Then $F(X)$ is a group, where the product of two words is the word defined by juxtaposition followed by reduction. The identity element for the group $F(X)$ is the empty string e. The inverse of the string $a_1^{\epsilon_1}a_2^{\epsilon_2}\cdots a_n^{\epsilon_n}$ is the string $a_n^{-\epsilon_n}\cdots a_2^{-\epsilon_2}a_1^{-\epsilon_1}$. There is a natural injection $\iota : X \to F(X)$ defined by $\iota(x) = x$.*

PROOF. By Lemma 2.5.7, if v and w are reduced words in $F(X)$, then the reduction of the word vw is uniquely defined. Since this binary operation does not depend on grouping by parentheses, it is associative. The rest is left to the reader. □

DEFINITION 2.5.9. The group $F(X)$ of Lemma 2.5.8 is called the *free group on the set X*.

THEOREM 2.5.10. *(Universal Mapping Property) Let X be a set and $\iota : X \to F(X)$ the natural injection map. For any group G and any function $j : X \to G$, there is a unique homomorphism $f : F(X) \to G$ such that the diagram*



commutes.

PROOF. Let $v = a_1^{\epsilon_1}a_2^{\epsilon_2}\cdots a_n^{\epsilon_n}$ be a reduced word in $F(X)$. Then we define $f(v)$ to be $j(a_1)^{\epsilon_1}j(a_2)^{\epsilon_2}\cdots j(a_n)^{\epsilon_n}$. Then f is a well defined function and $f\iota = j$. To see that f is a homomorphism of groups, let $w = b_1^{\phi_1}b_2^{\phi_2}\cdots b_p^{\phi_p}$ be another reduced word on the alphabet X. As in Lemma 2.5.7, factor $v = v_1v_2$, $w = w_1w_2$ such

that the reduction of vw is equal to v_1w_2. Since $f(v) = f(v_1v_2) = f(v_1)f(v_2)$, $f(w) = f(w_1w_2) = f(w_1)f(w_2)$, and $f(v_2)f(w_1) = e$, it follows that

$$f(vw) = f(v_1w_2) = f(v_1)f(w_2) = f(v_1)f(v_2)f(w_1)f(w_2) = f(v)f(w).$$

To prove the uniqueness claim, assume $g : F(X) \to G$ is another homomorphism and $g\iota = j$. Then $f(x) = g(x)$ for every $x \in X$. Since X is a generating set for the group $F(X)$, f is equal to g. □

COROLLARY 2.5.11. *Every group G is the homomorphic image of a free group.*

PROOF. In Theorem 2.5.10, take $X = G$ and $j : G \to G$ the identity map. Since j is onto, f is onto. □

DEFINITION 2.5.12. Let X be a set and Y a subset of $F(X)$. As in Exercise 2.3.49, let N be the normal subgroup of $F(X)$ generated by Y. Consider the quotient group $G = F(X)/N$. We say G is *defined by the generators X subject to the relations Y*. We denote the group $G = F(X)/N$ by $\langle X \mid Y\rangle$.

DEFINITION 2.5.13. Let G be a group. Then we say G is a *free group* if G has a generating set $X \subseteq G$ such that the natural map $F(X) \to G$ is an isomorphism. In this case, G has presentation $\langle X \mid \emptyset\rangle$. That is, a group G is free if there exists a relation-less or relation-free presentation of G.

EXAMPLE 2.5.14. In the notation of Theorem 2.3.27, let $A = \langle a\rangle$ be a cyclic group. If A is infinite, then a presentation of A in terms of generators and relations is $A = \langle a \mid \emptyset\rangle$. If A has order $n > 0$, then a presentation of A in terms of generators and relations is $A = \langle a \mid a^n\rangle$. It is common for the relations to be written as equations. Then $A = \langle a \mid a^n = e\rangle$.

EXAMPLE 2.5.15. Let $n > 2$ and D_n the dihedral group of order $2n$ of Example 2.1.17. Then D_n is generated by two elements, R and H. The order of R is n and the order of H is 2. The so-called commutator identity is $HRH = R^{-1}$. Therefore,

$$D_n = \langle R, H \mid H^2 = e,\ R^n = e,\ HRH = R^{-1}\rangle$$

is a presentation of D_n in terms of generators and relations.

EXAMPLE 2.5.16. Let V be the Klein 4-group of Example 2.1.22. Then V is an abelian group of order 4, generated by two elements of order two. Hence,

$$V = \langle a, b \mid a^2 = b^2 = e,\ ab = ba\rangle$$

is a presentation in terms of generators and relations.

EXAMPLE 2.5.17. Let $Q_8 = \{1, -1, i, -i, j, -j, k, -k\}$ be the quaternion eight group of Example 2.1.19. The multiplication rules are: $(-1)^2 = 1$, $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$. So we see that Q_8 is generated by i and j. Both i and j have order 4 and $-1 = i^2 = j^2$. The commutator relation for i and j is $ij = -ji = j^3i$. If we write a and b instead of i and j, then a presentation in terms of generators and relations is

$$Q_8 = \langle a, b \mid a^4 = e,\ b^4 = e,\ a^2 = b^2,\ ab = b^3a\rangle.$$

5.4. Exercises.

EXERCISE 2.5.18. The general linear group of 2-by-2 matrices over the field $\mathbb{Z}/3$, denoted $\mathrm{GL}_2(\mathbb{Z}/3)$, is the multiplicative group of invertible matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with entries in the field $\mathbb{Z}/3$. Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, $C = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}$, $P = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}$, and $Q = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ be matrices with entries in $\mathbb{Z}/3$. For the following computations, access to a computer algebra system such as [**28**] is not required, but will be beneficial, especially for parts (6) and (7).

(1) Show that A, B, C, P, and Q are in $\mathrm{GL}_2(\mathbb{Z}/3)$.
(2) Compute the cyclic subgroups $\langle A\rangle$, $\langle B\rangle$, $\langle C\rangle$, $\langle P\rangle$, $\langle Q\rangle$.
(3) Show that P is in the normalizer of $\langle A\rangle$. Show that P and A generate a subgroup of order 16.
(4) Show that P is in the normalizer of $\langle B\rangle$. Show that P and B generate a subgroup of order 16.
(5) Show that Q is in the normalizer of $\langle C\rangle$. Show that Q and C generate a subgroup of order 16.
(6) If $G = \mathrm{GL}_2(\mathbb{Z}/3)$, show that G has order 48. Show that G has 3 subgroups of order 16. Show that G has 4 subgroups of order 3.
(7) The special linear group of 2-by-2 matrices over $\mathbb{Z}/3$, denoted $\mathrm{SL}_2(\mathbb{Z}/3)$, is the subgroup of $\mathrm{GL}_2(\mathbb{Z}/3)$ consisting of those matrices with determinate equal to 1. Let $S = \mathrm{SL}_2(\mathbb{Z}/3)$. Show that S has order 24. Show that S has 3 subgroups of order 8. Show that every subgroup of order 8 is isomorphic to the quaternion 8-group, $Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$ of Example 2.1.19. . Show that S has 4 subgroups of order 3.

EXERCISE 2.5.19. Give an example of a group G and subgroups $N_1, N_2, \ldots, N_m$ of G satisfying:

(1) N_i is a normal subgroup of G for each i,
(2) $G = N_1 N_2 \cdots N_m$, and
(3) if $i \neq j$, then $N_i \cap N_j = \langle e\rangle$,

such that G is not the internal direct product of $N_1, N_2, \ldots, N_m$.

EXERCISE 2.5.20. Let G be a finite abelian group. Assume G is the internal direct product of cyclic subgroups $A = \langle a\rangle$ and $B = \langle b\rangle$ where a and b both have order 6.

(1) Show that $|G| = 36$.
(2) Show that $C = \langle ab^2\rangle$ has order 6.
(3) Compute $|AC|$.
(4) Show that AC is the internal direct product of A and $\langle b^2\rangle$.

EXERCISE 2.5.21. Let A and B be normal subgroups of G such that $G = AB$. Prove that $G/(A \cap B)$ is isomorphic to $G/A \times G/B$.

EXERCISE 2.5.22. Let G be a group containing subgroups A and B such that

(1) $G = AB$,
(2) $xy = yx$ for every $x \in A$ and $y \in B$, and
(3) $A \cap B = \langle e\rangle$.

Show that G is the internal direct product of A and B.

EXERCISE 2.5.23. Let A and B be groups.

(1) Let A_0 be a normal subgroup of A and B_0 a normal subgroup of B. Show that there is an isomorphism of groups
$$\frac{A \times B}{A_0 \times B_0} \cong \frac{A}{A_0} \times \frac{B}{B_0}.$$
(2) In the notation of Exercise 2.2.27, show that $\frac{A \times B}{A \times \langle e \rangle} \cong B$, and $\frac{A \times B}{\langle e \rangle \times B} \cong A$.

EXERCISE 2.5.24. This is a continuation of Exercise 2.3.52. Let F be a field and
$$U = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in M_2(F) \mid ad \neq 0 \right\}$$
the set of all upper triangular matrices in $\mathrm{GL}_2(F)$. Let T be the kernel of the homomorphism $U \to F^*$ defined by $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \mapsto d$. As in Example 2.3.38, let $\delta : F^* \to \mathrm{GL}_2(F)$ be the diagonal map. Let $Z = \operatorname{im} \delta$. Show that U is the internal direct product of T and Z.

EXERCISE 2.5.25. Let G be a group. Denote by $\mathbb{Z}^2$ the direct product $\mathbb{Z} \times \mathbb{Z}$ (see Exercise 2.1.27). In $\mathbb{Z}^2$ let $e_1 = (1, 0)$ and $e_2 = (0, 1)$. Prove the following generalization of Lemma 2.3.29.

(1) $\mathbb{Z}^2 = \langle e_1, e_2 \rangle$.
(2) Let $\theta : \mathbb{Z}^2 \to G$ be a homomorphism of groups.
   (a) $\operatorname{im}(\theta)$ is an abelian subgroup of G.
   (b) θ is completely determined by the two values $\theta(e_1)$ and $\theta(e_2)$.
   (c) $\theta(e_1)\theta(e_2) = \theta(e_2)\theta(e_1)$.
(3) If a and b are elements of G such that $ab = ba$, then there exists a group homomorphism $\theta : \mathbb{Z}^2 \to G$ such that $\theta(e_1) = a$ and $\theta(e_2) = b$.
(4) If G is a finite group with n conjugacy classes, then the number of distinct group homomorphisms $\theta : \mathbb{Z}^2 \to G$ is equal to $n|G|$.

EXERCISE 2.5.26. Let $\mathbb{R}^*$ be the group of all nonzero real numbers under multiplication. Let $(\mathbb{R}_{>0}, \cdot)$ be the group of all positive real numbers. Show that $\mathbb{R}^*$ is equal to the internal direct product of $\mathbb{R}_{>0}$ and the subgroup $\{1, -1\}$.

EXERCISE 2.5.27. Let $\mathbb{C}^*$ be the group of all nonzero complex numbers under multiplication. Let $S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$ be the subgroup of all complex numbers of absolute value 1 (see Exercise 2.3.23). Let $\mathbb{R}_{>0}$ be the group of all positive real numbers. Show that $\mathbb{C}^*$ is equal to the internal direct product of the subgroups S^1 and $\mathbb{R}_{>0}$.

EXERCISE 2.5.28. Let G be a group containing normal subgroups A and B such that $A \cap B = \langle e \rangle$. Prove:

(1) $ab = ba$ for every $a \in A$ and for every $b \in B$.
(2) The group AB is the internal direct product of A and B.

EXERCISE 2.5.29. The purpose of this exercise is to show that if A is an abelian group, then there exists a nonabelian group G such that A is isomorphic to $Z(G)$, the center of G.

(1) Show that the function defined by $\theta(x, y) = (y, x)$ defines an automorphism of $A \times A$.
(2) Show that $T = \langle \theta \rangle$ is a group of order 2.
(3) Let G be the semidirect product $(A \times A) \rtimes T$. Show that G is nonabelian.
(4) Show that $Z(G) = \{(x, x, 1) \mid x \in A\}$.
(5) Show that A is isomorphic to $Z(G)$.

EXERCISE 2.5.30. If G is the group of Exercise 2.5.29, determine the quotient group $G/Z(G)$.

6. Permutation Groups

The group of all permutations of $\mathbb{N}_n = \{1, 2, 3, \ldots, n\}$ is called the symmetric group on n letters and is denoted S_n. The reader is referred to Example 2.1.15 for the terminology and notation associated with the group S_n. Since S_n is a group of permutations of the set $\mathbb{N}_n$, we are in the context of Definition 2.4.2. In fact, S_n acts on the set $\mathbb{N}_n$. Given any permutation σ, the decomposition of the set $\mathbb{N}_n$ into orbits under σ gives rise to a factorization of σ into a product of disjoint cycles. Furthermore, using this cycle decomposition, we show that σ factors into a product of transpositions. In Section 2.6.2 we show that the set of all permutations that can be represented as a product involving an even number of transpositions is a subgroup of S_n. This subgroup, denoted A_n, has index two and is called the alternating group on n letters. We show in Corollary 2.6.15 below that A_n is the commutator subgroup of S_n and A_n is the only subgroup of index two in S_n. For $n \geq 3$, we know from Example 2.3.37 that the center of S_n is $\langle e \rangle$.

6.1. The Cycle Decomposition of a Permutation. Let $\alpha = (a_1, \ldots, a_s)$ be an s-cycle and $\beta = (b_1, \ldots, b_t)$ a t-cycle. We say α and β are *disjoint* if $\{a_1, \ldots, a_s\} \cap \{b_1, \ldots, b_t\} = \emptyset$. If this is the case, then $\beta(a_i) = a_i$ for each i, and $\alpha(b_j) = b_j$ for each j. Therefore, $\alpha\beta = \beta\alpha$. This proves Lemma 2.6.1.

LEMMA 2.6.1. *If α and β are disjoint cycles in S_n, then α and β commute. That is, $\alpha\beta = \beta\alpha$.*

EXAMPLE 2.6.2. Here is an example with $n = 6$. In S_6, let

$$\alpha = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 1 & 2 & 6 & 5 \end{bmatrix}, \quad \beta = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 5 & 4 & 3 & 1 & 2 \end{bmatrix}.$$

Then $A = \langle \alpha \rangle$ acts on $\{1, 2, 3, 4, 5, 6\}$. Given $x \in \{1, 2, 3, 4, 5, 6\}$, the orbit of x is $A * x$. We compute the orbit decomposition under this action. The reader should verify that $A * 1 = \{1, 3\}$, $A * 2 = \{2, 4\}$, $A * 5 = \{5, 6\}$. In Theorem 2.6.3 we find that from the orbit decomposition we can construct the factorization of α into cycles. For instance, $\alpha = (1, 3)(2, 4)(5, 6)$. Likewise, for $B = \langle \beta \rangle$, we find the disjoint orbits are $B * 1 = \{1, 6, 2, 5\}$, $B * 3 = \{3, 4\}$ and the factorization of β into cycles is $\beta = (1, 6, 2, 5)(3, 4)$.

THEOREM 2.6.3. *If $\sigma \in S_n$ is a permutation on n letters, then σ can be written as the product of disjoint cycles. This representation is unique in the sense that if $\sigma \neq e$ and $\sigma = \alpha_1\alpha_2 \cdots \alpha_k$ is a product of disjoint cycles all of length two or more and $\sigma = \beta_1\beta_2 \cdots \beta_\ell$ is another such representation, then $k = \ell$ and $\beta_1, \beta_2, \ldots, \beta_k$ can be relabeled such that $\alpha_i = \beta_i$ for each i.*

PROOF. Let $\sigma \in S_n$ and let $S = \langle \sigma \rangle$. Then S acts on $\mathbb{N}_n = \{1, 2, \ldots, n\}$. Let a be an arbitrary element of $\mathbb{N}_n$. We associate to the orbit of a under S a cyclic permutation α_a. Let S_a be the subgroup of S fixing a. Then S_a is a cyclic subgroup of S. If $[S : S_a] = w$, then by Theorem 2.3.27, S_a is the unique subgroup of S with index w and $S_a = \langle \sigma^w \rangle$. By Theorem 2.4.11, the length of the orbit of a is equal to w and the orbit of a is $\{a, \sigma(a), \sigma^2(a), \ldots, \sigma^{w-1}(a)\}$. On this set σ is equal to the cyclic permutation $\alpha_a = (a, \sigma(a), \sigma^2(a), \ldots, \sigma^{w-1}(a))$. We see that for every orbit under the S-action there is an associated cyclic permutation. If $\{a_1, a_2, \ldots, a_k\}$ is a full set of representatives for the orbits, then σ is equal to the product of cycles $\alpha_{a_1}\alpha_{a_1} \cdots \alpha_{a_k}$. The orbits are disjoint, hence so are the cycles in this factorization. The uniqueness claim follows from the fact that the cycle decomposition is determined by the orbit decomposition which is uniquely determined by σ. □

COROLLARY 2.6.4. *If $\alpha_1, \alpha_2, \ldots, \alpha_m$ are pairwise disjoint cycles in S_n, then the order of the product $\alpha_1\alpha_2 \ldots \alpha_m$ is equal to* $\operatorname{lcm}(|\alpha_1|, |\alpha_2|, \ldots, |\alpha_m|)$.

PROOF. Let $|\alpha_i| = k_i$ and let $k = \operatorname{lcm}(k_1, k_2, \ldots, k_m)$. By Lemma 2.6.1, the pairwise disjoint cycles commute. Therefore, $(\alpha_1\alpha_2 \ldots \alpha_m)^k = \alpha_1^k\alpha_2^k \ldots \alpha_m^k = e$. Suppose $\ell > 0$ and $e = (\alpha_1\alpha_2 \ldots \alpha_m)^\ell = \alpha_1^\ell\alpha_2^\ell \ldots \alpha_m^\ell$. The permutation $\alpha_2^\ell \ldots \alpha_m^\ell$ fixes point-wise every element of the orbit of α_1. Therefore, $\alpha_1^\ell = e$, hence $\ell \geq k_1$. By symmetry, $\ell \geq k_i$ for each i. □

COROLLARY 2.6.5. *Every $\pi \in S_n$ is a product of transpositions.*

PROOF. Let $k \geq 2$. By Theorem 2.6.3, it suffices to show that any k-cycle can be written as a product of transpositions. Notice that a 2-cycle (a_1a_2) is already a transposition, a 3-cycle $(a_1a_2a_3) = (a_1a_3)(a_1a_2)$ can be factored as a product of 2 transpositions, and a 4-cycle $(a_1a_2a_3a_4) = (a_1a_4)(a_1a_3)(a_1a_2)$ factors into 3 transpositions. In general, a k-cycle $(a_1a_2 \cdots a_k) = (a_1a_k) \cdots (a_1a_3)(a_1a_2)$ can be written as a product of $k - 1$ transpositions. □

6.2. The Sign of a Permutation. Let $n \geq 2$ and S_n the symmetric group on n letters. A permutation $\sigma \in S_n$ is said to be *even* if σ can be written as a product involving an even number of transpositions. If σ can be written as a product involving an odd number of transpositions, then we say σ is *odd*. We denote by A_n the subset of S_n consisting of all even permutations. The identity map e is even, and the product of even permutations is even. By Proposition 2.2.4, A_n is a subgroup, and is called the *alternating group on n letters*. In Lemma 2.6.6 below, we show that a permutation cannot be both even and odd. Although elementary, the proof is not trivial so we include many of the details.

LEMMA 2.6.6. *Let $n \geq 2$, S_n the symmetric group on n letters, and A_n the alternating group on n letters. A permutation σ in S_n cannot be both even and odd. The alternating group A_n is a normal subgroup of S_n, $[S_n : A_n] = 2$, the quotient group S_n/A_n is cyclic of order two, and if τ is any transposition in S_n, then the decomposition of S_n into left cosets is: $S_n = A_n \cup \tau A_n$.*

PROOF. Let $\sigma \in S_n$ and let $\sigma = \prod_{i=1}^{\ell} \sigma_i$ be the unique decomposition of σ into disjoint cycles (Theorem 2.6.3). Assume σ_i has length k_i. Define a function $N : S_n \to \mathbb{Z}$ by the formula $N(\sigma) = \sum_{i=1}^{\ell}(k_i - 1)$. By the proof of Corollary 2.6.5,

there exists a representation of σ as the product of $N(\sigma)$ transpositions. We proceed in two steps.

Step 1: Prove that if $\tau = (ab)$ is any transposition in S_n, then $N(\sigma\tau) = N(\sigma) \pm 1$. There are four cases.

Case 1: σ fixes a and b. Then clearly $N(\sigma\tau) = N(\sigma) + 1$.

Case 2: $\sigma(a) \neq a$ and $\sigma(b) = b$. Assume $\sigma_1 = (ac_2 \cdots c_{k_1})$. Any other cycle in σ fixes a and b. Then

$$\begin{aligned}\sigma\tau &= (ac_2 \cdots c_{k_1})(ab)\sigma_2 \cdots \sigma_\ell \\ &= (abc_2 \cdots c_{k_1})\sigma_2 \cdots \sigma_\ell\end{aligned}$$

which implies $N(\sigma\tau) = N(\sigma) + 1$.

Case 3: $\sigma(a) \neq a$ and $\sigma(b) \neq b$ and a and b belong to disjoint cycles in σ. Without loss of generality, assume $\sigma_1 = (ac_2 \cdots c_{k_1})$ and $\sigma_2 = (bd_2 \cdots d_{k_2})$. Then

$$\begin{aligned}\sigma\tau &= (ac_2 \cdots c_{k_1})(bd_2 \cdots d_{k_2})(ab)\sigma_3 \cdots \sigma_\ell \\ &= (ad_2 \cdots d_{k_2}bc_2 \cdots c_{k_1})\sigma_3 \cdots \sigma_\ell\end{aligned}$$

which implies $N(\sigma\tau) = k_1 + k_2 - 1 + k_3 - 1 + \cdots + k_\ell - 1 = N(\sigma) + 1$.

Case 4: a and b belong to the same cycle in σ. We split this case into two subcases. For simplicity's sake we assume σ is a cycle.

Subcase 4.1: a and b are not adjacent. Write $\sigma = (ac_1 \cdots c_i bd_1 \cdots d_j)$ where $i > 0$ and $j > 0$. Notice that $N(\sigma) = i + j + 1$. Then

$$\sigma\tau = (ac_1 \cdots c_i bd_1 \cdots d_j)(ab) = (ad_1 \cdots d_j)(bc_1 \cdots c_i)$$

which implies $N(\sigma\tau) = j + 1 + i + 1 - 2 = j + i = N(\sigma) - 1$.

Subcase 4.2: a and b are adjacent. Write $\sigma = (ac_1 \cdots c_i b)$. Notice that $N(\sigma) = i + 1$. Then

$$\sigma\tau = (ac_1 \cdots c_i b)(ab) = (bc_1 \cdots c_i)$$

which implies $N(\sigma\tau) = i = N(\sigma) - 1$.

Step 2: Suppose $\sigma = (a_1b_1) \cdots (a_mb_m)$ is a product of m transpositions. We show that m is congruent to $N(\sigma)$ modulo 2. A transposition has order 2, so

$$\begin{aligned}e &= \sigma\sigma^{-1} \\ &= \sigma(a_mb_m) \cdots (a_1b_1).\end{aligned}$$

Applying the formula from Step 1 above m times, we have

$$0 = N(e) = N(\sigma) + \sum_{i=1}^{m} \pm 1.$$

Reducing modulo 2, we get $N(\sigma) \equiv m \pmod 2$. Therefore, σ cannot be both even and odd. Using this result, we conclude that a transposition is odd. Hence $[S_n : A_n] \geq 2$. If σ and τ are both odd, then $\sigma\tau^{-1}$ is even, hence $[S_n : A_n] \leq 2$. Since $[S_n : A_n] = 2$, by Exercise 2.3.19, A_n is a normal subgroup of S_n. The quotient group S_n/A_n is cyclic of order two. Given any transposition $\tau \in S_n$, the decomposition of S_n into left cosets is: $S_n = A_n \cup \tau A_n$. □

PROPOSITION 2.6.7. *Let $n \geq 2$, S_n the symmetric group on n letters, and A_n the alternating group on n letters.*

(1) There is an epimorphism of multiplicative groups $\operatorname{sign} : S_n \to \{1, -1\}$.
(2) The kernel of sign *is A_n.*
(3) If $\sigma = \tau_1\tau_2 \cdots \tau_m$ is a product of transpositions, then $\operatorname{sign}(\sigma) = (-1)^m$.

PROOF. By Lemma 2.6.6, the assignment

$$\operatorname{sign}(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is even} \\ -1 & \text{if } \sigma \text{ is odd.} \end{cases}$$

defines a function sign : $S_n \to \{1, -1\}$. Identify $\{1, -1\}$ with the group μ_2 of square roots of unity in $\mathbb{C}$ (Exercise 2.3.23 (3). It is routine to check that sign is an onto homomorphism and the kernel is A_n. Since a transposition is odd, (3) follows from (1). □

EXAMPLE 2.6.8. Let σ be a k-cycle. By the proof of Corollary 2.6.5, $\sigma = (a_1 a_2 \cdots a_k) = (a_1 a_k) \cdots (a_1 a_3)(a_1 a_2)$ is a product of $k-1$ transpositions. By Proposition 2.6.7, $\operatorname{sign}(\sigma) = (-1)^{k-1}$. Hence σ is even if k is odd and odd if k is even.

We return to the study of the alternating group in Section 2.6.4.

6.3. Conjugacy Classes of the Symmetric Group. Let $n \geq 2$ and S_n the symmetric group on n letters. We view S_n as the group $\operatorname{Perm}(\mathbb{N}_n)$. The purpose of this section is to describe the conjugacy classes of S_n in terms of the partitions of the number n. If $\sigma \in S_n$, then we can write σ as a product of disjoint cycles $\sigma = \sigma_1 \sigma_2 \cdots \sigma_k$ where we assume $|\sigma_i| = s_i$ and $s_1 \geq s_2 \geq \cdots \geq s_k$. Furthermore, by adjoining 1-cycles if necessary, we assume $n = s_1 + s_2 + \cdots + s_k$. In other words, the sequence $s_1 \geq s_2 \geq \cdots \geq s_k$ is a partition of n. The next lemma shows that the conjugacy classes of S_n correspond to the partitions of n.

Let σ and θ be arbitrary permutations in S_n. Suppose $\sigma(i) = j$, $\theta(i) = k$, and $\theta(j) = \ell$. Then $\theta\sigma\theta^{-1}(k) = \theta\sigma(i) = \theta(j) = \ell$. This provides us with an algorithm to compute the cycle decomposition of the conjugation of σ by θ^{-1} given the cycle decomposition of σ: replace each letter by its image under θ. For instance, write $\sigma = \sigma_1 \sigma_2 \cdots \sigma_k$ as a product of disjoint cycles where $|\sigma_i| = s_i$, $s_1 \geq s_2 \geq \cdots \geq s_k$, and $n = s_1 + s_2 + \cdots + s_k$. Write $\sigma_i = (\sigma_{i1}, \sigma_{i2}, \ldots, \sigma_{is_i})$. Then $\theta\sigma_i\theta^{-1}$ is the cycle $(\theta(\sigma_{i1}), \theta(\sigma_{i2}), \ldots, \theta(\sigma_{is_i}))$. This shows that under conjugation the form of the cycle decomposition is preserved.

We illustrate this procedure by an example with $n = 10$. Let

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 3 & 8 & 4 & 5 & 1 & 10 & 9 & 7 & 6 & 2 \end{bmatrix}$$

$$\theta = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 5 & 4 & 10 & 1 & 7 & 3 & 9 & 8 & 6 & 2 \end{bmatrix}$$

Then

$$\theta\sigma\theta^{-1} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 7 & 4 & 2 & 8 & 10 & 3 & 5 & 9 & 6 & 1 \end{bmatrix}$$

As a product of disjoint cycles, we have $\sigma = (2, 8, 7, 9, 6, 10)(1, 3, 4, 5)$. Now compute the disjoint cycle form of the conjugate $\theta\sigma\theta^{-1}$. Because σ_1 starts with 2, and σ_2 starts with 1, we start the 6-cycle of $\theta\sigma\theta^{-1}$ with $\theta(2) = 4$, and the 4-cycle with $\theta(1) = 5$:

$$\begin{aligned} \theta\sigma\theta^{-1} &= (4, 8, 9, 6, 3, 2)(5, 10, 1, 7) \\ &= \big(\theta(2), \theta(8), \theta(7), \theta(9), \theta(6), \theta(10)\big)\big(\theta(1), \theta(3), \theta(4), \theta(5)\big). \end{aligned}$$

The last equation shows that the cycle decomposition can be obtained by applying θ to each letter in σ.

Now we show that every conjugacy class contains a canonical permutation. We continue to employ the notation established above. Consider the permutation

$$L = \begin{bmatrix} 1 & 2 & \dots & s_1 & s_1+1 & s_1+2 & \dots & s_1+s_2 & \dots & n \\ \sigma_{11} & \sigma_{12} & \dots & \sigma_{1s_1} & \sigma_{21} & \sigma_{22} & \dots & \sigma_{2s_2} & \dots & \sigma_{ks_k} \end{bmatrix}$$

where the second row is obtained by removing all of the parentheses from the product of disjoint cycles $\sigma_1\sigma_2\cdots\sigma_k$. Hence L is a permutation in S_n. Set $\tau = L^{-1}\sigma L$. Then the disjoint cycle decomposition of τ is obtained by inserting parentheses into $1, 2, \dots, n$ and splitting it into cycles with the lengths $s_1, \dots, s_k$.

We illustrate this algorithm on the example from above. Start with the permutation $\sigma = (2, 8, 7, 9, 6, 10)(1, 3, 4, 5)$ in S_{10}. Then

$$L = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 8 & 7 & 9 & 6 & 10 & 1 & 3 & 4 & 5 \end{bmatrix}$$

is the permutation whose second row is obtained by removing the parentheses from σ. Compute:

$$L^{-1}\sigma L = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 3 & 4 & 5 & 6 & 1 & 8 & 9 & 10 & 7 \end{bmatrix}.$$

We see that $L^{-1}\sigma L = (1, 2, 3, 4, 5, 6)(7, 8, 9, 10)$ in disjoint cycle form.

The two algorithms specified above combine to prove Lemma 2.6.9.

LEMMA 2.6.9. *Let $n \geq 2$ and S_n the symmetric group on n letters. Two permutations σ, τ in S_n are in the same conjugacy class if and only if they give rise to the same partition of n. The number of distinct conjugacy classes of S_n is equal to the number of distinct partitions of n.*

6.4. The Alternating Group. Let $n \geq 2$. The alternating group on n letters is denoted A_n and is defined to be the kernel of the homomorphism sign : $S_n \to \{1, -1\}$. That is, A_n is the subgroup of all even permutations. We have $[S_n : A_n] = 2$ and $|A_n| = n!/2$. Theorem 2.6.12, the main result of this section, is a proof that if $n \neq 4$, then A_n is simple. The proof we give is completely elementary. In Exercise 2.6.16 the reader is asked to prove that A_4 contains a normal subgroup of order 4, hence A_4 is not simple.

LEMMA 2.6.10. *If $n \geq 3$, then A_n is generated by 3-cycles.*

PROOF. By Corollary 2.6.5, a 3-cycle is even, so A_n contains every 3-cycle. Every permutation in A_n is a product of an even number of transpositions. It suffices to show that a typical product $(ab)(cd)$ factors into 3-cycles. If (ab) and (cd) are disjoint, then we see that

$$\begin{aligned}(ab)(cd) &= (ab)(ac)(ac)(cd) \\ &= (acb)(acd)\end{aligned}$$

is a product of 3-cycles. If $a = c$, then we have $(ab)(ad) = (adb)$. These are the only cases, so A_n is generated by 3-cycles. □

LEMMA 2.6.11. *Let $n \geq 3$. If N is a normal subgroup of A_n and N contains a 3-cycle, then $N = A_n$.*

PROOF. Without loss of generality assume $(123) \in N$. Then $(123)(123) = (132) \in N$. We assume $n > 3$, otherwise we are done. By Corollary 2.6.5, a 3-cycle is even, so A_n contains every 3-cycle. Let $3 < a \leq n$ be arbitrary. We use the fact that $\sigma^{-1}N\sigma \subseteq N$ for all $\sigma \in A_n$. Then $(1a3)(123)(13a) = (1a2)$ is in N. Also, $(1a2)^2 = (12a) \in N$. Similarly, we see that $(13a), (1a3), (23a), (2a3)$ are in N.

Now let $a \neq b$, $a > 2$, and $b > 2$. Then $(1b2)(12a)(12b) = (1ab)$ is in N. Similarly, we see that $(2ab), (3ab), (a1b), (a2b)$, etc. are in N.

Now let $a \neq b \neq c$, $a > 1$, $b > 1$, and $c > 1$. Then $(ac1)(a1b)(a1c) = (abc)$ is in N. So N contains every 3 cycle. By Lemma 2.6.10, $N = A_n$. □

THEOREM 2.6.12. *The alternating group A_n is simple if $n \neq 4$.*

PROOF. If $n = 2$, then $A_2 = \langle e \rangle$. If $n = 3$, then $A_3 = \langle (123) \rangle$ is a cyclic group of order 3, hence is simple. From now on assume $n > 4$, N is a normal subgroup of A_n and $N \neq \langle e \rangle$. We prove that $N = A_n$. The proof consists of a case-by-case analysis.

Case 1: If N contains a 3-cycle, then $N = A_n$, by Lemma 2.6.11.

Case 2: Assume N contains a permutation σ such that the cycle decomposition of σ has a cycle of length $r \geq 4$. Write $\sigma = (a_1 a_2 \cdots a_r)\tau$, where τ fixes each $a_1, \ldots, a_r$ element-wise. Let $\delta = (a_1 a_2 a_3)$. Then $\delta \in A_n$ and $\delta\sigma\delta^{-1} \in N$ since N is normal. The following computation

$$\begin{aligned}\sigma^{-1}\delta\sigma\delta^{-1} &= \tau^{-1}(a_1 a_r \cdots a_2)(a_1 a_2 a_3)(a_1 a_2 \cdots a_r)\tau(a_1 a_3 a_2) \\ &= (a_1 a_3 a_r)\end{aligned}$$

shows that Case 2 reduces to Case 1.

Case 3: Assume N has a permutation σ such that the cycle decomposition of σ has at least two disjoint 3-cycles. Write $\sigma = (a_1 a_2 a_3)(a_4 a_5 a_6)\tau$, where τ fixes each $a_1, a_2, a_3, a_4, a_5, a_6$ element-wise. Let $\delta = (a_1 a_2 a_4)$. Then $\delta \in A_n$ and $\delta^{-1}\sigma\delta \in N$ since N is normal. The following computation

$$\begin{aligned}\delta^{-1}\sigma\delta\sigma^{-1} &= (a_1 a_4 a_2)(a_1 a_2 a_3)(a_4 a_5 a_6)\tau(a_1 a_2 a_4)\tau^{-1}(a_1 a_3 a_2)(a_4 a_6 a_5) \\ &= (a_1 a_4 a_2 a_3 a_5)\end{aligned}$$

shows that Case 3 reduces to Case 2.

Case 4: Assume N has a permutation σ such that the cycle decomposition of σ consists of one 3-cycle and one or more 2-cycles. Write $\sigma = (a_1 a_2 a_3)\tau$, where τ is the product of the 2-cycles. Then $\sigma^2 = (a_1 a_3 a_2) \in N$, hence Case 4 reduces to Case 1.

Case 5: Assume every $\sigma \in N$ has a cycle decomposition that is a product of disjoint 2-cycles. Let $\sigma = (a_1 a_2)(a_3 a_4)\tau$ where τ is a product of 2-cycles and is disjoint from $(a_1 a_2)(a_3 a_4)$. Let $\delta = (a_1 a_2 a_3)$. Then $\delta \in A_n$ and $\delta^{-1}\sigma\delta \in N$ since N is normal. The following computation

$$\begin{aligned}\delta^{-1}\sigma\delta\sigma^{-1} &= (a_1 a_3 a_2)(a_1 a_2)(a_3 a_4)\tau(a_1 a_2 a_3)(a_1 a_2)(a_3 a_4)\tau \\ &= (a_1 a_4)(a_2 a_3)\end{aligned}$$

shows that $\beta = (a_1 a_4)(a_2 a_3)$ is in N. Since $n > 4$ (notice that this is the first time we have used this hypothesis), there exists $a_5 \notin \{a_1, a_2, a_3, a_4\}$. Let $\alpha = (a_1 a_4 a_5)$. The following computation

$$\begin{aligned}\alpha^{-1}\beta\alpha\beta &= (a_1 a_5 a_4)(a_1 a_4)(a_2 a_3)(a_1 a_4 a_5)(a_1 a_4)(a_2 a_3) \\ &= (a_1 a_4 a_5)\end{aligned}$$

shows that N contains a 3-cycle, hence Case 5 reduces to Case 1. □

COROLLARY 2.6.13. *If $n > 4$, the normal subgroups of S_n are $\langle e \rangle$, A_n, and S_n.*

PROOF. Let N be a normal subgroup of S_n. Then $N \cap A_n$ is a normal subgroup of A_n. By Theorem 2.6.12, $N \cap A_n$ is equal to either $\langle e \rangle$, or A_n. If $N \cap A_n = A_n$, then $[S_n : A_n] = 2$ implies $N = A_n$, or $N = S_n$. Suppose $N \cap A_n = \langle e \rangle$ and for contradiction's sake, suppose $N \neq \langle e \rangle$. Then N consists of e and odd permutations. If $\sigma \in N$ is an odd permutation, then σ^2 is even, hence $\sigma^2 \in N \cap A_n = \langle e \rangle$. Therefore, every element of N has order 2 or 1. Let $\sigma \in N$ and assume σ has order 2. By Corollary 2.6.4, σ decomposes into a product of disjoint transpositions. If $\sigma = (ab)$ is a transposition, then $(ab)(acb)(ab)(abc) = (acb)$ is in N, a contradiction. Assume $\sigma = (ab)(cd)\tau$, where τ is a product of disjoint transpositions that do not involve a, b, c, d. Let $\alpha = (acb)\sigma(abc) = (ac)(bd)\tau$. Then α is in N, and $\sigma\alpha = (ad)(bc)$ is in N. But $(ad)(bc)$ is even, which is a contradiction. □

COROLLARY 2.6.14. *Let $n > 4$. If H is a subgroup of S_n and $[S_n : H] < n$, then $H = A_n$ or $H = S_n$.*

PROOF. Let H be a subgroup of S_n, $m = [S_n : H]$, and assume $m < n$. Then S_n acts on G/H by left multiplication. If we identify $\mathrm{Perm}(G/H)$ with S_m, then there is a homomorphism of groups $\phi : S_n \to S_m$. By the Pigeonhole Principle (Exercise 1.1.11), $\ker \phi$ is a nontrivial normal subgroup of G. By Exercise 2.4.28, $\ker \phi$ is contained in H. By Corollary 2.6.13, $\ker \phi$ is either A_n or S_n. Therefore, H is either A_n or S_n. □

COROLLARY 2.6.15. *If $n \geq 2$, then the commutator subgroup of S_n is A_n. The subgroup A_n is the only subgroup of index two in S_n.*

PROOF. Let C denote the commutator subgroup of S_n. Since S_n/A_n is cyclic of order two, by Exercise 2.3.46 (3), C is a nontrivial normal subgroup of S_n and $C \leq A_n$. If $n \neq 4$, then Theorem 2.6.12 implies that $C = A_n$. If $n = 4$, Exercise 2.6.23 implies that $C = A_4$. Suppose H is a subgroup of S_n with index 2. Then H is normal by Exercise 2.3.19 and since S_n/H is abelian, H contains A_n, by Exercise 2.3.46 (3). □

6.5. Exercises.

EXERCISE 2.6.16. Let $G = A_4$ be the alternating group on 4 letters. The order of G is twelve.

(1) Viewing G as a group of permutations of $\{1, 2, 3, 4\}$, list the twelve elements of G using disjoint cycle notation. For each $x \in G$, compute the cyclic subgroup $\langle x \rangle$. Show that G has eight elements of order three and three elements of order two.
(2) Show that the subgroup of order 4 is the group of symmetries of a non-square rectangle (see Example 2.1.18).
(3) Show that G has four subgroups of order three. Show that the subgroup of order four is normal. Show that the center of G has order one. Construct the lattice of subgroups of G. Show that G has only one proper normal subgroup, namely the subgroup of order four.
(4) Show that the commutator subgroup of A_4 is the subgroup of order four.

EXERCISE 2.6.17. As in Exercise 2.6.16, the alternating group on four letters is denoted A_4. Let N be the normal subgroup of A_4 of order four. Show that G is isomorphic to the semidirect product of N and a cyclic subgroup of order three that acts on N by conjugation.

EXERCISE 2.6.18. Let A_4 be the alternating group on 4 letters (see Exercise 2.6.16). Compute the partition of A_4 into conjugacy classes.

EXERCISE 2.6.19. Show that the set of transpositions $\{(12), (23), \ldots, (n-1, n)\}$ generates S_n.

EXERCISE 2.6.20. Show that S_n is generated by a transposition $(1, 2)$ and an n-cycle $(123\cdots n)$.

EXERCISE 2.6.21. Compute the number of distinct k-cycles in S_n.

EXERCISE 2.6.22. Let $1 \leq k < n$. Show that for each k-subset $A = \{a_1, \ldots, a_k\}$ of $\mathbb{N}_n$ there is a subgroup of S_n isomorphic to $S_k \times S_{n-k}$. Show that any two such subgroups are conjugates of each other.

EXERCISE 2.6.23. Let $V = \{e, (12)(34), (13)(24), (14)(23)\}$ be the subgroup of order 4 in A_4. Show that V is a normal subgroup of S_4. Prove that S_4/V is a nonabelian group of order 6.

EXERCISE 2.6.24. Show that $\mathrm{Aut}(S_3)$ is isomorphic to S_3.

EXERCISE 2.6.25. Let $n \geq 2$. As in Section 1.5, if $\sigma \in S_n$, then P_σ is the n-by-n permutation matrix associated to σ. Show that $\{P_\sigma \mid \sigma \in S_n\}$ is a subgroup of $\mathrm{GL}_n(\mathbb{Z})$. Show that S_n is isomorphic to $\{P_\sigma \mid \sigma \in S_n\}$.

7. The Sylow Theorems

Throughout this section G will be a finite group and p will be a prime number that divides $|G|$, the order of G. If p is the only prime divisor of $|G|$, then we call G a p-group. Theorem 2.7.1, which plays a fundamental role in the study of p-groups, is proved. In particular, if $|G| = p^n$, then there is a chain of subgroups $\langle e\rangle \trianglelefteq G_1 \trianglelefteq G_2 \trianglelefteq \cdots \trianglelefteq G_{n-1} \trianglelefteq G$ where $|G_i| = p^i$. As an application, we give a second proof of Cauchy's Theorem, Theorem 2.7.3. A subgroup P of G is called a p-Sylow subgroup (pronounced *p-See-Low subgroup*), if P is a p-group and $|P|$ is the highest power of p that divides $|G|$. By Corollary 2.2.14, Lagrange's Theorem, if P is a p-Sylow subgroup of G, then $|P|$ is maximal among all p-subgroups of G.

The Sylow Theorems were first proved by P. Sylow, a nineteenth century Norwegian algebraist. In summary, the three Sylow Theorems prove that for every p that divides $|G|$, the following are true. The First Sylow Theorem (Theorem 2.7.4) shows that there exists at least one p-Sylow subgroup in G. The Second Sylow Theorem (Theorem 2.7.5) shows that two p-Sylow subgroups are conjugates of each other. The Third Sylow Theorem (Theorem 2.7.7) shows that the number of p-Sylow subgroups is a divisor of $|G|$ and is congruent to 1 modulo p.

7.1. p-Groups. Let p be a prime number. A finite group G is called a *p-group* if $|G| = p^r$ for some $r \geq 1$. We begin this section with the following fundamental theorem on p-groups. By Exercise 2.2.30, if $|G| = p$, then G is a finite simple cyclic abelian group.

THEOREM 2.7.1. *Let p be a prime and G a finite group of order p^n, where $n \geq 1$. Then the following are true.*

(1) $Z(G) \neq \langle e \rangle$.
(2) If G has order p or p^2, then G is abelian.
(3) If $n > 1$, then G has a proper normal subgroup N such that $\langle e \rangle \neq N \neq G$.
(4) (A finite p-group is solvable) There is a sequence of subgroups $G_0 \subseteq G_1 \subseteq \cdots \subseteq G_{n-1} \subseteq G_n$ such that
  (a) $G_0 = \langle e \rangle$, $G_n = G$,
  (b) for $0 \leq i \leq n$, $|G_i| = p^i$,
  (c) for $0 \leq i \leq n-1$, G_i is a normal subgroup of G_{i+1} and the quotient G_{i+1}/G_i is a cyclic group of order p.

  We call $G_0, G_1, \ldots, G_n$ a solvable series *for G.*
(5) Let X be a finite set and assume G acts on X as a group of permutations. Let $X_0 = \{x \in X \mid g * x = x \text{ for all } g \in G\}$. Then $|X| \equiv |X_0| \pmod{p}$.

PROOF. (5): If $x \in X$, then $x \in X_0$ if and only if $G*x = \{x\}$. If $X_0 = X$, there is nothing to prove. Let $x_1, \ldots, x_m$ be a full set of representatives of the orbits with length two or more. The orbit decomposition of X is $X_0 \cup (\bigcup_{i=1}^m G * x_i)$. Taking cardinalities and applying Theorem 2.4.11,

$$\begin{aligned} |X| &= |X_0| + \sum_{i=1}^{m} |G * x_i| \\ &= |X_0| + \sum_{i=1}^{m} [G : G_{x_i}]. \end{aligned}$$

Then $[G : G_{x_i}] \neq 1$ for each i and by Corollary 2.2.14, $[G : G_{x_i}]$ divides p^n. Reducing both sides of the equation modulo p, we get $|X| \equiv |X_0| \pmod{p}$.

(1): Let G act on itself by conjugation. Then $Z(G)$ is the set of all elements fixed by the group action. By Part (5), $0 \equiv |Z(G)| \pmod{p}$.

(2): By Part (1), $Z(G)$ has order p or p^2. Then $G/Z(G)$ has order 1 or p, hence is cyclic. By Exercise 2.3.42, G is abelian.

(3): By Part (1), if $Z(G) \neq G$, then $N = Z(G)$ works. If $Z(G) = G$, then G is abelian. In this case every subgroup of G is normal and by Corollary 2.2.21, G has a proper normal subgroup.

(4): The proof is by induction on n. If $n = 1$, then $G_0 = \langle e \rangle$, $G_1 = G$ is a solvable series. If $n = 2$, then by Part (3) $G_0 = \langle e \rangle$, $G_1 = N$, $G_2 = G$ is a solvable series.

Inductively, assume $n \geq 2$ and that a solvable series exists for any p-group of order less than p^n. By Part (3) there exists a proper normal subgroup N. Then $|N| = p^t$, where $1 \leq t < n-1$. By our induction hypothesis, let $G_0 = \langle e \rangle, G_1, \ldots, G_t = N$ be a solvable series for N. Let $H = G/N$. By Corollary 2.2.14, $|H| = p^{n-t}$. By our induction hypothesis, let $H_0 = \langle e \rangle, H_1, \ldots, H_{n-t-1}, H_{n-t} = H$ be a solvable series for $H = G/N$. By Theorem 2.3.15, we lift each H_i to a subgroup G_{i+t} of G and get a sequence $G_t = N \subseteq G_{t+1} \subseteq \cdots \subseteq G_{n-1} \subseteq G_n = G$. By Theorem 2.3.14, $G_{i+1+t}/G_{i+t} \cong H_{i+1}/H_i$ for each $0 \leq i \leq t$. Combining the two sequences, $G_0 \subseteq \cdots \subseteq G_t \subseteq \cdots \subseteq G_{n-1} \subseteq G_n = G$ is a solvable series for G. □

LEMMA 2.7.2. *Let G be a finite group and p a prime number that divides $|G|$. If H is a subgroup of G and H is a p-group, then the following are true:*
(1) $[N_G(H) : H] \equiv [G : H] \pmod{p}$.
(2) If p divides $[G : H]$, then p divides $[N_G(H) : H]$, hence $N_G(H) \neq H$.

PROOF. (1): As in Example 2.4.6, G acts on G/H by left multiplication. For $g \in G$, $g * xH$ is equal to $(gx)H$. Since H is a subgroup of G, this means H acts on G/H by left multiplication. Let $X = G/H$ and $X_0 = \{xH \in X \mid h * xH = xH \text{ for all } h \in H\}$. First we show that $xH \in X_0$ if and only if $x \in N_G(H)$. This follows from the following string of logical equivalences.

$$\begin{aligned} xH \in X_0 &\leftrightarrow (hx)H = xH \text{ for all } h \in H \\ &\leftrightarrow x^{-1}hx \in H \text{ for all } h \in H \\ &\leftrightarrow x^{-1}Hx = H \\ &\leftrightarrow xHx^{-1} = H \\ &\leftrightarrow x \in N_G(H). \end{aligned}$$

This shows that X_0 consists of those cosets xH such that $xH \subseteq N_G(H)$. Therefore, $|X_0| = [N_G(H) : H]$. By Theorem 2.7.1 (5), $|X| \equiv |X_0| \pmod{p}$, or $[G : H] \equiv [N_G(H) : H] \pmod{p}$.

(2): By Part (1), $0 \equiv [G : H] \equiv [N_G(H) : H] \pmod{p}$. Thus, $[N_G(H) : H]$ is a multiple of p. □

7.2. Cauchy's Theorem. As an application of Theorem 2.7.1 we give a second proof of Corollary 2.4.15, Cauchy's Theorem. The proof given below is due to J. McKay [**20**]. This has been the proof of choice used in [**6**], [**16**], and other introductory texts on this subject.

THEOREM 2.7.3. *(Cauchy's Theorem) Let G be a finite group of order n and p a prime divisor of n. Then G contains an element of order p.*

PROOF. Let $X = G^p = \prod_{i=1}^{p} G$ be the product of p copies of G. Elements of G^p are p-tuples $(x_1, \ldots, x_p)$ where each x_i is in G and $|X| = n^p$. In Exercise 2.7.10 the reader is asked to prove that the symmetric group S_p acts on X. For this proof, we require only a special case. Let ξ be the p-cycle $(12 \cdots p) \in S_p$. Then the cyclic subgroup $C = \langle \xi \rangle$ acts on X by

$$\xi^i * (x_1, \ldots, x_p) = \begin{cases} (x_p, x_1, \ldots, x_{p-1}) & \text{if } i = 1 \\ (x_{p-i+1}, \ldots, x_p, x_1, \ldots, x_{p-i}) & \text{if } 0 < i < p \\ (x_1, \ldots, x_p) & \text{if } i = 0 \text{ or } i = p. \end{cases}$$

Now define $Z = \{(x_1, \ldots, x_p) \in X \mid x_1 x_2 \cdots x_p = e\}$. Then Z is a subset of X. Given $(x_1, \ldots, x_p) \in Z$, notice that $x_p = (x_1 \cdots x_{p-1})^{-1}$, so $|Z| = n^{p-1}$. Since $x_p = (x_1 \cdots x_{p-1})^{-1}$ implies $x_p x_1 x_2 \cdots x_{p-1} = e$, it follows that $\xi * Z = Z$. Hence C acts on Z and there is a partition of Z into orbits. Let Z_0 be the set of all z in Z fixed by ξ. A p-tuple $z = (x_1, \ldots, x_p)$ is fixed by ξ if and only if $x_1 = x_2 = \cdots = x_p$. Since $(e, e, \ldots, e) \in Z_0$, we know $Z_0 \neq \emptyset$. By Theorem 2.7.1 (5), $|Z_0| \equiv 0 \pmod{p}$. Then $|Z_0| \geq p$, and there are at least p elements $g \in G$ such that $g^p = e$. One solution to $g^p = e$ is $g = e$. Any other solution is an element g of order p. □

7.3. The Sylow Theorems.

THEOREM 2.7.4. *(Sylow's First Theorem) Let G a finite group and p a prime number. If p^α divides $|G|$, then G contains a subgroup of order p^α.*

We give two proofs for Theorem 2.7.4. The first proof is due to H. Wielandt [**30**]. The proof is based on an elementary combinatorial and number theoretic argument. It has been the proof of choice used by [**6**], [**15**] and other introductory books on this subject. Recursively applying Lemma 2.7.2 and Theorem 2.7.3, the second proof constructs a tower of subgroups $P_{i-1} \subseteq P_i$ in G such that $|P_i| = p^i$. The idea for this proof comes from [**16**].

FIRST PROOF OF THEOREM 2.7.4. Write $|G| = p^\gamma r$ where p^γ is the highest power of p that divides $|G|$. Then $0 \leq \alpha \leq \gamma$, and we write $|G| = p^\alpha q$. If we let $\beta = \gamma - \alpha$, then p^β is the highest power of p that divides q. Let X be the set of all subsets of G of cardinality p^α. Then

$$|X| = \binom{p^\alpha q}{p^\alpha} = \frac{p^\alpha q}{p^\alpha} \cdot \frac{p^\alpha q - 1}{p^\alpha - 1} \cdots \frac{p^\alpha q - i}{p^\alpha - i} \cdots \frac{p^\alpha q - p^\alpha + 2}{p^\alpha - p^\alpha + 2} \cdot \frac{p^\alpha q - p^\alpha + 1}{p^\alpha - p^\alpha + 1}$$

where the factorization on the right hand side results from expanding the binomial coefficient using Lemma 1.1.4, Let $0 < i < p^\alpha$ and write $i = p^t k$ where $0 \leq t < \alpha$ and $\gcd(p, k) = 1$. Then $p^\alpha q - i = p^t(p^{\alpha - t} q - k)$ and $p^{\alpha-t} q - k \equiv -k \pmod{p}$. This implies the highest power of p that divides $p^\alpha q - i$ is p^t. Therefore, canceling all powers of p from the numerator and denominator we see that the highest power of p that divides $|X|$ is the same as the highest power of p that divides q, which is p^β. As in Example 2.4.3, G acts on itself by left multiplication. If $a \in G$, and $S \in X$, then aS has cardinality p^α. Therefore, $a * S = aS$ defines an action by G on X. Under this action, X is partitioned into orbits. Since $p^{\beta+1}$ does not divide $|X|$, we know there is an orbit, say $G * S$, such that $p^{\beta+1}$ does not divide $|G * S|$, the length of the orbit. Let $H = G_S$ be the stabilizer of S. Then $H = \{h \in G \mid hS = S\}$. So $hs \in S$ for each $h \in H$ and $s \in S$. For a fixed $s \in S$, this implies the right coset Hs is a subset of S. Hence $|H| \leq |S| = p^\alpha$. By Corollary 2.2.14, $|G * S| = |G|/|H| = (p^\alpha q)/|H|$. Thus $p^\alpha q = |H||G * S|$. Since $p^{\alpha+\beta}$ divides the left hand side, we have $p^{\alpha+\beta}$ divides $|H||G * S|$. Since $p^{\beta+1}$ does not divide $|G * S|$, this implies p^α divides $|H|$. This proves H is a subgroup of G order p^α. □

SECOND PROOF OF THEOREM 2.7.4. Write $|G| = p^\gamma r$ where p^γ is the highest power of p that divides $|G|$. We prove more than is required. In fact, we show that G has a sequence of subgroups $P_0 \trianglelefteq P_1 \trianglelefteq \cdots \trianglelefteq P_\gamma$ such that $|P_i| = p^i$. Thus, this gives us a new proof of Theorem 2.7.1 (4). Set $P_0 = \langle e \rangle$, which has order 1. If $\gamma \geq 1$, then by Theorem 2.7.3, there exists $a \in G$ such that $P_1 = \langle a \rangle$ has order p. The method of proof is to iteratively apply Cauchy's Theorem $\gamma - 1$ times.

Inductively assume $1 \leq i < \gamma$, and that we have already constructed the sequence of subgroups $P_0 \trianglelefteq P_1 \trianglelefteq \cdots \trianglelefteq P_i$ in G, where $|P_i| = p^i$. To finish the proof it suffices to show that G has a subgroup P_{i+1} of order p^{i+1} containing P_i as a normal subgroup. By Corollary 2.2.14, $[G : P_i] = p^{\gamma - i} r$ is a multiple of p. By Lemma 2.7.2, $P_i \neq N_G(P_i)$ and p divides $[N_G(P_i) : P_i]$. Since P_i is normal in $N_G(P_i)$, by Theorem 2.7.3, the group $N_G(P_i)/P_i$. has a subgroup P'_{i+1} of order p. By Theorem 2.3.15, $P'_{i+1} = P_{i+1}/P_i$ for a subgroup P_{i+1} of $N_G(P_i)$ such that $P_i \subseteq P_{i+1} \subseteq N_G(P_i)$. By Corollary 2.2.14, $|P_{i+1}| = |P'_{i+1}||P_i| = p^{i+1}$. Since P_i is normal in $N_G(P_i)$, P_i is normal in P_{i+1}. □

By Theorem 2.7.4, if p is a prime, G is a finite group, $\alpha \geq 1$, and p^α is the highest power of p that divides $|G|$, then G has a subgroup of order p^α, call it P. In this case, we say P is a *p-Sylow subgroup* of G. Therefore, a p-Sylow subgroup is a maximal member of the set of all subgroups of G that are p-groups.

THEOREM 2.7.5. *(Sylow's Second Theorem) Let G be a finite group and p a prime that divides $|G|$. Then any two p-Sylow subgroups of G are conjugates of each other.*

PROOF. Assume G is not a p-group, otherwise there is nothing to prove. By Theorem 2.7.4, a p-Sylow subgroup exists. Let P and Q be two p-Sylow subgroups of G. We prove that there exists $x \in G$ such that $x^{-1}Px = Q$. Let $X = G/Q$ be the set of left cosets of Q in G. By Example 2.4.6, G acts on X by left multiplication. Since P is a subgroup of G, P acts on X by left multiplication. Since P is a p-group, by Theorem 2.7.1 (5), $[G : Q] = |X| \equiv |X_0| \pmod{p}$. Since p does not divide $[G : Q]$, we know $|X_0| \neq 0$. Let $xQ \in X_0$. Then for each $a \in P$, $axQ = xQ$. Thus $x^{-1}ax \in Q$ for every $a \in P$, hence $x^{-1}Px \subseteq Q$. Since $|P| = |Q| = p^\alpha$, this implies $x^{-1}Px = Q$. □

COROLLARY 2.7.6. *Let G be a finite group and p a prime that divides $|G|$. Let P be a p-Sylow subgroup of G. Then the following are true.*

(1) For every $a \in G$, $a^{-1}Pa$ is a p-Sylow subgroup of G.
(2) In G, P is the unique p-Sylow subgroup if and only if P is a normal subgroup.
(3) $N_G(N_G(P)) = N_G(P)$.

PROOF. (1): Conjugation by a is an automorphism, hence $|P| = |a^{-1}Pa|$.

(2): The subgroup P is normal in G if and only if $P = a^{-1}Pa$ for all $a \in G$, which by (1) is true if and only if P is the unique p-Sylow subgroup of G.

(3): By Proposition 2.4.13, P is a normal subgroup of $N_G(P)$. By (2), P is the unique p-Sylow subgroup of $N_G(P)$. Let $z \in N_G(N_G(P))$. Then conjugation by z is an automorphism of $N_G(P)$, hence $zPz^{-1} = P$. This implies $z \in N_G(P)$. □

THEOREM 2.7.7. *(Sylow's Third Theorem) Let G be a finite group and p a prime that divides $|G|$. The number of p-Sylow subgroups in G is congruent to 1 modulo p and divides $|G|$. More precisely, let $|G| = p^\alpha r$ where $\alpha \geq 1$ and $\gcd(p, r) = 1$. If n is the number of p-Sylow subgroups in G, then n divides r and $n \equiv 1 \pmod{p}$.*

PROOF. By Theorem 2.7.4, a p-Sylow subgroup exists. Let P be a p-Sylow subgroup. As in Example 2.4.12, let G act by conjugation on 2^G, the power set of all subsets of G. By Theorem 2.7.5, the orbit of P is the set of all p-Sylow subgroups of G. The length of the orbit is $[G : N_G(P)]$, which divides $|G|$. By Theorem 2.2.13, $r = [G : P] = [G : N_G(P)][N_G(P) : P]$. Therefore the number of conjugates of P divides r.

Let X be the set of all p-Sylow subgroups of G. The number of p-Sylow subgroups in G is equal to $|X|$. Let P act on X by conjugation. By Theorem 2.7.1 (5), $|X| \equiv |X_0| \pmod{p}$. First note that $P \in X_0$. Suppose Q is another element of X_0. Then $a^{-1}Qa = Q$ for all $a \in P$. Therefore, $P \subseteq N_G(Q)$. In this case, both P and Q are p-Sylow subgroups of $N_G(Q)$. By Corollary 2.7.6 (2), $P = Q$. This proves $X_0 = \{P\}$. We have shown that $|X| \equiv 1 \pmod{p}$. □

PROPOSITION 2.7.8. *Let G be a finite group of order n where the unique factorization of n is $p_1^{e_1} \cdots p_m^{e_m}$. Assume for each p_i that G has a unique p_i-Sylow subgroup P_i. Then G is the internal direct product of $P_1, \ldots, P_m$.*

PROOF. By Corollary 2.7.6, each P_i is a normal subgroup of G. We use induction to show that $P_1, \ldots, P_m$ satisfy the criteria of Proposition 2.5.6 (4). If $m = 1$, there is nothing to prove. If $m = 2$, then $P_1 \cap P_2 = \langle e \rangle$ by Exercise 2.2.26. Then $G = P_1P_2$ by Theorem 2.2.16. By Proposition 2.5.6 (3), G is the internal direct product of P_1 and P_2.

Assume $m > 2$. The proof is by induction on $m - r$, where $1 < r < m$. Inductively assume that

(1) $P_r \cdots P_m$ is a subgroup of G, and
(2) $P_i \cap (P_{i+1} \cdots P_m) = \langle e \rangle$ for $i = r, \ldots, m-1$.

By Proposition 2.5.6, $P_r \cdots P_m$ is isomorphic to $P_r \times \cdots \times P_m$. Then $|P_r \cdots P_m| = p_r^{e_r} \cdots p_m^{e_m}$. Because P_{r-1} is normal in G, by Exercise 2.3.20, $P_{r-1}P_r \cdots P_m$ is a subgroup of G. Because p_{r-1} is relatively prime to $|P_r \cdots P_m|$, by by Exercise 2.2.26, we know that $P_{r-1} \cap (P_r \cdots P_m) = \langle e \rangle$. By Mathematical Induction on $m - r$, this proves $P_1 \cdots P_m$ is the internal direct product of $P_1, \ldots, P_m$. Since $|P_1 \cdots P_m| = |G|$, this proves the proposition. □

7.4. Exercises.

EXERCISE 2.7.9. Let G be a finite group and N a normal subgroup of G. Show that if p is a prime and $|N| = p^r$ for some $r \geq 1$, then N is contained in every p-Sylow subgroup of G. See Exercise 2.7.12 for an application of this exercise.

EXERCISE 2.7.10. Let $n \geq 1$, A a nonempty set, and $X = A^n$ the product of n copies of A. An element x of X is an n-tuple $(x_1, \ldots, x_n)$ where each $x_i \in A$. Alternatively, an n-tuple $x = (x_1, \ldots, x_n)$ can be viewed as a function $x : \mathbb{N}_n \to A$ (see Section 1.1.3) where $x(i) = x_i$. Show that the symmetric group S_n acts on X by the rule $\sigma * x = x\sigma^{-1}$ where $x\sigma^{-1}$ refers to the composition of functions:

$$\mathbb{N}_n \xrightarrow{\sigma^{-1}} \mathbb{N}_n \xrightarrow{x} A.$$

Therefore, $\sigma * x = (x_{\sigma^{-1}(1)}, \ldots, x_{\sigma^{-1}(n)})$.

EXERCISE 2.7.11. Let G be a group containing subgroups A and B such that $A \subseteq B \subseteq G$.

(1) Give an example such that B is normal in G, A is normal in B, and A is not normal in G. We say that normal over normal is not normal.
(2) Suppose G is finite and p is a prime number. Assume B is normal in G and A is normal in B and that A is a p-Sylow subgroup of B. Prove that A is normal in G.

EXERCISE 2.7.12. Let G be a group of order $2^r \cdot 7$, where $r \geq 5$. Apply Exercises 2.4.28 and 2.7.9 to show G contains a normal subgroup N satisfying: $2^{r-4} \leq |N| \leq 2^r$ and N is contained in every 2-Sylow subgroup of G.

EXERCISE 2.7.13. Let G be a finite group of order n.

(1) Show that for each n in the list: $30, 36, 40, 42, 44, 48, 50, 52, 54, 55, 56, 75, 3^2 \cdot 5^2$, $9 \cdot 37$, G is not a simple group.
(2) Show that for each n in the list: $45, 51, 5 \cdot 17, 5^2 \cdot 17, 5^2 \cdot 37$, G is abelian.

EXERCISE 2.7.14. Let G be a group of order p^2q, where p and q are distinct primes. Show that G is not simple.

EXERCISE 2.7.15. Let G be a group of order $(p-1)p^2$, where p is an odd prime. Prove the following.

(1) G has a unique p-Sylow subgroup.
(2) There are at least four groups of order $(p-1)p^2$ which are pairwise non-isomorphic.

EXERCISE 2.7.16. Show that a group of order 105 is a semidirect product of two cyclic groups. Show how to construct an example of a nonabelian group of order 105.

EXERCISE 2.7.17. Let p be a prime and G a finite p-group of order p^n, where $n \geq 1$. Show that if N is a proper normal subgroup of G, then $N \cap Z(G)$ is a proper subroup of G.

8. Finite Abelian Groups

The purpose of this section is to prove that a finite abelian group can be decomposed into an internal direct product of cyclic subgroups. The number of cyclic subgroups and their orders are unique. These numbers are called the invariants of the finite abelian group. The generators and the cyclic subgroups themselves are not unique. This is called the Basis Theorem for finite abelian groups. As an application, we show that the group of units modulo 2^a is the direct product of a group of order 2 with a cyclic group of order 2^{a-2}.

8.1. The nth Power Map. Let A be an abelian group written multiplicatively and $n \in \mathbb{Z}$. The nth power map $\pi^n : A \to A$ is defined by the rule $\pi^n(x) = x^n$.

By Exercise 2.3.18 (where the abelian group was written additively) we see that π^n is an endomorphism of A with kernel $\{x \in A \mid |x| \text{ divides } n\}$ and image $\{x^n \mid x \in A\}$. In the following, the kernel of π^n will be denoted $A(n)$ and the image will be denoted A^n. Then $A(n)$ and A^n are subgroups of A. By the Isomorphism Theorem, Theorem 2.3.14 (1), π^n induces an isomorphism $A/A(n) \cong A^n$.

LEMMA 2.8.1. *Let $\phi : A \to B$ be an isomorphism of abelian groups. Then for any $n \in \mathbb{Z}$, the following are true.*

(1) $\phi : A(n) \to B(n)$ is an isomorphism.
(2) $\phi : A^n \to B^n$ is an isomorphism.
(3) $\phi : A/A(n) \to B/B(n)$ is an isomorphism.
(4) $\phi : A/A^n \to B/B^n$ is an isomorphism.

PROOF. (1): Let $x \in A(n)$. Then $(\phi(x))^n = \phi(x^n) = \phi(e) = e$ implies $\phi(A(n)) \subseteq B(n)$. Given $y \in B(n)$, $y = \phi(x)$ for some $x \in A$. Then $e = y^n = (\phi(x))^n = \phi(x^n)$. So $x \in \ker(\phi) = \langle e \rangle$. This proves $\phi : A(n) \to B(n)$ is an isomorphism.

(2): Let $x \in A$. Then $\phi(x^n) = (\phi(x))^n$, so $\phi(A^n) \subseteq B^n$. Let $y^n \in B^n$. Then $y = \phi(x)$ for some $x \in A$, so $y^n = (\phi(x))^n = \phi(x^n)$, which proves $\phi : A^n \to B^n$ is an isomorphism.

(3): Consider the commutative diagram

$$\begin{array}{ccc} A & \xrightarrow{\phi} & B \\ \downarrow & & \downarrow \eta \\ A/\ker(\eta\phi) & \xrightarrow{\cong} & B/B(n) \end{array}$$

where all of the maps are onto. By Part (1), the kernel of $\eta\phi$ is $\phi^{-1}(B(n)) = A(n)$. By Theorem 2.3.14 (1), $\eta\phi$ factors through $A/A(n)$ giving the isomorphism: $A/A(n) \cong B/B(n)$.

(4): Consider the commutative diagram

$$\begin{array}{ccc} A & \xrightarrow{\phi} & B \\ \downarrow & & \downarrow \eta \\ A/\ker(\eta\phi) & \xrightarrow{\cong} & B/B^n \end{array}$$

where all of the maps are onto. By Part (2), the kernel of $\eta\phi$ is $\phi^{-1}(B^n) = A^n$. By Theorem 2.3.14 (1), $\eta\phi$ factors through A/A^n giving the isomorphism: $A/A^n \cong B/B^n$. □

LEMMA 2.8.2. *Let $A = \langle a \rangle$ be an infinite cyclic group. If $n \in \mathbb{N}$, then $A(n) = \langle e \rangle$ and A/A^n is cyclic of order n.*

PROOF. We have the isomorphism $\phi : \mathbb{Z} \to A$ which is defined on generators by the rule $\phi(1) = a$ (Theorem 2.3.27 (5)). The group $\mathbb{Z}$ is written additively as in Exercise 2.3.18, and instead of the nth power map π^n, we will use the "left multiplication by n" map $\lambda_n : \mathbb{Z} \to \mathbb{Z}$. The kernel of λ_n is $\langle 0 \rangle$ and the image of λ_n is $\langle n \rangle = n\mathbb{Z}$. Applying Lemma 2.8.1 we have $A(n) = \langle e \rangle$ and $A/A^n \cong \mathbb{Z}/n\mathbb{Z}$ is cyclic of order n. □

LEMMA 2.8.3. *Let $A = \langle a \rangle$ be a finite cyclic group of order m. If $n \in \mathbb{N}$ and $d = \gcd(m, n)$, then the following are true.*

(1) $A(n) = \langle a^{m/d} \rangle$ is cyclic of order d.
(2) $A/A(n) \cong A^n$ is cyclic of order m/d.
(3) A/A^n is cyclic of order d.

PROOF. We have $A = \{e, a, \ldots, a^{m-1}\}$.

(1): Suppose $0 \leq i < m$ and $(a^i)^n = e$. Then m divides ni and by Proposition 1.2.10, $\text{lcm}(m, n) = mn/d$ divides ni. This implies m/d divides i. Hence $A(n) \subseteq \langle a^{m/d} \rangle$. But $a^{m/d}$ has order d by Lemma 2.2.18. Since d divides n, $A(n) \supseteq \langle a^{m/d} \rangle$, proving (1).

(2) and (3): By Theorem 2.3.14 (1), $A/A(n) \cong A^n$. From Part (1) and Lagrange's Theorem (Corollary 2.2.14), we get (2). From Part (2) and Lagrange's Theorem, we get (3). □

LEMMA 2.8.4. *Let A and B be abelian groups. If $n \in \mathbb{Z}$, then the following are true.*

(1) $(A \times B)(n) = A(n) \times B(n)$.
(2) $(A \times B)^n = A^n \times B^n$.

PROOF. Let (a, b) be a typical element in $A \times B$. Part (2) follows immediately from the identity $(a, b)^n = (a^n, b^n)$. Part (1) follows from $(A \times B)(n) = \{(a, b) \mid (a, b)^n = (e, e)\} = \{(a, b) \mid a^n = e \text{ and } b^n = e\} = A(n) \times B(n)$. □

LEMMA 2.8.5. *Let A be a finite abelian group, p a prime, $r \in \mathbb{N}$, and assume p^r is the highest power of p that divides $|A|$. Then $A(p^r)$ is equal to the p-Sylow subgroup of A.*

PROOF. Since A is abelian, every subgroup is normal and by Corollary 2.7.6, A has a unique p-Sylow subgroup. Call it P. Then $|P| = p^r$. If $x \in P$, then $|x|$ divides p^r by Corollary 2.2.19. As a set, $A(p^r)$ consists of those elements $x \in A$ whose order divides p^r. Therefore, $P \subseteq A(p^r)$. If $x \in A(p^r)$, then by Exercise 2.7.9, x is in P. Therefore, $A(p^r) \subseteq P$. □

8.2. The Basis Theorem.

THEOREM 2.8.6. *Every finite abelian group G is isomorphic to an internal direct product of cyclic groups.*

PROOF. Since G is abelian, every subgroup of G is normal. It follows from Proposition 2.7.8 that G is isomorphic to the internal direct product of its Sylow subgroups. Therefore, it suffices to prove the theorem for a finite p-group. From now on, assume p is a prime and $[G : e] = p^n$, for some $n \in \mathbb{N}$.

The proof is by Mathematical Induction on n. If $n = 1$, then $G \cong \mathbb{Z}/p$ is cyclic. Assume inductively that $n > 1$ and that the theorem is true for all abelian groups of order p^i where $0 < i < n$.

Let $a \in G$ be an element of maximal order. If $|a| = p^n$, then $G = \langle a \rangle$ is cyclic and we are done. Assume $|a| = p^\alpha$, where $1 \leq \alpha < n$. Set $A = \langle a \rangle$. Look at the quotient G/A. We have $|G/A| = [G : A] = p^{n-\alpha}$. By our induction hypothesis, G/A is an internal direct product of cyclic groups. That is, there exist $b_1, \ldots, b_m \in G$ such that

$$G/A = \langle [b_1] \rangle \times \cdots \times \langle [b_m] \rangle \tag{8.1}$$

where we write $[b_i]$ for the left coset $b_i A$. Assume the order of $[b_i]$ in G/A is p^{β_i}. By Exercise 2.3.44, p^{β_i} divides the order of b_i in G . Since $|a|$ is maximal, $\alpha \geq \beta_i$ for each i. Because $(b_i A)^{p^{\beta_i}} = A$, $b_i^{p^{\beta_i}} \in A$. Therefore $b_i^{p^{\beta_i}} = a^{k_i}$ for some k_i. Because the order of every element of G divides p^α, we have

$$\left(a^{k_i}\right)^{p^{\alpha-\beta_i}} = \left(b_i^{p^{\beta_i}}\right)^{p^{\alpha-\beta_i}} = b_i^{p^\alpha} = e.$$

It follows that p^α divides $k_i p^{\alpha-\beta_i}$. Hence p^{β_i} divides k_i. Write $k_i = \ell_i p^{\beta_i}$. Set $a_i = b_i a^{-\ell_i}$. Then

$$a_i^{p^{\beta_i}} = \left(b_i a^{-\ell_i}\right)^{p^{\beta_i}} = b_i^{p^{\beta_i}} a^{-\ell_i p^{\beta_i}} = a^{k_i} a^{-k_i} = e$$

which implies $|a_i| \leq p^{\beta_i}$. Since $a_i \equiv b_i \pmod{A}$, the order of a_i is greater than or equal to the order of $[b_i]$ in G/A which is p^{β_i}. This implies $|a_i| = p^{\beta_i}$. Set $A_i = \langle a_i \rangle$.

To finish the proof, we show that G is the internal direct product of the cyclic subgroups $A, A_1, \ldots, A_m$. Let $x \in G$ be an arbitrary element of G. In G/A we can write the coset xA as a product $b_1^{e_1} A \cdots b_m^{e_m} A$. Since $b_i A = a_i A$, we see that $x = a_1^{e_1} \cdots a_m^{e_m} a^{e_0}$, for some $e_0 \in \mathbb{Z}$. This proves that $G = A A_1 \cdots A_m$. Suppose $e = a^{e_0} a_1^{e_1} \cdots a_m^{e_m}$. In G/A we have $[e] = [a_1]^{e_1} \cdots [a_m]^{e_m}$ which is equal to

$[b_1]^{e_1} \cdots [b_m]^{e_m}$. As in Eq. (8.1), G/A is a direct product so $[b_i]^{e_i} = [e]$ for each i. So p^{β_i} divides e_i for each i. Therefore, $a_i^{e_i} = e$ for each i. It follows that $e = a^{e_0}$, hence e has a unique representation. □

THEOREM 2.8.7. *(Basis Theorem for Finite Abelian Groups) Let G be an abelian group of finite order. Then the following are true.*

(1) G is the internal direct product of its Sylow subgroups.
(2) If p is a prime factor of $|G|$ and P is the unique p-Sylow subgroup of G, then there exist $a_1, \ldots, a_m$ in P such that P is the internal direct product of the cyclic subgroups $\langle a_1 \rangle, \ldots, \langle a_m \rangle$, the order of a_i is equal to p^{e_i}, and $e_1 \geq e_2 \geq \cdots \geq e_m$.
(3) G is uniquely determined by the prime factors p of $|G|$ and the integers e_i that occur in (2).

The prime powers p^{e_i} that occur in (2) are called the invariants *of G. Notice that if $|P| = p^n$, then $n = e_1 + \cdots + e_m$ is a partition of the integer n.*

PROOF. Part (1) follows from Proposition 2.7.8. Part (2) follows from Theorem 2.8.6.

(3): Let A and B be finite abelian groups. First we prove that if $\phi : A \to B$ is an isomorphism, then A and B have the same invariants. Because ϕ is a one-to-one correspondence, $|A| = |B|$. Let p be a prime that divides $|A|$ (and $|B|$). By Lemmas 2.8.5 and 2.8.1, the p-Sylow subgroups of A and B are isomorphic. Using Theorem 2.8.6 we can suppose the p-Sylow subgroup of A is the internal direct product of $A_1, \ldots, A_m$ where $A_i = \langle a_i \rangle$, $|a_i| = p^{e_i}$, and $e_1 \geq e_2 \geq \cdots \geq e_m \geq 1$. Likewise, assume the p-Sylow subgroup of B is the internal direct product of $B_1, \ldots, B_n$ where $B_i = \langle b_i \rangle$, $|b_i| = p^{f_i}$, and $f_1 \geq f_2 \geq \cdots \geq f_n \geq 1$. We have the isomorphism

$$\phi : A_1 \times \cdots \times A_m \to B_1 \times \cdots \times B_n. \tag{8.2}$$

Applying Lemma 2.8.1 and the pth power map to the isomorphism (8.2), we get the isomorphisms

$$\phi : (A_1 \times \cdots \times A_m)(p) \to (B_1 \times \cdots \times B_n)(p) \tag{8.3}$$

and

$$\phi : (A_1 \times \cdots \times A_m)^p \to (B_1 \times \cdots \times B_n)^p. \tag{8.4}$$

By Lemma 2.8.4 applied $m - 1$ times, we can write (8.3) as the isomorphism

$$\phi : A_1(p) \times \cdots \times A_m(p) \to B_1(p) \times \cdots \times B_n(p) \tag{8.5}$$

and (8.4) as the isomorphism

$$\phi : A_1^p \times \cdots \times A_m^p \to B_1^p \times \cdots \times B_n^p. \tag{8.6}$$

By Lemma 2.8.3, each side of (8.5) is a direct product of cyclic groups of order p. Comparing the orders of the groups on both sides of the isomorphism (8.5), we get that $p^m = p^n$. Therefore $m = n$. Inductively, assume the uniqueness claim is true for any finite p-group of order less than $p^{e_1 + \cdots + e_m}$. By Lemma 2.8.3, the invariants of the left hand side of (8.6) are $e_1 - 1 \geq \cdots \geq e_m - 1$ and the invariants of the right hand side of (8.6) are $f_1 - 1 \geq \cdots \geq f_n - 1$. By induction, $e_i = f_i$ for each i.

For the converse, suppose we are given the cyclic groups $A_1, \ldots, A_m, B_1, \ldots, B_n$, where $|A_i| = p^{e_i}$ for each i, and $|B_j| = p^{f_i}$ for each j. If $m = n$ and $e_i = f_i$ for each i, then clearly $A_i \cong B_i$ for each i and we have $A_1 \times \cdots \times A_m \cong B_1 \times \cdots \times B_m$. □

8.3. The Group of Units Modulo 2^a. As an application of Theorem 2.8.7 we compute the invariants of the group of units in $\mathbb{Z}/2^a$, for $a \geq 1$. In the notation of Lemma 1.2.12 and Example 2.1.3, let U_{2^a} denote the group of units modulo 2^a. Then $[U_{2^a} : 1] = \phi(2^a) = 2^a - 2^{a-1} = 2^{a-1}$. The congruence classes in U_{2^a} correspond to the odd numbers in $\{0, 1, \ldots, 2^a - 1\}$. By direct computation, the reader should verify the values appearing in Table 8.1.

TABLE 8.1. Groups of units modulo 2, 2^2, 2^3 and 2^4

a	U_{2^a}	$[U_{2^a} : 1]$	invariants
1	$U_2 = \langle 1 \rangle$	1	
2	$U_4 = \langle -1 \rangle$	2	2
3	$U_8 = \langle -1 \rangle \times \langle 1 + 2 \rangle$	4	2, 2
4	$U_{16} = \langle -1 \rangle \times \langle 1 + 2 \rangle$	8	2, 2^2

If $a \geq 3$, then U_{2^a} is not cyclic. It is a direct product of a group of order 2 and a cyclic group of order 2^{a-2}.

PROPOSITION 2.8.8. *If $a \geq 3$, then U_{2^a} is an abelian 2-group of order 2^{a-1} and has invariants 2, 2^{a-2}.*

PROOF. There are two steps to the proof. In Step 1 we show that U_{2^a} is the direct sum of two cyclic groups. In Step 2 we show that U_{2^a} contains an element of order 2^{a-2}.

Step 1: Prove that $U_{2^a} = \langle \alpha \rangle \times \langle \beta \rangle$ is the direct product of two cyclic groups and the subgroup annihilated by 2 is the internal direct product $\langle -1 \rangle \times \langle 1 + 2^{a-1} \rangle$. For this step of the proof, fix a and write U instead of U_{2^a}. An arbitrary square in U is $(1 + 2x)^2 = 1 + 2^2x + 2^2x^2 = 1 + 2^2x(1 + x)$. Then $1 + 2x$ is in $U(2)$, the subgroup annihilated by 2, if and only if 2^{a-2} divides $x(1 + x)$. There are only four possibilities for x because $0 < 1 + 2x < 2^a$. The four elements of order 2 or less are listed in Table 8.2 under the column with header $1 + 2x$. This proves $U(2) = \langle -1 \rangle \times \langle 1 + 2^{a-1} \rangle$. By Theorem 2.8.7, U is a product of the form $\langle \alpha \rangle \times \langle \beta \rangle$ where -1 is the element of order 2 in $\langle \alpha \rangle$ and $1 + 2^{a-1}$ is the element of order 2 in $\langle \beta \rangle$.

Step 2: We complete the proof. Let $\theta_a : \mathbb{Z}/2^a \to \mathbb{Z}/2^{a-1}$ be the natural map. The kernel of θ_a is the cyclic group $\langle 2^{a-1} \rangle$ of order 2. Then θ_a induces a map on the groups of units $\theta_a : U_{2^a} \to U_{2^{a-1}}$ and the kernel is the cyclic group $\langle 1 + 2^{a-1} \rangle$ of order 2. Since $\langle 1 + 2^{a-1} \rangle$ is the unique subgroup of order 2 in $\langle \beta \rangle$, we see that $|\beta| = 2|\theta_a(\beta)|$. Consider the sequence

$$U_{2^a} \xrightarrow{\theta_a} U_{2^{a-1}} \xrightarrow{\theta_{a-1}} \cdots \xrightarrow{\theta_5} U_{2^4} \xrightarrow{\theta_4} U_{2^3}$$

TABLE 8.2. The four elements of order 2 or less in U_{2^a}

x	$1 + 2x$	$\lvert 1 + 2x \rvert$
2^{a-2}	$1 + 2^{a-1}$	2
2^{a-1}	1	1
$2^{a-2} - 1$	$2^{a-1} - 1 = -(1 + 2^{a-1})$	2
$2^{a-1} - 1$	-1	2

of onto group homomorphisms. From Table 8.1, $U_{2^3} = \langle -1 \rangle \times \langle 1+2 \rangle$ and $1+2$ has order 2. Recursively, we have $|\beta| = 2^{a-3}|\theta_4\theta_5 \cdots \theta_a(\beta)| = 2^{a-3}|1+2| = 2^{a-2}$. This implies that $\alpha = \theta_a(-1)$ has order 2 and the invariants of U_{2^a} are 2 and 2^{a-2}. $\square$

8.4. Exercises.

EXERCISE 2.8.9. If G is any group, and $n \in \mathbb{N}$, the direct product of n copies of G is $G^n = \prod_{i=1}^n G$. Let $G, +$ be an abelian group. Using Exercise 2.3.18, show that an n-tuple $A \in (a_1, \ldots, a_n) \in \mathbb{Z}^n$ defines a homomorphism $A : G^n \to G$ by the rule $A(x_1, \ldots, x_n) = \sum_{i=1}^n a_i x_i$.

EXERCISE 2.8.10. Let $m, n \in \mathbb{N}$. Show that the direct product $\mathbb{Z}/m \times \mathbb{Z}/n$ is cyclic if and only if $\gcd(m, n) = 1$.

EXERCISE 2.8.11. If p is a prime, and $n \geq 1$, compute the following:

(1) Let $G = \prod_{i=1}^n \mathbb{Z}/2 = \mathbb{Z}/2 \times \cdots \times \mathbb{Z}/2$ be the direct product of n copies of $\mathbb{Z}/2$. How many subgroups of order 2 are there in G?
(2) Let $G = \prod_{i=1}^n \mathbb{Z}/p = \mathbb{Z}/p \times \cdots \times \mathbb{Z}/p$ be the direct product of n copies of $\mathbb{Z}/p$. How many elements of order p are there in G? How many subgroups of order p are there in G?
(3) Let $G = \prod_{i=1}^n \mathbb{Z}/p^{e_i} = \mathbb{Z}/p^{e_1} \times \cdots \times \mathbb{Z}/p^{e_n}$ where $e_i \geq 1$ for each i. How many elements of order p are there in G? How many subgroups of order p are there in G?

EXERCISE 2.8.12. Let G be a finite abelian group. Prove that the following are equivalent:

(1) G is cyclic.
(2) For every prime factor p of $|G|$, the p-Sylow subgroup of G is cyclic.
(3) For every prime factor p of $|G|$, $G(p)$ (see Exercise 2.3.18 for this notation) is cyclic.
(4) For every $n \in \mathbb{N}$, the order of $G(n)$ is at most n.
(5) For every $n \in \mathbb{N}$, the equation $x^n = e$ has at most n solutions in G.

EXERCISE 2.8.13. Let A and B be abelian groups written additively. The set of all homomorphisms from A to B is denoted $\mathrm{Hom}(A, B)$.

(1) If $f, g \in \mathrm{Hom}(A, B)$, then $f + g$ is the function defined by the rule: $(f + g)(x) = f(x) + g(x)$. Show that this additive binary operation makes $\mathrm{Hom}(A, B)$ into an abelian group.
(2) Now consider the case where $A = B$. Show that composition of functions defines a binary operation on $\mathrm{Hom}(A, A)$ satisfying the following.
   (a) $f(gh) = (fg)h$ for all f, g, h in $\mathrm{Hom}(A, A)$. In other words, composition of functions is associative.
   (b) $f(g+h) = fg + fh$ and $(f+g)h = fh + gh$ for all f, g, h in $\mathrm{Hom}(A, A)$. In other words, composition distributes over addition.

   Together with the two binary operations of addition and composition of functions, we call $\mathrm{Hom}(A, A)$ the *ring of endomorphisms of* A.

EXERCISE 2.8.14. Let $m, n \in \mathbb{N}$ be positive integers. Show that the abelian group $\mathrm{Hom}(\mathbb{Z}/m, \mathbb{Z}/n)$ is a cyclic group of order $\gcd(m, n)$.

EXERCISE 2.8.15. Show that if G is a finite group of order at least three, then $\mathrm{Aut}(G)$ has order at least two.

EXERCISE 2.8.16. Let p be a prime and G a finite group of order p^n, where $n \geq 2$. Show that if G is not a cyclic group, then G contains a normal subgroup N such that G/N is isomorphic to the direct product $\mathbb{Z}/p \times \mathbb{Z}/p$.

EXERCISE 2.8.17. Let A be an abelian group, written additively. Show that the assignment $f \mapsto f(1)$ induces an isomorphism of abelian groups $\mathrm{Hom}(\mathbb{Z}, A) \cong A$. See Exercise 4.1.27 for a generalization of this result.

EXERCISE 2.8.18. Suppose G is a finite abelian p-group. Prove that G is cyclic if and only if G has exactly $p-1$ elements of order p.

EXERCISE 2.8.19. Let $G = \langle a \rangle$ be a finite cyclic group of order $n > 1$. Find necessary and sufficient conditions on n such that the following statement is true:

If H and K are subgroups of G, then $H \cup K$ is a subgroup of G.

9. Classification of Finite Groups

This section consists of computations and applications of the theorems from the previous sections. The examples presented here are not only intended to classify all groups of a given order, but to illustrate how the various theorems of Group Theory are applied.

EXAMPLE 2.9.1. Before going on to the new examples below, we summarize here the results on the classification of finite groups from previous sections. In the following, p and q denote distinct primes.

(1) Finite abelian groups are classified by the Basis Theorem for Finite Abelian Groups, Theorem 2.8.7.
(2) A group of order p is a finite cyclic group and is simple (Exercise 2.2.30).
(3) A group of order p^2 is abelian (Theorem 2.7.1 (2)).
(4) If G is a group of order pq, then G is either a cyclic group or a nonabelian semidirect product (Proposition 2.4.20).

9.1. Groups of Order 12. We show in this example that up to isomorphism there are exactly five groups of order 12. Let G be a finite group of order $12 = 2^2 \cdot 3$. Let P be a 2-Sylow subgroup. Then P is either $\langle a \mid a^4 = e \rangle$, a cyclic group of order 4, or P is $\langle a, b \mid a^2 = b^2 = e, ab = ba \rangle$, an isomorphic copy of the Klein four group. In both cases P is abelian. By Theorem 2.7.7, the number of conjugates of P is odd and divides 3, hence P has either 1 or 3 conjugates. Let Q be a 3-Sylow subgroup. By Theorem 2.7.7, the number of conjugates of Q divides 4, hence Q has either 1 or 4 conjugates. We know that $Q = \langle c \mid c^3 = e \rangle$ is cyclic, hence abelian. Since $P \cap Q = \langle e \rangle$, by Theorem 2.2.16 we see that $PQ = G$. We consider the following four cases.

Case 1: Assume P and Q are both normal in G. By Theorem 2.7.8, G is the internal direct product of P and Q, hence G is abelian. By Theorem 2.8.7, G is isomorphic to either

$$\mathbb{Z}/3 \times \mathbb{Z}/4$$

or

$$\mathbb{Z}/3 \times \mathbb{Z}/2 \times \mathbb{Z}/2.$$

Case 2: Assume P is normal and Q has 4 conjugates. Then Q acts by conjugation on P and there is a homomorphism $\theta : Q \to \mathrm{Aut}(P)$, where $\theta(c) = \alpha_{c^{-1}}$ is

conjugation by c^{-1}. By Corollary 2.4.17, G is isomorphic to $P \rtimes Q$, the semidirect product of P and Q. There are two subcases to consider.

Subcase 2.1: Assume $P = \langle a \rangle$ is cyclic. Then $\mathrm{Aut}(P) \cong U_4$ is a group of order two. Since Q has order three, in this case $\operatorname{im}\theta = \langle e \rangle$. Then $cac^{-1} = a$, hence G must be abelian. In this case, G is the first group of Case 1.

Subcase 2.2: Assume $P = \langle a, b \mid a^2 = b^2 = e, ab = ba \rangle$ is not cyclic. Then $\mathrm{Aut}(P)$ is isomorphic to $\mathrm{GL}_2(\mathbb{Z}/2)$. We will prove this in Proposition 4.5.7. By Exercise 2.1.26, $\mathrm{GL}_2(\mathbb{Z}/2) \cong S_3$. There are two elements of order 3 in S_3. One element of order three in $\mathrm{Aut}(P)$ is the cyclic permutation π defined by $a \mapsto b \mapsto ab \mapsto a$. The other element of order three is π^{-1}. Therefore, if $\theta(c) = \pi$, then $\theta(c^{-1}) = \pi^{-1}$. Since Q is generated by either c, or c^{-1}, without loss of generality we assume $\theta(c) = \pi$. Then $cac^{-1} = b$ and $cbc^{-1} = ab$. The semidirect product $P \rtimes Q$ has presentation in terms of generators and relations

$$\langle a, b, c \mid a^2 = b^2 = c^3 = e, ab = ba, cac^{-1} = b, cbc^{-1} = ab \rangle.$$

This group is isomorphic to A_4 by the map defined by $a \mapsto (12)(34)$, $b \mapsto (14)(23)$, $c \mapsto (123)$. The reader should verify that

$$\begin{aligned}
&(123)(12)(34)(132) = (14)(23),\\
&(123)(14)(23)(132) = (13)(24), \text{ and}\\
&(123)(13)(24)(132) = (12)(34).
\end{aligned}$$

Case 3: Assume P has 3 conjugates and Q is normal. Then P acts on Q by conjugation and there is a homomorphism $\theta : P \to \mathrm{Aut}(Q)$. Then G is the semidirect product $Q \rtimes P$. By Theorem 2.3.30, $\mathrm{Aut}(Q) \cong U_3$ is a group of order 2. The automorphism of order two is defined by $c \mapsto c^{-1}$. There are two subcases to consider.

Subcase 3.1: Assume $P = \langle a \rangle$ is cyclic. Then there is one nontrivial possibility for θ. In this case, $aca^{-1} = c^{-1}$. The presentation of the semidirect product in terms of generators and relations is

$$\langle a, c \mid a^4 = c^3 = e, aca^{-1} = c^{-1} \rangle.$$

Subcase 3.2: Assume $P = \langle a, b \mid a^2 = b^2 = e, ab = ba \rangle$ is not cyclic. Then P has three subgroups of order two. Hence there are three possible homomorphisms from P onto $\mathrm{Aut}(Q)$. Therefore, one of a, b, ab commutes with c. Since P is generated by any two of the three, without loss of generality we assume $aca = c^{-1}$ and $bcb = c$. The semidirect product is described by

$$\langle a, b, c \mid a^2 = b^2 = c^3 = e, ab = ba, aca = c^{-1}, bc = cb \rangle.$$

This group is isomorphic to D_6 the element bc has order 6, and $a(bc)a = (bc)^{-1}$. Another way to view this group is as the internal direct product $\langle b \rangle \times \langle a, c \rangle$ which is isomorphic to $\mathbb{Z}/2 \times D_3$.

Case 4: Assume P has 3 conjugates and Q has 4 conjugates. Counting elements we find that each subgroup of order 3 has 2 elements of order 3. Therefore, G has 8 elements of order 3. The subgroup P has 4 elements. Since P is not normal, the group G has more than 12 elements, which is a contradiction. Case 4 cannot occur.

9.2. Groups of Order 30. In this example we show that up to isomorphism there are exactly 4 groups of order 30. Let G be a group of order $30 = 2 \cdot 3 \cdot 5$. Using Theorems 2.7.8 and 2.5.2 we see that if G is abelian, then G is cyclic. Let P be a 2-Sylow subgroup of G, Q a 3-Sylow subgroup, and R a 5-Sylow subgroup. By Theorem 2.7.7, Q is either normal or has 10 conjugates. The number of conjugates of R is either 1 or 6. By counting elements, we see that if G has 6 subgroups of order 5 then there are 24 elements of order 5. If G has 10 subgroups of order 3, then this includes 20 elements of order 3. Since $|G| = 30$, this implies either Q is normal or R is normal. By Exercise 2.3.20, QR is a subgroup of G. Since $Q \cap R = \langle e \rangle$, by Theorem 2.2.16, $|QR| = 15$. Since $[G : QR] = 2$, Exercise 2.3.19, implies QR is normal in G. By Theorem 2.5.2, QR is cyclic. Write $QR = \langle b \rangle$. Then P acts by conjugation on QR and there is a homomorphism $\theta : P \to \mathrm{Aut}(QR) \cong U_{15}$. The image of θ has order 1 or 2. The group U_{15} has order $\phi(15) = 8$. The reader should verify that there are 4 elements in U_{15} that satisfy $x^2 \equiv 1 \pmod{15}$, they are $1, 4, -1, -4$. Therefore, if $P = \langle a \rangle$, then $aba = b^s$, where $s \in \{1, 4, -1, -4\}$. Thus G is the semidirect product $QR \rtimes P$. The presentation in terms of generators and relations is

$$G = \langle a, b \mid a^2 = b^{15} = e, aba = b^s \rangle \tag{9.1}$$

where $s \in \{1, 4, -1, -4\}$. If $s = 1$, then a commutes with b, and G is abelian. If $s = -1$, then G is isomorphic to D_{15}. By Example 2.3.36, the center of D_{15} is $\langle e \rangle$.

If $s = 4$, then because $ab^5a = b^{20} = b^5$ we see that the center of G contains b^5, an element of order 3. Then $G/\langle b^5 \rangle$ has presentation $\langle a, b \mid a^2 = b^5 = e, aba = b^4 \rangle$ which is isomorphic to D_5. Since the center of D_5 is trivial, this proves the center of G is $Z = \langle b^5 \rangle$. Since $ab^3a = b^{12} = b^{-3}$ we see that the subgroup $D = \langle a, b^3 \rangle$ has order 10 and is isomorphic to D_5. generated by a and b^3. Using Exercise 2.5.22, we see that G is the internal direct product $D \times Z$, hence G is isomorphic to $D_5 \times \mathbb{Z}/3$.

If $s = -4$, then because $ab^3a = b^{-12} = b^3$ we see that the center of G contains b^3, an element of order 5. Then $G/\langle b^3 \rangle$ has presentation $\langle a, b \mid a^2 = b^3 = e, aba = b^{-1} \rangle$ which is isomorphic to D_3. Since the center of D_3 is trivial, this proves the center of G is $Z = \langle b^3 \rangle$. Since $ab^5a = b^{-20} = b^{-5}$ we see that the subgroup $D = \langle a, b^5 \rangle$ has order 6 and is isomorphic to D_3. Using Exercise 2.5.22, we see that G is the internal direct product $D \times Z$, hence G is isomorphic to $D_3 \times \mathbb{Z}/5$.

This proves that in (9.1) the four values of s give rise to four groups that are pairwise nonisomorphic.

9.3. Groups of Order 63. We show in this example that up to isomorphism there are exactly four groups of order 63. Let G be a finite group of order $63 = 7 \cdot 3^2$. If G is abelian, then by Theorem 2.8.7, G is isomorphic to either $\mathbb{Z}/7 \times \mathbb{Z}/9$, or $\mathbb{Z}/7 \times \mathbb{Z}/3 \times \mathbb{Z}/3$. Assume from now on that G is nonabelian. Let P be a 7-Sylow subgroup. The number of conjugates of P divides 9 and is of the form $1 + 7k$. Therefore, we conclude that $k = 0$ and P is normal. Let Q be a 3-Sylow subgroup. We know that Q is abelian. Since $P \cap Q = \langle e \rangle$, by Theorem 2.2.16 we see that $PQ = G$. By Corollary 2.4.17, $G = P \rtimes Q$ and the action by Q on P is conjugation. By Example 2.4.9, the homomorphism

$$\theta : Q \to \mathrm{Aut}(P) \cong U_7$$

is defined by $\theta(x) = \alpha_{x^{-1}}$, where $\alpha_{x^{-1}}$ is the inner automorphism of P corresponding to conjugation by x^{-1}. If the image of θ is $\langle 1 \rangle$, then every element of Q commutes

with every element of P and G is abelian. By our assumption, we can assume θ is not the trivial map. By Theorem 2.3.30, $\mathrm{Aut}(P) \cong U_7$ which is an abelian group of order $\phi(7) = 6$, hence is cyclic. Since Q has order 9, this implies $\ker(\theta)$ has order 3, and $\mathrm{im}(\theta)$ has order 3. Let $P = \langle a \rangle$. There are two cases.

Case 1: $Q = \langle b \rangle$ is cyclic. Then θ maps b to $\alpha_{b^{-1}}$, the inner automorphism defined by b^{-1}, which is an element of order 3 in U_7. There are two elements of order 3 in U_7, namely [2] and [4]. Therefore, $bab^{-1} = a^i$ where $i = 2$ or 4. Notice that $|b^2| = 9$ so $Q = \langle b^2 \rangle$. Since $b^2ab^{-2} = a^{2i}$, without loss of generality we can replace b with b^2 if necessary and assume $i = 2$. Then in this case,

$$G = \langle a, b \mid a^7 = b^9 = e,\, bab^{-1} = a^2 \rangle$$

is the presentation of G in terms of generators and relations.

Case 2: Q is a direct sum of two cyclic groups of order 3. Suppose $\ker(\theta) = \langle c \rangle$ and $b \in Q - \langle c \rangle$. Then $Q = \langle b, c \rangle$. As in Case 1, $bab^{-1} = a^i$ where $i = 2$ or 4. Again, we can replace b with b^{-1} if necessary and assume $bab^{-1} = a^2$. Then in this case,

$$G = \langle a, b, c \mid a^7 = b^3 = c^3 = e,\, bc = cb,\, bab^{-1} = a^2,\, cac^{-1} = a \rangle$$

is the presentation of G.

For a continuation of this example, see Exercise 2.9.7.

9.4. Groups of Order 171. We show in this example that up to isomorphism there are exactly five groups of order 171. Let G be a finite group of order $171 = 19 \cdot 3^2$. If G is abelian, then by Theorem 2.8.7, G is isomorphic to either $\mathbb{Z}/19 \times \mathbb{Z}/9$, or $\mathbb{Z}/19 \times \mathbb{Z}/3 \times \mathbb{Z}/3$. Assume from now on that G is nonabelian. Let P be a 19-Sylow subgroup. Then $P = \langle a \rangle$ is cyclic. The number of conjugates of P divides 9 and is of the form $1 + 19k$. Therefore, we conclude that $k = 0$ and P is normal. Let Q be a 3-Sylow subgroup. We know that Q is abelian. Since $P \cap Q = \langle e \rangle$, by Theorem 2.2.16 we see that $PQ = G$. By Corollary 2.4.17, $G = P \rtimes Q$ and the action by Q on P is conjugation. By Example 2.4.9, the homomorphism

$$\theta : Q \to \mathrm{Aut}(P) \cong U_{19}$$

is defined by $\theta(x) = \alpha_{x^{-1}}$, where $\alpha_{x^{-1}}$ is the inner automorphism of P corresponding to conjugation by x^{-1}. If the image of θ is $\langle 1 \rangle$, then every element of Q commutes with every element of P and G is abelian. By our assumption, we can assume θ is not the trivial map. By Theorem 2.3.30, $\mathrm{Aut}(P) \cong U_{19}$ which is an abelian group of order $\phi(19) = 18$. Since Q has order 9, this implies $\ker(\theta)$ has order 1 or 3, and $\mathrm{im}(\theta)$ has order 3 or 9. A direct computation shows that U_{19} is cyclic and has 6 elements of order 9, namely [4], [5], [6], [9], [16], and [17]. The 2 elements of order 3 are [7] and [11]. There are three cases.

Case 1: Assume $Q = \langle b \rangle$ is cyclic and $\mathrm{im}\,\theta$ has order 9. Then θ maps Q isomorphically onto the subgroup of order 9 in $\mathrm{Aut}(P)$. If necessary, we replace b with the generator of Q that maps to $[4] \in U_{19}$. We have $bab^{-1} = a^4$. The presentation of G in terms of generators and relations is

$$G = \langle a, b \mid a^{19} = b^9 = e,\, bab^{-1} = a^4 \rangle.$$

Case 2: Assume $Q = \langle b \rangle$ is cyclic and $\mathrm{im}\,\theta$ has order 3. Then the kernel of θ is the cyclic subgroup of order 3. Under θ, an element of order 9 is mapped onto one of the elements of order 3. If necessary, we replace b with a generator of Q

that maps to $[7] \in U_{19}$. We have $bab^{-1} = a^7$. The presentation of G in terms of generators and relations is

$$G = \langle a, b \mid a^{19} = b^9 = e,\ bab^{-1} = a^7 \rangle.$$

Case 3: Assume Q is a direct sum of two cyclic groups of order 3. Since U_{19} has a unique subgroup of order 3, the kernel of θ is a group of order 3. Suppose $\ker(\theta) = \langle c \rangle$. Because the image of θ contains both [7] and [11], we pick $b \in Q - \langle c \rangle$ such that $\theta(b) = [7]$. Then $Q = \langle b, c \rangle$, $cac^{-1} = a$, and $bab^{-1} = a^7$. Then in this case,

$$G = \langle a, b, c \mid a^{19} = b^3 = c^3 = e,\ bc = cb,\ bab^{-1} = a^7,\ cac^{-1} = a \rangle$$

is the presentation of G.

9.5. Groups of Order 225. In this example we show that there are at least six nonisomorphic groups of order 225. We show how to construct two nonisomorphic nonabelian groups of order $225 = 3^2 5^2$. Let G denote a group of order 225. Let P be a 5-Sylow subgroup of G. By Theorem 2.7.7, the number of conjugates of P divides 9 and is congruent to 1 modulo 5. We conclude that P is normal in G. Let Q be a 3-Sylow subgroup of G. The number of conjugates of Q divides 25 and is congruent to 1 modulo 3. Therefore, either Q is normal in G, or Q has 25 conjugates. By Theorem 2.7.1 (2), both P and Q are abelian.

Case 1: Assume P and Q are both normal in G. By Theorem 2.7.8, G is the internal direct product of P and Q, hence G is abelian. By Theorem 2.8.7, G is isomorphic to either

$$\mathbb{Z}/9 \times \mathbb{Z}/25$$

or

$$\mathbb{Z}/9 \times \mathbb{Z}/5 \times \mathbb{Z}/5$$

or

$$\mathbb{Z}/3 \times \mathbb{Z}/3 \times \mathbb{Z}/25$$

or

$$\mathbb{Z}/3 \times \mathbb{Z}/3 \times \mathbb{Z}/5 \times \mathbb{Z}/5.$$

Case 2: Assume P is normal and Q has 25 conjugates. Then Q acts by conjugation on P and there is a homomorphism of groups $\theta : Q \to \mathrm{Aut}(P)$. There are two subcases to consider.

Subcase 2.1: Assume P is cyclic. By Theorem 2.3.30, $\mathrm{Aut}(P) \cong U_{25}$ is an abelian group of order $\phi(25) = 20$. Since $\mathrm{Aut}(P)$ has no subgroup of order 3, θ is the trivial homomorphism. Therefore, every element of Q commutes with every element of P. By Exercise 2.5.22, G is the internal direct product of P and Q, hence this case reduces to Case 1.

Subcase 2.2: Assume $P \cong \mathbb{Z}/5 \times \mathbb{Z}/5$. Then $\mathrm{Aut}(P)$ is isomorphic to $\mathrm{GL}_2(\mathbb{Z}/5)$. We will prove this in Proposition 4.5.7. As seen in Exercise 2.9.11, there are subgroups of order 3 in $\mathrm{Aut}(P)$. Without being more specific, we end this example by showing how to construct two nonisomorphic nonabelian groups of order 225. Let $\alpha \in \mathrm{Aut}(P)$ be an automorphism of P of order 3. There are two cases for Q.

Subcase 2.2.1: Assume $Q = \langle a \mid a^9 = e \rangle$ is cyclic of order 9. Then $a \mapsto \alpha$ induces $\theta : Q \to \mathrm{Aut}(P)$. The kernel of θ has order 3, the image of θ has order 3. Then the semidirect product $P \rtimes Q$ is a nonabelian group of order 225.

Subcase 2.2.2: Assume $Q = \langle a, b \mid a^3 = b^3 = e \rangle$ is a noncyclic group of order 9. Then $a \mapsto \alpha$, $b \mapsto e$ induces $\theta : Q \to \mathrm{Aut}(P)$. The kernel of θ is $\langle b \rangle$, which has order

3, the image of θ is $\langle\alpha\rangle$, which has order 3. Then the semidirect product $P \rtimes Q$ is a nonabelian group of order 225.

9.6. Groups of Order p^3. Let p be an odd prime. In this example we show that up to isomorphism there are exactly five groups of order p^3. For the classification of groups of order 8, see Exercise 2.9.9. Let G be an arbitrary group of order p^3. If G is abelian, then by Theorem 2.8.7, G is isomorphic to either $\mathbb{Z}/p^3$, $\mathbb{Z}/p^2 \times \mathbb{Z}/p$, or $\mathbb{Z}/p \times \mathbb{Z}/p \times \mathbb{Z}/p$. Assume from now on that G is nonabelian. The proof is divided into two parts. First we show in Examples 2.9.2 and 2.9.3 below that there exist two nonisomorphic nonabelian groups of order p^3. The second part of the proof shows that there are at most two nonisomorphic nonabelian groups of order p^3.

9.6.1. *Existence of Two Nonabelian Groups.* We show by example that there exist two nonisomorphic nonabelian groups of order p^3. The group in Example 2.9.2 is a semidirect product of a cyclic group of order p^2 with a cyclic group of order p. The group in Example 2.9.3 is a semidirect product of a noncyclic p-group of order p^2 with a cyclic group of order p.

EXAMPLE 2.9.2. Let $A = \langle a\rangle$ be a cyclic group of order p^2. Then $\mathrm{Aut}(A)$ is isomorphic to U_{p^2}, an abelian group of order $\phi(p^2) = p(p-1)$ (Theorem 2.3.30). Therefore, there is a unique p-Sylow subgroup of U_{p^2} of order p. Let r be any integer in $\mathbb{N}_{p^2}$ such that the congruence class $[r]$ has order p in U_{p^2}. Let $\xi \in \mathrm{Aut}(A)$ be the automorphism of A of order p defined by $\xi(a) = a^r$. If C is the cyclic group $\langle\xi\rangle$, then the semidirect product $A \rtimes C$ is a nonabelian group of order p^3 which contains a normal subgroup isomorphic to A.

EXAMPLE 2.9.3. Let F be the prime field $\mathbb{Z}/p$. Let $V = F^2 = \{(x_1, x_2) \mid x_i \in F\}$ where the binary operation on V is written additively. Then V is isomorphic to $\mathbb{Z}/p \times \mathbb{Z}/p$. Let $\theta \in \mathrm{GL}_2(F)$ be the matrix $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. Then $\theta^2 = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$, $\theta^3 = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$, $\ldots$, $\theta^{p-1} = \begin{bmatrix} 1 & 0 \\ p-1 & 1 \end{bmatrix}$, $\theta^p = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. This shows that $C = \langle\theta\rangle$ is a cyclic subgroup of $\mathrm{GL}_2(F)$ of order p. Although we have not proved it yet, using matrices and properties of Hom we will prove in Proposition 4.5.7 that $\mathrm{Aut}(V) \cong \mathrm{GL}_2(F)$. Therefore, the semidirect product $V \rtimes C$ is a nonabelian group of order p^3 containing a normal subgroup isomorphic to V. Before ending this example, we show that every element of the semidirect product has order 1 or p. Let $i \in \mathbb{Z}$. Then

$$\begin{aligned} I_2 + \theta^i + \theta^{2i} + \cdots + \theta^{(p-1)i} &= \begin{bmatrix} p & 0 \\ 0 + i + 2i + \cdots + (p-1)i & p \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ ip(p-1)/2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

Let $z = (x, \theta^i)$ be a typical element of the semidirect product $V \rtimes C$. Then

$$\begin{aligned} z^2 &= (x,\theta^i)(x,\theta^i) = (x + \theta^i(x), \theta^{2i}) = \big((I_2 + \theta^i)(x), \theta^{2i}\big) \\ z^3 &= \big((I_2 + \theta^i)(x), \theta^{2i}\big)\,(x,\theta^i) = \big((I_2 + \theta^i + \theta^{2i})(x), \theta^{3i}\big) \\ &\vdots \\ z^p &= \Big((I_2 + \theta^i + \theta^{2i} + \cdots + \theta^{(p-1)i})(x), \theta^{pi}\Big) = (0, I_2). \end{aligned}$$

This shows z has order 1 or p. The group in Example 2.9.2 has elements of order p^2, hence is not isomorphic to the group in this example.

9.6.2. *Uniqueness of Two Nonabelian Groups.* To complete our classification, we show that there are at most two nonisomorphic groups of order p^3. The proof consists of a sequence of three lemmas. In what follows, G is a nonabelian group of order p^3. The center of G is $Z(G)$. By Exercise 2.3.46, the commutator subgroup of G, denoted G', is the subgroup of G generated by the set $\{x^{-1}y^{-1}xy \mid x, y \in G\}$.

LEMMA 2.9.4. *In the above context, the commutator subgroup G' is equal to the center $Z(G)$ and is a cyclic group of order p.*

PROOF. By Theorem 2.7.1, $Z(G)$ has order p or p^2 and $G/Z(G)$ is an abelian group. If $[G : Z(G)] = p$, then $G/Z(G)$ is cyclic and by Exercise 2.3.42 (4) this cannot happen since G is nonabelian. Therefore, $Z(G)$ is a cyclic group of order p. Since $G/Z(G)$ is abelian, Exercise 2.3.46 implies $G' \subseteq Z(G)$. Since $G' \neq \langle e \rangle$ and $Z(G)$ is simple, this proves $G' = Z(G)$. □

The next lemma contains some useful commutator identities for the group G.

LEMMA 2.9.5. *In the above context, if $x, y \in G$ and $n \geq 0$, then*

(1) $(x^{-1}y^{-1}xy)^n = x^{-1}y^{-n}xy^n = x^{-n}y^{-1}x^ny$, *and*
(2) $(xy)^n = x^ny^n(x^{-1}yxy^{-1})^{\binom{n}{2}}$.

PROOF. Both parts are trivial if $n \leq 1$.

(1): We prove the first identity. The proof of the second is left to the reader. Inductively assume $n \geq 1$ and that the identity holds for n. The string of equalities

$$\begin{aligned} (x^{-1}y^{-1}xy)^{n+1} &= (x^{-1}y^{-1}xy)(x^{-1}y^{-1}xy)^n \\ &= x^{-1}y^{-1}xy(x^{-1}y^{-n}xy^n) \\ &= x^{-1}(y^{-1}xyx^{-1})y^{-n}xy^n \\ &= x^{-1}y^{-n}(y^{-1}xyx^{-1})xy^n \\ &= x^{-1}y^{-n-1}xy^{n+1}. \end{aligned}$$

follow from our induction hypothesis, and Lemma 2.9.4.

(2): Inductively assume $n \geq 1$ and that the identity holds for n. The string of equalities

$$\begin{aligned}(xy)^{n+1} &= (xy)^n xy \\ &= x^n y^n (x^{-1}yxy^{-1})^{\binom{n}{2}} xy \\ &= x^n y^n xy (x^{-1}yxy^{-1})^{\binom{n}{2}} \\ &= x^n (xx^{-1}) y^n x (y^{-n}y^n) y (x^{-1}yxy^{-1})^{\binom{n}{2}} \\ &= x^{n+1} (x^{-1}y^n xy^{-n}) y^{n+1} (x^{-1}yxy^{-1})^{\binom{n}{2}} \\ &= x^{n+1} (x^{-1}yxy^{-1})^n y^{n+1} (x^{-1}yxy^{-1})^{\binom{n}{2}} \\ &= x^{n+1} y^{n+1} (x^{-1}yxy^{-1})^n (x^{-1}yxy^{-1})^{\binom{n}{2}} \\ &= x^{n+1} y^{n+1} (x^{-1}yxy^{-1})^{\binom{n+1}{2}}\end{aligned}$$

follow from our induction hypothesis, Lemma 2.9.4, and Part (1). □

LEMMA 2.9.6. *If p is an odd prime and G is a nonabelian group of order p^3, then in terms of generators and relations, G has presentation*

(1) $\langle a, b, c \mid a^p = b^p = c^p = e,\ ac = ca,\ bc = cb,\ c = a^{-1}b^{-1}ab \rangle$, *or*
(2) $\langle a, b \mid a^{p^2} = b^p = e,\ b^{-1}ab = a^{p+1} \rangle$.

PROOF. Every element of G has order 1, p, or p^2. We consider two mutually exclusive cases.

Case 1: Assume G has no element of order p^2. Let $a, b \in G$ be any two elements of G that do not commute with each other. Then $c = a^{-1}b^{-1}ab \neq e$ and by Lemma 2.9.4, $Z(G) = \langle c \rangle$. Since a is not central, $a \notin \langle c \rangle$. Since c commutes with a, the subgroup $\langle a, c \rangle$ is abelian of order p^2 and is normal. Since a and b do not commute, $b \notin \langle a, c \rangle$. Hence b maps to a generator of the cyclic group $G/\langle a, c \rangle$. Then G has presentation in Part (1).

Case 2: Assume there exists an element a in G of order p^2. Let b be in $G - \langle a \rangle$. First we show that b can be chosen such that $|b| = p$. Assume $|b| = p^2$. Since $G/\langle a \rangle$ is abelian, we know $Z(G) = G' \subseteq \langle a \rangle$, by Lemma 2.9.4. Therefore, $Z(G)$ is equal to $\langle a^p \rangle$, the unique subgroup of order p in $\langle a \rangle$. For the same reason, $Z(G) = \langle b^p \rangle$. There exists k such that $\gcd(k, p) = 1$ and $a^p = b^{kp}$. By Lemma 2.2.18, $\langle b^k \rangle = \langle b \rangle$. Replace b with b^k and assume $b^p = a^p$. This implies $a^p b^{-p} = e$. Since p is an odd prime, $\binom{p}{2}$ is a multiple of p (Exercise 1.2.21). The exponent of $Z(G) = G'$ is p. Together with Lemma 2.9.5 (2), we have

$$\begin{aligned}(ab^{-1})^p &= a^p b^{-p} (a^{-1}b^{-1}ab)^{\binom{p}{2}} \\ &= e.\end{aligned} \tag{9.2}$$

Now G is generated by a and b, and b is in $G - \langle a \rangle$. This implies G is generated by a and ab^{-1}. By (9.2), the order of ab^{-1} is p. Replace b with ab^{-1}. We now have: b is in $G - \langle a \rangle$, $|b| = p$, and G is generated by a and b. Let $c = a^{-1}b^{-1}ab$. Then c is a generator of $Z(G) = \langle a^p \rangle$. For some j, $\gcd(j, p) = 1$ and $c^j = a^p$. By Lemma 2.9.5 (1), $c^j = a^{-1}b^{-j}ab^j = a^p$. Since $\langle b \rangle = \langle b^j \rangle$, replace b with b^j and we have $b^{-1}ab = a^{p+1}$. This shows G has presentation in Part (2) of the lemma. □

9.7. Exercises.

EXERCISE 2.9.7. This exercise is a continuation of Example 9.3. Let G be a nonabelian group of order 63. Show that G contains a cyclic subgroup N of order 21 and N is normal in G. Show that the center of G is a cyclic group of order 3.

EXERCISE 2.9.8. Classify up to isomorphism all groups of order 99.

EXERCISE 2.9.9. Show that up to isomorphism there are 5 groups of order 8, namely $\mathbb{Z}/8$, $\mathbb{Z}/4\times\mathbb{Z}/2$, $\mathbb{Z}/2\times\mathbb{Z}/2\times\mathbb{Z}/2$, the dihedral group D_4, and the quaternion 8-group Q_8 of Example 2.1.19.

EXERCISE 2.9.10. (The square roots of unity in $\mathrm{GL}_2(\mathbb{Z}/5)$) The general linear group of 2-by-2 matrices over the field $\mathbb{Z}/5$, denoted $\mathrm{GL}_2(\mathbb{Z}/5)$, is the multiplicative group of invertible matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with entries in the field $\mathbb{Z}/5$. In this exercise the reader is asked to find all matrices M in $\mathrm{GL}_2(\mathbb{Z}/5)$, such that $M^2 = I_2$, where I_2 denotes the identity matrix. The following is a suggested outline to show that there are 31 elements of order two in $\mathrm{GL}_2(\mathbb{Z}/5)$.

(1) Let $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and assume $M^2 = I_2$. Show that a, b, c, d satisfy the equations: $a^2 - d^2 = 0$, $bc = 1 - a^2$.
(2) If $a = 0$, then M is of the form $\begin{bmatrix} 0 & b \\ b^{-1} & 0 \end{bmatrix}$, where $b = 1, 2, 3, 4$, so there are 4 such matrices.
(3) If $a = \pm 1$, then M has one of the forms $\pm \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\pm \begin{bmatrix} 1 & b \\ 0 & -1 \end{bmatrix}$, $\pm \begin{bmatrix} 1 & 0 \\ c & -1 \end{bmatrix}$, where $b = 0, 1, 2, 3, 4$, $c = 1, 2, 3, 4$. There are 20 such matrices, one of them has order 1, the rest order 2.
(4) If $a = \pm 2$, then M has one of the forms $\pm \begin{bmatrix} 2 & b \\ c & -2 \end{bmatrix}$, where $bc = 2$. There are 8 such matrices.

EXERCISE 2.9.11. (The cube roots of unity in $\mathrm{GL}_2(\mathbb{Z}/5)$) The general linear group of 2-by-2 matrices over the field $\mathbb{Z}/5$, denoted $\mathrm{GL}_2(\mathbb{Z}/5)$, is the multiplicative group of invertible matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with entries in the field $\mathbb{Z}/5$. In this exercise the reader is asked to find all matrices M in $\mathrm{GL}_2(\mathbb{Z}/5)$, such that $M^3 = I_2$, where I_2 denotes the identity matrix. The following is a three-step outline to show that there are 20 elements of order three in $\mathrm{GL}_2(\mathbb{Z}/5)$.

(1) Let $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Show that if $M^2 + M + I_2 = 0$, then $M^3 = I_2$.
(2) Show that a, b, c, d satisfy the equations: $bc = -(a^2 + a + 1)$, $d = 4 - a$.
(3) Show that there are 5 choices for a and for each a there are 4 choices for the ordered triple (b, c, d).
(4) This part assumes the reader has basic familiarity with field extensions. Show that every element of order three in the ring of 2-by-2 matrices over the field $\mathbb{Z}/5$ is a root of the polynomial equation $x^2 + x + 1 = 0$. Prove that every element of order 3 in $\mathrm{GL}_2(\mathbb{Z}/5)$ is in the list of Part (3).

EXERCISE 2.9.12. Show how to construct a nonabelian group of order 75.

10. Chain Conditions

10.1. Nilpotent Groups and Solvable Groups.

DEFINITION 2.10.1. Let G be a group. Set $Z^0 = \langle e \rangle$ and $Z^1 = Z(G)$, the center of G. Then $Z^1 = \{x \in G \mid xyx^{-1}y^{-1} \in Z^0 \text{ for all } y \in G\}$. By Exercise 2.3.42, Z^1 is an abelian normal subgroup of G. Inductively assume that $n \geq 1$ and we have the chain of normal subgroups $Z^0 \subseteq Z^1 \subseteq \cdots \subseteq Z^n$ in G. Let $\eta_n : G \to G/Z^n$ be the natural map. Then Z^{n+1} is defined by the rules

$$\begin{aligned} Z^{n+1} &= \eta_n^{-1}(Z(G/Z^n)) \\ &= \{x \in G \mid xyx^{-1}y^{-1} \in Z^n \text{ for all } y \in G\}. \end{aligned}$$

By Theorem 2.3.15, Z^{n+1} is a normal subgroup of G, $Z^n \subseteq Z^{n+1}$, and the quotient group Z^{n+1}/Z^n is isomorphic to $Z(G/Z^n)$, hence is abelian. The ascending chain of subgroups $Z^0 \subseteq Z^1 \subseteq Z^2 \subseteq \cdots \subseteq Z^n \subseteq Z^{n+1} \subseteq \cdots$ is called the *ascending central series of* G.

DEFINITION 2.10.2. Let G be a group. We say G is *nilpotent*, if the ascending central series of G converges to G. That is, if $Z^n = G$ for some $n \geq 1$.

LEMMA 2.10.3. *Let p be a prime and G a finite p-group. Then G is nilpotent.*

PROOF. By Theorem 2.7.1, G has a nontrivial center. If G is abelian, then $Z^1 = G$. Otherwise, $Z^1 \subsetneq G$, and the quotient G/Z^1 is a p-group of order less than $|G|$. Since G is finite, $Z^n = G$ for some $n \geq 1$. □

LEMMA 2.10.4. *If A and B are groups, then $Z^n(A \times B) = Z^n(A) \times Z^n(B)$.*

PROOF. The proof is by induction on n. By Exercise 2.3.42, $Z(A \times B) = Z(A) \times Z(B)$, so the result is true for $n = 1$. Assume inductively that $j \geq 1$ and $Z^j(A \times B) = Z^j(A) \times Z^j(B)$. By Exercise 2.5.23,

$$\frac{A \times B}{Z^j(A \times B)} = \frac{A \times B}{Z^j(A) \times Z^j(B)} = \frac{A}{Z^j(A)} \times \frac{B}{Z^j(B)}.$$

By Exercises 2.3.42 and 2.5.23,

$$\begin{aligned} Z\left(\frac{A \times B}{Z^j(A \times B)}\right) &= Z\left(\frac{A}{Z^j(A)} \times \frac{B}{Z^j(B)}\right) \\ &= Z\left(\frac{A}{Z^j(A)}\right) \times Z\left(\frac{B}{Z^j(B)}\right) \\ &= \frac{Z^{j+1}(A)}{Z^j(A)} \times \frac{Z^{j+1}(B)}{Z^j(B)} \\ &= \frac{Z^{j+1}(A) \times Z^{j+1}(B)}{Z^j(A) \times Z^j(B)} \\ &= \frac{Z^{j+1}(A) \times Z^{j+1}(B)}{Z^j(A \times B)}. \end{aligned}$$

This proves $Z^{j+1}(A \times B)/Z^j(A \times B) = \left(Z^{j+1}(A) \times Z^{j+1}(B)\right)/Z^j(A \times B)$. It follows from Theorem 2.3.15 that $Z^{j+1}(A \times B) = Z^{j+1}(A) \times Z^{j+1}(B)$. This completes the proof. □

PROPOSITION 2.10.5. *The direct product of a finite number of nilpotent groups is nilpotent.*

PROOF. Let A and B be nilpotent groups. We show that $A \times B$ is nilpotent. A finite induction argument proves the result for a general finite product. By hypothesis, there exists $n \geq 1$ such that $A = Z^m(A)$ and $B = Z^m(B)$. By Lemma $Z^m(A \times B) = Z^m(A) \times Z^m(B) = A \times B$. □

LEMMA 2.10.6. *Let G be a nilpotent group and H a proper subgroup of G. Then H is a proper subgroup of $N_G(H)$, the normalizer of H in G.*

PROOF. For some $n \geq 1$, we are given that $Z^n = G$. Let k be the largest integer such that $Z^k \subseteq H$. Let $a \in Z^{k+1} - H$. Then $aha^{-1} \equiv h \pmod{Z^k}$ implies there exists $z \in Z^k$ such that $aha^{-1} = zh$. But $zh \in H$, hence $a \in N_G(H) - H$. □

THEOREM 2.10.7. *Let G be a finite group. Then G is nilpotent if and only if G is the internal direct product of its Sylow subgroups.*

PROOF. Assume G is a finite nilpotent group. Let p be a prime divisor of $|G|$ and P a Sylow p-subgroup of G. First we show that P is a normal subgroup of G. By Corollary 2.7.6 (3), $N_G(N_G(P)) = N_G(P)$. By Lemma 2.10.6, $N_G(P) = G$. By Proposition 2.4.13, P is a normal subgroup of $N_G(P) = G$. By Proposition 2.7.8, G is the internal direct product of its Sylow subgroups. The converse follows from Lemma 2.10.3 and Proposition 2.10.5. □

DEFINITION 2.10.8. Let G be a group. By Exercise 2.3.46, the commutator subgroup of G, denoted G', is the subgroup of G generated by the set $\{x^{-1}y^{-1}xy \mid x, y \in G\}$. Moreover, G' is a normal subgroup of G and the quotient group G/G' is abelian. Set $G^{(0)} = G$ and $G^{(1)} = G'$. Recursively, for $n \geq 1$, define $G^{(n+1)}$ to be the commutator subgroup of $G^{(n)}$. Then $G^{(n+1)}$ is a normal subgroup of $G^{(n)}$ and the quotient group $G^{(n)}/G^{(n+1)}$ is an abelian group. The descending chain of subgroups $G^{(0)} \supseteq G^{(1)} \supseteq G^{(2)} \supseteq \cdots \supseteq G^{(n)} \supseteq G^{(n+1)} \supseteq \cdots \supseteq \langle e \rangle$ is called the *derived series of G*.

DEFINITION 2.10.9. A group G is said to be *solvable* if there is a descending chain of subgroups $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_m = \langle e \rangle$ starting with G and ending with $\langle e \rangle$ such that for $0 < i \leq m$, G_i is a normal subgroup of G_{i-1} and the quotient G_i/G_{i-1} is an abelian group. In this case, we say $G_0, G_1, \ldots, G_m$ is a *solvable series* for G.

EXAMPLE 2.10.10. It is proved in Theorem 2.7.1 that a finite p-group is solvable.

EXAMPLE 2.10.11. If G is a finite abelian group, then $\langle e \rangle \subseteq G$ is a solvable series for G.

LEMMA 2.10.12. *Let G be a group. If G is nilpotent, that is, if there exists $k \geq 1$ such that $Z^k = G$, then G is solvable.*

PROOF. Assume the ascending central series $\langle e \rangle = Z^0 \subseteq Z^1 \subseteq Z^2 \subseteq \cdots \subseteq Z^{k-1} \subseteq Z^k = G$ begins at $\langle e \rangle$ and ends at G. Since each quotient Z^{n+1}/Z^n is abelian, this is a solvable series. □

LEMMA 2.10.13. *Let G be a group. Then G has a solvable series if and only if for some $k \geq 1$, the kth derived subgroup $G^{(k)}$ is equal to $\langle e \rangle$. In other words, G is solvable if and only if the derived series converges to $\langle e \rangle$.*

PROOF. If $G^{(k)} = \langle e \rangle$, then the derived series is a solvable series. Conversely, assume $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_m = \langle e \rangle$ is a solvable series. Since G_1 is a normal subgroup of G and G/G_1 is abelian, by Exercise 2.3.46 (3), $G' \subseteq G_1$. Then $\{aba^{-1}b^{-1} \mid a, b \in G'\}$ is a subset of $\{aba^{-1}b^{-1} \mid a, b \in G_1\}$. So $G^{(2)} = G'' \subseteq G_1'$. But G_2 is a normal subgroup of G_1 and G_1/G_2 is abelian, so $G_1' \subseteq G_2$. Taken together, we have $G^{(2)} \subseteq G_2$. Iterating this argument shows that $G^{(m)} \subseteq G_m = \langle e \rangle$. □

COROLLARY 2.10.14. *The symmetric group S_n is solvable if and only if $n \leq 4$.*

PROOF. If n is less than 3, then S_n is abelian. A solvable series for S_3 is $\langle e \rangle \subseteq A_3 = \langle e, (123), (132) \rangle \subseteq S_3$. It follows from Exercise 2.6.16 that $\langle e \rangle \subseteq \langle e, (12)(34), (13)(24), (14)(23) \rangle \subseteq A_4 \subseteq S_4$ is a solvable series for S_4. Let $n \geq 5$ and let $G = S_n$. By Corollary 2.6.15, $G' = A_n$. By Theorem 2.6.12, A_n is nonabelian and simple. Therefore $G' = G^{(2)} = A_n$ which implies the derived series for G converges to A_n. By Lemma 2.10.13, G is not solvable. □

10.2. Composition Series.

DEFINITION 2.10.15. Let G be a group and suppose there is a strictly descending finite chain of subgroups

$$G = G_0 \supsetneq G_1 \supsetneq G_2 \supsetneq \cdots \supsetneq G_n = \langle e \rangle$$

starting with $G = G_0$ and ending with $G_n = \langle e \rangle$. The *length* of the chain is n. A *composition series* for G is a chain such that for $i = 1, \ldots, n$, G_i is a normal subgroup of G_{i-1} and G_{i-1}/G_i is simple. If G has no composition series, define $\ell(G) = \infty$. Otherwise, let $\ell(G)$ be the minimum of the lengths of all composition series of G.

LEMMA 2.10.16. *Let G be a finite group. Then G has a composition series.*

PROOF. The reader should verify that a strictly descending chain of subgroups of maximum length such that G_i is a normal subgroup of G_{i-1} is a composition series. □

10.3. Exercises.

EXERCISE 2.10.17. Let G be a group. Prove:

(1) For each $k \geq 1$, the kth derived subgroup, $G^{(k)}$, is a normal subgroup of G.
(2) If $\theta : G \to H$ is an epimorphism, then $\theta(G^{(k)}) = H^{(k)}$.

EXERCISE 2.10.18. Let G be a group. Prove:

(1) If G is solvable and H is a subgroup of G, then H is solvable.
(2) If G is solvable and $\theta : G \to H$ is an epimorphism, then H is solvable.
(3) Let N be a normal subgroup of G. If N and G/N are solvable, then G is solvable.
(4) If $G \neq \langle e \rangle$ and G is solvable, then there exists an abelian normal subgroup $A \subseteq G$, $A \neq \langle e \rangle$.

EXERCISE 2.10.19. Let $n \geq 3$.

(1) Show that there is a homomorphism $\theta : D_{2n} \to D_n$ from the dihedral group D_{2n} onto the dihedral group D_n and the kernel of θ is the center of D_{2n}.
(2) Let 2^m be the highest power of 2 that divides n. Show that the central ascending series of D_n is $Z^{(0)} \subseteq Z^{(1)} \subseteq \cdots \subseteq Z^{(m)}$, where $Z^{(i)} = \langle R^{n/2^i} \rangle$.
(3) Show that if n is odd, then D_{2n} is the internal direct sum of a cyclic subgroup of order two (the center) and a subgroup isomorphic to D_n.

EXERCISE 2.10.20. Let G be a finite solvable group. Prove:

(1) If G is abelian and $G = G_0 \supsetneq G_1 \supsetneq \cdots \supsetneq G_m = \langle e \rangle$ is a composition series, then G_{i-1}/G_i is a cyclic group and $[G_{i-1} : G_i]$ is a prime number.
(2) G has a composition series $G = G_0 \supsetneq G_1 \supsetneq \cdots \supsetneq G_m = \langle e \rangle$ such that G_{i-1}/G_i is a cyclic group and $[G_{i-1} : G_i]$ is a prime number.

EXERCISE 2.10.21. Let H and K be groups and assume K acts on H as a group of automorphisms. Show that the semidirect product $G = H \rtimes K$ is solvable if and only if H and K are both solvable. Use this to show that the following groups are solvable.

(1) D_n, for all $n \geq 3$.
(2) Any semidirect product $G = H \rtimes K$, where H and K are both abelian groups.

CHAPTER 3

Rings

A ring is an algebraic structure which has two binary operations called addition and multiplication. We have already seen concrete examples of rings. The prototypical example of a ring is the ring of integers, $\mathbb{Z}$. Its close relative is the ring of integers modulo n, $\mathbb{Z}/(n)$. The fields $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are rings. The set $M_n(\mathbb{R})$ all of n-by-n matrices over $\mathbb{R}$ is an example of a ring in which multiplication is not commutative. The set of polynomials, the set of rational functions, and the set of power series with coefficients over the field $\mathbb{R}$ are rings. The set of all continuous functions, differentiable functions, and integrable functions from $\mathbb{R}$ to $\mathbb{R}$ are rings. The set of all functions from $\mathbb{R}$ to $\mathbb{R}$ that are continuous at a specific point is a ring. If A is an abelian group, the set $\mathrm{Hom}(A, A)$ of all endomorphisms on A is a ring. Ring Theory can be viewed as the axiomatic abstraction of these examples.

1. Definitions and Terminology

DEFINITION 3.1.1. A *ring* is a nonempty set R with two binary operations, addition written $+$, and multiplication written $\cdot$ or by juxtaposition. Under addition $(R, +)$ is an abelian group with identity element 0. Under multiplication $(R, \cdot)$ is associative and contains an identity element, denoted by 1. Multiplication distributes over addition from both the left and the right. If $(R, \cdot)$ is commutative, then we say R is a *commutative ring*. The *trivial ring* is $\{0\}$, in which $0 = 1$. If R is not the trivial ring, the reader is asked to prove in Proposition 3.1.2 that $0 \neq 1$.

PROPOSITION 3.1.2. *Let R be a ring. In the following, a, b, $a_1, \ldots, a_n$, $b_1, \ldots, b_m$ all represent elements of R and n, m are natural numbers in $\mathbb{N}$. Then the following are true.*

(1) $0a = a0 = 0$.
(2) $(-a)b = a(-b) = -(ab)$.
(3) $(-a)(-b) = ab$.
(4) $(na)b = a(nb) = n(ab)$.
(5) $\left(\sum_{i=1}^{n} a_i\right)\left(\sum_{j=1}^{m} b_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j$.
(6) If R contains more than one element, then $0 \neq 1$.

PROOF. Is left to the reader. □

DEFINITION 3.1.3. Let R be a ring and $a \in R$. We say a is a *left zero divisor* if $a \neq 0$ and there exists $b \neq 0$ such that $ab = 0$. We say a is a *right zero divisor* if $a \neq 0$ and there exists $b \neq 0$ such that $ba = 0$. If a is both a left zero divisor and right zero divisor, then we say a is a *zero divisor*. We say a is *left invertible* in case there is $b \in R$ such that $ba = 1$. We say a is *right invertible* in case there is $b \in R$ such that $ab = 1$. If a is both left invertible and right invertible, then we say a is *invertible*. In this case, the left inverse and right inverse of a are equal and unique

(Exercise 2.1.23 (2)). An invertible element in a ring R is also called a *unit* of R. If $R \neq (0)$ and R has no zero divisors, then we say R is a *domain*. A commutative domain is called an *integral domain*. A domain in which every nonzero element is invertible is called a *division ring*. A commutative division ring is called a *field*. The set of all invertible elements in a ring R is a group which is denoted $\text{Units}(R)$ or R^* and is called *the group of units in R*.

REMARK 3.1.4. Notice that in Definition 3.1.3, we have explicitly required a domain to have at least two elements. The only ring with order one is the trivial ring (0). We will see in Example 3.2.2 (4) below that $\{0\}$ plays the role of a terminal object in the category of rings. Besides this, there is no significant result that can be proved about the ring $\{0\}$. It has no proper ideals, is not a subring of any larger ring, and there is no nontrivial module or algebra over $\{0\}$.

EXAMPLE 3.1.5. Standard examples of rings and fields are listed here.

(1) The ring of integers $\mathbb{Z}$ is an integral domain.
(2) Using Proposition 1.2.9, one can verify that the ring of integers modulo n, denoted $\mathbb{Z}/(n)$, is a commutative ring containing n elements. The group of units in $\mathbb{Z}/(n)$ is $U_n = \{[u] \mid \gcd(u, n) = 1\}$ (Lemma 1.2.12). If p is a prime number, then U_p is equal to the set of all nonzero congruence classes hence $\mathbb{Z}/(p)$ is a field.
(3) Denote by $\mathbb{Q}$ the field of rational numbers, by $\mathbb{R}$ the field of real numbers and by $\mathbb{C}$ the field of complex numbers (see Section 1.4).
(4) If k is a field and $n \geq 1$, the ring of n-by-n matrices over k is denoted by $M_n(k)$. If $n > 1$, then $M_n(k)$ is noncommutative. The group of units in the ring of matrices $M_n(k)$ is called the general linear group $\text{GL}_n(k)$. When $n = 2$, we showed in Example 2.1.21 that $\text{GL}_2(k) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(F) \mid ad - bc \neq 0 \right\}$. The general result for any $n \geq 2$ is proved in Corollary 6.3.10.
(5) As in Section 1.5, if R is any ring, the ring of n-by-n matrices over R is denoted by $M_n(R)$. It follows from Propositions 1.5.1 and 1.5.2 that $M_n(R)$ is a ring.

EXAMPLE 3.1.6. Let R be a commutative ring and G a finite multiplicative group. Assume the order of G is n and enumerate the elements $G = \{g_1, \ldots, g_n\}$, starting with the group identity $g_1 = e$. Let $R(G)$ be the set of all formal sums

$$R(G) = \{r_1 g_1 + \cdots + r_n g_n \mid r_i \in R\}.$$

Define two binary operations on $R(G)$. Addition is defined by

$$\sum_{i=1}^{n} r_i g_i + \sum_{i=1}^{n} s_i g_i = \sum_{i=1}^{n} (r_i + s_i) g_i$$

and multiplication by

$$\left(\sum_{i=1}^{n} r_i g_i \right) \left(\sum_{i=1}^{n} s_i g_i \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} (r_i s_j)(g_i g_j).$$

The additive identity is $0 = 0g_1 + 0g_2 + \cdots + 0g_n$. The multiplicative identity is $1 = 1g_1 + 0g_2 + \cdots + 0g_n$. Then $R(G)$ is a ring. We call $R(G)$ a *group ring*.

If R is a commutative ring and G is a group which is not necessarily finite, we can still define the group ring $R(G)$. In this case, take $R(G)$ to be the set of all finite formal sums

$$R(G) = \left\{ \sum_{g \in G} r_g g \mid r_g \in R \text{ and } r_g = 0 \text{ for all but finitely many } g \right\}.$$

If $g \in G$, then in $R(G)$ we have the identity $gg^{-1} = g^{-1}g = 1$. Therefore, we can view G as a subgroup of the group of units in the group ring $R(G)$.

EXAMPLE 3.1.7. If A is an abelian group, let $\mathrm{Hom}(A, A)$ be the set of all homomorphisms from A to A. Turn $\mathrm{Hom}(A, A)$ into a ring by coordinate-wise addition and composition of functions:

$$(f + g)(x) = f(x) + g(x)$$
$$(fg)(x) = f(g(x)).$$

See Exercise 2.8.13. For computations of $\mathrm{Hom}(\mathbb{Z}, \mathbb{Z}) \cong \mathbb{Z}$ and $\mathrm{Hom}(\mathbb{Z}/n, \mathbb{Z}/n) \cong \mathbb{Z}/n$, see Exercises 3.1.16 and 3.1.17.

DEFINITION 3.1.8. If R is any ring, the *opposite ring of* R is denoted R^o. As an additive abelian group, the opposite ring of R is equal to R. However, the multiplication of R^o is reversed from that of R. Writing the multiplication of R by juxtaposition and multiplication of R^o with the asterisk symbol, we have $x * y = yx$.

If G is a group and G^o denotes the opposite group, then G is isomorphic to G^o (Exercise 2.1.24). If R is a noncommutative ring, then R and R^o are not necessarily isomorphic to each other. To construct a ring R such that R is not isomorphic to R^o is a subject for a more advanced book on the theory of rings. There is an example in [**9**, Exercise 7.6.26] that is attributed to Lance Small. In Ring Theory the opposite ring plays an important role. We will see in Proposition 4.5.7 that the ring of endomorphisms of a finitely generated free module over a ring R is isomorphic to the ring of matrices over R^o. The opposite ring appears frequently throughout the study of separable algebras. The interested reader is referred to the book [**10**].

DEFINITION 3.1.9. If A is a ring and $B \subseteq A$, then we say B is a *subring* of A if B contains both 0 and 1 and B is a ring under the addition and multiplication rules of A. Let A be a ring. The *center* of A is the set

$$Z(A) = \{x \in A \mid xy = yx \, (\forall y \in A)\}.$$

The reader should verify that $Z(A)$ is a subring of A and $Z(A)$ is a commutative ring. If $x \in Z(R)$, then we say x is *central*.

EXAMPLE 3.1.10. Let $R = \mathbb{Z}/6 = \{0, 1, 2, 3, 4, 5\}$ be the ring of integers modulo 6. Let $B = \{0, 2, 4\}$ and $C = \{0, 3\}$. The reader should verify that B is a ring of order 3. In fact, B is isomorphic to the field $\mathbb{Z}/3$. Since B does not contain 1, B is not a subring of R. Likewise, C is a ring, isomorphic to the field $\mathbb{Z}/2$, but C is not a subring of R. The sets B and C are examples of ideals (see Example 3.2.1).

EXAMPLE 3.1.11. If $n > 1$, then the additive group $(\mathbb{Z}/n, +)$ is generated by 1. Therefore, the ring $\mathbb{Z}/n$ has no proper subring.

EXAMPLE 3.1.12. Let R be a commutative ring and $M_n(R)$ the ring of n-by-n matrices over R, where $n \geq 2$. Let

$$L = \{(r_{ij}) \mid r_{ij} = 0 \text{ if } i < j\}$$

be the set of all lower triangular matrices in $M_n(R)$. If $A = (a_{ij})$ and $B = (b_{ij})$ are lower triangular matrices in L, then the product AB is the matrix $C = (c_{ij})$ where

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} = \begin{cases} 0 & \text{if } i < j \\ a_{ii}b_{ii} & \text{if } i = j \\ \sum_{k=j}^{i} a_{ik}b_{kj} & \text{if } i > j. \end{cases}$$

This shows that the product of lower triangular matrices is lower triangular and the product of diagonal matrices is diagonal. It follows that L is a noncommutative subring of $M_n(R)$. Likewise, the set U of all upper triangular matrices is a noncommutative subring of $M_n(R)$. The intersection $D = L \cap U$ is the set of all diagonal matrices over R. Then D is a commutative subring of $M_n(R)$. See Example 3.2.12 for a continuation of this example.

EXAMPLE 3.1.13. Let R be a commutative ring and $M_2(R)$ the ring of two-by-two matrices over R. The proof given in Example 2.3.38 can be readily adapted to show that the center of the ring $M_2(R)$ is equal to the set of scalar matrices $\left\{ \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \mid a \in R \right\}$. Let $n \geq 2$. Using a different proof, we show that the center of the ring $M_n(R)$ is equal to the set of scalar matrices over R. Let $A = (a_{ij})$ be a central matrix. For each ordered pair (i, j), where $1 \leq i, j \leq n$, let e_{ij} be the elementary matrix with 1 in position (i, j) and 0 elsewhere. In the following, we use the following notation: $C_i(A)$ denotes column i of A, $R_j(A)$ denotes row j of A, and $M_{rs}(0)$ denotes the r-by-s matrix with 0 in every position. Then

$$e_{ij}A = \begin{pmatrix} M_{i-1,n}(0) \\ R_j(A) \\ M_{n-i,n}(0) \end{pmatrix}.$$

In words, row i of $e_{ij}A$ is equal to row j of A and all other entries of $e_{ij}A$ are equal to 0. The entry in position (i, j) of $e_{ij}A$ is a_{jj}. Also,

$$Ae_{ij} = \begin{pmatrix} M_{n,j-1}(0) & C_i(A) & M_{n,n-j}(0) \end{pmatrix}.$$

In words, column j of Ae_{ij} is equal to column i of A and all other entries of $e_{ij}A$ are equal to 0. The entry in position (i, j) of $e_{ij}A$ is a_{ii}. Since A commutes with e_{ij}, we conclude that all elements of A that are not on the diagonal are equal to 0. If we assume $i \neq j$, this also means $a_{jj} = a_{ii}$. Therefore, A is a scalar matrix. It is routine to check that a scalar matrix is central.

EXAMPLE 3.1.14. If F is a field the ring of quaternions over F is the four-dimensional vector space over F with basis $\{1, i, j, k\}$ with multiplication defined by extending these relations:

$$\begin{aligned} i^2 = j^2 = k^2 &= -1 \\ ij &= -ji = k \\ ik &= -ki = -j \end{aligned}$$

by associativity and distributivity. We denote the ring of quaternions by $\mathbb{H}(F)$, or $\mathbb{H}_F$. This terminology is due to W. R. Hamilton, who first discovered the ring of

real quaternions $\mathbb{H}_{\mathbb{R}}$. Notice that when F is a field with the property that $-1 \neq 1$, then under multiplication the set $\{1, -1, i, -i, j, -j, k, -k\}$ is Q_8, the quaternion 8-group of Example 2.1.19. The ring of quaternions $\mathbb{H}_F$ is a division ring if F is equal to either $\mathbb{Q}$ or $\mathbb{R}$ (Exercise 3.1.18). The ring of quaternions $\mathbb{H}_{\mathbb{C}}$ is isomorphic to $M_2(\mathbb{C})$ (Exercise 3.1.20). The ring of quaternions $\mathbb{H}(\mathbb{Z}/(2))$ is commutative (Exercise 3.1.19). The product formula for multiplying two quaternions $x = a + bi + cj + dk$ and $y = e + fi + gj + hk$ is

$$\begin{aligned} xy &= (a + bi + cj + dk)(e + fi + gj + hk) \\ &= (ae - bf - cg - dh) + (af + be + ch - dg)i \\ &\quad + (ag - bh + ce + df)j + (ah + bg - cf + de)k \end{aligned}$$

and is derived from the relations above. We identify F with $F \cdot 1$. Thus, F is a subring of $\mathbb{H}_F$. If $x \in F$, then $xy = yx$. That is, F is a subring of the center of $\mathbb{H}_F$. For a quaternion $x = a + bi + cj + dk$ define $\chi(x) = a - bi - cj - dk$. Using the product formula above, we find

$$\begin{aligned} \chi(y)\chi(x) &= (e - fi - gj - hk)(a - bi - cj - dk) \\ &= (ae - bf - cg - dh) - (af + be + ch - dg)i \\ &\quad - (ag - bh + ce + df)j - (ah + bg - cf + de)k \\ &= \chi(xy). \end{aligned}$$

Define the *norm* of x by

$$\begin{aligned} N(x) = x\chi(x) &= (a + bi + cj + dk)(a - bi - cj - dk) \\ &= (a^2 + b^2 + c^2 + d^2) + (-ab + ab + cd - cd)i \\ &\quad + (ac + bd - ac - bd)j + (-ad - bc + bc + ad)k \\ &= a^2 + b^2 + c^2 + d^2 \end{aligned}$$

which is an element of F. Using the formulas from above, we see that

$$N(xy) = xy\chi(xy) = xy\chi(y)\chi(x) = xN(y)\chi(x) = x\chi(x)N(y) = N(x)N(y)$$

hence $N : \mathbb{H}_F \to F$ is multiplicative. The function χ is an example of an involution.

DEFINITION 3.1.15. Let R and S be rings. A function $\theta : R \to S$ is called an *isomorphism of rings*, if θ is a one-to-one correspondence, $\theta(1) = 1$, $\theta(x + y) = \theta(x) + \theta(y)$, and $\theta(xy) = \theta(x)\theta(y)$ for all $x, y \in R$. In this case, we say R and S are *isomorphic* and write $R \cong S$. From an abstract algebraic point of view, isomorphic rings are indistinguishable.

1.1. Exercises.

EXERCISE 3.1.16. The point to this exercise is to compute the ring $\mathrm{Hom}(\mathbb{Z}, \mathbb{Z})$ of all endomorphisms of the infinite cyclic group $(\mathbb{Z}, +)$ (see Exercise 2.8.13). In the following, f and g always denote endomorphisms of $\mathbb{Z}$.

(1) Define $\phi : \mathrm{Hom}((\mathbb{Z}, +), (\mathbb{Z}, +)) \to \mathbb{Z}$ by $\phi(f) = f(1)$. Show that ϕ is an isomorphism of rings.
(2) Show that $\mathrm{Aut}((\mathbb{Z}, +))$ has order two.

EXERCISE 3.1.17. Let $n \in \mathbb{N}$. The object of this exercise is to compute the ring of all endomorphisms of the finite cyclic group $(\mathbb{Z}/n, +)$. As in Exercise 2.8.13, this

ring is denoted $\mathrm{Hom}((\mathbb{Z}/n,+),(\mathbb{Z}/n,+))$. In the following, f and g always denote endomorphisms of $(\mathbb{Z}/n,+)$.

(1) Define $\phi : \mathrm{Hom}((\mathbb{Z}/n,+),(\mathbb{Z}/n,+)) \to \mathbb{Z}/n$ by $\phi(f) = f(1)$. Prove that ϕ is an isomorphism of rings.
(2) Show that $\mathrm{Aut}((\mathbb{Z}/n,+)) \cong U_n$, where U_n is the group of units modulo n.

Exercise 3.1.18. Prove that the ring of quaternions (see Example 3.1.14) over $\mathbb{Q}$ (or $\mathbb{R}$) is a division ring.

Exercise 3.1.19. Let $G = \langle a, b \mid a^2 = b^2 = e,\ ab = ba\rangle$ be an elementary 2-group of order 4. Let $R = \mathbb{Z}/(2)$ be the field with 2 elements. For the definition of the ring of quaternions, see Example 3.1.14. For the definition of a group ring, see Example 3.1.6.

(1) Prove that the ring of quaternions over R is isomorphic to the group ring $R(G)$.
(2) Determine the group of units in $R(G)$.
(3) Determine the set of zero divisors in $R(G)$.
(4) Determine all elements in $R(G)$ that satisfy the equation $e^2 = e$. These elements are the so-called idempotents.

Exercise 3.1.20. Prove that the ring of quaternions over $\mathbb{C}$ is isomorphic to $M_2(\mathbb{C})$.

Exercise 3.1.21. Let R be the ring $M_2(\mathbb{Z}/(2))$ of two-by-two matrices over $\mathbb{Z}/(2)$.

(1) Determine the group of units in R.
(2) Determine the set of zero divisors in R.
(3) Determine all elements in R that satisfy the equation $e^2 = e$. These elements are the so-called idempotents in R.
(4) Show that R contains exactly two subrings that are fields. One is the image of the canonical homomorphism $\chi : \mathbb{Z} \to R$ which has order 2, and the other is a field of order 4.

Exercise 3.1.22. Let R be any ring. Let x and y be elements of R such that $xy = yx$. Prove the Binomial Theorem:

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

for any $n \geq 0$.

Exercise 3.1.23. Let $i \in \mathbb{C}$ be the square root of -1.

(1) Show that $\mathbb{Q}[i] = \{a + bi \mid a, b \in \mathbb{Q}\}$ is a subfield of $\mathbb{C}$.
(2) Show that $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$ is a subring of $\mathbb{Q}[i]$. The ring $\mathbb{Z}[i]$ is called the ring of *gaussian integers*.

Exercise 3.1.24. Consider the set

$$\mathbb{Z}/4[i] = \{a + bi \mid a, b \in \mathbb{Z}/4\}$$

where $i^2 = -1 \equiv 3 \pmod 4$. Addition and multiplication are defined as in the gaussian integers, where a and b are added and multiplied in the ring $\mathbb{Z}/4$. Show that $\mathbb{Z}/4[i]$ is a commutative ring of order 16. Show that the group of units in

$\mathbb{Z}/4[i]$ is isomorphic to U_{16}, the group of units modulo 16. Show that the rings $\mathbb{Z}/4[i]$ and $\mathbb{Z}/16$ are not isomorphic.

EXERCISE 3.1.25. Let R be a ring.

(1) Let I be an index set and $\{S_i \mid i \in I\}$ a family of subrings of R indexed by I. Show that $\bigcap_{i\in I} S_i$ is a subring of R.
(2) Let X be a subset of R and $\mathcal{F} = \{S \mid S \text{ is a subring of } R \text{ and } X \subseteq S\}$ the family of all subrings of R containing X. Show that $T = \bigcap_{S\in\mathcal{F}} S$ is the smallest subring of R containing X. We call T the *subring of R generated by X*.
(3) Show that the set of all subrings of R is a lattice.

EXERCISE 3.1.26. Let a and b be elements of a ring R. Denote the group of units of R by R^*. Prove that $ab \in R^*$ if and only if $a \in R^*$ and $b \in R^*$.

2. Homomorphisms and Ideals

Subgroups played an important role in our study of groups. Because a ring has two binary operations, there are two completely different kinds of structures that are for rings what subgroups are for groups. These structures are subrings and ideals. Just as subgroups are the building blocks of groups, we can think of subrings and ideals as the building blocks of rings. In general, a subring is not an ideal and an ideal is not a subring.

A subring of a ring R is a subset that contains 0 and 1 and which is itself a ring under the binary operations on R (Definition 3.1.9). One way we study rings is in terms of their subrings. In ring theory, in addition to subrings we have ideals. Moreover, we distinguish between left ideals, right ideals, and two-sided ideals. A left ideal of a ring R is a subset I such that $(I, +)$ is a subgroup of $(R, +)$ and for every element r in R, left multiplication by r defines a homomorphism of additive abelian groups $\lambda_r : I \to I$ where $\lambda_r(x) = rx$. Similarly, I is a right ideal of R if for every r in R, right multiplication by r defines a homomorphism of groups $\rho_r : I \to I$ where $\rho_r(x) = xr$. If I is both a left and right ideal, then I is called a two-sided ideal. A homomorphism of rings is a function $f : R \to S$ from a ring R to a ring S which is additive, multiplicative and maps the identity element $1 \in R$ to the identity element $1 \in S$. Therefore, f is a homomorphism from the group $(R, +)$ to the group $(S, +)$. The image of a homomorphism is a subring of S and the kernel is a two-sided ideal of R. For this reason subrings and ideals play important but different roles in ring theory. An ideal I is a two-sided ideal of R if and only if the set of cosets R/I is a ring. In our analogy with groups, two-sided ideals are the counterpart for rings of normal subgroups. Theorem 3.2.15 together with its corollaries are the counterparts for rings of the group theoretic theorems with the same names.

2.1. Definitions and First Properties. Let A be a ring. A *left ideal* of A is a nonempty subset $I \subseteq A$ such that $(I, +)$ is a subgroup of $(A, +)$ and $ax \in I$ for all $a \in A$ and all $x \in I$. A *right ideal* of A is a nonempty subset $I \subseteq A$ such that $(I, +)$ is a subgroup of $(A, +)$ and $xa \in I$ for all $a \in A$ and all $x \in I$. If I is both a left ideal and right ideal, we say I is an *ideal*. For emphasis we sometimes say I is a *two-sided ideal*.

EXAMPLE 3.2.1. Some important examples of ideals are listed here.

(1) If R is a commutative ring, then a left ideal is a two-sided ideal.
(2) In a ring R the trivial ideals are $\{0\}$ and R.
(3) If F is a field, the only ideals are $\{0\}$ and F. This is Exercise 3.2.33.
(4) Let R be a commutative ring and $M_n(R)$ the ring of n-by-n matrices over R, where $n \geq 2$. The set
$$L = \{(r_{ij}) \mid r_{ij} = 0 \text{ if } i < j\}$$
of all lower triangular matrices is a subring of $M_n(R)$ (Example 3.1.12). It is not an ideal, because the identity matrix I is in L.
(5) Let F be a field and $M_2(F)$ the ring of 2-by-2 matrices over F. Then
$$I = \left\{ \begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix} \mid a, b \in F \right\}$$
is a left ideal in $M_2(F)$, but not a right ideal.
(6) The subgroups of $\mathbb{Z}, +$ are the cyclic subgroups $\mathbb{Z}m$, where $m \in \mathbb{Z}$. Any such subgroup is an ideal. So the ideals of $\mathbb{Z}$ are of the form $\mathbb{Z}m$.

If R and S are rings, a *homomorphism* from R to S is a function $f\colon R \to S$ satisfying

(1) $f(x+y) = f(x) + f(y)$ for all $x, y \in R$,
(2) $f(xy) = f(x)f(y)$ for all $x, y \in R$, and
(3) $f(1) = 1$.

Notice that (1) implies $f : (R, +) \to (S, +)$ is a homomorphism of additive groups. The *kernel* of f is $\ker(f) = \{x \in R \mid f(x) = 0\}$ which is equal to the kernel of the homomorphism on additive groups. By Exercise 3.2.28, the kernel of f is an ideal in R. By Lemma 2.3.8, f is one-to-one if and only if $\ker f = (0)$. The *image* of the homomorphism f is $\operatorname{im}(f) = \{f(x) \in S \mid x \in R\}$. By Exercise 3.2.28, the image of f is a subring of S. As in Definition 3.1.15, an isomorphism is a homomorphism $f : R \to S$ that is one-to-one and onto. An *automorphism* of R is a homomorphism $f : R \to R$ that is one-to-one and onto. An *endomorphism* is a homomorphism from R to R. A *monomorphism* is a homomorphism that is one-to-one. An *epimorphism* is a homomorphism that is onto.

Example 3.2.2. Some important examples of homomorphisms are listed here.

(1) The natural projection $\mathbb{Z} \to \mathbb{Z}/(n)$ maps an integer to its congruence class modulo n. It is a homomorphism of rings which is onto. The kernel is the subgroup generated by n.
(2) If u is an invertible element of R, the *inner automorphism* of R defined by u is $\sigma_u\colon R \to R$ where $\sigma_u(x) = u^{-1}xu$. The reader should verify that σ_u is a homomorphism of rings and is a one-to-one correspondence.
(3) Suppose R is a commutative ring, H and G are groups and $\theta\colon H \to G$ is a homomorphism of groups. The action $rh \mapsto r\theta(h)$ induces a homomorphism of group rings $R(H) \to R(G)$ (see Example 3.1.6).
   (a) The homomorphism $\langle e \rangle \to G$ induces a homomorphism $\theta : R \to R(G)$. Notice that θ is one-to-one and the image of θ is contained in the center of $R(G)$.
   (b) The homomorphism $G \to \langle e \rangle$ induces $\epsilon : R(G) \to R$. Notice that ϵ is onto, and the kernel of ϵ contains the set of elements $D = \{1 - g \mid g \in G\}$. The reader should verify that the kernel of ϵ is the ideal

generated by D in $R(G)$ (see Definition 3.2.4). Sometimes ϵ is called the *augmentation map*.

(4) If R is a ring, then the zero mapping $R \to \{0\}$ is a homomorphism of rings. In the language of categories, this says that $\{0\}$ is a terminal object in the category of rings.

(5) If R is a ring, there is a unique homomorphism $\chi : \mathbb{Z} \to R$. In fact, by definition $\chi(1) = 1$ so $\chi(n) = n\chi(1) = n1$ for an arbitrary integer n. This says that in the category of rings, $\mathbb{Z}$ is an initial object. In the lattice of all subrings of R, the image of χ is the unique minimal member. If R is a domain, the image of χ is called the *prime ring of* R. The kernel of χ is a subgroup of $\mathbb{Z}$, hence is equal to (n) for some nonnegative integer n. We call n the *characteristic* of R and write $n = \text{char}(R)$.

Proposition 3.2.3 is the counterpart for ideals of Lemma 2.3.3.

PROPOSITION 3.2.3. *Let $\phi : R \to S$ be a homomorphism of rings. Then the following are true:*

(1) If J is a left ideal in S, then $\phi^{-1}(J)$ is a left ideal in R.

(2) If ϕ is onto and A is a left ideal of R, then $\phi(A)$ is a left ideal of S.

The corresponding statements are true, if left ideal is replaced by right ideal or by two-sided ideal.

PROOF. (1): We know from group theory that $(\phi^{-1}(J), +)$ is a subgroup of $(R, +)$ (see Lemma 2.3.3). Let $x \in \phi^{-1}(J)$, $r \in R$. Then $\phi(rx) = \phi(r)\phi(x) \in J$ since $\phi(x) \in J$. Therefore, $rx \in \phi^{-1}(J)$. Hence, $\phi^{-1}(J)$ is a left ideal in R. A similar proof applies if J is a right ideal in S.

(2): We know from group theory that $(\phi(A), +)$ is a subgroup of $(S, +)$ (see Lemma 2.3.3). Let $y \in \phi(A)$ and $s \in S = \phi(R)$. Then there exist $r \in R$ and $x \in A$ such that $s = \phi(r)$ and $y = \phi(x)$. If A is a left ideal, then $rx \in A$. We have $sy = \phi(r)\phi(x) = \phi(rx) \in \phi(A)$. So $\phi(A)$ is a left ideal in S. A similar proof applies if A is a right ideal in R. □

DEFINITION 3.2.4. Let R be any ring and $X \subseteq R$. The *left ideal generated by* X is

$$\left\{ \sum_{i=1}^{n} r_i x_i \mid n \geq 1,\ r_i \in R,\ x_i \in X \right\}.$$

The reader should verify that the left ideal generated by X is equal to the intersection of the left ideals containing X (see Exercise 3.2.35). The *ideal generated by* X is

$$\left\{ \sum_{i=1}^{n} r_i x_i s_i \mid n \geq 1,\ r_i, s_i \in R,\ x_i \in X \right\}.$$

The reader should verify that the ideal generated by X is equal to the intersection of the ideals containing X (see Exercise 3.2.35). If I is the ideal generated by X, we write $I = (X)$. If A and B are left ideals of R, then $A + B$ is the set $\{a + b \mid a \in A,\ b \in B\}$. The left ideal generated by the set $\{ab \mid a \in A,\ b \in B\}$ is denoted AB. The left ideal generated by X is sometimes denoted RX. A left ideal (or ideal) is *principal* if it is generated by a single element. If $a \in R$, the principal left ideal generated by a is Ra. A commutative ring R is called a *principal ideal ring* if every ideal is a principal ideal. A *principal ideal domain* is an integral domain in which every ideal is principal. Sometimes we say R is a PID.

PROPOSITION 3.2.5. *Let R be any ring. If A and B are left ideals in R, then the following are true.*

(1) $A + B$ is a left ideal of R. If A and B are ideals, then $A + B$ is an ideal.
(2) $A + B$ is the left ideal of R generated by the set $A \cup B$.
(3) $AB = \{\sum_{i=1}^{n} x_i y_i \mid n \geq 1,\ x_i \in A,\ y_i \in B\}$. If A and B are ideals, then AB is an ideal.
(4) If $X = \{a_1, \ldots, a_n\}$ is a finite subset of R, then (X), the ideal generated by X, is equal to $(a_1) + \cdots + (a_n)$.
(5) The set of all left ideals of R, ordered by set inclusion, is a lattice. The corresponding statements are true if left ideals are replaced by right ideals or by two-sided ideals.

PROOF. The proof is left to the reader. □

EXAMPLE 3.2.6. Additional examples of ideals are listed here.

(1) In any ring, the set $\{0\}$ is an ideal.
(2) In any ring R, if u is invertible, then for any $r \in R$ we see that $r = (ru^{-1})u$ is in the left ideal generated by u. That is, $(u) = R$. We call R the *unit ideal* of R. In R, the *trivial ideals* are $\{0\}$ and R. If R is a division ring, the only left ideals in R are the trivial ideals.
(3) The ideals in $\mathbb{Z}$ are precisely the subgroups of $(\mathbb{Z}, +)$. That is, I is an ideal of $\mathbb{Z}$ if and only if $I = (n)$ for some n. The ring $\mathbb{Z}$ is a principal ideal domain.

EXAMPLE 3.2.7. Let k be a field and $R = k[w, x, y, z]$ the polynomial ring in four variables over k. Let $A = (w, x)$ and $B = (y, z)$. Then $wy + xz \in AB$, but $wy + xz$ cannot be factored as uv, where $u \in A$ and $v \in B$. This shows that in general the set $\{uv \mid u \in A,\ v \in B\}$ is not an ideal.

EXAMPLE 3.2.8. A ring R is said to be a *simple ring* if the only two-sided ideals in R are the trivial ideals. If R is a division ring, then R is a simple ring because the only ideals in R are the trivial ideals, by Example 3.2.6 (2).

EXAMPLE 3.2.9. Let k be a field. In this example we prove that $R = M_2(k)$, the ring of 2-by-2 matrices over k is a simple ring. The same proof can be modified to show $M_n(k)$ has no proper ideal for any $n \geq 1$ (see Exercise 3.2.34). Let $I \neq (0)$ be an ideal in R. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a nonzero element of I. By Proposition 1.5.3, after multiplying A by suitable permutation matrices if necessary, we can assume $a \neq 0$. Let e_{ij} denote the elementary matrix with 1 in row i column j, and 0 elsewhere. Then $e_{11}Ae_{11} = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} \in I$. Multiplying by a^{-1} shows $e_{11} \in I$. Then $P_{12}e_{11} = e_{21} \in I$, $e_{11}P_{12} = e_{12} \in I$, and $P_{12}e_{12} = e_{22} \in I$. This proves I contains $\{e_{11}, e_{12}, e_{21}, e_{22}\}$ which is a k-vector space basis for R. Hence, $I = R$.

EXAMPLE 3.2.10. Let F be a field and $M_2(F)$ the ring of 2-by-2 matrices over F. The reader should verify that $\left\{ \begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix} \mid a, c \in F \right\}$ is the principal left ideal in $M_2(F)$ generated by the elementary matrix $e_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. The principal right ideal generated by e_{21} is $\left\{ \begin{pmatrix} 0 & 0 \\ c & d \end{pmatrix} \mid c, d \in F \right\}$.

LEMMA 3.2.11. *Let R be any ring and $a \in R$. The following are equivalent.*

(1) a has a left inverse in R.
(2) $1 \in Ra$.
(3) $Ra = R$.

PROOF. (1) implies (2): We have $a^{-1} \in R$ such that $1 = a^{-1}a$.

(2) implies (3): We have $1 = ra$ for some $r \in R$. For each $x \in R$, $(xr)a = x(ra) = x \in Ra$.

(3) implies (1): $1 \in R = Ra$ implies $1 = ra$ for some $r \in R$. □

EXAMPLE 3.2.12. This is a continuation of Example 3.1.12. Let R be a commutative ring, $n \geq 2$, $M_n(R)$ the ring of n-by-n matrices over R,

$$L = \{(r_{ij}) \mid r_{ij} = 0 \text{ if } i < j\}$$

the subring of all lower triangular matrices, and D the subring of all diagonal matrices. Define $\tau : L \to D$ to be the function which maps a lower triangular matrix $A = (a_{ij})$ to the diagonal matrix $\tau(A) = \text{diag}(a_{11}, \ldots, a_{nn})$. Using the product formula for lower triangular matrices given in Example 3.1.12, one can verify that $\tau(AB) = \tau(A)\tau(B)$. It is routine to check that $\tau(A+B) = \tau(A)+\tau(B)$. For any diagonal matrix C, we have $\tau(C) = C$. Therefore, τ is an epimorphism from L onto D. The kernel of τ is the ideal

$$N = \{(r_{ij}) \mid r_{ij} = 0 \text{ if } i \leq j\}.$$

This proves that L is not a simple ring, even when R is a field. The ring L is isomorphic to the opposite ring L^o. For a hint on how to prove this, see Exercise 4.5.27.

2.2. A Fundamental Theorem on Ring Homomorphisms. Theorem 3.2.15, which is of fundamental importance, shows that a homomorphism of rings $\theta : R \to S$ factors into an onto homomorphism followed by a one-to-one homomorphism. The construction of the factorization of θ mirrors the factorization of a homomorphism of groups presented in Theorem 2.3.12. The kernel of θ is a two-sided ideal of R. The image of θ is isomorphic to the residue class ring $R/\ker\theta$. Before defining the residue class ring, we list the fundamental properties of two-sided ideals in a ring R. Lemma 3.2.13 is the counterpart for ideals of Lemma 2.3.5. By R/I we denote the set of all left cosets of $(I, +)$ in $(R, +)$. Then the factor group R/I is an abelian group under addition and the natural map $\eta : R \to R/I$ is a homomorphism of additive groups.

LEMMA 3.2.13. *Let R be a ring and I a left ideal in R. The following are equivalent.*

(1) I is a two-sided ideal of R. That is, for each $r \in R$ and $x \in I$, we have $rx \in I$ and $xr \in I$.
(2) There is a well defined multiplicative binary operation $R/I \times R/I \to R/I$ on R/I defined by the rule $(x + I, y + I) \mapsto xy + I$.
(3) There is a multiplicative binary operation on R/I such that the natural map $\eta : R \to R/I$ is a homomorphism of rings.
(4) There exists a ring S and a homomorphism of rings $\theta : R \to S$ such that $I = \ker\theta$.

PROOF. (1) implies (2): We verify that multiplication of cosets is well defined. Say $x \equiv x' \pmod{I}$ and $y \equiv y' \pmod{I}$. Then $x - x' \in I$ implies that $xy - x'y =$

$(x - x')y \in I$. Likewise $y - y' \in I$ implies that $x'y - x'y' = x'(y - y') \in I$. Taken together, we have $xy \equiv x'y \equiv x'y' \pmod{I}$.

(2) implies (3): On R/I, the associative law for multiplication, the distributive laws and the fact that $1 + I$ is the multiplicative identity are routine to check. Therefore, R/I is a ring. Let $\eta\colon R \to R/I$ be the natural map defined by $x \mapsto x+I$. Then η is a homomorphism, $\operatorname{im}\eta = R/I$, and $\ker\eta = I$.

(3) implies (4): Take S to be R/I and for θ take the natural map η.

(4) implies (1): Let $x \in \ker\theta = I$ and $r \in R$. Then $\theta(rx) = \theta(r)\theta(x) = \theta(r)0 = 0$, by Proposition 3.1.2. Likewise, $\theta(xr) = \theta(x)\theta(r) = 0\theta(r) = 0$. This prove that xr and rx are in $\ker\theta = I$. □

DEFINITION 3.2.14. Let R be a ring and I an ideal in R. The *residue class ring* is the set $R/I = \{a + I \mid a \in R\}$ of all left cosets of I in R. We sometimes call R/I the factor ring, or quotient ring of R modulo I. We define addition and multiplication of cosets by the rules

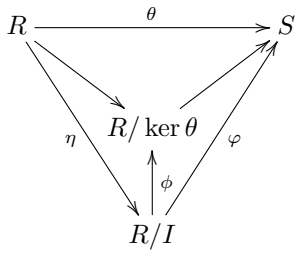
$$(a + I) + (b + I) = (a + b) + I$$
$$(a + I)(b + I) = ab + I.$$

By Lemma 3.2.13, R/I is a ring, the natural map $\eta : R \to R/I$ is a homomorphism of rings, η is onto, and $I = \ker\eta$.

Theorem 3.2.15 and Corollaries 3.2.17, 3.2.18, and 3.2.16, are the counterparts for homomorphisms of rings of Theorems 2.3.12, 2.3.14, 2.3.15, and Corollary 2.3.13.

THEOREM 3.2.15. *Let $\theta\colon R \to S$ be a homomorphism of rings. Let I be an ideal of R contained in $\ker\theta$. There exists a homomorphism $\varphi\colon R/I \to S$ satisfying the following.*

(1) $\varphi(a + I) = \theta(a)$, or in other words $\theta = \varphi\eta$.
(2) φ is the unique homomorphism from $R/I \to S$ such that $\theta = \varphi\eta$.
(3) $\operatorname{im}\theta = \operatorname{im}\varphi$.
(4) $\ker\varphi = \eta(\ker\theta) = \ker(\theta)/I$.
(5) φ is one-to-one if and only if $I = \ker\theta$.
(6) φ is onto if and only if θ is onto.
(7) There is a unique homomorphism $\phi\colon R/I \to R/\ker\theta$ such that the diagram

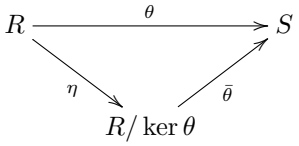


commutes.

PROOF. On the additive groups, this follows straight from Theorem 2.3.12. The map φ is multiplicative since θ is a homomorphism of rings. □

COROLLARY 3.2.16. *If $\theta : R \to S$ is a homomorphism of rings and $\eta : R \to R/\ker\theta$ is the natural map, then there exists a unique monomorphism $\bar{\theta}$ such that*

$\theta = \bar{\theta}\eta$. Hence θ factors into an epimorphism η followed by a monomorphism $\bar{\theta}$ and the diagram



commutes. There is an isomorphism of rings $\varphi : R/\ker\theta \to \operatorname{im}\theta$ where φ maps the coset $x + \ker\theta$ to $\theta(x)$

PROOF. This is Theorem 3.2.15 (5). □

COROLLARY 3.2.17. *Let R be a ring and $I \subseteq J \subseteq R$ a chain of ideals in R. Then J/I is an ideal in R/I and the natural map*

$$\frac{R/I}{J/I} \to R/J$$

sending the coset containing $x + I$ to the coset $x + J$ is an isomorphism of rings.

PROOF. This follows from Theorem 3.2.15 and Theorem 2.3.14 (3). □

COROLLARY 3.2.18. *(Correspondence Theorem) Let R be a ring and I an ideal in R. There is a one-to-one order-preserving correspondence between the ideals J such that $I \subseteq J \subseteq R$ and the ideals of R/I given by $J \mapsto J/I$.*

PROOF. This follows from Proposition 3.2.3, Theorem 3.2.15 and the Correspondence Theorem for Groups, Theorem 2.3.15. □

2.3. Prime Ideals and Integral Domains. This section focuses mostly on ideals in commutative rings. An ideal J in a commutative ring R is prime if the quotient ring R/J is an integral domain. If R/J is a field, then J is a maximal ideal. We show that a maximal ideal in R is a maximal proper ideal with respect to set inclusion. We show that for finite rings, prime ideals and maximal ideals are the same thing. In the setting of Corollary 3.2.18, prime ideals in R/I correspond to prime ideals in R containing I. The next lemma and its proof are written using symbolic expressions.

LEMMA 3.2.19. *Let R be a ring in which $0 \neq 1$. The following are equivalent, where a, b, c represent elements of R.*

(1) $(ab = 0) \to ((a = 0) \vee (b = 0))$
(2) $(a \neq 0) \to (((ab = ac) \to (b = c)) \wedge ((ba = ca) \to (b = c)))$
(3) $((a \neq 0) \wedge (b \neq 0)) \to (ab \neq 0)$

PROOF. (1) is equivalent to (3) by contraposition.
(1) implies (2):

$$\begin{aligned}((a \neq 0) \wedge (ab = ac)) &\to ((a \neq 0) \wedge (ab - ac = 0)) \\ &\to ((a \neq 0) \wedge (a(b - c) = 0)) \\ &\to ((a \neq 0) \wedge ((a = 0) \vee (b = c))) \\ &\to (b = c)\end{aligned}$$

Likewise, $((a \neq 0) \wedge (ba = ca)) \to (b = c)$.
(2) implies (1): $((a \neq 0) \wedge (ab = 0)) \to ((a \neq 0) \wedge (ab = a0)) \to (b = 0)$. □

As in Definition 3.1.3, a ring that satisfies the three equivalent statements of Lemma 3.2.19 is a domain. A commutative domain is called an integral domain.

EXAMPLE 3.2.20. If F is a field, then F is an integral domain.

(1) If R is a subring of F, then R is an integral domain.
(2) The ring of 2-by-2 matrices $M_2(F)$ is a noncommutative F-algebra. Since $M_2(F)$ contains zero divisors, it is not a domain. For example:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

THEOREM 3.2.21. *Let R be a finite domain. Then every $a \in R - \{0\}$ is invertible. In other words, R is a division ring. In particular, a finite integral domain is a field.*

PROOF. Let $a \in R - \{0\}$. Consider the "left multiplication by a" function $\lambda_a : R \to R$. Since R is an integral domain, λ_a is one-to-one, by Lemma 3.2.19 (2). Since R is finite, the Pigeonhole Principle (Exercise 1.1.11) implies λ_a is onto. So there exists $x \in R$ such that $ax = 1$. This proves a is left invertible. A symmetric argument using "right multiplication by a" shows that a is invertible. By Definition 3.1.3, R is a division ring. □

The proof of Theorem 3.2.21 shows that a finite domain is a division ring. By Wedderburn's Theorem, Theorem 5.5.11, a finite division ring is always commutative.

DEFINITION 3.2.22. Let R be a commutative ring. An ideal I in R is *prime* in case R/I is an integral domain. An ideal I in R is *maximal* in case R/I is a field. A field is an integral domain, so a maximal ideal is a prime ideal. By Definition 3.1.3, an integral domain has at least two elements, so the unit ideal is never prime.

EXAMPLE 3.2.23. Let R be a commutative ring and I an ideal in R.

(1) If R is an integral domain, then the zero ideal (0) is a prime ideal. The zero ideal is a maximal ideal in R if and only if R is a field (Exercise 3.2.33).
(2) By Theorem 3.2.21, if R/I is a finite ring, then I is a prime ideal in R if and only if I is a maximal ideal in R.
(3) If $R = \mathbb{Z}$, then by Example 3.2.1 (6) every ideal is principal, so $I = (n)$ for some $n \geq 0$. If $n > 0$, then by Example 3.1.5 (2), $\mathbb{Z}/(n)$ is a field if and only if n is prime. The maximal ideals in $\mathbb{Z}$ are the nonzero prime ideals (n), for prime numbers n.

Proposition 3.2.24 gives additional necessary and sufficient conditions for an ideal to be a prime ideal.

PROPOSITION 3.2.24. *Let R be a commutative ring and P an ideal of R. Assume $P \neq R$. The following are equivalent.*

(1) P is a prime ideal. That is, R/P is an integral domain.
(2) For all $x, y \in R$, if $xy \in P$, then $x \in P$ or $y \in P$.
(3) For any ideals I, J in R, if $IJ \subseteq P$, then $I \subseteq P$ or $J \subseteq P$.

PROOF. Is left to the reader. □

In Proposition 3.2.25 we see that the homomorphic preimage of a prime ideal is a prime ideal. This important property of prime ideals is central to most of modern

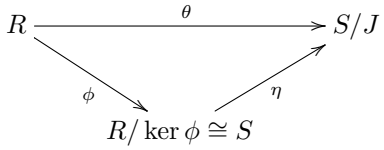
Algebraic Geometry. For a brief introduction to the subject, see the example in Section 7.4.

PROPOSITION 3.2.25. *Let $\phi : R \to S$ be a homomorphism of commutative rings. Let J be an ideal in S. Then the following are true:*

(1) If J is a prime ideal, then $\phi^{-1}(J)$ is a prime ideal.
(2) If ϕ is onto, and J is a maximal ideal, then $\phi^{-1}(J)$ is a maximal ideal.

PROOF. (1): This is Exercise 3.2.48.

(2): Let J be a maximal ideal of S. Consider the commutative diagram



where $\theta = \eta\phi$. Since ϕ and η are onto, θ is onto. By Corollary 3.2.16, $R/\ker\theta \cong S/J$. Since S/J is a field, $\ker\theta$ is a maximal ideal in R. Because ϕ is onto, we have $J = \phi\phi^{-1}(J)$. Then $\theta(\phi^{-1}(J)) = \eta(J/\ker\phi) = J$, so $\phi^{-1}(J) \subseteq \ker\theta$. If $\theta(x) = \phi(x)+J = J$, then $\phi(x) \in J$, hence $x \in \phi^{-1}(J)$. This proves $\ker\theta = \phi^{-1}(J)$. Hence $\phi^{-1}(J)$ is a maximal ideal. □

COROLLARY 3.2.26. *(Correspondence Theorem for Prime Ideals) Let R be a commutative ring and I an ideal in R. There is a one-to-one order-preserving correspondence between the ideals J such that $I \subseteq J \subseteq R$ and the ideals of R/I given by $J \mapsto J/I$. Under this correspondence prime ideals of R/I correspond to prime ideals of R that contain I.*

PROOF. The first part is Corollary 3.2.18. The preimage of a prime ideal is a prime ideal, by Proposition 3.2.25 (1). Corollary 3.2.17 shows that the image of a prime ideal that contains I is a prime ideal in R/I. □

Proposition 3.2.27 shows that an ideal M is maximal if and only if it is a maximal proper ideal with respect to the set inclusion relation. To show that maximal ideals exist, it is necessary to apply Zorn's Lemma.

PROPOSITION 3.2.27. *Let R be a commutative ring.*

(1) An ideal M is a maximal ideal in R if and only if M is not contained in a larger proper ideal of R.
(2) R contains a maximal ideal.
(3) If I is a proper ideal of R, then R contains a maximal ideal M such that $I \subseteq M$.

PROOF. (1): By Exercise 3.2.33 and Corollary 3.2.18 R/M is a field if and only if there is no proper ideal J such that $M \subsetneq J$.

(2): Let $\mathcal{S}$ be the set of all ideals I in R such that $I \neq R$. Then $(0) \in \mathcal{S}$. Order $\mathcal{S}$ by set inclusion. Let $\{A_\alpha\}$ be a chain in $\mathcal{S}$. The union $J = \bigcup A_\alpha$ is an ideal in R, by Exercise 3.2.35. Since 1 is not in any element of $\mathcal{S}$, it is clear that $1 \notin J$. Therefore, $J \in \mathcal{S}$ is an upper bound for the chain $\{A_\alpha\}$. By Zorn's Lemma, Proposition 1.3.3, $\mathcal{S}$ contains a maximal member. By Part (1), this ideal is a maximal ideal.

(3): By Part (2), R/I has a maximal ideal. By Corollary 3.2.26, there exists a maximal ideal of R containing I. □

2.4. Exercises.

EXERCISE 3.2.28. Let $\theta\colon R \to S$ be a homomorphism of rings. Prove the following.

(1) The image of θ is a subring of S.
(2) The kernel of θ is a two-sided ideal of R.
(3) If $\phi : A \to R$ is another homomorphism of rings, then the composite $\theta\phi : A \to S$ is a homomorphism of rings.

EXERCISE 3.2.29. Let $\theta\colon R \to S$ be a homomorphism of rings. Prove:

(1) θ is one-to-one if and only if $\ker\theta = (0)$.
(2) If R is a division ring, then θ is one-to-one.

EXERCISE 3.2.30. Let R be any ring.

(1) If $n = \operatorname{char} R$, then $nx = 0$ for any $x \in R$.
(2) If R is a domain, then the characteristic of R is either 0 or a prime number.

EXERCISE 3.2.31. Let R be any ring and suppose $p = \operatorname{char} R$ is a prime number. Let x and y be elements of R such that $xy = yx$. Prove:

(1) $(x + y)^p = x^p + y^p$.
(2) $(x - y)^p = x^p - y^p$.
(3) $(x - y)^{p-1} = \sum_{i=0}^{p-1} x^i y^{p-1-i}$.
(4) If $n \geq 0$, then $(x + y)^{p^n} = x^{p^n} + y^{p^n}$.

See Exercise 3.6.35 for an application of this exercise.

EXERCISE 3.2.32. Let R be a commutative ring and assume $\operatorname{char} R = p$ is a prime number. Define $\theta : R \to R$ by $x \mapsto x^p$. Show that θ is a homomorphism of rings. We call θ the *Frobenius homomorphism*, after F. G. Frobenius. For any $a \geq 1$, show that $\theta^a(x) = x^{p^a}$. If R is a field, show that θ is one-to-one.

EXERCISE 3.2.33. Prove:

(1) If R is a ring with no proper left ideal, then every nonzero element has a left inverse.
(2) If R is a ring with no proper left ideal, then R is a division ring.
(3) A commutative ring R is a field if and only if R has no proper ideal.

EXERCISE 3.2.34. This exercise is a continuation of Example 3.2.9. Let R be a ring and $M_n(R)$ the ring of n-by-n matrices over R where addition and multiplication are defined in the usual way.

(1) Let e_{ij} be the elementary matrix which has 0 in every position except in position (i, j) where there is 1. Determine the left ideal in $M_n(R)$ generated by e_{ij}.
(2) If $n \geq 2$, show that $M_n(R)$ has proper left ideals.
(3) If I is an ideal in $M_n(R)$, show that $I = M_n(J)$ for some ideal J in R.
(4) If D is a division ring, show that $M_n(D)$ has no proper ideal. We say that $M_n(D)$ is a *simple ring*.

EXERCISE 3.2.35. Let R be a ring, I an index set, and $\{A_i \mid i \in I\}$ a family of left ideals in R.

(1) Show that $\bigcap_{i\in I} A_i$ is a left ideal in R.

(2) Suppose $\{A_i \mid i \in I\}$ is an ascending chain of left ideals in R. That is, I is a partially ordered set that is a chain, and if $\alpha \leq \beta$ in I, then $A_\alpha \subseteq A_\beta$. Show that $\bigcup_{i\in I} A_i$ is a left ideal in R.

EXERCISE 3.2.36. Let U and V be ideals in the commutative ring R. As in Definition 3.2.4, UV is the ideal generated by the set $\{uv \mid u \in U, v \in V\}$. Prove the following.

(1) $UV \subseteq U \cap V$.
(2) If $U + V = R$, then $UV = U \cap V$.
(3) Show by counterexample that $UV = U \cap V$ is false in general.

EXERCISE 3.2.37. Let $n > 1$.

(1) Show that every prime ideal in $\mathbb{Z}/(n)$ is a maximal ideal.
(2) Let $n = \pi_1^{e_1} \cdots \pi_k^{e_k}$ be the unique factorization of n (Proposition 1.2.7). Determine the maximal ideals in $\mathbb{Z}/(n)$.

EXERCISE 3.2.38. An element x of a ring is said to be *nilpotent* if $x^n = 0$ for some $n > 0$. If R is a commutative ring, let $\operatorname{Rad}_R(0)$ denote the set of all nilpotent elements of R. We call $\operatorname{Rad}_R(0)$ the *nil radical* of R.

(1) Show that $\operatorname{Rad}_R(0)$ is an ideal.
(2) Let I be an ideal of R contained in $\operatorname{Rad}_R(0)$. Show that the nil radical of R/I is $\operatorname{Rad}_R(0)/I$, hence the nil radical of $R/\operatorname{Rad}_R(0)$ is the trivial ideal $(0 + \operatorname{Rad}_R(0))$.
(3) Show that $\operatorname{Rad}_R(0) \subseteq P$, if P is a prime ideal in R.

EXERCISE 3.2.39. Let $\theta\colon R \to S$ be a homomorphism of rings. Prove that θ induces a homomorphism $\theta\colon \operatorname{Units}(R) \to \operatorname{Units}(S)$ on the groups of units.

EXERCISE 3.2.40. Let R be a commutative ring, $\operatorname{Rad}_R(0)$ the nil radical of R, and $\eta\colon R \to R/\operatorname{Rad}_R(0)$ the natural map. Prove:

(1) If x is a nilpotent element of R, then $1 + x$ is a unit in R.
(2) If $\eta(r)$ is a unit in $R/\operatorname{Rad}_R(0)$, then r is a unit in R.
(3) Let I be an ideal of R contained in $\operatorname{Rad}_R(0)$. Then the natural map $\eta\colon \operatorname{Units}(R) \to \operatorname{Units}(R/I)$ is onto and the kernel of η is equal to the coset $1 + I$.

EXERCISE 3.2.41. Let I and J be ideals in the commutative ring R. The *ideal quotient* is $I : J = \{x \in R \mid xJ \subseteq I\}$. Prove that $I : J$ is an ideal in R.

EXERCISE 3.2.42. For the following, let I, J and K be ideals in the commutative ring R. Prove that the ideal quotient satisfies the following properties.

(1) $I \subseteq I : J$
(2) $(I : J)J \subseteq I$
(3) $(I : J) : K = I : JK = (I : K) : J$
(4) If $\{I_\alpha \mid \alpha \in S\}$ is a collection of ideals in R, then
$$\left(\bigcap_{\alpha\in S} I_\alpha\right) : J = \bigcap_{\alpha\in S} (I_\alpha : J)$$
(5) If $\{J_\alpha \mid \alpha \in S\}$ is a collection of ideals in R, then
$$I : \sum_{\alpha\in S} J_\alpha = \bigcap_{\alpha\in S} (I : J_\alpha)$$

EXERCISE 3.2.43. A *local ring* is a commutative ring R such that R has exactly one maximal ideal. If R is a local ring with maximal ideal $\mathfrak{m}$, then $R/\mathfrak{m}$ is called the *residue field* of R. If $(R, \mathfrak{m})$ and $(S, \mathfrak{n})$ are local rings and $f : R \to S$ is a homomorphism of rings, then we say f is a *local homomorphism of local rings* in case $f(\mathfrak{m}) \subseteq \mathfrak{n}$. Prove:

(1) A field is a local ring.
(2) If $(R, \mathfrak{m})$ is a local ring, then the group of units of R is equal to the set $R - \mathfrak{m}$.
(3) If $f : R \to S$ is a local homomorphism of local rings, then f induces a homomorphism of residue fields $R/\mathfrak{m} \to S/\mathfrak{n}$.

EXERCISE 3.2.44. Let R be a ring. If A and B are left ideals in R, then the product ideal AB is defined in Definition 3.2.4. The powers of A are defined recursively by the rule:

$$A^n = \begin{cases} R & \text{if } n = 0, \\ A & \text{if } n = 1, \\ AA^{n-1} & \text{if } n > 1. \end{cases}$$

The left ideal A is *nilpotent* if for some $n > 0$, $A^n = 0$. Let A and B be nilpotent left ideals of R. Prove:

(1) Assume $A^n = 0$. If $x_1, \ldots, x_n$ are elements of A, then $x_1 \cdots x_n = 0$.
(2) Every element x of A is nilpotent.
(3) $A + B$ is a nilpotent left ideal.

EXERCISE 3.2.45. Let R be a commutative ring and $\{x_1, \ldots, x_n\}$ a finite set of nilpotent elements of R. Show that $Rx_1 + \cdots + Rx_n$ is a nilpotent ideal.

EXERCISE 3.2.46. Let R be a ring. We say that a left ideal M of R is *maximal* if M is not equal to R and if I is a left ideal such that $M \subseteq I \subsetneq R$, then $M = I$. Let I be a left ideal of R which is not the unit ideal. Apply Zorn's Lemma, Proposition 1.3.3, to show that there exists a maximal left ideal M such that $I \subseteq M \subsetneq R$.

EXERCISE 3.2.47. Prove Proposition 3.2.24.

EXERCISE 3.2.48. Prove Proposition 3.2.25 (1).

EXERCISE 3.2.49. If R is a commutative ring, let $\mathrm{Aut}(R)$ denote the group of all ring automorphisms of R. Prove the following.

(1) $\mathrm{Aut}(\mathbb{Z}) = (1)$.
(2) $\mathrm{Aut}(\mathbb{Z}/(n)) = (1)$ for any n.

EXERCISE 3.2.50. Let R be a commutative ring and G a group. Show that the group ring $R(G)$ (see Example 3.1.6) is isomorphic to the opposite ring $R(G)^o$ (see Definition 3.1.8).

EXERCISE 3.2.51. Let R be a ring and $\mathrm{Aut}(R)$ the group of all ring automorphisms of R. Let R^* be the group of units of R. If $u \in R^*$, the inner automorphism defined by u is $\sigma_u : R \to R$ which is the function defined by $\sigma_u(x) = u^{-1}xu$ (see Example 3.2.2 (2)). Show that the assignment $\theta(u) = \sigma_{u^{-1}}$ defines a homomorphism of groups $\theta : R^* \to \mathrm{Aut}(R)$. Show that the image of θ is a normal subgroup of $\mathrm{Aut}(R)$. The image of θ is called the *group of inner automorphisms of* R and is denoted $\mathrm{Inn}(R)$.

EXERCISE 3.2.52. Let R be a ring. For every $r \in R$, let $\lambda_r : R \to R$ be "left multiplication by r". That is, $\lambda_r(x) = rx$. Similarly, let $\rho_r : R \to R$ be "right multiplication by r", where $\rho_r(x) = xr$. By Example 3.1.7, if I is an ideal (left, right or two-sided), then $\mathrm{Hom}(I, I)$ is a ring.

(1) Let I be a left ideal of R. Show that $\lambda : R \to \mathrm{Hom}(I, I)$ is a homomorphism of rings, where $\lambda(r) = \lambda_r$.
(2) Let I be a right ideal of R. As in Definition 3.1.8, R^o denotes the opposite ring of R. Show that $\rho : R^o \to \mathrm{Hom}(I, I)$ is a homomorphism of rings, where $\rho(r) = \rho_r$.

EXERCISE 3.2.53. Let R be a ring and I a proper left ideal in R. Assume the group $(I, +)$ is cyclic, isomorphic to $\mathbb{Z}/n$, where $n = 0$ is allowed. Prove that R contains a two-sided ideal A such that the ring R/A is isomorphic to the ring $\mathbb{Z}/n$.

3. Direct Product and Direct Sum of Rings

As with groups, we define the direct product of an arbitrary family of rings. For the definition of an internal direct sum, we limit our attention to a finite family of ideals. The direct product of a family of rings is a ring where the binary operations are coordinate-wise addition and coordinate-wise multiplication. A ring R is an internal direct sum of a finite family of ideals $A_1, \ldots, A_n$ provided each A_i is a ring and the summation map $A_1 \times \cdots \times A_n \to R$ is an isomorphism of rings. For rings there is a version of the Chinese Remainder Theorem that generalizes the theorem for the product of finite cyclic groups.

3.1. External Direct Product.

DEFINITION 3.3.1. Let $\{R_i \mid i \in I\}$ be a family of rings. For each $i \in I$, the same symbol 0 is used to denote the additive identity of each R_i. Likewise, 1 denotes the multiplicative identity of each R_i. As defined in Definition 1.3.4, the product of $\{R_i \mid i \in I\}$ is

$$\prod_{i \in I} R_i = \left\{ f : I \to \bigcup_{i \in I} R_i \mid f(i) \in R_i \right\}.$$

The product can be turned into a ring, if addition and multiplication operations are defined coordinate-wise:

$$(f + g)(i) = f(i) + g(i)$$
$$(fg)(i) = f(i)g(i).$$

Since each R_i contains 0, the constant function $0(i) = 0$ is the additive identity on the product. Since each R_i contains 1, the constant function $1(i) = 1$ is the multiplicative identity on the product. The other ring axioms hold in the product because they hold coordinate-wise. The ring $\prod_{i \in I} R_i$ is called the *direct product* of the family $\{R_i \mid i \in I\}$. The additive abelian group structure on $\prod_{i \in I} R_i$ is the direct product of the additive groups $\{(R_i, +) \mid i \in I\}$ as defined in Definition 2.5.1. It is routine to verify that for each $k \in I$ the canonical projection map

$$\pi_k : \prod_{i \in I} R_i \to R_k$$

is an onto homomorphism of rings. Also from Definition 2.5.1 there is a canonical one-to-one homomorphism of additive groups

$$\iota_k : R_k \to \prod_{i \in I} R_i$$

where $\iota_k(x)$ is equal to x in coordinate k, and 0 elsewhere. Moreover, ι_k is multiplicative, meaning $\iota_k(xy) = \iota_k(x)\iota_k(y)$ and we have $\pi_k \iota_k = 1_{R_k}$. The function ι_k is not a homomorphism of rings, since $\iota_k(1) \neq 1$.

If $I = \{1, 2, \ldots, n\}$, then

$$\prod_{i=1}^{n} R_i = R_1 \times R_2 \times \cdots \times R_n = \{(x_1, \ldots, x_n) \mid x_i \in R_i\}$$

and on n-tuples the binary operations are given by

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$
$$(x_1, \ldots, x_n)(y_1, \ldots, y_n) = (x_1 y_1, \ldots, x_n y_n).$$

3.2. Internal Direct Sum.

DEFINITION 3.3.2. Let $\{I_1, \ldots, I_n\}$ be a set of ideals in a ring R. In Definition 3.2.4 the sum $A + B$ of two ideals in R is defined. For $n \geq 2$, $I_1 + I_2 + \cdots + I_n$ is defined recursively to be $(I_1 + \cdots + I_{n-1}) + I_n$ and is called the *sum* of the ideals. The reader should verify that the sum of the ideals is equal to the ideal of R generated by the set $I_1 \cup I_2 \cup \cdots \cup I_n$. We say that R is the *internal direct sum* of the ideals in case

(1) $R = I_1 + I_2 + \cdots + I_n$, and
(2) for each $x \in R$, x has a unique representation as a sum $x = x_1 + x_2 + \cdots + x_n$ where $x_i \in I_i$.

We denote the internal direct sum by $R = I_1 \oplus I_2 \oplus \cdots \oplus I_n$. Notice that in this case the additive group $(R, +)$ is the internal direct product of the subgroups $\{(I_i, +) \mid 1 \leq i \leq n\}$ as defined in Definition 2.5.4. It is customary to say direct sum instead of direct product when the group is written additively.

DEFINITION 3.3.3. Let R be a ring. An *idempotent* of R is an element $e \in R$ that satisfies the equation $e^2 = e$. The elements 0 and 1 are called the trivial idempotents. A set $\{e_i \mid i \in I\}$ of idempotents in R is said to be *orthogonal* if $e_i e_j = 0$ for all $i \neq j$.

THEOREM 3.3.4. *If $A_1, \ldots, A_n$ are ideals in the ring R and $R = A_1 \oplus \cdots \oplus A_n$, then the following are true.*

(1) For each k, $A_k \cap \left(\sum_{j \neq k} A_j\right) = (0)$.
(2) If $x \in A_i$, $y \in A_j$ and $i \neq j$, then $xy = yx = 0$.
(3) For each i, A_i is a ring. If the identity element of A_i is denoted e_i, then $\{e_1, \ldots, e_n\}$ is a set of orthogonal idempotents in R. Moreover, each e_i is in the center of R and $A_i = Re_i$ is a principal ideal in R.
(4) R is isomorphic to the (external) direct product $A_1 \times \cdots \times A_n$.
(5) Suppose for each k that I_k is a left ideal in the ring A_k. Then $I = I_1 + I_2 + \cdots + I_n$ is a left ideal in R, where the sum is a direct sum.
(6) If I is a left ideal of R, then $I = I_1 \oplus I_2 \oplus \cdots \oplus I_n$ where each I_k is a left ideal in the ring A_k.

PROOF. (1): Assume $x \in A_k \cap \left(\sum_{j\neq k} A_j\right)$. Let $x_k = -x$. Since $x \in \sum_{j\neq k} A_j$, write $x = \sum_{j\neq k} x_j$ where each $x_j \in A_j$. Subtracting, $0 = x - x = x_1 + \cdots + x_k + \cdots + x_n$. By the uniqueness of the representation of 0 in the internal direct sum, it follows that $x = 0$.

(2): Notice that xy and yx are both in $A_i \cap A_j$ since the ideals are two-sided.

(3): Because A_i is an ideal, it is enough to show that A_i has a multiplicative identity. Write $1 = e_1 + e_2 + \cdots + e_n$. If $x \in A_i$, then multiply by x from the left and use Part (2) to get $x = x1 = \sum_{j=1}^{n} xe_j = xe_i$. Now multiply by x from the right and use Part (2) to get $x = 1x = \sum_{j=1}^{n} e_j x = e_i x$. This shows e_i is the multiplicative identity for A_i. Orthogonality of $\{e_1, \ldots, e_n\}$ is by Part (2). The rest is left to the reader.

(4): Define a function $f : A_1 \times A_2 \times \cdots \times A_n \to R$ from the external ring direct sum to R by the rule $(x_1, x_2, \ldots, x_n) \mapsto x_1 + x_2 + \cdots + x_n$. Then f is an isomorphism on additive groups since R is the internal direct sum of the ideals A_i. The reader should verify using Part (2) that f is multiplicative. By (3), $\phi(e_1, e_2, \ldots, e_n) = e_1 + e_2 + \cdots + e_n = 1$.

(5): Since each element r in $R = A_1 + A_2 + \cdots + A_n$ has a unique representation in the form $r = r_1 + r_2 + \cdots + r_n$, so does any element x in $I = I_1 + I_2 + \cdots + I_n$. So the sum is a direct sum and we can write $x = x_1 + x_2 + \cdots + x_n$ where each $x_k \in I_k$ is unique. Then $rx = r_1x_1 + r_2x_2 + \cdots + r_nx_n$ is in I, which shows I is a left ideal in R.

(6): By Part (3), for each k there is a central idempotent $e_k \in R$ such that $A_k = Re_k$. Let $I_k = e_kI$. Since e_k is central, $I_k = Ie_k$ is a left ideal in R. Since $I \subseteq R$ we have $I_k = Ie_k \subseteq Re_k = A_k$, so I_k is a left ideal in A_k. Since $1 = e_1 + \cdots + e_n$, we see that $I = I_1 + I_2 + \cdots + I_n$. The sum is a direct sum by Part (5). □

EXAMPLE 3.3.5. Let R be a ring and let $(R, +)$ denote the additive abelian group of R. By definition, a left ideal of R is a subgroup of $(R, +)$. In this example we show that subgroups of $(R, +)$ are not necessarily ideals of R.

(1) Let R be the ring $\mathbb{Z}/2 \times \mathbb{Z}/2$. Since $\mathbb{Z}/2$ is a field, it has only two ideals and both are generated by idempotents. By Theorem 3.3.4 (5), R has four ideals, namely the principal ideals generated by the four idempotents $(0,0)$, $(1,0)$, $(0,1)$, $(1,1)$.
(2) Now let G be the additive abelian group $(R, +)$ of the ring $R = \mathbb{Z}/2 \times \mathbb{Z}/2$. Then G has three elements of order 2, and each one generates a subgroup of order 2. There are five subgroups of G, namely $\langle(0,0)\rangle$, $\langle(1,0)\rangle$, $\langle(0,1)\rangle$, $\langle(1,1)\rangle$, G. Notice that the cyclic subgroup $\langle(1,1)\rangle$ of G is not an ideal in the ring R of Part (1).

PROPOSITION 3.3.6. *Suppose $A_1, \ldots, A_n$ are ideals in the ring R satisfying*

(1) $R = A_1 + A_2 + \cdots + A_n$ and
(2) for $k = 1, \ldots, n-1$, we have $A_k \cap (A_{k+1} + \cdots + A_n) = (0)$.

Then $R = A_1 \oplus A_2 \oplus \cdots \oplus A_n$.

PROOF. This follows from Part (4) implies Part (1) of Proposition 2.5.6. □

COROLLARY 3.3.7. *Let $R_1, \ldots, R_n$ be rings and $P = R_1 \times R_2 \times \cdots \times R_n$ the direct product. For $1 \leq k \leq n$, let e_k be the n-tuple in $R_1 \times R_2 \times \cdots \times R_n$ with 1 in coordinate k and 0 elsewhere.*

(1) $\{e_1, \ldots, e_n\}$ is a set of orthogonal idempotents in P, each e_k is in the center of P, $1 = e_1 + \cdots + e_n$, and Pe_k is a principal two-sided ideal in P.
(2) For each k, the kernel of the canonical projection map $\pi_k : P \to R_k$ is the principal ideal $P(1 - e_k)$. $P = Pe_k \times P(1 - e_k)$.
(3) The image of the canonical injection map $\iota_k : R_k \to P$ is the principal ideal Pe_k.

PROOF. (1): The proof of this part is left as an exercise for the reader.
(2): This follows from (1) and Exercise 3.3.15.
(3): This follows from (1). □

3.3. The Chinese Remainder Theorem for Rings. A generalization of Corollary 2.5.3 is proved for rings. By Lemma 1.2.5, two integers m and n are relatively prime if and only if the sum of the ideals $\mathbb{Z}m$ and $\mathbb{Z}n$ is the unit ideal $\mathbb{Z}$. Definition 3.3.8 generalizes this notion to two ideals in a ring R.

DEFINITION 3.3.8. If R is a ring and I and J are ideals in R, then we say I and J are *comaximal* if $I + J = R$.

Theorem 3.3.9 is a generalization of Corollary 2.5.3.

THEOREM 3.3.9. *(The Chinese Remainder Theorem) Let R be a ring and I, J comaximal ideals of R. Then there is an isomorphism of rings*

$$\frac{R}{I \cap J} \cong \frac{R}{I} \times \frac{R}{J}$$

induced by the natural projections $\eta_1 : R \to R/I$ and $\eta_2 : R \to R/J$. The isomorphism of rings induces an isomorphism

$$\text{Units}(R/(I \cap J)) \cong \text{Units}(R/I) \times \text{Units}(R/J)$$

of multiplicative groups of units.

PROOF. Step 1: Let $\phi : R \to R/I \times R/J$ be defined by $\phi(x) = (x + I, x + J)$. Since ϕ is defined in terms of the natural projections η_1, η_2, ϕ is a well defined homomorphism of rings.

Step 2: We prove that ϕ is onto. Let $a, b \in R$. We need to find $x \in R$ such that $\phi(x) = (a + I, b + J)$. Since I and J are comaximal, there exist $u \in I$, $v \in J$ such that $1 = u + v$. Then $u = 1 - v \equiv 1 \pmod{J}$ and $v = 1 - u \equiv 1 \pmod{I}$. Set $x = bu + av$. Then

$$\begin{aligned} x &\equiv bu + av \pmod{I} \\ &\equiv av \pmod{I} \\ &\equiv a \pmod{I}. \end{aligned}$$

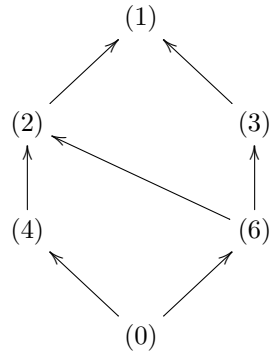
Likewise, $x \equiv b \pmod{J}$. Therefore, $\phi(x) = (a + I, b + J)$.

Step 3: Consider the kernel of ϕ, $\ker \phi = \{x \in R \mid x \in I \text{ and } x \in J\} = I \cap J$. By Theorem 3.2.15, this proves that the rings are isomorphic. By Exercise 3.2.39, ϕ induces an isomorphism on the groups of units of the rings. □

PROPOSITION 3.3.10. *Let R be a commutative ring. If I and J are comaximal ideals, then $IJ = I \cap J$.*

PROOF. If $x \in I$ and $y \in J$, then $xy \in I$ and $xy \in J$. Since IJ is generated by elements of the form xy, we have $IJ \subseteq I \cap J$. Let z be an arbitrary element of $I \cap J$. We show $z \in IJ$. Since $R = I + J$, there exist $u \in I$ and $v \in J$ such that $1 = u + v$. Now $zu \in IJ$ since $z \in J$ and $u \in I$. Also $zv \in IJ$ since $z \in I$ and $v \in J$. Then $z = zu + zv \in IJ$. □

EXAMPLE 3.3.11. Let $R = \mathbb{Z}$, $I = (3)$, $J = (4)$. Since $\gcd(3, 4) = 1$ we have $I + J = R$. By Proposition 3.3.10, $I \cap J = IJ = (3)(4) = (12)$. By Theorem 3.3.9, $\mathbb{Z}/(12) \cong \mathbb{Z}/(3) \times \mathbb{Z}/(4)$. By Theorem 3.3.4, R has a principal ideal Re_1 of order 3 and a principal ideal Re_2 of order 4, both idempotent generated. By Lemma 2.2.18, the subgroup of order 3 in $\mathbb{Z}/(12)$ is $(4) = \{0, 4, 8\}$. The subgroup of order 4 in $\mathbb{Z}/(12)$ is $(3) = \{0, 3, 6, 9\}$. The two idempotents in $\mathbb{Z}/(12)$ corresponding to the direct summands are 4 and 9 respectively. By Theorem 2.3.27 and Example 3.2.1 (6), for any $n \geq 1$, the the ideals of $\mathbb{Z}/(n)$ correspond to the divisors of n. Therefore, $\mathbb{Z}/(3)$ has 2 ideals, $\mathbb{Z}/(4)$ has 3 ideals, and $\mathbb{Z}/(12)$ has $6 = 2 \cdot 3$ ideals. The lattice of ideals in $\mathbb{Z}/(12)$ is



There are two maximal ideals in $\mathbb{Z}/(12)$.

COROLLARY 3.3.12. *Let R be a commutative ring. If I and J are comaximal ideals, then $R/IJ \cong R/I \times R/J$.*

COROLLARY 3.3.13. *Let R be any ring. If $I_1, \ldots, I_n$ are ideals in R and*

$$\phi : R \to R/I_1 \times R/I_2 \times \cdots \times R/I_n$$

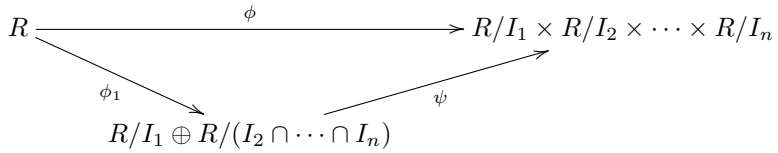
is the natural map given by $x \mapsto (x + I_1, \ldots, x + I_n)$, then the following are true.

(1) ϕ is a homomorphism of rings.
(2) The kernel of ϕ is equal to $I_1 \cap I_2 \cap \cdots \cap I_n$.
(3) ϕ is onto if and only if $n = 1$ or the ideals are pairwise comaximal, (that is, $I_i + I_j = R$ if $i \neq j$).

PROOF. We prove (3) and leave the rest to the reader. Assume ϕ is onto and $n > 1$. For each $1 \leq i \leq n$, consider the idempotent e_i in $R/I_1 \times \cdots \times R/I_n$ which is 1 in coordinate i and 0 in every other coordinate. Since ϕ is onto, there exists an element $a_i \in R$ such that $b_i = 1 - a_i \in I_i$ and $a_i \in I_j$ whenever $j \neq i$. Therefore, $1 = a_i + b_i$ is in $I_j + I_i$.

Now we prove the converse of (3). If $n = 1$, then this follows from Theorem 3.2.15. Assume $n > 1$ and the ideals are pairwise comaximal. If $n = 2$, this is Theorem 3.3.9. Inductively, assume $n > 2$ and that the result holds for a collection of $n - 1$ or fewer ideals. By our induction hypothesis, there is an isomorphism

of rings $\phi_2 : R/(I_2 \cap \cdots \cap I_n) \to R/I_2 \times \cdots \times R/I_n$. By Exercise 3.3.21, I_1 and $I_2 \cap \cdots \cap I_n$ are comaximal. Consider the diagram



where ϕ and ϕ_1 are homomorphisms of rings by (1). By induction, ϕ_1 is onto. The map ψ is defined by $(x, y) \mapsto (x, \phi_2(y))$ and it is easy to see that ψ is an isomorphism. The kernel of ϕ is equal to the kernel of ϕ_1, by (2). All of the maps are the natural maps, so the diagram commutes. Therefore, ϕ is onto. □

EXAMPLE 3.3.14. Let F be a field and

$$R = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \mid a, b, d \in F \right\}$$

the set of all upper triangular matrices in $M_2(F)$. As in Example 3.1.12, R is a noncommutative subring of $M_2(F)$. The proof given in Example 3.1.13 can be used to show that the center of R is the set of scalar matrices, which is isomorphic to F by the homomorphism $\delta : F \to R$ defined by $\delta(a) = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$. Define $\lambda : R \to F$ by $\lambda \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = a$. The reader should verify that λ is a homomorphism and $\lambda\delta(a) = a$ for all $a \in F$. We say F is a subfield of R and λ is a *section to* δ. The homomorphism $\rho : R \to F$ defined by $\rho \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = d$ also satisfies $\rho\delta(a) = a$, hence a section to δ is not unique. The kernels of λ and ρ are

$$\ker \lambda = \left\{ \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} \mid b, d \in F \right\}, \quad \ker \rho = \left\{ \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \mid a, b \in F \right\},$$

which are proper ideals in R. We say R is not a simple ring. Since F has no proper ideals, by Corollary 3.2.18, there is no proper ideal of R that contains $\ker \lambda$ or $\ker \rho$. The ideals $\ker \lambda$ and $\ker \rho$ are maximal proper ideals in R. Let $D = \left\{ \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} \mid a, d \in F \right\}$. The reader should verify that D is a subring of R. Define $\tau : R \to D$ by $\tau \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$. The reader should verify that τ is a homomorphism and for any matrix $A \in D$, $\tau(A) = A$. In other words, τ is a section to the inclusion map $D \to R$. The kernel of τ is the ideal

$$\ker \tau = \left\{ \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} \mid b \in F \right\}.$$

If $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ is an idempotent matrix, then a and d are idempotents in F. After looking at the possible cases, the reader should verify that the set of all idempotents in R is

$$\left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\}.$$

Only the two trivial idempotents, namely 0 and 1, are central. Therefore, R is not an internal direct sum of proper ideals. Let R^* be the group of units of R. By

Exercise 3.2.39, there are homomorphisms of groups $\delta^* : F^* \to R^*$ and $\rho^* : R^* \to F^*$. Let

$$T = \ker \rho^* = \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \mid a \in F^*, b \in F \right\},$$

and

$$Z = \delta(F^*) = \left\{ \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \mid a \in F^* \right\}.$$

By Exercise 2.5.24, the group of units of R is the internal direct product $R^* = T \times Z$ of the two proper normal subgroups T and Z. The ring R is an example of an *extension of a ring by a module*. Specifically, R is the extension of D by the module $\ker \tau$. The interested reader is referred to [**10**, Exercise 8.1.14] for the general construction.

3.4. Exercises.

Exercise 3.3.15. Suppose R is a ring and $e \in R$ is a central idempotent. Assume $e \neq 0$ and $e \neq 1$. Let I be the ideal generated by e. Prove that R is equal to the internal direct sum $I \oplus J$ for some ideal J.

Exercise 3.3.16. Consider the ring $R = \mathbb{Z}/(n)$.

(1) Suppose $n = 1105$.
   (a) Prove that R is isomorphic to a direct sum of fields.
   (b) Determine all maximal ideals in R.
   (c) Determine all idempotents in R.
(2) Suppose $n = 1800$.
   (a) Determine all maximal ideals in R.
   (b) Determine all idempotents in R.

Exercise 3.3.17. Assume the ring R is a direct product $R = R_1 \times R_2$. Let e_1, e_2 be the central idempotents corresponding to the factors (guaranteed by Corollary 3.3.7). Let D be a ring which has exactly two idempotents, namely 0 and 1. Let $\theta : R \to D$ be a homomorphism of rings. Prove that exactly one of the following is true:

(1) $\theta(e_1) = 1$ and $\theta(e_2) = 0$, or
(2) $\theta(e_1) = 0$ and $\theta(e_2) = 1$.

Exercise 3.3.18. Let R be any ring. Let I and J be ideals in R and $\phi : R \to R/I \oplus R/J$ the natural homomorphism of Theorem 3.3.9. Show that the image of ϕ is the subring of $R/I \oplus R/J$ defined by $\{(x + I, y + J) \mid x - y \in I + J\}$. See [**9**, Exercise 4.2.27] for an interpretation of this result in terms of modules.

Exercise 3.3.19. If $n > 1$, then we say n is square free if n is not divisible by the square of a prime number. Prove that the nil radical of $\mathbb{Z}/n$ is (0) if and only if n is square free. For the definition of nil radical, see Exercise 3.2.38.

Exercise 3.3.20. Let $n > 1$ and R a finite ring of order n. Suppose n is square free and the factorization of n into primes is $n = p_1 \cdots p_m$. Prove the following:

(1) $R \cong \mathbb{Z}/n$.
(2) R is commutative.
(3) R is a field, or a direct sum of fields.
(4) In terms of the prime factors of n, describe the maximal ideals of R.

EXERCISE 3.3.21. Let R be a ring and I, $I_1, \ldots, I_n$ a collection of two-sided ideals in R. Assume $I + I_j = R$ for $1 \leq j \leq n$. Show that $I + I_1 \cap \cdots \cap I_n = R$.

EXERCISE 3.3.22. Show that no two of the following rings are isomorphic to each other: (1) $\mathbb{Z}/16$. (2) $\mathbb{Z}/4 \times \mathbb{Z}/4$. (3) $\mathbb{Z}/2 \times \mathbb{Z}/8$. (4) $\mathbb{Z}/2 \times \mathbb{Z}/2 \times \mathbb{Z}/2$. (5) $\mathbb{Z}/4[i]$ which is the ring of Exercise 3.1.24. (6) $M_2(\mathbb{Z}/2)$ which is the ring of Exercise 3.1.21. (7) The group ring $R(G)$ where $R = \mathbb{Z}/2$, and $G = \mathbb{Z}/2 \times \mathbb{Z}/2$, of Exercise 3.1.19.

EXERCISE 3.3.23. Let R be a ring and e an idempotent in R. Prove that if $f : R \to S$ is a homomorphism of rings, then $f(e)$ is an idempotent in S.

EXERCISE 3.3.24. Let $\{R_i \mid i \in I\}$ be a family of rings indexed by the set I. Show that the group of units of the direct product $\prod_{i \in I} R_i$ is equal to the direct product of the family of groups $\{\text{Units}(R_i) \mid i \in I\}$. That is, show that $\text{Units}(\prod_{i \in I} R_i) = \prod_{i \in I} \text{Units}(R_i)$.

EXERCISE 3.3.25. Let $R_1, \ldots, R_n$ be commutative rings and $R = R_1 \times \cdots \times R_n$ the direct product. Show that R is a principal ideal ring if and only if R_i is a principal ideal ring for each i.

EXERCISE 3.3.26. Show that the following three rings are all isomorphic to each other: (1) The group ring $\mathbb{Z}(G)$, where $G = \langle \sigma \mid \sigma^2 = e \rangle$ is a cyclic group of order two. (2) $\left\{ \begin{bmatrix} a & b \\ b & a \end{bmatrix} \mid a, b \in \mathbb{Z} \right\}$, which is a subring of $M_2(\mathbb{Z})$. (3) $\{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid x \equiv y \pmod 2\}$, which is a subring of $\mathbb{Z} \times \mathbb{Z}$.

4. Factorization in Commutative Rings

Throughout this section every ring will be commutative unless specifically stated otherwise. Notions that arise in Number Theory are applied to the setting of rings. On the set of natural numbers the relation called "divides" is reflexive, antisymmetric and transitive, hence it is a partial order. On the ring of integers, it is not symmetric and not antisymmetric. If a and b are elements of a commutative ring R, then either a divides b or not. Thus "divides" is a binary relation on R that is reflexive and transitive but in general not symmetric or antisymmetric. Elements that are divisors of each other are called associates. The relation "is an associate" defines an equivalence relation. On associate classes, the relation "divides" does define a partial order. Using this partial order on associate classes we define the greatest common divisor of a subset X of R. For a general commutative ring R, greatest common divisors do not necessarily exist. In R we define what it means for an element to be prime, or irreducible. For the ring of integers these notions are equivalent, but for a general commutative ring R they are not (see Exercise 3.6.21). A unique factorization domain is an integral domain for which the conclusion of the Fundamental Theorem of Arithmetic holds. A euclidean domain is an integral domain which has a division algorithm. Some of the fundamental properties of a principal ideal domain are proved. For instance, a principal ideal domain is a unique factorization domain and satisfies the ascending chain condition on ideals.

DEFINITION 3.4.1. Let R be a commutative ring. Suppose a and b are elements of R. We say a *divides* b, and write $a \mid b$, in case there exists $c \in R$ such that $b = ac$. We also say that a is a *factor* of b, or b is a *multiple* of a.

DEFINITION 3.4.2. Let R be a commutative ring and suppose a and b are elements of R. If $a \mid b$ and $b \mid a$, then we say a and b are *associates*. In this case we write $a \sim b$. The reader should verify that the relation "a is an associate of b" is an equivalence relation on R.

LEMMA 3.4.3. *Let R be a commutative ring. Let $a, b, r, u \in R$.*

(1) The following are equivalent:
  (a) $a \mid b$.
  (b) $b \in Ra = (a)$.
  (c) $(a) \supseteq (b)$.
(2) a and b are associates if and only if $(a) = (b)$.
(3) If $a = bu$ and u is a unit, then a and b are associates.
(4) If R is an integral domain and a and b are associates, then $a = bu$ for some unit u.
(5) Let R be an integral domain. If $a \neq 0$ and $a \mid b$, then there exists a unique c such that $b = ac$. We write $c = ba^{-1}$, or $c = b/a$.

PROOF. (1): This follows straight from Definitions 3.2.4 and 3.4.1.

(5): Suppose $b = ac = ac'$. Subtract and distribute to get $a(c - c') = 0$. Since $a \neq 0$ and R is an integral domain, this means $c - c' = 0$, hence $c = c'$.

The rest of the proof is left to the reader. □

DEFINITION 3.4.4. Let R be a commutative ring and a an element of R which is not a unit and not a zero divisor. Then a is *irreducible* in case whenever $a = bc$, either b is a unit or c is a unit. We say that a is *prime* in case whenever $a \mid bc$, either $a \mid b$ or $a \mid c$.

In Lemma 3.4.5 below we show that every prime element in an integral domain is irreducible. For a general integral domain R an irreducible element is not prime (see Exercise 3.6.21). In Corollary 3.4.14 below we prove that in a unique factorization domain an element is prime if and only if it is irreducible.

LEMMA 3.4.5. *Let R be an integral domain and p an element of R.*

(1) p is prime if and only if (p) is a prime ideal.
(2) p is irreducible if and only if the principal ideal (p) is maximal among nonunit principal ideals of R.
(3) If p is prime, then p is irreducible.
(4) If p is irreducible and q is an associate of p, then q is irreducible.
(5) If p is prime and q is an associate of p, then q is prime.
(6) If p is irreducible, then the only divisors of p are units and associates of p.

PROOF. In the following, let $a, b, p, q, u \in R$.

(1): We have $ab \in (p)$ if and only if $p \mid ab$. Likewise, $a \in (p)$ if and only if $p \mid a$, and $b \in (p)$ if and only if $p \mid b$.

(2): Suppose p is irreducible and $(p) \subseteq (x)$. Then $x \mid p$. Since p is irreducible, either x is a unit or x and p are associates. By Lemma 3.2.11 and Lemma 3.4.3 (2), this implies either $(p) = (x)$ or $(x) = R$. Conversely, suppose (p) is maximal among all nonunit principal ideals. Suppose $p = xy$. Then $p \in (x)$ and $p \in (y)$. Since p is not a unit, either x or y is not a unit. Assume x is not a unit. Since (p) is maximal, this implies $(p) = (x)$. By Lemma 3.4.3 Parts (2), (4) and (5), this implies p and x are associates and y is a unit.

(3): Suppose p is prime and $p = ab$. Since p is prime we assume $p \mid a$. Therefore a and p are associates. By Lemma 3.4.3 (4), b is a unit in R.

(5): Assume p is prime, u is a unit, $q = pu$, and $q \mid ab$. For some $c \in R$, $ab = qc = puc$. Since p is prime we assume $p \mid a$. For some $d \in R$, $a = pd = (pu)(u^{-1}d)$, which shows $q \mid a$.

(4) and (6): The proofs are left to the reader. □

4.1. Greatest Common Divisors. Let R be a commutative ring. Two elements a and b are associates if $a \mid b$ and $b \mid a$, and we write $a \sim b$. Then "is an associate" defines an equivalence relation on R. On the set of equivalence classes the relation "divides" is a partial order. It is with respect to this partial order that we define the greatest common divisor of a set of elements in R.

Definition 3.4.6. Let R be a commutative ring and X a nonempty subset of R. An element $d \in R$ is a *greatest common divisor of* X if the following are satisfied:

(1) $d \mid x$ for all $x \in X$, and
(2) if $c \mid x$ for all $x \in X$, then $c \mid d$.

We sometimes write $d = \gcd(X)$ if d is a greatest common divisor of X. When $X = \{x_1, \ldots, x_n\}$ is finite, we write $d = \gcd(x_1, \ldots, x_n)$ for $\gcd(X)$.

In Lemma 3.4.7 we see that if d is a greatest common divisor, so is any associate of d. If $\gcd(X)$ exists, it is unique up to associates.

Lemma 3.4.7. *Let X be a nonempty subset of R.*

(1) If the greatest common divisor of X exists, then it is unique up to associates. In other words, the following are equivalent:
  (a) $d = \gcd(X)$ and $d \sim d'$.
  (b) $d = \gcd(X)$ and $d' = \gcd(X)$.
(2) Let R be an integral domain. Then d and d' are two greatest common divisors of X if and only if there exists a unit $u \in R^*$ such that $d' = du$.

Proof. (1): Suppose $d = \gcd(X)$ and $d \sim d'$. For each $x \in X$, we have $d' \mid d \mid x$. If $c \mid x$ for each $x \in X$, then $c \mid d \mid d'$. Therefore, $d' = \gcd(X)$. Conversely, if $d = \gcd(X)$ and $d' = \gcd(X)$, then $d \mid d'$ and $d' \mid d$. Thus d and d' are associates.

(2): By Lemma 3.4.3 (4), $d' = du$ for some $u \in R^*$. □

Proposition 3.4.8. *Let R be a commutative ring and X a nonempty subset of R.*

(1) If the ideal generated by X is principal and d is a generator for (X), then $d = \gcd(X)$.
(2) If $d = \gcd(X)$ exists and d is in the ideal (X), then $(d) = (X)$.

Proof. (1): If $(d) = (X)$, then $d \mid x$, for all $x \in X$. Also, $d = a_1x_1 + \cdots + a_nx_n$ for some $a_1, \ldots, a_n \in R$ and $x_1, \ldots, x_n \in X$. Suppose $c \mid x$ for each $x \in X$. Then $c \mid a_1x_1 + \cdots + a_nx_n = d$.

(2): This follows from Definition 3.4.6 and Exercise 3.4.26. □

Corollary 3.4.9. *(A PID is a Bézout domain) Let R be a principal ideal domain and X a nonempty subset of R. Then $d = \gcd(X)$, the greatest common divisor of X, exists and is unique up to associates. Any generator d of the ideal*

(X) is a greatest common divisor of a and b. In this case, $d = a_1x_1 + \cdots + a_nx_n$ for some $a_1, \ldots, a_n \in R$ and $x_1, \ldots, x_n \in X$.

PROOF. Since (X) is principal, there exists $d \in R$ such that $(d) = (X)$. Proposition 3.4.8 (1) implies $d = \gcd(X)$ exists and can be written in the form $d = a_1x_1 + \cdots + a_nx_n$ for some $a_1, \ldots, a_n \in R$ and $x_1, \ldots, x_n \in X$. By Lemma 3.4.7, d is unique up to associates. □

COROLLARY 3.4.10. *Let R be a principal ideal domain and $p \in R$ an irreducible element. Then the following are true.*

(1) p is prime. That is, if $p \mid ab$, then $p \mid a$ or $p \mid b$.
(2) If $x_1, x_2, \ldots, x_n$ in R and $p \mid x_1x_2 \cdots x_n$, then $p \mid x_i$ for some i.

PROOF. (1): Assume $p \mid ab$ and p does not divide b. We prove $p \mid a$. The ideal (p, b) is principal, hence is equal to (d), for some $d \in R$. Then $d \mid p$ and $d \mid b$. Since p is irreducible, d is a unit, or d is an associate of p (Lemma 3.4.5 (6)). We are assuming p does not divide b, hence d is not an associate of p, hence d is a unit. Therefore $(d) = (1)$. By Corollary 3.4.9, we can write $1 = px + by$. Multiply by a to get $a = pax + aby$. Since $p \mid ab$, this shows $p \mid a$.

(2) If $n = 1$, then take $i = 1$ and stop. Assume inductively that $n > 1$ and the result holds for a product of $n - 1$ factors. Then $p \mid (x_1 \cdots x_{n-1})x_n$. By Part (1), $p \mid x_n$, or $p \mid (x_1 \cdots x_{n-1})$. By the induction hypothesis, $p \mid x_i$ for some i. □

DEFINITION 3.4.11. Let R be an integral domain. Then R is a *unique factorization domain* if for every nonzero nonunit x in R, the following are satisfied:

(1) x has a representation as a product of irreducibles. That is, there exist irreducible elements $x_1, x_2, \ldots, x_n$ in R such that $x = x_1x_2 \cdots x_n$.
(2) In any factorization of x as in (1), the number of factors is unique.
(3) In any factorization of x as in (1), the irreducible factors are unique up to order and associates.

Sometimes we say R is a UFD.

EXAMPLE 3.4.12. The ring $\mathbb{Z}$ is a unique factorization domain, by the Fundamental Theorem of Arithmetic. We will prove in Theorem 3.4.15 that any principal ideal domain is a unique factorization domain.

COROLLARY 3.4.13. *Let R be a unique factorization domain. If $X = \{r_1, \ldots, r_n\}$ is a finite nonempty subset of R, then $d = \gcd(X)$ exists and is unique up to associates.*

PROOF. If $n = 1$, then by Proposition 3.4.8 (1), $r_1 = \gcd(X)$ exists. By Mathematical Induction and Exercise 3.4.27, it suffices to prove the $n = 2$ case. Assume $X = \{a, b\}$. If $a = 0$, then $(a, b) = (b)$ and by Proposition 3.4.8 (1), $b = \gcd(a, b)$ exists. If $(a, b) = (1)$, then by Proposition 3.4.8 (1), $1 = \gcd(a, b)$ exists. Assume a and b are both nonzero and nonunits. Then by Exercise 3.4.28, $\gcd(a, b)$ exists and we are done. □

COROLLARY 3.4.14. *Let R be a unique factorization domain and $p \in R - (0)$. Then the following are equivalent.*

(1) p is irreducible.
(2) p is prime.
(3) The principal ideal (p) is a prime ideal.

PROOF. By Lemma 3.4.5 (1), (2) is equivalent to (3). By Lemma 3.4.5 (3), (2) implies (1). We prove that (1) implies (2). Suppose p is irreducible and $p \mid ab$. If $a = 0$, then $p \mid a$. If $b = 0$, then $p \mid b$. Since p is not invertible, ab is not invertible. Write $ab = pc$ for some $c \in R$. Assume ab is nonzero and not invertible. Factor ab and pc into irreducibles. By uniqueness of factorization, p is an associate of one of the irreducible factors of a or b. □

4.2. Principal Ideal Domains. The fundamental properties of a principal ideal domain (a PID, for short) are derived in Theorem 3.4.15. In particular, every principal ideal domain is a unique factorization domain. Part (2) shows that a PID satisfies the ascending chain condition on ideals. A commutative ring with this property is said to be *noetherian*, after E. Noether.

THEOREM 3.4.15. *Let R be a principal ideal domain.*

(1) If p is an irreducible element, then p is a prime element.

(2) R satisfies the ascending chain condition *on ideals. That is, given a chain of ideals $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots \subseteq I_n \subseteq \cdots$, there exists $N \geq 1$ such that $I_N = I_{N+1} = \cdots$.*

(3) If $a \in R$ is a nonunit, nonzero element of R, then the set

$$\mathcal{S} = \{p \in R \mid p \text{ is irreducible and } p \mid a\}$$

contains only a finite number of associate classes. In other words, up to associates, a has only a finite number of irreducible factors.

(4) If I is an ideal in R which is not the unit ideal, then $\bigcap_{n\geq 1} I^n = (0)$.

(5) Suppose a is a nonzero element in R, p is irreducible and p is a factor of a. Then for some $n \geq 1$ we have $a \in (p^n)$ and $a \notin (p^{n+1})$.

(6) If $a \in R$ is a nonunit and a nonzero element, then there exists an irreducible element p such that $p \mid a$.

(7) R is a unique factorization domain.

PROOF. (1): This is Corollary 3.4.10.

(2): Let $I = \bigcup_{k=1}^{\infty} I_k$. By Exercise 3.2.35, I is an ideal in R. Since R is a PID, there exists $a \in R$ such that $I = (a)$. Given $a \in I$, we know $a \in I_N$ for some N. Then $I = (a) \subseteq I_N \subseteq I_{N+1} \subseteq \cdots$ and we are done.

(3): The proof is by contradiction. Assume $\{p_1, p_2, \dots\}$ is a sequence in $\mathcal{S}$ such that for each $n > 1$, p_n does not divide $p_1 p_2 \cdots p_{n-1}$. Write $a = p_1 a_1$. Then $p_2 \mid p_1 a_1$. By assumption, p_2 does not divide p_1. By Part (1), $p_2 \mid a_1$ and we write $a_1 = p_2 a_2$. Iteratively we arrive at the factorizations

$$a = p_1 a_1 = p_1 p_2 a_2 = \cdots = p_1 p_2 \cdots p_n a_n.$$

Applying one more step, we know $p_{n+1} \mid a$. Since p_{n+1} does not divide $p_1 p_2 \cdots p_n$, and p_{n+1} is prime, it follows that $p_{n+1} \mid a_n$. Write $a_n = p_{n+1} a_{n+1}$. Therefore $(a_n) \subseteq (a_{n+1})$ with equality if and only if a_n and a_{n+1} are associates. But p_{n+1} is not a unit, so by Lemma 3.4.3 (4), the chain of ideals

$$(a_1) \subseteq (a_2) \subseteq \cdots \subseteq (a_n) \subseteq (a_{n+1}) \subseteq \cdots$$

is strictly increasing. This contradicts Part (2).

(4): Because R is a PID, $I = (b)$ for some $b \in R$. If $I = 0$, then Part (4) is trivial, so we assume $b \neq 0$. Let $M = \bigcap_{n=1}^{\infty} I^n$. Then M is an ideal in R, so $M = (r)$ for some $r \in R$. Since M is an ideal, $bM \subseteq M$. To show that $bM = M$, assume $x \in M$. Then $x \in M \subseteq I$ implies $x = by$ for some $y \in R$. Let $n \geq 1$.

Then $x \in M \subseteq I^{n+1} = (b^{n+1})$ implies $x = b^{n+1}z$ for some $z \in R$. Since R is an integral domain and $b \neq 0$, $x = by = b^{n+1}z$ implies $y = b^n z \in I^n = (b^n)$. This proves $y \in \bigcap_{n\geq 1} I^n = M$. Therefore $x \in bM$, and $bM = M$. Since $bM = (br)$, Lemma 3.4.3 says br and r are associates. But b is not a unit, so $r = 0$, which proves (4).

(5): Set $I = (p)$. By assumption, $a \in (p)$ and $a \neq 0$. By Part (4), for some $n \geq 1$ we have $a \notin (p^{n+1})$ and $a \in (p^n)$.

(6): The proof is by contradiction. Suppose $a \in R$ is not a unit, and not divisible by an irreducible. Then a is not irreducible. There are nonunits a_1, b_1 in R such that $a = a_1 b_1$. By our assumption, a_1 and b_1 are not irreducible. By Lemma 3.4.3, $(a) \subsetneq (a_1)$. Since a_1 is not irreducible, there are nonunits a_2, b_2 in R such that $a_1 = a_2 b_2$. Since a_2 and b_2 are divisors of a, both are not irreducible. By Lemma 3.4.3, $(a) \subsetneq (a_1) \subsetneq (a_2)$. Recursively we construct a strictly increasing sequence of ideals $(a_i) \subsetneq (a_{i+1})$, contradicting Part (2).

(7): This proof is left to the reader. □

4.3. Euclidean Domains. In this section we define a family of rings called euclidean domains. The prototype of this family is the ring of integers, $\mathbb{Z}$. The ring of integers has a norm function, the absolute value function, and the Division Algorithm (Proposition 1.2.3). These are the properties of the ring $\mathbb{Z}$ that we investigate. A euclidean domain is an integral domain that has a norm function and a division algorithm.

Definition 3.4.16. Let R be an integral domain. Then R is called a *euclidean domain* if there is a function (called the *norm*) $\delta : R - (0) \to \mathbb{N}$ such that

(1) $\delta(ab) = \delta(a)\delta(b)$ for all $a, b \in R - (0)$, and
(2) for all $a, b \in R - (0)$ there exist $q, r \in R$ such that $a = bq + r$ and either $r = 0$ or $\delta(r) < \delta(b)$.

Example 3.4.17. The ring of integers $\mathbb{Z}$ is a euclidean domain with the norm function $\delta(x) = |x|$. The absolute value function is multiplicative, and property (2) is satisfied by the Division Algorithm on $\mathbb{Z}$ (Proposition 1.2.3).

Example 3.4.18. We will prove in Corollary 3.6.5 below that if F is a field, then the polynomial ring $F[x]$ is a euclidean domain.

Example 3.4.19. In this example we prove that the ring of gaussian integers $\mathbb{Z}[i]$ (see Exercise 3.1.23) is a euclidean domain. Let $\chi : \mathbb{C} \to \mathbb{C}$ be complex conjugation: $\chi(a + bi) = a - bi$. The norm function $\delta : \mathbb{C} - (0) \to \mathbb{R}$ is defined by $\delta(a + bi) = a^2 + b^2 = (a + bi)\chi(a + bi)$. Since $\delta = 1_{\mathbb{C}}\chi$ is defined by multiplying two automorphisms, δ is multiplicative. As in Exercise 3.1.23, we have $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$. Now we prove that Property (2) of Definition 3.4.16 holds. Let $\alpha, \beta \in \mathbb{Z}[i] - (0)$. Since $\mathbb{Q}[i]$ is a field, we can write $\alpha\beta^{-1} = u + vi$ where $u, v \in \mathbb{Q}$. Let $p, q \in \mathbb{Z}$ such that $|u - p| \leq 1/2$ and $|v - q| \leq 1/2$. Then $\gamma = p + qi \in \mathbb{Z}[i]$.

Define $\rho = \alpha - \beta\gamma = \beta((u-p)+(v-q)i)$. Then

$$\begin{aligned}\delta(\rho) &= \delta\left(\beta((u-p)+(v-q)i)\right)\\ &= \delta(\beta)\left((u-p)^2+(v-q)^2\right)\\ &\leq \delta(\beta)\left(\frac{1}{2^2}+\frac{1}{2^2}\right)\\ &\leq \frac{1}{2}\delta(\beta) < \delta(\beta)\end{aligned}$$

and $\alpha = \beta\gamma + \rho$.

PROPOSITION 3.4.20. *If R is a euclidean domain, then R is a principal ideal domain. Hence R is a unique factorization domain.*

PROOF. Let I be a nonzero ideal in R. Consider the nonempty set $S = \{\delta(a) \mid a \in I - (0)\}$. By the Well Ordering Principle for $\mathbb{N}$, S has a least element, say $\delta(b)$, for some $b \in I$. Let $a \in I$. Since R is a euclidean domain, there exist q and r in R such that $a = bq + r$ and either $r = 0$ or $\delta(r) < \delta(b)$. Since $a, b \in I$, we have $r \in I$. By the minimal choice of $\delta(b)$, we conclude that $r = 0$. Thus, $a \in Rb$. This shows $I = Rb$ is principal. By Theorem 3.4.15, R is a UFD. □

EXAMPLE 3.4.21. By Proposition 3.4.20, we have the following examples of principal ideal domains.

(1) We have not proved it yet, but if F is a field, then $F[x]$ is a principal ideal domain.
(2) By Example 3.4.19, the ring of gaussian integers $\mathbb{Z}[i]$ is a principal ideal domain.

PROPOSITION 3.4.22. *Let R be a euclidean domain with norm function $\delta : R - (0) \to \mathbb{N}$. Then the following are true:*

(1) $\delta(1) = 1$.
(2) If $u \in R^*$ is a unit in R, then $\delta(u) = 1$.
(3) If $\delta(u) = 1$, then $u \in R^*$ is a unit in R.
(4) The group of units of R is $R^* = \delta^{-1}\{1\}$.
(5) If p is a prime number, $x \in R - (0)$, and $\delta(x) = p$, then x is irreducible.

PROOF. (1) and (2): For any $u \in R - (0)$ we have $\delta(u) = \delta(u \cdot 1) = \delta(u)\delta(1)$. Therefore, $\delta(1) = 1$. Let $u \in R^*$. Then $1 = \delta(uu^{-1}) = \delta(u)\delta(u^{-1})$. Since the group of invertible elements of the ring $\mathbb{Z}$ is $\{1, -1\}$, we conclude that $\delta(u) = 1$.

(3): Assume $\delta(u) = 1$. Divide u into 1. There exist $q, r \in R$ such that $1 = uq + r$. Since 1 is the least element of $\mathbb{N}$, we conclude that $r = 0$. Thus, u is invertible.

(4): This part follows from (1), (2), and (3).

(5): Assume $x = ab$. Then $p = \delta(x) = \delta(a)\delta(b)$. Thus $\delta(a) = 1$ or $\delta(b) = 1$. By (4), $R^* = \{u \in R - (0) \mid \delta(u) = 1\}$. Hence a is a unit or b is a unit. □

We end this section with a proof that in a euclidean domain R the greatest common divisor of two elements a and b can be computed by the Euclidean Algorithm. In Corollary 3.4.24 we prove that the so-called Extended Euclidean Algorithm converges to a minimal solution (x, y) to the Bézout Identity $\gcd(a, b) = ax + by$.

PROPOSITION 3.4.23. *(The Euclidean Algorithm) Let R be a euclidean domain with with norm $\delta : R - (0) \to \mathbb{N}$. Let a and b be elements of R. The greatest common divisor of a and b exists and satisfies the following recursive formula:*

(1) (Basis) If $b = 0$, *then* $\gcd(a, b) = a$.
(2) (Recurrence) If $b \neq 0$, *then* $\gcd(a, b) = \gcd(b, r)$, *where* $a = bq + r$ *and either* $r = 0$ *or* $\delta(r) < \delta(b)$.

PROOF. If $b = 0$, then the ideals (a, b) and (a) are equal in R, and Corollary 3.4.9 implies $\gcd(a, b) = a$. If $b \neq 0$, then by Definition 3.4.16, $a = bq + r$, for elements q and r in R such that either $r = 0$ or $\delta(r) < \delta(b)$. Then the ideals (a, b) and (b, r) are equal in R. By Corollary 3.4.9, $\gcd(a, b) = \gcd(b, r)$. To see that the recursive algorithm converges, set $r_0 = b$ and successively apply Definition 3.4.16 to find a sequence of quotients $q_1, q_2, \ldots, q_{n+1}$ and a sequence of remainders $r_0, r_1, r_2, \ldots, r_n$ satisfying:

$$\begin{aligned}
a &= r_0 q_1 + r_1, \quad 0 < \delta(r_1) < \delta(r_0) \\
r_0 &= r_1 q_2 + r_2, \quad 0 < \delta(r_2) < \delta(r_1) \\
r_1 &= r_2 q_3 + r_3, \quad 0 < \delta(r_3) < \delta(r_2) \\
&\vdots \\
r_{n-3} &= r_{n-2} q_{n-1} + r_{n-1}, \quad 0 < \delta(r_{n-1}) < \delta(r_{n-2}) \\
r_{n-2} &= r_{n-1} q_n + r_n, \quad 0 < \delta(r_n) < \delta(r_{n-1}) \\
r_{n-1} &= r_n q_{n+1} + 0
\end{aligned}$$

where r_n is the last nonzero remainder. The algorithm converges for some n such that $0 \leq n \leq \delta(b)$ because $\delta(r_0) > \delta(r_1) > \delta(r_2) > \cdots > \delta(r_n) > 0$. As mentioned above,

$$\begin{aligned}
r_n = \gcd(r_n, r_{n-1}) = \gcd(r_n, r_{n-1}) = \gcd(r_{n-1}, r_{n-2}) \\
= \cdots = \gcd(r_3, r_2) = \gcd(r_2, r_1) = \gcd(r_1, r_0) = \gcd(a, b).
\end{aligned}$$

□

Let R be a euclidean domain, a, b elements of R, and $d = \gcd(a, b)$. In Corollary 3.4.24 below we prove the so-called Extended Euclidean Algorithm which is an efficient algorithm for computing a solution (x, y) to the equation $d = ax + by$. The proof we give is inspired by [**31**]. Using the method of Exercise 1.2.17, one can show that if (x_0, y_0) is a particular solution to the equation $d = ax + by$, then the general solution is given by $(x, y) = (x_0, y_0) + t(-b/d, a/d)$, where $t \in R$.

Let $q_1, q_2, \ldots, q_{n+1}$ and $r_0, r_1, r_2, \ldots, r_n, r_{n+1}$ be the sequences of quotients and remainders defined in the proof of Proposition 3.4.23. Then $r_n = \gcd(a, b)$ and $r_{n+1} = 0$. The Extended Euclidean Algorithm uses two additional sequences which are defined here since they will appear in Corollary 3.4.24 below. The sequence $s_0, s_1, \ldots, s_{n+1}$ is defined recursively:

$$s_i = \begin{cases} 0 & \text{if } i = 0 \\ 1 & \text{if } i = 1 \\ s_{i-2} - q_i s_{i-1} & \text{if } 2 \leq i \leq n+1. \end{cases} \tag{4.1}$$

Likewise, the sequence $t_0, t_1, \ldots, t_{n+1}$ is defined by the recursive formula:

$$t_i = \begin{cases} 1 & \text{if } i = 0 \\ -q_1 & \text{if } i = 1 \\ t_{i-2} - q_i t_{i-1} & \text{if } 2 \leq i \leq n+1. \end{cases} \tag{4.2}$$

COROLLARY 3.4.24. *(Extended Euclidean Algorithm) Let R be a euclidean domain with norm function $\delta : R - (0) \to \mathbb{N}$. Let a and b be elements of R. Let $q_1, \ldots, q_{n+1}$, $r_0, r_1, \ldots, r_{n+1}$ be the two sequences defined in the proof of Proposition 3.4.23. Let $s_0, s_1, \ldots, s_{n+1}$ and $t_0, t_1, \ldots, t_{n+1}$ be the sequences defined in Eq.* (4.1) *and Eq.* (4.2) *respectively. Then the following hold.*

(1) For $0 \leq i \leq n+1$, $r_i = as_i + bt_i$.
(2) $\gcd(a, b) = as_n + bt_n$.
(3) $\gcd(s_i, t_i) = 1$ for $i = 0, 1, \ldots, n+1$.
(4) Let $d = \gcd(a, b)$. Then b and ds_{n+1} are associates, and a and dt_{n+1} are associates.
(5) If $R = \mathbb{Z}$ is the ring of integers, $1 < a$, $1 < b$, a does not divide b, and b does not divide a, then $|s_n| \leq b/(2d)$ and $|t_n| \leq a/(2d)$. In other words, the Extended Euclidean Algorithm applied to a and b converges to a minimal solution (x, y) to the Bézout equation $d = ax + by$ in the sense that $|x| \leq b/(2d)$ and $|y| \leq a/(2d)$.

PROOF. (1): Since $r_0 = a0 + b1$ and $r_1 = a1 + b(-q_1)$, the formula in (1) holds for $i = 0$ and $i = 1$. Inductively assume $2 \leq k \leq n+1$ and that the formula in (1) holds for $0 \leq i < k$. Then

$$\begin{aligned} r_k &= r_{k-2} - q_k r_{k-1} \\ &= (as_{k-2} + bt_{k-2}) - q_k(as_{k-1} + bt_{k-1}) \\ &= a(s_{k-2} - q_k s_{k-1}) + b(t_{k-2} - q_k t_{k-1}) \\ &= as_k + bt_k. \end{aligned}$$

By Mathematical Induction, this proves (1).

(2): By Proposition 3.4.23, r_n is equal to $d = \gcd(a, b)$. This is a special case of (1).

(3): This part of the proof uses properties of 2-by-2 matrices. Let $M_i = \begin{bmatrix} s_{i-1} & t_{i-1} \\ s_i & t_i \end{bmatrix}$ and $A_i = \begin{bmatrix} 0 & 1 \\ 1 & -q_i \end{bmatrix}$. Then $M_1 = \begin{bmatrix} 0 & 1 \\ 1 & -q_1 \end{bmatrix} = A_1$ and for $i > 1$ we have:

$$\begin{aligned} A_i M_{i-1} &= \begin{bmatrix} 0 & 1 \\ 1 & -q_i \end{bmatrix} \begin{bmatrix} s_{i-2} & t_{i-2} \\ s_{i-1} & t_{i-1} \end{bmatrix} \\ &= \begin{bmatrix} s_{i-1} & t_{i-1} \\ s_{i-2} - q_i s_{i-1} & t_{i-2} - q_i t_{i-1} \end{bmatrix} \\ &= \begin{bmatrix} s_{i-1} & t_{i-1} \\ s_i & t_i \end{bmatrix} = M_i. \end{aligned}$$

Therefore, M_i factors into $M_i = A_i A_{i-1} \cdots A_1$. As was shown in Example 2.1.21, the determinant function $\det : M_2(\mathbb{Q}) \to \mathbb{Q}$ is multiplicative. Since $\det(A_i) = -1$ for each i, this implies $\det(M_i) = (-1)^i$. In particular,

$$\begin{aligned} (-1)^i &= \det(M_i) \\ &= \det \begin{bmatrix} s_{i-1} & t_{i-1} \\ s_i & t_i \end{bmatrix} \\ &= s_{i-1} t_i - s_i t_{i-1}. \end{aligned} \tag{4.3}$$

It follows from Eq. (4.3) that $Rs_i + Rt_i$ is the unit ideal. By Proposition 3.4.8, $\gcd(s_i, t_i) = 1$.

(4): From (1) with $i = n + 1$, we have $0 = as_{n+1} + bt_{n+1}$. From (3) and Exercise 3.4.25, we conclude that s_{n+1} divides b and t_{n+1} divides a. The reader should verify that the statement in (4) holds if $s_{n+1} = 0$ or $t_{n+1} = 0$. Now assume s_{n+1} and t_{n+1} are both nonzero. Write $b = s_{n+1}x$ and $a = t_{n+1}y$. Substituting into $as_{n+1} = -bt_{n+1}$, we get

$$t_{n+1}ys_{n+1} = -s_{n+1}xt_{n+1}. \tag{4.4}$$

From Eq. (4.4) and Lemma 3.2.19 (2), it follows that x and y are associates. Therefore $x \mid b$, $x \mid a$ and $\gcd(bx^{-1}, ax^{-1}) = \gcd(s_{n+1}, t_{n+1}) = 1$. By Exercise 3.4.29, $x = \gcd(a, b)$. Hence x and y are both associates of d.

(5): Except for the length of the sequences and the initial terms, the s_i and t_i sequences for a and b are the same as the t_i and s_i sequences obtained when the algorithm is applied with the roles of a and b swapped. Assume without loss of generality that $b < a$. Since b does not divide a, $0 < r_1 < b$ and $n \geq 1$. Then for $i = 1, 2, \ldots, n$, we have $1 \leq q_i$ and $0 < r_i < r_{i-1}$. Since $r_{n+1} = 0$ and $1 \leq r_n < r_{n-1}$, we have $2 \leq q_{n+1}$. We have $s_0 = 0$, $s_1 = 1$, $s_2 = -q_1 < 0$. It follows from induction and the recurrence relation $s_i = s_{i-2} - q_i s_{i-1}$ that $s_1, s_2, \ldots, s_n, s_{n+1}$ is an alternating sequence. Using this, it follows that $|s_i| = |s_{i-2}| + q_i|s_{i-1}| \geq q_i|s_{i-1}| \geq |s_{i-1}|$. Thus $|s_1| \leq |s_2| \leq \cdots \leq |s_n| \leq |s_{n+1}|$.

From Part (4) we have $d|s_{n+1}| = b$. Then

$$\begin{aligned} \frac{b}{d} &= |s_{n+1}| \\ &= |s_{n-1}| + q_{n+1}|s_n| \\ &\geq q_{n+1}|s_n| \\ &\geq 2|s_n| \end{aligned}$$

A similar proof using the sequence $\{t_0, t_1, \ldots, t_{n+1}\}$ proves that $a/d \geq 2|t_n|$. □

4.4. Exercises.

EXERCISE 3.4.25. Let a and b be elements of a commutative ring R. If $(a, b) = (1)$ and $a \mid bc$, then $a \mid c$.

EXERCISE 3.4.26. Let X be a nonempty subset of a commutative ring R. If $d \in (X)$ and $d \mid x$ for all $x \in X$, then $(d) = (X)$.

EXERCISE 3.4.27. Let $X = \{x_1, \ldots, x_n\}$ be a nonempty finite subset of a commutative ring R, with $n \geq 2$. If $e = \gcd(x_1, \ldots, x_{n-1})$ and $d = \gcd(e, x_n)$, then $d = \gcd(x_1, \ldots, x_n)$.

EXERCISE 3.4.28. (Exponential Notation in a UFD) Let a and b be elements of a unique factorization domain R. Assume a and b are both nonzero and nonunits.

(1) Show that there are representations of a and b:

$$\begin{aligned} a &= x_1^{e_1} \cdots x_m^{e_m}, \text{ and} \\ b &= u x_1^{f_1} \cdots x_m^{f_m}, \end{aligned}$$

where $x_1, \ldots, x_m$ are irreducible elements in R such that x_i and x_j are associates of each other if and only if $i = j$, u is a unit in R, and $e_1, \ldots, e_m$, $f_1, \ldots, f_m$ are nonnegative integers.

(2) Show that in the notation from (1) that $a \mid b$ if and only if $e_i \leq f_i$ for each i.

(3) In the notation from (1), for $j = 1, \ldots, m$, let ℓ_j be the least element in the set $\{e_j, f_j\}$. Prove that $d = x_1^{\ell_1} x_2^{\ell_2} \cdots x_m^{\ell_m} = \gcd(a, b)$.

EXERCISE 3.4.29. Let R be an integral domain and X a nonempty subset of R.

(1) Assume $d = \gcd(X)$ exists and $d \neq 0$. Let $Y = \{xd^{-1} \mid x \in X\}$ (see Lemma 3.4.3 (5) for this notation). Prove that $1 = \gcd(Y)$.
(2) Suppose $d \in R - \{0\}$ and $d \mid x$ for every $x \in X$. As in (1), let $Y = \{xd^{-1} \mid x \in X\}$. Prove that if $1 = \gcd(Y)$, then $d = \gcd(X)$.

EXERCISE 3.4.30. Let R be a principal ideal domain and P a nonzero prime ideal of R. Show that P is a maximal ideal.

EXERCISE 3.4.31. Let a, b, and c be a elements of a commutative ring R.

(1) Show that if $(a, c) = (p)$ is a principal ideal, $(b, c) = (q)$ is a principal ideal, and $(a, b) = (1)$ is the unit ideal R, then (ab, c) is equal to the principal ideal (pq).
(2) Show that if $(a, c) = (d)$ is a principal ideal and (b, c) is the unit ideal R, then (ab, c) is equal to the principal ideal (d).

5. The Quotient Field of an Integral Domain

An integral domain R is an abstraction of the ring of integers $\mathbb{Z}$. Just as the ring of integers can be extended to the field of rational numbers, the ring R can be extended to a field K. In other words, R can be assumed to be a subring of a field K. We build the field K from R in the same way that the field of rational numbers $\mathbb{Q}$ is constructed from the ring of integers $\mathbb{Z}$. The field $\mathbb{Q}$ is the set of all quotients, or fractions, of integers. The field $\mathbb{Q}$ contains $\mathbb{Z}$ as a subring. The definition of $\mathbb{Q}$ as an extension of $\mathbb{Z}$ generalizes to the setting of any integral domain R. Following the construction of the field of rational numbers, the field K of all quotients, or fractions, of elements from R is defined and the ring R embeds in a natural way as a subring of the field K.

Let R be an integral domain. Define a relation on $R \times (R - (0))$ by the rule: $(r, v) \sim (s, w)$ if and only if $rw = sv$. We show that $\sim$ is an equivalence relation. Clearly $\sim$ is reflexive and symmetric. Let us show that it is transitive. Suppose $(r, u) \sim (s, v)$ and $(s, v) \sim (t, w)$. Then $rv = su$ and $sw = tv$. Multiply the first equality by w and the second by u to get $rvw = suw = tvu$. Then $rvw = tvu$. Canceling v, $rw = tu$, which implies $(r, u) \sim (t, w)$. We have shown that $\sim$ is an equivalence relation on $R \times (R - (0))$. The set of equivalence classes, $(R \times (R - (0)))\,/\,\sim$, is called the *quotient field*, or *field of fractions of* R. The equivalence class containing (r, w) is denoted by the fraction r/w.

LEMMA 3.5.1. *Let R be an integral domain and $K = (R \times (R - (0)))\,/\,\sim$ the quotient field of R. Then K is a field with the binary operations*

$$\frac{r}{v} + \frac{s}{w} = \frac{rw + sv}{vw}, \quad \frac{r}{v}\frac{s}{w} = \frac{rs}{vw}.$$

The additive identity is $0/1$, the multiplicative identity is $1/1$. There is a natural map $\theta : R \to K$ defined by $r \mapsto r/1$ which is a one-to-one homomorphism of rings. If R is a field, then θ is an isomorphism.

PROOF. Assume $\frac{r}{v} = \frac{r_1}{v_1}$ and $\frac{s}{w} = \frac{s_1}{w_1}$. Then

$$rv_1 = r_1 v \tag{5.1}$$

$$sw_1 = s_1 w. \tag{5.2}$$

Multiply (5.1) by ww_1 and (5.2) by vv_1 to get the identities $rv_1ww_1 = r_1vww_1$ and $sw_1vv_1 = s_1wvv_1$. From these we derive

$$\begin{aligned}(rw + sv)v_1w_1 &= rv_1ww_1 + sw_1vv_1 \\ &= r_1vww_1 + s_1wvv_1 \\ &= (r_1w_1 + s_1v_1)vw.\end{aligned}$$

This is the center equation in:

$$\frac{r}{v} + \frac{s}{w} = \frac{rw + sv}{vw} = \frac{r_1w_1 + s_1v_1}{v_1w_1} = \frac{r_1}{v_1} + \frac{s_1}{w_1}.$$

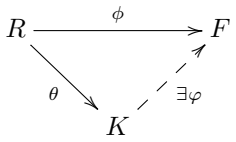
Hence, addition of fractions is well defined. Multiply (5.1) by sw_1 and (5.2) by r_1v to get the identities $rsv_1w_1 = r_1vsw_1$ and $sw_1r_1v = s_1wr_1v$. Taken together, we have $rsv_1w_1 = r_1vsw_1 = s_1wr_1v$. This is the center equation in:

$$\frac{r}{v}\frac{s}{w} = \frac{rs}{vw} = \frac{r_1s_1}{v_1w_1} = \frac{r_1}{v_1}\frac{s_1}{w_1}.$$

Hence, multiplication of fractions is well defined. It is routine to check that the associative and distributive laws hold, that K is a field, and that θ is a one-to-one homomorphism of rings. The details are left to the reader. □

5.1. Exercises.

EXERCISE 3.5.2. (Universal Mapping Property) Let R be an integral domain with field of fractions K. Let F be a field and $\phi : R \to F$ a one-to-one homomorphism of rings. Prove that there is a unique homomorphism of fields $\varphi : K \to F$ such that the diagram



commutes where θ is the natural map of Lemma 3.5.1.

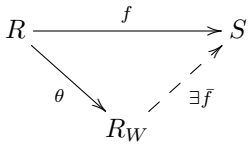
EXERCISE 3.5.3. Let R be a commutative ring. A subset W of R is called a *multiplicative subset of* R, if the following three properties hold:

(a) $1 \in W$.
(b) W contains no zero divisor of R.
(c) If x and y are in W, then $xy \in W$.

If W is a multiplicative subset of R, do the following:

(1) Define a relation on $R \times W$ by the rule: $(r, v) \sim (s, w)$ if and only if $rw = sv$. Show that $\sim$ is an equivalence relation. Denote the set of equivalence classes by R_W.
(2) Show how to make R_W into a commutative ring by imitating the construction of the quotient field of an integral domain in Lemma 3.5.1. The ring R_W is called the *localization* of R at W.
(3) Show that there is a one-to-one homomorphism of rings $\theta : R \to R_W$.

(4) (Universal Mapping Property) Let S be a commutative ring and $f : R \to S$ a homomorphism such that $f(W) \subseteq \text{Units}(S)$. Show that there exists a unique homomorphism $\bar{f} : R_W \to S$



such that $f = \bar{f}\theta$.

EXERCISE 3.5.4. Let R be a commutative ring and W the set of all elements of R that are not zero divisors.

(1) Show that W is a multiplicative subset. In this case, the localization R_W is called the *total ring of quotients* of R.
(2) Let S be the total ring of quotients of R. Show that S is a commutative ring with the property that every element of S is either a unit or a zero divisor.

EXERCISE 3.5.5. Let R be a finite ring in which $0 \neq 1$, and $x \in R$. Show that if x is not a zero divisor, then x is invertible.

EXERCISE 3.5.6. Let D be an integer that is not a square. Let $\sqrt{D}$ be the complex number given by Proposition 1.4.3 (5).

(1) Show that $\mathbb{Q}[\sqrt{D}] = \{r + s\sqrt{D} \mid r, s \in \mathbb{Q}\}$ is a subfield of $\mathbb{C}$. The field $\mathbb{Q}[\sqrt{D}]$ is an example of an *algebraic number field*.
(2) Show that $\mathbb{Z}[\sqrt{D}] = \{a + b\sqrt{D} \mid a, b \in \mathbb{Z}\}$ is a subring of $\mathbb{Q}[\sqrt{D}]$.
(3) Show that $\mathbb{Q}[\sqrt{D}]$ is equal to the quotient field of $\mathbb{Z}[\sqrt{D}]$.

EXERCISE 3.5.7. Let R be a commutative ring. Denote by R^* the group of units in R. Show that the following are equivalent.

(1) R is a local ring (see Exercise 3.2.43).
(2) For every $r \in R$, either $r \in R^*$ or $1 - r \in R^*$.
(3) For every pair r, s in R, if $r + s = 1$, then either $r \in R^*$ or $s \in R^*$.

EXERCISE 3.5.8. Prove that if R is a local ring, then 0 and 1 are the only idempotents in R.

6. Polynomial Rings

Given a commutative ring R, the ring of all polynomials in one variable x is a commutative ring that contains R as a subring. The construction is purely formal. In other words, a polynomial in x with coefficients $a_0, a_1, \ldots, a_m$ in R is not defined as a function on R, but as a formal sum of the form $a_0 + a_1x + \cdots + a_mx^m$. The variable x is also called an *indeterminate*. We tacitly assume $a_i = 0$ for all $i > m$. In this case, two polynomials $f = a_0 + a_1x + \cdots + a_mx^m$ and $g = b_0 + b_1x + \cdots + b_nx^n$ are equal if and only if $a_i = b_i$ for each $0 \leq i$. The *polynomial ring in one variable x with coefficients in R* is

$$R[x] = \left\{ \sum_{i=0}^{n} a_i x^i \mid n \geq 0, a_i \in R \right\}.$$

The sum and product of two polynomials are defined in the usual way. Suppose $f = a_0 + a_1x + \cdots + a_mx^m$ and $g = b_0 + b_1x + \cdots + b_nx^n$ are polynomials in $R[x]$. By inserting terms with zero coefficients to f or g if necessary, we can assume $m = n$. The sum of f and g is defined by adding the coefficients of the corresponding powers of x:

$$\begin{aligned} f + g &= (a_0 + a_1x + \cdots + a_nx^n) + (b_0 + b_1x + \cdots + b_nx^n) \\ &= (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n. \end{aligned} \tag{6.1}$$

The product of f and g is

$$\begin{aligned} fg &= \left(\sum_{i=0}^{m} a_ix^i\right)\left(\sum_{i=0}^{n} b_ix^i\right) \\ &= a_0b_0 + (a_0b_1 + a_1b_0)x + \cdots + \left(\sum_{j=0}^{k} a_jb_{k-j}\right)x^k + \cdots + a_mb_nx^{m+n}. \end{aligned} \tag{6.2}$$

Notice that the indeterminate x commutes with elements of R. The reader should verify that the binary operations of addition and multiplication are commutative, associative, and that multiplication distributes over addition. The polynomial with every coefficient $a_i = 0$ is simply denoted 0. The polynomial with $a_0 = 1$ and $a_i = 0$ for all $i > 0$ is denoted 1. The polynomial 0 is the additive identity and the polynomial 1 is the multiplicative identity. Therefore, the set $R[x]$ of polynomials is a commutative ring.

If $a \in R$ and $n \geq 0$, the polynomial ax^n is called a *monomial*. If $a \in R - (0)$, the *degree* of the monomial ax^n is n. For convenience, the degree of 0 is taken to be $-\infty$. The *degree* of a polynomial $f = \sum_{i=0}^{n} a_ix^i$ in $R[x]$ is the maximum of the degrees of the terms $a_0x^0, \ldots, a_nx^n$. The degree of f is denoted $\deg(f)$. If f is nonzero of degree n, the *leading coefficient* of f is a_n. We say that f is *monic* if the leading coefficient of f is 1.

PROPOSITION 3.6.1. *Let R be a commutative ring. Let $f = \sum_{i=0}^{m} a_ix^i$ and $g = \sum_{i=0}^{n} b_ix^i$ be polynomials in $R[x]$. If $\deg(f) = m$ and $\deg(g) = n$, then the following are true.*

(1) $\deg(f + g) \leq \max(\deg(f), \deg(g))$.
(2) $\deg(fg) \leq m + n$.
(3) If a_m or b_n is not a zero divisor in R, then $\deg(fg) = \deg(f) + \deg(g)$.
(4) R is an integral domain if and only if $R[x]$ is an integral domain. In general, R is an integral domain if and only if $R[x_1, \ldots, x_t]$ is an integral domain.

PROOF. (1): This follows from (6.1).

(2): This follows from (6.2).

(3): If one of the leading coefficients a_m or b_n is not a zero divisor in R, it follows from (6.2) that a_mb_n is the leading coefficient of fg.

(4): Since R is a subring of $R[x]$, if R has a nonzero zero divisor, so does $R[x]$. If R is an integral domain, then it follows from (3) that $R[x]$ is an integral domain. If $t > 1$, the proof follows by Mathematical Induction. □

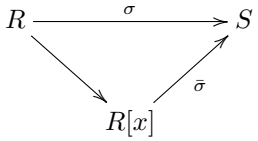
We view R as the subring of all polynomials in $R[x]$ of degree less than or equal to 0. The natural mapping $R \to R[x]$ which maps $a \in R - (0)$ to the polynomial of degree zero is a monomorphism. The polynomial ring over R in several variables

is defined by iterating the one-variable construction. If $t > 1$ and $x_1, \ldots, x_t$ are indeterminates, then $R[x_1, \ldots, x_t] = R[x_1, \ldots, x_{t-1}][x_t]$. See Section 3.6.1.

THEOREM 3.6.2. *Let R be a commutative ring and $\sigma : R \to S$ a homomorphism of rings.*

(1) If S is a commutative ring, the definition $\bar{\sigma}(\sum r_i x^i) = \sum \sigma(r_i) x^i$ extends σ to a homomorphism on the polynomial rings $\bar{\sigma} : R[x] \to S[x]$. If $K = \ker(\sigma)$, then the kernel of $\bar{\sigma}$ is the set $K[x]$ consisting of those polynomials $f \in R[x]$ such that every coefficient of f is in K.

(2) (Universal Mapping Property) Let s be an element of S such that $s\sigma(r) = \sigma(r)s$ for every $r \in R$. Then there is a unique homomorphism $\bar{\sigma}$ such that $\bar{\sigma}(x) = s$ and the diagram



commutes. We say $\bar{\sigma}$ is the evaluation homomorphism *defined by $x \mapsto s$.*

PROOF. The proof is left to the reader. □

THEOREM 3.6.3. *(The Division Algorithm) Let R be a commutative ring. Let $f, g \in R[x]$ and assume the leading coefficient of g is a unit of R. There exist unique polynomials $q, r \in R[x]$ such that $f = qg + r$ and $\deg r < \deg g$.*

PROOF. (Existence.) If $\deg f < \deg g$, then set $q = 0$ and $r = f$. Otherwise assume $f = \sum_{i=0}^{m} a_i x^i$ where $a_m \neq 0$ and $g = \sum_{i=0}^{n} b_i x^i$ where $b_n \neq 0$ and b_n is a unit in R. If $m = 0$, then $n = 0$ so $q = a_0 b_0^{-1}$ and $r = 0$. Proceed by induction on m. The leading coefficient of $\left(a_m b_n^{-1} x^{m-n}\right) g$ is a_m. Set $h = f - \left(a_m b_n^{-1} x^{m-n}\right) g$. Then $\deg h < \deg f$. By induction, $h = q_1 g + r$ where $\deg r < \deg g$. Now

$$\begin{aligned} f &= \left(a_m b_n^{-1} x^{m-n}\right) g + q_1 g + r \\ &= \left(a_m b_n^{-1} x^{m-n} + q_1\right) g + r \end{aligned}$$

so take $q = a_m b_n^{-1} x^{m-n} + q_1$.

(Uniqueness.) Assume $f = gq + r = gq_1 + r_1$ where $\deg r < \deg g$ and $\deg r_1 < \deg g$. Subtracting, we have $g(q - q_1) = r_1 - r$. The degree of the right hand side is $\deg(r_1 - r) \leq \max(\deg r_1, \deg r) < \deg g$. The degree of the left hand side is $\deg g + \deg(q - q_1)$. If $q - q_1 \neq 0$, this is impossible. So $q_1 = q$ and $r = r_1$. Hence the quotient and remainder are unique. □

COROLLARY 3.6.4. *(Synthetic Division) Let R be a commutative ring, $f(x) = \sum_{i=0}^{m} r_i x^i$ a polynomial in $R[x]$, and a an element in R. Then there exists a unique polynomial $q(x) \in R[x]$ such that $f(x) = q(x)(x - a) + f(a)$ where $f(a) = \sum_{i=0}^{m} r_i a^i \in R$.*

PROOF. Upon dividing $x - a$ into $f(x)$, this follows straight from Theorem 3.6.3. □

COROLLARY 3.6.5. *If F is a field, then $F[x]$ is a euclidean domain. It follows that $F[x]$ is a principal ideal domain and a unique factorization domain.*

PROOF. Define the norm function by the exponential formula: $\delta(f) = 2^{\deg f}$ for all $f \in F[x] - (0)$. Then $\delta(fg) = 2^{\deg fg} = 2^{\deg f + \deg g} = 2^{\deg f} 2^{\deg g} = \delta(f)\delta(g)$, hence δ is multiplicative. In Definition 3.4.16, property (2) is the division algorithm on $F[x]$. □

If k is a field, and $R = k[x]$, then the quotient field of $k[x]$, denoted $k(x)$, is called the field of rational functions over k. If S is a ring and R a subring, then by Theorem 3.6.2 we can view $R[x]$ as a subring of $S[x]$.

EXAMPLE 3.6.6. Let R be a commutative ring and $g \in R[x]$ a monic polynomial of degree n. Consider the residue class ring $R[x]/(g)$. Given any $f \in R[x]$, by the Division Algorithm, Theorem 3.6.3, there is a unique polynomial $r \in R[x]$ such that $f + (g) = r + (g)$ and $\deg r < n$. Therefore, the set of polynomials $\{r \in R[x] \mid \deg r < n\}$ is a complete set of coset representatives for $R[x]/(g)$.

DEFINITION 3.6.7. Let R be a commutative ring, $u \in R$, and $f = \sum_{i=0}^{m} r_i x^i \in R[x]$. We say that u is a *root* of f in case $f(u) = \sum_{i=0}^{m} r_i u^i = 0$.

LEMMA 3.6.8. *Let R be a commutative ring, $u \in R$, and $f \in R[x]$. The following are equivalent.*

(1) u is a root of f.
(2) f is in the kernel of the evaluation homomorphism $R[x] \to R$ defined by $x \mapsto u$.
(3) There exists $q \in R[x]$ such that $f = (x - u)q$.

PROOF. The proof is left to the reader. □

COROLLARY 3.6.9. *If R is an integral domain, and $f \in R[x]$ has degree $d \geq 0$, then the following are true:*

(1) If u is a root of f in R, then there exists $m \geq 1$ such that $f = (x - u)^m q$ and $q(u) \neq 0$.
(2) f has at most d roots in R.

PROOF. (1): Apply Lemma 3.6.8 and induction on the degree.

(2): If $d = 0$, then f has no root. Inductively assume $d \geq 1$ and that the result holds for any polynomial of degree in the range $0, \ldots, d-1$. If f has no root, then we are done. Suppose u is a root of f. By Part (1) we can write $f = (x - u)^m q$, where $\deg q = d - m$. If $v \neq u$ is another root of f, then $0 = f(u) = (v - u)^m q(u)$. Since R is an integral domain, this means u is a root of q. By induction, there are at most $d - m$ choices for v. □

COROLLARY 3.6.10. *(Lagrange Interpolation) Let F be a field and $n \geq 1$. Given $n + 1$ distinct elements of F: $\alpha_0, \ldots, \alpha_n$, and $n + 1$ arbitrary elements of F: $\beta_0, \ldots, \beta_n$, there exists a unique polynomial $f \in F[x]$ such that $\deg f \leq n$ and $f(\alpha_i) = \beta_i$ for each i.*

PROOF. (Existence.) The Lagrange basis polynomials with respect to the set $\{\alpha_0, \ldots, \alpha_n\}$ are

$$L_0(x) = \frac{(x-\alpha_1)\cdots(x-\alpha_n)}{(\alpha_0-\alpha_1)\cdots(\alpha_0-\alpha_n)}$$

$$\vdots$$

$$L_j(x) = \frac{(x-\alpha_0)\cdots(x-\alpha_{j-1})(x-\alpha_{j+1})\cdots(x-\alpha_n)}{(\alpha_j-\alpha_0)\cdots(\alpha_j-\alpha_{j-1})(\alpha_j-\alpha_{j+1})\cdots(\alpha_j-\alpha_n)}$$

$$\vdots$$

$$L_n(x) = \frac{(x-\alpha_0)\cdots(x-\alpha_{n-1})}{(\alpha_n-\alpha_0)\cdots(\alpha_n-\alpha_{n-1})}.$$

Notice that $L_j(x)$ has degree n and

$$L_j(\alpha_k) = \begin{cases} 0 & \text{if } k \neq j \\ 1 & \text{if } k = j. \end{cases}$$

Set

$$f(x) = \sum_{j=0}^{n} \beta_j L_j(x).$$

Then $f(\alpha_k) = \beta_k$ for each $k = 0, \ldots, n$ and $\deg f \leq n$.

(Uniqueness.) Suppose f and g are two polynomials in $F[x]$ such that $\deg f \leq n$, $\deg g \leq n$ and $f(\alpha_k) = \beta_k = g(\alpha_k)$ for each $k = 0, \ldots, n$. Then $\deg(f-g) \leq n$ and $f-g$ has $n+1$ roots, namely $\alpha_0, \ldots, \alpha_n$. By Corollary 3.6.9 (2), $f-g = 0$. □

COROLLARY 3.6.11. *Let R be an integral domain. Let $n > 1$ be an integer. The group of nth roots of unity in R, $\mu_n = \{u \in R \mid u^n = 1\}$, is a cyclic group of order at most n.*

PROOF. The set μ_n is the kernel of the nth power map $\pi^n : R^* \to R^*$ (see Exercise 2.3.18). Therefore, μ_n is a subgroup of R^*. The order of μ_n is at most n, by Corollary 3.6.9 (2). For every divisor d of n, the equation $x^d = 1$ has at most d solutions in R^*. By Exercise 2.8.12, μ_n is a cyclic group. □

COROLLARY 3.6.12. *Let F be a finite field of order q. Then F^* is a cyclic abelian group of order $q-1$.*

PROOF. In a field the nonzero elements make up an abelian group. The group F^* has order $q-1$. By Corollary 2.2.19, every $u \in F^*$ satisfies the equation $u^{q-1} = 1$. By Corollary 3.6.11, F^* is a cyclic group of order $q-1$. □

EXAMPLE 3.6.13. If F is a field, the ring $F[x,y]$ is not a PID. The ideal $(x,y) = \{ux + vy \mid u, v \in F[x,y]\}$ is not a principal ideal.

DEFINITION 3.6.14. If R is an integral domain, $f \in R[x]$, and u is a root of f, then the *multiplicity* of u as a root of f is the positive number m given by Corollary 3.6.9 (1). We say that u is a *simple root* if $m = 1$. If $m > 1$, then u is called a *multiple root*.

A significant portion of Chapter 5 deals with the subject of separable field extensions. In Section 5.2.2, the definition of a separable field extension is based on polynomials that have only simple roots. We also mention that in Numerical Analysis, the problem of approximating a solution to a polynomial equation is theoretically more challenging when the root is not a simple root.

DEFINITION 3.6.15. If R is any ring and $f = \sum_{i=0}^{n} a_i x^i \in R[x]$, then the *formal derivative* of f is defined to be

$$f' = \sum_{i=1}^{n} i a_i x^{i-1}$$

which is also in $R[x]$. The reader should verify the usual identities satisfied by the derivative operator. In particular, $(af + bg)' = af' + bg'$ and $(fg)' = f'g + fg'$. If R is commutative, then $(f^n)' = nf^{n-1}f'$.

Proposition 3.6.16 presents necessary and sufficient conditions for an element to be a simple root of a polynomial. This is sometimes called a jacobian criterion because it is based on the derivative.

PROPOSITION 3.6.16. *Suppose S is an integral domain and R is a subring of S. Let f be a nonconstant polynomial in $R[x]$ and $u \in S$. Then u is a multiple root of f if and only if $f'(u) = f(u) = 0$.*

PROOF. Suppose u is a multiple root of f. Write $f = (x-u)^2 q$ for some $q \in S[x]$ and compute $f' = 2(x-u)q + (x-u)^2 q'$. It is immediate that $f'(u) = 0$. Conversely, assume $f(u) = f'(u) = 0$. Write $f = (x-u)q$ for some $q \in S[x]$ and compute $f' = q + (x-u)q'$. It is immediate that $q(u) = 0$, so $f = (x-u)^2 q_2$ for some $q_2 \in S[x]$. □

Part (1) of Theorem 3.6.17 presents useful sufficient criteria for a polynomial to have no multiple roots. When the characteristic of the coefficient field is positive, Part (2) has useful necessary and sufficient conditions for the existence of multiple roots in case the characteristic of the ground field is positive.

THEOREM 3.6.17. *Let k be a subfield of the integral domain S and f a nonconstant polynomial in $k[x]$.*

(1) Assume
  (a) $\gcd(f, f') = 1$, or
  (b) f is irreducible in $k[x]$ and $f' \neq 0$ in $k[x]$, or
  (c) f is irreducible in $k[x]$ and k has characteristic zero (see Example 3.2.2 (5)).

  Then f has no multiple root in S.

(2) Suppose p denotes the characteristic of k. Assume u is a root of f in S.
  (a) If f is irreducible in $k[x]$ and u is a multiple root of f, then $p > 0$ and $f \in k[x^p]$.
  (b) If $p > 0$ and $f \in k[x^p]$, then u is a multiple root of f.

PROOF. (1): Assuming $\gcd(f, f') = 1$, by Corollary 3.4.9 there exist $s, t \in k[x]$ such that $1 = fs + f't$. It is clear that f and f' do not have a common root in S. By Proposition 3.6.16, f has no multiple root in S. Case (b) reduces immediately to case (a). Case (c) reduces immediately to case (b).

(2) (a): If $u \in S$ is a multiple root of f, then because f is irreducible in $k[x]$, Part (1) implies $p > 0$ and $f' = 0$. The reader should verify that under these conditions $f \in k[x^p]$.

(2) (b): If k has characteristic $p > 0$ and $f \in k[x^p]$, then clearly $f' = 0$. If $u \in S$ is a root of f, then by Proposition 3.6.16, u is a multiple root of f. □

6.1. Polynomials in Several Variables. The polynomial ring over R in several variables is defined by iterating the one-variable construction. If $m > 1$ and $x_1, \ldots, x_m$ are indeterminates, then $R[x_1, \ldots, x_m] = R[x_1, \ldots, x_{m-1}][x_m]$. A *monomial* in $S = R[x_1, \ldots, x_m]$ is a polynomial of the form $M = rx_1^{e_1} \cdots x_m^{e_m}$, where $r \in R$ is the *coefficient* and each exponent e_i is a nonnegative integer. The *degree* of a monomial is $-\infty$ if $r = 0$, otherwise it is the sum of the exponents. If $M \neq 0$, then $\deg M = e_1 + \cdots + e_m$. If M_1 and M_2 are monomials with coefficients r_1, r_2, then M_1M_2 is a monomial with coefficient r_1r_2. So $M_1M_2 = 0$ if and only if $r_1r_2 = 0$. If $M_1M_2 \neq 0$, then $\deg M_1M_2 = \deg M_1 + \deg M_2$. A polynomial f in S is a sum $f = \sum_{j=1}^{d} M_j$ where each M_j is a monomial. A polynomial $f \in S$ is said to be *homogeneous* if f can be written as a sum of monomials all of the same degree. Let $S_0 = R$ be the set of all polynomials in S of degree less than or equal to 0. For all $n \geq 1$, let S_n be the R-submodule generated by the set of all homogeneous polynomials in S of degree n. If f is homogeneous of degree d and g is homogeneous of degree e, then we see fg is homogeneous of degree $d + e$. A polynomial $f \in S$ can be written $f = f_0 + f_1 + \cdots + f_d$ where each f_i is homogeneous of degree i. We call f_i the *homogeneous component of f of degree i*. This representation of f as a sum of homogeneous polynomials is unique. The *degree* of a polynomial is the maximum of the degrees of the homogeneous components. If k is a field, then $k[x_1, \ldots, x_m]$ is an integral domain. The quotient field of $k[x_1, \ldots, x_m]$, denoted $k(x_1, \ldots, x_m)$, is called the field of rational functions over k in m variables.

In Exercise 1.2.23 the lexicographical order $\leq$ is defined on the set of all m-tuples of nonnegative integers $\prod_{i=1}^{m} \mathbb{Z}_{\geq 0} = \{(e_1, \ldots, e_m) \mid x_i \in \mathbb{Z}_{\geq 0}\}$. Under this partial ordering $\prod_{i=1}^{m} \mathbb{Z}_{\geq 0}$ is a chain. This notion induces the *lexicographical order* on the set of nonzero monomials in $R[x_1, \ldots, x_m]$. If $M_1 = r_1x_1^{a_1} \cdots x_m^{a_m}$, and $M_2 = r_2x_1^{b_1} \cdots x_m^{b_m}$ are two nonzero monomials, then $M_1 < M_2$ if and only if $(a_1, \ldots, a_m) < (b_1, \ldots, b_m)$. We see that M_1 and M_2 are comparable if $(a_1, \ldots, a_m)$ is not equal to $(b_1, \ldots, b_m)$.

LEMMA 3.6.18. *Let R be a ring and $S = R[x_1, \ldots, x_m]$.*

1. *A nonzero polynomial f in S can be written as a sum $f = \sum_{j=1}^{d} M_j$ where each M_j is a nonzero monomial such that $M_1 < M_2 < \cdots < M_d$. This representation as a sum of strictly increasing monomials is unique. The monomial M_d is called the* leading term *of f.*
2. *Let f and g be nonzero polynomials in S. Let $L(f)$ be the leading term of f and $L(g)$ the leading term of g. Then the leading term of fg is equal to $L(f)L(g)$.*
3. *If U is a nonempty set of nonzero monomials in S, then there exists an element $\alpha \in U$ with the property that if $\beta \in U$ and β is comparable to α, then $\alpha < \beta$. If U has the property that any two distinct elements are comparable, then there exists $\alpha \in U$ such that if $\beta \in U - \{\alpha\}$, then $\alpha < \beta$.*

PROOF. (1): Given a nonzero polynomial f, write $f = \sum_{j=1}^{d} M_j$ where each M_j is a nonzero monomial. By adding coefficients, all monomials that are incomparable

can be combined. Hence we can assume the monomials appearing in the sum are comparable. After rearranging if necessary, we can assume $M_1 < M_2 < \cdots < M_d$. Conversely, if $M_1 < M_2 < \cdots < M_d$ is a strictly increasing sequence of monomials, then the sum $f = \sum_{j=1}^{d} M_j$ is nonzero. The uniqueness claim follows from this fact.

(2): The proof of this part is left to the reader.

(3): This follows from Exercise 1.2.23 (3). □

6.2. The Group of Units Modulo p^a. As an application of Theorem 2.8.7 and Corollary 3.6.12, we show that the group of units modulo p^a is a cyclic group if p is an odd prime and $a \geq 1$. The finite ring $\mathbb{Z}/p^a$ is a principal ideal ring. If $a = 1$, then by Corollary 3.6.12, the group of units in the prime field $\mathbb{Z}/p$ is a cyclic group of order $p-1$. In the notation of Lemma 1.2.12 and Example 2.1.3, let U_{p^a} denote the group of units modulo p^a. Then $[U_{p^a} : 1] = \phi(p^a) = p^a - p^{a-1} = p^a(p-1)$. The congruence classes in U_{p^a} correspond to the numbers in $\{0, 1, \ldots, p^a - 1\}$ that are not divisible by p.

PROPOSITION 3.6.19. *Let p be an odd prime and $a \geq 1$. If $a = 1$, the group U_p is cyclic of order $p-1$. If $a > 1$, the natural map $\eta : U_{p^a} \to U_p$ is onto and the kernel of η is $1 + \langle p \rangle$, which is a cyclic group of order p^{a-1}. Hence U_{p^a} is isomorphic to the direct product $(1 + \langle p \rangle) \times U_p$ and is a cyclic group of order $p^{a-1}(p-1)$.*

PROOF. It follows from Corollary 3.6.12 that the group of units in the field $\mathbb{Z}/p$ is cyclic of order $p-1$. Assume from now on that $a > 1$. Let $\eta : \mathbb{Z}/p^a \to \mathbb{Z}/p$ be the natural map. The kernel of η is the principal ideal $\langle p \rangle$. The order of $\langle p \rangle$ is p^{a-1}. Then η induces a surjective map on the groups of units $\eta : U_{p^a} \to U_p$ and the kernel is the subgroup $1 + \langle p \rangle$ which also has order p^{a-1}. By Exercise 2.8.12, to prove that $1 + \langle p \rangle$ is a cyclic group it suffices to show that there are at most $p-1$ elements of order p. Let $1 + p^k x$ be an arbitrary element of $1 + \langle p \rangle$. Assume $\gcd(p, x) = 1$. This implies $k \geq 1$. By the Binomial Theorem we have

$$\begin{aligned}
(1 + p^k x)^p &= 1 + \binom{p}{1} p^k x + \binom{p}{2} (p^k x)^2 + \binom{p}{3} (p^k x)^3 + \cdots + (p^k x)^p \\
&= 1 + p^{k+1} x + \frac{p(p-1)}{2} p^k p^k x^2 + \frac{p(p-1)(p-2)}{3 \cdot 2} p^k p^{2k} x^3 + \cdots + p^{pk} x^p \\
&= 1 + p^{k+1} \left(x + \frac{p-1}{2} p^k x^2 + \frac{(p-1)(p-2)}{3 \cdot 2} p^{2k} x^3 + \cdots + p^{pk-(k+1)} x^p \right) \\
&= 1 + p^{k+1} v.
\end{aligned}$$

Since $k \geq 1$, this means v is relatively prime to p. In this case, $(1 + p^k x)^p \equiv 1 \pmod{p^a}$ if and only if $k + 1 \geq a$. So the subgroup annihilated by p in $1 + \langle p \rangle$ is $1 + \langle p^{a-1} \rangle$, a group of order p. By Exercise 2.8.12, $1 + \langle p \rangle$ is a cyclic group of order p^{a-1}. By Theorem 2.8.7, $U_{p^a} = (1 + \langle p \rangle) \times V$ is the internal direct product where V is a subgroup of order $p-1$. However, V is isomorphic to the factor group $U_{p^a}/(1 + \langle p \rangle)$ which is isomorphic to the cyclic group U_p. By Theorem 2.5.2, U_{p^a} is cyclic. □

6.3. Exercises.

EXERCISE 3.6.20. Let k be a field of characteristic different from 2. Let $f = x^2 - 1$. Show that $k[x]/(f)$ is isomorphic to a direct sum of fields.

EXERCISE 3.6.21. Let k be a field. Let $R = k[x^2, x^3]$ be the subring of $k[x]$ consisting of all polynomials such that the coefficient of x is zero. Prove:

(1) R is an integral domain.
(2) R is not a UFD.
(3) R is not a PID.
(4) The converse of Lemma 3.4.5 (3) is false.
(5) Show that the quotient field of $k[x^2, x^3]$ is equal to $k(x)$. In other words, show that R and $k[x]$ have the same quotient field.

EXERCISE 3.6.22. Let R be a commutative ring and $I = (a)$ a principal ideal in R. Show that for any $n \geq 1$, $I^n = (a^n)$.

EXERCISE 3.6.23. Prove that if R is an integral domain, then the homomorphism $R \to R[x]$ induces an isomorphism on the groups of units $\mathrm{Units}(R) \to \mathrm{Units}(R[x])$.

EXERCISE 3.6.24. Let R be a commutative ring. Prove:

(1) The nil radical of $R[x]$ is equal to $\mathrm{Rad}_R(0)[x]$. That is, a polynomial is nilpotent if and only if every coefficient is nilpotent.
(2) The kernel of $R[x] \to (R/\mathrm{Rad}_R(0))[x]$ is equal to the nil radical of $R[x]$.
(3) The group of units of $R[x]$ consists of those polynomials of the form $f = a_0 + a_1x + \cdots + a_nx^n$, where a_0 is a unit in R and $f - a_0 \in \mathrm{Rad}_R(0)[x]$.
(4) If $\mathrm{Rad}_R(0) = (0)$, then the homomorphism $R \to R[x]$ induces an isomorphism on the groups of units $\mathrm{Units}(R) \to \mathrm{Units}(R[x])$.

EXERCISE 3.6.25. Let R be an integral domain and $a \in R$. Prove that the linear polynomial $x - a$ is a prime element in $R[x]$.

EXERCISE 3.6.26. Let R be a commutative ring and $a \in R$. Show that there is an automorphism $\theta : R[x] \to R[x]$ such that $\theta(x) = x + a$ and for all $r \in R$, $\theta(r) = r$.

EXERCISE 3.6.27. Let R be an integral domain and a an irreducible element of R. Prove that a is an irreducible element in $R[x]$.

EXERCISE 3.6.28. Let k be a field and $A = k[x]$. Prove:

(1) If $I = (x)$ is the ideal in A generated by x, then $I^n = (x^n)$.
(2) Let $n \geq 1$. The nil radical of $k[x]/(x^n)$ consists of those cosets represented by polynomials of the form $\alpha_1x + \cdots + \alpha_{n-1}x^{n-1}$.
(3) The group of units of $k[x]/(x^n)$ consists of those cosets represented by polynomials of the form $\alpha_0 + \alpha_1x + \cdots + \alpha_{n-1}x^{n-1}$, where α_0 is a unit in k.

EXERCISE 3.6.29. Let R be an integral domain.

(1) A polynomial f in $R[x]$ defines a function $f : R \to R$. If R is infinite, show that f is the zero function (that is, $f(a) = 0$ for all $a \in R$) if and only if f is the zero polynomial.

(2) A polynomial f in $R[x_1, \ldots, x_r]$ defines a function $f : R^r \to R$. If R is infinite, use induction on r to show f is the zero function if and only if f is the zero polynomial.

EXERCISE 3.6.30. Let R be a commutative ring and $S = R[x]$ the polynomial ring in one variable over R. If $W = \{1, x, x^2, \ldots\}$, then the localization S_W is called the *Laurent polynomial ring* over R (see Exercise 3.5.3). The ring of Laurent polynomials over R is named for P. A. Laurent, and is usually denoted $R[x, x^{-1}]$.

(1) Show that every element of $R[x, x^{-1}]$ has a unique representation in the form $f(x)/x^n$ where $f(x) \in R[x]$ and $n \geq 0$.
(2) If R is an integral domain, prove that the group of units in $R[x, x^{-1}]$ is equal to the set $\{ux^e \mid u \in R^* \text{ and } e \in \mathbb{Z}\}$.
(3) If R is an integral domain, prove that the group of units in $R[x, x^{-1}]$ is the internal direct product $R^* \times \langle x \rangle$.
(4) Let k be a field. Prove that $k[x, x^{-1}]$ is a PID.
(5) Let R be a UFD. Prove that $R[x, x^{-1}]$ is a UFD.

EXERCISE 3.6.31. Let R be a UFD and P a nonzero prime ideal of R. Prove that P contains a prime element π of R.

EXERCISE 3.6.32. (GCD is invariant under a change of base field) Let $k \subseteq F$ be a tower of fields such that k is a subfield of F. In this case we view $k[x]$ as a subring of $F[x]$. Let $f, g \in k[x]$. Prove that if d is the greatest common divisor of f and g in $k[x]$, then d is the greatest common divisor of f and g in $F[x]$.

EXERCISE 3.6.33. Let F be a field of positive characteristic p. Let $\theta : F[y] \to F[y]$ be the evaluation mapping given by $y \mapsto y^p$. Let $F[y^p]$ denote the image of θ. Prove that θ extends to a homomorphism $\chi : F(y) \to F(y)$ and let $F(y^p)$ be the image of χ. Prove that $F(y^p)$ is the quotient field of $F[y^p]$ and that the diagram

$$\begin{array}{ccc} F[y] & \longrightarrow & F(y) \\ \uparrow & & \uparrow \\ F[y^p] & \longrightarrow & F(y^p) \end{array}$$

commutes where each of the four maps is the set inclusion homomorphism.

EXERCISE 3.6.34. Let $K = F(y^p)$ be the subfield of $L = F(y)$ defined as in Exercise 3.6.33. We say that L/K is an extension of fields. Show that the polynomial $f = x^p - y^p$ is irreducible in $K[x]$, but that $f = (x - y)^p$ in $L[x]$.

EXERCISE 3.6.35. Let p be a prime number and R a commutative ring of characteristic p. Let $R[x, y]$ be the ring of polynomials in two variables with coefficients in R. Prove:

(1) If $n \geq 0$, then $(x + y)^{p^n} = x^{p^n} + y^{p^n}$ in $R[x, y]$.
(2) If $n > 0$ and $0 < k < p^n$, then $\binom{p^n}{k}$ is divisible by p.

EXERCISE 3.6.36. Assume R is a commutative ring and $\theta : R \to A$ is a homomorphism of rings such that the image of θ is a subring of the center of A. Let $a \in A$ and $\sigma : R[x] \to A$ the evaluation map defined by $x \mapsto a$. Let $R[a]$ denote the image of σ. Show that $R[a]$ is the smallest subring of A containing $\theta(R)$ and a. Show that $R[a]$ is commutative.

EXERCISE 3.6.37. Let R be a commutative ring and $a \in R$. Prove that $R[x]/(x-a) \cong R$.

EXERCISE 3.6.38. Let k be an infinite field and assume there exists a monic irreducible polynomial of degree d in $k[x]$. Show that there are infinitely many monic irreducible polynomials of degree d in $k[x]$.

7. Polynomials over a Unique Factorization Domain

Throughout this section, R denotes a unique factorization domain with quotient field K . This section contains four important theorems on the ring of polynomials $R[x]$ over R. The first main result is the Rational Root Theorem. It gives necessary conditions for a polynomial f in $R[x]$ to have a root in the field K. An important application of this theorem is the fact that if f is a monic polynomial in $R[x]$ and $u \in K$ is a root of f, then u is actually in R (see Exercise 3.7.20). This is an important property of R. In the terminology of Commutative Algebra, we say that R is an integrally closed integral domain. The second main theorem of this section is Gauss' Lemma, which shows that a primitive polynomial f in $R[x]$ is irreducible when viewed as an element of the ring $R[x]$ if and only if it is irreducible when viewed as an element of the larger ring $K[x]$. The third main result, which is a corollary to Gauss' Lemma, shows that the ring $R[x]$ is a unique factorization domain. The fourth main theorem of this section is Eisenstein's Irreducibility Criterion, which provides us with sufficient conditions such that a primitive polynomial in $R[x]$ is irreducible. As an application of Eisenstein's Criterion, we show that if p is a prime number, then the cyclotomic polynomial $1 + x + x^2 + \cdots + x^{p-1}$, is irreducible in $\mathbb{Q}[x]$. As another application of the theorems of this section, we show that for any square free integer D such that $D \equiv 1 \pmod 4$, the ring $\mathbb{Z}[\sqrt{D}]$ is a subring of $\mathbb{C}$ which is not a unique factorization domain. In Section 3.7.1 a version of Eisenstein's Criterion is derived for the polynomial ring $K[y]$, where $K = k(x)$ is the field of rational functions in one variable over a field k.

Theorem 3.7.1 is usually stated and proved in the context where R is $\mathbb{Z}$, the ring of integers, and K is $\mathbb{Q}$, the field of rational numbers. This explains why it goes by the name Rational Root Theorem.

THEOREM 3.7.1. *(The Rational Root Theorem) Suppose R is a UFD with quotient field K and $u = b/c$ is an element of K such that $\gcd(b,c) = 1$. If $f = a_0 + a_1x + \cdots + a_dx^d \in R[x]$ and u is a root of f, then $b \mid a_0$ and $c \mid a_d$.*

PROOF. If $f(b/c) = 0$, then

$$a_0 + \frac{a_1 b}{c} + \frac{a_2 b^2}{c^2} + \cdots + \frac{a_d b^d}{c^d} = 0.$$

Multiply by c^d

$$a_0 c^d + a_1 b c^{d-1} + a_2 b^2 c^{d-2} + \cdots + a_d b^d = 0.$$

Since b divides the last d terms, it follows that $b \mid a_0 c^d$. Since c divides the first d terms, it follows that $c \mid a_d b^d$. Since $\gcd(b,c) = 1$ and R is a UFD, it follows that $b \mid a_0$ and $c \mid a_d$. □

Let R be a unique factorization domain, or UFD for short. Suppose f is a nonzero polynomial in $R[x]$. If we write $f = a_0 + a_1x + \cdots + a_nx^n$, then the *content* of f, written $C(f)$, is defined to be $\gcd(a_0, a_1, \ldots, a_n)$. By Corollary 3.4.13, $C(f)$

is unique up to associates, which means $C(f)$ is unique up to multiplication by a unit of R. If $C(f) = 1$, then we say f is *primitive*.

LEMMA 3.7.2. *Let R be a UFD and f a nonzero polynomial in $R[x]$. If $c_1 = C(f)$, then f factors as $f = c_1 f_1$ where $f_1 \in R[x]$ is primitive. The factors c_1 and f_1 of f are unique up to associates in $R[x]$.*

PROOF. By Exercise 3.4.29, if we factor out the content, then $f = C(f)f_1$ where $C(f_1) = 1$. Both $C(f)$ and $C(f_1)$ are unique up to multiplication by units of R. But units of $R[x]$ correspond to the units of R by Exercise 3.6.23. So f_1 is unique up to associates in $R[x]$. □

LEMMA 3.7.3. *Let R be a UFD with quotient field K. Let f and g be nonzero polynomials in $R[x]$.*

(1) If f and g are primitive, then fg is primitive.
(2) $C(fg) = C(f)C(g)$.
(3) Suppose f and g are primitive. Then f and g are associates in $R[x]$ if and only if they are associates in $K[x]$.

PROOF. (1): Assume f and g are nonzero elements of $R[x]$ and fg is not primitive. Then $C(fg)$ is not a unit in R. Let p be an irreducible factor of $C(fg)$ in R. Under the natural map $\eta : R[x] \to R/(p)[x]$ of Theorem 3.6.2 (1), we have $\eta(fg) = \eta(f)\eta(g) = 0$. By Corollary 3.4.14, (p) is a prime ideal, so $R/(p)$ is an integral domain. Thus $R/(p)[x]$ is an integral domain, which implies one of $\eta(f)$ or $\eta(g)$ is zero. That is, p divides the content of f or the content of g. That is, either f or g is not primitive.

(2): As in Lemma 3.7.2, we factor $f = C(f)f_1$, $g = C(g)g_1$, where f_1 and g_1 are primitive. Then $fg = C(f)C(g)f_1g_1$. By Part (1), f_1g_1 is primitive. By Lemma 3.7.2, $C(fg) = C(f)C(g)$.

(3): We are given that $1 = C(f) = C(g)$. Assume f and g are associates in $K[x]$. By Exercise 3.6.23, a unit in $K[x]$ is a nonzero constant polynomial. Suppose $f = ug$ where $u = r/s$ is a unit in K and $\gcd(r, s) = 1$. Then $sf = rg$ implies $sC(f) = rC(g)$, which implies r and s are associates. Therefore u is a unit in R. The converse is trivial, since $R \subseteq K$. □

THEOREM 3.7.4. *(Gauss' Lemma) Let R be a UFD with quotient field K. Suppose $f \in R[x]$ is primitive. Then f is irreducible in $R[x]$ if and only if f is irreducible in $K[x]$.*

PROOF. If f has a nontrivial factorization in $R[x]$, then this factorization still holds in $K[x]$. Assume $f = pq$ is a factorization in $K[x]$, where we assume $m = \deg p \geq 1$, and $n = \deg q \geq 1$. Write

$$p = \sum_{i=0}^{m} \frac{a_i}{b_i} x^i, \quad q = \sum_{i=0}^{n} \frac{c_i}{d_i} x^i$$

and set $b = b_0 b_1 \cdots b_m$, $d = d_0 d_1 \ldots d_m$. Then $b(a_i/b_i) = \alpha_i \in R$ and $d(c_i/d_i) = \gamma_i \in R$ for each i, so we get

$$bp = \sum_{i=0}^{m} \alpha_i x^i, \quad dq = \sum_{i=0}^{n} \gamma_i x^i$$

are both in $R[x]$. Applying Lemma 3.7.2, let $\alpha = C(bp)$ and factor $bp = \alpha p_1$, where p_1 is primitive. Set $\gamma = C(dq)$ and factor $dq = \gamma q_1$ where q_1 is primitive (Lemma 3.7.2). Combining all of this, we have $(bd)f = (\alpha\gamma)(p_1 q_1)$. By Lemma 3.7.3, it follows that bd and $\alpha\gamma$ are associates in R. Up to a unit in R, $f = p_1 q_1$. □

THEOREM 3.7.5. *Let R be a UFD. Then $R[x_1, \ldots, x_n]$ is a UFD.*

PROOF. By finite induction, it is enough to show $R[x]$ is a UFD.

(Existence.) Let $f \in R[x]$ be a nonunit nonzero. If f has degree zero, then we can view f as an element of R and factor f into irreducibles in R. By Exercise 3.6.27, this is a factorization of f into irreducibles in $R[x]$.

Assume $\deg f \geq 1$ and factor $f = C(f)f_1$ where f_1 is primitive and $C(f) \in R$. Since $C(f)$ can be factored into irreducibles, we can reduce to the case where f is primitive. Let K be the quotient field of R. We know that $K[x]$ is a UFD, by Corollary 3.6.5. Let $f = p_1 \cdots p_n$ be the unique factorization of f into a product of irreducibles in $K[x]$. By Theorem 3.7.4, for each i we can write

$$p_i = \frac{a_i}{b_i} q_i$$

where $a_i, b_i \in R$, and $q_i \in R[x]$ is primitive and irreducible. Set $\alpha = a_1 \cdots a_n$ and $\beta = b_1 \cdots b_n$. Multiplying,

$$f = \frac{\alpha}{\beta} q_1 q_2 \cdots q_n.$$

By Lemma 3.7.3 (3) we conclude that α and β are associates in R. Up to associates, we have factored $f = q_1 q_2 \cdots q_n$ into irreducibles in $R[x]$.

(Uniqueness.) Let f be a nonzero nonunit element of $R[x]$. Then f can be factored into a product of irreducibles $f = (c_1 \cdots c_m)(p_1 p_2 \cdots p_n)$ where each p_i is a primitive irreducible polynomial in $R[x]$ and each c_i is an irreducible element of R. Up to associates, $C(f) = c_1 c_2 \cdots c_m$ is uniquely determined by f. Since R is a UFD, the factorization $C(f) = c_1 c_2 \cdots c_m$ is unique in R. In $K[x]$ the factorization $p_1 p_2 \cdots p_n$ is uniquely determined up to associates. By Lemma 3.7.3 (3), the factorization is unique in $R[x]$. □

THEOREM 3.7.6. *(Eisenstein's Irreducibility Criterion) Let R be UFD and $f = a_0 + a_1 x + \cdots + a_n x^n$ a primitive polynomial of degree $n \geq 1$ in $R[x]$. Let p be a prime in R such that $p \nmid a_n$, $p \mid a_i$ for $i = 0, 1, \ldots, a_{n-1}$, and $p^2 \nmid a_0$. Then f is irreducible.*

PROOF. Let $P = (p)$. Then P is a prime ideal in R by Corollary 3.4.14. The proof is by contraposition. Assume $a_n \notin P$, $(a_0, \ldots, a_{n-1}) \subseteq P$ and f is reducible. We prove that $p^2 \mid a_0$. By assumption, there is a factorization $f = gh$, where $\deg g = s \geq 1$, $\deg h = t \geq 1$, and $s + t = n$. By Theorem 3.6.2 (1) the natural map $\eta : R \to R/P$ induces $\bar{\eta} : R[x] \to R/P[x]$. Under this homomorphism, $\bar{\eta}(f) = \bar{\eta}(g)\bar{\eta}(h)$. By hypothesis, $\bar{\eta}(f) = \eta(a_n)x^n$ has degree n. If we write $g = b_0 + b_1 x + \cdots + b_s x^s$ and $h = c_0 + c_1 x + \cdots + c_t x^t$, then

$$\eta(a_n)x^n = (\eta(b_0) + \eta(b_1)x + \cdots + \eta(b_s)x^s)(\eta(c_0) + \eta(c_1)x + \cdots + \eta(c_t)x^t) \tag{7.1}$$

holds in $R/P[x]$. Since P is prime, R/P is an integral domain. Let K denote the quotient field of R/P. The factorization of $\bar{\eta}(f)$ in (7.1) holds in $K[x]$. By Corollary 3.6.5, $K[x]$ is a UFD. We conclude that $(b_0, b_1, \ldots, b_{s-1}) \subseteq P$ and

$(c_0, c_1, \ldots, c_{t-1}) \subseteq P$. In particular, $p \mid b_0$ and $p \mid c_0$. The constant term of f is equal to $a_0 = b_0 c_0$ which is divisible by p^2. □

EXAMPLE 3.7.7. Let k be a field and $f \in k[x]$. Assume $\deg f \geq 2$. The set of zeros of $y^2 - f$ in k^2 is called an affine hyperelliptic curve. Assume f is square free. By Theorem 3.7.6, $y^2 - f$ is irreducible in $k[x, y]$.

EXAMPLE 3.7.8. Let p be a prime number. Let $\Phi(x) = x^p - 1 \in \mathbb{Z}[x]$. Consider $\phi(x) = \Phi(x)/(x-1) = x^{p-1} + x^{p-2} + \cdots + x + 1$. By Exercise 3.6.26, the change of variable $x = y + 1$ induces an isomorphism $\mathbb{Z}[x] \cong \mathbb{Z}[y]$. Applying the Binomial Theorem (Exercise 3.1.22) we see that

$$\begin{aligned}\phi(y+1) &= \frac{\Phi(y+1)}{y} \\ &= \frac{(y+1)^p - 1}{y} \\ &= y^{p-1} + \binom{p}{1} y^{p-2} + \cdots + \binom{p}{p-2} y + \binom{p}{p-1}.\end{aligned}$$

By Exercise 1.2.21, p divides $\binom{p}{i}$ if $1 \leq i \leq p-1$. By Theorem 3.7.6, $\phi(y+1)$ is irreducible in $\mathbb{Z}[y]$. Therefore, $\phi(x)$ is irreducible in $\mathbb{Z}[x]$ and by Gauss' Lemma (Theorem 3.7.4), $\phi(x)$ is irreducible in $\mathbb{Q}[x]$.

EXAMPLE 3.7.9. In this example we apply Gauss' Lemma, Theorem 3.7.4, to construct a large class of rings of the form $\mathbb{Z}[\sqrt{D}]$ which are not unique factorization domains. Let D be a square free integer such that $D \equiv 1 \pmod 4$. Let $u = \sqrt{D}$ be the complex number given by Proposition 1.4.3 (5). If $f(x) = x^2 - D$, then by Theorem 3.7.6, $f(x)$ is irreducible in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$, hence u is not in $\mathbb{Q}$. If $S = \mathbb{Z}[u]$ and $L = \mathbb{Q}[u]$, then by Exercise 3.5.6, S is an integral domain and L is equal to the quotient field of S. In L, let $\alpha = (1+u)/2$ and $\beta = (1-u)/2$. Since u is not in $\mathbb{Q}$, we see that α and β are not in S. Since $D \equiv 1 \pmod 4$, there exists an integer k such that $1 = D + 4k$. Then $\alpha\beta = (1-u^2)/4 = (1-D)/4 = k$ and $\alpha + \beta = 1$. Consider the polynomial $g(y) = (y-\alpha)(y-\beta) = y^2 - y + k$ in $L[y]$. We conclude that $g(y)$ is irreducible in S, but factors in $L[y]$. By Theorem 3.7.4, this implies S is not a unique factorization domain.

7.1. Rational Function Fields. Let k be a field and x, y indeterminates. Let $K = k(x)$ be the field of rational functions over k in the variable x. A rational function $\phi \in K$ can be written as a quotient $\phi = p/q$ where $p, q \in k[x]$ are polynomials and $\gcd(p, q) = 1$. By unique factorization in $k[x]$, the polynomials p and q are uniquely determined up to associates. If $u \in k$, and $q(u) \neq 0$, then $\phi(u) = p(u)q(u)^{-1}$ is an element of k. The *pole set* of ϕ is the set of roots of q and the *zero set* of ϕ is the set of roots of p. If u is not a pole of ϕ, then $f(u) = p(u)q(u)^{-1}$ is a well defined element of k. So if the pole set of ϕ is not equal to k, ϕ defines a function on the complement of its pole set. The next theorem provides an Eisenstein irreducibility criterion for polynomials in $K[y]$. It first appeared in [**21**].

THEOREM 3.7.10. *Let k be a field and x, y indeterminates. Let $K = k(x)$ be the field of rational functions over k in the variable x. Let $f(y) = f_0 + f_1 y + f_2 y^2 + \cdots + f_n y^n$ be a polynomial in $K[y]$ where $n \geq 1$ and $f_n \neq 0$. If*

(1) each f_i is a polynomial in $k[x]$,

(2) x divides each of $f_0, f_1, \ldots, f_{n-1}$ and x does not divide f_n, and
(3) x^2 does not divide f_0,

then f is irreducible in $K[y]$.

PROOF. For sake of contradiction, suppose

$$f = (a_0 + a_1 y + \cdots + a_r y^r)(b_0 + b_1 y + \cdots + b_s y^s) \tag{7.2}$$

where $r \geq 1$, $s \geq 1$, and each a_i and b_j is in $K = k(x)$. We have

$$\begin{aligned} f_0 &= a_0 b_0 \\ f_1 &= a_0 b_1 + a_1 b_0 \\ f_2 &= a_0 b_2 + a_1 b_1 + a_2 b_0 \\ &\vdots \\ f_n &= a_0 b_n + \cdots + a_n b_0 \end{aligned}$$

By hypothesis (2), $0 = f_0(0) = f_1(0) = \cdots = f_{n-1}(0)$ and $f_n(0) \neq 0$. We start with $0 = f_0(0) = (a_0 b_0)(0)$. Write $a_0 = p/q$, $b_0 = g/h$, where p, q, g, h are polynomials in $k[x]$ and $\gcd(p, q) = \gcd(g, h) = 1$. Then $pq = f_0 qh$ in $k[x]$. Since $x \mid f_0$ we have $x \mid p$ or $x \mid q$. Suppose for contradiction's sake that $x \mid p$ and $x \mid q$. Then x does not divide q and x does not divide h. Thus x^2 divides f_0, a contradiction. Assume from now on that $x \mid a_0$ and x does not divide b_0. Equivalently, assume $a_0(0) = 0$ and $b_0(0) \neq 0$. Now we consider

$$0 = f_1(0) = (a_0 b_1)(0) + (a_1 b_0)(0). \tag{7.3}$$

From step one, $a_0(0) = 0$, $b_0(0) \neq 0$, hence (7.3) reduces to $0 = a_1(0)$. Now look at

$$0 = f_2(0) = (a_0 b_2)(0) + (a_1 b_1)(0) + (a_2 b_0)(0) \tag{7.4}$$

which reduces to $a_2(0) = 0$ by applying the first two steps. Iterating this argument, we see that $0 = a_0(0) = a_1(0) = a_2(0) = \cdots = a_r(0)$. This implies $f_n(0) = 0$, a contradiction. □

7.2. Exercises.

EXERCISE 3.7.11. Let $n \in \mathbb{Z}$ and consider the polynomial $f(x) = x^3 + nx - 2$. Show that $f(x)$ is reducible over $\mathbb{Q}$ if and only if n is in the set $\{1, -3, -5\}$.

EXERCISE 3.7.12. Let $f(x) = 20x^5 + 35x^4 - 42x^3 + 21x^2 + 70$ and $g(x) = 80x^5 + 18x^3 - 24x - 15$. Let $F = \mathbb{Q}[x]/(f)$ and $G = \mathbb{Q}[x]/(g)$. Show that F and G are fields.

EXERCISE 3.7.13. Modify the method of Example 3.7.8 to show that the following polynomials are irreducible over $\mathbb{Q}$.

(1) $x^4 + 1$
(2) $x^4 + a^2$, where $a \in \mathbb{Z}$ is odd.
(3) $x^8 + 1$
(4) $x^9 + 2$
(5) $x^{2^n} + a^2$, where $a \in \mathbb{Z}$ is odd and $n \geq 1$.
(6) $x^{p^n} + p - 1$, where p is prime and $n \geq 1$.

EXERCISE 3.7.14. Let k be a field. If $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ and $a_n \neq 0$, then the *reverse* of f is the polynomial $f^r(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n$.

(1) Show that $f^r(x) = x^n f(x^{-1})$.
(2) If $a_0 \neq 0$, show that f is irreducible over k if and only if f^r is irreducible over k.

EXERCISE 3.7.15. Let $f = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1} + a_nx^n$ be a polynomial of degree $n \geq 1$ in $\mathbb{Z}[x]$. Let p be a prime and $[f] = [a_0] + [a_1]x + [a_2]x^2 + \cdots + [a_{n-1}]x^{n-1} + [a_n]x^n$ be the polynomial over the prime field $\mathbb{Z}/(p)$ achieved by reducing the coefficients of f modulo p.

(1) If $[f]$ has degree n and is irreducible over $\mathbb{Z}/(p)$, then f is irreducible over $\mathbb{Q}$. Proof:
(2) Show by counterexample that (1) is false if the degree of $[f]$ is less than n.
(3) Show by counterexample that the converse of (1) is false.

EXERCISE 3.7.16. Let $f = x^3 + 1$. Prove that there is an isomorphism $\theta : \mathbb{Q}[x]/(f) \to F_1 \oplus F_2$ where F_1 and F_2 are fields. Carefully describe the fields F_1 and F_2, and the map θ.

EXERCISE 3.7.17. Let k be a field, a, b, c some elements of k and assume $a \neq b$. Let $f = (x-a)(x-b)$ and $g = (x-c)^2$. Prove:

(1) The ring $k[x]/(x-a)$ is isomorphic to k.
(2) There is an isomorphism of rings $k[x]/(f) \cong k \oplus k$.
(3) There is an isomorphism of rings $k[x]/(g) \cong k[x]/(x^2)$.
(4) If h is a monic irreducible quadratic polynomial in $k[x]$, then the rings $k[x]/(f)$, $k[x]/(g)$, and $k[x]/(h)$ are pairwise nonisomorphic.

EXERCISE 3.7.18. (Partial Fractions) Let k be a field. In this exercise we outline a proof that a rational function in one variable over k has a partial fraction decomposition. Prove:

(1) If f and g are two nonzero polynomials in $k[x]$ and $d = \gcd(f, g)$, then there exist polynomials u, v in $k[x]$ such that $d = fu + gv$, $\deg u < \deg g$, and $\deg v < \deg f$.
(2) If $1 = \gcd(f, g)$ and $\deg h < \deg(fg)$, then there exist unique polynomials u and v satisfying:
$$\frac{h}{fg} = \frac{u}{f} + \frac{v}{g},$$
$\deg u < \deg f$, and $\deg v < \deg g$.
(3) Let g be a polynomial of degree at least one. Let
$$g = p_1^{e_1} \cdots p_n^{e_n}$$
be the unique factorization of g where $p_1, \ldots, p_n$ are distinct irreducibles, $n \geq 2$, and $e_i \geq 1$ for each i. If f is a polynomial and $\deg f < \deg g$, then there exist unique polynomials $q_1, \ldots, q_n$ satisfying:
$$\frac{f}{g} = \frac{q_1}{p_1^{e_1}} + \cdots + \frac{q_n}{p_n^{e_n}},$$
and for each i, $\deg q_i < \deg p_i^{e_i}$.
(4) Let g be a polynomial of degree at least one, $n \geq 1$, and f a polynomial satisfying $\deg f < \deg g^n$. Then there exist unique polynomials $f_0, \ldots, f_{n-1}$ satisfying:
$$f = f_0 + f_1 g + \cdots + f_{n-1} g^{n-1}$$

and for each i, $\deg f_i < \deg g$.

(5) Let g be a polynomial of degree at least one, $n \geq 1$, and f a polynomial satisfying $\deg f < \deg g^n$. Then there exist unique polynomials $f_0, \ldots, f_{n-1}$ satisfying:
$$\frac{f}{g^n} = \frac{f_0}{g^n} + \frac{f_1}{g^{n-1}} + \cdots + \frac{f_{n-1}}{g},$$
and for each i, $\deg f_i < \deg g$.

EXERCISE 3.7.19. Let R be a UFD with quotient field K. Let f be a monic irreducible polynomial in $R[x]$.

(1) Show that $S = R[x]/(f)$ is an integral domain and $L = K[x]/(f)$ is a field.
(2) Show that there is a commutative square
$$\begin{array}{ccc} S & \longrightarrow & L \\ \uparrow & & \uparrow \\ R & \longrightarrow & K \end{array}$$
where each arrow is the natural map and each arrow is one-to-one.
(3) Show that L is the quotient field of S.

EXERCISE 3.7.20. Let R be a unique factorization domain with quotient field K. Let $p(x)$ be a monic polynomial in $R[x]$ and $u \in K$. Show that if u is a root of $p(x)$, then u is in R.

EXERCISE 3.7.21. Let k be a field. In Algebraic Geometry, the ring $k[x^2, x^3]$ of Exercise 3.6.21 corresponds to a cuspidal cubic curve and is not a UFD. The ring $k[x^2, x + x^3]$ corresponds to a nodal cubic curve.

(1) Show that the quotient field of $k[x^2, x + x^3]$ is $k(x)$. In other words, $k[x^2, x + x^3]$ and $k[x]$ are birational.
(2) Prove that $k[x^2, x + x^3]$ is not a UFD.

EXERCISE 3.7.22. Let k be a field and $A = k[x]$ the polynomial ring in one variable over k. Let R denote a subring of A which contains k as a subring such that $k \subsetneq R \subsetneq A$. Give specific examples of R satisfying the following.

(1) R is a principal ideal domain.
(2) R is not a principal ideal domain.
(3) The quotient field of R is equal to the quotient field of A.
(4) The quotient field of R is not equal to the quotient field of A.

CHAPTER 4

Modules, Vector Spaces, Algebras, Matrices

"What makes Linear Algebra linear?" is an important question that every student of this subject should be prepared to answer. I have not forgotten the first time I was asked this question. It was the beginning of the semester when I was taking my first undergraduate course on Linear Algebra. I was living on campus, and at the dining hall one evening one of the people at our table asked the above question. The event has stuck with me because I did not have an answer for my friend. Here is the answer to the question, and the response you should give when you are asked. Algebra is the study of polynomial equations and in this light, Linear Algebra is the study of linear equations.

As much as possible, we study linear algebra over a general ring. Nevertheless, because of the introductory nature of this book, many of the results assume the ground ring is commutative. We hope that a reasonable balance has been achieved between accessibility and generality of results. We define a module over an arbitrary ring and a vector space over a division ring. Algebras are defined over commutative rings. In Proposition 4.5.7 below, the isomorphism between the ring of endomorphisms of a finitely generated free module and the ring of matrices is constructed for an arbitrary commutative ring. The basis theorem for finitely generated modules over a principal ideal domain is proved in both the invariant factor form and the elementary divisor form.

1. Modules

1.1. Definitions and First Properties. In this section we introduce the notion of a module over an arbitrary ring R. An abelian group M is an R-module if multiplication by elements of R turns R into a ring of endomorphisms of M.

DEFINITION 4.1.1. If R is a ring, a *left R-module* is an additive abelian group M together with a left multiplication action by R such that for all $r, s \in R$ and $x, y \in M$ the rules

(1) $r(x + y) = rx + ry$
(2) $r(sx) = (rs)x$
(3) $(r + s)x = rx + sx$
(4) $1x = x$

are satisfied. If R is a field or a division ring, then M is called a *vector space*.

In the following text, an R-module is by default assumed to be a left R-module. This is in agreement with our convention that functions act from the left (Section 1.1.2). There will be occasions (for example, in Section 4.5.2) when we will utilize right R-modules. A *right R-module* is an additive abelian group M together with a right multiplication action by R such that for all $r, s \in R$ and $x, y \in M$ the rules

(1) $(x+y)r = xr + yr$
(2) $(xr)s = x(rs)$
(3) $x(r+s) = xr + xs$
(4) $x1 = x$

are satisfied.

In Lemma 2.4.1 we saw that a group G acts on a set X if and only if there is a homomorphism of G into $\mathrm{Perm}(X)$. Lemma 4.1.2 is the counterpart of this notion in the context of modules. By Exercise 2.8.13, if M is an abelian group, then the set of all endomorphisms of M, $\mathrm{Hom}(M, M)$, is a ring. Endomorphisms are added point-wise and multiplication is composition of functions.

LEMMA 4.1.2. *Let R be a ring and M an additive abelian group.*

(1) M is a left R-module if and only if there is a homomorphism of rings $\theta : R \to \mathrm{Hom}(M, M)$.
(2) M is a right R-module if and only if there is a homomorphism of rings $\theta : R^o \to \mathrm{Hom}(M, M)$.

PROOF. We prove (1). The proof of (2) is similar and left to the reader.

First assume there is a homomorphism of rings $\theta : R \to \mathrm{Hom}(M, M)$. Instead of $\theta(r)(x)$ we will write $r * x$. This defines a left multiplication action by R on M. Then

$$r * (x + y) = \theta(r)(x + y) = \theta(r)(x) + \theta(r)(y)) = r * x + r * y$$

is Part (1) of Definition 4.1.1,

$$r * (s * x) = \theta(r)(\theta(s)(x) = (\theta(r)\theta(s))(x) = \theta(rs)(x)) = (rs) * x$$

is Part (2),

$$(r + s) * x = \theta(r + s)(x) = (\theta(r) + \theta(s))(x) = \theta(r)(x) + \theta(s)(x) = r * x + s * x$$

is Part (3), and lastly,

$$1 * x = \theta(1)(x) = 1_M(x) = x$$

is Part (4).

Conversely, assume M is a left R-module. For each $r \in R$, define $\lambda_r : M \to M$ to be the "left multiplication by r" function defined by $\lambda_r(x) = rx$. By the first distributive law, $\lambda_r(x + y) = r(x + y) = rx + ry = \lambda_r(x) + \lambda_r(y)$, so $\lambda_r \in \mathrm{Hom}(M, M)$. Define $\theta : R \to \mathrm{Hom}(M, M)$ by $\theta(r) = \lambda_r$. The associative law implies $\lambda_{rs}(x) = (rs)x = r(sx)$, so $\theta(rs) = \theta(r)\theta(s)$ and θ is multiplicative. By the second distributive law, $\lambda_{r+s}(x) = (r + s)x = rx + sx = \lambda_r(x) + \lambda_s(x)$, so $\theta(r + s) = \theta(r) + \theta(s)$ and θ is additive. Lastly, $\lambda_1 = 1_M$, so $\theta(1) = 1$, hence θ is a homomorphism of rings. □

DEFINITION 4.1.3. Let R be a ring, M an R-module, and $\theta : R \to \mathrm{Hom}(M, M)$ the homomorphism of Lemma 4.1.2. The kernel of θ is denoted $\mathrm{annih}_R(M)$ and is called the *annihilator of M in R*. Then $\mathrm{annih}_R(M)$ is equal to $\{r \in R \mid rx = 0 \text{ for all } x \in M\}$. Since θ is a homomorphism of rings, $\mathrm{annih}_R(M)$ is a two-sided ideal in R. If θ is one-to-one, then we say M is a *faithful R-module*.

EXAMPLE 4.1.4. Standard examples of modules are listed here.

(1) If R is any ring, and I is a left ideal in R, then R acts on I from the left. If $x \in I$ and $r \in R$, then $rx \in I$. The associative and distributive laws in R apply. Thus I is an R-module. In particular, R is a left R-module.

(2) Let M be any additive abelian group. By Exercise 2.3.18, for any $n \in \mathbb{Z}$, left multiplication by n defines a homomorphism $\lambda_n : M \to M$. We can make M into a $\mathbb{Z}$-module by the following rule: for any $x \in M$, define nx to be $\lambda_n(x)$. It is important to mention that this is the only way to make M into a $\mathbb{Z}$-module. By Example 3.2.2 (5), there is a unique homomorphism of rings $\chi : \mathbb{Z} \to \mathrm{Hom}(M, M)$. Consequently, by Lemma 4.1.2, there is a unique way to make M into a $\mathbb{Z}$-module. It is routine to verify that $\chi(n) = \lambda_n$.

(3) Let A be an abelian group written additively. Let $m > 1$ be an integer and assume $mx = 0$ for all $x \in A$. It follows from Exercise 4.1.18 that A is a $\mathbb{Z}/m$-module by the action $[n]x = nx$. In particular, if p is a prime and $px = 0$ for all $x \in A$, then A is a vector space over the field $\mathbb{Z}/p$.

(4) Let $\phi : R \to S$ be a homomorphism of rings. By (1), S is a left S-module, and by Lemma 4.1.2, there is a homomorphism of rings $\theta : S \to \mathrm{Hom}(S, S)$ where $\theta(s) : S \to S$ is defined by $\theta(s)x = sx$ for all $s, x \in S$. By Exercise 3.2.28, the composite function $\theta\phi : R \to \mathrm{Hom}(S, S)$ is a homomorphism of rings. Hence, S is an R-module and R acts on S by the multiplication rule $rx = \phi(r)x$, for $r \in R$ and $x \in S$.

(5) Let $\phi : R \to S$ be a homomorphism of rings. If M is an S-module, then by Lemma 4.1.2, there is a homomorphism of rings $\theta : S \to \mathrm{Hom}(M, M)$. By Exercise 3.2.28, the composite function $\theta\phi : R \to \mathrm{Hom}(M, M)$ is a homomorphism of rings. Therefore, M is an R-module and R acts on M by the multiplication rule $rx = \phi(r)x$, for $r \in R$ and $x \in M$.

LEMMA 4.1.5. *Let M be an R-module, $x \in M$, and $r \in R$. Then the following are true:*

(1) $r0 = 0$.
(2) $0x = 0$.
(3) $-1x = -x$.

PROOF. (1): $r0 = r(0 + 0) = r0 + r0$. Since $M, +$ is a group, we cancel $r0$ to get $r0 = 0$.

(2): $0x = (0 + 0)x = 0x + 0x$. Since $M, +$ is a group, we cancel $0x$ to get $0x = 0$.

(3): $0 = (1 - 1)x = 1x + (-1)x = x + (-1)x$. Since $M, +$ is a group, we get $-x = (-1)x$. □

1.2. Submodules and Homomorphisms.

DEFINITION 4.1.6. Let R be a ring and M an R-module. A *submodule* of M is a nonempty subset $N \subseteq M$ such that N is an R-module under the operation by R on M. If $X \subseteq M$, the *submodule of M generated by X* is

$$\left\{ \sum_{i=1}^{n} r_i x_i \mid n \geq 1, r_i \in R, x_i \in X \right\}.$$

The reader should verify that the submodule generated by X is equal to the intersection of the submodules of M containing X. A submodule is *principal*, or *cyclic*, if it is generated by a single element. The submodule generated by X is denoted (X). If $X = \{x_1, x_2, \ldots, x_n\}$ is finite, we sometimes write $(X) = Rx_1 + Rx_2 + \cdots + Rx_n$. We say M is *finitely generated* if there exists a finite subset $\{x_1, \ldots, x_n\} \subseteq M$ such that $M = Rx_1 + \cdots + Rx_n$.

DEFINITION 4.1.7. If I is a left ideal of R and M is an R-module, then IM denotes the R-submodule of M generated by the set $\{rx \mid r \in I, x \in M\}$. Notice that a typical element of IM is not a product rx, but a finite sum of the form $r_1x_1 + \cdots + r_nx_n$.

DEFINITION 4.1.8. Let R be a ring and M an R-module. If A and B are R-submodules of M, then $A+B$ denotes the R-submodule generated by the set $A \cup B$. The reader should verify that the set of all submodules of M is a lattice.

DEFINITION 4.1.9. If M and N are R-modules, an *R-module homomorphism* from M to N is a function $f : M \to N$ satisfying

(1) $f(x + y) = f(x) + f(y)$ and
(2) $f(rx) = rf(x)$

for all $x, y \in M$ and $r \in R$. The *kernel* of the homomorphism f is $\ker(f) = \{x \in M \mid f(x) = 0\}$. The *image* of the homomorphism f is $\operatorname{im}(f) = \{f(x) \in N \mid x \in M\}$. An *epimorphism* is a homomorphism that is onto. A *monomorphism* is a homomorphism that is one-to-one. An *isomorphism* is a homomorphism $f : M \to N$ that is one-to-one and onto. In this case we say M and N are *isomorphic*. An *endomorphism* of M is a homomorphism from M to M.

PROPOSITION 4.1.10. *If $f : M \to N$ is an R-module homomorphism, then the following are true:*

(1) The kernel of f is a submodule of M.
(2) f is one-to-one if and only if $\ker(f) = (0)$.
(3) If A is a submodule of M, then $f(A)$, the image of A under f, is a submodule of N.
(4) If B is a submodule of N, then $f^{-1}(B)$, the preimage of B under f, is a submodule of M.
(5) If $g : N \to P$ is an R-module homomorphism, then the composite map $gf : M \to P$ is an R-module homomorphism.

PROOF. Let A be a submodule of M and B a submodule of N. Since f is a homomorphism of additive groups, $\ker(f)$ is a subgroup of $M, +$, $f(A)$ is a subgroup of $N, +$, and $f^{-1}(B)$ is a subgroup of $M, +$, by Lemma 2.3.3. Part (2) follows from the corresponding result for group homomorphisms, Lemma 2.3.8. Let $x \in \ker(f)$ and $r \in R$. Then $f(rx) = rf(x) = r0 = 0$ by Lemma 4.1.5. This completes Part (1). If x is an arbitrary element of A, then $f(x)$ represents a typical element of $f(A)$. Then $rf(x) = f(rx) \in f(A)$, which completes Part (3). Let $x \in M$ such that $f(x) \in B$. Then x represents a typical element of $f^{-1}(B)$. Then $f(rx) = rf(x) \in B$, which completes Part (4). By Lemma 2.3.8 (1), $gf(x+y) = gf(x)+gf(y)$. If $r \in R$, then $gf(rx) = g(r(f(x)) = rg(f(x))$, which proves Part (5). □

DEFINITION 4.1.11. Let R be a ring, M an R-module and S a submodule. The *factor* module of M modulo S is the set $M/S = \{a+S \mid a \in M\}$ of all left cosets of S in M. We sometimes call M/S the quotient module of M modulo S. We define addition and scalar multiplication of cosets by the rules

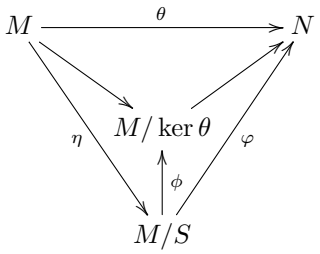
$$(a + S) + (b + S) = (a + b) + S$$
$$r(a + S) = ra + S.$$

The reader should verify that M/S is an R-module. Let $\eta : M \to M/S$ be the natural map defined by $x \mapsto x + S$. Then η is a homomorphism, $\operatorname{im}\eta = M/S$, and $\ker\eta = S$.

Theorem 4.1.12, Corollary 4.1.14, and Theorem 4.1.15 are the counterparts for modules of Theorems 2.3.12, 2.3.14 and 2.3.15.

THEOREM 4.1.12. *Let $\theta : M \to N$ be a homomorphism of R-modules. Let S be a submodule of M contained in $\ker\theta$. There exists a homomorphism $\varphi : M/S \to N$ satisfying the following.*

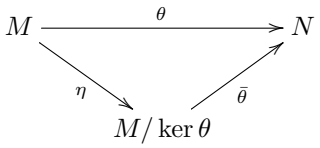
(1) $\varphi(a + S) = \theta(a)$, or in other words $\theta = \varphi\eta$.
(2) φ is the unique homomorphism from $M/S \to N$ such that $\theta = \varphi\eta$.
(3) $\operatorname{im}\theta = \operatorname{im}\varphi$.
(4) $\ker\varphi = \eta(\ker\theta) = \ker(\theta)/S$.
(5) φ is one-to-one if and only if $S = \ker\theta$.
(6) φ is onto if and only if θ is onto.
(7) There is a unique homomorphism $\phi : M/S \to M/\ker\theta$ such that the diagram



commutes.

PROOF. On the additive groups, this follows straight from Theorem 2.3.12. The rest is left to the reader. □

COROLLARY 4.1.13. *If $\theta : M \to N$ is a homomorphism of modules, then there exists a unique monomorphism $\bar\theta$ such that $\theta = \bar\theta\eta$. Hence θ factors into an epimorphism η followed by a monomorphism $\bar\theta$ and the diagram*



commutes.

PROOF. This is Theorem 4.1.12 (5). □

COROLLARY 4.1.14. *(The Isomorphism Theorems) Let M be an R-module with submodules A and B.*

(1) If $\theta : M \to N$ is a homomorphism of modules, then the map $\varphi : M/\ker\theta \to \operatorname{im}\theta$ sending the coset $x + \ker\theta$ to $\theta(x)$ is an isomorphism of modules.
(2) The natural map

$$\frac{A}{A \cap B} \to \frac{A + B}{B}$$

sending the coset $x + A \cap B$ to the coset $x + B$ is an isomorphism.

(3) If $A \subseteq B$, then B/A is a submodule of M/A and the natural map

$$\frac{M/A}{B/A} \to M/B$$

sending the coset containing $x + A$ to the coset $x + B$ is an isomorphism.

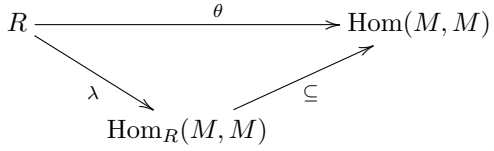
PROOF. This follows from Theorem 4.1.12 and Theorem 2.3.14, its counterpart for groups. □

THEOREM 4.1.15. *(The Correspondence Theorem) Let M be an R-module and A a submodule of M. There is a one-to-one order-preserving correspondence between the submodules B such that $A \subseteq B \subseteq M$ and the submodules of M/A given by $B \mapsto B/A$.*

PROOF. This follows from Proposition 4.1.10 and The Correspondence Theorem for Groups, Theorem 2.3.15. □

DEFINITION 4.1.16. If M and N are R-modules, the set of all R-module homomorphisms from M to N is denoted $\mathrm{Hom}_R(M, N)$. Modules are additive abelian groups and an abelian group has a natural structure as a $\mathbb{Z}$-module (Example 4.1.4 (2)). The set of all group homomorphisms from M to N is denoted $\mathrm{Hom}(M, N)$ or $\mathrm{Hom}_{\mathbb{Z}}(M, N)$. By Exercise 2.8.13, $\mathrm{Hom}_{\mathbb{Z}}(M, N)$ is an abelian group where addition of functions is defined point-wise. Since an R-module homomorphism $\phi : M \to N$ is a homomorphism of additive abelian groups, there is a set containment $\mathrm{Hom}_R(M, N) \subseteq \mathrm{Hom}_{\mathbb{Z}}(M, N)$. Hence $\mathrm{Hom}_R(M, N)$ is an abelian group. The reader should be advised that when R is noncommutative, $\mathrm{Hom}_R(M, N)$ is not an R-module per se. If $M = N$, then in Exercise 4.1.19 the reader is asked to prove that $\mathrm{Hom}_R(M, M)$ is a ring. In general, $\mathrm{Hom}_R(M, M)$ is a noncommutative ring.

EXAMPLE 4.1.17. Let R be a commutative ring and M an R-module. If $r \in R$, then "left multiplication by r" is the function $\lambda_r : M \to M$, where $\lambda_r(x) = rx$. As in Lemma 4.1.2, there is a homomorphism of rings $\theta : R \to \mathrm{Hom}(M, M)$ defined by $\theta(r) = \lambda_r$. Since R is commutative, if $r, s \in R$, then $\lambda_r(sx) = r(sx) = (rs)x = (sr)x = s(rx) = s\lambda_r(x)$. Therefore, λ_r is an R-module homomorphism from M to M. This shows that the homomorphism θ factors through a homomorphism $\lambda : R \to \mathrm{Hom}_R(M, M)$ which we call the *left regular representation* of R in $\mathrm{Hom}_R(M, M)$. The diagram of ring homomorphisms



commutes. For any $\phi \in \mathrm{Hom}_R(M, M)$, $r \in R$, and $x \in M$, $r\phi(x) = \phi(rx)$. Therefore, $\lambda_r\phi = \phi\lambda_r$, which implies the image of R under the homomorphism λ is a subring of the center of $\mathrm{Hom}_R(M, M)$. By λ, $\mathrm{Hom}_R(M, M)$ is turned into an R-algebra (see Definition 4.4.1).

1.3. Exercises.

EXERCISE 4.1.18. Let R be a commutative ring, I an ideal of R, and M an R-module. As in Definition 4.1.7, IM denotes the R-submodule of M generated by the set $\{rx \mid r \in I, x \in M\}$. Prove that M/IM is an R/I-module under the action $(r+I)(x+IM) = rx + IM$.

EXERCISE 4.1.19. This exercise is based on Exercise 2.8.13. Let M be an R-module, where R is any ring. Follow the outline below to show that the set $\mathrm{Hom}_R(M, M)$ of all R-module endomorphisms of M is a ring.

(1) If $f, g \in \mathrm{Hom}_R(M, M)$, then $f + g$ is the function defined by the rule: $(f+g)(x) = f(x) + g(x)$. Show that this additive binary operation makes $\mathrm{Hom}_R(M, M)$ into an abelian group.
(2) Show that composition of functions is a binary operation on $\mathrm{Hom}_R(M, M)$ and the following are satisfied:
   (a) $f(gh) = (fg)h$ for all f, g, h in $\mathrm{Hom}_R(M, M)$. In other words, composition of functions is associative.
   (b) $f(g + h) = fg + fh$ and $(f + g)h = fh + gh$ for all f, g, h in $\mathrm{Hom}_R(M, M)$. In other words, composition distributes over addition.

Together with the two binary operations of addition and composition of endomorphisms, we call $\mathrm{Hom}_R(M, M)$ the *ring of endomorphisms of* M.

EXERCISE 4.1.20. This exercise is based on Exercise 4.1.19. Let M be an R-module, where R is any ring. Let $S = \mathrm{Hom}_R(M, M)$ be the ring of R-module endomorphisms of M. Show that M is a left S-module under the action $\phi x = \phi(x)$, for all $\phi \in S$ and $x \in M$.

EXERCISE 4.1.21. Let R be a commutative ring and I an ideal in R. The natural ring homomorphism $\eta : R \to R/I$ turns R/I into an R-module (Example 4.1.4). Define

$$\phi : \mathrm{Hom}_R(R/I, R/I) \to R/I$$

by $\phi(f) = f(1 + I)$. Show that ϕ is an isomorphism of rings.

EXERCISE 4.1.22. Let R be a ring and M an R-module. Then M is said to be *simple* if its only submodules are (0) and M.

(1) Prove that any simple R-module is cyclic.
(2) Let M be a nonzero simple R-module. Prove that any R-module homomorphism $h : M \to M$ is either an automorphism of M, or $h(m) = 0$ for every $m \in M$.
(3) (Schur's Lemma) Let M be a nonzero simple R-module. Prove that $\mathrm{Hom}_R(M, M)$ is a division ring.
(4) Say $R = F$ is a field, $M = V$ is a finite dimensional F-vector space. Find necessary and sufficient conditions for V to be simple. Calculate $\mathrm{Hom}_F(V, V)$ for a nonzero simple F-vector space V.
(5) Say $R = \mathbb{Z}$ and M is a finitely generated $\mathbb{Z}$-module. Find necessary and sufficient conditions for M to be simple. Calculate $\mathrm{Hom}_{\mathbb{Z}}(M, M)$ for a nonzero simple $\mathbb{Z}$-module M.

EXERCISE 4.1.23. Let R be a ring. The opposite ring of R is defined in Definition 3.1.8. Show that there exists an isomorphism of rings $\mathrm{Hom}_R(R, R) \cong R^o$, where R is viewed as a left R-module and R^o denotes the opposite ring.

EXERCISE 4.1.24. (Module version of Finitely Generated over Finitely Generated is Finitely Generated) Let $R \to S$ be a homomorphism of rings such that S is finitely generated as an R-module. If M is a finitely generated S-module, prove that M is finitely generated as an R-module.

EXERCISE 4.1.25. Let $\theta : R \to S$ be a homomorphism of rings. Let M and N be S-modules. Via θ, M and N can be viewed as R-modules (see Example 4.1.4 (5)). Show that θ induces a well defined $\mathbb{Z}$-module monomorphism

$$H_\theta : \mathrm{Hom}_S(M, N) \to \mathrm{Hom}_R(M, N).$$

EXERCISE 4.1.26. Let $\theta : R \to S$ be a homomorphism of commutative rings. If M is an S-module, show that there is a commutative diagram of ring homomorphisms

$$\begin{array}{ccc} R & \xrightarrow{\lambda_R} & \mathrm{Hom}_R(M, M) \\ {\scriptstyle\theta}\downarrow & & \uparrow{\scriptstyle H_\theta} \\ S & \xrightarrow{\lambda_S} & \mathrm{Hom}_S(M, M) \end{array}$$

where λ_R and λ_S are the left regular representations of Example 4.1.17 and H_θ is one-to-one.

EXERCISE 4.1.27. Let R be a ring and M a left R-module. Follow the following outline to show that $\mathrm{Hom}_R(R, M)$ is isomorphic to M as a left R-module.

(1) Prove the following generalization to modules of Lemma 2.3.29 (2) (a). Let x be an element of M. Define $\rho_x : R \to M$ to be "right multiplication by x". That is, $\rho_x(a) = ax$ for every a in R. Then ρ_x is an R-module homomorphism in $\mathrm{Hom}_R(R, M)$.
(2) Show that the assignment $f \mapsto f(1)$ defines an isomorphism of additive abelian groups $\mathrm{Hom}_R(R, M) \to M$. This is a generalization of Exercise 2.8.17.
(3) Show that $\mathrm{Hom}_R(R, M)$ can be turned into a left R-module by the action $(rf)(x) = f(xr)$ for every $r \in R$ and $f \in \mathrm{Hom}_R(R, M)$.
(4) Show that $\mathrm{Hom}_R(R, M)$ and M are isomorphic as left R-modules.

2. Free Modules

Given a ring R and a family of R-modules $\{M_i \mid i \in I\}$, the R-module direct product $\prod_{i\in I} M_i$ is defined. The construction is based on the direct product of the underlying abelian groups together with coordinate-wise scalar multiplication. For each i in I, the module M_i is mapped isomorphically onto a submodule of the direct product under the canonical injection map ι_i. Thus, we identify M_i with the submodule $\iota_i(M_i)$. The R-module direct sum of the family $\{M_i \mid i \in I\}$ is the submodule of $\prod_{i\in I} M_i$ generated by the submodules M_i and is denoted $\bigoplus_{i\in I} M_i$. For a finite index set I, the direct sum is equal to the direct product. In general the direct sum and direct product are not equal. A free R-module is the direct sum of copies of the left R-module R. Free modules play a fundamental role in the sense that every module is the homomorphic image of a free module. A vector space is a free module over a field, or more generally, over a division ring.

2.1. Direct Product and Direct Sum of a Family of Modules. As mentioned above, we define the direct product and the direct sum of a family of R-modules $\{M_i \mid i \in I\}$ over an arbitrary index set I.

DEFINITION 4.2.1. Let R be a ring, I an index set and $\{M_i \mid i \in I\}$ a family of R-modules indexed by I. By Definition 2.5.1, the direct product $\prod_{i\in I} M_i = \{f : I \to \bigcup_{i\in I} \mid f(i) \in M_i\}$ is an abelian group. The binary operation is coordinate-wise addition: $(f+g)(i) = f(i) + g(i)$. The identity element, denoted 0, is the constant function $0(i) = 0$. The inverse of f is defined by $(-f)(i) = -f(i)$. We turn the direct product $\prod_{i\in I} M_i$ into an R-module by defining the R-action coordinate-wise: $(rf)(i) = rf(i)$. The R-module $\prod_{i\in I} M_i$ is called the *direct product* of $\{M_i \mid i \in I\}$. As in Definition 2.5.1, for each $k \in I$ there are the canonical injection and projection maps

$$M_k \xrightarrow{\iota_k} \prod_{i\in I} M_i \xrightarrow{\pi_k} M_k$$

such that $\pi_k\iota_k = 1_{M_k}$. The reader should verify that each ι_k and π_k is an R-module homomorphism.

The *direct sum* of $\{M_i \mid i \in I\}$ is denoted $\bigoplus_{i\in I} M_i$ and is defined to be the submodule of the direct product generated by the set $\bigcup_{k\in I} \iota_k(M_k)$. By Definition 4.1.6, it is routine to check that

$$\bigoplus_{i\in I} M_i = \Big\{ f : I \to \bigcup_{i\in I} M_i \mid f(i) \in M_i \text{ and } f(i) = 0 \text{ for all but finitely many } i \in I \Big\}.$$

For each $k \in I$ the canonical injection map ι_k factors through the direct sum. That is, $\iota_k : M_k \to \bigoplus_{i\in I} M_i$ is a one-to-one homomorphism of R-modules. All of the maps

$$M_k \xrightarrow{\iota_k} \bigoplus_{i\in I} M_i \xrightarrow{\subseteq} \prod_{i\in I} M_i \xrightarrow{\pi_k} M_k$$

are R-module homomorphisms. The restriction of π_k to the direct sum is an onto homomorphism of R-modules $\pi_k : \bigoplus_{i\in I} M_i \to M_k$. We have $\pi_k\iota_k = 1_{M_k}$. The direct sum $\bigoplus_{i\in I} M_i$ is sometimes called the *external direct sum* to distinguish it from the internal direct sum of submodules defined in Definition 4.2.3 below.

If the index set I is $\{1, \ldots, n\}$ and $M_1, \ldots, M_n$ are R-modules, then the direct product and the direct sum are equal. Both are the R-module with underlying set $M_1 \times \cdots \times M_n$ and with addition and R-action defined coordinate-wise on n-tuples:

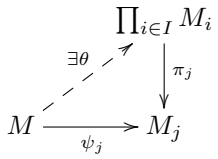
$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$
$$r(x_1, \ldots, x_n) = (rx_1, \ldots, rx_n).$$

In this case, the direct sum is sometimes denoted $M_1 \oplus M_2 \oplus \cdots \oplus M_n$.

PROPOSITION 4.2.2. *Let R be a ring, I an index set and $\{M_i \mid i \in I\}$ a family of R-modules indexed by I. Let M be an R-module.*

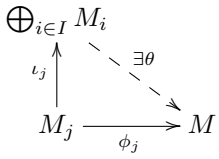
(1) (Universal Mapping Property) Given any family $\{\psi_i : M \to M_i \mid i \in I\}$ of R-module homomorphisms, there exists a unique R-module homomorphism

$\theta : M \to \prod_{i\in I} M_i$ *such that for each* $j \in I$ *the diagram*



commutes and $\pi_j\theta = \psi_j$.

(2) *(Universal Mapping Property) Given any family* $\{\phi_i : M_i \to M \mid i \in I\}$ *of* R*-module homomorphisms, there exists a unique* R*-module homomorphism* $\theta : \bigoplus_{i\in I} M_i \to M$ *such that for each* $j \in I$ *the diagram*



commutes and $\theta\iota_j = \phi_j$.

PROOF. (1): Given $x \in M$, define $\theta(x)$ to be the choice function in $\prod_{i\in I} M_i$ defined by $\theta(x)(i) = \psi_i(x)$. It is routine to check that θ is an R-module homomorphism, the diagram commutes, and θ is unique.

(2): Define $\theta : \bigoplus_{i\in I} M_i \to M$ by $\theta(f) = \sum_{i\in I} \phi_i(f(i))$. This is a well defined R-module homomorphism since $f(i)$ is nonzero on a finite subset of I. It is routine to check that the diagram commutes and θ is unique. □

DEFINITION 4.2.3. Let I be an index set and $\{S_i \mid i \in I\}$ a family of submodules of the R-module M. The submodule of M generated by the set $\bigcup_{i\in I} S_i$ is called the *sum* of the submodules and is denoted $\sum_{i\in I} S_i$. This is a generalization of Definition 4.1.8. Let $\bigoplus_{i\in I} S_i$ be the external direct sum of the R-modules $\{S_i \mid i \in I\}$. By Proposition 4.2.2 there exists an R-module homomorphism $\phi : \bigoplus_{i\in I} S_i \to M$ defined by $\phi(f) = \sum_{i\in I} f(i)$. Therefore the image of ϕ is equal to the sum $\sum_{i\in I} S_i$. We say that M is the *internal direct sum* of the submodules $\{S_i \mid i \in I\}$ in case ϕ is an isomorphism. In this case we write $M = \bigoplus_{i\in I} S_i$.

Proposition 4.2.4 lists some useful necessary and sufficient conditions for a module M to be the internal direct sum of a family of submodules.

PROPOSITION 4.2.4. *Let* I *be an index set and* $\{S_i \mid i \in I\}$ *a family of submodules of the* R*-module* M*. Then the following are equivalent.*

(1) $M = \bigoplus_{i\in I} S_i$ *is the internal direct sum of the submodules* $\{S_i \mid i \in I\}$.

(2) *For each* $x \in M$ *there is a unique representation of* x *in the form* $x = \sum_{i\in I} x_i$ *where each* x_i *comes from* S_i *and for all but finitely many* $i \in I$ *we have* $x_i = 0$.

(3) *The following are satisfied:*
   (a) $M = \sum_{i\in I} S_i$ *is the sum of the submodules* $\{S_i \mid i \in I\}$, *and*
   (b) *for every finite subset* $\{k_1, \ldots, k_n\}$ *of* I, *if* $x_{k_i} \in S_{k_i}$ *for* $1 \le i \le n$, *and* $0 = \sum_{i=1}^{n} x_{k_i}$, *then* $x_{k_i} = 0$ *for each* i.

(4) *The following are satisfied:*
   (a) $M = \sum_{i\in I} S_i$ *is the sum of the submodules* $\{S_i \mid i \in I\}$, *and*
   (b) *for every* $k \in I$, $S_k \cap \sum_{i\in I-\{k\}} S_i = \{0\}$.

PROOF. The proof that (1), (2) and (3) are equivalent to each other is left to the reader.

(2) implies (4): Suppose $x \in S_k \cap \sum_{i \in I-\{k\}} S_i$. Then there is a finite subset $\{k_1, \ldots, k_n\}$ of $I - \{k\}$ such that $x_{k_i} \in S_{k_i}$ for $1 \leq i \leq n$, and $x = \sum_{i=1}^{n} x_{k_i}$. Then $x - \sum_{i=1}^{n} x_{k_i} = 0$ and by (2) this implies $x = 0$.

(4) implies (3): Suppose there is a finite subset $\{k_1, \ldots, k_n\}$ of I such that $x_{k_i} \in S_{k_i}$ for $1 \leq i \leq n$ and $0 = \sum_{i=1}^{n} x_{k_i}$. Then $x_{k_1} = -\sum_{i=2}^{n} x_{k_i}$. So $x_{k_1} \in S_{k_1} \cap \sum_{i \in I-\{k_1\}} S_i = \{0\}$. This proves $x_{k_1} = 0$. The same argument shows $x_{k_i} = 0$ for each i. □

For convenience, Proposition 4.2.5 contains a version of Proposition 4.2.4 for the special case where the index set I is finite. It is the module theoretic version of Proposition 2.5.6.

PROPOSITION 4.2.5. *Suppose R is a ring, M is an R-module, and $S_1, \ldots, S_n$ are submodules of M. Then the following are equivalent.*

(1) $M = S_1 \oplus \cdots \oplus S_n$ is the internal direct sum of the submodules $\{S_1, \ldots, S_n\}$.
(2) For each $x \in M$ there is a unique representation of x in the form $x = x_1 + \cdots + x_n$ where x_i comes from S_i for each i.
(3) The following are satisfied:
  (a) $M = S_1 + \cdots + S_n$, and
  (b) if $x_i \in S_i$ for each i and $0 = x_1 + \cdots + x_n$, then $x_i = 0$ for each i.
(4) The following are satisfied:
  (a) $M = S_1 + \cdots + S_n$, and
  (b) for every $1 \leq k \leq n$, $S_k \cap \sum_{i \neq k} S_i = \{0\}$.

DEFINITION 4.2.6. Let M be an R-module. If N is a submodule of M, then N is called a *direct summand* of M if there is a submodule L of M such that $M = N \oplus L$.

EXAMPLE 4.2.7. Let A be a finite abelian group. By Example 4.1.4 (2), A is a $\mathbb{Z}$-module. We proved in Theorem 2.8.6 that if a is an element of maximal order in G, then the cyclic subgroup $\langle a \rangle$ is a direct summand of A.

PROPOSITION 4.2.8. *Let R be a ring, M an R-module, and N an R-submodule of M. The following are equivalent.*

(1) N is a direct summand of M.
(2) There is an R-module homomorphism $\pi : M \to N$ such that $\pi(x) = x$ for every $x \in N$.
(3) There is an R-module homomorphism $\phi : M \to M$ such that ϕ is an idempotent in the ring $\mathrm{Hom}_R(M, M)$ (that is, $\phi^2 = \phi$), and $\mathrm{im}(\phi) = N$.

PROOF. (1) implies (2): There is a submodule L of M such that $M = N \oplus L$. The canonical projection map $\pi : M \to N$ is an R-module homomorphism and $\pi(x) = x$ for every $x \in N$.

(2) implies (3): Let $\iota : N \to M$ be the set inclusion map, and ϕ the composite map $\iota\pi$.

(3) implies (1): Let $L = \ker \phi$. Given $z \in M$, let $x = \phi(z)$ and $y = z - x$. Then $\phi(y) = \phi(z) - \phi(x) = x - x = 0$ implies $y \in L$. This shows $M = N + L$. Let $z \in N \cap L$. Then $z \in L$ implies $\phi(z) = 0$ and $z \in N$ implies $z = \phi(x)$ for some $x \in M$. Hence $\phi(z) = \phi(\phi(x)) = \phi(x) = z$. This shows $N \cap L = (0)$. By Proposition 4.2.5, $M = N \oplus L$. □

2.2. Free Modules. An R-module is free if it is isomorphic to a direct sum of copies of R. A free module has a linearly independent generating set, or a free basis. A free basis has no dependence relation, hence is "relation-free", or "relation-less". Hence, as in Definition 2.5.13, the descriptor "free" is applied to signify that a module has a relation-free basis. We saw in Section 2.5.3 that free groups play a fundamental role in abstract group theory. Likewise, we show in the present section that free modules are fundamental. Free modules satisfy a universal mapping property. This implies every module is the homomorphic image of a free module.

DEFINITION 4.2.9. Let R be any ring. As defined in Definition 4.1.6, an R-module M is finitely generated if there exist elements $x_1, \ldots, x_n$ in M such that for each $m \in M$ there exist $r_1, \ldots, r_n$ in R such that $m = r_1x_1 + \cdots + r_nx_n$. Equivalently, M is finitely generated if there is a finite subset $\{x_1, \ldots, x_n\}$ of M such that $M = Rx_1 + \cdots + Rx_n$. Thus, M is finitely generated if and only if M is equal to the sum of a finite number of cyclic submodules. If M has a finite generating set, then by the Well Ordering Principle, there exists a generating set with minimal cardinality. We call such a generating set a *minimal generating set*. The *rank of* M, written $\mathrm{Rank}(M)$, is defined to be the number of elements in a minimal generating set.

EXAMPLE 4.2.10. If k is a field and V is a finite dimensional k-vector space, then we will see in Theorem 4.3.4 below that the rank of V as defined in Definition 4.2.9 is equal to $\dim_k(V)$, the dimension of V over k.

DEFINITION 4.2.11. Let R be a ring and I any index set. For $i \in I$, let $R_i = R$ as R-modules. By Example 4.1.4 (1), R is a left R-module. Denote by R^I the R-module direct sum $\bigoplus_{i\in I} R_i$. Let M be an R-module. We say M is *free* if M is isomorphic to R^I for some index set I. If $I = \{1, 2, \ldots, n\}$, then we will write $R^{(n)}$ for the direct sum $R \oplus \cdots \oplus R$ of n copies of R. An R module M is said to be *free of finite rank* n if M is isomorphic to $R^{(n)}$ for some n. In particular, $\mathbb{Z}^{(n)}$ is a free $\mathbb{Z}$-module of rank n.

DEFINITION 4.2.12. Let M be an R module. If $X = \{x_1, \ldots, x_n\}$ is a finite subset of M, define $\Sigma_X : R^{(n)} \to M$ by $\Sigma_X(r_1, \ldots, r_n) = r_1x_1 + \ldots r_nx_n$. Using Exercise 4.1.27 (1) and Proposition 4.2.2 (2), the reader should verify that Σ_X is an R-module homomorphism and the image of Σ_X is the R-submodule of M generated by X. We say X is a *linearly independent set* in case Σ_X is one-to-one. An arbitrary subset $Y \subseteq M$ is a *linearly independent set* if every finite subset of Y is linearly independent.

DEFINITION 4.2.13. The function $\delta : I \times I \to \{0, 1\}$ defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

is called the Kronecker *delta function*, and is named after L. Kronecker. The *standard basis* for R^I is $\{e_i \in R^I \mid i \in I\}$ where $e_i(j) = \delta_{ij}$. The reader should verify that the standard basis is a linearly independent generating set for R^I. The standard basis for $R^{(n)}$ is the set $\{e_i \mid 1 \leq i \leq n\}$ where e_i is the n-tuple with 1 in coordinate i and 0 elsewhere. A linearly independent generating set for M is called

a *free basis* (or simply *basis*) for M. Lemma 4.2.14 shows that an R-module M is free if and only if M has a free basis.

LEMMA 4.2.14. *Let R be any ring and M a nonzero R-module. Then M is free if and only if M has a free basis. That is, M is free if and only if there exists a subset $X = \{b_i \mid i \in I\} \subseteq M$ which is a linearly independent generating set for M.*

PROOF. Given a linearly independent spanning set $X = \{b_i \mid i \in I\}$ for M, define $\Sigma_X : R^I \to M$ by $\Sigma_X(f) = \sum_{i \in I} f(i)b_i$. By Exercise 4.1.27 (1) and Proposition 4.2.2 (2), Σ_X is a well defined R-module homomorphism and the image is equal to the submodule of M generated by X. Because X generates M and is linearly independent, Σ_X is one-to-one and onto. The converse is left to the reader. □

LEMMA 4.2.15. *Let R be any ring and M a nonzero R-module. Let $X = \{x_1, \ldots, x_n\}$ be a nonempty subset of M. Then X is a linearly independent subset of M if and only if every v in the span of X has a unique representation as a linear combination of the form $v = \alpha_1 x_1 + \cdots + \alpha_n x_n$ where $\alpha_1, \ldots, \alpha_n$ are elements of R.*

PROOF. This follows straight from Definition 4.2.12. □

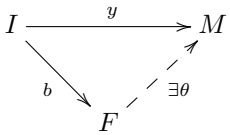
EXAMPLE 4.2.16. We have already seen examples of free modules. Let R be a commutative ring.

(1) The natural mapping $R \to R[x]$ makes the ring of polynomials $R[x]$ into an R-module. In fact, $R[x]$ is a free R-module and the set $\{1, x, x^2, x^3, \ldots\}$ is a free basis.
(2) If G is a group, and $R(G)$ the group ring (see Example 3.1.6), then $R(G)$ is a free R module with free basis $\{g \mid g \in G\}$.

The first part of Theorem 4.2.17 proves that a homomorphism on a free R-module is completely determined by its values on a basis. The second part shows that every R-module is the homomorphic image of a free R-module.

THEOREM 4.2.17. *Let R be a ring and M an R-module.*

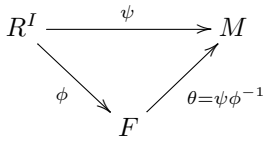
(1) (Universal Mapping Property) Let F be a free R-module and $\{b_i \mid i \in I\}$ a basis for F. For any function $y : I \to M$, there exists a unique R-module homomorphism $\theta : F \to M$ such that $\theta(b_i) = y(i)$ for each $i \in I$ and the diagram



commutes.
(2) There exists a free R-module F and a surjective homomorphism $F \to M$.
(3) M is finitely generated if and only if M is the homomorphic image of a free R-module $R^{(n)}$ for some n.

PROOF. (1): Since $\{b_i \mid i \in I\}$ is a basis for F, there is an isomorphism of R-modules $\phi : R^I \to F$ defined by $\phi(f) = \sum_{i \in I} f(i)b_i$. Suppose $y : I \to M$. Define $\psi : R^I \to M$ by $\psi(f) = \sum_{i \in I} f(i)y(i)$. Using Exercise 4.1.27 (1) and

Proposition 4.2.2 (2), it is routine to check that ψ is an R-module homomorphism and the image of ψ is the R-submodule of M generated by the image of y.



Take θ to be $\psi\phi^{-1}$. This is the existence part of the proof. The uniqueness of θ follows from the fact that ψ is defined in terms of y.

(2) and (3): Let X be a generating set for M and F the free R-module on X. Map the basis elements of R^X to the generators for M. If M is finitely generated, X can be taken to be finite. Conversely, if $R^{(n)} \to M$ is an epimorphism, then a basis for $R^{(n)}$ maps to a generating set for M. □

EXAMPLE 4.2.18. Let A be any nontrivial ring. Let $R = \prod_{i=1}^{\infty} A$ be the ring direct sum of infinitely many copies of A. Let $I = \bigoplus_{i=1}^{\infty} A$ be the direct sum of copies of A as a left A-module with index set $\mathbb{N}$. Then we view I as a proper subset of R. It is routine to check that I is a two-sided ideal in R. Notice that the ideal I is not generated by any finite subset. The ring R is an example of a non-noetherian ring.

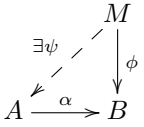
2.3. Projective Modules. Except for the superficial reference in Proposition 4.6.5, the topics and results of this section will not be used in the rest of the text until Section 7.2. Proposition 4.2.19 lists three fundamental properties of a projective module. The definition follows the proposition.

PROPOSITION 4.2.19. *Let R be a ring and M an R-module. The following are equivalent.*

(1) There is a free R-module F and M is isomorphic to a direct summand of F.

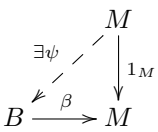
(2) For every epimorphism $\beta : B \to M$ of R-modules there exists an R-module homomorphism $\psi : M \to B$ such that $\beta\psi = 1_M$.

(3) For any diagram of R-module homomorphisms



with α onto, there exists an R-module homomorphism $\psi : M \to A$ such that $\alpha\psi = \phi$.

PROOF. (3) implies (2): Consider the diagram

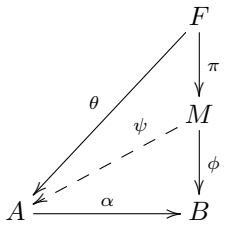


of R-module homomorphisms where $1_M : M \to M$ is the identity map. By (3) there exists $\psi : M \to B$ such that $\beta\psi = 1_M$.

(2) implies (1): By Theorem 4.2.17 there is a free R-module F and an R-module epimorphism $\phi : F \to M$. If M is finitely generated, we can assume F is finitely

generated. By (2) there exists an R-module homomorphism $\psi : M \to F$ such that $\phi\psi = 1_M$. Then M is isomorphic to the image of ψ which is a direct summand of F, by Proposition 4.2.8.

(1) implies (3): Let $F = R^I$ be a free R-module and assume M is a direct summand of F. By Proposition 4.2.8, there is an R-module homomorphism $\pi : F \to M$ such that $\pi(x) = x$ for all $x \in M$. Suppose $\phi : M \to B$ and $\alpha : A \to B$ are R-module homomorphisms and α is onto. Let $X = \{x_i \mid i \in I\}$ be a basis for F and set $Y = \{y_i = \phi(x_i) \mid i \in I\}$. Since α is onto, $\alpha^{-1}(y_i)$ is nonempty for each $i \in I$. By The Axiom of Choice (Proposition 1.3.5), pick $Z = \{z_i \mid i \in I\} \subseteq A$ such that $\alpha(z_i) = y_i$ for each i. By Theorem 4.2.17 there is a unique R-module homomorphism $\theta : F \to A$ such that $\theta(x_i) = z_i$. Since $\phi\pi(x_i) = y_i = \alpha\theta(x_i)$ and $X = \{x_i \mid i \in I\}$ is a generating set for F, we have $\alpha\theta(x) = \phi\pi(x)$ for all $x \in F$. The outer triangle in the diagram



commutes. Define $\psi : M \to A$ to be the restriction of θ to M. If $x \in M$, then $\pi(x) = x$, so $\alpha\psi(x) = \phi(x)$. □

Definition 4.2.20. If R is a ring and M is an R-module satisfying any of the equivalent conditions of Proposition 4.2.19, then we say M is a *projective R-module*.

Example 4.2.21. Here are some examples of modules that are projective and modules that are not projective.

(1) A free R-module of finite rank satisfies Proposition 4.2.19 (1), hence a finitely generated free R-module is a projective R-module. In particular, R is a free R-module of rank 1.
(2) Let R be a ring containing proper two-sided ideals I and J such that $R = I \oplus J$. Then I and J are direct summands of the free R-module R, hence are projective R-modules by Proposition 4.2.19 (1). By Theorem 3.3.4, $I = Re_1$ and $J = Re_2$, where e_1, e_2 is a set of orthogonal idempotents. Then $e_1e_2 = 0$ is a nontrivial dependence relation. This implies $0 \in J$ does not have a unique representation in terms of any generating set for J. Hence I and J are not free R-modules.
(3) Let p and q be distinct prime numbers. By the Chinese Remainder Theorem, Theorem 1.2.11, $\mathbb{Z}/(pq) \cong \mathbb{Z}/(p) \oplus \mathbb{Z}/(q)$. By Part (2), $\mathbb{Z}/(p)$ is a projective $\mathbb{Z}/(pq)$-module which is not a free $\mathbb{Z}/(pq)$-module.
(4) If R is a division ring, in particular if R is a field, then any R-module is an R-vector space, hence is free. This is proved in Corollary 4.3.3 when M is finitely generated. For the general case, see Exercise 4.3.18.
(5) We show in Proposition 4.6.5 below that a finitely generated projective module over a principal ideal domain is free.
(6) For a list of more examples of rings for which projective modules are free see [**9**, Example 6.2.6].

EXAMPLE 4.2.22. Let F be a field and $M_2(F)$ the ring of two-by-two matrices over F. Let $D = \left\{ \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} \mid a, d \in F \right\}$ be the subring of $M_2(F)$ consisting of all diagonal matrices. Let $R = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \mid a, b, d \in F \right\}$ be the ring of Example 3.3.14. In this example we state facts about the rings D, R and $M_2(F)$, leaving most of the proofs to the reader. By Example 4.1.4 (4), $M_2(F)$ is a left R-module and R is a left D-module. As in Example 3.1.13, let e_{ij} be the elementary matrix with 1 in position (i, j) and 0 elsewhere. Then e_{11} and e_{22} are idempotents, $e_{11} + e_{22} = 1$, and by Exercise 3.3.15, D is the internal direct sum of the ideals De_{11} and De_{22}. Let $c_1 = e_{11} + e_{21}$. Then $Dc_1 = C_1 = \left\{ \begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix} \mid a, c \in F \right\}$. As a D-module, C_1 is free of rank 1 with basis $\{c_1\}$. Likewise, if $c_2 = e_{12} + e_{22}$, then $Dc_2 = C_2 = \left\{ \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} \mid b, d \in F \right\}$ is a free D-module of rank 1 with basis $\{c_2\}$. Therefore, $M_2(F)$ is a free D-module of rank 2 with basis c_1, c_2. In a similar way, $M_2(F)$ is the internal direct sum of the D-submodules De_{11}, De_{12}, De_{21}, De_{22}. By Proposition 4.2.19, each submodule De_{ij} is a projective D-module. An argument similar to the one used in Example 4.2.21 (2) can be used to show none of the modules De_{ij} is a free D-module. As a D-module, R is the internal direct sum of the D-submodules De_{11}, De_{12}, De_{22}. This shows R is a projective D-module which is not a free D-module. As an R-module, R is equal to the internal direct sum of the R-submodules Re_{11} and Re_{22}. The R-module $M_2(F)$ is equal to the internal direct sum of the R-submodules Re_{21} and Re_{22}.

2.4. Exercises.

EXERCISE 4.2.23. Let F be a field and $R = M_2(F)$ the ring of two-by-two matrices over F. Let

$$e_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Follow the following outline to prove that the ideals Re_1 and Re_2 are finitely generated projective R-modules but not free R-modules.

(1) Show $e_1^2 = e_1$, $e_2^2 = e_2$, $e_1e_2 = e_2e_1 = 0$. We say e_1 and e_2 are orthogonal idempotents.
(2) Show that Re_1 is the set of all matrices with second column consisting of zeros.
(3) Show that Re_2 is the set of all matrices with first column consisting of zeros.
(4) Show that $R = Re_1 \oplus Re_2$ as R-modules. Show that Re_i is a finitely generated projective R-module for $i = 1, 2$.
(5) For $i = 1, 2$, show that Re_i is not a free R-module.

EXERCISE 4.2.24. Let R be any ring and M a free R-module of rank n with basis $X = \{x_1, \ldots, x_n\}$. Use Theorem 4.2.17 to show that the group of units in the ring $\mathrm{Hom}_R(M, M)$ contains a subgroup isomorphic to S_n, the symmetric group on n letters.

EXERCISE 4.2.25. Let R be a UFD with quotient field K. Let a be an element of R which is not a square in R and let $f = x^2 - a \in R[x]$.

(1) Show that $S = R[x]/(f)$ is an integral domain and $L = K[x]/(f)$ is a field.
(2) Show that S is a free R-module, $\text{Rank}_R(S) = 2$, and $\dim_K(L) = 2$.

EXERCISE 4.2.26. Let R be a commutative ring and $f \in R[x]$ a monic polynomial of degree n. Show that $S = R[x]/(f)$ is a free R-module of rank n and the set $\{1, x, x^2, \ldots, x^{n-1}\}$ is a free basis.

EXERCISE 4.2.27. Let R_1 and R_2 be rings and $R = R_1 \oplus R_2$.

(1) If M_1 and M_2 are left R_1 and R_2-modules respectively, show how to make $M_1 \oplus M_2$ into a left R-module.
(2) If M is a left R-module, show that there are R-submodules M_1 and M_2 of M such that $M = M_1 \oplus M_2$ and for each i, M_i is a left R_i-module.

EXERCISE 4.2.28. Let G be a group and H a subgroup. For any commutative ring R, let $\theta : R(H) \to R(G)$ be the homomorphism of rings induced by the set inclusion man $H \to G$ (see Example 3.2.2 (3)). Show that $R(G)$ is a free $R(H)$-module.

EXERCISE 4.2.29. Let R be a commutative ring and F a free R-module with basis $\{b_1, \ldots, b_n\}$. Prove that if J is a proper ideal of R and $\pi : F \to F/JF$ is the natural homomorphism, then F/JF is a free R/J-module with basis $\{\pi(b_1), \ldots, \pi(b_n)\}$.

3. Vector Spaces

A vector space is a module over a division ring. A submodule of a vector space is called a *subspace*. Elements of a vector space are called a *vectors*. If D is a division ring and V, W are D-vector spaces, then a homomorphism $\phi \in \text{Hom}_D(V, W)$ is called a *linear transformation*. A generating set for V as a D-module is called a *spanning set*.

LEMMA 4.3.1. *Let V be a vector space over a division ring D.*

(1) If v is a nonzero vector in V, then $\{v\}$ is a linearly independent set. Equivalently, if $v \in V - (0)$, $\alpha \in D$ and $\alpha v = 0$, then $\alpha = 0$.
(2) If $X = \{x_1, \ldots, x_n\}$ is a linearly independent set in V and $v \in V - (X)$, then $X \cup \{v\}$ is a linearly independent set in V

PROOF. (1): Assume $\alpha v = 0$ and $\alpha \neq 0$. By Lemma 4.1.5, we have $0 = \alpha^{-1}0 = \alpha^{-1}\alpha v = 1v = v$.
(2): Apply Exercise 4.3.21. □

LEMMA 4.3.2. *Let D be a division ring and V a nonzero finitely generated vector space over D. If $B \subseteq V$, then the following are equivalent.*

(1) B is a basis for V. That is, B is a linearly independent spanning set for V.
(2) B is a spanning set for V and no proper subset of B is a spanning set for V.

PROOF. (1) implies (2): For sake of contradiction, suppose there is a proper subset $B_1 \subsetneq B$ and B_1 is also a spanning set for V. Let $v \in B - B_1$. Since B_1 is a spanning set, there exist $x_1, \ldots, x_n$ in B_1 and $\alpha_1, \ldots, \alpha_n$ in D such that $v = \alpha_1 x_1 + \cdots + \alpha_n x_n$. Then $v - \alpha_1 x_1 - \cdots - \alpha_n x_n = 0$ is a dependency relation in B, which is a contradiction.

(2) implies (1): Assume $B = \{x_1, \ldots, x_n\}$ is a spanning set. We prove that if B is linearly dependent, then there is a proper subset of B that is a spanning set. Since V is nonzero and B is a spanning set, we know B is nonempty. If $0 \in B$, then

the span of B is equal to the span of $B - \{0\}$. From now on we assume each x_i is nonzero. Assume $\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$ where $(\alpha_1, \ldots, \alpha_n)$ is a nonzero vector in $D^{(n)}$. Let k be the largest integer satisfying: $\alpha_k \neq 0$ and if $i > k$, then $\alpha_i = 0$. By Lemma 4.3.1, $k > 1$. Then

$$x_k = -\alpha_k^{-1}(\alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1})$$

is in the subspace spanned by $x_1, \ldots, x_{k-1}$. Therefore, $B - x_k$ is a spanning set for V. □

COROLLARY 4.3.3. *If V is a finitely generated vector space over a division ring D, then V has a basis.*

PROOF. As in Definition 4.2.9, a minimal generating set exists. It follows from Lemma 4.3.2 that a minimal generating set is a basis. □

THEOREM 4.3.4. *Let V be a finitely generated vector space over the division ring D and $B = \{b_1, \ldots, b_n\}$ a basis for V.*

(1) If $Y = \{y_1, \ldots, y_m\}$ is a linearly independent set in V, then $m \leq n$. We can re-order the elements of B such that $\{y_1, \ldots, y_m, b_{m+1}, \ldots, b_n\}$ is a basis for V.
(2) Every basis for V has n elements.

PROOF. Step 1: Write $y_1 = \alpha_1 b_1 + \cdots + \alpha_n b_n$ where each $\alpha_i \in D$. For some i, $\alpha_i \neq 0$. Re-order the basis elements and assume $\alpha_1 \neq 0$. Solve for b_1 to get $b_1 = \alpha_1^{-1} y_1 - \sum_{i=2}^{n} \alpha_1^{-1} \alpha_i b_i$. Therefore $B \subseteq Dy_1 + Db_2 + \cdots + Db_n$, hence $\{y_1, b_2, \ldots, b_n\}$ is a spanning set for V. Suppose $0 = \beta_1 y_1 + \beta_2 b_2 + \cdots + \beta_n b_n$. Then

$$\begin{aligned} 0 &= \beta_1 (\alpha_1 b_1 + \cdots + \alpha_n b_n) + \beta_2 b_2 + \cdots + \beta_n b_n \\ &= \beta_1 \alpha_1 b_1 + (\beta_1 \alpha_2 + \beta_2) b_2 + \cdots + (\beta_1 \alpha_n + \beta_n) b_n, \end{aligned}$$

from which it follows that $\beta_1 \alpha_1 = 0$, hence $\beta_1 = 0$. Now $0 = \beta_2 b_2 + \cdots + \beta_n b_n$ implies $0 = \beta_2 = \cdots = \beta_n$. We have shown that $\{y_1, b_2, \ldots, b_n\}$ is a basis for V.

Step j: Inductively, assume $j \geq 2$ and that $\{y_1, y_2, \ldots, y_{j-1}, b_j, \ldots, b_n\}$ is a basis for V. Write $y_j = \alpha_1 y_1 + \cdots + \alpha_{j-1} y_{j-1} + \alpha_j b_j + \cdots + \alpha_n b_n$ where each $\alpha_i \in D$. Since the set $\{y_1, \ldots, y_j\}$ is linearly independent, for some $i \geq j$, $\alpha_i \neq 0$. Re-order the basis elements and assume $\alpha_j \neq 0$. Solve for b_j and by a procedure similar to that used in Step 1, we see that $\{y_1, \ldots, y_j, b_{j+1}, \ldots, b_n\}$ is a basis for V.

By finite induction, Part (1) is proved. For Part (2), assume $\{c_1, \ldots, c_m\}$ is another basis for V. By applying Part (1) from both directions, it follows that $m \leq n$ and $n \leq m$. □

DEFINITION 4.3.5. Suppose D is a division ring and V is a vector space over D. If V is finitely generated and nonzero, then we define the *dimension* of V, written $\dim_D(V)$, to be the number of elements in a basis for V. If $V = (0)$, set $\dim_D(V) = 0$ and if V is not finitely generated, set $\dim_D(V) = \infty$.

COROLLARY 4.3.6. *Let V be a finitely generated vector space over the division ring D and $X = \{x_1, \ldots, x_n\}$ a spanning set for V. Then the following are true:*

(1) There is a subset of X that is a basis for V.
(2) $\dim_D V \leq n$.

PROOF. Assume V is nonzero. Then X contains a nonzero vector. Without loss of generality assume $x_1 \neq 0$. By Lemma 4.3.1, $\{x_1\}$ is a linearly independent set. Let S be the set of all subsets of X that are linearly independent. Choose $B \in S$ such that B has maximal cardinality. We show B is a spanning set for V. For sake of contradiction, assume $(B) \neq V$. Since X is a spanning set for V, this implies X is not a subset of (B). Assume $x_n \notin (B)$. By Lemma 4.3.1, $B \cup \{x_n\}$ is a linearly independent set, which contradicts the maximality of B. □

DEFINITION 4.3.7. Let R be a commutative ring and M a free R-module with a finite basis $\{b_1, \ldots, b_n\}$. By Exercise 4.3.15, any other basis of M has n elements. We call n the *rank* of M and write $\text{Rank}_R M = n$.

PROPOSITION 4.3.8. *(Free over Free is Free) Let $\theta : R \to S$ be a homomorphism of rings such that S is a finitely generated free R-module. Let M be a finitely generated free S-module. As in Example 4.1.4 (4), we view M as an R-module. In this context, M is a finitely generated free R module. If R and S are both commutative, then* $\text{Rank}_R(M) = \text{Rank}_S(M)\,\text{Rank}_R(S)$.

PROOF. Let $X = \{s_1, \ldots, s_m\}$ be a basis for S over R and $Y = \{y_1, \ldots, y_n\}$ a basis for M over S. Let $Z = \{s_i y_j \mid i = 1, \ldots, m \text{ and } j = 1, \ldots, n\}$. We show Z is basis for M over R.

Step 1: Z is a spanning set for M as an R-module. Let x be an arbitrary element of M. There exist $b_1, \ldots, b_n$ in S such that $x = \sum_{j=1}^{n} b_j y_j$. For each j there exist $a_{1j}, \ldots, a_{mj}$ in R such that $b_j = \sum_{i=1}^{m} a_{ij} s_i$. Taken together, we have

$$x = \sum_{j=1}^{n} b_j y_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} s_i \right) y_j = \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} \left(s_i y_j \right)$$

which shows Z is a spanning set for M over R.

Step 2: Z is linearly independent over R. Assume there is a dependence relation $0 = \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} \left(s_i y_j \right)$ where the elements a_{ij} are in R. Since Y is a basis for M over S, for each j we have $\sum_{i=1}^{m} a_{ij} s_i = 0$ in S. Since X is a basis for S over R, we have $a_{ij} = 0$ for every i and for every j.

The cardinality of Z is equal to $|Z| = |X||Y|$, which proves the rank formula. □

3.1. Exercises.

EXERCISE 4.3.9. Suppose D is a division ring, V is a finite dimensional vector space over D, and W is a subspace of V. Prove:

(1) W is finite dimensional and $\dim_D(W) \leq \dim_D(V)$.
(2) There is a subspace U of V such that $V = U \oplus W$ is an internal direct sum and $\dim_D(V) = \dim_D(U) + \dim_D(W)$.
(3) $\dim_D(V/W) = \dim_D(V) - \dim_D(W)$.

EXERCISE 4.3.10. Suppose $\phi \in \text{Hom}_D(V, W)$, where V and W are vector spaces over the division ring D. Prove:

(1) If V is finite dimensional, then the kernel of ϕ is finite dimensional and the image of ϕ is finite dimensional.
(2) If the kernel of ϕ is finite dimensional and the image of ϕ is finite dimensional, then V is finite dimensional.

EXERCISE 4.3.11. (The Rank-Nullity Theorem) Suppose $\phi \in \mathrm{Hom}_k(V, W)$, where V and W are vector spaces over the field k. The *rank* of ϕ, written $\mathrm{Rank}\,(\phi)$, is defined to be the dimension of the image of ϕ. The *nullity* of ϕ, written $\mathrm{Nullity}\,(\phi)$, is defined to be the dimension of the kernel of ϕ. Prove that if V is finite dimensional, then $\dim_k(V) = \mathrm{Rank}(\phi) + \mathrm{Nullity}(\phi)$.

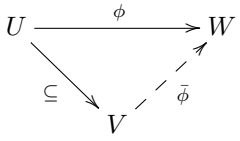
EXERCISE 4.3.12. Suppose $\phi \in \mathrm{Hom}_D(V, V)$, where V is a finite dimensional vector space over the division ring D. Prove that the following are equivalent:

(1) ϕ is invertible.
(2) $\mathrm{Nullity}(\phi) = 0$.
(3) $\mathrm{Rank}(\phi) = \dim_D(V)$.

EXERCISE 4.3.13. Let V be a finite dimensional vector space over a division ring D. Let ϕ, ψ be elements of $\mathrm{Hom}_D(V, V)$. Prove:

(1) $\mathrm{Rank}(\phi\psi) \leq \mathrm{Rank}(\phi)$.
(2) $\mathrm{Rank}(\phi\psi) \leq \mathrm{Rank}(\psi)$.
(3) $\mathrm{Rank}(\phi\psi) \leq \min(\mathrm{Rank}(\phi), \mathrm{Rank}(\psi))$.
(4) If ϕ is invertible, $\mathrm{Rank}(\phi\psi) = \mathrm{Rank}(\psi\phi) = \mathrm{Rank}(\psi)$.

EXERCISE 4.3.14. Let D be a division ring and V and W finitely generated vector spaces over D. Suppose U is a subspace of V and $\phi : U \to W$ an element of $\mathrm{Hom}_D(U, W)$. Show that there exists an element $\bar{\phi}$ of $\mathrm{Hom}_D(V, W)$ such that the diagram



commutes. That is, $\bar{\phi}$ is an extension of ϕ.

EXERCISE 4.3.15. Let R be a commutative ring and F a finitely generated free R-module. Show that any two bases for F have the same number of elements.

EXERCISE 4.3.16. Let V be a finitely generated vector space over a division ring D. Let $X \subseteq V$ be a spanning set for V. Show that there is a subset of X that is a basis for V. Do not assume X is finite.

EXERCISE 4.3.17. Let V be a vector space over a division ring D. Suppose there exists a positive number n such that every linearly independent subset of V has cardinality less than or equal to n. Show that V is finitely generated and $\dim_D(V) \leq n$.

EXERCISE 4.3.18. Let D be a division ring and V a nonzero vector space over D. As in Definition 4.2.13, a subset $X \subseteq V$ is a basis for V if X is a linearly independent spanning set for V. Apply Zorn's Lemma (Proposition 1.3.3) to prove the following.

(1) Every linearly independent subset of V is contained in a basis for V.
(2) If $S \subseteq V$ is a spanning set for V, then S contains a basis for V.
(3) V is a free D-module.

EXERCISE 4.3.19. Let D be a division ring and V a vector space over D. Let A and B be finite dimensional subspaces of V. Prove:

(1) $A + B$ is finite dimensional.
(2) $\dim_D(A + B) = \dim_D(A) + \dim_D(B) - \dim_D(A \cap B)$.

EXERCISE 4.3.20. Let F be a field and $M_2(F)$ the ring of two-by-two matrices over F. Let $R = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \mid a, b, d \in F \right\}$ be the subring of $M_2(F)$ of Example 3.3.14. Let $A = \left\{ \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \mid a, b \in F \right\}$, which is a subring of R. We identify F with the subring of scalar matrices $\left\{ \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \mid a \in F \right\}$. There is a chain of subrings: $F \subseteq A \subseteq R \subseteq M_2(F)$. By Example 4.1.4, any ring is a left module over any of its subrings. Prove:

(1) $\dim_F(A) = 2$, $\dim_F(R) = 3$, $\dim_F(M_2(F)) = 4$.
(2) R is not a free A-module.
(3) $M_2(F)$ is a free A-module.
(4) Show that A is a local ring.
(5) Show that R is not a projective A-module.

EXERCISE 4.3.21. Let R be a ring, M an R-module, and $X = \{x_1, \ldots, x_n\}$ a linearly independent subset of M. Let $\eta : M \to M/(X)$ be the natural map. Let $Y = \{y_1, \ldots, y_m\}$ be another subset of M. Show that if $\{\eta(y_1), \ldots, \eta(y_m)\}$ is a linearly independent subset of $M/(X)$, then $X \cup Y$ is a linearly independent subset of M.

EXERCISE 4.3.22. State and prove a version of Exercise 4.3.21 in which X and Y are not necessarily finite subsets of V.

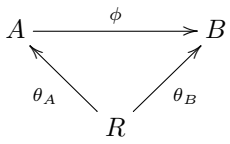
EXERCISE 4.3.23. Let R be a ring. Show:

(1) If F_1 and F_2 are free R-modules, then $F_1 \oplus F_2$ is a free R-module.
(2) If for each i, F_i is a finitely generated R-module, then $F_1 \oplus F_2$ is a finitely generated R-module.
(3) If R is commutative and for each i, F_i is a finitely generated free R-module of rank n_i, then $F_1 \oplus F_2$ is a finitely generated free R-module of rank $n_1 + n_2$.

4. Algebras

DEFINITION 4.4.1. Let R be a commutative ring, A a ring and $\theta : R \to A$ a homomorphism of rings such that $\theta(R)$ is a subring of the center of A. Then we say A is an *R-algebra* and θ is the *structure homomorphism*. If A and B are two R-algebras, then an *R-algebra homomorphism from A to B* is a function $\phi : A \to B$ satisfying:

(1) ϕ is a ring homomorphism from A to B, and
(2) if $\theta_A : R \to A$ and $\theta_B : R \to B$ are the structure homomorphisms for A and B respectively, then the diagram



commutes. That is, $\phi\theta_A = \theta_B$.

An *R-algebra isomorphism from A to B* is a homomorphism $\phi : A \to B$ that is one-to-one and onto. An *R-algebra automorphism of A* is a homomorphism from A to A that is one-to-one and onto. The set of all R-algebra automorphisms is a group and is denoted $\mathrm{Aut}_R(A)$.

If k is a field and A is a k-algebra, then the structure homomorphism is necessarily one-to-one, so it is convenient to identify k as a subring of the center of A. In this case, A is a left k-vector space by virtue of the multiplication and addition operations on A.

EXAMPLE 4.4.2. Let R be a commutative ring, and M an R-module. By Example 4.1.17, the left regular representation $\lambda : R \to \mathrm{Hom}_R(M, M)$ is a ring homomorphism that makes the endomorphism ring $\mathrm{Hom}_R(M, M)$ into an R-algebra.

EXAMPLE 4.4.3. Important examples of algebras over a field are listed here.

(1) If F and k are fields and k is a subfield of F, then we say F/k is an *extension of fields*. In this case F is a k-algebra.
(2) The ring of polynomials $k[x]$ is a k-algebra where we identify k with the constant polynomials. Because $1, x, x^2, \ldots$ are linearly independent over k, $\dim_k(k[x]) = \infty$.
(3) Let $q \in k[x]$ be a polynomial of degree $n > 0$. In Lemma 4.4.5 below we prove that the quotient ring $k[x]/(q)$ is a commutative k-algebra of dimension n.

EXAMPLE 4.4.4. Let R be a commutative ring and $A = M_n(R)$ the ring of n-by-n matrices over R. By Example 3.1.13, the center of A is the subring of scalar matrices. Therefore, A is an R-algebra. Let e_{ij} be the elementary matrix with 1 in position (i, j) and 0 elsewhere. In Lemma 4.5.2 the reader is asked to prove that the set $\{e_{ij} \mid 1 \leq i \leq n, 1 \leq j \leq n\}$ is a free R-basis. That is, A is a free R-module of rank n^2.

Let k be a field, x an indeterminate, and q a polynomial in $k[x]$. The principal ideal generated by q is $(q) = \{fq \mid f \in k[x]\}$, which is equal to the set of all polynomials that are divisible by q. By 3.2.14, $k[x]/(q)$ is a commutative ring.

LEMMA 4.4.5. *In the above context, the following are true.*

(1) $k[x]/(q)$ is a commutative k-algebra.
(2) $k[x]/(q)$ is a k-vector space.
(3) $\dim_k\left(k[x]/(q)\right) = \begin{cases} \infty & \text{if } q = 0 \\ \deg q & \text{if } q \neq 0. \end{cases}$
(4) If $(q) \neq k[x]$, then $k[x]/(q)$ is a k-algebra.
(5) $k[x]/(q)$ is a field if and only if q is irreducible.

PROOF. Since k is a subring of $k[x]$, $k[x]$ is a k-algebra. If $q = 0$, then $k[x]/(q) = k[x]$ is not finite dimensional (Example 4.4.3 (3)). If $q \neq 0$ and $n = \deg q$, then by Exercise 4.2.26, $k[x]/(q)$ is a k-vector space and $\{[1], [x], \ldots, [x^{n-1}]\}$ is a k-basis for $k[x]/(q)$. Since $k[x]$ is a PID, $k[x]/(q)$ is a field if and only if q is irreducible, by Corollary 3.4.14 and Exercise 3.4.30. If $\deg q = 0$, then $k[x]/(q)$ is the trivial ring and is not a k-algebra. Otherwise, $k[x]/(q)$ is a k-algebra. □

DEFINITION 4.4.6. Let A be a k-algebra. If $X \subseteq A$, then by $k[X]$ we denote the k-subalgebra of A generated by k and X. Thus $k[X]$ is the smallest subring of A that contains both k and X.

DEFINITION 4.4.7. Let k be a field, A a k-algebra, and α an element of A. If there is a nonzero polynomial $f \in k[x]$ and $f(\alpha) = 0$, then we say α is *algebraic over* k. Otherwise we say α is *transcendental over* k. We say A is *algebraic over* k if every $\alpha \in A$ is algebraic over k.

Theorem 4.4.8 will play a fundamental role in the study of algebraic elements in a k-algebra.

THEOREM 4.4.8. *Let k be a field, A a k-algebra, and $\alpha \in A - \{0\}$. There is a k-algebra homomorphism $\tau : k[x] \to A$ satisfying the following.*

(1) $\tau(x) = \alpha$.
(2) The kernel of τ is $I(\alpha) = \{p \in k[x] \mid p(\alpha) = 0\}$. There is a polynomial $f \in k[x]$ such that $I(\alpha)$ is equal to the principal ideal (f) generated by f.
(3) The image of τ is $k[\alpha]$, the subalgebra of A generated by k and α.
(4) α is transcendental over k if and only if $I(\alpha) = (0)$.
(5) α is algebraic over k if and only if $I(\alpha) \neq (0)$. In this case, $\deg f > 0$, $\dim_k k[\alpha] = \deg f$, f can be taken to be monic, and if $p \in I(\alpha)$, then $f \mid p$.
(6) $k[\alpha] \cong k[x]/(f)$.
(7) $k[\alpha]$ is a commutative principal ideal ring.

The polynomial f is called the minimal polynomial of α *and is denoted* $\min.\,\mathrm{poly}_k(\alpha)$. *If α is algebraic and f is taken to be monic, then f is uniquely determined by α.*

PROOF. Given $\alpha \in A$, the evaluation homomorphism $\tau : k[x] \to A$, is a k-algebra homomorphism determined by $x \mapsto \alpha$ (Theorem 3.6.2). Since $k[x]$ is a principal ideal domain (Corollary 3.6.5), there exists a polynomial $f \in k[x]$ which generates the kernel of τ. The image of τ is denoted $k[\alpha]$. By Exercise 3.6.36, $k[\alpha]$ is a commutative principal ideal ring and is the smallest subring of A containing k and α. By Proposition 3.2.15, $k[\alpha] \cong k[x]/(f)$. By Definition 4.4.7, α is transcendental if and only if $I(\alpha) = (0)$. In this case, τ is one-to-one and $k[\alpha] \cong k[x]$. If $I(\alpha) \neq (0)$, then $\deg f \geq 1$ and f is unique up to associates in $k[x]$. Hence if f is taken to be monic, then f is unique. Let $f = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be the minimal polynomial of α, where $n \geq 1$. Exercise 4.2.26 says $k[\alpha]$ is a k-vector space of dimension n spanned by $1, \alpha, \ldots, \alpha^{n-1}$. □

EXAMPLE 4.4.9. If x is an indeterminate, and $k(x)$ is the field of rational functions over k, then $k[x] \to k(x)$ is one-to-one (Lemma 3.5.1). Hence x is transcendental over k.

EXAMPLE 4.4.10. Let k be a field and A a k-algebra. Let $\alpha \in A - \{0\}$ be an algebraic element and let $f = \min.\,\mathrm{poly}_k(\alpha)$.

(1) Assume α is nilpotent. Then there exists $n > 0$ such that $\alpha^n = 0$ and $\alpha^{n-1} \neq 0$. Let $p = x^n$. Since $p(\alpha) = 0$, by Theorem 4.4.8 (5), f divides p. Since n is the minimal power that annihilates α, it is clear that $f = x^n$.
(2) Assume α is a nontrivial idempotent. Let $p = x^2 - x$. Then $p(\alpha) = 0$ and by Theorem 4.4.8 (5), f is a divisor of p. Since $\alpha \neq 1$ and $\alpha \neq 0$, this implies $f = x^2 - x$.

COROLLARY 4.4.11. *If k is a field and A is a finite dimensional k-algebra, then A is algebraic over k. If $\alpha \in A$ and $\dim_k(A) = n$, then the degree of* $\min.\,\mathrm{poly}_k(\alpha)$ *is less than or equal to n.*

PROOF. Let $\alpha \in A$, and $\dim_k(A) = n$. It follows from Theorem 4.3.4 that the set $\{u^n, u^{n-1}, \ldots, u, 1\}$ is linearly dependent. A dependence relation $0 = a_n u^n + a_{n-1}u^{n-1} + \cdots + a_1 u + a_0$ over k shows that u is algebraic over k. □

EXAMPLE 4.4.12. We use the notation of Section 1.4. The field of complex numbers, $\mathbb{C}$, is a vector space of dimension two over the field of real numbers, $\mathbb{R}$. Let $\zeta = a+bi$ be a nonreal complex number. Since ζ is nonreal, by Corollary 4.4.11, $\min.\,\text{poly}_{\mathbb{R}}(\zeta)$ has degree two. The complex conjugate of ζ is $\bar{\zeta} = a - bi$. Then $\zeta + \bar{\zeta} = 2a$ and $\zeta\bar{\zeta} = a^2 + b^2$ are real numbers. Let x be an indeterminate. The polynomial $f(x) = (x-\zeta)(x-\bar{\zeta}) = x^2 - 2ax - (a^2+b^2)$ has coefficients in $\mathbb{R}$. The roots of $f(x)$ are ζ and $\bar{\zeta}$. This implies that $f(x)$ is equal to the minimal polynomial of ζ over $\mathbb{R}$.

COROLLARY 4.4.13. *Let k be a field and A a k-algebra. If $\alpha \in A$ is algebraic over k, then $k[\alpha]$ is algebraic over k.*

PROOF. By Theorem 4.4.8 (5), $k[\alpha]$ is finite dimensional over k. □

COROLLARY 4.4.14. *Let k be a field, A a k-algebra, and u an element of A that is algebraic over k. Then u is an invertible element of A if and only if* $\min.\,\text{poly}_k(u)$ *has a nonzero constant term.*

PROOF. Let $f(x) = \min.\,\text{poly}_k(u) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. If $u \in k$, then $f(x) = x - u$ and in this case the result holds. Assume $n \geq 2$. We have $f(u) = u^n + a_{n-1}u^{n-1} + \cdots + a_1 u + a_0 = 0$. Solving for a_0 and factoring, we get

$$-a_0 = u(u^{n-1} + a_{n-1}u^{n-2} + \cdots + a_1). \tag{4.1}$$

Assume $a_0 = 0$ and for sake of contradiction assume u is invertible. Then multiplying (4.1) by u^{-1} on both sides we get $u^{n-1} + a_{n-1}u^{n-2} + \cdots + a_1 = 0$, which contradicts the definition of the minimal polynomial of u in Theorem 4.4.8. Conversely, assume $a_0 \neq 0$. Multiplying both sides of (4.1) by $-a_0^{-1}$, we get

$$1 = u(-a_0^{-1})(u^{n-1} + a_{n-1}u^{n-2} + \cdots + a_1)$$

which shows u is invertible in A. □

THEOREM 4.4.15. *Let k be a field and A a k-algebra which is algebraic over k.*

(1) If $u \in A$ and u is not a zero divisor, then u is invertible.
(2) If A is a domain (that is, A has no zero divisors), then A is a division ring.

PROOF. The proof is by contraposition. Assume A contains a nonzero element u which is not invertible. We show that u is a zero divisor in A. Let $f = \min.\,\text{poly}_k(u) \in k[x]$. By Corollary 4.4.14, u is invertible if and only if $f(0) \in k - (0)$. Assume $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x$ has zero constant term. By Eq. (4.1),

$$0 = u(u^{n-1} + a_{n-1}u^{n-2} + \cdots + a_1).$$

Since the minimum polynomial for u has degree n, we know $u^{n-1} + a_{n-1}u^{n-2} + \cdots + a_1 \neq 0$. This shows u is a zero divisor in A. □

As an application of the previous results, we compute in Proposition 4.4.16 below the minimal polynomial for a 2-by-2 matrix α over a field k. If $A = M_2(k)$ is the ring of all 2-by-2 matrices over k, then by Example 4.4.4, A is a k-algebra and $\dim_k(A) = 4$.

PROPOSITION 4.4.16. *Let k be a field, $A = M_2(k)$, the ring of all 2-by-2 matrices over k, and $\alpha = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ an element of A. Let λ be an indeterminate. Define the polynomial $f(\lambda) = \lambda^2 - T\lambda + D$, where $T = a + d$ is the trace of α (see Exercise 6.3.26) and $D = ad - bc$ is the determinant of α (see Example 2.1.21). Then the following are true.*

(1) (Cayley-Hamilton) $f(\alpha) = 0$.
(2) If $\alpha \notin k$, then $f(\lambda) = \text{min. poly}_k(\alpha)$.
(3) The following are equivalent.
  (a) α is invertible.
  (b) $D \neq 0$.
  (c) α is not a zero divisor.
(4) If $\alpha \notin k$, then α is nilpotent if and only if $T = D = 0$.
(5) If $\alpha \notin k$, then α is idempotent if and only if $T = 1$ and $D = 0$.

PROOF. By direct computation, $f(\lambda) = \lambda^2 - (a+d)\lambda + (ad - bc) = (a - \lambda)(d - \lambda) - bc$ in $k[\lambda]$. Therefore,

$$\begin{aligned} f(\alpha) &= (a - \alpha)(d - \alpha) - bc \\ &= \begin{bmatrix} 0 & -b \\ -c & a - d \end{bmatrix} \begin{bmatrix} d - a & -b \\ -c & 0 \end{bmatrix} - \begin{bmatrix} bc & 0 \\ 0 & bc \end{bmatrix} \\ &= \begin{bmatrix} bc & 0 \\ 0 & bc \end{bmatrix} - \begin{bmatrix} bc & 0 \\ 0 & bc \end{bmatrix} \\ &= 0 \end{aligned}$$

which is (1). By Theorem 4.4.8, $\text{min. poly}_k(\alpha)$ is a factor of the quadratic polynomial $f(\lambda)$. But $\alpha \in k$ if and only if $\text{min. poly}_k(\alpha)$ has degree one. This proves (2). For (3), apply Corollary 4.4.14. If $\alpha \notin k$, then by (1) we see that α is nilpotent if and only if $f(\lambda) = \lambda^2$ and α is idempotent if and only if $f(\lambda) = \lambda^2 - \lambda$. This proves (4) and (5). □

4.1. Exercises.

EXERCISE 4.4.17. Let R be a commutative ring and A an R-algebra. Suppose $\alpha \in A$ is a root of the polynomial $p \in R[x]$. Prove:

(1) If B is another R-algebra and $\phi : A \to B$ is an R-algebra homomorphism, then $\phi(\alpha)$ is a root of p.
(2) If u is a unit in A, then $u^{-1}\alpha u$ is a root of p.

EXERCISE 4.4.18. (Universal Mapping Property) Let R be a commutative ring, G a finite group, and $R(G)$ the group ring (see Example 3.1.6). Let A be an R-algebra and $h : G \to A^*$ a homomorphism from G to the group of units of A. Show that there is a unique homomorphism of R-algebras $\phi : R(G) \to A$ such that diagram

$$\begin{array}{ccc} G & & \\ {\scriptstyle\subseteq}\downarrow & \searrow^{h} & \\ R(G) & \xrightarrow{\phi} & A \end{array}$$

commutes. Show that the same result holds if G is a group that is not necessarily finite.

EXERCISE 4.4.19. Let R be a commutative ring and M a finitely generated free R-module of rank n. Using Exercises 4.2.24 and 4.4.18, show that there exists an R-algebra homomorphism $\phi : R(S_n) \to \mathrm{Hom}_R(M, M)$ from the group ring to the ring of endomorphisms. Show that in general ϕ is not one-to-one.

EXERCISE 4.4.20. Let k be a field, A a k-algebra, and $\alpha \in A$. Assume α is algebraic over k.

(1) Let $\theta : A \to B$ be a k-algebra homomorphism.
   (a) Show that $\theta(\alpha)$ is algebraic over k and the minimal polynomial of $\theta(\alpha)$ divides the minimal polynomial of α.
   (b) If θ is one-to-one, show that the minimal polynomial of $\theta(\alpha)$ is equal to the minimal polynomial of α.
(2) If u is an invertible element in A, show that $u^{-1}\alpha u$ is algebraic over k and the minimal polynomial of α is equal to the minimal polynomial of $u^{-1}\alpha u$.

EXERCISE 4.4.21. Let F be a field, $M_2(F)$ the ring of 2-by-2 matrices over F, and $L = \left\{ \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \mid a, b, c \in F \right\}$ the subring of all lower triangular matrices (see Example 3.2.12). Follow the following outline to completely classify all ideals in L.

(1) Show that L has exactly three proper two-sided ideals, namely:
   (a) $\left\{ \begin{pmatrix} 0 & 0 \\ b & 0 \end{pmatrix} \mid b \in F \right\} = L \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$.
   (b) $\left\{ \begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix} \mid a, b \in F \right\} = L \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.
   (c) $\left\{ \begin{pmatrix} 0 & 0 \\ b & c \end{pmatrix} \mid b, c \in F \right\} = L \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + L \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} L$.
(2) Let $\alpha \in F$ and $I_\alpha = \left\{ \begin{pmatrix} 0 & 0 \\ k\alpha & k \end{pmatrix} \mid k \in F \right\} = L \begin{pmatrix} 0 & 0 \\ \alpha & 1 \end{pmatrix}$.
   (a) Show that I_α is a left ideal in L but not a right ideal.
   (b) Show that every left ideal of L that is not a right ideal is equal to I_α for some $\alpha \in F$.
(3) Let $\alpha \in F$ and $J_\alpha = \left\{ \begin{pmatrix} k & 0 \\ k\alpha & 0 \end{pmatrix} \mid k \in F \right\} = \begin{pmatrix} 1 & 0 \\ \alpha & 0 \end{pmatrix} L$.
   (a) Show that J_α is a right ideal in L but not a left ideal.
   (b) Show that every right ideal of L that is not a left ideal is equal to J_α for some $\alpha \in F$.

EXERCISE 4.4.22. Let k be a field and A a finite dimensional k-algebra. If $\alpha \in A$ and A is noncommutative, prove:

(1) $k[\alpha] \neq A$.
(2) If $f = \text{min. poly}_k(\alpha)$, then $1 \leq \deg f < \dim_k(A)$.

EXERCISE 4.4.23. Let $k = \mathbb{Z}/2$ be the field of order two. Let A be the ring $M_2(k)$. Show:

(1) $k[x]/(x^2)$ is isomorphic to $k[x]/(x^2 + 1)$ as k-algebras.
(2) A has exactly seven subrings of order four, namely:
   (a) There is one subfield of order four, and it is isomorphic to $k[x]/(x^2 + x + 1)$. Call it F.
   (b) There are three subrings, each of which is isomorphic to $k[x]/(x^2+1)$. Call them N_1, N_2, N_3.

(c) There are three subrings, each of which is isomorphic to $k[x]/(x^2+x)$. Call them E_1, E_2, E_3.

(3) A has exactly three subrings of order eight. Call them U, L, M.

(4) Let $u \in \mathrm{GL}_2(\mathbb{Z}/2)$. As in Example 3.2.2 (2), let $\sigma_u : A \to A$ be the inner automorphism of A defined by u.
   (a) $\sigma_u : F \to F$. That is, σ_u restricts to a k-automorphism of F. The action by σ_u is nontrivial if and only if u has order two.
   (b) σ_u acts as a permutation of $\{N_1, N_2, N_3\}$. $\mathrm{GL}_n(\mathbb{Z}/2)$ acts as a group of permutations on $\{N_1, N_2, N_3\}$.
   (c) σ_u acts as a permutation of $\{E_1, E_2, E_3\}$. $\mathrm{GL}_n(\mathbb{Z}/2)$ acts as a group of permutations on $\{E_1, E_2, E_3\}$.
   (d) σ_u acts as a permutation of $\{U, L, M\}$. $\mathrm{GL}_n(\mathbb{Z}/2)$ acts as a group of permutations on $\{U, L, M\}$.

EXERCISE 4.4.24. Let p be a prime number. Consider the quotient ring $R = \left(\mathbb{Z}/p^2\right)[x]/(px, x^2 - p)$. In the following, cosets in the ring R are written without brackets or any extra adornment. Prove:

(1) R has order p^3 and characteristic p^2.
(2) Denote by (x) the principal ideal generated by x. Then (x) has order p^2 and (x) is equal to $\mathrm{Rad}_R(0)$, the nil radical of R.
(3) R is a local ring, the maximal ideal is (x).
(4) The ideals (x^2) and (p) are equal and they both have order p.
(5) Find the invariants (Theorem 2.8.7) of the abelian groups $(R, +)$ and $(Rx, +)$.
(6) If $p = 2$, find the invariants of the group of units R^*.

EXERCISE 4.4.25. Let k be a field and $k[x, y]$ the polynomial ring over k in two variables. Consider the quotient ring $R = k[x, y]/(x^2, xy, y^2)$. In the following, cosets in the ring R are written without brackets or any extra adornment. Prove:

(1) The only prime ideal in R is $\mathfrak{m} = Rx + Ry$.
(2) $\mathfrak{m}$ is equal to $\mathrm{Rad}_R(0)$, the nil radical of R.
(3) R is a local ring with maximal ideal $\mathfrak{m}$.
(4) $\dim_k(R) = 3$.
(5) R is isomorphic to the subring $\left\{ \begin{bmatrix} \alpha & 0 & 0 \\ \beta & \alpha & 0 \\ \gamma & 0 & \alpha \end{bmatrix} \mid \alpha, \beta, \gamma \in k \right\}$ of $M_3(k)$.

EXERCISE 4.4.26. Let k be a field and $k[x, y]$ the polynomial ring over k in two variables. Consider the quotient ring $R = k[x, y]/(x^2 - y, xy, y^2)$. In the following, cosets in the ring R are written without brackets or any extra adornment. Prove:

(1) The only prime ideal in R is $\mathfrak{m} = Rx + Ry$.
(2) $\mathfrak{m}$ is equal to $\mathrm{Rad}_R(0)$, the nil radical of R.
(3) R is a local ring with maximal ideal $\mathfrak{m}$.
(4) $\dim_k(R) = 3$.
(5) R is isomorphic to the subring $\left\{ \begin{bmatrix} \alpha & 0 & 0 \\ \beta & \alpha & 0 \\ \gamma & \beta & \alpha \end{bmatrix} \mid \alpha, \beta, \gamma \in k \right\}$ of $M_3(k)$.

EXERCISE 4.4.27. Let k be a field and A a finite dimensional k-algebra. Let $\mathfrak{m}_1, \ldots, \mathfrak{m}_n$ be maximal ideals in A such that $\mathfrak{m}_i \neq \mathfrak{m}_j$, if $i \neq j$. Prove that $\dim_k(A) \geq n$.

5. Matrix Theory

If R is a ring and M and N are finitely generated free modules over R, then we show that any R-module homomorphism $\phi : M \to N$ can be represented as a matrix. The matrix representation of ϕ depends on a choice of bases for M and N. When R is commutative, matrix multiplication corresponds to composition of functions. This is proved in Proposition 4.5.5. If P is a free R-module and $\psi : N \to P$ an R-module homomorphism, then upon fixing bases for M, N, and P, the matrix of the composite $\psi\phi$ is the product of the matrices for ψ and ϕ. If M is free of rank n, then we show in Proposition 4.5.7 that there is an isomorphism of R-algebras $\mathrm{Hom}_R(M, M) \cong M_n(R)$. The analogs for the above results when R is a noncommutative ring are also proven.

5.1. The Matrix of a Linear Transformation.

Definition 4.5.1. Let R be any ring and m, n positive integers. By $M_{nm}(R)$ we denote the set of all n-by-m matrices over R. If $m = n$, then we simply write $M_n(R)$ instead of $M_{nn}(R)$. Addition of matrices is coordinate-wise $(\alpha_{ij}) + (\beta_{ij}) = (\alpha_{ij} + \beta_{ij})$. We can multiply by elements of R from the left $r(\alpha_{ij}) = (r\alpha_{ij})$. If $(\alpha_{ij}) \in M_{nm}(R)$ and $(\beta_{jk}) \in M_{mp}(R)$, then the matrix product is defined by $(\alpha_{ij})(\beta_{jk}) = (\gamma_{ik}) \in M_{np}(R)$, where $\gamma_{ik} = \sum_{j=1}^{m} \alpha_{ij}\beta_{jk}$. In Corollary 4.5.6 below we prove that $M_n(R)$ is a ring that contains R as a subring. If R is a commutative ring, $M_n(R)$ is an R-algebra. If e_{ij} is the matrix with 1 in position (i, j) and 0 elsewhere, then e_{ij} is called an elementary matrix (see Section 1.5).

Lemma 4.5.2. *For a ring R, the set $M_{nm}(R)$ of n-by-m matrices over R is a free R-module. The set $\{e_{ij} \mid 1 \le i \le n, 1 \le j \le m\}$ of elementary matrices is a free basis with nm elements.*

Proof. See Definition 4.2.11 for the definition of free module. As an R-module, $M_{nm}(R)$ can be identified with $R^{(nm)}$ and the set $\{e_{ij} \mid 1 \le i \le n, 1 \le j \le m\}$ of elementary matrices can be identified with the standard basis. The rest of the proof is left to the reader. □

Definition 4.5.3. Let R be any ring, M a free R-module of rank m and N a free R-module of rank n. Let $X = \{x_1, \ldots, x_m\}$ be a basis for M and $Y = \{y_1, \ldots, y_n\}$ a basis for N. Given $\phi \in \mathrm{Hom}_R(M, N)$, ϕ maps $x_j \in X$ to a linear combination of Y. That is,

$$\phi(x_j) = \sum_{i=1}^{n} \phi_{ij} y_i$$

where the elements ϕ_{ij} are in R. The *matrix of ϕ with respect to the bases X and Y* is defined to be $M(\phi, X, Y) = (\phi_{ij})$, which is a matrix in $M_{nm}(R)$.

Proposition 4.5.4. *Let R be any ring. If M is a free R-module of rank m, and N is a free R-module of rank n, then there is a $\mathbb{Z}$-module isomorphism $\mathrm{Hom}_R(M, N) \cong M_{nm}(R)$. If R is a commutative ring, then this is an R-module isomorphism and $\mathrm{Hom}_R(M, N)$ is a free R-module of rank mn.*

Proof. Let $X = \{x_1, \ldots, x_m\}$ be a basis for M and $Y = \{y_1, \ldots, y_n\}$ a basis for N. The assignment $\phi \mapsto M(\phi, X, Y)$ defines a $\mathbb{Z}$-module homomorphism

$$M(\cdot, X, Y) : \mathrm{Hom}_R(M, N) \to M_{nm}(R).$$

Conversely, if $(\alpha_{ij}) \in M_{nm}(R)$, define α in $\mathrm{Hom}_R(M,N)$ by

$$\alpha(x_j) = \sum_{i=1}^{n} \alpha_{ij} y_i.$$

The rest is left to the reader. □

PROPOSITION 4.5.5. *Let R be any ring. Let M, N, and P denote free R-modules, each of finite rank. Let X, Y and Z be bases for M, N, and P respectively. Let $\phi \in \mathrm{Hom}_R(M,N)$ and $\psi \in \mathrm{Hom}_R(N,P)$. If the matrices $M(\psi, Y, Z)$ and $M(\phi, X, Y)$ are treated as having entries from the ring R^o, the opposite ring of R, then*

$$M(\psi\phi, X, Z) = M(\psi, Y, Z)M(\phi, X, Y).$$

PROOF. The opposite ring R^o is defined as in Definition 3.1.8. Let $X = \{x_1, \ldots, x_m\}$, $Y = \{y_1, \ldots, y_n\}$, and $Z = \{z_1, \ldots, z_p\}$. Let $M(\phi, X, Y) = (\phi_{ij})$, $M(\psi, Y, Z) = (\psi_{ij})$. It follows from

$$\psi\phi(x_j) = \psi\left(\sum_{i=1}^{n} \phi_{ij} y_i\right) = \sum_{i=1}^{n} \phi_{ij} \sum_{k=1}^{p} \psi_{ki} z_k = \sum_{k=1}^{p} \left(\sum_{i=1}^{n} \phi_{ij}\psi_{ki}\right) z_k$$

that $M(\psi\phi, X, Z) = (\gamma_{kj})$, where $\gamma_{kj} = \sum_{i=1}^{n} \phi_{ij}\psi_{ki}$. Computing the product of the two matrices over R^o, we get $M(\psi, Y, Z)M(\phi, X, Y) = (\tau_{kj})$, where

$$\tau_{kj} = \sum_{i=1}^{n} \psi_{ki} * \phi_{ij} = \sum_{i=1}^{n} \phi_{ij}\psi_{ki}.$$

□

COROLLARY 4.5.6. *Let R be any ring. With the binary operations defined in Definition 4.5.1, $M_n(R)$ is a ring with identity element $I_n = e_{11} + \cdots + e_{nn}$. The set $R \cdot I_n$ of all scalar matrices in $M_n(R)$ is a subring which is isomorphic to R. The center of the ring $M_n(R)$ is equal to the center of the subring $R \cdot I_n$. If R is commutative, the matrix ring $M_n(R)$ is an R-algebra and the center of $M_n(R)$ is equal to $R \cdot I_n$.*

PROOF. Use Proposition 4.5.5 to show that matrix multiplication is associative. If R is commutative, as shown in Example 3.1.13, the center of $M_n(R)$ is equal to the set of scalar matrices. The same proof can be used to prove that the center of $M_n(R)$ is equal to the center of the subring $R \cdot I_n$. The rest is left to the reader. □

PROPOSITION 4.5.7. *Let R be any ring. If M is a free R-module of rank n, then there is an isomorphism of rings $\mathrm{Hom}_R(M,M) \cong M_n(R^o)$. If R is commutative, this is an isomorphism of R-algebras.*

PROOF. Pick a basis for M. The map of Proposition 4.5.4 defines an isomorphism of abelian groups. It is multiplicative by Proposition 4.5.5. □

DEFINITION 4.5.8. Let R be a commutative ring and $n \geq 1$. If A, B are matrices in $M_n(R)$ and P is an invertible matrix in $M_n(R)$ such that $A = P^{-1}BP$, then we say A and B are *similar*. The reader should verify that this defines an equivalence relation on $M_n(R)$. By Exercise 4.4.20, if R is a field, then two similar matrices in $M_n(R)$ have the same minimal polynomial.

PROPOSITION 4.5.9. *Let R be a commutative ring and M a free R-module of rank n. Let X and Y be two bases for M. If $\phi \in \mathrm{Hom}_R(M,M)$, then the matrix $M(\phi,X,X)$ of ϕ with respect to X and the matrix $M(\phi,Y,Y)$ of ϕ with respect to Y are similar. In fact, if $1 \in \mathrm{Hom}_R(M,M)$ is the identity map, then $M(1,X,Y)^{-1} = M(1,Y,X)$ and $M(\phi,X,X) = M(1,Y,X)M(\phi,Y,Y)M(1,X,Y)$.*

PROOF. Let $I \in M_n(R)$ be the identity matrix. It follows from Proposition 4.5.5 that $I = M(1,X,X) = M(1,Y,Y)$, $M(1,X,Y)M(1,Y,X) = I$, and $M(1,Y,X)M(1,X,Y) = I$. Also

$$\begin{aligned} M(\phi,X,Y) &= M(1,X,Y)M(\phi,X,X) \\ &= M(\phi,Y,Y)M(1,X,Y). \end{aligned}$$

□

EXAMPLE 4.5.10. Let R be a commutative ring and $A \in M_{mn}(R)$. Elements of R^n can be viewed as n-by-1 column matrices in M_{n1}. As in Proposition 4.5.4, multiplication by A from the left defines an element in $\mathrm{Hom}_R(R^n,R^m)$. In particular, if k is a field and $A \in M_n(k)$, then left multiplication by A defines a linear transformation from k^n to k^n. We define the rank of A and the nullity of A as in Exercise 4.3.11. Define the *column space* of A to be the subspace of k^n spanned by the columns of A. The rank of A is seen to be the dimension of the column space of A.

5.2. The Transpose of a Matrix and the Dual of a Module. If $A = (a_{ij})$ is a matrix in $M_{nm}(R)$, then the *transpose* of A is the matrix in $M_{mn}(R)$ whose entry in row i column j is equal to a_{ji}. The transpose of A is denoted A^T.

DEFINITION 4.5.11. Let R be a commutative ring. Let M be a left R-module. The *dual of M* is defined to be $M^* = \mathrm{Hom}_R(M,R)$. We turn M^* into a right R-module by the action $(fr)(x) = (f(x))r$, for $r \in R$, $f \in M^*$, $x \in M$. The reader should verify that this is a well defined right R-module action on M^*. If N is another left R-module, and $\psi \in \mathrm{Hom}_R(M,N)$, define $\psi^* : N^* \to M^*$ by the rule $\psi^*(f) = f \circ \psi$, for any $f \in N^*$.

LEMMA 4.5.12. *Let R be a commutative ring. Let M and N be left R-modules. If $\psi : M \to N$ is a homomorphism of left R-modules, then $\psi^* : N^* \to M^*$ is a homomorphism of right R-modules. If L is another R-module, and $\phi \in \mathrm{Hom}_R(L,M)$, then $(\psi\phi)^* = \phi^*\psi^*$.*

PROOF. Let $f,g \in N^*$ and $a \in R$. The reader should verify that $\psi^*(f+g) = \psi^*(f) + \psi^*(g)$. If $x \in M$, then

$$(\psi^*(fa))(x) = (fa)(\psi(x)) = (f(\psi(x)))a = (\psi^*(f)(x))a = (\psi^*(f)a)(x).$$

Lastly, $\phi^*\psi^*(f) = (\psi\phi)^*(f)$. □

DEFINITION 4.5.13. Let R be a commutative ring. Let M be a left R-module which is free of finite rank. If $B = \{v_1,\ldots,v_n\}$ is a basis for M, then define $v_1^*,\ldots,v_n^*$ in M^* by the rules

$$v_i^*(v_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

PROPOSITION 4.5.14. *Let R be a commutative ring. If M is a free left R-module with basis $B = \{v_1, \ldots, v_n\}$, then M^* is a free right R-module with basis $B^* = \{v_1^*, \ldots, v_n^*\}$.*

PROOF. By Proposition 4.5.4, M^* is isomorphic to $M_{1n}(R)$ as $\mathbb{Z}$-modules. Under this isomorphism, v_i^* is mapped to the row matrix e_{1i} which has 1 in position i and zeros elsewhere. This is therefore a homomorphism of right R-modules. □

THEOREM 4.5.15. *Let R be a commutative ring. Let M and N be free R-modules, each of finite rank. Let X be a basis for M, and Y a basis for N. Let X^* and Y^* be the corresponding bases for M^* and N^*. Given $\phi \in \mathrm{Hom}_R(M, N)$,*

$$M(\phi^*, Y^*, X^*) = M(\phi, X, Y)^T.$$

That is, the matrix of ϕ^* with respect to Y^* and X^* is the transpose of the matrix of ϕ with respect to X and Y.

PROOF. Let $X = \{u_1, \ldots, u_m\}$ and $Y = \{v_1, \ldots, v_n\}$. Let $M(\phi, X, Y) = (\phi_{ij})$. Consider $\phi^*(v_l^*)(u_j) = v_l^*(\phi(u_j)) = v_l^*(\sum_{i=1}^n \phi_{ij} v_i) = \phi_{lj}$. Now consider $(\sum_{i=1}^m \phi_{li} u_i^*)(u_j) = \phi_{lj}$. Therefore, $\phi^*(v_l^*) = \sum_{i=1}^m \phi_{li} u_i^*$ as elements of $M^* = \mathrm{Hom}_R(M, R)$ because they agree on a basis of M. This also shows that column l of the matrix $M(\phi^*, Y^*, X^*)$ is the transpose of $(\phi_{l1}, \phi_{l2}, \ldots, \phi_{lm})$, which is row l of $M(\phi, X, Y)$ □

DEFINITION 4.5.16. If k is a field, the space $V^{**} = \mathrm{Hom}_k(V^*, k)$ is called the *double dual* of V. Given $v \in V$, let $\varphi_v : V^* \to k$ be the "evaluation at v" map. That is, if $f \in V^*$, then $\varphi_v(f) = f(v)$. The reader should verify that φ_v is an element of V^{**}, and that the assignment $v \mapsto \varphi_v$ is a homomorphism of k-vector spaces $V \to V^{**}$.

THEOREM 4.5.17. *Let V be a finitely generated vector space over a field k. The map $V \to V^{**}$ which sends a vector $v \in V$ to φ_v is a vector space isomorphism.*

PROOF. Let v be a nonzero vector in V. By Theorem 4.3.4, we can extend $\{v\}$ to a basis for V, say $B = \{v, v_2, \ldots, v_n\}$. Define $f \in V^*$ to be the projection mapping onto the v-coordinate. Then $f(v) = 1$, and $f(v_i) = 0$ for $2 \le i \le n$. Then $\varphi_v(f) = f(v) = 1$. This proves $V \to V^{**}$ is one-to-one. If V is finite dimensional, then $V \to V^{**}$ is onto since $\dim_k(V) = \dim_k(V^{**})$. □

Theorem 4.5.17 extends to finitely generated projective modules over any ring (see [**9**, Exercise 6.5.22]).

THEOREM 4.5.18. *Let D be a field and V and W finitely generated D-vector spaces. Let $\phi \in \mathrm{Hom}_D(V, W)$. Let $\phi^* : W^* \to V^*$ be the associated homomorphism of right D-vector spaces.*

(1) If ϕ is one-to-one, then ϕ^* is onto.
(2) If ϕ is onto, then ϕ^* is one-to-one.
(3) The rank of ϕ is equal to the rank of ϕ^*.

PROOF. (1): Assume ϕ is one-to-one. Let $f : V \to D$ be in V^*. By Exercise 4.3.14 there is $\bar{f} : W \to D$ in W^* such that $f = \bar{f}\phi = \phi^*(\bar{f})$.

(2): Assume ϕ is onto. A typical element of W is of the form $w = \phi(v)$, for some $v \in V$. Assume $g \in W^*$ and $g\phi = 0$. Then $g(w) = g(\phi(v)) = 0$.

(3): Let $n = \dim_D(V)$. By Proposition 4.5.4, $\dim_D(V^*) = n$. Let $U = \ker\phi$. Let $\psi : U \to V$ be the inclusion map. By (1), ψ^* is onto. Then $\text{Rank}(\psi^*) = \dim(U^*) = \dim(U) = \text{Nullity}(\phi) = n - \text{Rank}\,\phi$. By Lemma 4.5.12, $\text{im}\,\phi^* \subseteq \ker\psi^*$. We prove the reverse inclusion. Suppose $f \in V^*$ and $\psi^*(f) = f\psi = 0$. Then f factors through $V/\ker\phi = \text{im}\,\phi$. There is $\bar{f} : \text{im}\,\phi \to D$ such that $f = \bar{f}\phi$. By Exercise 4.3.14, $\bar{f}$ extends to W, so f is in the image of ϕ^*. This proves $\text{Rank}\,\phi^* = \text{Nullity}\,\psi^* = n - \text{Rank}\,\psi^* = \text{Rank}\,\phi$. □

Let k be a field and $A \in M_{nm}(k)$. The *column rank of* A is defined to be the dimension of the subspace of k^n spanned by the column vectors of A. The *row rank of* A is defined to be the dimension of the subspace of k^m spanned by the row vectors of A.

COROLLARY 4.5.19. *Let k be a field and $A \in M_{nm}(k)$. The row rank of A is equal to the column rank of A.*

PROOF. As in Proposition 4.5.4, define α in $\text{Hom}_k(k^m, k^n)$ to be "left multiplication by A". Let α^* be the associated map on dual spaces. By Theorem 4.5.15 the matrix of α^* is A^T. The column rank of A is equal to $\text{Rank}\,\alpha$ which is equal to $\text{Rank}\,\alpha^*$, by Theorem 4.5.18. But $\text{Rank}\,\alpha^*$ is equal to the column rank of A^T, which is the row rank of A. □

5.3. Exercises.

EXERCISE 4.5.20. Let k be a field and V a finite dimensional vector space over k. Show that $\text{Hom}_k(V, V)$ is a commutative ring if and only if $\dim_k(V) \leq 1$.

EXERCISE 4.5.21. Suppose $\phi \in \text{Hom}_D(V, V)$, where V is a finite dimensional vector space over the field D. Prove:

(1) There is a chain of subspaces $\ker(\phi) \subseteq \ker(\phi^2) \subseteq \ker(\phi^3) \subseteq \cdots$.
(2) There is a chain of subspaces $\phi(V) \supseteq \phi^2(V) \supseteq \phi^3(V) \supseteq \cdots$.
(3) The kernel of $\phi : \phi(V) \to \phi^2(V)$ is equal to $\ker(\phi) \cap \phi(V)$. More generally, if $m \geq 1$, the kernel of $\phi^m : \phi^m(V) \to \phi^{2m}(V)$ is equal to $\ker(\phi^m) \cap \phi^m(V)$.
(4) If $m \geq 1$ and $\phi^m(V) = \phi^{m+1}(V)$, then $\phi^m(V) = \phi^{m+i}(V)$ for all $i \geq 1$.
(5) If $n = \dim_D(V)$, then there exists m such that $1 \leq m \leq n$ and $\phi^m(V) = \phi^{m+1}(V)$.
(6) If $n = \dim_D(V)$, then there exists m such that $1 \leq m \leq n$ and $\ker(\phi^m) \cap \phi^m(V) = (0)$.

EXERCISE 4.5.22. Let R be a commutative ring. Let $A \in M_{nm}(R)$ and $B, C \in M_{ml}(R)$. Prove:

(1) $(A^T)^T = A$.
(2) $(B + C)^T = B^T + C^T$.
(3) $(AB)^T = B^T A^T$.

EXERCISE 4.5.23. If R is a commutative ring, show that the mapping $M_n(R) \to M_n(R)^o$ defined by $A \mapsto A^T$ is an isomorphism of R-algebras.

EXERCISE 4.5.24. If R is any ring, show that the mapping $M_n(R) \to M_n(R^o)^o$ defined by $A \mapsto A^T$ is an isomorphism of rings. Using the Morita Theorems, a very general version of this is proved in [**9**, Corollary 6.9.3].

EXERCISE 4.5.25. Let R be any ring, M and N finitely generated R-modules, and $\phi \in \mathrm{Hom}_R(M, N)$. Show that there exist positive integers m and n, epimorphisms $f : R^{(m)} \to M$, $g : R^{(n)} \to N$, and $\theta \in \mathrm{Hom}_R(R^{(m)}, R^{(n)})$ such that the diagram

$$\begin{array}{ccc} R^{(m)} & \xrightarrow{\theta} & R^{(n)} \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\ M & \xrightarrow{\phi} & N \end{array}$$

commutes. Therefore, given generators for M and N, ϕ can be represented as a matrix.

EXERCISE 4.5.26. Let R be a commutative ring and $n > 1$. If $A = (a_{ij})$ is a matrix in $M_n(R)$, then the transpose of A with respect to the anti-diagonal is the matrix $A^\tau = (b_{ij})$, where $b_{ij} = a_{n+1-j,n+1-i}$. This notation and terminology are from [**13**]. Let σ be the permutation in S_n which reverses the ordered list $1, 2, \ldots, n$. In the notation of Example 2.1.15, σ can be represented an array of two rows:

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & \ldots & n \\ n & n-1 & n-2 & \ldots & 1 \end{bmatrix}.$$

As in Section 1.5, let P_σ be the permutation matrix in $M_n(R)$ associated to σ.

(1) Show that $A^\tau = P_\sigma A^T P_\sigma$.
(2) Show that the mapping $M_n(R) \to M_n(R)^o$ defined by $A \mapsto A^\tau$ is an isomorphism of R-algebras.

EXERCISE 4.5.27. Let R be a commutative ring and $n \geq 2$. As in Example 3.1.12, we denote by $L = \{(r_{ij}) \mid r_{ij} = 0 \text{ if } i < j\}$ the subring of $M_n(R)$ consisting of all lower triangular matrices. Show that the isomorphism of Exercise 4.5.26 (2) maps L isomorphically onto L^o.

EXERCISE 4.5.28. Let k be a field, A a k-algebra, and M a left A-module. Assume that A is a simple ring and the dimension of M as a k-vector space is $\dim_k(M) = n$.

(1) Prove that $\dim_k(A) \leq n^2$.
(2) If I is a left ideal in $M_n(k)$, prove that $\dim_k(I) \geq n$.

EXERCISE 4.5.29. (The Dual Basis Lemma) Let R be a ring and M a finitely generated R-module. Let $\{x_1, \ldots, x_n\}$ be a generating set for M over R. Suppose $\{f_1, \ldots, f_n\}$ is a subset of $M^* = \mathrm{Hom}_R(M, R)$ such that for every $x \in M$, $x = \sum_{i=1}^n f_i(x)x_i$. Then $\{(x_i, f_i) \mid 1 \leq i \leq n\}$ is called a *dual basis* for M. Prove that a finitely generated R-module M is projective if and only if M has a dual basis.

6. Finitely Generated Modules over a Principal Ideal Domain

Throughout this section, R is a principal ideal domain, or PID for short. The main result of this section is the proof that a finitely generated R-module M is the internal direct sum of cyclic submodules. The cyclic submodules forming this direct sum decomposition are not unique, nevertheless we show that it is always possible to find cyclic submodules of M such that the resulting factorization of M is in so-called canonical form. In fact we derive two such canonical form decompositions of M, called the Invariant Factor Form and the Elementary Divisor Form. When

the direct sum decomposition of the R module M is in one of the canonical forms, the number of direct summands and the order of each summand are unique.

In Theorem 2.8.6 we proved that a finite abelian group is equal to the internal direct sum of cyclic subgroups. Every abelian group is a $\mathbb{Z}$-module and cyclic subgroups correspond to cyclic submodules. Therefore, we have already proved that every finite $\mathbb{Z}$-module is equal to the internal direct sum of cyclic submodules.

The canonical form decomposition of a finitely generated R-module M consists of two parts, which are sometimes called the free part and the torsion part. The module M is the internal direct sum of two submodules, $M = F \oplus M_t$. The submodule F is a free R-module, the submodule M_t consists of so-called torsion elements and is not free. This decomposition is not unique. The torsion subgroup M_t is uniquely defined, the free part F is not. The lead-up to the proofs of the Basis Theorems is therefore split into two parts. First the important properties of finitely generated free R-modules are proved. This allows us to show that M is equal to an internal direct sum $M = F \oplus M_t$. The second phase is focused on deriving the canonical form for the torsion subgroup M_t. This step is the counterpart for R-modules of the Basis Theorem for finite abelian groups. In Theorem 2.8.7, the invariants of the finite abelian group G are a special case of the elementary divisors that appear in Theorem 4.6.12 below.

6.1. Finitely Generated Free Modules. Recall that if F is a free module over R with a finite basis, then by Exercise 4.3.15, any two bases for F have the same number of elements, namely $\text{Rank}_R(F)$.

THEOREM 4.6.1. *Let R be a principal ideal domain and F a free R-module with a finite basis. If M is a submodule of F, then M is a free R-module and* $\text{Rank}_R(M) \leq \text{Rank}_R(F)$.

PROOF. Assume $M \neq (0)$. Let $\{b_1, \ldots, b_n\}$ be a free basis for F over R. Let Rb_1 be the submodule of F spanned by b_1. By Theorem 4.2.17, there is a homomorphism $\Sigma : R \to F$ defined by the assignment $1 \mapsto b_1$. This induces an isomorphism of R-modules $R \cong Rb_1$. By Exercise 4.6.17, the nonzero submodules of R are free R-modules of rank 1. If $M_1 = M \cap Rb_1$, then M_1 is an R-submodule of the free R-module Rb_1. Then M_1 is equal to the image under θ of an ideal $I = Ra$ for some $a \in R$. In other words, $M_1 = Rab_1$. If $a = 0$, then $M_1 = 0$. Otherwise, there is an isomorphism $R \cong M_1$ given by the assignment $1 \mapsto ab_1$. For each j in the range $1 \leq j \leq n$ define $M_j = M \cap (Rb_1 + \cdots + Rb_j)$. The proof is by induction on n. If $n = 1$, we are done. Assume $1 \leq j < n$ and that M_j is a free R-module on j or fewer generators. We now prove that $M_{j+1} = M \cap (Rb_1 + \cdots + Rb_{j+1})$ is free of rank $j + 1$ or less. Assume $M \neq M_j$. By Lemma 4.2.14, $Rb_1 + \cdots + Rb_{j+1}$ is a free R-module. Let $\pi : Rb_1 + \cdots + Rb_{j+1} \to Rb_{j+1}$ be the projection onto the last summand. The image of M_{j+1} under π is a submodule of the free R-module Rb_{j+1}. By the basis step, $\pi(M_{j+1}) = Rab_{j+1}$ for some $a \in R$. If $a = 0$, then $M_j = M_{j+1}$, and we stop. If $a \neq 0$, then Rab_{j+1} is free of rank 1. Let $b \in M_{j+1}$ such that $\pi(b) = ab_{j+1}$. By Theorem 4.2.17, there is a homomorphism $\sigma : Rab_{j+1} \to M_{j+1}$ defined by $ab_{j+1} \mapsto b$. So $\pi\sigma = 1$ and $\sigma\pi : M_{j+1} \to Rab_{j+1}$ such that $\sigma\pi(rb) = rb$ for all $r \in R$. By Proposition 4.2.8, $M_{j+1} = M_j \oplus Rb$. By Exercise 4.3.23 and Mathematical Induction, we are done. □

PROPOSITION 4.6.2. *Let R be a ring and M an R-module. The following are equivalent.*

(1) *Every submodule of M is finitely generated.*
(2) *M satisfies the ascending chain condition (ACC) on submodules. That is, given a chain of submodules $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots \subseteq I_n \subseteq \cdots$, there exists $N \geq 1$ such that $I_N = I_{N+1} = \cdots$.*
(3) *M satisfies the maximum condition on submodules. That is, any nonempty family of submodules in M has a supremum.*

PROOF. (1) implies (2): The reader should verify that the union $U = \bigcup_{i \geq 1} I_i$ is a submodule of M. By (1) there is a finite generating set $\{x_1, \ldots, x_m\}$ for U. There is some $N \geq 1$ such that I_N contains every x_i. Then $U \subseteq Rx_1 + \cdots + Rx_m \subseteq I_N \subseteq I_{N+1} \subseteq \cdots \subseteq U$. This proves that the ACC is satisfied by M.

(2) and (3) are equivalent by Exercise 1.3.7.

(3) implies (1): Let A be a submodule of M and let $\mathfrak{S}$ be the set of all finitely generated submodules of A. Let B be a maximal member of $\mathfrak{S}$. If $B = A$, then we are done. Otherwise, let x be an arbitrary element of $A - B$. So $B + Rx$ is a finitely generated submodule of A which properly contains B. This contradicts the maximality of B. □

COROLLARY 4.6.3. *Let R be a principal ideal domain and M a finitely generated R-module. Then every submodule of M is finitely generated, M satisfies the ACC on submodules, and M satisfies the maximum condition on submodules.*

PROOF. By Theorem 4.2.17, there is is a surjection $\psi : R^{(n)} \to M$. If N is a submodule of M, then $\psi^{-1}(N)$ is a submodule of $R^{(n)}$, which by Theorem 4.6.1 is free of rank n or less. This shows N is the homomorphic image of a finitely generated free R-module. By Theorem 4.2.17, N is finitely generated. The rest follows from Proposition 4.6.2. □

DEFINITION 4.6.4. Let R be an integral domain and M an R-module. If $x \in M$, then we say x is a *torsion element* of M in case there exists a nonzero $r \in R$ such that $rx = 0$. If every element of M is torsion, then we say M is torsion. Since R is an integral domain, by Exercise 4.6.16 the set of all torsion elements in M is a submodule of M, which is denoted M_t. If $M_t = 0$, then we say M is *torsion free.*

PROPOSITION 4.6.5. *Let R be a PID and M a finitely generated R-module. The following are equivalent.*
(1) *M is torsion free.*
(2) *M is free.*
(3) *M is projective.*

PROOF. (2) implies (1): Is left to the reader.

(2) is equivalent to (3): By Proposition 4.2.19 (1), if M is a free R-module, then M is projective. Also, if M is projective, then M is isomorphic to a direct summand of a finitely generated free R-module. By Theorem 4.6.1, this implies M is free.

(1) implies (2): Assume $M = Ry_1 + \cdots + Ry_n$. Let $\{v_1, \ldots, v_m\}$ be a linearly independent subset of $\{y_1, \ldots, y_n\}$ such that m is maximal. If $N = Rv_1 + \cdots + Rv_m$, then N is a free R-module. By the choice of $\{v_1, \ldots, v_m\}$, for each $j = 1, \ldots, n$, there is a nontrivial dependence relation

$$c_j y_j = \sum_{i=1}^{m} a_{ij} v_i$$

such that $c_j, a_{1j}, \ldots, a_{mj}$ are in R and $c_j \neq 0$. Since R is a domain, if $c = c_1 c_2 \cdots c_n$, then $c \neq 0$. For each j, c factors into $c = c_j d_j$. Consider the submodule $cM = \{cx \mid x \in M\}$ of M. A typical element of $cM = c(Ry_1 + \cdots + Ry_n)$ looks like

$$\begin{aligned} cx &= c\sum_{j=1}^{n} r_j y_j \\ &= \sum_{j=1}^{n} r_j c y_j \\ &= \sum_{j=1}^{n} r_j d_j c_j y_j \\ &= \sum_{j=1}^{n} \left(r_j d_j \sum_{i=1}^{m} a_{ij} y_j \right) \end{aligned}$$

which is in N. Since N is free of rank m, Theorem 4.6.1 says that cM is free of rank no more than m. Because c is nonzero and M is torsion free, the assignment $x \mapsto cx$ defines an isomorphism $M \to cM$. □

COROLLARY 4.6.6. *Let R be a PID and M a finitely generated R-module. Let M_t denote the submodule consisting of all torsion elements of M. Then there is a finitely generated free submodule F such that M is the internal direct sum $M = F \oplus M_t$. The rank of F is uniquely determined by M.*

PROOF. By Exercise 4.6.16, M/M_t is torsion free, and Proposition 4.6.5 implies M/M_t is a finitely generated free R-module. Consider the exact sequence

$$0 \to M_t \to M \xrightarrow{\eta} M/M_t \to 0$$

of R-modules. By Proposition 4.2.19, there is an R-module homomorphism $\psi : M/M_t \to M$ such that $\eta\psi = 1$. Let $F = \operatorname{im}\psi$. By Proposition 4.2.8, M is the internal direct sum $M = F \oplus M_t$. The rank of F is equal to the rank of M/M_t, which is uniquely determined by M. □

6.2. Finitely Generated Torsion Modules. The results of this section are similar to those we proved for finite abelian groups in Section 2.8.2. For example, a finite abelian group is the direct sum of its Sylow subgroups. For torsion R-modules, Theorem 4.6.10 is the corresponding theorem. In a finite abelian group an element of maximal order generates a cyclic direct summand. In Lemma 4.6.11 we show that a torsion element of maximal order in a finitely generated torsion R-module M generates a cyclic submodule that is a direct summand.

DEFINITION 4.6.7. Let R be a principal ideal domain, M an R-module and $x \in M$. The cyclic submodule generated by x is Rx. Define $\theta_x : R \to M$ by $\theta(r) = rx$. Then θ_x is an R-module homomorphism. Denote by I_x the kernel of θ_x. That is,

$$I_x = \{r \in R \mid rx = 0\}$$

which is an ideal in R, hence is principal. So $I_x = Ra$ and up to associates in R, a is uniquely determined by x. We call a the *order of* x. The image of θ_x is Rx and by Theorem 4.1.12, $Rx \cong R/(I_x) \cong R/Ra$.

DEFINITION 4.6.8. Let R be a unique factorization domain and M a finitely generated R-module. By Example 4.1.17, the left regular representation $\lambda : R \to \mathrm{Hom}_R(M, M)$ is a homomorphism of rings that maps $r \in R$ to $\ell_r : M \to M$, where $\ell_r(x) = rx$ is "left multiplication by r". Let π be a prime element in R and n a positive integer. The kernel of ℓ_{π^n} is contained in the kernel of $\ell_{\pi^{n+1}}$. Therefore the union

$$\begin{aligned} M(\pi) &= \bigcup_{n>0} \ker\left(\ell_{\pi^n}\right) \\ &= \{x \in M \mid \text{there exists } n > 0 \text{ such that } \pi^n x = 0\} \end{aligned}$$

is a submodule of M.

LEMMA 4.6.9. *Assume R is a PID, π is a prime in R, and M is an R-module.*

(1) If $(\pi, q) = 1$, then $\ell_q : M(\pi) \to M(\pi)$ is one-to-one.
(2) If $M \cong R/(\pi^e R)$ is a cyclic R-module of order π^e, where $e \geq 1$, then
  (a) πM is cyclic of order π^{e-1}, and
  (b) $M/\pi M$ is a vector space of dimension one over the field $R/\pi R$.

PROOF. (1): Suppose $x \in \ker(\ell_q)$ and $\pi^n x = 0$. Then $(\pi^n, q) = 1$, so there exist $a, b \in R$ such that $1 = qa + \pi^n b$. Therefore, $x = aqx + b\pi^n x = 0$.
(2): Is left to the reader. □

A finite abelian group decomposes into an internal direct sum of its Sylow subgroups (Theorem 2.8.7 (1)). Theorem 4.6.10 is the counterpart of this theorem for a finitely generated torsion R-module.

THEOREM 4.6.10. *If R is a PID and M a finitely generated torsion R-module, then there exists a finite set $\pi_1, \ldots, \pi_n$ of primes in R such that $M = M(\pi_1) \oplus \cdots \oplus M(\pi_n)$.*

PROOF. Let $x \in M$ and let a be the order of x. Since M is torsion, $a \neq 0$. Since R is a UFD, we factor a into primes, $a = \pi_1^{e_1} \cdots \pi_m^{e_m}$ where each $e_i > 0$. For each π_i, let $q_i = a/\pi_i^{e_i}$. Then $Rq_1 + \cdots + Rq_m = 1$. There exist $s_1, \ldots s_m \in R$ such that $1 = s_1 q_1 + \cdots + s_m q_m$. This means $x = s_1 q_1 x + \cdots + s_m q_m x$. Note that $\pi_i^{e_i} q_i x = ax = 0$ so $q_i x \in M(\pi_i)$. This proves $x \in M(\pi_1) + \cdots + M(\pi_m)$. Since M is finitely generated, if necessary we add more π_i so that $\pi_1, \ldots, \pi_n$ is a finite set of distinct primes and for every x in a finite generating set, x is in the sum $M(\pi_1) + \cdots + M(\pi_n)$.

To show that the sum is direct, we apply Proposition 4.2.4 (4). Assume $n > 1$, fix $1 \leq k \leq n$ and consider $x \in M(\pi_k) \bigcap \Big(\sum_{i \neq k} M(\pi_i) \Big)$. Because x is in the sum $\sum_{i \neq k} M(\pi_i)$, for t sufficiently large, if $s = \prod_{i \neq k} \pi_i^t$, then $sx = 0$. But $(s, \pi_k) = 1$ and Lemma 4.6.9 says $\ell_s : M(\pi_k) \to M(\pi_k)$ is one-to-one. This implies $x = 0$. □

In Theorem 2.8.6 we proved that a finite abelian group G is an internal direct sum of cyclic subgroups. The key step of the proof showed that an element of maximal order generates a subgroup which is a direct summand of G. Lemma 4.6.11 is the generalization of this result to R-modules.

LEMMA 4.6.11. *Let R be a PID and M a finitely generated torsion R-module such that the annihilator of M in R is $R\pi^n$, where π is a prime and $n > 0$. Then there exists an element $a \in M$ of order π^n such that the cyclic submodule Ra is a direct summand of M.*

PROOF. There exists $a \in M$ such that $\pi^n a = 0$ and $\pi^{n-1}a \neq 0$. If $Ra = M$, then we are done. Otherwise continue.

Step 1: We show that there exists $b \in M$ such that $\pi b = 0$, $b \neq 0$ and $Ra \cap Rb = 0$. Start with any element c in $M - Ra$. Pick the least positive integer j such that $\pi^j c \in Ra$. Then $1 \leq j \leq n$. Let $\pi^j c = r_1 a$. Since R is factorial, write $r_1 = r\pi^k$ and assume $(r, \pi) = 1$. Now $0 = \pi^n c = \pi^{n-j}\pi^j c = r\pi^{n-j}\pi^k a$. By Lemma 4.6.9, $\pi^{n-j+k}a = 0$. Since the order of a is π^n, this implies $0 \leq -j + k$, so we have $1 \leq j \leq k$. Set $b = \pi^{j-1}c - r\pi^{k-1}a$. Since $\pi^{j-1}c \notin Ra$ but $r\pi^{k-1}a \in Ra$ we know $b \neq 0$. Also, $\pi b = \pi^j c - r\pi^k a = 0$. Now check that $Ra \cap Rb = 0$. Assume otherwise. Then for some $s \in R$ we have $sb \in Ra$ and $sb \neq 0$. Since the order of b is π, this implies $(s, \pi^n) = 1$. For some $x, y \in R$ we can write $xs + y\pi^n = 1$. In this case $b = xsb + y\pi^n b = xsb \in Ra$ which is a contradiction.

Step 2: We show that Ra is a direct summand of M. Let $\mathcal{S}$ be the set of all submodules S of M such that $S \cap Ra = 0$. By Step 1, $\mathcal{S}$ is nonempty. By Corollary 4.6.3, $\mathcal{S}$ has a maximal member, call it C. To complete the proof, it suffices to show $C + Ra = M$, which is equivalent to showing M/C is generated by $a + C$. For contradiction's sake, assume $M \neq C + Ra$. Since $C \cap Ra = 0$, the order of $a + C$ in M/C is π^n. By Step 1, there exists $b + C \in M/C$ such that $b + C \neq C$, $\pi b + C = C$, and $(Ra + C) \cap (Rb + C) = C$. It suffices to show that $Rb + C$ is in $\mathcal{S}$. Suppose $x \in (Rb + C) \cap Ra$. We can write x in two ways, $x = rb + c \in Rb + C$, and $x = sa \in Ra$. Hence $rb \equiv sa \pmod{C}$. The choice of b implies $\pi \mid r$. Then $x = (r/\pi)\pi b + c$ is an element of C. So $x \in C \cap Ra = 0$, which says $x = 0$. This says $Rb + C$ is in $\mathcal{S}$, which contradicts the choice of C. □

6.3. The Basis Theorems. We state and prove two forms of the Basis Theorem for a finitely generated module over a principal ideal domain. The elementary divisors that appear in Theorem 4.6.12 below are a generalization of the group invariants of Theorem 2.8.7. Theorem 4.6.12 will play a central role in Section 6.2.2 when we define the Jordan canonical form of a linear transformation.

THEOREM 4.6.12. *(Basis Theorem – Elementary Divisor Form) Let R be a PID and M a finitely generated R-module. In the notation established above, the following are true.*

(1) $M = F \oplus M_t$, where F is a free submodule of finite rank. The rank of F is uniquely determined by M.
(2) $M_t = \bigoplus_\pi M(\pi)$ where π runs through a finite set of primes in R.
(3) For each prime π such that $M(\pi) \neq 0$, there exists a basis $\{a_1, \ldots, a_m\}$ such that $M(\pi) = Ra_1 \oplus Ra_2 \oplus \cdots \oplus Ra_m$ where the order of a_i is equal to π^{e_i} and $e_1 \geq e_2 \geq \cdots \geq e_m$.
(4) M_t is uniquely determined by the primes π that occur in (2) and the integers e_i that occur in (3).

The prime powers π^{e_i} that occur are called the elementary divisors *of M.*

PROOF. (1): This is Corollary 4.6.6.

(2): This is Theorem 4.6.10.

(3): Since $M(\pi)$ is a submodule of M, it follows from Corollary 4.6.3 that $M(\pi)$ is finitely generated. Let $x_1, \ldots, x_n$ be a generating set. Let k be the maximum integer in the set $\{k_i \mid x_i \text{ has order } \pi^{k_i}\}$. Then $\pi^k M(\pi) = 0$. There exists $e_1 > 0$ such that $\pi^{e_1} M(\pi) = 0$ and $\pi^{e_1 - 1} M(\pi) \neq 0$. By Lemma 4.6.11, there exists

$a_1 \in M(\pi)$ such that a_1 has order π^{e_1} and $M = Ra_1 \oplus C_1$. If $C_1 \neq 0$, then we can apply Lemma 4.6.11 and find $a_2 \in C_1$ such that a_2 has order π^{e_2}, where $e_1 \geq e_2 \geq 1$ and $C_1 = Ra_2 \oplus C_2$. Notice that $R/\pi R$ is a field, and

$$M(\pi)/\pi M(\pi) = (Ra_1)/(\pi Ra_1) \oplus (Ra_2)/(\pi Ra_2) \oplus C_2/\pi C_2.$$

is a finite dimensional vector space. Since $(Ra_i)/(\pi Ra_i)$ is a vector space of dimension one, the number of times we can apply Lemma 4.6.11 is bounded by the dimension of the vector space $M/\pi M$. After a finite number of iterations we arrive at (3).

(4): Fix a prime π in R such that $M(\pi)$ is nonzero. In the proof of Step (3) we saw that the integer m is uniquely determined since it is equal to the dimension of the vector space $M/\pi M$ over the field $R/\pi R$. Suppose there are two decompositions of $M(\pi)$ into direct sums of cyclic submodules

$$M(\pi) = Ra_1 \oplus Ra_2 \oplus \cdots \oplus Ra_m = Rb_1 \oplus Rb_2 \oplus \cdots \oplus Rb_m,$$

where the order of a_i is equal to π^{e_i} where $e_1 \geq e_2 \geq \cdots \geq e_m$, and the order of b_i is equal to π^{f_i}, where $f_1 \geq f_2 \geq \cdots \geq f_m$. We must show that $e_i = f_i$ for each i. Consider the submodule

$$\pi M(\pi) = \pi Ra_1 \oplus \pi Ra_2 \oplus \cdots \oplus \pi Ra_m = \pi Rb_1 \oplus \pi Rb_2 \oplus \cdots \oplus \pi Rb_m.$$

By Lemma 4.6.9, the order of the cyclic module πRa_i is $\pi^{e_i - 1}$. If $e_1 = 1$, then $\pi M(\pi) = 0$ which implies $f_1 = 1$. The proof follows by induction on e_1. □

Theorem 4.6.13 will play an important role in Section 6.2.1 when we define the rational canonical form of a linear transformation.

THEOREM 4.6.13. *(Basis Theorem – Invariant Factor Form) Let R be a PID and M a finitely generated R-module. The following are true.*

(1) $M = F \oplus M_t$, where F is a free submodule of finite rank. The rank of F is uniquely determined by M.

(2) There exist $r_1, \ldots, r_\ell \in R$ such that $r_1 \mid r_2 \mid r_3 \mid \cdots \mid r_\ell$ and

$$M_t \cong R/(r_1 R) \oplus \cdots \oplus R/(r_\ell R).$$

The integer ℓ is uniquely determined by M. Up to associates in R, the elements r_i are uniquely determined by M.

The elements $r_1, \ldots, r_\ell$ are called the invariant factors *of M.*

PROOF. By Theorem 4.6.12, there is a finite set of primes $\{\pi_i \mid 1 \leq i \leq k\}$ and a finite set of nonnegative integers $\{e_{ij} \mid 1 \leq i \leq k,\ 1 \leq j \leq \ell\}$ such that

$$M_t \cong \bigoplus_{i=1}^{k} \bigoplus_{j=1}^{\ell} R/(\pi_i^{e_{ij}} R).$$

For each i we assume $e_{i1} \geq e_{i2} \geq \cdots \geq e_{i\ell} \geq 0$. Also assume for at least one of the primes π_i that $e_{i\ell} \geq 1$. For each j such that $1 \leq j \leq \ell$, set $r'_j = \prod_{i=1}^{k} \pi_i^{e_{ij}}$. Then $r'_\ell \mid \cdots \mid r'_2 \mid r'_1$. Reverse the order by setting $r_1 = r'_\ell, r_2 = r'_{\ell-1}, \ldots, r_\ell = r'_1$. By Exercise 4.6.19 (3),

$$R/(r'_j) \cong \bigoplus_{i=1}^{k} R/(\pi_i^{e_{ij}} R)$$

from which it follows that $M_t \cong R/(r_1 R) \oplus \cdots \oplus R/(r_\ell R)$. This proves the existence claim of Part (2).

For the uniqueness claim, suppose we are given the elements $r_1, \ldots, r_\ell$ in R. By unique factorization in R, $r_\ell = \pi_1^{e_{1\ell}} \cdots \pi_k^{e_{k\ell}}$. Likewise, factor each of the other r_i. By stepping through the existence proof backwards, we get

$$M_t \cong \bigoplus_{i=1}^{k} \bigoplus_{j=1}^{\ell} R/(\pi_i^{e_{ij}} R).$$

The uniqueness of the primes and the exponents follows from Theorem 4.6.12. This gives the uniqueness of the r_i. □

EXAMPLE 4.6.14. Suppose M is a finitely generated abelian group of rank n. Consider the cases that can arise when $n \leq 3$.

(1) If $n = 1$, then M is cyclic. There are two cases: $M \cong \mathbb{Z}$, or $M \cong \mathbb{Z}/r_1$, for some $r_1 > 1$.
(2) If $n = 2$, then there are three cases:
   (a) $M \cong \mathbb{Z} \oplus \mathbb{Z}$, or
   (b) $M \cong \mathbb{Z}/r_1 \oplus \mathbb{Z}$, where $1 < r_1$, or
   (c) $M \cong \mathbb{Z}/r_1 \oplus \mathbb{Z}/r_2$, where $1 < r_1 \leq r_2$, and $r_1 \mid r_2$.
(3) If $n = 3$, then there are four cases:
   (a) $M \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$, or
   (b) $M \cong \mathbb{Z}/r_1 \oplus \mathbb{Z} \oplus \mathbb{Z}$, where $1 < r_1$, or
   (c) $M \cong \mathbb{Z}/r_1 \oplus \mathbb{Z}/r_2 \oplus \mathbb{Z}$, where $1 < r_1 \leq r_2$, and $r_1 \mid r_2$, or
   (d) $M \cong \mathbb{Z}/r_1 \oplus \mathbb{Z}/r_2 \oplus \mathbb{Z}/r_3$, where $1 < r_1 \leq r_2 \leq r_3$, and $r_1 \mid r_2 \mid r_3$.

COROLLARY 4.6.15. *In the context of Theorem 4.6.13, there exist elements $x_1, \ldots, x_n$ in M such that M is the internal direct sum of the cyclic submodules $M = Rx_1 \oplus \cdots \oplus Rx_n$, where $n \geq \ell$, $I_{x_i} = Rr_i$, $r_1 \mid r_2 \mid \cdots \mid r_n$, and $r_i = 0$ if and only if $i > \ell$.*

PROOF. The proof is left to the reader. □

The "Basis" of M mentioned in Theorem 4.6.13 is the set $\{x_1, \ldots, x_n\}$ of Corollary 4.6.15. The invariant factors r_i are zero for the basis elements $x_{\ell+1}, \ldots, x_n$ corresponding to the basis of the free part F.

6.4. Exercises.

EXERCISE 4.6.16. Let R be an integral domain and M an R-module. Let M_t be the set of all torsion elements in M (see Definition 4.6.4). Show that M_t is a submodule of M. Show that M/M_t is a torsion free R-module.

EXERCISE 4.6.17. Let R be a PID. Show that every nonzero ideal of R is a free R-module of rank 1.

EXERCISE 4.6.18. Let R be a PID. Let π be an irreducible element of R, $e > 0$ and $A = R/(\pi^e)$. Prove:

(1) Every ideal in A is principal.
(2) A is a field if and only if $e = 1$.
(3) A is a local ring, the unique maximal ideal is generated by π.
(4) A has exactly $e + 1$ ideals, namely: $(0) \subseteq (\pi^{e-1}) \subseteq \cdots \subseteq (\pi^2) \subseteq (\pi) \subseteq A$.

EXERCISE 4.6.19. Let R be a PID. Let $\pi_1, \ldots, \pi_n$ be irreducible elements of R that are pairwise nonassociates. Let $e_1, \ldots, e_n$ be positive integers. If $x = \pi_1^{e_1} \pi_2^{e_2} \cdots \pi_n^{e_n}$, and $A = R/(x)$, prove:

(1) The ideals in A correspond to the divisors of x. Including the two trivial ideals (0) and A, there are exactly $(e_1+1)(e_2+1)\cdots(e_n+1)$ ideals in A.
(2) A has exactly n maximal ideals, namely $(\pi_1),\ldots,(\pi_n)$.
(3) A is isomorphic to the direct sum of the local rings $\bigoplus_i R/(\pi_i^{e_i})$.

EXERCISE 4.6.20. (The abelian group $\mathbb{Q}/\mathbb{Z}$) This exercise is a continuation of Exercises 2.2.31 and 2.3.23. For any integer $r \geq 1$, let $\ell_r : \mathbb{Q}/\mathbb{Z} \to \mathbb{Q}/\mathbb{Z}$ be the left multiplication by r map. Prove the following.

(1) Show that ℓ_r is onto for all $r \geq 1$. We say $\mathbb{Q}/\mathbb{Z}$ is a divisible abelian group.
(2) $\mathbb{Q}/\mathbb{Z}$ is a torsion $\mathbb{Z}$-module.
(3) The kernel of ℓ_r is a cyclic group of order r.
(4) If H is a finite subgroup of $\mathbb{Q}/\mathbb{Z}$, then H is cyclic.
(5) If H is a finite subgroup of $\mathbb{Q}/\mathbb{Z}$, then $(\mathbb{Q}/\mathbb{Z})/H$ is isomorphic to $\mathbb{Q}/\mathbb{Z}$.

EXERCISE 4.6.21. (The p-torsion subgroup of $\mathbb{Q}/\mathbb{Z}$) Let p be a prime number. As in Section 4.6, let

$$\mathbb{Q}/\mathbb{Z}(p) = \bigcup_{n>0} \ker(\ell_{p^n})$$

be the subgroup of $\mathbb{Q}/\mathbb{Z}$ consisting of all elements annihilated by some power of p. Some authors denote the group $\mathbb{Q}/\mathbb{Z}(p)$ by $\mathbb{Z}(p^\infty)$. Prove the following.

(1) Every proper subgroup of $\mathbb{Q}/\mathbb{Z}(p)$ is a finite cyclic group.
(2) $\mathbb{Q}/\mathbb{Z}(p)$ is a divisible group (see Exercise 4.6.20 (1)).
(3) $\mathbb{Q}/\mathbb{Z}$ is equal to the internal direct sum $\bigoplus_{p\in P} \mathbb{Q}/\mathbb{Z}(p)$, where P is the set of all prime numbers.
(4) If H is a proper subgroup of $\mathbb{Q}/\mathbb{Z}(p)$, then the quotient $\mathbb{Q}/\mathbb{Z}(p)/H$ is isomorphic to $\mathbb{Q}/\mathbb{Z}(p)$.

EXERCISE 4.6.22. Let R be a local ring (see Exercises 3.2.43 and 3.5.7). Let n denote the characteristic of R. Prove that $n = 0$ or there is a prime p and $t > 0$ such that $n = p^t$.

EXERCISE 4.6.23. Let M be a finitely generated module over a principal ideal domain R. State and prove a version of Corollary 4.6.15 for Theorem 4.6.12, the elementary divisor form of the Basis Theorem.

CHAPTER 5

Fields

A field is a commutative ring in which $0 \neq 1$ and every nonzero element is invertible. In this chapter, the study of an arbitrary field F is always in relation to its subfields. That is, F will be viewed as an extension of a subfield k. In this context, F is a k-algebra. Therefore, elements of F are either algebraic or transcendental over k. A central theme of this book is that Algebra is the study of polynomial equations. To study a polynomial equation over a field k, we consider those extension fields F that contain solutions to the given polynomial equation. The algebraic properties of the field F provide information about the polynomials over k.

If $p(x)$ is a polynomial with coefficients over a field k, then we show in Kronecker's Theorem (see Theorem 5.2.4 and its corollary) that there is an extension field of k which contains all of the roots of $p(x)$. A minimal extension field of k containing all of the roots of $p(x)$ is called a splitting field of $p(x)$ and is unique up to isomorphism. A polynomial $p(x)$ over k is called separable if every irreducible factor of $p(x)$ has only simple roots in a splitting field. An extension field of k that is the splitting field of a separable polynomial is called a Galois extension of k. If F is a Galois extension of k, the group $G = \mathrm{Aut}_k(F)$ of k-algebra automorphisms is called the Galois group. For a Galois extension, the group G acts as a group of automorphisms of the field F, and the subset fixed by G is the field k. Groups arise as permutation groups of the roots of the polynomial $p(x)$. Since a polynomial has only a finite number of roots, in this chapter we restrict our attention to finite groups. There is a connection between the groups acting on the roots of $p(x)$ and the intermediate fields between k and F. This relationship is encapsulated in the Fundamental Theorem of Galois Theory.

In Section 5.5 we consider Galois extensions in some particular cases. A cyclotomic extension of order n over the field k is the splitting field F of the polynomial $x^n - 1$. Thus, F is the smallest extension field of k containing all of the nth roots of unity. Section 5.5.5 contains an introduction to the study of finite fields.

A Galois extension F of a field k is called a cyclic extension if the Galois group G is a cyclic group. Section 5.6 includes a theorem of E. Artin and O. Schreier on cyclic extensions of degree p of a field with characteristic p. There is a theorem on Kummer Theory that classifies cyclic extensions of a cyclotomic extension. Radical extensions of a field k arise as intermediate fields of the splitting field of a polynomial of the form $x^n - a$, where $a \in k$.

Section 5.7 contains an introduction to the theory of symmetric rational functions. In this context we prove a theorem of Abel that a general polynomial of degree $n \geq 5$ is not solvable by radicals. That is, there is no formula involving only square roots, cube roots, fourth roots, ..., nth roots for factoring a general polynomial of degree n.

If k is a field, there is a unique homomorphism $\eta : \mathbb{Z} \to k$ and the kernel of η is either (0), or (p) for some prime p (Example 3.2.2 (5) and Exercise 3.2.30 (2)). If η is one-to-one, then the characteristic of k is zero and k contains the quotient field of $\operatorname{im}\eta$, which is isomorphic to the field of rational numbers $\mathbb{Q}$ (Exercise 3.5.2). Otherwise, the characteristic of k is positive and the image of η is a finite field isomorphic to $\mathbb{Z}/p$, where $p = \operatorname{char} k$. The image of η is contained in every subring of k. The *prime subfield* of k is the smallest subfield P of k. Since P contains the image of η, if $\operatorname{char} k = 0$, then P is isomorphic to $\mathbb{Q}$. Otherwise, $\operatorname{char} k = p$ is positive and P is isomorphic to $\mathbb{Z}/p$.

1. Field Extensions

This section serves as the preparation site for the rest of the chapter. The results in Section 5.1.1 are basic and of a foundational nature. Section 5.1.1 contains an illustration of how Algebra can be applied to Geometry. Using field extensions, three questions of antiquity involving straightedge and compass constructions are answered in Theorem 5.1.17.

1.1. Algebraic Extensions and Transcendental Extensions. Let k and F be fields. If k is a subring of F, then we say F is an *extension* of k, k is a *subfield* of F, or that F/k is an *extension of fields*. An *intermediate field of* F/k is a field E such that $k \subseteq E \subseteq F$, k is a subfield of E, and E is subfield of F.

Definition 5.1.1. Let F/k be an extension of fields. Then F is a k-algebra, and in particular F is a vector space over k. If $X \subseteq F$, then as in Definition 4.4.6 we denote by $k[X]$ the k-subalgebra of F generated by k and X. By $k(X)$ we denote the *subfield of F generated by k and X*. If $F = k(u_1, \ldots, u_n)$, then we say F is a *finitely generated field extension of k*. If $F = k(u)$, then we say F is a *simple extension* of k and u is a *primitive element*. The *degree* of the extension F/k is the dimension of F as a k-vector space.

Example 5.1.2. Let F be a finite field of order q. Let k be the prime subfield of F. If F has characteristic p, then k is isomorphic to $\mathbb{Z}/p$. If $\dim_k F = n$, then $q = p^n$. By Corollary 3.6.12, the group of units of F is a cyclic group of order $q-1$. Let $\zeta \in F^*$ be an element of order $q-1$. Then $F = k(\zeta)$ is a simple extension and ζ is a primitive element.

Lemma 5.1.3. *Let F/k be an extension of fields and $X \subseteq F$. Then*

$k[X] = \{g(u_1, \ldots, u_n) \mid n \geq 1, u_i \in X, g \in k[x_1, \ldots, x_n]\}$, *and*

$$k(X) = \left\{ \frac{g(u_1, \ldots, u_n)}{h(v_1, \ldots, v_n)} \mid n \geq 1, u_i, v_j \in X, g, h \in k[x_1, \ldots, x_n], h(v_1, \ldots, v_n) \neq 0 \right\}.$$

As k-algebras, the quotient field of $k[X]$ is isomorphic to $k(X)$.

Proof. Is left to the reader. □

Let F/k be an extension of fields and $u \in F$. By Definition 4.4.7, u is algebraic over k if there is a nonzero polynomial $f \in k[x]$ and $f(u) = 0$. Otherwise, u is transcendental over k. If each element of F is algebraic over k, then F/k is an algebraic extension.

Theorem 5.1.4. *Let F/k be an extension of fields. Let $u \in F$ be an element that is algebraic over k. Let x be an indeterminate. Then the following are true.*

(1) $k[u] = k(u)$.
(2) $k[u] \cong k[x]/(f)$ *where* f *is a polynomial in* $k[x]$ *satisfying:*
  (a) f *is monic and irreducible,*
  (b) $f(u) = 0$, *and*
  (c) if $g \in k[x]$ *and* $g(u) = 0$, *then* $f \mid g$. *The polynomial* f *is uniquely determined by* u. *We call* f *the* irreducible polynomial of u *and write* $f = \operatorname{Irr.poly}_k(u)$. *Sometimes we call* f *the* minimal polynomial of u *and write* $f = \min.\operatorname{poly}_k(u)$.
(3) If $f = \operatorname{Irr.poly}_k(u)$, *and* $\deg f = n$, *then* $\{1, u, \ldots, u^{n-1}\}$ *is a basis for* $k[u]$ *as a* k*-vector space.*
(4) $\dim_k k[u] = n$.

Proof. Since u is algebraic, we know from Theorem 4.4.8 that $\deg f > 0$. If $f = gh$, then $0 = f(u) = g(u)h(u)$. Since F is a field, this implies $g(u) = 0$ or $h(u) = 0$. Theorem 4.4.8 implies that $f \mid g$ or $f \mid h$. So $\deg g = \deg f$ or $\deg h = \deg f$. This proves f is irreducible. The rest follows from Theorem 4.4.8 and Lemma 4.4.5. □

Theorem 5.1.5. *Let* F/k *be an extension of fields and* $u \in F$ *an element that is transcendental over* k. *Let* x *be an indeterminate. Then* $k(x) \cong k(u)$ *by a* k*-algebra isomorphism that maps* x *to* u.

Proof. Define $\tau : k[x] \to F$ to be the "evaluation at u" map. By Theorem 4.4.8, τ maps $k[x]$ isomorphically onto $k[u]$. By Exercise 3.5.2, τ factors through $k(x)$. Hence there is a k-algebra isomorphism $k(x) \cong k(u)$. □

Theorem 5.1.6. *Let* F/k *be an extension of fields and* $u \in F$. *Assume* L/K *is another extension of fields and* $v \in L$. *Let* $\sigma : k \to K$ *be an isomorphism of fields and assume either*

(1) u *is transcendental over* k *and* v *is transcendental over* K, *or*
(2) there exists an irreducible polynomial $f \in k[x]$ *such that* $f(u) = 0$ *and* $(\sigma f)(v) = 0$.

Then there is an isomorphism $\tau : k(u) \to K(v)$ *such that* $\tau(u) = v$ *and* $\tau|_k = \sigma$.

Proof. (1): Follows straight from Theorem 5.1.5.

(2): Because σ is an isomorphism of fields, by Theorem 3.6.2 (1), we have an isomorphism of polynomial rings $\sigma : k[x] \to K[x]$, where $\sigma\left(\sum a_i x^i\right) = \sum \sigma(a_i) x^i$. Therefore, $\sigma(f)$ is irreducible in $K[x]$. Then $\ker \eta\sigma = (f)$ and the diagram

$$\begin{array}{ccc} k[x] & \xrightarrow{\sigma} & K[x] \\ \downarrow & & \downarrow \eta \\ \frac{k[x]}{(f)} & \xrightarrow{\tau} & \frac{K[x]}{(\sigma f)} \end{array}$$

commutes. By Corollary 3.2.16, τ is an isomorphism. The rest follows from Theorem 5.1.4. □

The next two corollaries play a fundamental role in Galois Theory.

Corollary 5.1.7. *Let* F/k *be an extension of fields and assume* $u, v \in F$. *Assume either*

(1) u *and* v *are transcendental over* k, *or*

(2) u and v are algebraic and satisfy the same irreducible polynomial.
Then there is a k-algebra isomorphism $\tau : k(u) \to k(v)$ such that $\tau(u) = v$.

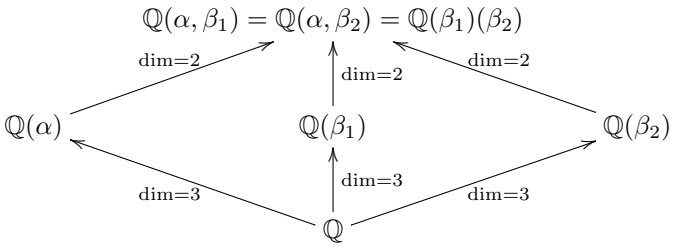
COROLLARY 5.1.8. *Let F/k be an extension of fields. Assume $u, v \in F$ are algebraic over k and that there is a k-algebra isomorphism $\tau : k(u) \to k(v)$ such that $\tau(u) = v$. Then u and v satisfy the same irreducible polynomial.*

PROOF. Let $\phi : k[x] \to k[u]$ where $\phi(x) = u$. Let $\psi : k[x] \to k[v]$ where $\psi(x) = v$. The diagram of k-algebra homomorphisms

$$\begin{array}{ccc} k[x] & \xrightarrow{\phi} & k[u] \\ {\scriptstyle =}\downarrow & & \downarrow{\scriptstyle \tau} \\ k[x] & \xrightarrow{\psi} & k[v] \end{array}$$

commutes. Let $\ker(\phi) = (f)$, where f is the monic irreducible polynomial for u. The diagram commutes, so $f \in \ker(\psi)$. It follows that $f(v) = 0$. By Theorem 5.1.4, it follows that $\ker(\psi)$ is generated by f. □

EXAMPLE 5.1.9. In $\mathbb{Q}[x]$, let $p(x) = x^3+2x+1$. By the Rational Root Theorem, $p(1) = 4$ and $p(-1) = -2$ imply $p(x)$ has no root in $\mathbb{Q}$. Therefore, p is irreducible. Since $p'(x) = 3x^2 + 2$ is positive, we see that $p(x)$ has exactly one real root, call it α. In $\mathbb{C}$ there are two nonreal roots of $p(x)$, call them β_1, β_2. By Example 4.4.12, β_1 and β_2 are complex conjugates of each other. By Corollary 5.1.7, the fields $\mathbb{Q}(\alpha), \mathbb{Q}(\beta_1), \mathbb{Q}(\beta_2)$ are pairwise isomorphic to each other. Since $\mathbb{Q}(\alpha) \subseteq \mathbb{R}$ and $\beta_i \notin \mathbb{R}$, we know that as subsets of $\mathbb{C}$, $\mathbb{Q}(\alpha)$ is not equal to $\mathbb{Q}(\beta_i)$. By Corollary 3.6.9, the polynomial $p(x)$ factors over the field $\mathbb{Q}(\alpha)$ into $p(x) = (x-\alpha)q(x)$, where $q(x)$ is an irreducible quadratic with roots β_1, β_2. This implies $\mathbb{Q}(\alpha)(\beta_1)$ has degree 2 over $\mathbb{Q}(\alpha)$. Any field that contains two of the three roots contains the third, since $p(x)$ has degree three. Therefore, $\mathbb{Q}(\alpha)(\beta_1) = \mathbb{Q}(\alpha)(\beta_2) = \mathbb{Q}(\beta_1)(\beta_2)$.



This shows that although the fields $\mathbb{Q}(\alpha)$, $\mathbb{Q}(\beta_1)$, $\mathbb{Q}(\beta_2)$ are pairwise isomorphic, no two of them are equal as sets. Using Galois Theory, we will see in Example 5.3.23 below that there is a fourth intermediate field which is a quadratic extension of $\mathbb{Q}$.

PROPOSITION 5.1.10. *Let F/k be an extension of fields.*

(1) (Finite Dimensional is Finitely Generated and Algebraic) If F is finite dimensional over k, then F is finitely generated and algebraic over k.
(2) (Finitely Generated and Algebraic is Finite Dimensional) Given a finite subset $X = \{u_1, \ldots, u_n\} \subseteq F$ such that each u_i is algebraic over k, it follows that the field extension $k(X)/k$ is finite dimensional.
(3) If $F = k(X)$ and every element of X is algebraic over k, then F is algebraic over k.

(4) (Algebraic over Algebraic is Algebraic) Let E be an intermediate field of F/k. If F/E is algebraic and E/k is algebraic, then F/k is algebraic.
(5) (Algebraic Closure of k in F Exists) If $E = \{u \in F \mid u \text{ is algebraic over } k\}$, then E is an intermediate field of F/k.

PROOF. (1): Since F is finite dimensional over k, F is finitely generated (Definition 4.3.5). By Corollary 4.4.11, F is algebraic over k.

(2): By Theorem 4.4.8 (5), $\dim_k k(u_1) < \infty$. Now use induction and Proposition 4.3.8.

(3): Let $u \in k(X)$. By Lemma 5.1.3 there exist $u_1, \ldots, u_m, v_1, \ldots, v_n$ in X and polynomials f, g over k such that

$$u = \frac{f(u_1, \ldots, u_m)}{g(v_1, \ldots, v_n)}.$$

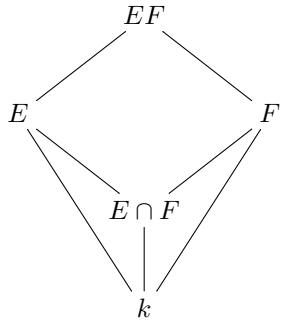
This shows $u \in k(u_1, \ldots, u_m, v_1, \ldots, v_n)$. By Parts (2) and (1) this shows u is algebraic over k.

(4): Let $u \in F$. There is a polynomial $f = \sum_{i=0}^{n} a_i x^i$ in $E[x]$ such that $f(u) = 0$. Let $K = k(a_0, \ldots, a_n)$. Then u is algebraic over K and $\dim_K K(u) < \infty$. Since each a_i is algebraic over k, by Part (2), $\dim_k K < \infty$. By Proposition 4.3.8, $\dim_k K(u) < \infty$. By Part (1), u is algebraic over k.

(5): Let u, v be algebraic over k. By Part (3), $k(u, v)$ is an algebraic extension of k. So $k(u, v) \subseteq E$. Therefore, $u + v$, $u - v$, uv, u/v are all in E. It follows that E is a field. □

DEFINITION 5.1.11. Let K/k be an extension of fields. Let E and F be intermediate fields. That is, $k \subseteq E \subseteq K$ and $k \subseteq F \subseteq K$. The *composite* of E and F, denoted EF, is $k(E \cup F)$. The reader should verify that the set of all intermediate fields of K/k is a lattice.

THEOREM 5.1.12. *Let K/k be an extension of fields. Let E and F be intermediate fields.*



Assume $\dim_k F = n$ is finite and that $\{v_1, \ldots, v_n\}$ is a basis for F as a k-vector space. The following are true.

(1) As a vector space over E, EF is spanned by $\{v_1, \ldots, v_n\}$.
(2) $\dim_E (EF) \leq \dim_k F$.
(3) If $\dim_k E = m$ is finite and $\{u_1, \ldots, u_m\}$ is a basis for E as a k-vector space, then $\dim_k EF \leq \dim_k E \dim_k F$ and as a vector space over k, EF is spanned by $\{u_i v_j \mid 1 \leq i \leq m, 1 \leq j \leq n\}$.
(4) If $\dim_k E$ and $\dim_k F$ are both finite and relatively prime to each other, then $\dim_k EF = \dim_k F \dim_k E$.

(5) If $\dim_k EF = \dim_k F \dim_k E$*, then* $k = E \cap F$*.*

PROOF. (1): We have $F = k(v_1, \ldots, v_n)$. It follows that $EF = k(E \cup F) = k(E)(F) = E(F) = E(k(v_1, \ldots, v_n)) = E(v_1, \ldots, v_n)$. By Exercise 5.1.22, a typical element u in EF is a linear combination $u = e_1 M_1 + \cdots + e_r M_r$ where each e_i is in E and each M_i is a monomial of the form $M_i = v_1^{\epsilon_{i,1}} \cdots v_n^{\epsilon_{i,n}}$, where $\epsilon_{i,j} \geq 0$ for each i, j. In the field F, each monomial M_i can be written as a k-linear combination in the form $M_i = a_{i,1} v_1 + \cdots + a_{i,n} v_n$, where $a_{i,j} \in k$ for each i, j. Therefore,

$$\begin{aligned} u &= e_1 M_1 + \cdots + e_r M_r \\ &= \sum_{i=1}^{r} e_i \left(\sum_{j=1}^{n} a_{i,j} v_j \right) \\ &= \sum_{i=1}^{r} \left(\sum_{j=1}^{n} e_i a_{i,j} v_j \right) \\ &= \sum_{j=1}^{n} \left(\sum_{i=1}^{r} e_i a_{i,j} \right) v_j \end{aligned}$$

This proves (1) since each $e_i a_{i,j}$ is in E.

(2): This follows from (1) and Corollary 4.3.6.

(3): This follows from (2) and Proposition 4.3.8.

(4): We have $\dim_k(E) = m$ and $\dim_k(F) = n$ both divide $\dim_k(EF)$. Since m and n are relatively prime, it follows that mn is the least common multiple of m and n. Thus $mn \leq \dim_k(EF)$. This and (3) proves (4).

(5): We have $\dim_k(EF) = \dim_k(F) \dim_k(E) = \dim_E(EF) \dim_k(E)$, which implies $\dim_E(EF) = \dim_k(F)$. By this and (2), $\dim_E(EF) = \dim_k(F) \leq \dim_{E \cap F}(F)$. It follows from Proposition 4.3.8 that $k = E \cap F$. □

1.2. Classical Straightedge and Compass Constructions. In this section we apply field extensions to answer three questions of antiquity on geometric constructions using straightedge and compass. The results of this section are not applied anywhere else in the book.

A real number a in $\mathbb{R}$ is *constructible* if by use of straightedge and compass we can construct a line segment of length $|a|$. We are given that 1 is constructible. Ruler and compass constructions involve:

(1) Drawing lines through two points.
(2) Intersecting two lines.
(3) Drawing a circle with a given center and radius.
(4) Intersecting a line and a circle.
(5) Intersecting two circles.

LEMMA 5.1.13. *The set of all constructible numbers is a subfield of* $\mathbb{R}$ *containing* $\mathbb{Q}$.

PROOF. Using the straightedge we can construct the x-axis. Given the unit length 1 and compass we can construct any $n \in \mathbb{Z}$. In fact, for any constructible numbers a and b, the compass can be used to construct $a \pm b$. Using the straightedge and compass we can construct the y-axis, by erecting a perpendicular to the x-axis at the number 0. The line L through the points $(0, 0)$ and $(1, b)$ in $\mathbb{R}^2$ is the set

of solutions to $y = bx$. The point (a, ab) is the intersection of L with the vertical line through $(a, 0)$. If $b \neq 0$, the point $(a/b, b)$ is the intersection of L with the horizontal line through $(0, b)$. Therefore, ab and a/b are constructible. □

Let F be any subfield of $\mathbb{R}$. Let $F^2 = \{(x, y) \mid x, y \in F\}$ be the *plane over* F, which we view as a subset of the euclidean plan $\mathbb{R}^2$. A linear equation over F in two variables is an equation of the form $ax+by+c = 0$, where a and b are in F and are not both equal to 0. A *line* in F^2 is the set of solutions $(x, y) \in F^2$ to a linear equation over F. A *circle* in F^2 is the set of solutions $(x, y) \in F^2$ to a quadratic equation of the form $x^2 + y^2 + ax + by + c = 0$, where $a, b, c \in F$.

LEMMA 5.1.14. *The following are true.*

(1) *Given $A_0 = (x_0, y_0)$ and $A_1 = (x_1, y_1)$ in F^2, if $A_0 \neq A_1$, there is a line L in F^2 passing through A_0 and A_1.*
(2) *Given a point $A_0 = (x_0, y_0)$ in F^2 and a positive $r \in F$, there is a circle in F^2 with center A_0 and radius r.*
(3) *If L_1 and L_2 are non-parallel lines in F^2, then $L_1 \cap L_2$ is a point in F^2.*
(4) *If L is a line and C a circle, both in F^2, and $L \cap C$ is nonempty in $\mathbb{R}^2$, then $L \cap C$ is nonempty in the plane over $F(\sqrt{\gamma})$, for some $\gamma \in F$, $\gamma \geq 0$.*
(5) *If C_0 and C_1 are circles in F^2, and $C_0 \cap C_1$ is nonempty in $\mathbb{R}^2$, then $C_0 \cap C_1$ is nonempty in the plane over $F(\sqrt{\gamma})$, for some $\gamma \in F$, $\gamma \geq 0$.*

PROOF. (1), (2) and (3): Proofs are left to the reader.

(4): Suppose the equation for C is $x^2 + y^2 + ax + by + c = 0$, and the equation for L is $dx+ey+f = 0$, where $a, b, c, d, e, f \in F$. Without loss of generality, assume $e \neq 0$. Solve for y on the line L to get $y = -(f + dx)/e$. Substituting into C,

$$x^2 + (f + dx)^2/e^2 + ax - b(f + dx)/e + c = 0.$$

This is a quadratic equation over F of the form $Ax^2 + Bx + C = 0$, where $A = (e^2 + d^2)/e^2 > 0$. In the field of complex numbers $\mathbb{C}$ the solutions are

$$x = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}.$$

Let $\gamma = B^2 - 4AC$. Then $\gamma \in F$. If $\gamma = 0$, then $L \cap C$ consists of one point in F^2. If $\gamma < 0$, then in $\mathbb{R}^2$, $L \cap C = \emptyset$. If $\gamma > 0$, then there are two points in $L \cap C$, and both belong to the plane over $F(\sqrt{\gamma})$.

(5): Suppose the equation for C_0 is $x^2+y^2+a_0x+b_0y+c_0 = 0$, and the equation for C_1 is $x^2 + y^2 + a_1x + b_1y + c_1 = 0$. If $C_0 = C_1$, then take γ to be 1. Otherwise subtract to get $(a_0 - a_1)x + (b_0 - b_1)y + (c_0 - c_1) = 0$. If $a_0 = a_1$ and $b_0 = b_1$, then $C_0 \cap C_1 = \emptyset$. Otherwise the linear equation $(a_0 - a_1)x + (b_0 - b_1)y + (c_0 - c_1) = 0$ defines a line, which we call L. Then $C_0 \cap L = C_1 \cap L = C_0 \cap C_1$, and we reduce to part (4). □

PROPOSITION 5.1.15. *If $u \in \mathbb{R}$ is constructible, then for some $r \geq 0$, $\dim_{\mathbb{Q}}(\mathbb{Q}(u))$ is equal to 2^r.*

PROOF. To construct u, a finite sequence of straightedge and compass constructions are performed. By Lemma 5.1.14, u belongs to a field extension of $\mathbb{Q}$ obtained by a finite number of quadratic extensions, each of which is inside $\mathbb{R}$. There exist positive real numbers $\gamma_1, \ldots, \gamma_n$ such that u belongs to $\mathbb{Q}(\gamma_1)\cdots(\gamma_n)$, a subfield of $\mathbb{R}$. Moreover, $\gamma_1^2 \in \mathbb{Q}$ and for $1 < i \leq n$, $\gamma_i^2 \in \mathbb{Q}(\gamma_1, \ldots, \gamma_{i-1})$. By

Proposition 4.3.8, degrees of consecutive extensions multiply. The degree of each consecutive extension is either 1 or 2. This means $\dim_{\mathbb{Q}}(\mathbb{Q}(\gamma_1, \ldots, \gamma_n))$ is 2^s for some $s \geq 0$. Since $\dim_{\mathbb{Q}}(\mathbb{Q}(u))$ divides 2^s, we are done. □

COROLLARY 5.1.16. *Suppose $u \in \mathbb{R}$ is algebraic over $\mathbb{Q}$ and the degree of* $\operatorname{Irr.poly}_{\mathbb{Q}}(u)$ *has degree d. If d is not of the form 2^r, then u is not constructible.*

THEOREM 5.1.17. *It is impossible by straightedge and compass alone to*

(1) trisect the angle $60°$ (that is, $\cos 20°$ is not constructible),
(2) double the cube (that is, $\sqrt[3]{2}$ is not constructible), or
(3) square the circle (that is, $\sqrt{\pi}$ is not constructible).

PROOF. (1): Take θ to be $60°$. Then $\cos\theta = \frac{1}{2}$. By trigonometry, $\cos\theta = 4\cos^3\left(\frac{\theta}{3}\right) - 3\cos\left(\frac{\theta}{3}\right)$. Let $u = 2\cos 20°$. Then u satisfies $u^3 - 3u - 1 = 0$. The irreducible polynomial for u over $\mathbb{Q}$ is $x^3 - 3x - 1$, which has degree 3. Then u is not constructible, $\cos 20°$ is not constructible, and it is impossible to trisect $60°$.

(2): The irreducible polynomial for $\sqrt[3]{2}$ over $\mathbb{Q}$ is $x^3 - 2$, which has degree 3.

(3): We have not proved it here, but π is transcendental. Hence $\sqrt{\pi}$ is not constructible. □

1.3. Exercises.

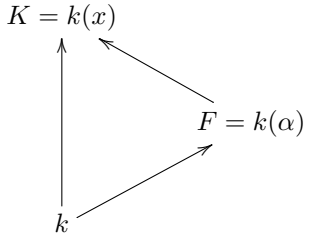
EXERCISE 5.1.18. Let p be an odd prime and $k = \mathbb{Z}/p$ the field of order p. Show that there are $(p-1)/2$ elements $\alpha \in U_p$ such that $\phi_\alpha = x^2 - \alpha$ is irreducible. Show that in this case $k[x]/(\phi_\alpha)$ is a field of order p^2.

EXERCISE 5.1.19. Let $k = \mathbb{Z}/3$ be the field of order 3. Show that $f = x^2 + 1$ is irreducible over k. Let $F = k[x]/(f)$. Let $u \in F$ be the coset represented by x. By Corollary 3.6.12, the group F^* is cyclic. A generator for F^* is called a primitive element. Show that $u + 1, u - 1, -u + 1, -u - 1$ are the four primitive elements in F^*.

EXERCISE 5.1.20. Let $p(x) = x^3 - 3x - 1 \in \mathbb{Q}[x]$. Show that p is irreducible and let $F = \mathbb{Q}[x]/(p)$ be the quotient. Let u denote the element of F corresponding to the coset containing x.

(1) Exhibit a basis for F as a $\mathbb{Q}$-vector space.
(2) Write the following in terms of the basis given in (1): u^{-1}, $u^4 + 2u^3 + 3$, u^{-2}.

EXERCISE 5.1.21. Let k be a field, x an indeterminate, and $K = k(x)$ the field of rational functions. Let α denote the rational function $x^4/(4x^3 - 1)$ in K. Then $F = k(\alpha)$ is a field extension of k and K is a field extension of F. There is a lattice of subfields



where an arrow denotes set containment. Show that K is algebraic over F. Determine the minimal polynomial of x over F and the dimension $\dim_F(K)$.

EXERCISE 5.1.22. Let K/k be an extension of fields and $u_1, \ldots, u_n$ elements of K, where $n \geq 1$. As in Definition 5.1.1, $k[u_1, \ldots, u_n]$ is the k-subalgebra of K generated by k and $u_1, \ldots, u_n$. Show that a typical element in $k[u_1, \ldots, u_n]$ can be written as a sum of the form $k_1M_1 + \cdots + k_rM_r$ where $a_i \in k$ for each i and each M_i is a product of the form $M_i = u_1^{\epsilon_{i,1}} \cdots u_n^{\epsilon_{i,n}}$ where $\epsilon_{i,j} \geq 0$ for each i, j.

EXERCISE 5.1.23. Let F/k be a finite dimensional extension of fields. If E is an intermediate field of F/k, show that F/E is finite dimensional, E/k is finite dimensional, and $\dim_k(F) = \dim_k(E) \dim_E(F)$.

EXERCISE 5.1.24. Let K/k be an extension of fields and u an element in K that is algebraic over k. Prove that if $\dim_k k(u)$ is odd, then $k(u^2) = k(u)$.

2. Algebraic Field Extensions

There are two main results in this section. Let k be a field and f a polynomial over k. The main result of Section 5.2.1 is the proof that there is a unique extension F/k generated by adjoining the roots of f to k (Corollary 5.2.8). The main result of Section 5.2.2 is the Primitive Element Theorem (Theorem 5.2.14) which contains sufficient conditions for an algebraic extension of fields to be a simple extension.

2.1. Existence and Uniqueness of a Splitting Field. Let k be a field and p a polynomial in $k[x]$ of positive degree. If F/k is an extension of fields, then we say that p *splits* in F if each irreducible factor of p in $F[x]$ is linear. Equivalently, p factors in $F[x]$ into a product of linear polynomials.

LEMMA 5.2.1. *Let F be a field. The following are equivalent.*

(1) Every nonconstant polynomial $p \in F[x]$ has a root in F.
(2) Every nonconstant polynomial $p \in F[x]$ splits in F.
(3) Every irreducible polynomial $p \in F[x]$ has degree 1.
(4) If K/F is an algebraic extension of fields, then $F = K$.
(5) F contains a subfield k such that F/k is algebraic and every polynomial in $k[x]$ splits in F.

PROOF. (1), (2), and (3) are clearly equivalent.

(2) implies (5): Is trivial.

To show (3) and (4) are equivalent, use Theorem 5.1.4.

(5) implies (4): If K/F is algebraic, then by Proposition 5.1.10 (4), K/k is algebraic. If $u \in K$, then the irreducible polynomial of u over k splits in F. Therefore $u \in F$. □

DEFINITION 5.2.2. If F is a field that satisfies any of the equivalent statements of Lemma 5.2.1, then we say F is *algebraically closed*. If F/k is an extension of fields, we say F is an *algebraic closure* of k in case F is algebraic over k, and F is algebraically closed.

DEFINITION 5.2.3. Let F/k be an extension of fields and p a nonconstant polynomial in $k[x]$. We say that F is a *splitting field* of p if

(1) p splits in F, and
(2) $F = k(u_1, \ldots, u_n)$ where $p(u_i) = 0$ for each i.

THEOREM 5.2.4. *(Kronecker's Theorem) Let k be a field and f a polynomial of positive degree in $k[x]$. There exists an extension field F of k and an element $u \in F$ satisfying*

(1) u *is a root of* f*,*
(2) $\dim_k(k[u]) \leq \deg(f)$*, and*
(3) if f *is irreducible, then* $\dim_k(k[u]) = \deg(f)$ *and* $k[u]$ *is unique up to a* k*-algebra isomorphism.*

PROOF. Let p be an irreducible factor of f. Write $f = pq$. Let $F = k[x]/(p)$ and take u to be the coset represented by x in F. Then $p(u) = p([x]) = [p(x)] = [0]$. Then $f(u) = p(u)q(u) = 0$. The rest follows from Theorems 5.1.4 and 5.1.6. □

EXAMPLE 5.2.5. Let p be a prime and k a field of characteristic p. Let $\alpha \in k$ and $f = x^p - \alpha$. In this example we show that f is either irreducible, or splits. The Frobenius homomorphism $\theta : k \to k$ is defined by $a \mapsto a^p$ (Exercise 3.2.32). If $\alpha = a^p$ for some $a \in k$, then $f = x^p - a^p = (x - a)^p$ by (Exercise 3.2.31). This shows that f splits over k if f has a root in k. Now assume that α is not in the image of the Frobenius map. Thus f does not have a root in k. For sake of contradiction assume f is reducible over k. By our assumption, an irreducible factor of f has degree at least 2. This also implies $2 < p$. Let $f = gg_1$ where g is irreducible and $\deg g = m$ where $1 < m < p$. Let $F = k[x]/(g)$. By Theorem 5.2.4, F is an extension field of k containing a root u of g. Every root of g is a root of f. By the first part, $f = (x - u)^p$ in $F[x]$. By Corollary 3.6.5, $F[x]$ is a UFD. This implies $g = (x - u)^m$ in $F[x]$. But $g \in k[x]$. By the Binomial Theorem, $g = x^m - mux^{m-1} + \cdots + (-u)^m$, which implies $mu \in k$. But $\gcd(m, p) = 1$ implies $u \in k$. This contradicts our original assumption that f does not have a root in k. We have shown that $f = x^p - \alpha$ is either irreducible, or splits.

COROLLARY 5.2.6. *If* k *is a field and* f *a polynomial in* $k[x]$ *of positive degree* n*, then there exists a splitting field* F/k *for* f *such that* $\dim_k(F) \leq n!$*.*

PROOF. Factor $f = p_1 \dots p_m$ in $k[x]$ where each p_i is irreducible. If $\deg p_i = 1$ for each i, then take $F = k$ and stop. Otherwise, assume $\deg p_1 > 1$. By Kronecker's Theorem (Theorem 5.2.4), there is an extension field F_1/k such that $F_1 = k(\alpha)$ and $p_1(\alpha) = 0$. Note that $f(\alpha) = 0$ and $\dim_k(F_1) = \deg p_1 \leq n$. Factor $f = (x - \alpha)g$ in $F_1[x]$. By induction on n, there exists a splitting field F/F_1 for g and $\dim_{F_1}(F) \leq (n - 1)!$. So f splits in F and there exist roots $u_1, \dots, u_m$ of f such that $F = F_1(u_1, \dots, u_m) = k(\alpha, u_1, \dots, u_m)$. Lastly, $\dim_k(F) = \dim_k(F_1) \dim_{F_1}(F) \leq n!$, by Proposition 4.3.8. □

LEMMA 5.2.7. *Let* k *be a field,* f *a polynomial in* $k[x]$ *of positive degree* n*, and* F *a splitting field for* f *over* k*. Let* $\sigma : k \to K$ *be an isomorphism of fields,* $\sigma(f)$ *the image of* f *in* $K[x]$*. Let* L/K *be an extension field such that* $\sigma(f)$ *splits in* L*. Then* σ *extends to a homomorphism of* k*-algebras* $\bar{\sigma} : F \to L$ *making a commutative*

$$\begin{array}{ccc} F & \xrightarrow{\bar{\sigma}} & L \\ \uparrow & & \uparrow \\ k & \xrightarrow{\sigma} & K \end{array}$$

diagram. Every root of f *in* F *is mapped by* $\bar{\sigma}$ *to a root of* $\sigma(f)$ *in* L*. If* L *is a splitting field for* $\sigma(f)$*, then* $\bar{\sigma}$ *is an isomorphism.*

PROOF. If $F = k$, then take $\bar{\sigma} = \sigma$ and stop. Otherwise, $\dim_k(F) > 1$ and there is an irreducible factor g of f such that $\deg g > 1$. Let α be a root of g in

F and β a root of $\sigma(g)$ in L. By Theorem 5.1.6 there is a k-algebra isomorphism $\tau : k(\alpha) \to K(\beta)$ such that $\tau(\alpha) = \beta$ and the bottom square of the diagram

$$\begin{array}{ccc} F & \overset{\exists \bar{\sigma}}{\dashrightarrow} & L \\ \uparrow & & \uparrow \\ k(\alpha) & \xrightarrow[\cong]{\tau} & K(\beta) \\ \uparrow & & \uparrow \\ k & \xrightarrow[\cong]{\sigma} & K \end{array}$$

commutes. Also, F is a splitting field for f over $k(\alpha)$, and $\dim_{k(\alpha)}(F) < \dim_k(F)$. By induction on $\dim_k(F)$, τ can be extended to a k-algebra homomorphism $\bar{\sigma} : F \to L$ such that the entire diagram above commutes. A root of f is mapped under $\bar{\sigma}$ to a root of $\sigma(f)$. Since f splits in F, $\sigma(f)$ splits in $\bar{\sigma}(F)$. The polynomial $\sigma(f)$ has at most $\deg(f)$ roots in L by Corollary 3.6.9, and they all belong to $\bar{\sigma}(F)$. If $\lambda \in L$ is a root of $\sigma(f)$, then $\lambda \in \bar{\sigma}(F)$. If L/K is generated by roots of $\sigma(f)$, then $L \subseteq \bar{\sigma}(F)$ and $\bar{\sigma}$ is an isomorphism. □

COROLLARY 5.2.8. *Let k be a field and $f \in k[x]$. A splitting field for f exists and is unique up to k-algebra isomorphism.*

PROOF. This follows straight from Corollary 5.2.6 and Lemma 5.2.7. □

EXAMPLE 5.2.9. Let $n \geq 2$. In $\mathbb{C}$, let $\zeta = e^{2\pi i/n}$. Then ζ is a primitive nth root of unity. That is, $\{\zeta^k \mid 0 \leq k \leq n-1\}$ are the n distinct roots of $x^n - 1$ in $\mathbb{C}$. Therefore, in $\mathbb{C}[x]$

$$x^n - 1 = (x-1)(x-\zeta)(x-\zeta^2)\cdots(x-\zeta^{n-1})$$

is the unique factorization of $x^n - 1$. For each k, $\zeta^k \in \mathbb{Q}(\zeta)$. This shows that $\mathbb{Q}(\zeta)$ is a splitting field for $x^n - 1$ over $\mathbb{Q}$. Consider the polynomial

$$\Phi_n(x) = 1 + x + \cdots + x^{n-1} = \frac{x^n - 1}{x - 1}$$

of degree $n-1$. The distinct roots of Φ_n in $\mathbb{C}$ are $\zeta, \zeta^2, \ldots, \zeta^{n-1}$. By the same reasoning as above, $\mathbb{Q}(\zeta)$ is a splitting field for Φ_n over $\mathbb{Q}$. If p is a prime, then by Example 3.7.8, Φ_p is irreducible over $\mathbb{Q}$. By Theorem 5.1.4, $\Phi_p = \operatorname{Irr.poly}_{\mathbb{Q}}(\zeta)$, $\mathbb{Q}(\zeta) = \mathbb{Q}[x]/(\Phi_p)$, and $\{1, \zeta, \zeta^2, \ldots, \zeta^{p-2}\}$ is a basis for $\mathbb{Q}(\zeta)$ as a $\mathbb{Q}$-vector space. The polynomial $\Phi_p(x)$ is called the *pth cyclotomic polynomial*.

2.2. The Primitive Element Theorem. Let k be a field, $f \in k[x]$, and F/k a splitting field for f. By Corollary 5.2.8, F exists and is unique up to a k-algebra isomorphism. We say f is *separable* in case for every irreducible factor p of f, every root of p in F is a simple root. If K/k is a extension of fields, then we say K/k is a *separable extension* if every $u \in K$ is the root of a separable polynomial in $k[x]$. If $u \in K$ is the root of a separable polynomial in $k[x]$, then we say u is *separable*. A separable extension is an algebraic extension. If $\operatorname{char} k = 0$, then by Theorem 3.6.17 (1), every polynomial $f \in k[x]$ is separable. The purpose of this section is to prove Theorem 5.2.14 which shows that a finite separable extension of fields is a simple extension.

EXAMPLE 5.2.10. Let k be a field of prime characteristic p. The Frobenius homomorphism $\theta : k \to k$ is defined by $a \mapsto a^p$ (Exercise 3.2.32). The image of θ is denoted k^p. Assume θ is not onto and let $\alpha \in k - k^p$. As shown in Example 5.2.5, the polynomial $f = x^p - \alpha$ is irreducible in $k[x]$ but is not separable.

LEMMA 5.2.11. *Let k be a field and f an irreducible polynomial in $k[x]$. The formal derivative of f is denoted f' (see Definition 3.6.15).*

(1) The following are equivalent:
  (a) f is separable.
  (b) $\gcd(f, f') = 1$.
  (c) $f' \neq 0$.

(2) If f is not separable, then $\operatorname{char} k = p$ is a prime number and there exists a polynomial $g(x) \in k[x]$ such that $f(x) = g(x^p)$.

PROOF. This follows from Theorem 3.6.17. □

In Theorem 5.2.12 we assemble various properties of finite fields. In particular, a finite field F is a simple separable extension of its prime subfield k and is uniquely determined by its order.

THEOREM 5.2.12. *Let F be a finite field with $\operatorname{char} F = p$. Let k be the prime subfield of F and $n = \dim_k(F)$.*

(1) The group of units of F is a cyclic group.
(2) $F = k(u)$ is a simple extension, for some $u \in F$.
(3) The order of F is p^n.
(4) F/k is a separable extension.
(5) F is the splitting field for the separable polynomial $x^{p^n} - x$ over k.
(6) Any two finite fields of order p^n are isomorphic as fields.

PROOF. As a k-vector space, F is isomorphic to k^n, which has cardinality $|k|^n$, by Exercise 1.1.12. By Corollary 3.6.12, the group of units of F is a finite cyclic group of order $p^n - 1$. If u is a generator for F^*, then $F = k(u)$. The polynomial $x^{p^n} - x = x(x^{p^n-1} - 1)$ has p^n distinct roots in F. Therefore F is the splitting field for the separable polynomial $x^{p^n} - x$ over k and every element of F is separable over k. By Corollary 5.2.8, F is unique up to k-algebra isomorphism. □

The Primitive Element Theorem is proved in Theorem 5.2.14 below. Our proof is by Mathematical Induction on the number of generators. The inductive step is proved in Lemma 5.2.13.

LEMMA 5.2.13. *Let F/k be an extension of fields. Let α and β be elements of F that are algebraic over k. If β is separable over k, then there exists $\gamma \in F$ such that $k(\alpha, \beta) = k(\gamma)$.*

PROOF. First we prove the lemma for some special cases. Let $K = k(\alpha, \beta)$. If $\alpha \in k$, then $K = k(\beta)$, so set $\gamma = \beta$. If $\beta \in k$, then $K = k(\alpha)$, so set $\gamma = \alpha$. If k is a finite field, then K is a finite field by Proposition 5.1.10 (2). In this case $K = k(\gamma)$ is a simple extension, by Theorem 5.2.12 (2). Assume from now on that $\alpha \notin k$, $\beta \notin k$, and k is infinite. The proof of the general case is split into a sequence of three steps.

Step 1 is to define a candidate for γ. Let $f = \text{min. poly}_k(\alpha)$ be the minimal polynomial for α and $g = \text{min. poly}_k(\beta)$ the minimal polynomial for β. Let F_1 be

a splitting field for fg over F. Let $\alpha = \alpha_1, \alpha_2, \ldots, \alpha_m$ be the distinct roots of f in F_1. Let $\beta = \beta_1, \beta_2, \ldots, \beta_n$ be the distinct roots of g in F_1. By our hypotheses, m and n are positive, and since g is separable we know $n \geq 2$. Consider

$$S = \left\{ \frac{\alpha_1 - \alpha_i}{\beta_j - \beta_1} \mid i = 1, \ldots, m \text{ and } j = 2, \ldots, n \right\}$$

which is a finite subset of F_1. Since k is infinite, there exists $c \in k^*$ such that $c \notin S$. Set $\gamma = \alpha + c\beta$. So $\gamma \in k(\alpha, \beta)$. To finish, it is enough to show $\alpha \in k(\gamma)$ and $\beta \in k(\gamma)$.

Step 2 is to show that $\gamma = \alpha_i + c\beta_j$ if and only if $i = j = 1$. If $1 \leq i \leq m$ and $1 \leq j \leq n$ and $\gamma = \alpha_i + c\beta_j$, then $\alpha_i + c\beta_j = \alpha + c\beta$. So $c(\beta_j - \beta) = \alpha - \alpha_i$. If $j = 1$, then $i = 1$. If $j \neq 1$, then $c = (\alpha - \alpha_i)/(\beta_j - \beta)$. This contradicts the choice of c. This completes Step 2.

Step 3 is to show that $k(\alpha, \beta) \subseteq k(\gamma)$. Define $h(x) \in k(\gamma)[x]$ by $h(x) = f(\gamma - cx)$. Then $h(\beta) = f(\gamma - c\beta) = f(\alpha) = 0$. If $j > 1$, then $\gamma - c\beta_j \neq \alpha_i$ for any i. Thus $h(\beta_j) = f(\gamma - c\beta_j) \neq 0$. Thus, $\beta_2, \ldots, \beta_n$ are not roots of $h(x)$. Let $g_1 = \text{min. poly}_{k(\gamma)}(\beta)$. Since $h(\beta) = 0$, by Theorem 5.1.4 we know $g_1 \mid h$. Likewise, $g(\beta) = 0$ implies $g_1 \mid g$. Every root of g_1 is a root of h and g. We proved that the only root g and h have in common is β. Since g is separable, β is a simple root. It follows that $\gcd(g, h) = x - \beta$. Hence g_1 is linear with one root, β, which implies $\beta \in k(\gamma)$. Moreover, $\alpha = \gamma - c\beta \in k(\beta, \gamma) = k(\gamma)$. □

THEOREM 5.2.14. *(The Primitive Element Theorem) Let F/k be a finite dimensional separable extension of fields. Then there is a separable element $u \in F$ such that $F = k(u)$.*

PROOF. Let $\dim_k(F) = n$. Let $\alpha_1, \ldots, \alpha_n$ be a basis for F as a k-vector space. For $i = 1, \ldots, n$, let $F_i = k(\alpha_1, \ldots, \alpha_i)$. Then $F_2 = k(\alpha_1, \alpha_2)$. Lemma 5.2.13 implies there exists $\gamma_2 \in F$ such that $F_2 = k(\gamma_2)$. By the same argument, $F_3 = F_2(\alpha_3) = k(\gamma_2, \alpha_3)$ and there exists $\gamma_3 \in F$ such that $F_3 = k(\gamma_3)$. Iterate this process $n - 1$ times. Hence $F = F_n = k(\gamma_n)$ for some γ_n. □

2.3. Exercises.

EXERCISE 5.2.15. Show that two finite fields E and F are isomorphic if and only if the order of E is equal to the order of F.

EXERCISE 5.2.16. Let $\alpha = \sqrt[3]{2}$ be the cube root of 2 in $\mathbb{R}$ and $\zeta = e^{2\pi i/3}$ a primitive cube root of 1 in $\mathbb{C}$.

(1) Show that the splitting field for $x^3 - 2$ over $\mathbb{Q}$ is $\mathbb{Q}(\zeta, \alpha)$ and that the dimension of the extension is $\dim_{\mathbb{Q}} \mathbb{Q}(\zeta, \alpha) = 6$.
(2) Show that $\mathbb{Q}(\zeta, \alpha)$ is equal to the composite field EF where E and F are any two fields from this list: $\mathbb{Q}(\zeta)$, $\mathbb{Q}(\alpha)$, $\mathbb{Q}(\zeta\alpha)$, $\mathbb{Q}(\zeta^2\alpha)$.
(3) Show that $\text{Irr. poly}_{\mathbb{Q}(\zeta)}(\alpha)$ has degree 3. Show that $\text{Irr. poly}_{\mathbb{Q}(\zeta)}(\zeta\alpha)$ has degree 3. Show that $\text{Irr. poly}_{\mathbb{Q}(\zeta)}(\zeta^2\alpha)$ has degree 3.
(4) Show that $\text{Irr. poly}_{\mathbb{Q}(\zeta\alpha)}(\alpha)$ has degree 2. Show that $\text{Irr. poly}_{\mathbb{Q}(\zeta^2\alpha)}(\alpha)$ has degree 2.

EXERCISE 5.2.17. Let F/k be an extension of fields and assume $\dim_k F = p$ is prime. Let u be any element of F that is not in k. Prove that $F = k(u)$.

EXERCISE 5.2.18. Let F/k be an extension of fields and assume $\dim_k F = 2$. We say F/k is a quadratic extension of fields. Let u be an element of F that is not in k and $f = \text{Irr. poly}_k u$. Show that F is a splitting field for f over k.

EXERCISE 5.2.19. Let K/k be an extension of fields. Let F_1, F_2 be two intermediate fields where $k \subseteq F_i \subseteq K$ and $\dim_k F_i = 2$ for each i. Suppose there exists a k-algebra isomorphism $\sigma : F_1 \to F_2$. Show that F_1 and F_2 are equal as sets.

EXERCISE 5.2.20. Let $k = \mathbb{F}_2$ be the field of order 2. In $k[x]$, let $f = x^2$, $g = x^2 + x$, and $h = x^2 + x + 1$. Show that the following four rings are distinct in the sense that no two are isomorphic to each other: $\mathbb{Z}/(4)$, $k[x]/(f)$, $k[x]/(g)$, $k[x]/(h)$. For a continuation of this exercise, see Exercise 5.5.25.

EXERCISE 5.2.21. Let k be a field and A a finite dimensional k-algebra. Prove that if $\dim_k(A) = 2$, then A is commutative. For an example of a noncommutative k-algebra L such that $\dim_k(L) = 3$, see Exercise 4.4.21.

EXERCISE 5.2.22. True or False. Justify your answers.

(1) $\mathbb{Q}(\sqrt{2}) \cong \mathbb{Q}(\sqrt{3})$
(2) $\mathbb{R}(\sqrt{-2}) \cong \mathbb{R}(\sqrt{-3})$

EXERCISE 5.2.23. Let k be a field and $K = k(x)$ the field of rational functions over k in the variable x. Let $\sigma : K \to K$ be the function which maps a typical rational function $f(x) \in K$ to the rational function $f(x^{-1})$. Show that σ is an automorphism of the field K.

EXERCISE 5.2.24. Let R be a unique factorization domain with quotient field K. Assume $\text{char}(R) \neq 2$. Let F/K be a quadratic extension of fields. In other words, assume $\dim_K F = 2$. Show that there exists a square free element $a \in R$ such that $F = K[x]/(x^2 - a) = K(\sqrt{a})$.

EXERCISE 5.2.25. Let k be a field, f an irreducible polynomial in $k[x]$ and F the splitting field of f. Assume that the degree of f is at least 2 and that a and b are distinct roots of f in F. Show that there exists a k-algebra automorphism $\sigma \in \text{Aut}_k(F)$ such that $\sigma(a) = b$.

3. Galois Theory

In this section we study Galois Theory for fields. For the most part, we follow the traditional presentation which is attributed to Emil Artin [**2**].

Let k be a field, $f \in k[x]$ a separable polynomial, and F a splitting field for f over k. The roots of f are the solutions to the algebraic equation $f(x) = 0$. The field extension F/k is generated by the roots of f. As in Definition 4.4.1, by $\text{Aut}_k(F)$ we denote the group of all k-algebra automorphisms of F. In Theorem 5.3.18 we show that F/k is a so-called Galois extension. For a Galois extension, the group $\text{Aut}_k(F)$ acts not only on F, but on the set of roots of $f(x)$. Moreover the action of the group $\text{Aut}_k(F)$ on F is entirely determined by its action on the roots of $f(x)$. In the Fundamental Theorem of Galois Theory (Theorem 5.3.21), we show that there is a one-to-one correspondence between the intermediate fields of F/k and the subgroups of $\text{Aut}_k(F)$. By this theorem, the study of the roots of the polynomial equation $f(x) = 0$ is reduced to the study of the action of a finite group acting on the set of roots. It was Galois himself who emphasized the importance of studying the set of roots of a polynomial under the action by a finite group of permutations (see [**12**]).

3.1. A Group Acting on a Field. In this section we will be using some results as well as some terminology from Group Theory. For instance, if a group of permutations G acts on a set X, there is the well defined notion of the subset of X fixed by G. Also, for any subset S of X there is the subgroup of G fixing S. The reader is referred to Section 2.4.1, especially Definition 2.4.10. While the underlying theory applies, the notation and terminology in the present context are slightly different than that of Chapter 2. Proposition 5.3.1 extends to the context of field extensions these important notions from Group Theory.

PROPOSITION 5.3.1. *Let F/k be an extension of fields and $G = \mathrm{Aut}_k(F)$.*

(1) If H is a subset of G, then

$$F^H = \{v \in F \mid \sigma(v) = v \text{ for all } \sigma \in H\}$$

is an intermediate field of F/k which is called the fixed field of H. *Note that $H \subseteq \mathrm{Aut}_{F^H}(F)$.*

(2) If E is an intermediate field of F/k, then

$$G_E = \{\sigma \in G \mid \sigma(v) = v \text{ for all } v \in E\}$$

is a subgroup of G which is called the subgroup of G fixing E. Note that $G_E = \mathrm{Aut}_E(F)$.

PROOF. The proof is left to the reader. □

PROPOSITION 5.3.2. *Let F/k be an extension of fields.*

(1) Let $f \in k[x]$, $\sigma \in \mathrm{Aut}_k(F)$, and $u \in F$. If $f(u) = 0$, then $f(\sigma(u)) = 0$.

(2) Assume $u \in F$ is algebraic over k and $E = k[u]$. If $\sigma \in \mathrm{Aut}_k(E)$, then σ is completely determined by $\sigma(u)$.

PROOF. (1): If $f = \sum_{i=0}^{n} a_i x^i$, then

$$f(\sigma(u)) = \sum_{i=0}^{n} a_i(\sigma(u))^i = \sum_{i=0}^{n} \sigma(a_i u^i) = \sigma(\sum_{i=0}^{n} a_i u^i) = \sigma(f(u)) = \sigma(0) = 0.$$

(2): By Theorem 5.1.4, there is a k-basis for E of the form $1, u, u^2, \ldots, u^{n-1}$ where $n = \dim_k(E)$. □

EXAMPLE 5.3.3. Let $\mathbb{F}_2 = \{0, 1\}$ be the field of order 2, which is isomorphic to the ring $\mathbb{Z}/2$. Let $p(x) = x^2 + x + 1 \in \mathbb{F}_2[x]$. Since $p(0) = p(1) = 1$, $p(x)$ has no root in $\mathbb{F}_2$ and is irreducible in $\mathbb{F}_2[x]$. Let F be the splitting field of $p(x)$. Then F has order 4. Let α be a root of $p(x)$ in F. Then $\alpha^2 = \alpha + 1$ and by Theorem 5.1.4, $F = \{0, 1, \alpha, \alpha + 1\}$. By Theorem 5.2.12, F is unique up to isomorphism. Let $\phi \in \mathrm{Aut}(F)$. Then $\phi(0) = 0$, $\phi(1) = 1$ and $\phi(\alpha)$ is equal to α or $\alpha + 1$. If $\phi(\alpha) = \alpha$, then ϕ is equal to $1 \in \mathrm{Aut}(F)$, the identity function. By Proposition 5.3.2, ϕ is determined by the value of $\phi(\alpha)$. Therefore, $\mathrm{Aut}(F)$ has order at most 2. We prove that there is an automorphism of order two in $\mathrm{Aut}(F)$. By Exercise 3.2.32, the Frobenius homomorphism $\theta : F \to F$ defined by $\theta(a) = a^2$ is a homomorphism. Since F is a finite field, θ is necessarily one-to-one and onto (Exercises 3.2.29 and 1.1.11). Since $\theta(\alpha) = \alpha^2 = \alpha + 1$, we have shown that $\mathrm{Aut}(F)$ has order two.

EXAMPLE 5.3.4. The polynomial $p(x) = x^2 + 1$ is irreducible in $\mathbb{Q}[x]$. The roots of $p(x)$ in $\mathbb{C}$ are $i, -i$. Let $F = \mathbb{Q}(i) = \mathbb{Q}(i)$ be the splitting field for $p(x)$ over $\mathbb{Q}$. By Theorem 5.1.4, a basis for F over $\mathbb{Q}$ is $1, i$. By Corollary 5.1.7, there exists an automorphism $\chi : \mathbb{Q}(i) \to \mathbb{Q}(i)$ such that $\chi(i) = -i$. The automorphism

χ is usually called *complex conjugation* (see Section 1.4). By Proposition 5.3.2 (1), if $\phi \in \mathrm{Aut}_{\mathbb{Q}}(F)$, then $\phi(i)$ is equal to either i or $-i$. By Proposition 5.3.2 (2), this implies $\mathrm{Aut}_{\mathbb{Q}}(F)$ has order at most two. This proves $\mathrm{Aut}_{\mathbb{Q}}(F) = \langle\sigma\rangle$ is a cyclic group of order two.

EXAMPLE 5.3.5. Let $t > 2$ be an odd integer and $f(x) = x^t + 3x - 6$. By Theorem 3.7.6, f is irreducible over $\mathbb{Q}$. Consider $f : \mathbb{R} \to \mathbb{R}$ as a continuous real valued function. Since $f'(x) = tx^{t-1} + 3$ is a sum of squares in $\mathbb{R}$, $f'(x) > 0$ and f is increasing. Therefore, f has exactly one real root, call it u. The other $t-1$ roots of f are nonreal. Then $\mathbb{Q}(u)$ is a subfield of $\mathbb{R}$ and $\mathbb{Q}(u)$ contains exactly one root of f. If $\sigma \in \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(u))$, then by Proposition 5.3.2, $\sigma(u) = u$ and σ is the identity function. This shows $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(u)) = \langle 1\rangle$.

LEMMA 5.3.6. *Let F/k be an extension of fields and $\mathrm{Hom}_k(F,F)$ the ring of k-linear endomorphisms of F.*

(1) There is a one-to-one homomorphism of rings $F \to \mathrm{Hom}_k(F,F)$ defined by $a \mapsto \ell_a$, where ℓ_a is "left multiplication by a". That is, $\ell_a(x) = ax$.
(2) The automorphism group $\mathrm{Aut}_k(F)$ is a subgroup of the group of units of the ring $\mathrm{Hom}_k(F,F)$.
(3) If $\dim_k(F) = n$ is finite, then $\mathrm{Hom}_k(F,F)$ is a finite dimensional F-vector space of dimension n.

PROOF. (1): By Exercise 4.1.26 there is a commutative diagram of k-algebra homomorphisms

$$\begin{array}{ccc} k & \xrightarrow{\lambda_k} & \mathrm{Hom}_k(F,F) \\ {\scriptstyle\theta}\downarrow & & \uparrow{\scriptstyle H_\theta} \\ F & \xrightarrow{\lambda_F} & \mathrm{Hom}_F(F,F) \end{array}$$

where λ_k and λ_F are the left regular representations of Example 4.1.17. All of the maps are one-to-one.

(2): Since each $\sigma \in \mathrm{Aut}_k(F)$ is a k-linear transformation of F, we view $\mathrm{Aut}_k(F)$ as a subgroup of the group of units of the ring $\mathrm{Hom}_k(F,F)$.

(3): By Proposition 4.5.4, $\mathrm{Hom}_k(F,F)$ is a k-vector space of dimension n^2. By Proposition 4.3.8, $\mathrm{Hom}_k(F,F)$ is an F-vector space of dimension n. □

Theorems 5.3.7, 5.3.9, and 5.3.11 play key roles in Galois Theory. The proof we give below of the Fundamental Theorem (Theorem 5.3.21) relies heavily on these three theorems. The statements of the theorems concern a field and its group of automorphisms. Because the topic of this section is Galois Theory, this is to be expected. But, upon a close look at the proofs of these results, the common feature that stands out is that they are strictly of a linear algebra nature. From this perspective, we see that Galois Theory is inherently a part of Linear Algebra.

THEOREM 5.3.7. *Let F be a field and $\sigma_1, \ldots, \sigma_n$ a finite set of distinct automorphisms of F. If $u_1, \ldots, u_n$ are elements of F and*

$$u_1\sigma_1(x) + \cdots + u_n\sigma_n(x) = 0 \tag{3.1}$$

for all $x \in F$, then each u_i is equal to zero.

PROOF. For sake of contradiction assume a nontrivial dependence relation of the type (3.1) exists. Pick one such relation involving a minimal number of the automorphisms. If necessary, relabel the automorphisms and assume

$$u_1\sigma_1 + \cdots + u_r\sigma_r = 0 \tag{3.2}$$

where $u_1, \ldots, u_r$ are all nonzero and r is minimal. Since $\sigma_i(1) = 1$ for each i, in (3.2) we have $r \geq 2$. For some $y \in F$ we have $\sigma_1(y) \neq \sigma_r(y)$. Evaluating (3.2) at yx, we have:

$$u_1\sigma_1(y)\sigma_1(x) + \cdots + u_r\sigma_r(y)\sigma_r(x) = 0 \tag{3.3}$$

for all $x \in F$. Multiplying (3.2) by $\sigma_r(y)$, we have:

$$u_1\sigma_r(y)\sigma_1(x) + \cdots + u_r\sigma_r(y)\sigma_r(x) = 0 \tag{3.4}$$

for all $x \in F$. Subtracting (3.3) and (3.4), we have:

$$u_1(\sigma_1(y) - \sigma_r(y))\sigma_1(x) + \cdots + u_{r-1}(\sigma_{r-1}(y) - \sigma_r(y))\sigma_{r-1}(x) = 0$$

which is a shorter dependence relation, a contradiction. □

COROLLARY 5.3.8. *Let F/k be an extension of fields. Then* $\mathrm{Aut}_k(F)$ *is a linearly independent subset of the F-vector space* $\mathrm{Hom}_k(F, F)$.

PROOF. This follows from Lemma 5.3.6, Theorem 5.3.7, and Definition 4.2.12. □

THEOREM 5.3.9. *Let F/k be a finite dimensional extension of fields. Then the order of the group of automorphisms* $\mathrm{Aut}_k(F)$ *is less than or equal to* $\dim_k(F)$.

We present two proofs of Theorem 5.3.9. Both are based on Theorem 5.3.7. The first is deceptively brief. The second is more traditional.

FIRST PROOF OF THEOREM 5.3.9. Let $n = \dim_k(F)$. By Lemma 5.3.6, the ring $\mathrm{Hom}_k(F, F)$ is an F-vector space of dimension n. By Corollary 5.3.8, $\mathrm{Aut}_k(F)$ is a linearly independent subset of the F-vector space $\mathrm{Hom}_k(F, F)$. By Theorem 4.3.4, $\mathrm{Aut}_k(F)$ has order less than or equal to n. □

SECOND PROOF OF THEOREM 5.3.9. If $\mathrm{Aut}_k(F) = \langle 1 \rangle$, then there is nothing to prove. Let $r = \dim_k(F)$. For sake of contradiction assume $\mathrm{Aut}_k(F)$ contains a set of $r+1$ distinct automorphisms, which we enumerate: $\sigma_0, \ldots, \sigma_r$. Let $v_1, \ldots, v_r$ be a basis for F as a k-vector space. By Theorem 4.3.4, the $r+1$ vectors

$$\begin{aligned} x_0 &= \big(\sigma_0(v_1), \sigma_0(v_2), \ldots, \sigma_0(v_r)\big), \\ x_1 &= \big(\sigma_1(v_1), \sigma_1(v_2), \ldots, \sigma_1(v_r)\big), \\ &\vdots \\ x_r &= \big(\sigma_r(v_1), \sigma_r(v_2), \ldots, \sigma_r(v_r)\big), \end{aligned}$$

in F^r are linearly dependent over F. Hence there exists a nonzero vector in F^{r+1}, call it $(c_0, c_1, \ldots, c_r)$, such that

$$c_0\sigma_0(v_j) + c_1\sigma_1(v_j) + \cdots + c_r\sigma_r(v_j) = 0 \tag{3.5}$$

for $j = 1, \ldots, r$. Let u be an arbitrary element of F. In terms of the k-basis, u has a representation $u = a_1 v_1 + \cdots + a_r v_r$ for unique $a_1, \ldots, a_r$ in k. For each σ_i we have:

$$\sigma_i(u) = a_1\sigma_i(v_1) + \cdots + a_r\sigma_i(v_r). \tag{3.6}$$

Consider:

$$\begin{aligned} \sum_{i=0}^{r} c_i\sigma_i(u) &= \sum_{i=0}^{r} c_i \left(a_1\sigma_i(v_1) + \cdots + a_r\sigma_i(v_r)\right) \\ &= \sum_{j=1}^{r} a_j \left(\sum_{i=0}^{r} c_i\sigma_i(v_j)\right) \\ &= 0 \end{aligned} \tag{3.7}$$

where the last equation follows from (3.5). Since u was arbitrary, (3.7) is a contradiction to Theorem 5.3.7. □

EXAMPLE 5.3.10. Let $F = \mathbb{F}_q$ be a finite field with order q and $\operatorname{char}(F) = p$. If $k = \mathbb{F}_p$ is the prime subfield and $\dim_k(F) = n$, then $q = p^n$. By Exercise 3.2.32, the Frobenius homomorphism $\theta : F \to F$ defined by $\theta(x) = x^p$ is a homomorphism. Since F is a finite field, θ is necessarily one-to-one and onto (Exercises 3.2.29 and 1.1.11). Let α be a generator for the group of units of F (Corollary 3.6.12). Then in F^*, the order of α is $|\alpha| = p^n - 1$. Therefore $\alpha^{p^n} = \alpha$ and if $1 < i < p^n$, then $\alpha^i \neq \alpha$. It follows from $\theta(\alpha) = \alpha^p \neq \alpha$, $\theta^2(\alpha) = \theta(\alpha^p) = (\alpha^p)^p = \alpha^{p^2} \neq \alpha$, $\ldots$, $\theta^i(\alpha) = \alpha^{p^i} \neq \alpha$, $\ldots$, $\theta^n(\alpha) = \alpha^{p^n} = \alpha$ that θ has order n in $\operatorname{Aut}(F)$. By Theorem 5.3.9, $\operatorname{Aut}(F)$ is cyclic of order n and the Frobenius homomorphism θ is a generator.

If G is a group and H is a subgroup, the index of H in G is denoted $[G : H]$. The order of G is $[G : 1]$.

THEOREM 5.3.11. *Let F/k be an extension of fields, G a finite subgroup of $\operatorname{Aut}_k(F)$, and $K = F^G$. Then F/K is finite dimensional and $\dim_K(F) \leq [G : 1]$.*

PROOF. Assume $[G : 1] = n$ and $G = \{\sigma_1, \ldots, \sigma_n\}$. For sake of contradiction, assume the statement of the theorem is false. By Exercise 4.3.17, there exists a subset $\{v_0, \ldots, v_n\} \subseteq F$ which is linearly independent over K. By Theorem 4.3.4, the $n + 1$ vectors

$$\begin{aligned} x_0 &= \big(\sigma_1(v_0), \sigma_2(v_0), \ldots, \sigma_n(v_0)\big), \\ x_1 &= \big(\sigma_1(v_1), \sigma_2(v_1), \ldots, \sigma_n(v_1)\big), \\ &\vdots \\ x_n &= \big(\sigma_1(v_n), \sigma_2(v_n), \ldots, \sigma_n(v_n)\big) \end{aligned}$$

in F^n are linearly dependent over F. Let V be the subspace of F^n spanned by $X = \{x_0, x_1, \ldots, x_n\}$. Then $\dim_F(V) \leq n$ so a linearly independent subset of X has cardinality at most n. By Corollary 4.3.6, there is a linearly independent subset of X that is a spanning set for V. If necessary, reorder the vectors in X such that x_0 is in the linear span of $\{x_1, \ldots, x_n\}$. If c_0 is an arbitrary element of F, then

there exist n elements $c_1, \ldots, c_n$ in F such that $0 = c_0x_0 + c_1x_1 + \cdots + c_nx_n$. This is equivalent to

$$0 = \sum_{i=0}^{n} c_i\sigma_j(v_i) \tag{3.8}$$

for $j = 1, \ldots, n$. For each $i = 0, \ldots, n$, consider

$$a_i = \sigma_1(c_i) + \cdots + \sigma_n(c_i).$$

By Theorem 5.3.7, $\sigma_1, \ldots, \sigma_n$ are linearly independent so we can find c_0 in F such that $a_0 \neq 0$. By the comment above, we can pick $c_1, \ldots, c_n$ so that (3.8) holds for $j = 1, \ldots, n$. Since G is a group,

$$\begin{aligned}\sigma_j(a_i) &= \sigma_j\sigma_1(c_i) + \cdots + \sigma_j\sigma_n(c_i)\\ &= \sigma_1(c_i) + \cdots + \sigma_n(c_i)\\ &= a_i\end{aligned}$$

implies $a_i \in K = F^G$, for $i = 0, 1, \ldots, n$. Consider

$$\begin{aligned}\sum_{i=0}^{n} a_iv_i &= \sum_{i=0}^{n}\left(\sum_{j=1}^{n}\sigma_j(c_i)\right)v_i\\ &= \sum_{i=0}^{n}\left(\sum_{j=1}^{n}\sigma_j(c_i)\sigma_j(\sigma_j^{-1}(v_i))\right)\\ &= \sum_{j=1}^{n}\sigma_j\left(\sum_{i=0}^{n}c_i\sigma_j^{-1}(v_i)\right)\\ &= 0\end{aligned} \tag{3.9}$$

where the last 0 is from (3.8). The left hand side of (3.9) is a nontrivial K-linear combination of $v_0, v_1, \ldots, v_n$. This is a contradiction. □

3.2. Galois Extensions. In this section useful necessary and sufficient conditions for an extension of fields F/k to be a Galois extension are derived. As an application, in Corollary 5.3.20 we prove the important result that any finite separable extension can be embedded as an intermediate field of a Galois extension.

DEFINITION 5.3.12. Let F/k be an extension of fields and G a finite subgroup of $\mathrm{Aut}_k(F)$. If $k = F^G$, then we say F/k is a *Galois* extension with Galois group G. We also say F is a G-Galois extension of k.

PROPOSITION 5.3.13. *Let F/k be an extension of fields and G a finite subgroup of $\mathrm{Aut}_k(F)$.*

(1) If $K = F^G$, then $\dim_K(F) = [G : 1]$, $G = \mathrm{Aut}_K(F)$, and F is a G-Galois extension of K.
(2) If F is a G-Galois extension of k, then $\dim_k(F) = [G : 1]$ and $G = \mathrm{Aut}_k(F)$.

PROOF. (1): Since $K = F^G$, by Proposition 5.3.1, we have G is a subgroup of $\mathrm{Aut}_K(F)$. By Lemma 5.3.6 there is a commutative diagram

$$\begin{array}{ccccc} & & \mathrm{Aut}_k(F) & \longrightarrow & \mathrm{Hom}_k(F,F) \\ & & \uparrow & & \uparrow \\ G & \longrightarrow & \mathrm{Aut}_K(F) & \longrightarrow & \mathrm{Hom}_K(F,F) \end{array}$$

where all of the maps are one-to-one. The left vertical arrow is a homomorphism of groups, the right vertical arrow is a homomorphism of rings. By Theorem 5.3.11, we have $\dim_K(F) \leq [G:1]$. By Theorem 5.3.9, we have $[G:1] \leq [\mathrm{Aut}_K(F):1] \leq \dim_K(F)$.

(2): Since $k = F^G$, this is a special case of (1). □

PROPOSITION 5.3.14. *Let F be a G-Galois extension of k and $\alpha \in F$. The subgroup of G fixing α is denoted G_α (Definition 2.4.10). If $G_\alpha = \langle 1 \rangle$, then $F = k(\alpha)$.*

PROOF. Let $f = \min.\mathrm{poly}_k(\alpha)$. The orbit of α under the group G is $R = \{\sigma(\alpha) \mid \sigma \in G\}$. By Theorem 2.4.11, the length of the orbit is equal to the index of the stabilizer. That is, $|R| = [G:1]$. By Proposition 5.3.2, every element of R is a root of f. So $\deg f \geq [G:1]$. By Theorem 5.1.4, $\dim_k k(\alpha) = \deg f$. By Proposition 5.3.13, all of the numbers in the string of inequalities:

$$[G:1] \leq \deg f = \dim_k k(\alpha) \leq \dim_k(F)$$

are equal. Hence $k(\alpha) = F$. □

PROPOSITION 5.3.15. *Let F/k be a finite dimensional extension of fields and $\sigma_1, \ldots, \sigma_n$ a finite set of distinct automorphisms in $\mathrm{Aut}_k(F)$. If $\dim_k(F) = n$, then F/k is Galois with group $\mathrm{Aut}_k(F) = \{\sigma_1, \ldots, \sigma_n\}$.*

PROOF. We have $\{\sigma_1, \ldots, \sigma_n\} \subseteq \mathrm{Aut}_k(F)$, hence $n \leq [\mathrm{Aut}_k(F):1]$. By Theorem 5.3.9, $n = \dim_k(F) \geq [\mathrm{Aut}_k(F):1] \geq n$. Therefore, $\mathrm{Aut}_k(F) = \{\sigma_1, \ldots, \sigma_n\}$. In particular, this proves the set $\{\sigma_1, \ldots, \sigma_n\}$ is a group. For notational simplicity, let $G = \mathrm{Aut}_k(F)$ and $K = F^G$. By Proposition 5.3.13, F/K is a G-Galois extension and $\dim_K(F) = n$. By Exercise 5.1.23 applied to the tower of fields: $k \subseteq K \subseteq F$, we conclude that $k = K = F^G$. □

EXAMPLE 5.3.16. Let $F = \mathbb{F}_q$ be a finite field with characteristic $\mathrm{char}\,(F) = p$ and order q. If $k = \mathbb{F}_p$ is the prime subfield and $\dim_k(F) = n$, then $q = p^n$. By Example 5.3.10, $\mathrm{Aut}_k(F) = \langle \theta \rangle$ is cyclic of order n where θ is the Frobenius homomorphism defined by $\theta(x) = x^p$. By Proposition 5.3.15, $\mathbb{F}_q/\mathbb{F}_p$ is Galois with cyclic Galois group.

DEFINITION 5.3.17. Let F/k be an algebraic extension of fields. We say F/k is a *normal* extension if every irreducible polynomial over k that has a root in F actually splits over F.

Theorem 5.3.18 provides very useful necessary and sufficient conditions for an extension of fields to be Galois.

THEOREM 5.3.18. *Let F/k be a finite dimensional extension of fields. The following are equivalent.*

(1) F/k is a Galois extension.
(2) F/k is normal and separable.
(3) F is the splitting field over k of a separable polynomial in $k[x]$.

PROOF. (1) implies (2): Suppose F/k is Galois with group $G = \{\sigma_1, \ldots, \sigma_n\}$. We prove F/k is normal and separable. Let $f(x) \in k[x]$ be an irreducible polynomial and suppose $u \in F$ is a root of f. Look at the orbit of α under G: $R = \{\sigma_1(u), \ldots, \sigma_n(u)\}$. Suppose R has r elements which we enumerate: $R = \{u_1, \ldots, u_r\}$. Then G acts as a group of permutations of R. The polynomial $g(x) = (x - u_1)(x - u_2)\cdots(x - u_r)$ is in $F[x]$ and is fixed by every element of G. Since $k = F^G$, we have $g(x) \in k[x]$. Now $u \in R$, so $g(u) = 0$. Since $f(x)$ is the irreducible polynomial of u, by Theorem 5.1.4 we have $f \mid g$. This proves f splits over F. Since g is separable, so is f. We have proved that F/k is normal. Let v be an arbitrary element of F. Then by the previous argument, $\text{min. poly}_k(v)$ is separable. This proves F/k is separable.

(2) implies (1): By Theorem 5.2.14, $F = k(\alpha)$ for some $\alpha \in F$. If $f = \text{Irr. poly}_k(\alpha)$, then f is separable and splits over F. If $n = \deg(f)$, then by Theorem 5.1.4, $n = \dim_k(F)$. Let $\alpha_1, \ldots, \alpha_n$ be the distinct roots of f in F. Then for each i we have $f = \text{Irr. poly}_k(\alpha_i)$. Since $k(\alpha_i)$ is an intermediate field of F/k and $\dim_k k(\alpha_i) = \dim_k F$, we have $F = k(\alpha_i)$. By Corollary 5.1.7 there is a k-automorphism $\sigma_i : F \to F$ such that $\sigma_i(\alpha) = \alpha_i$. By Proposition 5.3.2 (2), $\sigma_1, \ldots, \sigma_n$ are distinct elements of $\text{Aut}_k(F)$. By Proposition 5.3.15, F/k is Galois.

(2) implies (3): By Theorem 5.2.14, The Primitive Element Theorem, $F = k(\alpha)$ for some $\alpha \in F$. If $f = \text{Irr. poly}_k(\alpha)$, then f is separable and splits in F.

(3) implies (1): Suppose $f \in k[x]$ is separable and F is the splitting field for f over k. Let $n = \dim_k(F)$. If $n = 1$, then $F = k$, so F/k is Galois with group $\langle 1 \rangle$. Inductively, assume $n > 1$ and that (3) implies (1) for any extension of fields of dimension less than n. Let $G = \text{Aut}_k(F)$. To finish the proof, we show $F^G = k$. Let g be a monic irreducible factor of the polynomial f and assume $\deg g = d > 1$. Since g is separable and splits in F, there are d distinct roots $\alpha_1, \ldots, \alpha_d$ in F and $g = (x - \alpha_1)\cdots(x - \alpha_d)$. Now $k(\alpha_1)$ is an intermediate field of F/k and F is a splitting field of the separable polynomial f over $k(\alpha_1)$. By the induction hypothesis, we can assume $F/k(\alpha_1)$ is a Galois extension with group H and $[H : 1] = \dim_{k(\alpha_1)}(F)$. By Corollary 5.1.7, for each i, there is a k-algebra isomorphism $\sigma_i : k(\alpha_1) \to k(\alpha_i)$. By Lemma 5.2.7, each σ_i extends to an automorphism also denoted σ_i, in $G = \text{Aut}_k(F)$. Let θ be an arbitrary element of F^G. Since H is a subgroup of $G = \text{Aut}_k(F)$, $\theta \in F^H = k(\alpha_1)$. By Theorem 5.1.4(3) there are $c_0, c_1, \ldots, c_{d-1}$ in k such that

$$\theta = c_0 + c_1\alpha_1 + \cdots + c_{d-1}\alpha_1^{d-1}. \tag{3.10}$$

Applying σ_i to (3.10) we have

$$\theta = c_0 + c_1\alpha_i + \cdots + c_{d-1}\alpha_i^{d-1} \tag{3.11}$$

since θ is fixed by G. Let $p(x) = (c_0 - \theta) + c_1x + \cdots + c_{d-1}x^{d-1} \in k(\alpha_i)[x]$. Then in F, there are d distinct roots $\alpha_1, \ldots, \alpha_d$ of $p(x)$. Since $\deg p(x) \leq d - 1$, we must have $p = 0$. In particular, $\theta = c_0$ is in k. □

COROLLARY 5.3.19. *Let k be a field, f a separable polynomial in $k[x]$, and F a splitting field for f over k. If $\alpha_1, \ldots, \alpha_n$ are the distinct roots of f in F, then the following are true:*

(1) F/k is a Galois extension with group $G = \mathrm{Aut}_k(F)$.
(2) G acts as a group of permutations of the set $\{\alpha_1, \ldots, \alpha_n\}$.
(3) G is isomorphic to a subgroup of S_n, the symmetric group on n letters.

PROOF. By Theorem 5.3.18, F/k is Galois with group $G = \mathrm{Aut}_k(F)$. By Proposition 5.3.2, every $\sigma \in G$ is a permutation of the set of roots of f. By Lemma 2.4.1, there is a homomorphism $\theta : G \to S_n$. Since $F = k(\alpha_1, \ldots, \alpha_n)$, if two automorphisms define the same permutation of $\alpha_1, \ldots, \alpha_n$, they define the same automorphism of F. This proves θ is one-to-one. □

COROLLARY 5.3.20. *(Embedding Theorem for Fields) Let F/k be a finite dimensional extension of fields. If F/k is separable, then there exists a finite dimensional Galois extension K/k which contains F as an intermediate field.*

PROOF. Pick a finite set of separable elements $u_1, \ldots, u_n$ that generate F/k. For each i, if $f_i = \mathrm{Irr.poly}_k(u_i)$, then f_i is separable over k. Let K be the splitting field for $f_1 \cdots f_n$ over k. So K contains a generating set for F, hence F is an intermediate field of K/k. By Theorem 5.3.18, K/k is a Galois extension. □

3.3. The Fundamental Theorem of Galois Theory. In this section, we prove the Fundamental Theorem of Galois Theory. To illustrate the theorem, nontrivial examples are given for which the Galois group is completely determined.

THEOREM 5.3.21. *(The Fundamental Theorem of Galois Theory) Let F/k be a Galois extension of fields with finite group G. There is a one-to-one order inverting correspondence between the subgroups H of G and the intermediate fields E of F/k. A subgroup H corresponds to the fixed field F^H. An intermediate field E corresponds to the subgroup of G fixing E, G_E. If E is an intermediate field of F/k, then*
(1) $\dim_E(F) = [G_E : 1]$, $\dim_k(E) = [G : G_E]$, $G_E = \mathrm{Aut}_E(F)$,
(2) F/E is a Galois extension with group G_E, and
(3) E/k is a Galois extension if and only if G_E is a normal subgroup of G and in this case, $G/G_E \cong \mathrm{Aut}_k(E)$.

PROOF. By Proposition 5.3.1 there are functions

$$\{H \mid H \text{ is a subgroup of } G\} \underset{\lambda}{\overset{\rho}{\rightleftarrows}} \{E \mid E \text{ is an intermediate field of } F/k\}$$

defined by $\rho(H) = F^H$ and $\lambda(E) = G_E$. It is clear that if $H_1 \subseteq H_2$, then $\rho(H_1) \supseteq \rho(H_2)$. Likewise, if $E_1 \subseteq E_2$, then $\lambda(E_1) \supseteq \lambda(E_2)$. Suppose A and B are two subgroups of G such that $F^A = F^B$. Let $E = F^A = F^B$. Proposition 5.3.13 says $\dim_E(F) = [A : 1] = [B : 1]$. For contradiction's sake, suppose there exists $\sigma \in B - A$. Then $F^\sigma \supseteq F^B = F^A$. If we write $H = \{\sigma\} \cup A$, then $F^H = F^A \cap F^\sigma = F^A$. This contradicts Theorem 5.3.9. So $B \subseteq A$. Similarly, $A \subseteq B$. This shows ρ is one-to-one. Let E be an intermediate field of F/k. Since F/k is Galois, by Theorem 5.3.18, F is the splitting field of a separable polynomial f in $k[x]$. Then f is a separable polynomial in $E[x]$ and F is the splitting field of f over E. Since F/E is finite dimensional, Theorem 5.3.18 implies F/E is Galois. By Proposition 5.3.2, $G_E = \mathrm{Aut}_E(F)$ is the subgroup of $\mathrm{Aut}_k(F)$ fixing E and $\dim_E(F) = [G_E : 1]$. This implies $E = \rho\lambda(E)$, so ρ is a one-to-one correspondence. By Lagrange's Theorem (Corollary 2.2.14), $[G : 1] = [G : G_E][G_E : 1]$. By Exercise 5.1.23 $\dim_k(F) = \dim_k(E) \dim_E(F)$. This says $\dim_k(E) = [G : G_E]$. We have proved (1) and (2).

The rest of the proof is devoted to proving (3). Assume E/k is Galois. We prove that $G_E = \mathrm{Aut}_E(F)$ is a normal subgroup of $G = \mathrm{Aut}_k(F)$ and $\mathrm{Aut}_k(E)$ is isomorphic to the quotient G/G_E. First we show that there is a homomorphism of groups:

$$G = \mathrm{Aut}_k(F) \xrightarrow{h} \mathrm{Aut}_k(E)$$

defined by $\phi \mapsto \phi|_E$. The binary operation in both groups is composition of functions, so it suffices to show that if $\phi \in G$, then $\phi(E) = E$. By Theorem 5.2.14, $E = k(\xi)$ is a simple extension. Say $g(x) = \mathrm{min.poly}_k(\xi)$ and $\deg g = m$. Since E/k is normal, g splits over E and has m distinct roots in E, call them $\xi_1, \ldots, \xi_m$. Given $\phi \in \mathrm{Aut}_k(F)$, $\phi(\xi) = \xi_j$ for some j, by Proposition 5.3.2. Therefore, $\phi(E) = \phi(k(\xi)) \subseteq E$. Since ϕ is one-to-one, $\phi(E) = E$ by, Exercise 4.3.11. From this is follows that $\phi|_E$ is an automorphism of E, and h is a homomorphism of groups. The kernel of h is G_E, the set of all $\phi \in G$ fixing E. Therefore, $G_E = \mathrm{Aut}_E(F)$ is a normal subgroup of $G = \mathrm{Aut}_k(F)$. To show that $\mathrm{Aut}_k(E)$ is isomorphic to the quotient G/G_E, it suffices to show h is onto (Theorem 2.3.14). We are given that E/k is a Galois extension. This and (1) yield $[\mathrm{Aut}_k(E) : 1] = \dim_k(E) = [G : G_E]$. Theorem 2.3.14 and Lagrange's Theorem (Corollary 2.2.14) yield: $[G : G_E] = [\mathrm{im}\, h : 1]$. Therefore, $[\mathrm{im}\, h : 1] = [\mathrm{Aut}_k(E) : 1]$. Since the groups are finite, h is onto.

Conversely, assume $G_E = \mathrm{Aut}_E(F)$ is a normal subgroup of G and prove E/k is Galois. First we show that there is a homomorphism of groups

$$G = \mathrm{Aut}_k(F) \xrightarrow{h} \mathrm{Aut}_k(E)$$

defined by $\psi \mapsto \psi|_E$. To show that h is well defined, we use the fact that $\psi^{-1}\,\mathrm{Aut}_E(F)\psi = \mathrm{Aut}_E(F)$ (Lemma 2.3.5). Let $\phi \in \mathrm{Aut}_E(F)$. Then $\psi^{-1}\phi\psi = \phi_1 \in \mathrm{Aut}_E(F)$. Let y be an arbitrary element of E. Then $\psi^{-1}\phi\psi(y) = \phi_1(y) = y$. Therefore, $\phi\psi(y) = \psi(y)$. This shows $\psi(y)$ is fixed by each ϕ in $\mathrm{Aut}_E(F)$. By (2), this means $\psi(y) \in E$, hence h is well defined. The kernel of h is $G_E = \mathrm{Aut}_E(F)$, the subgroup of $\mathrm{Aut}_k(F)$ fixing E. By Theorem 2.3.12, the diagram

$$\begin{array}{ccc} \mathrm{Aut}_k(F) & \xrightarrow{\quad h \quad} & \mathrm{Aut}_k(E) \\ & {\scriptstyle\eta}\searrow \quad \nearrow{\scriptstyle\bar{h}} & \\ & \mathrm{Aut}_k(F)/\,\mathrm{Aut}_E(F) & \end{array}$$

commutes and $\bar{h}$ is one-to-one. From (1) and Lagrange's Theorem,

$$\begin{aligned} \dim_k(E) &= \dim_k(F)/\dim_E(F) \\ &= [\mathrm{Aut}_k(F) : 1]/[\mathrm{Aut}_E(F) : 1] \\ &= [\mathrm{im}(h) : 1] \\ &\leq [\mathrm{Aut}_k(E) : 1] \end{aligned}$$

By Proposition 5.3.15, E/k is Galois. □

EXAMPLE 5.3.22. This is an example of a Galois extension of $\mathbb{Q}$ with Galois group the full symmetric group S_p. Let p be a prime number and $f \in \mathbb{Q}[x]$ an irreducible polynomial of degree p such that f has exactly two nonreal roots. In this example we show that the Galois group of f is isomorphic to S_p, the symmetric group on p letters. Let F be the splitting field for f in $\mathbb{C}$. By Theorem 5.3.18, F is

Galois over $\mathbb{Q}$. Let a and b be the nonreal roots of f. If $p = 2$, then $F = \mathbb{Q}(a)$ has degree two over $\mathbb{Q}$ and $\mathrm{Aut}_{\mathbb{Q}}(F)$ has order two hence is isomorphic to S_2. Assume $p > 2$ and let c be a real root of f. Then $\dim_{\mathbb{Q}} \mathbb{Q}(c) = p$ and by Theorem 5.3.21, p divides the order of $\mathrm{Aut}_{\mathbb{Q}}(F)$. By Cauchy's Theorem (Corollary 2.4.15), $\mathrm{Aut}_{\mathbb{Q}}(F)$ contains an element σ of order p. By Corollary 5.3.19, we know that $\mathrm{Aut}_{\mathbb{Q}}(F)$ is a group of permutations of the roots of f. By Corollary 2.6.4 we know that σ is a p-cycle and can be written in the form $\sigma = (s_1 s_2 \cdots s_p)$. For some i and j we have $a = s_i$ and $b = s_j$. Then $\sigma^{j-i}(s_i) = s_j$. By Lemma 2.2.18, the order of σ^{j-i} is p. Therefore, we can write σ^{j-i} in the cycle form $(abt_3 \cdots t_p)$. Let χ be the automorphism of $\mathbb{C}$ defined by complex conjugation (Example 5.3.4). Then χ maps F to F. Also, $\chi(a) = b$ and χ fixes every real root of f. So χ corresponds to the transposition $\chi = (ab)$. By Exercise 2.6.20, the group S_p is generated by the transposition (12) and the p-cycle $(123 \cdots p)$. Therefore, $\mathrm{Aut}_{\mathbb{Q}}(F)$ is generated by χ and σ^{j-i}, hence is isomorphic to S_p. For the polynomial f, the upper bound of Corollary 5.2.6 is attained.

EXAMPLE 5.3.23. In $\mathbb{Q}[x]$, let $p(x) = x^3 + 2x + 1$. In Example 5.1.9 we saw that $p(x)$ is irreducible, has one real root α and two nonreal roots β_1, β_2. The splitting field of $p(x)$ over $\mathbb{Q}$ is $F = \mathbb{Q}(\alpha)(\beta_1)$. By Example 5.3.22, the Galois group $G = \mathrm{Aut}_{\mathbb{Q}}(F)$ is isomorphic to S_3. Under the Galois correspondence, the three subgroups of S_3 of order two correspond to the three intermediate fields $\mathbb{Q}(\alpha)$, $\mathbb{Q}(\beta_1)$, and $\mathbb{Q}(\beta_2)$. Since S_3 has one subgroup of order three, there is an intermediate field of $F/\mathbb{Q}$ that has dimension two over $\mathbb{Q}$. For completeness' sake, we give a formula for this quadratic extension. Set $\Delta = (\alpha - \beta_1)(\alpha - \beta_2)(\beta_1 - \beta_2)$. Identify G with the symmetric group S_3. Given any $\sigma \in S_3$, $\sigma(\Delta) = \mathrm{sign}(\sigma)\Delta$. Therefore, $\Delta \notin \mathbb{Q}$, $\Delta^2 \in \mathbb{Q}$. Consequently, $\mathbb{Q}(\Delta)$ is an intermediate field of $F/\mathbb{Q}$ and has dimension two over $\mathbb{Q}$. The formula for Δ is an example of a discriminant formula. See Section 5.7.4 for the general construction.

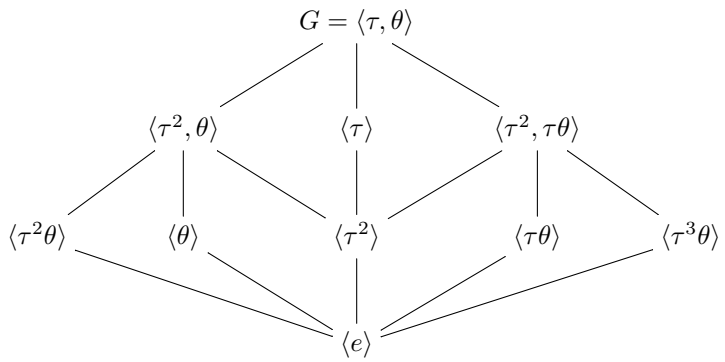
EXAMPLE 5.3.24. In $\mathbb{Q}[x]$, let $f(x) = x^4 - 2$. Let u be the positive real number such that $u^4 = 2$ and let $i \in \mathbb{C}$ be a root of $x^2 + 1$. Then the four roots of $f(x)$ in $\mathbb{C}$ are

$$\{u, -u, ui, -ui\}. \tag{3.12}$$

Let $F = \mathbb{Q}(u, ui)$ be the splitting field of f over $\mathbb{Q}$. Then F is a Galois extension of $\mathbb{Q}$ with group $G = \mathrm{Aut}_{\mathbb{Q}}(F)$. The group G is called the Galois group of f. First we determine generators for the Galois group G. The elements of G are described as permutations of the set (3.12) and G is identified with a subgroup of S_4. The subgroup lattice of G is then presented. Lastly, the corresponding lattice of subfields is determined.

By Theorem 5.1.4, $(\mathbb{Q}(u) : \mathbb{Q}) = (\mathbb{Q}(ui) : \mathbb{Q}) = 4$. Since $u \in \mathbb{R}$ is real and ui is nonreal, we know $\mathbb{Q}(u) \neq \mathbb{Q}(ui)$. Over $\mathbb{Q}(u^2)$ we have the factorization $f = (x^2 - u^2)(x^2 + u^2)$ into irreducibles. The irreducible polynomial for ui over $\mathbb{Q}(u)$ is $x^2 + u^2$. The irreducible polynomial for u over $\mathbb{Q}(ui)$ is $x^2 - u^2$. Then $(F : \mathbb{Q}(u)) = (F : \mathbb{Q}(ui)) = 2$. By Corollary 5.1.7, there is an isomorphism $\sigma : \mathbb{Q}(u) \to \mathbb{Q}(ui)$ which is given by $\sigma(u) = ui$. By Lemma 5.2.7, σ can be extended to an isomorphism $F = \mathbb{Q}(u)(ui) \to \mathbb{Q}(ui)(u) = F$ which is defined by sending ui to one of u or $-u$. Let τ be the automorphism of F defined by $\tau(u) = ui$, $\tau(ui) = -u$. Let θ be the automorphism of F defined by $\theta(u) = ui$, $\theta(ui) = u$. By Theorem 5.3.18, F is Galois over $\mathbb{Q}$ with group $G = \mathrm{Aut}_{\mathbb{Q}}(F)$. By Exercise 5.3.32, we can view G

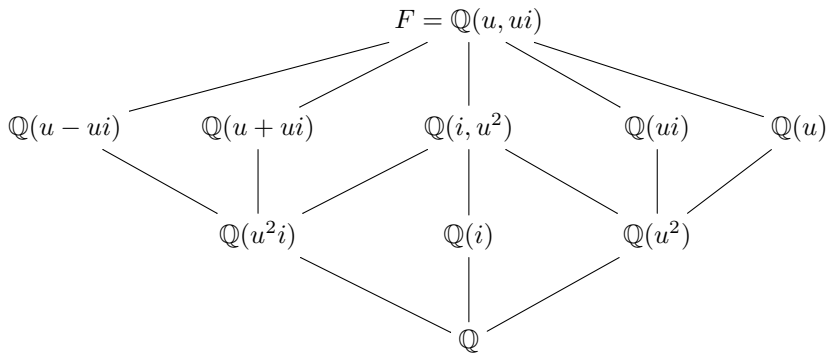
as a subgroup of S_4. Using the ordering of the roots given in (3.12), the cycle representations of τ and θ are $\tau = (1324)$, $\theta = (13)(24)$. We can now compute the elements of G: $\langle e, \tau = (1324), \tau^2 = (12)(34), \tau^3 = (1423), \theta = (13)(24), \tau\theta = (12), \tau^2\theta = (14)(23), \tau^3\theta = (34)\rangle$. Therefore, G is isomorphic to the dihedral group D_4 (Example 2.1.17). The subgroup lattice of G was computed in Example 2.3.41:



By Example 2.3.36, the center of G is $\langle \tau^2 \rangle$ which is normal. The three subgroups of order four are normal. The other four subgroups of order two, $\langle \tau^2\theta \rangle$, $\langle \theta \rangle$, $\langle \tau\theta \rangle$, and $\langle \tau^3\theta \rangle$, are not normal. Notice that $\tau^3\theta$ is complex conjugation. The reader should verify the following.

$$\begin{aligned} F^{\langle \tau^2 \rangle} &= \mathbb{Q}(i, u^2) \\ F^{\langle \tau\theta \rangle} &= \mathbb{Q}(ui) \\ F^{\langle \tau^3\theta \rangle} &= \mathbb{Q}(u) \\ F^{\langle \tau^2, \theta \rangle} &= \mathbb{Q}(u^2 i) \\ F^{\langle \tau \rangle} &= \mathbb{Q}(i) \\ F^{\langle \tau^2, \tau\theta \rangle} &= \mathbb{Q}(u^2) \\ F^{\langle \tau^2\theta \rangle} &= \mathbb{Q}(u - ui) \\ F^{\langle \theta \rangle} &= \mathbb{Q}(u + ui). \end{aligned}$$

The subfield lattice of F is



Notice that $\mathbb{Q}(u)$ is Galois over $\mathbb{Q}(u^2)$, $\mathbb{Q}(u^2)$ is Galois over $\mathbb{Q}$, but $\mathbb{Q}(u)$ is not Galois over $\mathbb{Q}$. This example shows that the property of being Galois is not transitive. In

other words, Galois over Galois is not Galois. The analogous statement for groups is also true. Namely, normal over normal is not normal.

3.4. Exercises.

EXERCISE 5.3.25. Show that the group of automorphisms of a prime field is trivial. In other words, prove: $\mathrm{Aut}(\mathbb{Q}) = \langle 1 \rangle$ and $\mathrm{Aut}(\mathbb{Z}_p) = \langle 1 \rangle$.

EXERCISE 5.3.26. Let F be a field, k the prime field of F, and σ an automorphism of F. Show that $\sigma(a) = a$ for every $a \in k$.

EXERCISE 5.3.27. This exercise outlines a proof that $\mathrm{Aut}(\mathbb{R}) = \langle 1 \rangle$. In the following, assume a, b, c are real numbers and r, s are rational numbers. For this exercise you can assume that if $a < b$, then there exists a rational number r such that $a < r < b$. Let σ be an automorphism of $\mathbb{R}$. Prove:

(1) $\sigma(a^2) = \sigma(a)^2$.
(2) If $b > 0$, then $\sigma(b) > 0$.
(3) If $r < c < s$, then $r < \sigma(c) < s$.
(4) For every $c \in \mathbb{R}$, $\sigma(c) = c$.

EXERCISE 5.3.28. Let $f(x) = x^3 + 3x + 3$. Show that f is irreducible in $\mathbb{Q}[x]$ and f has exactly one real root and two nonreal roots. Let $\alpha \in \mathbb{R}$ be the real root and β_1, β_2 be the nonreal roots of $f(x)$. Show that $\mathbb{Q}[\alpha, \beta_1]$ is the splitting field for f over $\mathbb{Q}$ and $\dim_{\mathbb{Q}} \mathbb{Q}[\alpha, \beta_1] = 6$. Show that $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}[\alpha]) = \langle 1 \rangle$. Show that $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}[\alpha, \beta_1])$ is isomorphic to S_3, the group of permutations of $\{\alpha, \beta_1, \beta_2\}$.

EXERCISE 5.3.29. Prove the following for $f = x^3 + x - 1$.

(1) f is irreducible in $\mathbb{Q}[x]$.
(2) If $F = \mathbb{Q}[x]/(f)$ and σ is an automorphism of F, then σ is the identity function.
(3) In $\mathbb{R}[x]$, f factors into a product of a linear polynomial and an irreducible quadratic.
(4) If F is the splitting field of f over $\mathbb{Q}$, then the Galois group $\mathrm{Aut}_{\mathbb{Q}}(F)$ is a nonabelian group of order six.

EXERCISE 5.3.30. Let F be the splitting field of $f = x^3 - 5$ over $\mathbb{Q}$.

(1) Show that the Galois group $\mathrm{Aut}_{\mathbb{Q}}(F)$ is a nonabelian group of order six.
(2) Find all intermediate fields K between $\mathbb{Q}$ and F.
(3) Prove or give a counterexample: Each intermediate field K is a Galois extension of $\mathbb{Q}$.

EXERCISE 5.3.31. Let F be the splitting field of $f = (x^2 - 2)(x^2 - 3)$ over $\mathbb{Q}$.

(1) Show that the Galois group $\mathrm{Aut}_{\mathbb{Q}}(F)$ is a noncyclic abelian group of order four.
(2) Find all intermediate fields K between $\mathbb{Q}$ and F.
(3) Prove or give a counterexample: Each intermediate field K is a Galois extension of $\mathbb{Q}$.

EXERCISE 5.3.32. Let $f \in k[x]$ be an irreducible separable polynomial of degree n over the field k. Let F/k be the splitting field for f over k and let $G = \mathrm{Aut}_k(F)$ be the Galois group. We call G the *Galois group* of f. Prove the following.

(1) G acts transitively on the roots of f. That is, given two roots α, β of f, there is $\sigma \in G$ such that $\sigma(\alpha) = \beta$.
(2) n divides $[G : 1]$.

EXERCISE 5.3.33. Consider the polynomial $f = x^4 + p^2$ in $\mathbb{Q}[x]$, where p is a prime number. Determine the following.

(1) The splitting field of f over $\mathbb{Q}$. Call this field K.
(2) The Galois group of f over $\mathbb{Q}$.
(3) The lattice of intermediate fields of $K/\mathbb{Q}$. Determine which intermediate fields are normal over $\mathbb{Q}$.

EXERCISE 5.3.34. Let F be a field and $f(x)$ a polynomial in $F[x]$ such that $f'(x) = 0$. That is, the derivative of $f(x)$ is the zero polynomial.

(1) If F has characteristic 0, show that $f(x) = \alpha$, for some $\alpha \in F$.
(2) If F has characteristic $p > 0$, show that there exists $g(x) \in F[x]$ such that $f(x) = g(x^p)$.

EXERCISE 5.3.35. Let F/k be an extension of fields. Let $\alpha \in F$. Prove that $F(\alpha)$ is a separable extension of k if and only if α is separable over k.

EXERCISE 5.3.36. Let F/k be a quadratic extension of fields. That is, $\dim_k(F) = 2$. Prove that if F/k is separable, then F/k is a Galois extension.

EXERCISE 5.3.37. Let F be a field, $\phi \in \mathrm{Aut}(F)$ and $k = F^{\langle \phi \rangle}$. Let $f \in F[x]$ be a polynomial satisfying:

(1) f is monic,
(2) f splits in $F[x]$,
(3) f has no repeated root, and
(4) if $\alpha \in F$ and $f(\alpha) = 0$, then $f(\phi(\alpha)) = 0$.

Show that $f \in k[x]$.

EXERCISE 5.3.38. Let F/k be an extension of fields where $\mathrm{char}\, k = p > 0$. Let $\alpha \in F$. Prove that α is separable over k if and only if $k(\alpha) = k(\alpha^p)$.

EXERCISE 5.3.39. Determine the group of automorphisms $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\sqrt{2}))$.

EXERCISE 5.3.40. Let p be a prime number and $\zeta = e^{2\pi i/p}$ a primitive pth root of unity in $\mathbb{C}$. Show that the group of automorphisms $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta))$ is isomorphic to the group of units in the ring $\mathbb{Z}/p$, hence is a cyclic group of order $p - 1$. This is a special case of Corollary 5.5.9 (3).

EXERCISE 5.3.41. Let F be the splitting field for $x^3 - 2$ over $\mathbb{Q}$ (see Exercise 5.2.16). Show that the group of automorphisms $\mathrm{Aut}_{\mathbb{Q}}(F)$ is isomorphic to the symmetric group S_3.

EXERCISE 5.3.42. Let F/k be a Galois extension of fields with finite group G. Let α be an arbitrary element of F, and set

$$g = \prod_{\sigma \in G} (x - \sigma(\alpha)).$$

Show that $g \in k[x]$ and the only irreducible factor of g in $k[x]$ is $\mathrm{Irr.poly}_k(\alpha)$.

EXERCISE 5.3.43. Determine the Galois group of the polynomial $x^4 + x^2 - 6$ over $\mathbb{Q}$.

EXERCISE 5.3.44. Determine the smallest Galois extension $K/\mathbb{Q}$ containing $2^{1/2}+2^{1/3}$. Determine $\mathrm{Aut}_{\mathbb{Q}}(K)$.

EXERCISE 5.3.45. Determine the Galois group of the polynomial x^6-8 over each of these fields: $\mathbb{Q}$, $\mathbb{Q}\big(\sqrt{2}\big)$, and $\mathbb{Q}\,(\zeta)$, where $\zeta = e^{2\pi i/3}$ is a primitive third root of 1 in $\mathbb{C}$.

EXERCISE 5.3.46. For the polynomial $\big(x^2-2\big)\big(x^3+2\big)$ determine the Galois group over each of these fields: $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Q}\big(\sqrt{2}\big)$, $\mathbb{Q}\big(\sqrt[3]{2}\big)$, and $\mathbb{Q}\,(\zeta)$, where $\zeta = e^{2\pi i/6}$ is a primitive third root of -1 in $\mathbb{C}$.

EXERCISE 5.3.47. Let F denote the splitting field of x^8-1 over the field $\mathbb{Q}$ of rational numbers. Determine the lattice of subfields and show that the Galois group $\mathrm{Aut}_{\mathbb{Q}}(F)$ is a noncyclic group of order 4. This is a special case of Corollary 5.5.9 (3).

EXERCISE 5.3.48. Let k be a field of characteristic zero and f an irreducible polynomial in $k[x]$. Let F/k be an extension of fields and assume f splits over F. Prove that if $\alpha \in F$ and $f(\alpha)=0$, then $f(\alpha+1)\neq 0$.

EXERCISE 5.3.49. Let F/k be an extension of fields and G a finite subgroup of $\mathrm{Aut}_k(F)$. As in Corollary 5.3.8 we view G as a linearly independent subset of the F-vector space $\mathrm{Hom}_k(F,F)$. Let $\Delta(F/k,G)$ denote the F-vector subspace of $\mathrm{Hom}_k(F,F)$ spanned by $\{\sigma \mid \sigma \in G\}$. Show that $\Delta(F/k,G)$ is a k-subalgebra of $\mathrm{Hom}_k(F,F)$. The ring $\Delta(F/k,G)$ is an example of a trivial crossed product (see [**10**, Section 12.1]).

EXERCISE 5.3.50. For the polynomial $f=x^4-5$, find the Galois group of f over over each of these fields: $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Q}(\sqrt{5})$, $\mathbb{Q}(i\sqrt{5})$.

EXERCISE 5.3.51. Let $f=(2x^2-4x+1)(x^4+1)$. Find a splitting field and determine the Galois group of f over each of these fields: $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Q}(\sqrt{2})$, $\mathbb{Q}(i)$, and $\mathbb{Q}\,(\zeta)$, where $\zeta = e^{2\pi i/8}$ is a primitive eighth root of 1 in $\mathbb{C}$.

EXERCISE 5.3.52. Let $f=(4x^2+2x+1)(x^6-1)$. Find a splitting field and determine the Galois group of f over each of the following fields: $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Q}(\sqrt{3})$, $\mathbb{Q}(i)$, and $\mathbb{Q}\,(\zeta)$, where $\zeta = e^{2\pi i/6}$ is a primitive sixth root of 1 in $\mathbb{C}$.

4. Separable Closure

Given an algebraic extension of fields F/k we construct the separable closure of k in F. This result is then applied to show that the property of being separable is transitive. As another application, we prove in Theorem 5.4.5 a characterization of perfect fields. As we saw in Example 5.3.24, the property of being Galois is not transitive. Nevertheless, we prove in Theorem 5.4.6 that the property of being Galois is preserved under a change of base field. As an application of Galois Theory, in Theorem 5.4.10 we give a proof of the Fundamental Theorem of Algebra.

4.1. The Existence of a Separable Closure.

LEMMA 5.4.1. *Let F/k be an extension of fields and assume* $\operatorname{char} k = p > 0$. *Let $u \in F$ and assume u is algebraic over k. There exists $n \geq 0$ such that u^{p^n} is separable over k.*

PROOF. If u is separable over k, then take $n = 0$. Let $f = \text{Irr. poly}_k(u)$ and use induction on the degree of f. Assume f is not separable and $d = \deg f > 1$. By Lemma 5.2.11, there exists $g \in k[x]$ such that $f(x) = g(x^p)$. Because f is irreducible, so is g. Therefore, $f(u) = g(u^p) = 0$, u^p is algebraic over k, and the degree of $\text{Irr. poly}_k(u^p)$ is equal to d/p. By induction on d, there is some $n \geq 0$ such that $(u^p)^{p^n}$ is separable over k. □

THEOREM 5.4.2. *Let F/k be an algebraic extension of fields. If*

$$S = \{u \in F \mid u \text{ is separable over } k\},$$

then S is an intermediate field of F/k, and S/k is separable. The field S is called the separable closure of k in F.

PROOF. It is enough to show S is a field. Let α and β be elements of $S - k$. If $f = \text{Irr. poly}_k(\alpha)$, then f is separable and irreducible over k. Likewise, $g = \text{Irr. poly}_k(\beta)$ is separable and irreducible over k. By Theorem 5.3.18, if E is the splitting field over k of fg, then E/k is a separable extension of fields. Since $k(\alpha, \beta)$ is an intermediate field of E/k, it is itself a separable extension of k. Therefore, S contains $\alpha + \beta$, $\alpha - \beta$, $\alpha\beta$, α/β. It follows that S is a field. □

THEOREM 5.4.3. *(Separable over Separable is Separable) Let $k \subseteq F \subseteq K$ be a tower of algebraic field extensions. If F is separable over k and K is separable over F, then K is separable over k.*

PROOF. By Proposition 5.1.10 (4), K is algebraic over k. If $\text{char}\, k = 0$, then an algebraic extension is separable, so assume $\text{char}\, k = p > 0$. By Theorem 5.4.2, let S be the separable closure of k in K. Then $F \subseteq S \subseteq K$. It is enough to show $S = K$. Let $u \in K$. By Lemma 5.4.1, there exists $n \geq 0$ such that $\alpha = u^{p^n}$ is in S. Then u satisfies the polynomial $x^{p^n} - \alpha \in S[x]$ and in $K[x]$ we have the factorization $x^{p^n} - \alpha = (x - u)^{p^n}$. If $f = \text{Irr. poly}_S(u)$, then f divides $(x - u)^{p^n}$ in $K[x]$. If $g = \text{Irr. poly}_F(u)$, then g is separable and since f divides g in $S[x]$, we know that f has no multiple roots in K. So $f = x - u$ and $u \in S$. □

DEFINITION 5.4.4. A field k is said to be *perfect* if $\text{char}\, k = 0$, or $\text{char}\, k = p$ is a prime number and the Frobenius homomorphism $\theta : k \to k$ by $a \mapsto a^p$ is onto (see Exercise 3.2.32).

We will show in Lemma 5.5.12 below that a finite field is a perfect field.

THEOREM 5.4.5. *Let k be a field. The following are equivalent.*

(1) k is a perfect field.
(2) Every irreducible polynomial in $k[x]$ is separable.
(3) Every algebraic extension of k is separable over k.

PROOF. If $\text{char}\, k = 0$, then this is immediate.

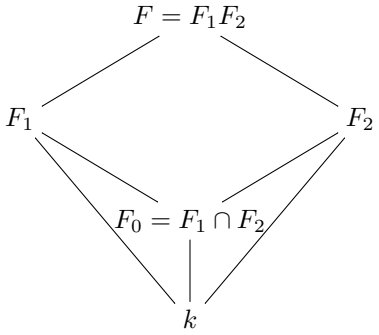
(2) is equivalent to (3): This is Exercise 5.4.12.

(3) implies (1): Assume k has positive characteristic p and every algebraic extension of k is separable. Let $\theta : k \to k$ be the Frobenius homomorphism (Exercise 3.2.32). Let $\alpha \in k$. We show $\alpha = \theta(u)$ for some $u \in k$. Consider the polynomial $x^p - \alpha$ in $k[x]$. Let F be an extension of k containing a root u of $x^p - \alpha$. In $F[x]$ we have the factorization $x^p - \alpha = (x - u)^p$. By assumption, F/k is separable, which implies this factorization occurs in $k[x]$. That is, $u \in k$ and $\alpha = \theta(u)$.

(1) implies (3): Assume $\operatorname{char} k = p > 0$ and the Frobenius homomorphism $\theta : k \to k$ is an automorphism of k. Let F/k be an algebraic extension. Let $\alpha \in F - k$. We show that $k(\alpha)$ is a separable extension of k. Let $f \in k[x]$ be the irreducible polynomial of α over k. By Theorem 3.6.2, $\theta(f) = g$ is an irreducible polynomial in $k[x]$ such that $\deg g = \deg f$. Since $g(\alpha^p) = (f(\alpha))^p = 0$, we see that $k(\alpha^p)$ is a field extension of k which is an intermediate field of $k(\alpha)/k$ such that $\dim_k(k(\alpha^p)) = \dim_k(k(\alpha))$. It follows that $k(\alpha^p) = k(\alpha)$, hence the Frobenius homomorphism is an automorphism $\theta : k(\alpha) \to k(\alpha)$. For any $m > 0$, $\theta^m(x) = x^{p^m}$. Since $k[\alpha] = k(\alpha)$, a typical element in $k(\alpha)$ can be represented in the form $u = \sum_i a_i \alpha^i$ where $a_i \in k$. Therefore $\theta^m(u) = \sum_i a_i^{p^m} (\alpha^{p^m})^i$ is in $k(\alpha^{p^m})$. This shows $k(\alpha^{p^m}) = k(\alpha)$ for all $m > 0$. By Theorem 5.4.2, let S be the separable closure of k in $k(\alpha)$. For some $n \geq 0$, $\alpha^{p^n} \in S$. Therefore $k(\alpha) = k(\alpha^{p^n}) \subseteq S$ so $k(\alpha)$ is a separable extension of k. □

4.2. A Change of Base Theorem for a Galois Extension. Theorem 5.4.6 is what is called a "change of base" theorem for a Galois extension. It says that if F_1/k is a Galois extension and F_2/k is a finite field extension, then $F = F_1F_2$ is a Galois extension of F_2. The base field is extended from k to F_2. This useful result also gives sufficient conditions such that the Galois group is preserved.

THEOREM 5.4.6. *Let K/k be a finite dimensional extension of fields. Let F_1 and F_2 be intermediate fields. Set $F = F_1F_2$ and $F_0 = F_1 \cap F_2$.*



(1) If F_1 is a Galois extension of k, then F is a Galois extension of F_2 and there is an isomorphism of groups $\operatorname{Aut}_{F_2}(F) \cong \operatorname{Aut}_{F_0}(F_1)$ defined by the assignment $\phi \mapsto \phi|_{F_1}$.

(2) If F_1 and F_2 are both Galois extensions of k, then F is a Galois extension of k. If $F_1 \cap F_2 = k$, then $\operatorname{Aut}_k(F) \cong \operatorname{Aut}_{F_1}(F) \times \operatorname{Aut}_{F_2}(F)$.

PROOF. (1): By Theorems 5.3.18 and 5.2.14, $F_1 = k(u)$ is a simple extension. Let $f = \operatorname{Irr.poly}_k(u)$. By Theorem 5.1.12, $F = F_2(u)$. Let $g = \operatorname{Irr.poly}_{F_2}(u)$. Theorem 5.1.4 implies g divides f. Then every root of g is in F, hence F is a splitting field for g. By Theorem 5.3.18, F/F_2 is a Galois extension. If $\phi \in \operatorname{Aut}_{F_2} F$, then ϕ is completely determined by the value of $\phi(u)$. But $\phi(u)$ is a root of f. Since F_1 is a splitting field for f, $\phi(F_1) \subseteq F_1$. Since ϕ fixes F_2 point-wise, ϕ fixes k point-wise. Therefore, $\theta : \operatorname{Aut}_{F_2}(F) \to \operatorname{Aut}_k(F_1)$ is a homomorphism of groups. If ϕ fixes F_1 point-wise, then $\phi(u) = u$ and ϕ is the identity function on F. This proves θ is one-to-one. Using θ, we identify $\operatorname{Aut}_{F_2}(F)$ with a subgroup of $\operatorname{Aut}_k F_1$. Let $E = F_1^{\operatorname{Aut}_{F_2}(F)}$. By Theorem 5.3.21, F_1/E is a Galois extension and $\dim_E(F_1) = |\operatorname{Aut}_{F_2}(F)| = \dim_{F_2}(F)$. Since $F_1 \subseteq F$, we have

$E \subseteq F^{\mathrm{Aut}_{F_2}(F)} = F_2$. Since $\dim_{F_2}(F) = \dim_E(F_1)$, Exercise 5.1.23 implies that $\dim_E(F) = \dim_E(F_1)\dim_E(F_2)$. By Theorem 5.1.12 (5), we have $E = F_1 \cap F_2$, which completes the proof.

(2): This is Exercise 5.4.13. □

4.3. Examples. In this section we include some examples that did not seem to fit in elsewhere.

EXAMPLE 5.4.7. This is an example of a Galois extension of $\mathbb{Q}$ with abelian Galois group of order 8. Let a be a positive odd integer and $f = x^8 + a^4$. By Exercise 3.7.13, f is irreducible over $\mathbb{Q}$. Let ζ be the complex number $e^{2\pi i/16}$. Then $\zeta^8 = -1$. Let α be the positive real number such that $\alpha^2 = a$. For any integer k, $f(\zeta^{2k+1}\alpha) = \zeta^8\zeta^{16k}\alpha^8 + a^4 = 0$. Therefore the eight roots of f in $\mathbb{C}$ are $S = \{\zeta^{2k+1}\alpha \mid 0 \le k \le 7\}$. By Theorem 5.1.4, the set $\{1, \zeta\alpha, \zeta^2\alpha^2, \ldots, \zeta^7\alpha^7\}$ is a basis for $\mathbb{Q}(\zeta\alpha)$ as a $\mathbb{Q}$-vector space. Since $(\zeta\alpha)^{2k+1} = \zeta^{2k+1}a^k\alpha$, we see that $S \subseteq \mathbb{Q}(\zeta\alpha)$. Hence $\mathbb{Q}(\zeta\alpha)$ is a splitting field for f. By Corollary 5.1.7 applied to $\zeta\alpha$ and $\zeta^3\alpha$, there is an automorphism $\tau \in \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha))$ such that $\tau(\zeta\alpha) = \zeta^3\alpha$. Since $\zeta^2\alpha^2 = \zeta^2 a$, it follows that $\zeta^2 \in \mathbb{Q}(\zeta\alpha)$. We have $\tau(\zeta^2) = \tau((\zeta\alpha)^2a^{-1}) = \tau(\zeta\alpha)^2a^{-1} = (\zeta^3\alpha)^2a^{-1} = (\zeta^6 a)a^{-1} = \zeta^6$. Using this it is now possible to compute the action of τ on S: $\tau(\zeta\alpha) = \zeta^3\alpha$, $\tau(\zeta^3\alpha) = -\zeta\alpha$, $\tau(-\zeta\alpha) = -\zeta^3\alpha$, $\tau(-\zeta^3\alpha) = \zeta\alpha$, $\tau(\zeta^5\alpha) = -\zeta^7\alpha$, $\tau(-\zeta^7\alpha) = -\zeta^5\alpha$, $\tau(-\zeta^5\alpha) = \zeta^7\alpha$, $\tau(\zeta^7\alpha) = \zeta^5\alpha$. So τ has two disjoint orbits, each of length four. Fix this ordering of the 8 elements of S:

$$S = \{\zeta\alpha, \zeta^3\alpha, -\zeta\alpha, -\zeta^3\alpha, \zeta^7\alpha, \zeta^5\alpha, -\zeta^7\alpha, -\zeta^5\alpha\}. \tag{4.1}$$

Then τ has the cycle representation $\tau = (1234)(5678)$ (see Example 2.1.15). Let $\chi : \mathbb{C} \to \mathbb{C}$ be complex conjugation (see Example 5.3.4). Then χ restricts to a permutation of S, hence defines an automorphism of $\mathbb{Q}(\zeta\alpha)$. Based on the ordering of S in (4.1), $\chi = (17)(28)(35)(46)$ is the disjoint cycle representation of χ. By direct computation, we see that $\tau\chi = (1836)(2547) = \chi\tau$. By Exercise 2.5.22, τ and χ generate an abelian group, call it G, isomorphic to $\mathbb{Z}/4 \oplus \mathbb{Z}/2$. Since $\dim_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha)) = 8 = [G : 1]$, by Proposition 5.3.15, $\mathbb{Q}(\zeta\alpha)$ is Galois over $\mathbb{Q}$ and the Galois group is $G = \langle\tau, \chi\rangle$. This also shows $G = \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha))$.

EXAMPLE 5.4.8. This is a generalization of Example 5.4.7. In this example we construct a Galois extension over $\mathbb{Q}$ such that the Galois group is isomorphic to the group of units in $\mathbb{Z}/(2^{n+1})$. As in Example 2.1.3, the set of invertible elements in the ring $\mathbb{Z}/(2^{n+1})$ is denoted $U_{2^{n+1}}$ and the order of this group is 2^n. Let a be a positive odd integer and $n \ge 2$. Let $f = x^{2^n} + a^{2^{n-1}}$. When $n = 3$, this example agrees with Example 5.4.7. By Exercise 3.7.13, f is irreducible over $\mathbb{Q}$. Let ζ be the complex number $e^{2\pi i/2^{n+1}}$, a primitive 2^{n+1}th root of unity. Then $\zeta^{2^{n+1}} = 1$ and $\zeta^{2^n} = -1$. Let α be the positive real number such that $\alpha^2 = a$. For any integer k,

$$f(\zeta^{2k-1}\alpha) = (\zeta^{2k-1}\alpha)^{2^n} + a^{2^{n-1}} = \zeta^{-2^n}(\zeta^{2^{n+1}})^k\alpha^{2^n} + a^{2^{n-1}} = -a^{2^{n-1}} + a^{2^{n-1}} = 0.$$

Therefore the 2^n roots of f in $\mathbb{C}$ are

$$S = \{\zeta^{2k-1}\alpha \mid 1 \le k \le 2^n\} = \{\zeta\alpha, \zeta^3\alpha, \ldots, \zeta^{2^{n+1}-1}\alpha\}.$$

By Theorem 5.1.4, the set

$$\{(\zeta\alpha)^j \mid 0 \le j < 2^n\} = \{1, \zeta\alpha, (\zeta\alpha)^2, \ldots, (\zeta\alpha)^{2^n-1}\}$$

is a basis for $\mathbb{Q}(\zeta\alpha)$ as a $\mathbb{Q}$-vector space. Since $(\zeta\alpha)^{2k+1} = \zeta^{2k+1}a^k\alpha$, we see that $S \subseteq \mathbb{Q}(\zeta\alpha)$. Hence $\mathbb{Q}(\zeta\alpha)$ is a splitting field for f. Let t be an arbitrary odd integer. By Corollary 5.1.7 applied to $\zeta\alpha$ and $\zeta^t\alpha$, there is an automorphism $\tau_t \in \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha))$ such that $\tau_t(\zeta\alpha) = \zeta^t\alpha$. Let s be another odd integer. Since ζ is a primitive 2^{n+1}th root of unity, Proposition 5.3.2 (2) implies that $\tau_t = \tau_s$ if and only if $s \equiv t \pmod{2^{n+1}}$. Since $\zeta^2\alpha^2 = \zeta^2 a$, it follows that $\zeta^2 \in \mathbb{Q}(\zeta\alpha)$. We have

$$\tau_t(\zeta^2) = \tau_t((\zeta\alpha)^2a^{-1}) = \tau_t(\zeta\alpha)^2a^{-1} = (\zeta^t\alpha)^2a^{-1} = (\zeta^{2t}a)a^{-1} = \zeta^{2t}.$$

Using this, we see that

$$\tau_t(\zeta^{2k+1}\alpha) = \tau_t((\zeta^2)^k\zeta\alpha) = (\zeta^{2t})^k(\zeta^t\alpha) = (\zeta^{2k+1})^t\alpha$$

and

$$\tau_s\tau_t(\zeta\alpha) = \tau_s(\zeta^t\alpha) = \zeta^{ts}\alpha = \tau_{ts}(\zeta\alpha).$$

Let σ denote an arbitrary automorphism in $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha))$. Then Proposition 5.3.2 (1) implies $\sigma(\zeta\alpha) = \zeta^t\alpha$ for a unique $t \in \{1, 3, \ldots, 2^{n+1}-1\}$. By Proposition 5.3.2 (2), σ is equal to τ_t. The computations above show that the assignment $\theta(t) = \tau_t$ defines an isomorphism of groups $\theta : U_{2^{n+1}} \to \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha))$. Since $\dim_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha)) = 2^n$, Proposition 5.3.15 implies $\mathbb{Q}(\zeta\alpha)$ is Galois over $\mathbb{Q}$ and the Galois group is isomorphic to $U_{2^{n+1}}$. See Corollary 5.5.9 for a related result concerning cyclotomic extensions.

The next proposition shows that for a Galois extension F/k, if f is an irreducible separable polynomial in $k[x]$, then the irreducible factors of f in $F[x]$ all have the same degree.

PROPOSITION 5.4.9. *Let F/k be a Galois extension of fields and f an irreducible separable polynomial in $k[x]$. If the unique factorization of f in $F[x]$ is $f = f_1 \cdots f_m$, then* $\deg f_1 = \deg f_2 = \cdots = \deg f_m$.

PROOF. We prove this in two steps.

Step 1: Suppose K/k is a Galois extension of fields with group G. Assume f splits in $K[x]$. Let N be a normal subgroup of G and assume $F = K^N$. We prove that the irreducible factors of f in $F[x]$ all have the same degree. Let $X = \{\alpha_1, \ldots, \alpha_n\}$ be the roots of f in K. If $L = k(X)$ is the splitting field for f in K, then L/k is Galois by Theorem 5.3.18. By Exercise 5.3.32, $\mathrm{Aut}_k(L)$ acts transitively on X. By Theorem 5.3.21, $\mathrm{Aut}_k(L)$ is a homomorphic image of G, hence G acts transitively on X. Let a, b be two arbitrary elements of X. Let $\tau \in G$ such that $\tau(a) = b$. Since N is normal, $\tau N = N\tau$. Therefore $\tau Na = N\tau a = Nb$. This shows the orbit containing a is in one-to-one correspondence with the orbit containing b. Let $O_1, \ldots, O_m$ be the orbits of N acting on X. Then $|O_1| = \cdots = |O_m|$. For each $1 \le i \le m$, set $f_i = \prod_{a \in O_i}(x - a)$. We have

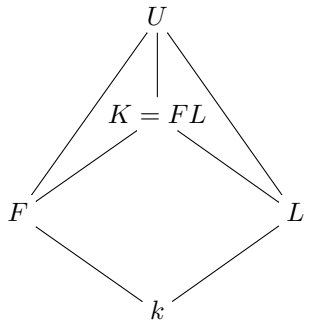
$$\begin{aligned} f &= \prod_{a \in X}(x - a) \\ &= \prod_{i=1}^{m}\prod_{a \in O_i}(x - a) \\ &= f_1 \cdots f_m. \end{aligned}$$

Since $\deg f_i = |O_i|$, all of the f_i have the same degree. Now we prove that each f_i is in $F[x]$. If $\tau \in N$, then $\tau O_i = O_i$, hence

$$\tau(f_i) = \prod_{a \in O_i} (x - \tau(a)) = \prod_{a \in O_i} (x - a) = f_i$$

so the coefficients of f_i are fixed by N. Hence $f_i \in F[x]$. Now we prove that each f_i is irreducible in $F[x]$. Fix one element of O_i, say a_i. If $p_i = \text{Irr. poly}_F(a_i)$, then by Theorem 5.1.4 we have $p_i \mid f_i$. For each $\tau \in N$, $p_i(\tau a_i) = \tau(p_i(a_i)) = 0$ shows that every element of O_i is a root of p_i. Therefore, $\deg p_i \geq \deg f_i$. This proves $f_i = p_i$ and in particular, f_i is irreducible over F. We have proved that $f = f_1 \cdots f_m$ is the factorization of f into irreducibles in the ring $F[x]$ and all of the factors f_i have the same degree.

Step 2. In the context of the proposition, assume F/k is a Galois extension. Let U/F be a splitting field for f over F. Let $X = \{\alpha_1, \ldots, \alpha_n\}$ be the roots of f in U. Let $L = k(X)$ be the splitting field for f over k in U. Then L/k is Galois by Theorem 5.3.18.



By Theorem 5.4.6, $K = FL$ is a Galois extension of k containing K. By Theorem 5.3.21, Step 2 reduces to Step 1. □

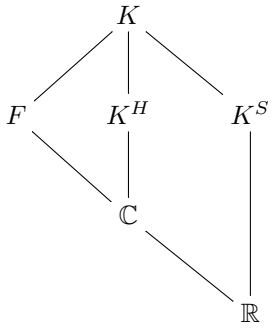
4.4. The Fundamental Theorem of Algebra. The purpose of this section is to apply Galois Theory and some facts about the completion of the metric space $\mathbb{R}$ to prove the Fundamental Theorem of Algebra.

As in Section 1.4, the field of real numbers is denoted $\mathbb{R}$ and the field of complex numbers is denoted $\mathbb{C}$. The proof of the Fundamental Theorem of Algebra utilizes results from Calculus. By Theorem 1.4.2, an irreducible polynomial of odd degree in $\mathbb{R}[x]$ is linear. By Proposition 1.4.3 (5), the ring $\mathbb{C}[x]$ contains no irreducible quadratic polynomial.

THEOREM 5.4.10. *The field of complex numbers is algebraically closed. In particular, an irreducible polynomial over $\mathbb{C}$ is linear.*

PROOF. By Lemma 5.2.1, we show that every irreducible polynomial over $\mathbb{C}$ is linear. Let F be a finite dimensional extension field of $\mathbb{C}$. By Theorem 5.2.4, it suffices to show that $F = \mathbb{C}$. Since F is a finite dimensional separable extension field of $\mathbb{R}$, by Corollary 5.3.20, there is a finite dimensional Galois extension $K/\mathbb{R}$ which contains F as an intermediate field. Let G be the Galois group of K over $\mathbb{R}$. Let S be a Sylow-2 subgroup of G. Then K^S is an extension field of $\mathbb{R}$ and $\dim_{\mathbb{R}} K^S$ is odd. If $\alpha \in K^S$, then $\dim_{\mathbb{R}} \mathbb{R}(\alpha)$ divides $\dim_{\mathbb{R}} K^S$, hence is odd. By Theorem 5.1.4, the degree of $\text{Irr. poly}_{\mathbb{R}}(\alpha)$ is odd. By Theorem 1.4.2, an irreducible polynomial of odd degree in $\mathbb{R}[x]$ is linear. Therefore, $K^S = \mathbb{R}$. This proves $S = G$

is a 2-group. For sake of contradiction, assume $\mathrm{Aut}_{\mathbb{C}}(K)$ is a nontrivial 2-group. By Theorem 2.7.1, there exists a normal subgroup H of $\mathrm{Aut}_{\mathbb{C}}(K)$ of index 2. Then K^H is a field extension of $\mathbb{C}$ of degree 2. This is a contradiction, because by Proposition 1.4.3 (5), the ring $\mathbb{C}[x]$ contains no irreducible quadratic polynomial.



□

THEOREM 5.4.11. *An irreducible polynomial in $\mathbb{R}[x]$ has degree 1 or 2. If f is a monic polynomial of positive degree in $\mathbb{R}[x]$, then the unique factorization of f into irreducible polynomials has the general form*

$$f = (x-u_1)^{m_1}\cdots(x-u_{r_1})^{m_{r_1}}\, q_1^{n_1}\cdots q_{r_2}^{n_{r_2}}$$

where $u_1,\ldots,u_{r_1}$ are the distinct real roots of f, $r_1 \geq 0$, each $m_i \geq 1$, $q_1,\ldots,q_{r_2}$ are the distinct irreducible monic quadratic factors of f in $\mathbb{R}[x]$, $r_2 \geq 0$, and each $n_j \geq 1$.

PROOF. In $\mathbb{C}[x]$, f factors into linear factors. Let $z = a + bi$ be a nonreal complex number. By Example 4.4.12, the irreducible polynomial of z over $\mathbb{R}$ is $\mathrm{Irr.poly}_{\mathbb{R}}(z) = (x-z)(x-\bar{z}) = x^2 - 2ax - (a^2+b^2)$. The nonreal roots of f come in conjugate pairs. The rest of the proof is left to the reader. □

4.5. Exercises.

EXERCISE 5.4.12. Prove that (2) is equivalent to (3) in Theorem 5.4.5.

EXERCISE 5.4.13. Prove Theorem 5.4.6 (2).

5. Galois Extensions, Some Special Cases

We begin this section by showing that to a Galois extension F/k there is associated a k-linear functional $T_k^F : F \to k$, called the trace, and a multiplicative map $N_k^F : F \to k$, called the norm. In Section 5.5.2, Hilbert's Theorem 90 is proved for the special case when the Galois group $\mathrm{Aut}_k(F)$ is a cyclic group. When F is the splitting field over k of the polynomial $x^n - 1$, we say F/k is a cyclotomic extension of k. Cyclotomic extensions are the subject of Section 5.5.3. In Section 5.5.4 we prove Wedderburn's Theorem which says that a finite division ring is a field. The proof uses results from Chapters 2 and 5. In Section 5.5.5, many results and theorems about finite fields are assembled into a fundamental theorem on finite fields. Results from Section 5.5 will be applied in Section 5.6.

5.1. The Trace Map and Norm Map. In this section we show that to a Galois extension F/k are associated the trace map $T_k^F : F \to k$ and the norm map $N_k^F : F \to k$. By Lemma 5.3.6, the left regular representation embeds F as a subring of $\mathrm{Hom}_k(F, F)$. Using this embedding, we show how to make the group of linear functionals $\mathrm{Hom}_k(F, k)$ into an F-vector space. Since F/k is Galois, we show that $\mathrm{Hom}_k(F, k)$ has dimension one over the field F and the trace map is a generator.

Let F/k be a Galois extension with finite group G. For $x \in F$, define

$$T_k^F(x) = \sum_{\sigma \in G} \sigma(x) \tag{5.1}$$

and

$$N_k^F(x) = \prod_{\sigma \in G} \sigma(x). \tag{5.2}$$

Since G is a group, for any $\tau \in G$,

$$\begin{aligned} \tau\Big(\sum_{\sigma \in G} \sigma(x)\Big) &= \sum_{\sigma \in G} \tau\sigma(x) \\ &= \sum_{\sigma \in G} \sigma(x) \end{aligned}$$

so the right hand side of (5.1) is fixed by every $\tau \in G$. Likewise,

$$\begin{aligned} \tau\Big(\prod_{\sigma \in G} \sigma(x)\Big) &= \prod_{\sigma \in G} \tau\sigma(x) \\ &= \prod_{\sigma \in G} \sigma(x) \end{aligned}$$

so the right hand side of (5.2) is fixed by G as well. Since $F^G = k$, this means that both T_k^F and N_k^F are mappings from F to k. We call the mapping T_k^F the *trace from F to k* and the mapping N_k^F is called the *norm from F to k*. If $x, y \in F$ and $a, b \in k$, then

$$\begin{aligned} T_k^F(ax + by) &= \sum_{\sigma \in G} \sigma(ax + by) \\ &= a \sum_{\sigma \in G} \sigma(x) + b \sum_{\sigma \in G} \sigma(y) \\ &= aT_k^F(x) + bT_k^F(y). \end{aligned}$$

Therefore, the trace is k-linear and represents an element of $\mathrm{Hom}_k(F, k)$. Also

$$\begin{aligned} N_k^F(xy) &= \prod_{\sigma \in G} \sigma(xy) \\ &= \prod_{\sigma \in G} \sigma(x) \prod_{\sigma \in G} \sigma(y) \\ &= N_k^F(x) N_k^F(y). \end{aligned}$$

Hence, the norm induces a homomorphism of multiplicative groups $F^* \to k^*$.

LEMMA 5.5.1. *Let F/k be a Galois extension of fields with finite group G. Let $\mathrm{Hom}_k(F, F)$ be the ring of k-linear transformations of F.*

(1) If $n = [G : 1]$, then $\mathrm{Hom}_k(F, F)$ is an F-vector space of dimension n and $\{\sigma \mid \sigma \in G\}$ is a basis.
(2) There exists $c \in F$ such that $T_k^F(c) = 1$.
(3) $\mathrm{Hom}_k(F, k)$ is an F-vector space of dimension 1 and $\{T_k^F\}$ is a basis.

PROOF. (1): Let $n = \dim_k(F)$. By Proposition 5.3.13, $n = [G : 1]$. By Lemma 5.3.6, the left regular representation embeds F as a subring of $\mathrm{Hom}_k(F, F)$ and by Proposition 4.3.8, and the F-vector space $\mathrm{Hom}_k(F, F)$ has dimension n. By Corollary 5.3.8, the set $G = \mathrm{Aut}_k(F)$ is a linearly independent subset of the F-vector space $\mathrm{Hom}_k(F, F)$. By Theorem 4.3.4, G is a basis for the F-vector space $\mathrm{Hom}_k(F, F)$.

(2): By Theorem 5.3.7, there exists $y \in F$ such that $x = \sum_{\sigma \in G} \sigma(y) \neq 0$. Since G is a group, $\tau(x) = x$ for every $\tau \in G$. Therefore, $x \in F^G = k$. Define $c = x^{-1}y$. Then $T_k^F(c) = \sum_{\sigma \in G} \sigma(x^{-1}y) = x^{-1} \sum_{\sigma \in G} \sigma(y) = 1$.

(3): Using $\lambda : F \to \mathrm{Hom}_k(F, F)$ we can turn $\mathrm{Hom}_k(F, k)$ into an F-vector space. For every $f \in \mathrm{Hom}_k(F, k)$ and $a \in F$, define af to be $f \circ \ell_a$. By Proposition 4.3.8, $\mathrm{Hom}_k(F, k)$ is an F-vector space of dimension one. As an F-vector space, any nonzero element $f \in \mathrm{Hom}_k(F, k)$ is a generator. By (2), the trace mapping T_k^F is a generator for $\mathrm{Hom}_k(F, k)$ over F. This implies for every $f \in \mathrm{Hom}_k(F, k)$ there is a unique $\alpha \in F$ such that $f(x) = T_k^F(\alpha x)$ for all $x \in F$. The mapping $F \to \mathrm{Hom}_k(F, k)$ given by $\alpha \mapsto T_k^F \circ \ell_\alpha$ is an isomorphism of k-vector spaces. □

PROPOSITION 5.5.2. *Suppose F/k is G-Galois where the group G has order $[G : 1] = n$. Then there exist elements $a_1, \ldots, a_n, y_1, \ldots, y_n$ in F such that*
(1) $T_k^F(y_j a_i) = \delta_{ij}$ (Kronecker delta), and
(2) for each $\sigma \in G$: $a_1\sigma(y_1) + \cdots + a_n\sigma(y_n) = \begin{cases} 1 & \text{if } \sigma = 1 \\ 0 & \text{if } \sigma \neq 1 \end{cases}$.

PROOF. Let $\{a_1, \ldots, a_n\}$ be a k-basis for F. For each $j = 1, 2, \ldots, n$, let $f_j : F \to k$ be the projection onto coordinate j. That is, $f_j(a_i) = \delta_{ij}$ (Kronecker delta). For each $x \in F$,

$$x = \sum_{j=1}^{n} f_j(x) a_j.$$

We say $\{(a_j, f_j) \mid j = 1, \ldots, n\}$ is a *dual basis* for F. By Lemma 5.5.1 (3), T_k^F is a generator for $\mathrm{Hom}_k(F, k)$ over F. There exist unique $y_1, \ldots, y_n$ in F such that for each $x \in F$, $f_j(x) = T_k^F(y_j x) = \sum_{\sigma \in G} \sigma(y_j x)$. Part (1) follows by substituting $x = a_i$. Combining these facts,

$$\begin{aligned} x &= \sum_{j=1}^{n} f_j(x) a_j \\ &= \sum_{j=1}^{n} \sum_{\sigma \in G} \sigma(y_j x) a_j \\ &= \sum_{\sigma \in G} \Big(\sigma(x) \sum_{j=1}^{n} \sigma(y_j) a_j \Big). \end{aligned}$$

By Lemma 5.5.1 (1), G is a basis for $\mathrm{Hom}_k(F, F)$ over F. Therefore, $\sum_{j=1}^{n} \sigma(y_j) a_j = \delta_{\sigma,1}$, which is (2). □

LEMMA 5.5.3. *Suppose F/k is a Galois extension of fields with finite group G. If H is a normal subgroup of G and $E = F^H$, then $T_k^F = T_k^E \circ T_E^F$ and $N_k^F = N_k^E \circ N_E^F$.*

PROOF. Let $x \in F$. Then

$$\begin{aligned} T_k^E\Big(T_E^F(x)\Big) &= T_k^E\Big(\sum_{\sigma\in H}\sigma(x)\Big) \\ &= \sum_{\tau\in G/H}\tau\Big(\sum_{\sigma\in H}\sigma(x)\Big) \\ &= \sum_{\tau\in G/H}\sum_{\sigma\in H}\tau\sigma(x) \\ &= \sum_{\rho\in G}\rho(x) \\ &= T_k^F(x). \end{aligned}$$

The proof of the second identity is left to the reader. □

5.2. Hilbert's Theorem 90. In Theorem 5.5.4 we prove Hilbert's Theorem 90 for the special case where F/k is a Galois extension with finite cyclic Galois group. For generalizations of Theorem 5.5.4, see [**9**, Theorem 12.5.25]

THEOREM 5.5.4. *(Hilbert's Theorem 90) Let F/k be a Galois extension of fields with finite group G. Assume $G = \langle\sigma\rangle$ is cyclic and $u \in F$. Then*

(1) $T_k^F(u) = 0$ if and only if $u = v - \sigma(v)$ for some $v \in F$.
(2) $N_k^F(u) = 1$ if and only if $u = v/\sigma(v)$ for some $v \in F^*$.

PROOF. Throughout the proof, assume $G = \{1, \sigma, \sigma^2, \ldots, \sigma^{n-1}\}$ and $\sigma^n = 1$.

(1): If $v \in F$, then $T(\sigma(v)) = \sum_{\tau\in G}\tau\sigma(v) = \sum_{\rho\in G}\rho(v) = T(v)$. It follows that $T(v - \sigma(v)) = 0$. Conversely, assume $T(u) = 0$. By Lemma 5.5.1 (2), there exists $w \in F$ with $T(w) = 1$. Starting with

$$\begin{aligned} v = uw + (u + \sigma(u))\sigma(w) + (u + \sigma(u) + \sigma^2(u))\sigma^2(w) + \ldots \\ + (u + \sigma(u) + \sigma^2(u) + \cdots + \sigma^{n-2}(u))\sigma^{n-2}(w), \end{aligned}$$

apply σ to get

$$\begin{aligned} \sigma(v) = \sigma(u)\sigma(w) + (\sigma(u) + \sigma^2(u))\sigma^2(w) + \ldots \\ + (\sigma(u) + \sigma^2(u) + \cdots + \sigma^{n-1}(u))\sigma^{n-1}(w). \end{aligned}$$

Subtract $\sigma(v)$ from v. Use the identities $T(u) = u + \sigma(u) + \cdots + \sigma^{n-1}(u) = 0$ and $T(w) = 1$ to simplify

$$\begin{aligned} v - \sigma(v) &= uw + u\sigma(w) + u\sigma^2(w) + \cdots + u\sigma^{n-2}(w) \\ &\quad - \big(\sigma(u) + \sigma^2(u) + \cdots + \sigma^{n-1}(u)\big)\,\sigma^{n-1}(w) \\ &= u\big((w + \sigma(w) + \sigma^2(w) + \cdots + \sigma^{n-2}(w)\big) - (-u)\sigma^{n-1}(w) \\ &= u\big((w + \sigma(w) + \sigma^2(w) + \cdots + \sigma^{n-2}(w) + \sigma^{n-1}(w)\big) \\ &= uT(w) = u. \end{aligned}$$

(2): If $v \in F^*$, then $N(\sigma(v)) = \prod_{\tau \in G} \tau\sigma(v) = N(v)$. This shows $N(v/\sigma(v)) = 1$. Conversely, assume $N(u) = 1$. By Theorem 5.3.7 we know that

$$v = ux + u\sigma(u)\sigma(x) + u\sigma(u)\sigma^2(u)\sigma^2(x) + \cdots + u\sigma(u)\sigma^2(u)\cdots\sigma^{n-1}(u)\sigma^{n-1}(x)$$

is nonzero for some $x \in F$. In this case, we have

$$u^{-1}v = x + \sigma(u)\sigma(x) + \sigma(u)\sigma^2(u)\sigma^2(x) + \cdots + \sigma(u)\sigma^2(u)\cdots\sigma^{n-1}(u)\sigma^{n-1}(x)$$

and

$$\begin{aligned}\sigma(v) &= \sigma(u)\sigma(x) + \sigma(u)\sigma^2(u)\sigma^2(x) + \cdots + \sigma(u)\sigma^2(u)\cdots\sigma^n(u)\sigma^n(x)\\ &= \sigma(u)\sigma(x) + \sigma(u)\sigma^2(u)\sigma^2(x) + \cdots + N(u)x\\ &= \sigma(u)\sigma(x) + \sigma(u)\sigma^2(u)\sigma^2(x) + \cdots + x.\end{aligned}$$

This shows $\sigma(v) = u^{-1}v$, hence $u = v/\sigma(v)$. □

5.3. Cyclotomic Extensions. Let k be a field. We say F is a *cyclotomic extension of k of order n* if F is the splitting field over k of $x^n - 1$. If $\operatorname{char} k = p > 0$, then we can factor $n = p^e m$ where $(m, p) = 1$. Then $x^n - 1 = (x^m)^{p^e} - 1^{p^e} = (x^m - 1)^{p^e}$, so the splitting field of $x^n - 1$ is equal to the splitting field of $x^m - 1$. For this reason, we assume n is relatively prime to $\operatorname{char} k$ and $x^n - 1$ is separable. In the following, $\phi(n)$ denotes the Euler ϕ-function.

LEMMA 5.5.5. *Let k be any field. If m and n are positive integers and $m \mid n$, then $x^m - 1$ divides $x^n - 1$ in the ring $k[x]$. Conversely, if the characteristic of k does not divide m and $x^m - 1$ divides $x^n - 1$, then $m \mid n$.*

PROOF. We are given that $n = km$, for some $k \geq 1$. Use Mathematical Induction on k. If $k = 1$, then this is trivial. Assume $1 < k$ and that the result holds for $k - 1$. Consider

$$x^n - 1 = (x^m - 1)x^{n-m} + (x^{n-m} - 1).$$

Since $n - m = (k - 1)m$, by Mathematical Induction, $x^m - 1$ divides $x^{n-m} - 1$. Therefore, $x^m - 1$ divides the right hand side.

For the converse, let F be a field extension of k containing all of the roots of $x^n - 1$. By hypothesis, we can factor $x^n - 1 = (x^m - 1)q(x)$ for some $q(x) \in k[x]$. If we let $f = x^m - 1$, then f splits over F. Since $\operatorname{char} k$ does not divide m, we have $\gcd(f, f') = 1$. By Theorem 3.6.17 (1), $f = x^m - 1$ has m distinct roots in F. By Corollary 3.6.11, the set of roots of $x^m - 1$ is a cyclic subgroup of F^* of order m. That is, there exists an element $\alpha \in F^*$ such that α has order m. Then $\alpha^n - 1 = (\alpha^m - 1)q(\alpha) = 0$ says $\alpha^n = 1$. By Lemma 2.2.18, we have $m \mid n$. □

THEOREM 5.5.6. *Let F be a cyclotomic extension of k of order n. If $\operatorname{char} k = p > 1$, assume $\gcd(n, p) = 1$. Then*

(1) $F = k(\zeta)$ where ζ is a primitive nth root of 1 over k.
(2) F is a Galois extension of k and $\operatorname{Aut}_k(F)$ is a subgroup of the group of units in $\mathbb{Z}/n$. The dimension $\dim_k(F)$ is a divisor of $\phi(n)$.

PROOF. (1): By assumption, $x^n - 1$ is separable, and the group μ_n of nth roots of unity in F is a cyclic group of order n, by Corollary 3.6.11. Let ζ be a primitive nth root of unity in F. Therefore $F = k(\zeta)$ is a simple extension.

(2): Since F is the splitting field of a separable polynomial, F/k is Galois by Theorem 5.3.18. The Galois group $G = \operatorname{Aut}_k(F)$ acts on the cyclic group of order n generated by ζ (Corollary 5.3.19). This defines a homomorphism $G \to \operatorname{Aut}(\langle\zeta\rangle)$.

Since $F = k(\zeta)$, this mapping is one-to-one. By Theorem 2.3.30, the order of $\mathrm{Aut}(\langle\zeta\rangle)$ is $\phi(n)$. □

Let F be a cyclotomic extension of k of order n. If $\mathrm{char}\, k = p > 1$, assume $(n, p) = 1$. By Theorem 5.5.6 (1), the group μ_n of nth roots of unity in F is a cyclic group of order n. There are $\phi(n)$ generators of μ_n. The *nth cyclotomic polynomial over* k is

$$\Phi_n(x) = (x - \zeta_1) \cdots (x - \zeta_{\phi(n)})$$

where $\zeta_1, \ldots, \zeta_{\phi(n)}$ are the $\phi(n)$ primitive nth roots of unity in μ_n. We have seen in Examples 5.2.9 and 3.7.8 that if p is a prime number and $k = \mathbb{Q}$, then $\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$ is irreducible in $\mathbb{Q}[x]$.

PROPOSITION 5.5.7. *Assume k is the prime subfield of F and F is a cyclotomic extension of k of order n. If* $\mathrm{char}\, k = p > 1$, *assume $(n, p) = 1$. Then*

(1) $x^n - 1 = \prod_{d|n} \Phi_d(x)$.
(2) $\Phi_n(x) \in k[x]$.
(3) If $k = \mathbb{Q}$, then $\Phi_n(x) \in \mathbb{Z}[x]$.

PROOF. (1): By Theorem 2.3.27, we can partition μ_n into disjoint subsets

$$\mu_n = \bigcup_{d|n} \{\zeta \in \mu_n \mid |\zeta| = d\}.$$

The set elements of order d in μ_n has cardinality $\phi(d)$. The corresponding factorization of $x^n - 1$ is $x^n - 1 = \prod_{d|n} \Phi_d(x)$.

(2): The proof is by induction on n. For $n = 1$, $\Phi_1(x) = x - 1$ is in $k[x]$. Assume $n > 1$ and that (2) is true for all $1 \leq m < n$. Define $g(x) = \prod_{\substack{d|n \\ d \neq n}} \Phi_d(x)$. By our induction hypothesis, $g(x) \in k[x]$. By (1), $x^n - 1 = g(x)\Phi_n(x)$. By the Division Algorithm, Theorem 3.6.3, $\Phi_n(x) \in k[x]$.

(3): In the proof of (2), by the induction hypothesis, $g(x) \in \mathbb{Z}[x]$. Moreover, $g(x)$ is monic, so Theorem 3.6.3 implies $\Phi_n(x) \in \mathbb{Z}[x]$. □

PROPOSITION 5.5.8. *If $\Phi_n(x)$ is the nth cyclotomic polynomial over $\mathbb{Q}$, then $\Phi_n(x)$ is irreducible.*

PROOF. Let F be a cyclotomic extension of order n over the field $\mathbb{Q}$ and $\Phi_n(x)$ the nth cyclotomic polynomial over $\mathbb{Q}$. We know from Proposition 5.5.7 that $\Phi_n(x)$ is a monic polynomial in $\mathbb{Z}[x]$ and has degree $\phi(n)$. By Theorem 3.7.4 (Gauss' Lemma) it suffices to show that $\Phi_n(x)$ is irreducible in $\mathbb{Z}[x]$. Let $f(x)$ be a monic irreducible factor of $\Phi_n(x)$ in $\mathbb{Z}[x]$ and write $\Phi_n(x) = f(x)g(x)$. To complete the proof, we show that $\Phi_n(x) = f(x)$. To do this, we show that $f(x)$ has degree $\phi(n)$. Let $\zeta \in F$ be a root of f. Then ζ is a root of $\Phi_n(x)$, hence is a primitive nth root of unity. By Theorem 2.3.27, a typical primitive nth root of unity is of the form ζ^d, where $0 < d < n$ and $\gcd(d, n) = 1$. We show that each such ζ^d is a root of f. We do this is several steps.

First let p be a prime divisor of d. Then ζ^p is a root of $\Phi_n(x) = f(x)g(x)$. We show ζ^p is a root of f. For contradiction's sake, assume $g(\zeta^p) = 0$. Then ζ is a root of $g(x^p)$. Since $f(x)$ is irreducible, $f = \mathrm{Irr.poly}_{\mathbb{Q}}(\zeta)$ and by Theorem 5.1.4 we have $f(x) \mid g(x^p)$ in $\mathbb{Q}[x]$. By Theorem 3.6.3 applied over $\mathbb{Z}$ and $\mathbb{Q}$, we have $g(x^p) = f(x)h(x)$ where $h \in \mathbb{Z}[x]$. We apply Theorem 3.6.2 (1) to reduce the coefficients of the polynomials modulo p. The image of the polynomial $g(x^p) = f(x)h(x)$

under the natural map $\mathbb{Z}[x] \to \mathbb{Z}/(p)[x]$ will be denoted $[g(x^p)] = [f(x)][h(x)]$. The Frobenius homomorphism $\mathbb{Z}/(p)[x] \to \mathbb{Z}/(p)[x]$ of Exercise 3.2.32 fixes the field $\mathbb{Z}/(p)$, hence $[g(x^p)] = [g(x)]^p = [f(x)][h(x)]$. By unique factorization, some irreducible factor of $[f(x)]$ divides $[g(x)]$. By Proposition 5.5.7 (1), for some $q(x) \in \mathbb{Z}[x]$ we have $x^n - 1 = \Phi_n(x)q(x) = f(x)g(x)q(x)$. Reduce modulo p to get $x^n - 1 = [f(x)][g(x)][q(x)]$. Since $[f(x)]$ and $[g(x)]$ have a common factor, this proves $x^n - 1$ is not separable, a contradiction. This proves ζ^p is a root of $f(x)$.

Now assume $0 < d < n$ and $\gcd(d, n) = 1$. We show that ζ^d is a root of f. Factor $d = p_1 \cdots p_m$ into a product of primes. If $m = 1$, then by the first step, ζ^d is a root of f. Inductively assume $m > 1$ and that $\zeta^{p_1 \cdots p_{m-1}}$ is a root of f. Then by the first step, $\zeta^d = (\zeta^{p_1 \cdots p_{m-1}})^{p_m}$ is a root of f. Since there are $\phi(n)$ choices for d, we have shown f has $\phi(n)$ roots, hence $\Phi_n(x) = f(x)$ is irreducible. □

COROLLARY 5.5.9. *Let F be a cyclotomic extension of order n over the field $\mathbb{Q}$ and $\Phi_n(x)$ the nth cyclotomic polynomial over $\mathbb{Q}$. Then the following are true.*

(1) If $\zeta \in F$ is a primitive nth root of unity, then $\Phi_n(x) = \text{Irr. poly}_{\mathbb{Q}}(\zeta)$.
(2) $F \cong \mathbb{Q}[x]/(\Phi_n)$.
(3) F is a Galois extension of $\mathbb{Q}$, the Galois group $\text{Aut}_{\mathbb{Q}}(F)$ is isomorphic to the group of units in the ring $\mathbb{Z}/(n)$, and $\dim_{\mathbb{Q}}(F) = \phi(n)$.

COROLLARY 5.5.10. *If $n > 1$, then $x^n - 1 = \prod_{d|n} \Phi_d(x)$ is the factorization of $x^n - 1$ into irreducible factors in the unique factorization domain $\mathbb{Z}[x]$.*

5.4. Wedderburn's Theorem. In Theorem 5.5.11, we apply Corollary 5.5.10 and the Class Equation (Corollary 2.4.14) to prove that every finite division ring is a field. This is a classic result of J. H. M. Wedderburn, [**19**]. The proof given here is due to E. Witt, [**32**].

THEOREM 5.5.11. *(Wedderburn's Theorem) Let D be a ring. The following are equivalent.*

(1) D is a finite field.
(2) D is a finite domain.
(3) D is a finite division ring.

PROOF. (2) implies (3): This is Theorem 3.2.21.

(1) implies (3) and (3) implies (2): These are by Definition 3.1.3.

(3) implies (1): Let k be the center of D. It is routine to check that k is a field. Let q denote the order of k. Then D is a finite k-vector space of dimension n, for some $n \geq 1$. Hence, the order of D is q^n. We prove that $n = 1$, hence $k = D$. For contradiction's sake, assume $n > 1$. Then D^*, the group of units of D, is a nonabelian group of order $q^n - 1$. The first part of the proof is to write a formula for the class equation of D^* in terms of the number q. Given a noncentral element $u \in D - k$, we wish to represent the order of the normalizer of u in D^* as a function of q. The inner automorphism defined by u is $\sigma_u : D \to D$ which is the function defined by $\sigma_u(x) = u^{-1}xu$ (see Example 3.2.2 (2)). Let $N_u = \{x \in D \mid ux = xu\}$ be the subset of D fixed by σ_u. As in Proposition 5.3.1, it is routine to check that N_u is a subring of D and N_u is a division ring. Since u is not central, N_u is a proper subring of D which contains k. Then N_u is a finite dimensional k-vector space, and has order q^r, where $1 \leq r < n$. Also, D is a finite dimensional vector space over N_u. If m is the dimension of D as a vector space over N_u, then by Proposition 4.3.8,

$q^n = (q^r)^m$, which implies r divides n. The group of units N_u^* is the normalizer of u in the group D^*. Since N_u is a division ring, the order of N_u^* is $q^r - 1$. The index $[D^* : N_u^*]$ is equal to $(q^n - 1)/(q^r - 1)$ (Corollary 2.2.14). Let $u_1, \ldots, u_t$ be a complete set of representatives for the noncentral conjugacy classes of D^*. For each i, N_{u_i} has order q^{r_i}, for some r_i such that $1 \leq r_i < n$. By Corollary 2.4.14,

$$|D^*| = |k^*| + \sum_{i=1}^{t} [D^* : N_{u_i}^*] \tag{5.3}$$

is the class equation for D^*. From the above, in terms of q, (5.3) is

$$q^n - 1 = (q-1) + \sum_{i=1}^{t} \frac{q^n - 1}{q^{r_i} - 1}. \tag{5.4}$$

This completes the first part of the proof. The second part of the proof is to show that (5.4) leads to a contradiction, if $n > 1$. By the factorization formula of Corollary 5.5.10, we see that the integer $\Phi_n(q)$ divides the left hand side of (5.4) as well as each term $(q^n - 1)/(q^{r_i} - 1)$ appearing in the summation on the right hand side. Therefore, $\Phi_n(q)$ divides $q - 1$. The proof is complete after we show that this is impossible. Since $q \geq 2$, for any primitive nth root of unity ζ in $\mathbb{C}^*$, the reader should verify that $|q - \zeta| > q - 1$. Therefore, if $\zeta_1, \ldots, \zeta_{\phi(n)}$ are the primitive nth roots of unity in $\mathbb{C}^*$, then $|\Phi_n(q)| = |q - \zeta_1| \cdots |q - \zeta_{\phi(n)}| > (q-1)$. Therefore, $\Phi_n(q)$ is not a divisor of $q - 1$. □

5.5. Finite Fields. A finite field has positive characteristic and is finite dimensional over its prime subfield. We prove in Theorem 5.5.14 (9) that a finite extension of finite fields is a cyclic extension.

LEMMA 5.5.12. *Let F be a field and assume* $\operatorname{char} F = p$ *is positive. For any $r > 0$, the mapping $\varphi : F \to F$ defined by $x \mapsto x^{p^r}$ is a homomorphism of fields. If F is finite, then φ is an automorphism and F is a perfect field. If $r = 1$, then φ is called the* Frobenius homomorphism.

PROOF. It follows from Exercise 3.2.32 that φ is a monomorphism. If F is finite, then φ is an automorphism, by Exercise 1.1.11. By Definition 5.4.4, F is a perfect field. □

LEMMA 5.5.13. *For each prime number p and for every $n \geq 1$, there exists a field F of order p^n.*

PROOF. Let k denote the field $\mathbb{Z}/p$. Let $f = x^{p^n} - x \in k[x]$. Let F be the splitting field of f over k. Since $f' = -1$, by Theorem 3.6.17, f has no multiple roots in F. Therefore, f is separable and there are p^n distinct roots of f in F. Let $\varphi : F \to F$ be the automorphism of F defined by $x \mapsto x^{p^n}$. If $u \in F$ is a root of f, then $\varphi(u) = u$. By Exercise 5.3.25, the prime field k is fixed by φ. Since F is generated over k by roots of f, F is fixed point-wise by φ. Every u in F is a root of f, and F has order p^n. □

THEOREM 5.5.14. *Let F be a finite field with* $\operatorname{char} F = p$. *Let k be the prime subfield of F and $n = \dim_k(F)$.*

(1) The group of units of F is a cyclic group.
(2) $F = k(u)$ is a simple extension, for some $u \in F$.
(3) The order of F is p^n.

(4) *F is the splitting field for the separable polynomial $x^{p^n} - x$ over k.*
(5) *F/k is a separable extension. F is a perfect field.*
(6) *Any two finite fields of order p^n are isomorphic as fields.*
(7) *F/k is a Galois extension.*
(8) *The Galois group $\mathrm{Aut}_k(F)$ is cyclic of order n and is generated by the Frobenius homomorphism $\varphi : F \to F$ defined by $\varphi(x) = x^p$.*
(9) *If d is a positive divisor of n, then $E = \{u \in F \mid u^{p^d} = u\}$ is an intermediate field of F/k which satisfies the following.*
   (a) *$\dim_E(F) = n/d$, and $\dim_k(E) = d$.*
   (b) *If φ is the generator for $\mathrm{Aut}_k(F)$, then $\mathrm{Aut}_E(F)$ is the cyclic subgroup generated by φ^d.*
   (c) *E/k is Galois and $\mathrm{Aut}_k(E)$ is the cyclic group of order d generated by the restriction $\varphi|_E$.*
(10) *If E is an intermediate field of F/k, and $d = \dim_k(E)$, then d divides n and E is the field described in Part (9).*

PROOF. Parts (1) – (6) are from Theorem 5.2.12 and Lemma 5.5.12. Parts (7) and (8) are from Example 5.3.16. The proofs of Parts (9) and (10) follow straight from Theorem 5.3.21 and Part (8). □

5.5.1. *Irreducible Polynomials over Finite Fields.* Throughout this section, p will be a fixed prime number and $\mathbb{F}_p = \mathbb{Z}/p$ is the prime field of order p.

THEOREM 5.5.15. *The factorization of the polynomial $x^{p^n} - x$ in $\mathbb{F}_p[x]$ into irreducible factors is equal to the product of all the monic irreducible polynomials of degree d where d runs through all divisors of n.*

PROOF. Is left to the reader. □

THEOREM 5.5.16. *Let $\psi(n)$ denote the number of distinct monic irreducible polynomials of degree n in $\mathbb{F}_p$.*

(1) *If μ is the Möbius function, then $\psi(n) = \dfrac{1}{n}\displaystyle\sum_{d|n} \mu(d)p^{n/d} = \frac{1}{n}\sum_{d|n} \mu\left(\frac{n}{d}\right) p^d$.*
(2) *$\psi(n) > \dfrac{p^n}{2n}$.*

PROOF. (1): By Theorem 5.5.15, $p^n = \displaystyle\sum_{d|n} d\psi(d)$. Now apply the Möbius Inversion Formula (Theorem 1.2.16).

(2): The reader should verify the identities:

$$\begin{aligned} n\psi(n) &= p^n + \sum_{d|n, d<n} \mu\left(\frac{n}{d}\right) p^d \\ &\geq p^n - \sum_{d|n, d<n} p^d \\ &\geq p^n - \sum_{1\leq d\leq n/2} p^d \\ &\geq p^n - p^{\lfloor n/2 \rfloor + 1} \end{aligned}$$

where $\lfloor n/2 \rfloor$ is the greatest integer less than $n/2$. If $n > 2$, then $\lfloor n/2 \rfloor + 1 \leq n - 1$, so

$$\psi(n) > \frac{1}{n}\left(p^n - p^{n-1}\right) = \frac{p^n}{n}\left(1 - \frac{1}{p}\right) \geq \frac{p^n}{2n}.$$

If $n = 2$, the formula can be derived from $\psi(2) = (1/2)(p^2 - p)$. □

5.6. Exercises.

EXERCISE 5.5.17. Let k be a field. Show that for any $n \geq 1$ there exists a polynomial $f \in F[x]$ of degree n such that f has no repeated roots.

EXERCISE 5.5.18. Let F/k be a Galois extension of fields with finite group G. Assume $G = \langle \sigma \rangle$ is cyclic.

(1) Show that the function $D : F^* \to F^*$ defined by $D(u) = u/\sigma(u)$ is a homomorphism of abelian groups.
(2) Show that the kernel of D is k^*, and the image of D is the kernel of $N_k^F : F^* \to k^*$.

EXERCISE 5.5.19. For the cyclic Galois extension $\mathbb{C}/\mathbb{R}$ of degree two, determine the image of the norm map $N_{\mathbb{R}}^{\mathbb{C}} : \mathbb{C}^* \to \mathbb{R}^*$ and show that it is a subgroup of $\mathbb{R}^*$ of index two.

EXERCISE 5.5.20. Let F/k be a Galois extension of fields with finite group G. Assume $G = \langle \sigma \rangle$ is cyclic. This exercise outlines another proof of Part (1) of Hilbert's Theorem 90. Without using Theorem 5.5.4, prove:

(1) The function $D : F \to F$ defined by $D(x) = x - \sigma(x)$ is a homomorphism of k-vector spaces.
(2) The kernel of D is k, and the image of D is the kernel of the trace map $T_k^F : F \to k$.

EXERCISE 5.5.21. Let F/k be an extension of fields and assume $\dim_k F = n$ is finite. As in Lemma 5.3.6, the left regular representation $\lambda : F \to \mathrm{Hom}_k(F, F)$ makes $\mathrm{Hom}_k(F, F)$ into a left F-vector space. Prove:

(1) If $\{v_1, \ldots, v_n\}$ is a k-basis for F and $\{\phi_1, \ldots, \phi_n\}$ is an F-basis for the ring of endomorphisms $\mathrm{Hom}_k(F, F)$, then the matrix $(\phi_i(v_j))$ is invertible in $M_n(F)$.
(2) If F/k is a Galois extension of fields with group $G = \{\sigma_1, \ldots, \sigma_n\}$, then the matrix $(\sigma_i(v_j))$ in $M_n(F)$ is invertible.

EXERCISE 5.5.22. Prove Theorem 5.5.15.

EXERCISE 5.5.23. Let K be a finite field of order p^d. As in Theorem 5.5.16, let $\psi(n)$ be the number of irreducible monic polynomials of degree n in $\mathbb{F}_p[x]$. If $d \mid n$, show that there are at least $\psi(n)$ irreducible monic polynomials of degree n/d in $K[x]$.

EXERCISE 5.5.24. Let k be a finite field and K/k a finite dimensional extension of fields, with $\dim_k K = d$. Let n be an arbitrary positive integer and $A = K \oplus \cdots \oplus K$ the direct sum of n copies of K.

(1) Show that if there exists a surjective k-algebra homomorphism $f : k[x] \to A$, then there exist at least n distinct irreducible monic polynomials in $k[x]$ of degree d.

(2) Find an example of k and A such that the k-algebra A is not the homomorphic image of $k[x]$.
(3) Show that for some integer $m \geq 1$, there exist n distinct irreducible monic polynomials $h_1, \ldots, h_n$ in $k[x]$ such that each h_i has degree md.
(4) Show that for some integer $m \geq 1$, if F/k is a finite extension field with $\dim_k F = md$, then the direct sum $F \oplus \cdots \oplus F$ of n copies of F is the homomorphic image of $k[x]$. Show that m can be chosen to be relatively prime to d.
(5) Show that there is a separable polynomial $g \in k[x]$ such that A is isomorphic to a subalgebra of $k[x]/(g)$.

EXERCISE 5.5.25. Classify up to isomorphism all finite rings of order four. For a generalization of this result to rings of order p^2, p a prime number, see Exercise 5.5.26. The reader interested in rings that do not necessarily contain a unit element is referred to the classification obtained in [**26**].

EXERCISE 5.5.26. Let p be a prime number and A a finite ring of order p^2.

(1) Prove that either A is isomorphic to $\mathbb{Z}/(p^2)$, or the characteristic of A is p and A is isomorphic as $\mathbb{Z}/p$-algebras to $(\mathbb{Z}/p)[x]/(\phi)$, for some monic quadratic polynomial ϕ with coefficients in the field $\mathbb{Z}/p$.
(2) Prove that A is commutative.
(3) Prove that A is isomorphic to exactly one of the following four rings:
   (a) $\mathbb{Z}/(p^2)$ (if $\mathrm{char}(A) = p^2$).
   (b) $\mathbb{Z}/p \oplus \mathbb{Z}/p$ (if $\mathrm{char}(A) = p$ and ϕ factors and is separable).
   (c) $(\mathbb{Z}/p)[x]/(x^2)$ (if $\mathrm{char}(A) = p$ and ϕ is a square).
   (d) a finite field of order p^2 (if $\mathrm{char}(A) = p$ and ϕ is irreducible).

EXERCISE 5.5.27. If F/k is an extension of finite fields, show that the image of the norm map $N_k^F : F^* \to k^*$ is equal to k^*.

EXERCISE 5.5.28. Let $k = \mathbb{F}_2$ be the finite field of order 2. Show that the factor ring $F = k[x, y]/(x^2 + x + 1, y^3 + y + 1)$ is a field. Determine the order of F.

6. Cyclic Galois Extensions

We say a finite Galois extension of fields F/k is *cyclic of degree* n if the group $\mathrm{Aut}_k(F)$ is a cyclic group of order n.

6.1. Artin-Schreier Theorem. Let p be a prime number and k a field of characteristic p. In [**3**], E. Artin and O. Schreier described the Galois extensions F/k of degree p. Their results are summarized in Theorem 5.6.2 below, the so-called Artin-Schreier Theorem.

EXAMPLE 5.6.1. Let k be a field of positive characteristic p. For any $a \in k$, the polynomial $f = x^p - x - a \in k[x]$ is separable over k. To see this, assume u is a root of f in any extension field F/k. Let $i \in \mathbb{Z}/p$ be any element of the prime field of k. Then $f(u + i) = (u + i)^p - (u + i) - a = u^p + i - u - i - a = f(u) = 0$. Therefore, f has p distinct roots in F, namely $u, u + 1, \ldots, u + p - 1$.

THEOREM 5.6.2. *(Artin-Schreier) Let k be a field of positive characteristic p.*

(1) If F/k is a cyclic Galois extension of degree p, then there exists $a \in k$ such that $f = x^p - x - a$ is an irreducible separable polynomial over k and F is the splitting field for f over k. In this case $F = k(u)$, where u is any root of f.

(2) If $a \in k$ and $f = x^p - x - a$, then
  (a) f is separable, and
  (b) either f is irreducible over k, or splits in $k[x]$.
(3) If $a \in k$ and $f = x^p - x - a$ is irreducible over k, then
  (a) $F = k[x]/(f)$ is a splitting field for f, and
  (b) F/k is a cyclic Galois extension of k of degree p.

PROOF. (1): Let $G = \mathrm{Aut}_k(F) = \langle\sigma\rangle$. Since G is simple and abelian (Exercise 2.2.30), there are no proper intermediate fields for F/k. Since $\mathrm{char}(k) = \dim_k(F) = p$, $T_k^F(1) = p = 0$. By Theorem 5.5.4, there is $v \in F$ such that $v - \sigma(v) = 1$. If $u = -v$, then $\sigma(u) = 1 + u$. This shows $u \notin k$, hence $F = k(u)$. Note that $\sigma(u^p) = (\sigma(u))^p = (1+u)^p = 1 + u^p$, and $\sigma(u^p - u) = \sigma(u^p) - \sigma(u) = (1+u^p) - (u+1) = u^p - u$. If $a = u^p - u$, then $a \in k$ and u satisfies the polynomial $f = x^p - x - a$. Since the dimension of $k(u)$ over k is p, this implies f is equal to the irreducible polynomial of u. By Example 5.6.1, f is separable and splits in F.

(2) and (3): Let $f = x^p - x - a$ in $k[x]$. Let F be a splitting field for f. As was shown in Example 5.6.1, f is separable and if $u \in F$ is a root of f, then the p distinct roots of f are $u, u+1, \ldots, u+p-1$, hence $F = k(u)$. By Theorem 5.3.18, F/k is a Galois extension. For any τ in $\mathrm{Aut}_k(F)$, by Proposition 5.3.2, $\tau(u)$ is a root of f. Thus, $\tau(u) - u$ is an element of the prime field $\mathbb{Z}/p$. Define a function $\theta : \mathrm{Aut}_k(F) \to \mathbb{Z}/p$ by $\theta(\tau) = \tau(u) - u$. If σ is another element of $\mathrm{Aut}_k(F)$, then $\sigma(\tau(u) - u) = \tau(u) - u$. Hence $\sigma\tau(u) - \sigma(u) = \tau(u) - u$. From this we see that

$$\sigma\tau(u) - u = \sigma(u) + \tau(u) - u - u. \tag{6.1}$$

The left hand side of (6.1) is $\theta(\sigma\tau)$, the right hand side is $\theta(\sigma) + \theta(\tau)$. This shows θ is a homomorphism from the group $\mathrm{Aut}_k(F)$ to the additive cyclic group $\mathbb{Z}/p$. By Proposition 5.3.2, θ is one-to-one. Since $\mathbb{Z}/p$ is a simple group, either $\mathrm{Aut}_k(F)$ has order 1 or p. By Theorem 5.3.21, if $\mathrm{Aut}_k(F)$ has order 1, then $F = k$ and f splits in $k[x]$. If $\mathrm{Aut}_k(F)$ has order p, then $\dim_k(F) = p$. Since $F = k(u)$, by Theorem 5.1.4, $\mathrm{Irr.poly}_k(u)$ has degree p. Therefore, $f = \mathrm{Irr.poly}_k(u)$. □

6.2. Kummer Theory. Kummer Theory is the study of cyclic Galois extensions of a field containing sufficiently many roots of unity and is named after E. Kummer, a nineteenth century German number theorist. If $\zeta \in k^*$ and ζ generates a subgroup of order n in k^*, then we say ζ is a *primitive nth root of 1 in k* and write $\zeta = \sqrt[n]{1}$. There are at most n solutions to $x^n - 1$ in k, so by Theorem 2.3.27 the subgroup $\langle\zeta\rangle$ has $\phi(n)$ generators. That is, if k contains a primitive nth root of 1, then k contains $\phi(n)$ primitive nth roots of 1. A cyclic extension F/k of degree n is called a *Kummer extension* if $\sqrt[n]{1} \in k$.

THEOREM 5.6.3. *Let $n > 0$ and assume k is a field containing a primitive nth root of 1. The following are equivalent.*
(1) F/k is a cyclic Galois extension of degree d, for some positive divisor d of n.
(2) F is a splitting field over k of $x^n - a$ for some $a \in k^*$.
(3) F is a splitting field over k of $x^d - a$ for some $a \in k^*$ and some positive divisor d of n.

PROOF. Throughout the proof, let $\zeta = \sqrt[n]{1}$ be a primitive nth root of 1 in k.

(2) implies (1): Let α be a root of $x^n - a$ in F. For each $i \geq 0$ we have $\left(\zeta^i\alpha\right)^n = \left(\zeta^n\right)^i \alpha^n = a$, so the roots of $x^n - a$ in F are $\{\zeta^i\alpha \mid 0 \leq i < n\}$. This

shows $x^n - a$ is separable. Also, since $\zeta \in k$, this implies $F = k(\alpha)$ is a simple extension. If $\sigma \in G = \mathrm{Aut}_k(F)$, then $\sigma(\alpha) = \zeta^i \alpha$ for some i such that $0 \leq i < n$. As σ runs through the nonidentity elements of G, consider the positive numbers i such that $\sigma(\alpha) = \zeta^i \alpha$ and pick the smallest. Fix $\sigma \in G$, such that $\sigma(\alpha) = \zeta^i \alpha$ and i is minimal. We prove that G is generated by σ. Let τ be any element of G. Then $\tau(\alpha) = \zeta^j \alpha$ and we can assume $0 < i \leq j < n$. Dividing, $j = iq + r$, where $0 \leq r < i$. Now $\sigma^q(\alpha) = \zeta^{qi}\alpha$. Therefore, $\sigma^{-q}\tau(\alpha) = \sigma^{-q}(\zeta^j \alpha) = \zeta^j \sigma^{-q}(\alpha) = \zeta^j \zeta^{-qi}\alpha = \zeta^r \alpha$. By the choice of i we conclude that $r = 0$, so $\tau = \sigma^q$. The order of G is equal to the order of ζ^i, which is a divisor of n.

(3) implies (2): Assume F is the splitting field of $x^d - a$ where d is a divisor of n, and $a \in k$. Let $\rho = \zeta^{n/d}$. Then $\rho = \sqrt[d]{1}$. Let $\alpha \in F$ satisfy $\alpha^d = a$. Then $x^d - a$ factors in $F[x]$ as $(x - \alpha)(x - \rho\alpha) \cdots (x - \rho^{d-1}\alpha)$. This implies $F = k(\alpha)$, because $\rho \in k$. Consider the polynomial $x^n - a^{n/d}$. For any i such that $0 \leq i < n$ we see that $\left(\zeta^i \alpha\right)^n = \alpha^n = \left(\alpha^d\right)^{n/d} = a^{n/d}$. So $x^n - a^{n/d}$ splits in F.

(1) implies (3): Assume F/k is cyclic of degree d and that σ is a generator for $G = \mathrm{Aut}_k(F)$. Since $\rho = \zeta^{n/d} = \sqrt[d]{1}$ is in k, the norm of ρ is $N(\rho) = \rho^d = 1$. By Theorem 5.5.4, there is $u \in F^*$ such that $\rho = u/\sigma(u)$. Setting $v = u^{-1}$, we have $\rho = v^{-1}\sigma(v)$, or $\sigma(v) = \rho v$. Then $\sigma(v^d) = (\rho v)^d = v^d$. This says $v^d \in k$ and v satisfies the polynomial $x^d - v^d$. The roots of $x^d - v^d$ are $\{v, \rho v, \ldots, \rho^{d-1} v\}$. Note that $\sigma^i(v) = \rho^i v$, for all i such that $0 \leq i < d$. If f is the irreducible polynomial for v, then f has d roots in F. Therefore $\deg(f) = d$ and $f = x^d - v^d$. We have shown that F is the splitting field of f and $F = k(v)$. □

6.3. Radical Extensions. Let α and β be elements of a field F, such that $\alpha^n = \beta$, for some $n \geq 2$. Then β is the nth power of α, and α is an nth root of β. In this case we say α is a *radical* of β. Before stating the general definition of solvability by radicals, we consider a case familiar to everyone, the solution of a quadratic equation. If k is a field with characteristic not equal to 2 and $f = ax^2 + bx + c$ is a polynomial of degree 2 over k, then the roots of f in a splitting field for f over k are given by the quadratic formula. They are $\left(-b \pm \sqrt{b^2 - 4ac}\right)/(2a)$. If we set $u = \sqrt{b^2 - 4ac}$, then the field $k(u)$ is a splitting field for f. The element u is a radical of $b^2 - 4ac$ and we say $k(u)$ is a radical extension of k. Since the roots of f lie in a radical extension, the quadratic polynomial f is said to be solvable by radicals.

Definition 5.6.4. Let k be a field. A *radical tower* over k is a tower of field extensions

$$k = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_n$$

such that for each $i \geq 1$, $F_i = F_{i-1}(u_i)$ and $u_i^{r_i} \in F_{i-1}$ for some positive integer r_i. We say F_n is a *radical extension* of k. Notice that $F_i = k(u_1, \ldots, u_i)$, for $i = 1, \ldots, n$. If $f(x) \in k[x]$, we say f is *solvable by radicals* in case there is a radical extension F/k such that f splits over F.

In Lemma 5.6.5 below we prove that when the Embedding Theorem, Corollary 5.3.20, is applied to embed a radical extension F/k in a Galois extension K/k, then the extension K/k is also a radical extension.

Lemma 5.6.5. *Let F/k be a finite dimensional separable extension of fields. Then there is a field K satisfying the following.*

(1) $k \subseteq F \subseteq K$ *is a tower of field extensions.*
(2) K/k *is a Galois extension.*
(3) There exist intermediate fields $F_1, \ldots, F_m$ *of* K/k *such that*
  (a) each F_i *is isomorphic to* F *as a* k*-algebra, and*
  (b) $K = F_1 F_2 \cdots F_m$.
(4) If F/k *is a radical extension, then* K/k *is a radical extension.*

PROOF. Write $F = k(u_1, \ldots, u_n)$. For each i in $\{1, \ldots, n\}$, let f_i be the irreducible polynomial $\operatorname{Irr.poly}_k(u_i)$. Let K be the splitting field for $f_1 \cdots f_n$ over F. By Corollary 5.3.20, the field K satisfies parts (1) and (2). We prove that K satisfies (3). Let $\alpha \in K$ be an arbitrary root of $f_1 \cdots f_n$. Then α is a root of f_i, for some i. By Theorem 5.1.5, there is a k-algebra isomorphism $\theta : k(u_i) \to k(\alpha)$. By Lemma 5.2.7, θ extends to a k-algebra automorphism $\bar{\theta} : K \to K$. Then $\bar{\theta}(F)$ is an intermediate field of K/k which is k-isomorphic to F and contains α. Since K/k is generated by the roots α of $f_1 \cdots f_n$, there is a finite number of fields of the form $\bar{\theta}(F)$ that generate K.

(4): We are given $F = k(u_1, \ldots, u_n)$, where $u_i^{r_i}$ is in $k(u_1, \ldots, u_{i-1})$. Let $F_1, \ldots, F_m$ be as in (3). For each i, there is a k-algebra isomorphism $F_i \cong F$. Therefore, F_i is a radical extension of k. For each j we have $F_j = k(u_{j1}, \ldots, u_{jn})$, where $u_{ji}^{r_i}$ is in $k(u_{j1}, \ldots, u_{j,i-1})$. Therefore

$$K = F_1 F_2 \cdots F_m = k(u_{11}, \ldots, u_{1n}, u_{21}, \ldots, u_{2n}, \ldots, u_{m1}, \ldots, u_{mn})$$

is a radical extension of k. □

Although Theorems 5.6.6 and 5.6.7 below are true in more generality, we assume in both that the base field k has characteristic zero. Moreover, in Theorem 5.6.6 we assume that the polynomial $x^n - 1$ splits in k for each $n \geq 1$. These restrictions are for the sake of brevity of the presentation as well as the simplification of the proofs. In Theorem 5.6.6 we show that if f is a polynomial over k that is solvable by radicals, then the Galois group of f is necessarily a solvable group. In other words, if the Galois group of f over k is not solvable, then f is not solvable by radicals. In Theorem 5.6.7 we show that if the Galois group of f is solvable, then f is solvable by radicals. We know from Corollary 2.10.14 that the symmetric group S_n is solvable if and only if $n \leq 4$. By Corollary 5.3.19, the splitting field of an irreducible polynomial of degree 4 or less embeds as a subgroup of S_4. Hence, we prove that in characteristic zero a polynomial of degree 4 or less is solvable by radicals. The famous theorem of Abel says that the general polynomial of degree 5 or higher is not solvable by radicals. This is proved below in Corollary 5.7.11.

THEOREM 5.6.6. *Let* k *be a field of characteristic zero and assume for each* $n > 0$ *that* $x^n - 1$ *splits over* k. *Let* $p(x) \in k[x]$. *If* $p(x)$ *is solvable by radicals over* k, *then the Galois group of* $p(x)$ *is a solvable group.*

PROOF. Since $p(x)$ is solvable by radicals, there is a radical tower

$$k = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_n$$

positive integers $r_1, \ldots, r_n$ such that $F_i = F_{i-1}(u_i)$, $u_i^{r_i} \in F_{i-1}$, and $p(x)$ splits over F_n. By Lemma 5.6.5, we can assume F_n is a Galois extension over k. By Kummer Theory (Theorem 5.6.3), F_i is a Galois extension of F_{i-1} and $\operatorname{Aut}_{F_{i-1}} F_i$ is cyclic. By the Fundamental Theorem of Galois Theory (Theorem 5.3.21), F_n is

Galois over F_i, $\mathrm{Aut}_{F_i}(F_n)$ is a normal subgroup of $\mathrm{Aut}_{F_{i-1}}(F_n)$ and

$$\mathrm{Aut}_{F_{i-1}} F_i \cong \mathrm{Aut}_{F_{i-1}}(F_n)/\mathrm{Aut}_{F_i}(F_n).$$

Therefore the series of groups

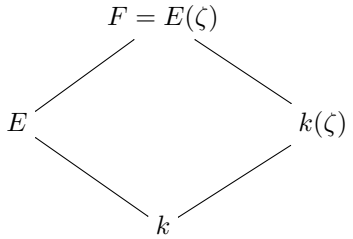
$$\begin{aligned}\mathrm{Aut}_k(F_n) \supseteq \mathrm{Aut}_{F_1}(F_n) \supseteq \\ \cdots \supseteq \mathrm{Aut}_{F_{i-1}}(F_n) \supseteq \mathrm{Aut}_{F_i}(F_n) \supseteq \cdots \\ \supseteq \mathrm{Aut}_{F_{n-1}}(F_n) \supseteq \langle e\rangle\end{aligned}$$

is a normal series and at each step the quotient is an abelian group. So the series is a solvable series for $\mathrm{Aut}_k F_n$. Let E be the splitting field for $p(x)$ over k in F_n. Then E is an intermediate field. By Theorem 5.3.18, E is a Galois extension of k. By the Fundamental Theorem of Galois Theory, $\mathrm{Aut}_k E$ is a homomorphic image of $\mathrm{Aut}_k F_n$. By Exercise 2.10.18 , $\mathrm{Aut}_k E$ is solvable. □

Theorem 5.6.7 is a partial converse to Theorem 5.6.6. In characteristic zero, if f is a polynomial with solvable Galois group, then f is solvable by radicals.

THEOREM 5.6.7. *Let k be a field of characteristic zero, $f \in k[x]$ a separable polynomial and E a splitting field for f. If $\mathrm{Aut}_k(E)$ is solvable, then f is solvable by radicals. That is, there exists a radical extension of k that contains E.*

PROOF. Let $n = \dim_k(E)$. Let $F = E(\zeta)$ be a cyclotomic extension of E of order n. That is, ζ is a primitive nth root of unity over k.



By Theorem 5.3.18, E/k is a Galois extension and by hypothesis $\mathrm{Aut}_k(E)$ is a solvable group. By Theorem 5.4.6, $F = E(\zeta)$ is a Galois extension of $k(\zeta)$ and $G = \mathrm{Aut}_{k(\zeta)}(F)$ embeds as a subgroup of $\mathrm{Aut}_k(E)$. By Exercise 2.10.18, G is a solvable group. By Exercise 2.10.20, G has a composition series $G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_m = \langle e\rangle$ where the factor group G_i/G_{i+1} is cyclic of order $[G_i : G_{i+1}]$, a prime divisor of $|G|$. By Theorem 5.3.21 there is a tower of field extensions $F = F_0 \supseteq F_1 \supseteq F_2 \supseteq \cdots \supseteq F_m = k(\zeta)$ and F_i/F_{i+1} is a cyclic extension, hence a Kummer extension. By Theorem 5.6.3, $F_i = F_{i+1}(v_i)$ is a radical extension. Since $k(\zeta)$ is a radical extension, this proves F/k is a radical extension. □

6.4. Exercises.

EXERCISE 5.6.8. Let k be a field, $n \geq 1$ and $a \in k$. Let $f = x^n - a$ and F/k a splitting field for f. Show that the following are equivalent

(1) Every root of f in F is a simple root.
(2) $F[x]/(f)$ is a direct sum of fields.
(3) $n = 1$ or $na \neq 0$.

EXERCISE 5.6.9. This exercise is a continuation of Exercise 4.2.25. Let R be a UFD with quotient field K. Assume the characteristic of R is not equal to 2. Let $a \in R$ be an element which is not a square in R and $f = x^2 - a \in R[x]$. Let $S = R[x]/(f)$, $L = K[x]/(f)$.

(1) Show that there is a commutative square

$$\begin{array}{ccc} S & \longrightarrow & L \\ \uparrow & & \uparrow \\ R & \longrightarrow & K \end{array}$$

where each arrow is the natural map and each arrow is one-to-one.
(2) Show that L is the quotient field of S.
(3) $\mathrm{Aut}_K L = \langle \sigma \rangle$ is a cyclic group of order two and L/K is a Galois extension.
(4) If $\sigma : L \to L$ is the automorphism of order two, then σ restricts to an R-automorphism of S.
(5) The norm map $N_K^L : L \to K$ restricts to a norm map $N_R^S : S \to R$.

EXERCISE 5.6.10. Let p be a prime number, and F/k an extension of fields which is cyclic of degree p^n. If E is an intermediate field such that $F = E(a)$, and E/k is cyclic of degree p^{n-1}, then $F = k(a)$.

EXERCISE 5.6.11. Let k be a field of positive characteristic p.

(1) The map $a \mapsto a^p - a$ defines a homomorphism of additive groups $\varphi : k \to k$. Prove that a cyclic extension field E/k of degree p exists if and only if the map φ is not onto.
(2) In this exercise, we outline a proof that a cyclic extension field E/k of degree p^{n-1} can be embedded in a cyclic extension field F/k of degree p^n. For the complete classification of cyclic extensions F/k of degree p^n, the interested reader is referred to [**1**]. Assume $n > 1$, E/k is cyclic of degree p^{n-1}, and $\mathrm{Aut}_k(E) = \langle \sigma \rangle$.
   (a) Show that there exists $a, b \in E$ satisfying: $T_k^E(a) = 1$ and $\sigma(b) - b = a^p - a$.
   (b) Show that $x^p - x - a$ is irreducible in $E[x]$.
   (c) Let $F = E[x]/(x^p - x - a)$. Show that F/E is cyclic of degree p and F/k is cyclic of degree p^n.

EXERCISE 5.6.12. Let K be a finite extension field of $\mathbb{Q}$. Prove that K contains only a finite number of roots of unity.

7. Transcendental Field Extensions

For a finite extension of fields K/k we prove that a transcendence base exists and any two transcendence bases have the same number of elements. Therefore, the transcendence degree of the extension is well defined. These notions play important roles in Algebraic Geometry. The field of rational functions K on an algebraic variety V is a finite extension of the ground field k. The transcendence degree of K/k is equal to the dimension of V. In other words, the number of topological degrees of freedom on V is equal to the number of algebraic degrees of freedom in K. In a fundamental theorem on symmetric rational functions we prove that the transcendence degree of the field of symmetric rational functions in n variables over k is equal to n. Moreover, we show that the field of symmetric rational functions is

generated by the elementary symmetric polynomials, hence a transcendence base is constructed. In a fundamental theorem on symmetric polynomials we prove that the ring of symmetric polynomials is generated by the elementary symmetric polynomials. In fact, we show that the ring of symmetric polynomials contains a transcendence base for the field of symmetric rational functions. This is called a globalization result, because rational functions in general have a nonempty pole set, but polynomials do not. There is a version of Emmy Noether's Normalization Lemma (see [**10**, Theorem 8.4.6]) that says under certain sufficient conditions a transcendence base can be constructed globally.

7.1. Transcendence Bases.

DEFINITION 5.7.1. Let F/k be an extension of fields and $\Xi \subseteq F$. We say Ξ is *algebraically dependent* over k if there exist n distinct elements $\xi_1, \ldots, \xi_n$ in Ξ and a nonzero polynomial $f \in k[x_1, \ldots, x_n]$ such that $f(\xi_1, \ldots, \xi_n) = 0$. Otherwise we say Ξ is *algebraically independent*. A *transcendence base* for F/k is a subset $\Xi \subseteq F$ which satisfies

(1) Ξ is algebraically independent over k and
(2) if $\Xi \subseteq Z$ and Z is algebraically independent over k, then $\Xi = Z$.

LEMMA 5.7.2. *Let F/k be an extension of fields and Ξ a subset of F which is algebraically independent over k. For $u \in F - k(\Xi)$, the following are equivalent*

(1) $\Xi \cup \{u\}$ is algebraically independent over k.
(2) u is transcendental over $k(\Xi)$.

PROOF. (2) implies (1): Suppose there exist a polynomial f in $k[x_1, \ldots, x_n]$ and elements $\xi_1, \ldots, \xi_{n-1}$ in Ξ such that $f(\xi_1, \ldots, \xi_{n-1}, u) = 0$. Expand f as a polynomial in x_n with coefficients in $k[x_1, \ldots, x_{n-1}]$, say $f = \sum_j h_j x_n^j$. Then $0 = f(\xi_1, \ldots, \xi_{n-1}, u) = \sum_j h_j(\xi_1, \ldots, \xi_{n-1}) u^j$. But u is transcendental over $k(\Xi)$, so $h_j(\xi_1, \ldots, \xi_{n-1}) = 0$ for each j. But Ξ is algebraically independent, so each polynomial $h_j = 0$. Thus $f = 0$.

(1) implies (2): We prove the contrapositive. Assume u is algebraic over $k(\Xi)$ and $f = \text{min. poly}_{k(\Xi)}(u) = x^m + h_{m-1}x^{m-1} + \cdots + h_1 x + h_0$. Each h_j is in $k(\Xi)$, so there is a finite subset $\xi_1, \ldots, \xi_n$ of Ξ and polynomials $\alpha_0, \ldots, \alpha_m$, $\beta_0, \ldots, \beta_m$ in $k[x_1, \ldots, x_n]$ such that $h_j = \alpha_j(\xi_1, \ldots, \xi_n)/\beta_j(\xi_1, \ldots, \xi_n)$. Multiply across by the least common multiple, β, of the denominators to get

$$f(x)\beta(\xi_1, \ldots, \xi_n) = \sum_j \gamma_j(\xi_1, \ldots, \xi_n) x^j$$

where $\beta(\xi_1, \ldots, \xi_n) \neq 0$ and each γ_j is in $k[x_1, \ldots, x_n]$. Since $(f\beta)(\xi_1, \ldots, \xi_n, u) = 0$, we are done. □

LEMMA 5.7.3. *Let F/k be an extension of fields and Ξ a subset of F which is algebraically independent over k. The following are equivalent.*

(1) F is algebraic over $k(\Xi)$.
(2) Ξ is a transcendence base for F over k.

PROOF. (1) implies (2): Suppose Z is linearly independent, $Z \supseteq \Xi$, and $z \in Z$. Then z is algebraic over $k(\Xi)$, so by Lemma 5.7.2, $\Xi \cup \{z\}$ is linearly dependent. Therefore, $z \in \Xi$, which implies $Z = \Xi$.

(2) implies (1): We prove the contrapositive. Suppose $u \in F - k(\Xi)$ and u is transcendental over $k(\Xi)$. By Lemma 5.7.2, $\Xi \cup \{u\}$ is algebraically independent, so Ξ is not a transcendence base. □

LEMMA 5.7.4. *Let F be a finitely generated field extension of k. Then the following are true:*

(1) If Ξ is a finite subset of F such that F is algebraic over $k(\Xi)$, then there is a subset of Ξ that is a transcendence base for F/k.
(2) There is a finite transcendence base for F/k.

PROOF. We prove (1). The reader should verify (2). Let Ξ be a finite subset of F such that F is algebraic over $k(\Xi)$. Consider the finite set

$$S = \{Z \subseteq \Xi \mid Z \text{ is algebraically independent over } k\}$$

ordered by set containment. Then S contains a maximal member, call it X. Given $u \in \Xi$, by Lemma 5.7.2, u is algebraic over $k(X)$. By Proposition 5.1.10 (3), $k(\Xi)$ is algebraic over $k(X)$. By Proposition 5.1.10 (4), F is algebraic over $k(X)$. By Lemma 5.7.3, X is a transcendence base. □

THEOREM 5.7.5. *Let F/k be an extension of fields and assume $\Xi = \{\xi_1, \ldots, \xi_n\}$ is a transcendence base for F over k. If Z is another transcendence base for F over k, then Z also has cardinality n.*

PROOF. Step 0: If $n = 0$, then by Exercise 5.7.19, F/k is an algebraic extension. Since Z is algebraically independent over k, we conclude that $Z = \emptyset$. Assume from now on that $n > 0$.

Step 1: There exists $\zeta_1 \in Z$ such that $\zeta_1, \xi_2, \ldots, \xi_n$ is a transcendence base for F/k. First we show that there exists $\zeta \in Z$ such that ζ is transcendental over $K = k(\xi_2, \ldots, \xi_n)$. Assume the contrary. Then F is algebraic over $K(Z)$ and $K(Z)$ is algebraic over K, hence F is algebraic over K. Then ξ_1 is algebraic over K, which contradicts Lemma 5.7.2. Suppose $\zeta_1 \in Z$ and ζ_1 is transcendental over K. By Lemma 5.7.2, $\{\zeta_1, \xi_2, \ldots, \xi_n\}$ is algebraically independent over k. The set $\{\zeta_1, \xi_2, \ldots, \xi_n\} \cup \{\xi_1\}$ is algebraically dependent, so Lemma 5.7.2 says ξ_1 is algebraic over $k(\zeta_1, \xi_2, \ldots, \xi_n)$. In this case, the field $k(\Xi)(\zeta_1) = k(\zeta_1, \xi_2, \ldots, \xi_n)(\xi_1)$ is algebraic over $k(\zeta_1, \xi_2, \ldots, \xi_n)$ and F is algebraic over $k(\Xi)(\zeta_1) = k(\zeta_1, \xi_2, \ldots, \xi_n)(\xi_1)$, hence by Proposition 5.1.10 (4), F is algebraic over $k(\zeta_1, \xi_2, \ldots, \xi_n)$. By Lemma 5.7.3, the set $\zeta_1, \xi_2, \ldots, \xi_n$ is a transcendence base for F/k.

Step 2: Iterate Step 1 $n-1$ more times to get a subset $\{\zeta_1, \ldots, \zeta_n\}$ of Z which is a transcendence base for F/k. By Definition 5.7.1, this implies $Z = \{\zeta_1, \ldots, \zeta_n\}$. □

DEFINITION 5.7.6. Let F/k be an extension of fields such that a finite transcendence base exists. The *transcendence degree* of F/k, denoted $\mathrm{tr.deg}_k(F)$, is the number of elements in any transcendence base of F over k.

THEOREM 5.7.7. *Suppose $k \subseteq F \subseteq K$ is a tower of field extensions. Assume $\Xi = \{\xi_1, \ldots, \xi_n\}$ is a transcendence base for F/k and $Z = \{\zeta_1, \ldots, \zeta_m\}$ is a transcendence base for K/F. Then*

(1) $\Xi \cup Z$ is a transcendence base for K/k, and
(2) $\mathrm{tr.deg}_k(K) = \mathrm{tr.deg}_k(F) + \mathrm{tr.deg}_F(K)$.

PROOF. (2): Follows straight from (1).

(1): The reader should verify that K is algebraic over $k(Z \cup \Xi)(F)$ and $k(Z \cup \Xi)(F)$ is algebraic over $k(Z \cup \Xi)$. Therefore, K is algebraic over $k(Z \cup \Xi)$. Let f be a polynomial in $k[x_1, \ldots, x_n][z_1, \ldots, z_m]$ such that $f(\xi_1, \ldots, \xi_n, \zeta_1, \ldots, \zeta_m) = 0$. Since Z is algebraically independent over F, this implies $f(\xi_1, \ldots, \xi_n, z_1, \ldots, z_m)$ is the zero polynomial in the ring $k(\xi_1, \ldots, \xi_n)[z_1, \ldots, z_m]$. Therefore, each coefficient of $f(\xi_1, \ldots, \xi_n, z_1, \ldots, z_m)$ is an algebraic relation over k involving $\xi_1, \ldots, \xi_n$. Because $\xi_1, \ldots, \xi_n$ are algebraically independent over k, we conclude that $f = 0$. This proves $Z \cup \Xi$ is algebraically independent over k. By Lemma 5.7.3 we are done. □

7.2. Symmetric Rational Functions. Throughout this section $n \geq 2$, k is a field and $A = k[x_1, \ldots, x_n]$ is the ring of polynomials over k in the variables $x_1, \ldots, x_n$ (see Section 3.6.1). The field of rational functions in $x_1, \ldots, x_n$ over k is denoted $K = k(x_1, \ldots, x_n)$. Let S_n be the symmetric group on $\{1, 2, \ldots, n\}$. The group S_n acts on A as a group of k-algebra automorphisms in the following way. By Exercise 2.7.10, S_n acts on the set $\{x_1, \ldots, x_n\}$ by the rule $\sigma * x_i = x_{\sigma^{-1}(i)}$, for any permutation $\sigma \in S_n$. Therefore, if $f(x_1, \ldots, x_n)$ is any polynomial in A, define $\sigma(f)$ to be the polynomial $f(x_{\sigma^{-1}(1)}, \ldots, x_{\sigma^{-1}(n)})$. Using Theorem 3.6.2 we see that σ defines an automorphism of A that fixes each element of k. By Exercise 3.5.2, the permutation σ induces an automorphism of K and S_n can be viewed as a group of automorphisms of K. Then K is a Galois extension of K^{S_n} with group S_n. The degree of the extension K/K^{S_n} is equal to the order of the group S_n, which is $n!$, by Example 2.1.15. The fixed field K^{S_n} is called the *field of symmetric rational functions in n variables over k*. The subring of A fixed by S_n is denoted A^{S_n}. We call A^{S_n} the *ring of symmetric polynomials in n variables over k*. Let λ be another indeterminate. Consider the polynomial

$$\Lambda = (\lambda - x_1)(\lambda - x_2) \cdots (\lambda - x_n)$$

in $A[\lambda]$. Notice that Λ is symmetric in $x_1, \ldots, x_n$. In other words, if we extend the action by S_n on A to an action on the ring $A[\lambda]$, then Λ is fixed by S_n. Therefore, the coefficients of Λ are symmetric polynomials and belong to the ring A^{S_n}. The *elementary symmetric polynomial of degree i in the variables* $x_1, \ldots, x_n$, denoted $\sigma_{i,n}$, is the coefficient of λ^{n-i} in the expansion of Λ:

$$\Lambda = \lambda^n - \sigma_{1,n}\lambda^{n-1} + \sigma_{2,n}\lambda^{n-2} - \cdots + (-1)^i \sigma_{i,n}\lambda^{n-i} + \cdots + (-1)^n \sigma_{n,n}.$$

We see that

$$\begin{aligned}
\sigma_{1,n} &= x_1 + x_2 + \cdots + x_n \\
\sigma_{2,n} &= \sum_{i_1 < i_2} x_{i_1} x_{i_2} \\
\sigma_{3,n} &= \sum_{i_1 < i_2 < i_3} x_{i_1} x_{i_2} x_{i_3} \\
&\vdots \\
\sigma_{n,n} &= x_1 x_2 \ldots x_n
\end{aligned}$$

By Exercise 5.7.23, if $1 < i < m \leq n$, then the polynomials $\sigma_{i,m}$ satisfy the recurrence relation: $\sigma_{i,m} = \sigma_{i,m-1} + x_m \sigma_{i-1,m-1}$. Therefore, we have the tower of fields: $k(\sigma_{1,n}, \ldots, \sigma_{n,n}) \subseteq k(x_1, \ldots, x_n)^{S_n} \subseteq k(x_1, \ldots, x_n)$.

THEOREM 5.7.8. *Let k be a field and $k(x_1, \ldots, x_n)$ the field of rational functions in the variables $x_1, \ldots, x_n$ over k. Let S_n be the symmetric group on $\{1, \ldots, n\}$ and $k(x_1, \ldots, x_n)^{S_n}$ the field of symmetric rational functions in the variables $x_1, \ldots, x_n$ over k. Then the following are true.*

(1) $k(x_1, \ldots, x_n)$ is a Galois extension of $k(x_1, \ldots, x_n)^{S_n}$ with Galois group S_n.
(2) The degree of the extension $k(x_1, \ldots, x_n)/k(x_1, \ldots, x_n)^{S_n}$ is $n!$.
(3) If $\sigma_{1,n}, \ldots, \sigma_{n,n}$ are the elementary symmetric polynomials in $x_1, \ldots, x_n$, then $k(x_1, \ldots, x_n)^{S_n} = k(\sigma_{1,n}, \ldots, \sigma_{n,n})$.
(4) $k(x_1, \ldots, x_n)$ is the splitting field of the polynomial

$$\Lambda = \lambda^n - \sigma_{1,n}\lambda^{n-1} + \sigma_{2,n}\lambda^{n-2} - \cdots + (-1)^i\sigma_{i,n}\lambda^{n-i} + \cdots + (-1)^n\sigma_{n,n}$$

over the field $k(x_1, \ldots, x_n)^{S_n} = k(\sigma_{1,n}, \ldots, \sigma_{n,n})$.

PROOF. Parts (1) and (2) were proved in the paragraph preceding this theorem. By definition, $\Lambda = (\lambda - x_1)(\lambda - x_2)\cdots(\lambda - x_n)$ splits over $k(x_1, \ldots, x_n)$ and $k(x_1, \ldots, x_n)$ is generated by the roots of Λ. This proves $k(x_1, \ldots, x_n)$ is the splitting field for Λ over $k(\sigma_{1,n}, \ldots, \sigma_{n,n})$, which is (4). By Corollary 5.2.6 and and Corollary 5.2.8, the dimension of $k(x_1, \ldots, x_n)$ as a vector space over $k(\sigma_{1,n}, \ldots, \sigma_{n,n})$ is at most $n!$. Part (2) and Exercise 5.1.23 imply $k(x_1, \ldots, x_n)^{S_n} = k(\sigma_{1,n}, \ldots, \sigma_{n,n})$, which proves (3). □

COROLLARY 5.7.9. *Let k be a field and $k[x_1, \ldots, x_n]$ the ring of polynomials in the variables $x_1, \ldots, x_n$ over k. If $\sigma_{1,n}, \ldots, \sigma_{n,n}$ are the elementary symmetric polynomials in $x_1, \ldots, x_n$ and $t_1, \ldots, t_n$ are indeterminates, then the k-algebra homomorphism $k[t_1, \ldots, t_n] \to k[\sigma_{1,n}, \ldots, \sigma_{n,n}]$ defined by $t_i \mapsto \sigma_{i,n}$ is an isomorphism.*

PROOF. By Exercise 5.7.21, $K = k(x_1, \ldots, x_n)$ has transcendence degree n over k. By Theorem 5.7.8, K is algebraic over $k(s_{1,n}, \ldots, s_{n,n})$. By Lemma 5.7.4 and Theorem 5.7.5, $\{s_{1,n}, \ldots, s_{n,n}\}$ is a transcendence base for K over k. Therefore, the k-algebra homomorphism $k[t_1, \ldots, t_n] \to k[s_{1,n}, \ldots, s_{n,n}]$ defined by $t_i \mapsto \sigma_{i,n}$ is a k-algebra isomorphism. □

COROLLARY 5.7.10. *If G is a finite group, then there exists a Galois field extension with Galois group isomorphic to G.*

PROOF. Let $[G : 1] = n$. By Cayley's Theorem, Theorem 2.4.5, we can identify G with a subgroup of S_n. By Theorem 5.7.8 and Theorem 5.3.21, $k(x_1, \ldots, x_n)$ is a Galois extension of $k(x_1, \ldots, x_n)^G$ with Galois group G. □

7.3. The General Polynomial is Not Solvable by Radicals. Let k be a field of characteristic zero and assume $x^d - 1$ splits over k, for each $d > 1$. Let $t_0, t_1, \ldots, t_{n-1}$ be indeterminates, and $K = k(t_0, t_1, \ldots, t_{n-1})$ the field of rational functions over k. The general polynomial of degree n over the field k is

$$p(x) = x^n - t_{n-1}x^{n-1} + \ldots (-1)^{n-1}t_1x + (-1)^n t_0$$

which is an element of the ring $K[x]$.

COROLLARY 5.7.11. *(Abel) If $n \geq 5$, the general polynomial of degree n is not solvable by radicals.*

PROOF. Let $\sigma_1, \ldots, \sigma_n$ be the elementary symmetric polynomials in the n variables $x_1, \ldots, x_n$. By Theorem 5.7.8, $K = k(x_1, \ldots, x_n)$ is the splitting field of the polynomial

$$\begin{aligned}\Lambda &= (\lambda - x_1)(\lambda - x_2)\cdots(\lambda - x_n)\\ &= \lambda^n - \sigma_{1,n}\lambda^{n-1} + \cdots + (-1)^{n-1}\sigma_{n-1,n}\lambda + (-1)^n\sigma_{n,n}.\end{aligned}$$

over $k(\sigma_{1,n}, \ldots, \sigma_{n,n})$. By Corollary 5.7.9, the field $k(\sigma_{1,n}, \ldots, \sigma_{n,n})$ is isomorphic to the field of rational functions $k(t_0, t_1, \ldots, t_{n-1})$ in n variables over k. Therefore, Λ is a general polynomial of degree n over k. The Galois group of K over $k(\sigma_{1,n}, \ldots, \sigma_{n,n})$ is S_n, the symmetric group on n letters. By Corollary 2.10.14, S_n is not solvable. By Theorem 5.6.6, Λ is not solvable by radicals, □

7.4. The Discriminant. Throughout this section the characteristic of the ground field k is assumed to be different from 2. In the context of Theorem 5.7.8, the field of rational functions $k(x_1, \ldots, x_n)$ is a Galois extension of the field of symmetric rational functions $k(x_1, \ldots, x_n)^{S_n}$, which is equal to $k(\sigma_{1,n}, \ldots, \sigma_{n,n})$. The Galois group is S_n. The alternating group A_n is the only subgroup of S_n of index two. By Theorem 5.3.21 there is a unique quadratic extension (an extension of degree 2) of $k(\sigma_{1,n}, \ldots, \sigma_{n,n})$ in $k(x_1, \ldots, x_n)$. In Theorem 5.7.12 below we define the discriminant polynomial Δ^2, show that it is a symmetric rational polynomial in $k[x_1, \ldots, x_n]^{S_n}$, and show that under the Galois correspondence the intermediate field corresponding to A_n is the quadratic extension of $k(x_1, \ldots, x_n)^{S_n}$ in $k(x_1, \ldots, x_n)$ obtained by adjoining Δ.

THEOREM 5.7.12. *In the above context, let k be a field of characteristic different from 2, $n \geq 2$, and $\Delta = \prod_{i=1}^{n-1}\prod_{j=i+1}^{n}(x_i - x_j)$.*

(1) For any $\sigma \in S_n$, $\sigma(\Delta) = \operatorname{sign}(\sigma)\Delta$.
(2) $\Delta \notin k(x_1, \ldots, x_n)^{S_n}$ and $\Delta^2 \in k(x_1, \ldots, x_n)^{S_n}$.
(3) There is exactly one quadratic extension of $k(x_1, \ldots, x_n)^{S_n}$ in $k(x_1, \ldots, x_n)$. It is the field obtained by adjoining Δ to $k(x_1, \ldots, x_n)^{S_n}$.

PROOF. Consider the polynomial $\Delta = \prod_{i=1}^{n-1}\prod_{j=i+1}^{n}(x_i - x_j)$ in $k[x_1, \ldots, x_n]$. For any $\sigma \in S_n$, we have

$$\begin{aligned}\sigma(\Delta) &= \prod_{i=1}^{n-1}\prod_{j=i+1}^{n}\sigma(x_i - x_j)\\ &= \prod_{i=1}^{n-1}\prod_{j=i+1}^{n}(x_{\sigma^{-1}(i)} - x_{\sigma^{-1}(j)})\\ &= \prod_{i=1}^{n-1}\prod_{j=i+1}^{n}\pm(x_i - x_j)\\ &= \pm\Delta.\end{aligned}$$

This proves that under the group action by S_n, the orbit of Δ has length at most two. To see that the orbit has length at least two, consider the transposition

$\tau = (12)$ acting on Δ. Only the factors $x_i - x_j$ containing x_1 or x_2 are transformed.

$$\begin{aligned}
\tau(\Delta) &= \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \tau(x_i - x_j) \\
&= \left(\prod_{i=1}^{2} \prod_{j=i+1}^{n} \tau(x_i - x_j) \right) \left(\prod_{i=3}^{n-1} \prod_{j=i+1}^{n} (x_i - x_j) \right) \\
&= (x_2 - x_1)(x_2 - x_3) \cdots (x_2 - x_n)(x_1 - x_3) \cdots (x_1 - x_n) \prod_{i=3}^{n-1} \prod_{j=i+1}^{n} (x_i - x_j) \\
&= -\Delta.
\end{aligned}$$

This proves that the orbit of Δ is equal to $\{\Delta, -\Delta\}$. By Theorem 2.4.11, the stabilizer of Δ is a subgroup of S_n of index 2. By Corollary 2.6.15, the only subgroup of S_n of index two is the alternating group A_n. Thus, the stabilizer of Δ is A_n. If $\sigma \in S_n$, it follows that $\sigma(\Delta) = \text{sign}(\sigma)\Delta$. In particular, $\sigma\left(\Delta^2\right) = \Delta^2$, which implies $\Delta^2 \in k(x_1, \ldots, x_n)^{S_n}$. We have proved parts (1) and (2). Since there is only one subgroup of S_n of index 2, by Theorem 5.3.21, there is only one quadratic extension of $k(x_1, \ldots, x_n)^{S_n}$ in $k(x_1, \ldots, x_n)$. By (2), the irreducible polynomial of Δ is $x^2 - \Delta^2$, which has degree 2. By Theorem 5.1.4, the field extension obtained by adjoining Δ is a quadratic extension. □

DEFINITION 5.7.13. Let k be a field with characteristic different from 2. Let f be a separable polynomial in $k[x]$ and F is a splitting field for f over k. Assume the distinct roots of f in F are $\alpha_1, \ldots, \alpha_n$ and that $n \geq 2$. Let

$$\Delta = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} (\alpha_i - \alpha_j),$$

which is an element of F. Then the *discriminant of f* is defined to be Δ^2, which by Theorem 5.7.14 below is an element of k.

THEOREM 5.7.14. *Let k be a field with characteristic different from 2, f a separable polynomial in $k[x]$, F a splitting field for f over k, and assume the distinct roots of f in F are $\alpha_1, \ldots, \alpha_n$, where $n \geq 2$. If $\Delta = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} (\alpha_i - \alpha_j)$, then the following are true.*

(1) F/k is a Galois extension and the group $G = \text{Aut}_k(F)$ is isomorphic to a subgroup of S_n.
(2) If $\sigma \in \text{Aut}_k(F)$, then $\sigma(\Delta) = \text{sign}(\sigma)\Delta$.
(3) The discriminant of f, Δ^2, is an element of k.
(4) Under the Galois correspondence of Theorem 5.3.21, the intermediate field $k(\Delta)$ corresponds to the subgroup $G \cap A_n$.

PROOF. (1): By Theorem 5.3.18, F/k is a Galois extension. The Galois group $\text{Aut}_k(F)$ acts on the set $\alpha_1, \ldots, \alpha_n$ and can be identified with a subgroup of S_n, by Corollary 5.3.19.

(2): This follows from Theorem 5.7.12 (1).

(3): For any $\sigma \in \text{Aut}_k(F)$, $\sigma(\Delta^2) = \Delta^2$. Thus Δ^2 is in the fixed field F^G which is equal to k by Part (1).

(4): By Part (2), σ is in the subgroup of G fixing Δ if and only if $\text{sign}(\sigma) = 1$. The set of all even permutations in G is equal to $G \cap A_n$. □

When $n = 2$, Theorem 5.7.14 simplifies to the familiar "quadratic formula". This is summarized in Corollary 5.7.15 below. When $n = 3$, Theorem 5.7.14 shows that the Galois group of an irreducible cubic is either the cyclic group of order three, A_3, or the nonabelian group of order six, S_3. This result is summarized in Corollary 5.7.16 below.

COROLLARY 5.7.15. *Let k be a field with characteristic different from 2 and f a monic separable polynomial in $k[x]$ of degree 2. Let F be a splitting field for f over k. Then the following are equivalent.*

(1) f is irreducible in $k[x]$.
(2) $\dim_k(F) = 2$ and the Galois group $\text{Aut}_k(F)$ is a cyclic group of order two.
(3) If α_1, α_2 are the roots of f in F, then $\Delta = \alpha_1 - \alpha_2$ is not in k.

PROOF. The roots of f in F are given by the quadratic formula. For instance, let $f(x) = x^2 + bx + c$. After completing the square, we see that solving $f(x) = 0$ is equivalent to adjoining the square root of $b^2 - 4c$ to k. Therefore, α_1 and α_2 are equal to $(-b \pm \sqrt{b^2 - 4c})/2$ and $\Delta = \alpha_1 - \alpha_2$ is equal to $\pm\sqrt{b^2 - 4c}$. We see that Δ is in k if and only if f splits in k. The rest is left to the reader. □

COROLLARY 5.7.16. *Let k be a field with characteristic different from 2 and f an irreducible separable polynomial in $k[x]$ of degree 3. Let F be a splitting field for f over k and Δ^2 the discriminant of f. Then the following are true.*

(1) $\Delta \in k$ if and only if $\dim_k(F) = 3$. In this case the Galois group $\text{Aut}_k(F)$ is isomorphic to A_3.
(2) $\Delta \notin k$ if and only if $\dim_k(F) = 6$. In this case the Galois group $\text{Aut}_k(F)$ is isomorphic to S_3.

PROOF. By Corollary 5.3.19, we identify the Galois group $G = \text{Aut}_k(F)$ with a subgroup of S_3. By Lemma 4.4.5, $\dim_k(F) \geq 3$, which implies $\dim_k(F)$ is a multiple of 3. Therefore 3 divides the order of the group G. It follows that G is either A_3 or S_3. By Theorem 5.7.14 (4), $\Delta \in k$ if and only if $G = A_3$. □

7.5. Symmetric Polynomials. Theorem 5.7.8 (3) says that every symmetric rational function is a rational function in the elementary symmetric polynomials. In Theorem 5.7.17, which is due to Gauss, we improve this result by proving that every symmetric polynomial is a polynomial in the elementary symmetric polynomials.

THEOREM 5.7.17. *Let k be a field and $k[x_1, \ldots, x_n]$ the ring of polynomials in the variables $x_1, \ldots, x_n$ over k. Let S_n be the symmetric group on $\{1, \ldots, n\}$ and $k[x_1, \ldots, x_n]^{S_n}$ the ring of symmetric polynomials in the variables $x_1, \ldots, x_n$ over k. If $\sigma_{1,n}, \ldots, \sigma_{n,n}$ are the elementary symmetric polynomials in $x_1, \ldots, x_n$, then the following are true.*

(1) If f is a nonzero symmetric polynomial, then there exists a polynomial $g \in k[t_1, \ldots, t_n]$ such that $f = g(\sigma_{1,n}, \ldots, \sigma_{n,n})$.
(2) $k[x_1, \ldots, x_n]^{S_n} = k[\sigma_{1,n}, \ldots, \sigma_{n,n}]$.
(3) The polynomial g in (1) is unique.

The proof of the theorem will utilize the following lemma.

LEMMA 5.7.18. *In the context of Theorem 5.7.17, let f be a nonzero symmetric polynomial in $k[x_1, \ldots, x_n]^{S_n}$. If the leading term of f (see Lemma 3.6.18) is $M = rx_1^{e_1} \cdots x_n^{e_n}$, then $e_1 \geq e_2 \geq \cdots \geq e_n$.*

PROOF. For sake of contradiction assume $1 \leq i < j \leq n$ and $e_i < e_j$. Apply the transposition $\tau = (i, j)$ to f. Since $\tau f = f$, we know that f has the monomial

$$\tau M = rx_1^{e_1} \cdots x_{i-1}^{e_{i-1}} x_j^{e_i} x_{i+1}^{e_{i+1}} \cdots\cdots x_{j-1}^{e_{j-1}} x_i^{e_j} x_{j+1}^{e_{j+1}} \cdots x_n^{e_n} = rx_1^{e_1} \cdots x_i^{e_j} \cdots x_j^{e_i} \cdots x_n^{e_n}.$$

Thus in the monomial τM, the exponents of x_i and x_j are swapped. But

$$M = rx_1^{e_1} \cdots x_i^{e_i} \cdots x_j^{e_j} \cdots x_n^{e_n} < rx_1^{e_1} \cdots x_i^{e_j} \cdots x_j^{e_i} \cdots x_n^{e_n} = \tau M.$$

This is a contradiction, since M is the leading term of f. □

PROOF OF THEOREM 5.7.17. (1) and (2): Let f be a nonzero symmetric polynomial in $k[x_1, \ldots, x_n]^{S_n}$ and assume the leading term of f is $r_1 x_1^{e_1} \cdots x_n^{e_n}$. By Lemma 5.7.18, $e_1 \geq e_2 \geq \cdots \geq e_n$. Set $d_1 = e_1 - e_2$, $d_2 = e_2 - e_3$, $\ldots$, $d_{n-1} = e_{n-1} - e_n$, and $d_n = e_n$. By Exercise 5.7.25, the leading term of $s_{1,n}^{d_1} s_{2,n}^{d_2} \cdots s_{n,n}^{d_n}$ is equal to

$$x_1^{d_1+d_2+\cdots+d_n} x_2^{d_2+\cdots+d_n} \cdots x_n^{e_n} = x_1^{e_1} x_2^{e_2} \cdots x_n^{e_n}.$$

Let $g_1 = r_1 s_{1,n}^{d_1} s_{2,n}^{d_2} \cdots s_{n,n}^{d_n}$. Then $g_1 \in k[s_{1,n}, \ldots, s_{n,n}]$ and $f_1 = f - g_1$ is a symmetric polynomial in $k[x_1, \ldots, x_n]^{S_n}$. The leading terms of f and g_1 are equal, so if f_1 is nonzero, the leading term of f_1 is less than the leading term of f in the lexicographical order (see Section 3.6.1). If f_1 is nonzero, then we repeat the above steps to get $g_2 \in k[s_{1,n}, \ldots, s_{n,n}]$ with the same leading term as f_1. Hence $f_2 = f_1 - g_2$ is either zero, or has a leading term less than the leading term of f_1. Iterating, we get a sequence of symmetric polynomials $f, f_1, f_2, \ldots$ such that the leading terms form a strictly decreasing sequence. By Lemma 3.6.18 (3), after a finite number of iterations we have $f_m = 0$. This shows that $f = g_1 + g_2 + \cdots + g_m$ is in $k[s_{1,n}, \ldots, s_{n,n}]$, proving (1) and (2).

(3): This follows from Corollary 5.7.9, because the map induced by sending t_i to $\sigma_{i,n}$ is a k-algebra isomorphism $k[t_1, \ldots, t_n] \cong k[s_{1,n}, \ldots, s_{n,n}]$. □

7.6. Exercises.

EXERCISE 5.7.19. If F/k is an extension of fields, show that $\emptyset$ is a transcendence base if and only if F/k is an algebraic extension.

EXERCISE 5.7.20. If F/k is an extension of fields, and $\Xi \subseteq F$ is algebraically independent over k, show that there exists a transcendence base Z such that $Z \supseteq \Xi$.

EXERCISE 5.7.21. Let k is a field, and $x_1, \ldots, x_n$ a set of indeterminates. Show that $\operatorname{tr.deg}_k k(x_1, \ldots, x_n) = n$ and $\{x_1, \ldots, x_n\}$ is a transcendence base for $k(x_1, \ldots, x_n)$ over k.

EXERCISE 5.7.22. If F is a finitely generated extension field of the field k, show that $\operatorname{tr.deg}_k(F)$ is equal to the least integer n such that there exist $\xi_1, \ldots, \xi_n$ in F and F is algebraic over $k(\xi_1, \ldots, \xi_n)$.

EXERCISE 5.7.23. Let $x_1, \ldots, x_n$ be a set of indeterminates. If $1 \leq i \leq m \leq n$, let $\sigma_{i,m}$ be the elementary symmetric polynomial of degree i in the variables $x_1, \ldots, x_m$. Prove the following recursive formula:

$$\sigma_{i,m} = \begin{cases} x_1 + x_2 + \cdots + x_m & \text{if } i = 1, \\ x_1 x_2 \ldots x_m & \text{if } i = m, \\ \sigma_{i,m-1} + x_m \sigma_{i-1,m-1} & \text{if } 1 < i < m \leq n. \end{cases}$$

EXERCISE 5.7.24. Let S_n be the symmetric group on $\{1, 2, \ldots, n\}$ and S_{n-1} the symmetric group on $\{1, 2, \ldots, n-1\}$. We view S_{n-1} as a subgroup of S_n. Let k be a field. Prove that if $f(x_1, \ldots, x_n) \in k[x_1, \ldots, x_n]^{S_n}$, then $f(x_1, \ldots, x_{n-1}, 0) \in k[x_1, \ldots, x_{n-1}]^{S_{n-1}}$. Show that there exists a commutative diagram

$$\begin{array}{ccc} A_n = k[x_1, \ldots, x_n] & \xrightarrow{\alpha} & A_{n-1} = k[x_1, \ldots, x_{n-1}] \\ {\scriptstyle a}\uparrow{\scriptstyle \subseteq} & & {\scriptstyle b}\uparrow{\scriptstyle \subseteq} \\ A_n^{S_n} & \xrightarrow{\beta} & A_{n-1}^{S_{n-1}} \\ {\scriptstyle c}\uparrow{\scriptstyle \subseteq} & & {\scriptstyle d}\uparrow{\scriptstyle =} \\ k[\sigma_{1,n}, \ldots, \sigma_{n,n}] & \xrightarrow{\gamma} & k[\sigma_{1,n-1}, \ldots, \sigma_{n-1,n-1}] \end{array}$$

of commutative rings satisfying the following:

(1) The maps a, b, c, d are homomorphisms defined by set inclusion.
(2) The epimorphism α is defined by $x_n \mapsto 0$.
(3) The homomorphism β is the restriction of α to $A_n^{S_n}$.
(4) The epimorphism γ is the restriction of α to $k[\sigma_{1,n}, \ldots, \sigma_{n,n}]$.

EXERCISE 5.7.25. Let $e_i \geq 0$ for each i. In the context of Theorem 5.7.17, show that the leading term of $s_{1,m}^{e_1} s_{2,m}^{e_2} \cdots s_{m,m}^{e_m}$ is equal to $x_1^{e_1+e_2+\cdots+e_m} x_2^{e_2+\cdots+e_m} \cdots x_m^{e_m}$.

EXERCISE 5.7.26. Follow the steps below to show that the map γ in Exercise 5.7.24 has a section.

(1) Show that there is a k-algebra homomorphism

$$\epsilon : k[\sigma_{1,n-1}, \ldots, \sigma_{n-1,n-1}] \to k[\sigma_{1,n}, \ldots, \sigma_{n,n}]$$

defined by $\sigma_{i,n-1} \mapsto \sigma_{i,n}$.
(2) Show that $\gamma\epsilon$ is the identity map on $k[\sigma_{1,n-1}, \ldots, \sigma_{n-1,n-1}]$.

EXERCISE 5.7.27. Let F/k be an extension of fields. Apply Zorn's Lemma, Proposition 1.3.3, to prove: If Ξ is a subset of F such that F is algebraic over $k(\Xi)$, then Ξ contains a subset which is a transcendence base for F over k.

EXERCISE 5.7.28. Determine whether each of the following polynomials f is a symmetric polynomial or not. If yes, then write f in terms of the elementary symmetric polynomials (Theorem 5.7.17).

(1) $f = x_1^2 - x_1 x_2 + x_2^2 - 1$
(2) $f = (x_1 + x_2)(x_1 + x_3)(x_2 + x_3)$
(3) $f = x_1^2 - x_1 x_2 - x_1 x_3 - x_2 x_3 + x_2^2 + x_3^2$

CHAPTER 6

Linear Transformations

This subject of this chapter are R-module homomorphisms $\phi : M \to N$ where M and N are free R-modules of finite rank. We showed in Section 4.5 that if R is a ring and M and N are free R-modules of finite rank, then a homomorphism ϕ from M to N can be represented as a matrix. The matrix of ϕ depends only on the choice of bases for M and N. If $M = N$, and $\phi : M \to M$, then the matrix of ϕ is unique up to a similarity transformation. That is, two matrix representations are similar and the similarity transformation corresponds to a change of basis for M. If k is a field and V is a finite dimensional k-vector space, then associated to a linear transformation ϕ from V to V are important invariants. The first that we define are the so-called invariant factors and elementary divisors. These invariants are defined using the corrsponding basis theorems of Section 4.6.3 for finitely generated modules over the principal ideal domain $k[x]$. Given the invariant factors of ϕ, we show that there is a matrix representation which is in so-called canonical form. For any square matrix A over k, there is a unique member of the similarity class of A which is in canonical form. In Definition 6.3.11 we define the determinant function from $\mathrm{Hom}_R(M, M)$ to R, for any commutative ring R and free R-module M of finite rank. The characteristic polynomial of ϕ is defined using the determinant function. The Normal Basis Theorem for a cyclic Galois extension is proved in Section 6.4. The proof uses results from Chapters 5 and 6.

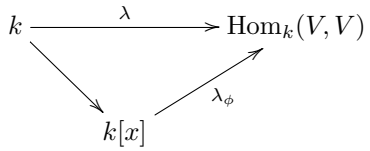
1. A Linear Transformation on a Vector Space

In this section we study properties of a linear transformation ϕ on a finite dimensional vector space V over a field k. In Section 6.1.1 we show that ϕ can be used to turn V into a module over the principal ideal domain $k[x]$. This is called the $k[\phi]$-module structure on V and is denoted V_ϕ. Since V is finitely generated over k, V_ϕ is a finitely generated $k[x]$-module. Therefore, the important theorems on finitely generated modules over a principal ideal domain of Section 4.6 apply to V_ϕ. First, we show that the $k[\phi]$-module V_ϕ decomposes into an internal direct sum of cyclic $k[\phi]$-modules (Proposition 6.1.2). The results of this section and the basis theorems of Section 4.6.3 are applied to the module V_ϕ in Section 6.2. The invariant factors and the elementary divisors associated to V_ϕ allow us to define important invariants of the linear transformation ϕ. The basis theorems permit us to define two matrix canonical forms for ϕ. These are the rational canonical form of ϕ, which always exists, and if the minimal polynomial of ϕ splits, then we can define the Jordan canonical form. In Section 6.1.2 we define the eigenvalues and eigenvectors of ϕ. These correspond to $k[\phi]$-submodules of V_ϕ that have dimension one over k. The eigenvalues of ϕ correspond to the roots of the minimal polynomial of ϕ. In Theorem 6.1.12 we show that the minimal polynomial of ϕ splits in k and has no multiple roots if and only if there is a basis for V such that the matrix of ϕ

is a diagonal matrix. In this case the diagonal entries are the eigenvalues of ϕ. In Section 6.1.3 we show that the minimal polynomial splits in k if and only if there is a basis for V such that the matrix of ϕ is lower triangular and the diagonal entries are the eigenvalues of ϕ. As an application we show that the degree of the minimal polynomial of ϕ is less than or equal to the dimension of V.

1.1. A Vector Space as a $k[\phi]$-Module. Let k be a field and V a finite dimensional k-vector space. We begin by reviewing properties of the ring $\mathrm{Hom}_k(V,V)$ of k-linear transformations of V. By Example 4.4.2, $\mathrm{Hom}_k(V,V)$ is a k-algebra. By Proposition 4.5.4, $\mathrm{Hom}_k(V,V)$ is finite dimensional as a k-vector space. By Corollary 4.4.11, $\mathrm{Hom}_k(V,V)$ is algebraic over k. By Theorem 4.4.8, every ϕ in $\mathrm{Hom}_k(V,V)$ has a minimal polynomial, $f = \mathrm{min.\,poly}_k(\phi)$. By Proposition 4.5.7, if $\dim_k(V) = n$, then the ring of matices $M_n(k)$ and the ring $\mathrm{Hom}_k(V,V)$ are isomorphic as k-algebras. If X is a basis for V, and $A = M(\phi, X, X)$, then $\mathrm{min.\,poly}_k(\phi)$ is equal to $\mathrm{min.\,poly}_k(A)$ (Exercise 4.4.20).

By Exercise 4.1.20, V is a left $\mathrm{Hom}_k(V,V)$-module by the action $\psi v = \psi(v)$, for any $\psi \in \mathrm{Hom}_k(V,V)$ and $v \in V$. Let $\phi \in \mathrm{Hom}_k(V,V)$. Using this ϕ, we make V into a left $k[x]$-module. By Theorem 3.6.2, the evaluation homomorphism $\lambda_\phi : k[x] \to \mathrm{Hom}_k(V,V)$ which maps x to ϕ is a homomorphism of rings.



If $p(x) = a_0 + a_1 x + \cdots + a_n x^n \in k[x]$, then $\lambda_\phi(p(x)) = a_0 + a_1\phi + \cdots + a_n\phi^n$. The image of $k[x]$ under λ_ϕ is the commutative subring of $\mathrm{Hom}_k(V,V)$ denoted $k[\phi]$. The kernel of λ_ϕ is the principal ideal generated by $f = \mathrm{min.\,poly}_k(\phi)$ and there is a k-algebra isomorphism $k[x]/(f) \cong k[\phi]$ (Theorem 4.4.8). Since $k[x]$ is a principal ideal domain, by Corollary 3.2.18 every ideal in $k[\phi]$ is a principal ideal. The ideals in $k[\phi]$ correspond up to associates to the divisors of f in $k[x]$ (see Exercises 4.6.18 and 4.6.19).

By Example 4.1.4 (4), λ_ϕ turns V into a $k[x]$-module. For any $v \in V$ and $p(x) \in k[x]$, the left multiplication of v by $p(x)$ is given by the formula:

$$\begin{aligned} p(x)v &= \lambda_\phi(p(x))v \\ &= (a_0 + a_1\phi + \cdots + a_n\phi^n)v \\ &= a_0 v + a_1\phi(v) + \cdots + a_n\phi^n(v). \end{aligned}$$

Definition 6.1.1. We denote by V_ϕ the left $k[x]$-module structure on V induced by λ_ϕ. A $k[x]$-submodule of V_ϕ is also called a ϕ-invariant subspace of V.

Given a linear transformation $\phi : V \to V$ on a finite dimensional vector space V over a field k, we have the associated finitely generated module V_ϕ over the principal ideal domain $k[x]$. We are in the context of Section 4.6. In Proposition 6.1.2 we exploit the fact that $k[x]$ is a principal ideal domain to show that the $k[x]$-module V_ϕ decomposes into an internal direct sum of cyclic submodules.

Proposition 6.1.2. *Let k be a field, V a finite dimensional k-vector space, and $\phi \in \mathrm{Hom}_k(V,V)$. The $k[x]$-module V_ϕ is the internal direct sum of cyclic*

submodules. That is, there exist $v_1, \ldots, v_q$ in V such that $V_\phi = (v_1) \oplus \cdots \oplus (v_q)$, where $(v_i) = k[\phi]v_i$.

PROOF. Since V is finitely generated as a k-vector space, V_ϕ is finitely generated as a $k[x]$-module. By Corollary 3.6.5, $k[x]$ is a principal ideal domain. By Corollary 4.6.15, V_ϕ is the internal direct sum of cyclic submodules. That is, there exist $v_1, \ldots, v_q$ in V such that $V_\phi = (v_1) \oplus \cdots \oplus (v_q)$, where $(v_i) = k[\phi]v_i$. □

If the minimal polynomial of ϕ has more than one distinct irreducible factor, then Proposition 6.1.3 shows that the ring $k[\phi]$ decomposes into a direct sum of idempotent generated ideals and there is a corresponding decomposition of V_ϕ into $k[x]$-submodules.

PROPOSITION 6.1.3. *Let k be a field, V a finite dimensional k-vector space, $\phi \in \mathrm{Hom}_k(V, V)$, and $f = \min.\,\mathrm{poly}_k(\phi)$. In the polynomial ring $k[x]$, let $f = p_1^{e_1} \cdots p_r^{e_r}$ be the unique factorization of f where $p_1, \ldots, p_r$ are distinct monic irreducible polynomials, $r \geq 1$, and $e_i \geq 1$ for each i. Then there exist submodules $V_1, \ldots, V_r$ of the $k[x]$-module V_ϕ such that the following are true.*

(1) The minimal polynomial of the restriction of ϕ to a linear transformation $\phi : V_i \to V_i$ is $p_i^{e_i}$.
(2) $V_\phi = V_1 \oplus \cdots \oplus V_r$.

PROOF. By the Chinese Remainder Theorem (Corollary 3.3.13),

$$k[\phi] = k[x]/(f) \cong k[x]/(p_1^{e_1}) \oplus \cdots \oplus k[x]/(p_1^{e_1}).$$

By Exercise 4.2.27 there is a corresponding internal direct sum decomposition $V_\phi = V_1 \oplus \cdots \oplus V_r$ into $k[x]$-submodules. Moreover, for each i, the left regular representation $\lambda_\phi : k[x] \to \mathrm{Hom}_k(V_i, V_i)$ factors through $k[x]/(p_i^{e_i})$. The minimal polynomial of ϕ restricted to V_i is $p_i^{e_i}$. □

PROPOSITION 6.1.4. *Let k be a field, V a k-vector space of dimension n, and ϕ a nonzero linear transformation in $\mathrm{Hom}_k(V, V)$. Let V_ϕ be the $k[x]$-module structure on V induced by the ring homomorphism $k[x] \to \mathrm{Hom}_k(V, V)$ which maps x to ϕ. If V_ϕ is a cyclic $k[x]$-module with generator u, then the following are true.*

(1) The set $B = \{u, \phi u, \phi^2 u, \ldots, \phi^{n-1}u\}$ is a k-basis for V.
(2) As $k[x]$-modules, $V_\phi \cong k[x]/(f)$.
(3) If $\min.\,\mathrm{poly}_k(\phi) = f$, then $\deg f = n$ and f is the monic polynomial of minimal degree such that $f(\phi)u = 0$.

PROOF. If $f = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ is the minimal polynomial of ϕ, then a k-basis for $k[\phi]$ is $\{\phi^{n-1}, \ldots, \phi, 1\}$ (Theorem 4.4.8). If $u \in V$, the cyclic $k[x]$-submodule of V_ϕ generated by u is therefore equal to

$$k[\phi]u = \{p(\phi)u \mid p \in k[x]\} = k\phi^{n-1}u + \cdots + k\phi u + ku.$$

Since ϕ maps this subspace to itself, we say $k[\phi]u$ is *ϕ-invariant*. If u is nonzero, the $k[x]$-module homomorphism $k[x] \to k[\phi]u$ is onto. The kernel is a principal ideal $I_u = (q)$, and we have

$$k[\phi]u \cong k[x]/(q).$$

The polynomial q is called the *order of u*. Since u is nonzero and $k[\phi]u$ is finite dimensional over k, by Lemma 4.4.5 we know q is a monic polynomial of positive degree. In fact, q is the polynomial of minimal degree such that $q(\phi)u = 0$. By Exercise 6.1.27, q is a divisor of the minimal polynomial f of ϕ. Because the

dimension of the k-vector space $k[\phi]u$ is equal to the degree of q, we see that q is the minimal polynomial of the restriction of ϕ to the ϕ-invariant subspace $k[\phi]u$. $\square$

The ring of matrices $M_n(k)$ is a k-algebra where we identify k with the set of scalar matrices. The center of the ring of matrices is k. By Proposition 4.5.4, $\dim_k(M_n(k)) = n^2$. Since $M_n(k)$ is finite dimensional over k, every matrix $A \in M_n(k)$ has a minimal polynomial $\text{min. poly}_k(A)$ (see Theorem 4.4.8). We will show in Corollary 6.1.19 that the degree of the minimal polynomial is at most n. The evaluation homomorphism $\theta : k[x] \to M_n(R)$ which is defined by $x \mapsto A$ maps $k[x]$ onto the commutative subring $k[A]$ of $M_n(R)$. The kernel of θ is the principal ideal generated by $f = \text{min. poly}_k(A)$.

$$\begin{array}{ccccc} k[x] & \xrightarrow{\theta} & k[A] & \xrightarrow{\subseteq} & M_n(R) \\ & \searrow^{\eta} & \uparrow \cong & & \\ & & k[x]/(f) & & \end{array}$$

In Examples 6.1.5 and 6.1.6 we compute the minimal polynomials of some matrices.

EXAMPLE 6.1.5. Let k be a field, $n \geq 2$, and $A = M_n(k)$ the ring of n-by-n matrices over k. Let e_{st} be the elementary matrix with 1 in position (s,t) and 0 elsewhere (see Section 1.5). Notice that

$$e_{st}e_{uv} = \begin{cases} e_{sv} & \text{if } t = u, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, $e_{st}e_{st} = 0$ if $s \neq t$ and $e_{ss}e_{ss} = e_{ss}$. From this it follows that

$$\text{min. poly}_k(e_{st}) = \begin{cases} x^2 - x & \text{if } s = t, \\ x^2 & \text{if } s \neq t. \end{cases}$$

In both cases we see that the minimal polynomial of e_{st} is not irreducible.

$$k[e_{st}] \cong \begin{cases} k[x]/x^2 - x & \text{if } s = t, \\ k[x]/x^2 & \text{if } s \neq t. \end{cases}$$

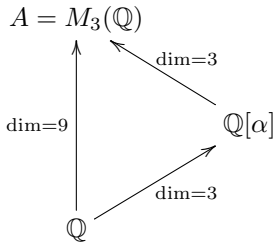
Therefore, $k[e_{st}]$ is not a field.

EXAMPLE 6.1.6. Let k be a field, $a \in k$, $A = M_3(k)$ the ring of 3-by-3 matrices over k, and $\alpha = \begin{bmatrix} 0 & 0 & a \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. Notice that $\alpha^2 = \begin{bmatrix} 0 & a & 0 \\ 0 & 0 & a \\ 1 & 0 & 0 \end{bmatrix}$ and $\alpha^3 = \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{bmatrix} = aI_3$. Therefore, α^3 is in k. Let $p(x) = x^3 - a$. Then $p(\alpha) = 0$. Let $f(x) = \text{min. poly}_k(\alpha)$. Then $f(x)$ divides $p(x)$. To show that $f(x)$ is equal to $p(x)$, it suffices to show $f(x)$ has degree greater than 2. First, since α is not a diagonal matrix we know $f(x)$ has degree greater than 1. For contradiction's sake, suppose $f(x) = x^2 + bx + c$ for some $b, c \in k$. Then $\alpha^2 + b\alpha \in k$. But

$$\alpha^2 + b\alpha = \begin{bmatrix} 0 & a & 0 \\ 0 & 0 & a \\ 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & ab \\ b & 0 & 0 \\ 0 & b & 0 \end{bmatrix} = \begin{bmatrix} 0 & a & ab \\ b & 0 & a \\ 1 & b & 0 \end{bmatrix}$$

is not a diagonal matrix. This contradiction implies $f(x)$ has degree greater than 2, hence $\text{min. poly}_k(\alpha) = x^3 - a$. This example is a special case of Exercise 6.3.17.

The matrix α is called the companion matrix of the polynomial $x^3 - a$. Notice that $k[\alpha] \cong k[x]/(x^3 - a)$ is a field if and only if $x^3 - a$ is irreducible in $k[x]$. For instance, if $k = \mathbb{Q}$, and $a = 8$, then $x^3 - 8 = (x-2)(x^2 + 2x + 4)$ is not irreducible, hence $\mathbb{Q}[\alpha]$ is not a field. On the other hand, if $k = \mathbb{Q}$ and $a = 10$, then α is a root of $x^3 - 10$ in $M_3(\mathbb{Q})$, $\mathbb{Q}[\alpha]$ is an extension field of k inside of A, and there is a lattice of subrings



where an arrow denotes set containment. Using the fact that $\mathbb{Q}[\alpha]$ is a subring of A we can view A as a vector space over $\mathbb{Q}[\alpha]$. We have $9 = (A : \mathbb{Q}) = (\mathbb{Q}[\alpha] : \mathbb{Q})(\mathbb{Q}[\alpha] : \mathbb{Q}) = 3 \cdot 3$. Notice that $\mathbb{Q}[\alpha]$ is not contained in the center of A, hence A is not an algebra over $\mathbb{Q}[\alpha]$.

1.2. Eigenvalues. Let k be a field, V a finite dimensional k-vector space, and $\phi \in \mathrm{Hom}_k(V, V)$. Suppose there exists a nonzero vector $v \in V$ such that the cyclic submodule $k[\phi]v$ has dimension one over k. Then this implies $\phi(v) = \lambda v$ for some scalar λ. We say λ is an eigenvalue of ϕ and v is called an eigenvector corresponding to λ. There is a correspondence between eigenvalues of ϕ and roots of the minimal polynomial. In particular, there is a basis for V consisting of eigenvectors, if and only if the minimal polynomial of ϕ splits in k and has no repeated root.

DEFINITION 6.1.7. Let k be a field, V a finite dimensional k-vector space, and $\phi \in \mathrm{Hom}_k(V, V)$. If ϕ is not invertible, then we say ϕ is *singular*.

Theorem 6.1.8 is stated here for reference. In it we assemble many results that have already been proven.

THEOREM 6.1.8. *Let k be a field, V a finite dimensional k-vector space, and $\phi \in \mathrm{Hom}_k(V, V)$. The following are true.*

(1) $\phi \neq 0$ if and only if there exists $v \in V$ such that $\phi(v) \neq 0$.
(2) $\min.\mathrm{poly}_k(\phi)$ has degree less than or equal to n^2.
(3) $k[\phi]$ is a commutative k-subalgebra of $\mathrm{Hom}_k(V, V)$.
(4) The following are equivalent.
  (a) ϕ is singular.
  (b) The constant term of $\min.\mathrm{poly}_k(\phi)$ is zero.
  (c) There exists $\sigma \in \mathrm{Hom}_k(V, V)$ such that $\sigma \neq 0$ and $\phi\sigma = \sigma\phi = 0$.
  (d) There exists $v \in V - (0)$ such that $\phi(v) = 0$.
(5) The following are equivalent.
  (a) ϕ is invertible.
  (b) $\mathrm{Rank}(\phi) = \dim_k(V)$.
  (c) $\mathrm{Nullity}(\phi) = 0$.

PROOF. For the proof, apply Proposition 4.5.4, Theorems 4.4.8 and 4.4.15, Corollaries 4.4.11 and 4.4.14, and Exercise 4.3.11. □

DEFINITION 6.1.9. Let k be a field, V a finite dimensional k-vector space. If $\phi \in \mathrm{Hom}_k(V,V)$ and $\lambda \in k$, then λ is called an *eigenvalue* or *characteristic root* of ϕ if $\phi - \lambda$ is singular. The set $U(\lambda) = \ker(\phi - \lambda) = \{x \in V \mid \phi(x) = \lambda x\}$ is called the *eigenspace* of λ. By Theorem 6.1.8 (4), $U(\lambda) \neq (0)$. If $v \in U(\lambda)$ and $v \neq 0$, then $\phi(v) = \lambda v$ and we say v is an *eigenvector* corresponding to λ. Since ϕ restricts to an endomorphism $\phi : U(\lambda) \to U(\lambda)$, $U(\lambda)$ is a ϕ-invariant subspace of V.

THEOREM 6.1.10. *Let k be a field, V a finite dimensional vector space over k and $\phi \in \mathrm{Hom}_k(V,V)$. Then the eigenvalues of ϕ are precisely the roots of the minimal polynomial of ϕ.*

PROOF. Let $\lambda \in k$ and $f(x) = \mathrm{min.poly}_k(\phi)$. By Synthetic Division (Corollary 3.6.4), $f(x) = q(x)(x - \lambda) + f(\lambda)$. Then $f(\phi) = 0$ implies

$$f(\lambda) = -q(\phi)(\phi - \lambda) = -(\phi - \lambda)q(\phi).$$

If λ is an eigenvalue of ϕ, then there exists a nonzero $v \in V$ such that $(\phi - \lambda)(v) = 0$. Therefore, $f(\lambda)v = 0$ and Lemma 4.3.1 implies $f(\lambda) = 0$. Conversely, assume $f(\lambda) = 0$. Since $\deg(q) < \deg(f)$, we know $q(\phi) \neq 0$. By Theorem 6.1.8 (1), there exists $u \neq 0$ such that $v = q(\phi)u \neq 0$. Then $0 = (\phi - \lambda)q(\phi)u = (\phi - \lambda)v$. Theorem 6.1.8 (4) implies $\phi - \lambda$ is singular, hence λ is an eigenvalue of ϕ. □

THEOREM 6.1.11. *Let k be a field, V a finite dimensional k-vector space, and $\phi \in \mathrm{Hom}_k(V,V)$. Suppose $\{\lambda_1, \ldots, \lambda_n\}$ is a set of n distinct eigenvalues of ϕ in k. For $i = 1, \ldots, n$, assume v_i is an eigenvector in V corresponding to λ_i. Then $\{v_1, \ldots, v_n\}$ is a linearly independent set in V.*

PROOF. Assume for contradiction's sake that there exists a nonzero vector $(\alpha_1, \ldots, \alpha_n)$ in $k^{(n)}$ such that $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$. Out of all such dependency relations, pick one such that the number of nonzero coefficients is minimal. Without loss of generality, rearrange the lists and assume

$$\alpha_1 v_1 + \cdots + \alpha_m v_m = 0 \tag{1.1}$$

where $1 \leq m \leq n$, $\alpha_1, \ldots, \alpha_m$ are all nonzero, and m is minimal. By Lemma 4.3.1, we know $1 < m$. Apply ϕ to (1.1), multiply (1.1) by λ_m, and subtract. This results in

$$\alpha_1(\lambda_1 - \lambda_m)v_1 + \cdots + \alpha_{m-1}(\lambda_{m-1} - \lambda_m)v_{m-1} = 0. \tag{1.2}$$

Since the λ_i are distinct elements of k, (1.2) is a nontrivial dependence relation of length $m - 1$. This contradicts the minimal choice of m. □

THEOREM 6.1.12. *Let k be a field, V a finite dimensional vector space over k and $\phi \in \mathrm{Hom}_k(V,V)$. Then the following are equivalent.*

(1) There is a basis B for V such that $M(\phi, B)$ is diagonal.
(2) There is a basis of V consisting of eigenvectors of ϕ.
(3) The minimal polynomial $\mathrm{min.poly}_k(\phi)$ factors into a product of linear factors in $k[x]$ and has no multiple roots.

PROOF. (1) is equivalent to (2): This follows straight from Definitions 4.5.3 and 6.1.9.

(2) implies (3): Let $\{v_1, \ldots, v_n\}$ be a basis for V such that each v_j is an eigenvector of ϕ. Let $\{\lambda_1, \ldots, \lambda_m\}$ be the distinct eigenvalues of ϕ, where $1 \leq m \leq n$, by Theorem 6.1.11. Let $g(x) = (x - \lambda_1) \cdots (x - \lambda_m)$. We show that $g(\phi) = 0$.

It suffices to show $g(\phi)v_j = 0$, for each $1 \leq j \leq n$. For some $1 \leq \ell \leq m$, v_j is an eigenvector corresponding to λ_ℓ. Hence $\phi v_j = \lambda_\ell v_j$. Since the ring $k[\phi]$ is commutative, we have

$$\begin{aligned} g(\phi)v_j &= (\phi - \lambda_1)\cdots(\phi - \lambda_m)v_j \\ &= \Big(\prod_{i\neq\ell}(\phi - \lambda_i)\Big)(\phi - \lambda_\ell)v_j \\ &= 0. \end{aligned}$$

By Theorem 4.4.8, $\text{min. poly}_k(\phi)$ divides g. It follows from Theorem 6.1.10 that $g = \text{min. poly}_k(\phi)$.

(3) implies (1): Suppose $f = \text{min. poly}_k(\phi) = (x - \lambda_1)\cdots(x - \lambda_m)$, where f has no repeated roots. By Proposition 6.1.3, V decomposes into $k[\phi]$-submodules $V_1 \oplus \cdots \oplus V_m$ where the action of ϕ on V_i is multiplication by λ_i. In other words, $V_i \subseteq U(\lambda_i)$, for each i. If B_i is a basis for V_i, then B_i consists of eigenvectors of ϕ. Concatenating we get a basis $B_1 + \cdots + B_m$ for V. □

PROPOSITION 6.1.13. *Let k be a field and V a finite dimensional vector space over k. Let ϕ and ψ be linear transformations in $\text{Hom}_k(V, V)$ and assume ψ is invertible. Then the following are true.*

(1) $\text{min. poly}_k(\phi) = \text{min. poly}_k(\psi^{-1}\phi\psi)$.
(2) If $\lambda \in k$, then λ is a eigenvalue of ϕ if and only if λ is an eigenvalue of $\psi^{-1}\phi\psi$.

PROOF. (1): This follows from Exercise 4.4.20.

(2): If λ is an eigenvalue of ϕ with corresponding eigenvector v, then

$$\left(\psi^{-1}\phi\psi\right)\left(\psi^{-1}v\right) = \psi^{-1}\left(\phi v\right) = \psi^{-1}\left(\lambda v\right) = \lambda\left(\psi^{-1}v\right).$$

This shows λ is an eigenvalue of $\psi^{-1}\phi\psi$ with corresponding eigenvector $\psi^{-1}v$. The converse follows by a symmetric argument. □

DEFINITION 6.1.14. Let k be a field, and A a matrix in $M_n(k)$. With respect to the standard basis on $k^{(n)}$, left multiplication by A defines a linear transformation ℓ_A in $\text{Hom}_k(k^{(n)}, k^{(n)})$. The eigenvalues of A are defined to be the corresponding eigenvalues of ℓ_A. By Proposition 6.1.13, similar matrices have the same eigenvalues.

1.3. Triangular Matrices. In this section we prove that a linear transformation ϕ on a finite dimensional k-vector space V has a matrix representation that is lower triangular if and only if the minimal polynomial of ϕ splits in k. As an application, we show that the degree of the minimal polynomial of ϕ is less than or equal to the dimension $\dim_k(V)$. We first prove Lemma 6.1.15 which shows that the minimal polynomial of a matrix is invariant under change of base field.

LEMMA 6.1.15. *Let F/k be an extension of fields. Then we view $k[x]$ as a subring of $F[x]$ and $M_n(k)$ as a subring of $M_n(F)$. If A is a matrix in the ring $M_n(k)$, then the minimal polynomials* $\text{min. poly}_k(A)$ *and* $\text{min. poly}_F(A)$ *are equal.*

PROOF. By $k[A]$ we denote the commutative subring of $M_n(k)$ generated by k and A. By $F[A]$ we denote the commutative subring of $M_n(F)$ generated by F

and A. We have the commutative diagram of rings,

$$\begin{array}{ccccc} F & \longrightarrow & F[A] & \longrightarrow & M_n(F) \\ \uparrow & & \uparrow & & \uparrow \\ k & \longrightarrow & k[A] & \longrightarrow & M_n(k) \end{array}$$

where each arrow is the set inclusion map. If $f = \text{min. poly}_k(A)$, and $\deg f = d$, then $k[A] \cong k[x]/(f)$ is a k-vector space of dimension d, and the set $\{1, A, \ldots, A^{d-1}\}$ is a k-basis. If $g = \text{min. poly}_F(A)$, then $\dim(F[A]) = \deg g$. Since $f(A) = 0$, by Theorem 4.4.8, we have $g \mid f$. This implies $\dim_F(F[A]) = \deg g \leq d$. By Exercise 6.1.20, the set $\{1, A, \ldots, A^{d-1}\}$ is a linearly independent set of vectors in the F-vector space $M_n(F)$. This implies $\dim_F(F[A]) \geq d$. We have shown that $\deg g = \dim_F(F[A]) = d$. Since $g \mid f$, this implies $g = f$. □

The following two technical lemmas characterize the rank of a lower triangular matrix and the eigenvalues in terms of the diagonal entries.

LEMMA 6.1.16. *Let k be a field, and $A = (a_{ij})$ a lower triangular matrix in $M_n(k)$. If $n \geq 2$, then assume $a_{22}, \ldots, a_{nn}$ are all nonzero elements in k. For $j = 1, \ldots, n$, let A_j denote column j of A. Then $\{A_1, A_2, \ldots, A_n\}$ is a linearly independent set of vectors in $k^{(n)}$ if and only if $a_{11} \neq 0$.*

PROOF. Let $e_1, \ldots, e_n$ denote the standard basis vectors for $k^{(n)}$. For $1 \leq m \leq n$, view $k^{(m)}$ as the subspace of $k^{(n)}$ spanned by $e_{n-m+1}, \ldots, e_n$. For $j = 1, \ldots, n$, we have A_j is in the span of $e_j, \ldots, e_n$. The proof follows by Lemma 4.3.1 and a standard induction argument. The basis for the induction is Lemma 4.3.1 (1). Assume $n > 1$ and that the span of $\{A_2, \ldots, A_n\}$ is equal to the span of $\{e_2, \ldots, e_n\}$. If $a_{11} \neq 0$, then A_1 is not in the span of $\{A_2, \ldots, A_n\}$, so by Lemma 4.3.1 (2), the set $\{A_1, A_2, \ldots, A_n\}$ is linearly independent. If $a_{11} = 0$, then A_1 is in the span of $\{A_2, \ldots, A_n\}$. □

LEMMA 6.1.17. *Let k be a field and $A = (a_{ij})$ a lower triangular matrix in $M_n(k)$.*

(1) A is invertible if and only if $a_{ii} \neq 0$ for each $1 \leq i \leq n$.
(2) The eigenvalues of A are $\{a_{ii} \mid 1 \leq i \leq n\}$.

PROOF. (1): As in Proposition 4.5.7, define α in $\text{Hom}_k(k^{(n)}, k^{(n)})$ to be "left multiplication by A". Then A is invertible if and only if α is invertible. By Theorem 6.1.8 (5), A is invertible if and only if α has rank n. The rank of α is equal to the column rank of A. By Lemma 6.1.16, A has column rank n if and only if $a_{ii} \neq 0$ for each $1 \leq i \leq n$.

(2): Let $\lambda \in k$. Then λ is an eigenvalue of A if and only if the matrix $A - \lambda$ is singular. But $A - \lambda$ is lower triangular and by Part (1), $A - \lambda$ is singular if and only if λ is equal to a_{ii} for some i. □

Next we show that a linear transformation has a matrix in triangular form if and only if the minimal polynomial splits.

THEOREM 6.1.18. *Let k be a field, V a finite dimensional k-vector space, and $\phi \in \text{Hom}_k(V, V)$. Then there exists a basis B for V such that $M(\phi, B)$ is lower triangular if and only if* $\text{min. poly}_k(\phi)$ *splits in k.*

PROOF. Let $n = \dim_k(V)$. If $B = \{v_1, \ldots, v_n\}$ is a basis for V, then the reader should verify that the matrix $M(\phi, B)$ is lower triangular if and only if for each i, $\phi(v_i)$ is in the span of $\{v_j \mid i \leq j \leq n\}$.

Let $f = \min.\,\mathrm{poly}_k(\phi)$ and assume f splits in k. Let λ be a root of f. By Theorem 6.1.10, λ is an eigenvalue of ϕ. Let v_n be an eigenvector corresponding to λ. By Lemma 4.3.1, $\{v_n\}$ is a linearly independent set. If $n = 1$, then take $B = \{v_n\}$ and stop. Otherwise, assume inductively that the triangular basis exists for a linear transformation on a vector space of dimension $n - 1$, whenever the minimal polynomial splits. If (v_n) is the subspace of V spanned by v_n, then (v_n) is a $k[\phi]$-submodule of V. If $\eta : V \to V/(v_n)$ is the natural map, then (v_n) is in the kernel of $\eta\phi$, so there exists a linear transformation $\bar{\phi}$ such that the diagram

$$\begin{array}{ccc} V & \xrightarrow{\phi} & V \\ {\scriptstyle\eta}\downarrow & & \downarrow{\scriptstyle\eta} \\ V/(v_n) & \xrightarrow{\bar{\phi}} & V/(v_n) \end{array}$$

commutes. The dimension of $V/(v_n)$ is $n - 1$. Since $f(\bar{\phi}) = 0$, the minimal polynomial of $\bar{\phi}$ divides f hence splits in k. By the induction hypothesis, there is a set of vectors $v_1, \ldots, v_{n-1}$ in V such that $\{v_i + (v_n) \mid 1 \leq i \leq n-1\}$ is a basis for $V/(v_n)$ and $\phi(v_i)$ is in the span of $\{v_j \mid i \leq j \leq n\}$. By Exercise 4.3.21, the set $B = \{v_1, \ldots, v_{n-1}, v_n\}$ is a basis for V. Moreover, the matrix $M(\phi, B)$ is lower triangular.

For the converse, suppose $B = \{v_1, \ldots, v_n\}$ is a basis for V such that the matrix $A = M(\phi, B)$ is lower triangular. By Lemma 6.1.17, if $A = (a_{ij})$, then the eigenvalues of ϕ are $\{a_{ii} \mid 1 \leq i \leq n\}$. Let $f = \min.\,\mathrm{poly}_k(\phi)$. If F is a splitting field for f containing k, then by Theorem 6.1.10 and Lemma 6.1.15, the roots of f are $\{a_{ii} \mid 1 \leq i \leq n\}$. Therefore, f splits in k. □

COROLLARY 6.1.19. *Let k be a field, V a finite dimensional k-vector space, and $\phi \in \mathrm{Hom}_k(V, V)$. If $\dim_k(V) = n$, then*

(1) the degree of $\min.\,\mathrm{poly}_k(\phi)$*, and*
(2) the dimension of the k-vector space $k[\phi]$

are both less than or equal to n.

PROOF. By Lemma 6.1.15 we assume $f = \min.\,\mathrm{poly}_k(\phi)$ splits in k. We use the notation from the proof of Theorem 6.1.18. There exists a basis $B = \{v_1, \ldots, v_n\}$ for V such that for each i, $\phi(v_i)$ is in the span of $\{v_j \mid i \leq j \leq n\}$. There is nothing to prove if $n = 1$. Inductively assume $n > 1$ and that the result is true for any linear transformation on a vector space of dimension $n - 1$. By induction on n applied to $\bar{\phi} \in \mathrm{Hom}_k(V/(v_n), V/(v_n))$, there is a polynomial $p(x) \in k[x]$ such that $\deg p \leq n-1$ and $p(\bar{\phi})(v_j + (v_n)) = 0 + (v_n)$, for $1 \leq j \leq n-1$. Thus $p(\phi)(V) \subseteq (v_n)$. Since v_n is an eigenvector of ϕ, there exists $\lambda \in k$ such that $\phi(v_n) = \lambda v_n$. Consider the polynomial $q(x) = (x - \lambda)p(x)$ which has degree n. We see that

$$\begin{aligned} q(\phi)(V) &= (\phi - \lambda)p(\phi)(V) \\ &= (\phi - \lambda)(v_n) \\ &= (0). \end{aligned}$$

By Theorem 4.4.8, f divides q and $\dim_k k[\phi] = \deg f \leq \deg q = n$. □

1.4. Exercises.

EXERCISE 6.1.20. Let F/k be an extension of fields. View the k-vector space $k^{(n)}$ as a subset of $F^{(n)}$ and the ring of matrices $M_n(k)$ as a subring of $M_n(F)$. Prove the following.

(1) Let $A \in M_n(k)$. Then A is invertible in $M_n(k)$ if and only if A is invertible in $M_n(F)$.
(2) If X is a basis for the k-vector space $k^{(n)}$, then X is a basis for the F-vector space $F^{(n)}$.
(3) If $Y = \{y_1, \ldots, y_m\}$ is a linearly independent set of vectors in the k-vector space $k^{(n)}$, then Y is a linearly independent set of vectors in the F-vector space $F^{(n)}$.

EXERCISE 6.1.21. Let F/k be an extension of fields. In the context of Exercise 6.1.20, let $A \in M_n(k)$ and assume $A \neq 0$ and $A \neq 1$. Prove the following.

(1) A is nilpotent in $M_n(k)$ if and only if A is nilpotent in $M_n(F)$.
(2) A is an idempotent in $M_n(k)$ if and only if A is an idempotent in $M_n(F)$.

EXERCISE 6.1.22. Let k be a field, V a finite dimensional k-vector space and $\phi \in \mathrm{Hom}_k(V, V)$. If $\lambda \in k$ is an eigenvalue of ϕ and $v \in V$ is a corresponding eigenvector, prove:

(1) For any $n \geq 0$, $\phi^n v = \lambda^n v$.
(2) For any $f(x) \in k[x]$, $f(\lambda)$ is an eigenvalue of $f(\phi)$.

EXERCISE 6.1.23. Let k be a field and $A = (a_{ij})$ a lower triangular matrix in $M_n(k)$. Let $f \in k[x]$. Prove that the eigenvalues of $f(A)$ are $\{f(a_{ii}) \mid 1 \leq i \leq n\}$.

EXERCISE 6.1.24. Let k be a field and $A \in M_n(k)$. Let $\lambda_1, \ldots, \lambda_m$ be the (not necessarily distinct) eigenvalues of A in k. Let $f \in k[x]$. Assume the minimal polynomial of A splits in k. Show that the eigenvalues of $f(A)$ are $f(\lambda_1), \ldots, f(\lambda_m)$.

EXERCISE 6.1.25. Let k be a field. As in Example 3.2.12, let N be the set of all lower triangular matrices $A = (a_{ij})$ in $M_n(k)$ such that $a_{ii} = 0$ for every diagonal entry. Prove the following.

(1) Every A in N is nilpotent.
(2) If A is a nilpotent matrix in $M_n(k)$, then there is an invertible matrix S in $M_n(k)$ such that $S^{-1}AS$ is in N.

EXERCISE 6.1.26. Let k be a field, $\{v_1, \ldots, v_n\}$ a basis for k^n, and $(\lambda_1, \ldots, \lambda_n) \in k^n$. Show how to construct a matrix $A \in M_n(k)$ such that λ_i is an eigenvalue for A with corresponding eigenvector v_i. That is, $Av_i = \lambda_i v_i$ for each $1 \leq i \leq n$.

EXERCISE 6.1.27. Let k be a field, V a finite dimensional k-vector space, u a nonzero vector in V, and $\phi \in \mathrm{Hom}_k(V, V)$. Let $f \in k[x]$ be the monic polynomial of minimal degree such that $f(\phi)u = 0$. Prove that f divides $\mathrm{min.poly}_k(\phi)$.

2. The Canonical Form of a Linear Transformation

If k is a field, V a finite dimensional k-vector space, and $\phi : V \to V$ a linear transformation, then we show that there is basis for V such that the matrix of ϕ is in so-called rational canonical form. As in Definition 6.1.1, V_ϕ denotes the $k[x]$-module structure on V associated to ϕ. Since V_ϕ is a finitely generated $k[x]$-module, we are in the context of Section 4.6. In particular, the invariant factor form

of the basis theorem for finitely generated modules over a principal ideal domain, Theorem 4.6.13, applies. Starting with the $k[x]$-basis for V_ϕ that results from applying Theorem 4.6.13, we derive a basis for V over k such that the matrix of ϕ is in rational canonical form (Corollary 6.2.5). Assuming the minimal polynomial for ϕ splits over k, we show that there is a basis for V such that the matrix of ϕ is in so-called Jordan canonical form (Corollary 6.2.7). The proof is an application of the elementary divisor form of the basis theorem, Theorem 4.6.12. With respect to the standard basis, a matrix in $M_n(k)$ defines a linear transformation on $k^{(n)}$. By treating a matrix A as a linear transformation, in Section 6.2.3, we define the rational canonical form for A. The canonical form is a unique matrix in the similarity class containing A. Two matrices are similar if and only if they have the same canonical form. In Section 6.2.4 we show that a matrix over a field has a unique reduced row echelon form.

2.1. Rational Canonical Form. This section should be treated as a continuation of Section 6.1.1. Given a linear transformation $\phi : V \to V$ on a finite dimensional vector space V over a field k, we have the $k[\phi]$-module V_ϕ (Definition 6.1.1). By Proposition 6.1.3 we can decompose V_ϕ into cyclic submodules over $k[x]$. Theorem 4.6.13, the invariant factor form of the basis theorem for a finitely generated module over a principal ideal domain, associates to ϕ a direct sum decomposition of V_ϕ into cyclic submodules and additionally a corresponding set of invariant factors $q_1, \dots q_r$ in $k[x]$. In Proposition 6.1.4 we showed that a generator for a cyclic $k[x]$-module gives rise to a k-basis. This allows us to show that there exists a basis B for the k-vector space V such that the matrix $M(\phi, B)$ is in so-called rational canonical form.

THEOREM 6.2.1. *If V is a finite dimensional vector space over the field k, and ϕ is a nonzero linear transformation in* $\mathrm{Hom}_k(V, V)$*, then there is a basis $\{u_1, u_2, \dots, u_r\}$ for the $k[\phi]$-module V_ϕ such that the following are true.*

(1) The $k[\phi]$-module V_ϕ is equal to the internal direct sum $U_1 \oplus U_2 \oplus \cdots \oplus U_r$ where $U_i = k[\phi]u_i$ is the cyclic submodule of V_ϕ spanned by u_i.
(2) $U_i \cong k[x]/(q_i)$ where q_i is the order of u_i and $q_1 \mid q_2 \mid \cdots \mid q_r$.
(3) U_i is a ϕ-invariant subspace of V and the minimal polynomial of $\phi|_{U_i}$ is q_i.
(4) The minimal polynomial of ϕ is q_r.
(5) The sequence of polynomials $(q_1, q_2, \dots, q_r)$ is uniquely determined by ϕ.

The polynomials $q_1, \dots, q_r$ are called the invariant factors *of ϕ.*

PROOF. Apply Theorem 4.6.13 to the finitely generated $k[x]$-module V_ϕ. □

Before we derive the rational canonical form of a linear transformation ϕ, we consider the possible sequences of invariant factors that can arise when the minimal polynomial $\mathrm{min.poly}_k(\phi)$ has at most two irreducible factors.

EXAMPLE 6.2.2. Let k be a field and V a k-vector space of dimension $n \geq 2$. Let $\phi \in \mathrm{Hom}_k(V, V)$ and assume the minimal polynomial of ϕ is irreducible. Say $\mathrm{min.poly}_k(\phi) = q$, and $\deg(q) = d$. What are the possible invariant factors of ϕ? We use the notation of Theorem 6.2.1. Say the decomposition of V_ϕ into cyclic submodules is $V_\phi = U_1 \oplus \cdots \oplus U_r$ where $U_i = k[\phi]u_i$ and u_i has order q_i. Then $q_1 \mid q_2 \mid \cdots \mid q_r$. Since $q_r = q$ is irreducible, this means each q_i is equal to q. Since $\dim_k(U_i) = \deg(q)$, this implies $n = dr$. This shows that if the minimal polynomial

of ϕ is an irreducible polynomial q of degree d, then $d \mid n$, and the invariants of ϕ are $q_1 = \cdots = q_r = q$, where $r = n/d$.

EXAMPLE 6.2.3. Let k be a field and V a k-vector space of dimension $n > 2$. Let $\phi \in \mathrm{Hom}_k(V, V)$ and assume the minimal polynomial of ϕ factors into two distinct monic irreducible factors. Suppose $q = \text{min. poly}_k(\phi) = \pi_1\pi_2$, where $\deg(\pi_i) = d_i$ and $\deg(q) = d_1 + d_2 = d$. What are the possible invariant factors of ϕ? We use the notation of Theorem 6.2.1. Say the decomposition of V_ϕ into cyclic submodules is $V_\phi = U_1 \oplus \cdots \oplus U_r$ where $U_i = k[\phi]u_i$ and u_i has order q_i. Then $q_1 \mid q_2 \mid \cdots \mid q_r$. Since $q_r = \pi_1\pi_2$, this means each q_i is one of π_1, π_2, or $\pi_1\pi_2$. There are three general configurations for the sequence $q_1, \ldots, q_r$, depending on whether q_1 is π_1, π_2, or $\pi_1\pi_2$.

(1) If $q_1 = \pi_1\pi_2$, then this means $q_1 = q_2 = \cdots = q_r = \pi_1\pi_2$. In this case, $\dim_k(U_i) = d$ for each i. This implies $n = dr$.
(2) If $q_1 = \pi_1$, then there exists some $p < r$ such that $q_1 = \cdots = q_p = \pi_1$, and $q_{p+1} = \cdots = q_r = \pi_1\pi_2$.
(3) If $q_1 = \pi_2$, then there exists some $p < r$ such that $q_1 = \cdots = q_p = \pi_2$, and $q_{p+1} = \cdots = q_r = \pi_1\pi_2$.

EXAMPLE 6.2.4. Let k be a field and V a k-vector space of dimension $n > 2$. Let $\phi \in \mathrm{Hom}_k(V, V)$ and assume the minimal polynomial of ϕ is the square of a monic irreducible polynomial. Suppose $q = \text{min. poly}_k(\phi) = \pi^2$, where $\deg(\pi) = d$. What are the possible invariant factors of ϕ? We use the notation of Theorem 6.2.1. Say the decomposition of V_ϕ into cyclic submodules is $V_\phi = U_1 \oplus \cdots \oplus U_r$ where $U_i = k[\phi]u_i$ and u_i has order q_i. Then $q_1 \mid q_2 \mid \cdots \mid q_r$. Since $q_r = \pi^2$, this means each q_i is either π or π^2. There are two general configurations for the sequence $q_1, \ldots, q_r$, depending on whether q_1 is π or π^2.

(1) If $q_1 = \pi^2$, then this means $q_1 = q_2 = \cdots = q_r = \pi^2$. In this case, $\dim_k(U_i) = 2d$ for each i. This implies $n = 2dr$.
(2) If $q_1 = \pi$, then there exists some $p < r$ such that $q_1 = \cdots = q_p = \pi$, and $q_{p+1} = \cdots = q_r = \pi^2$. In this case, $\dim_k(U_i) = d$ for $1 \le i \le p$ and $\dim_k(U_i) = 2d$ for $p < i \le r$. This implies $n = d(2r - p)$.

Now we determine a canonical form for the matrix of ϕ. In other words, we try to find a basis B of V for which the matrix $M(\phi, B)$ is simplified. If V and ϕ are as in Theorem 6.2.1, then $V = U_1 \oplus \cdots \oplus U_r$ where $\phi(U_i) \subseteq U_i$ for each i. Then each U_i is a k-subspace of V. We can pick a k-basis B_i for each subspace U_i and concatenate to get a k-basis $B = B_1 + \cdots + B_r$ for V. It is clear that the matrix of ϕ with respect to B is the block diagonal matrix

$$M(\phi, B) = \mathrm{diag}(M(\phi|_{U_1}, B_1), \ldots, M(\phi|_{U_r}, B_r))$$

where there are r blocks and block i is the matrix with respect to B_i of the restriction of ϕ to U_i. Based on this observation, we consider the case where $V_\phi = k[\phi]u$ is a cyclic module over the ring $k[\phi]$. We are in the context of Proposition 6.1.4. Suppose the minimal polynomial of ϕ is $\text{min. poly}_k(\phi) = p = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$. The $k[x]$-module homomorphism $k[x] \to k[\phi]u$ defined by $1 \mapsto u$ is surjective and the kernel is the principal ideal $I_u = (p)$ generated by p. Therefore, as a $k[x]$-module, V_ϕ is isomorphic to $k[x]/(p)$. Applying the division algorithm, we see that $1, x, x^2, \ldots, x^{n-1}$ is a k-basis for $k[x]/(p)$. Therefore, a k-basis for V is $B =$

$\{u, \phi u, \phi^2 u, \ldots, \phi^{n-1}u\}$. Introduce the notation $x_i = \phi^{i-1}u$. The action of ϕ on $B = \{x_1, x_2, \ldots, x_n\}$ determines the matrix $M(\phi, B)$. Computing, we get

$$\begin{aligned}
\phi x_1 &= \phi u = x_2 \\
\phi x_2 &= \phi\phi u = x_3 \\
&\vdots \\
\phi x_{n-1} &= \phi^{n-1}u = x_n \\
\phi x_n &= \phi^n u = -a_{n-1}\phi^{n-1}u - \cdots - a_1\phi^1 u - a_0 u = -a_0 x_1 - a_1 x_2 - \cdots - a_{n-1}x_n
\end{aligned}$$

so the matrix is

$$M(\phi, B) = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & 0 & -a_0 \\ 1 & 0 & 0 & \ldots & 0 & 0 & -a_1 \\ 0 & 1 & 0 & \ldots & 0 & 0 & -a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \\ 0 & 0 & 0 & \ldots & 0 & 0 & -a_{n-3} \\ 0 & 0 & 0 & \ldots & 1 & 0 & -a_{n-2} \\ 0 & 0 & 0 & \ldots & 0 & 1 & -a_{n-1} \end{bmatrix}. \tag{2.1}$$

We call (2.1) the *companion matrix* of the polynomial $p = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. If $p \in k[x]$ is a monic polynomial of degree $n \geq 1$, denote the companion matrix of p in $M_n(k)$ by $C(p)$. Conversely, by Exercise 6.2.22, the minimal polynomial of (2.1) is again $p = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$.

COROLLARY 6.2.5. *If V is a finite dimensional vector space over the field k, $\phi \in \mathrm{Hom}_k(V, V)$, and $q_1, q_2, \ldots, q_r$ are the invariant factors of ϕ, then there is a basis B for V such that the matrix of ϕ with respect to B is the block diagonal matrix*

$$M(\phi, B) = \mathrm{diag}\,(C(q_1), C(q_2), \ldots, C(q_r))$$

where block i is the companion matrix of q_i. The matrix $M(\phi, B)$ is called the rational canonical form *for ϕ.*

2.2. Jordan Canonical Form. Given a finite dimensional vector space V over a field k, and a linear transformation $\phi : V \to V$, we apply the Elementary Divisor Form of the Basis Theorem for a Finitely Generated Module over a Principal Ideal Domain to associate to ϕ a set of elementary divisors $\{\pi_i^{e_{ij}} \mid 1 \leq i \leq s; 1 \leq j \leq \nu_i\}$ in $k[x]$. Assuming the minimal polynomial of ϕ splits in k, we show that there exists a basis B for the k-vector space V such that the matrix $M(\phi, B)$ is in so-called Jordan canonical form.

THEOREM 6.2.6. *If V is a finite dimensional vector space over the field k, and ϕ is a nonzero linear transformation in $\mathrm{Hom}_k(V, V)$, then there exist positive integers $s, \nu_1, \ldots, \nu_s$ and a basis $\{u_{ij} \mid 1 \leq i \leq s; 1 \leq j \leq \nu_i\}$ for the $k[\phi]$-module V_ϕ such that the following are true.*

(1) The $k[\phi]$-module V_ϕ is equal to the internal direct sum

$$V_\phi = \bigoplus_{i=1}^{s} \bigoplus_{j=1}^{\nu_i} U_{ij}$$

where $U_i = k[\phi]u_{ij}$ is the cyclic submodule of V_ϕ spanned by u_{ij}.

(2) $U_{ij} \cong k[x]/(\pi_i^{e_{ij}})$ *where*
  (a) $\pi_1, \ldots, \pi_s$ *are distinct monic irreducible polynomials,*
  (b) the order of u_{ij} *is* $\pi_i^{e_{ij}}$*, and*
  (c) $e_{i1} \geq e_{i2} \geq \cdots \geq e_{i\nu_i} \geq 1$.

(3) U_{ij} *is a* ϕ*-invariant subspace of* V *and the minimal polynomial of* $\phi|_{U_{ij}}$ *is* $\pi_i^{e_{ij}}$.

(4) The minimal polynomial of ϕ *is*

$$\text{min. poly}_k(\phi) = \prod_{i=1}^{s} \pi_i^{e_{i1}}$$

(5) The sequence of irreducible polynomials $(\pi_1, \pi_2, \ldots, \pi_s)$ *and the positive integers* $\{e_{ij}\}$ *are uniquely determined by* ϕ.

The polynomials $\pi_i^{e_{ij}}$ *are called the* elementary divisors *of* ϕ.

PROOF. Apply Theorem 4.6.12 to the finitely generated $k[x]$-module V_ϕ. □

Using the basis for V_ϕ given by Theorem 6.2.6, we determine a canonical form for the matrix of ϕ. The minimal polynomial for ϕ restricted to U_{ij} is a power of the irreducible polynomial π_i. We assume each π_i is a linear polynomial, because the canonical form of ϕ in this case is particularly simplified. This case will occur if and only if the minimal polynomial of ϕ factors into a product of linear polynomials in $k[x]$. The k-bases for the individual ϕ-invariant subspaces U_{ij} can be concatenated for a basis of V. We now determine a canonical form for the matrix of ϕ under the following assumptions.

(1) V_ϕ is a cyclic $k[\phi]$-module spanned by u.
(2) $\text{min. poly}_k(\phi) = (x-b)^n$ is a power of a linear polynomial.

We are in the context of Proposition 6.1.4. Since $k[\phi] = k[\phi - b]$, it follows that V_ϕ is a cyclic $k[\phi - b]$-module, spanned by u. If $\theta : k[x] \to \text{Hom}_k(V, V)$ is defined by $x \mapsto \phi$, then $\ker\theta$ is the principal ideal generated by $(x-b)^n$. If $\tau : k[x] \to \text{Hom}_k(V, V)$ is defined by $x \mapsto \phi - b$, then the minimal polynomial of $\psi = \phi - b$ is the monic generator of $\ker\tau$, which is x^n. Therefore $B = \{u, \psi u, \psi^2 u, \ldots, \psi^{n-1}u\}$ is a k-basis for V. The matrix of $\psi = \phi - b$ with respect to the basis B is

$$M(\phi - b, B) = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 & 0 \\ \vdots & & & \vdots & & & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & \ldots & 1 & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 & 0 \end{bmatrix}$$

which is the companion matrix of the polynomial x^n. The matrix of ϕ with respect to the basis B is equal to $M(\phi, B) = M(\phi - b, B) + M(b, B)$. Therefore,

$$M(\phi, B) = \begin{bmatrix} b & 0 & 0 & \ldots & 0 & 0 & 0 \\ 1 & b & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & b & \ldots & 0 & 0 & 0 \\ \vdots & & & \vdots & & & \vdots \\ 0 & 0 & 0 & \ldots & b & 0 & 0 \\ 0 & 0 & 0 & \ldots & 1 & b & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 & b \end{bmatrix}. \tag{2.2}$$

We denote the n-by-n matrix (2.2) by $J_n(b)$ and refer to it as the basic *Jordan block* for the polynomial $(x-b)^n$. The matrix $J_n(b)$ and the canonical form of Corollary 6.2.7 are named after Camille Jordan.

COROLLARY 6.2.7. *Assume V is a finite dimensional vector space over the field k, $\phi \in \mathrm{Hom}_k(V,V)$, and that the minimal polynomial* $\mathrm{min.poly}_k(\phi)$ *factors into a product of linear factors in $k[x]$. If $b_1, \ldots, b_s$ are the distinct roots of* $\mathrm{min.poly}_k(\phi)$ *and $\{e_{ij}\}$ is the set of exponents of the elementary divisors of ϕ, then there is a basis B for V such that the matrix of ϕ with respect to B is the block diagonal matrix*

$$M(\phi, B) = \mathrm{diag}\left(J_{e_{11}}(b_1), J_{e_{12}}(b_1), \ldots, J_{e_{ij}}(b_i), \ldots\right)$$

where the block corresponding to the ordered pair (i,j) is the Jordan matrix of $(x-b_i)^{e_{ij}}$. The matrix $M(\phi, B)$ is called the Jordan canonical form *for ϕ and B is called a* Jordan basis.

EXAMPLE 6.2.8. Let k be a field and V a vector space of dimension n over k. Let $\phi : V \to V$ be a linear transformation and assume ϕ is a nontrivial idempotent in the ring $\mathrm{Hom}_k(V,V)$. Then $\phi^2 - \phi = 0$, $\phi \neq 1$, $\phi \neq 0$. Therefore, $\mathrm{min.poly}_k(\phi) = x^2 - x$. Let $V_1 = \ker(\phi)$ and $V_2 = \mathrm{im}(\phi)$. For each $v \in V_1$, we have $\phi(v) = 0$. For each $v \in V_2$, we have $\phi(v) = v$. Then each V_i is a ϕ-invariant subspace of V. By Proposition 4.2.8, $V = V_1 \oplus V_2$. Let B_1 be a basis for V_1 and B_2 a basis for V_2. If $n_i = \dim_k(V_i)$, then $n = n_1 + n_2$. The matrix of ϕ with respect to the basis $B = B_1 + B_2$ is $M(\phi, B) = \mathrm{diag}(0, \ldots, 0, 1, \ldots, 1) = \sum_{i=n_1+1}^{n} e_{ii}$. By Corollary 6.2.7, the matrix $M(\phi, B)$ is is the Jordan canonical form of ϕ, and B is a Jordan basis.

2.3. Canonical Form of a Matrix. Let k be a field, and A a matrix in $M_n(k)$. With respect to the standard basis on $k^{(n)}$, left multiplication by A defines a linear transformation ℓ_A in $\mathrm{Hom}_k(k^{(n)}, k^{(n)})$. The invariant factors, elementary divisors, rational canonical form, and the Jordan canonical form of A are defined to be the corresponding invariants of ℓ_A.

EXAMPLE 6.2.9. Let k be a field, $n \geq 2$ and e_{st} the elementary matrix in $M_n(k)$ where $s \neq t$. As in Example 6.1.5, the matrix e_{st} is nilpotent and the minimal polynomial is $\mathrm{min.poly}_k(e_{st}) = x^2$. Left multiplication by e_{st} defines the linear transformation ϕ_{st} in $\mathrm{Hom}_k(k^{(n)}, k^{(n)})$ defined by

$$\phi_{st}(e_i) = e_{st}e_i = \begin{cases} e_t & \text{if } i = s \\ 0 & \text{otherwise.} \end{cases}$$

Let $B_1 = \{e_1, \ldots, e_n\} - \{e_s, e_t\}$ be the standard basis with e_s and e_t deleted. Since each vector in B_1 is in the kernel of ϕ_{st}, if U_1 is the subspace of $k^{(n)}$ spanned by B_1, then U_1 is ϕ_{st}-invariant. The matrix $M(\phi_{st} \mid_{U_1}, B_1)$ is the zero matrix in $M_{n-2}(k)$. Let $B_2 = \{e_s, e_t\}$. Since $\phi_{st}(e_s) = e_t$ and $\phi_{st}(e_t) = 0$, if U_2 is the subspace of $k^{(n)}$ spanned by B_2, then U_2 is ϕ_{st}-invariant. The matrix $M(\phi_{st} \mid_{U_2}, B_2)$ is the matrix $J_2(0) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ in $M_2(k)$. If $B = B_1 + B_2$, then the matrix of ϕ_{st} with respect to B is the block diagonal matrix $M(\phi_{st}, B) = \mathrm{diag}(0, J_2(0))$. This shows that the Jordan canonical form of e_{st} is equal to the elementary matrix $e_{n,n-1}$ and B is a Jordan basis.

EXAMPLE 6.2.10. Let k be a field, $n \geq 2$ and e_{ss} the elementary matrix in $M_n(k)$. As in Example 6.1.5, e_{ss} is idempotent. By Example 6.2.8, the Jordan canonical form of e_{ss} is equal to the elementary matrix $e_{n,n}$.

LEMMA 6.2.11. *Let V be a finite dimensional vector space over the field k. Let ϕ and ψ be linear transformations in $\mathrm{Hom}_k(V,V)$. The $k[x]$-modules V_ϕ and V_ψ are isomorphic if and only if there exists an invertible linear transformation ρ in $\mathrm{Hom}_k(V,V)$ such that $\phi = \rho^{-1}\psi\rho$.*

PROOF. Let $f : V_\phi \to V_\psi$ be an isomorphism of $k[x]$-modules. Then f is an isomorphism of k-vector spaces. That is, $f = \rho$ for some invertible element ρ in $\mathrm{Hom}_k(V,V)$. For each $u \in V$ we have $f(\phi u) = \psi f(u)$. Therefore, $\phi = \rho^{-1}\psi\rho$. Conversely, if $\phi = \rho^{-1}\psi\rho$, define $f : V_\phi \to V_\psi$ by $f(u) = \rho u$. For $i \geq 1$, we have $\rho\phi^i = \psi^i\rho$. Then $f(\phi^i u) = \rho\phi^i u = \psi^i \rho u = \psi^i f(u)$. The rest follows from the fact that ρ is k-linear. □

COROLLARY 6.2.12. *Let k be a field, and A and B two matrices in $M_n(k)$. The following are equivalent.*

(1) A and B are similar.
(2) A and B have the same invariant factors.
(3) A and B have the same rational canonical form.

PROOF. If A and B have the same invariant factors, say $q_1, q_2, \ldots, q_r$, then they are both similar to the block diagonal matrix $C = \mathrm{diag}\,(C(q_1), C(q_2), \ldots, C(q_r))$. The matrix C is in rational canonical form. The reader should verify that the invariant factors of C are $q_1, \ldots, q_r$. If A and B are similar, then by Proposition 4.5.7 and Lemma 6.2.11, the $k[x]$-modules that they induce on k^n are isomorphic. So they have the same invariant factors. □

EXAMPLE 6.2.13. Consider the matrix $A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix}$ over the field $\mathbb{Q}$. Let $S = \{e_1, e_2, e_3\}$ be the standard basis for $V = \mathbb{Q}^{(3)}$. By Proposition 4.5.4, $A = M(\phi, S, S)$, where ϕ is the linear transformation in $\mathrm{Hom}_{\mathbb{Q}}(V,V)$ defined by multiplication by A from the left. Notice that $A^2 = \begin{bmatrix} 1 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, and $A^3 = 0$. Thus, A is nilpotent and the index of nilpotency is 3. This proves that $\mathrm{min.\,poly}(A) = x^3$. Since the minimal polynomial of A has only one root and is split, the rational canonical form of A is equal to the Jordan canonical form, which is $J_3(0) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. Let $u_1 = (1,0,0)^T$, $u_2 = Au_1 = (1,-1,1)^T$, and $u_3 = Au_2 = (1,-1,0)^T$. Then $B = \{u_1, u_2, u_3\}$ is a Jordan basis for ϕ. If $P = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -1 \\ 0 & 1 & 0 \end{bmatrix}$ is the matrix with columns u_1, u_2, u_3, the reader should verify that $P^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}$ and $P^{-1}AP = J_3(0)$.

2.4. Reduced Row Echelon Form. In this section we show that any matrix over a field has a unique reduced row echelon form. This canonical form exists whether the matrix is square or not. Using gaussian elimination and elementary row operations, an algorithm which is not included in this book, the reduced row echelon form can be efficiently computed. The application to the augmented matrix associated to a system of linear equations is in Proposition 6.2.20.

DEFINITION 6.2.14. Let k be a field and $R \in M_{mn}(k)$ an m-by-n matrix. We say R is in *reduced row echelon form*, if the following conditions are satisfied:

(1) Any row that consists only of zeros is below any nonzero row.
(2) The left-most nonzero entry of a row is equal to 1. We call this 1 a *leading* 1.
(3) The leading ones form a staggered, or echelon pattern from left to right and top to bottom. That is, if $i < j$ and rows i and j are nonzero, then the leading 1 in row i is to the left of the leading 1 in row j.
(4) Above and below any leading 1 are zeros.

LEMMA 6.2.15. *Let k be a field and $R \in M_{mn}(k)$ an m-by-n matrix in reduced row echelon form.*

(1) The rank of R is equal to the number of nonzero rows.
(2) The rank of R is equal to the number of leading ones.
(3) The nullity of R is equal to the number of columns that do not contain a leading 1.
(4) The columns that contain a leading one make up a basis for the column space of R.
(5) Let $R_1, \ldots, R_n$ be the columns of R. If R_j does not contain a leading 1*, then R_j is a unique linear combination of the columns in the set $\{R_1, \ldots, R_{j-1}\}$ that contain a leading one. In other words, there is a unique vector in the kernel of R of the form $(x_1, \ldots, x_{j-1}, 1, 0, \ldots, 0)$ such that $x_i = 0$, if $1 \leq i < j$ and R_i does not contain a leading* 1*.*

PROOF. The $(m-1)$-by-n submatrix of R consisting of rows $2, \ldots, m$ is in reduced row echelon form. Each nonzero row of R contains a leading one. For $1 \leq i < m$, since R is in row echelon form, if row i has a leading one, then row i is not in the span of rows $i+1, \ldots, m$. Therefore, (1) and (2) follow from induction on the number of nonzero rows. By Corollary 4.5.19 and Exercise 4.3.11, the nullity of R plus the rank of R is equal to n. Therefore, (3) follows from (2). The columns that contain a leading one are a subset of the standard basis $\{e_1, \ldots, e_m\}$ for $k^{(m)}$. By (2), the number of columns that contain a leading one is equal to the rank of R. This proves (4). Let R_j be a column that does not contain a leading one. Then R_j is a unique linear combination of the columns that contain a leading one. There is a unique vector $(x_1, \ldots, x_{j-1}, 1, x_{j+1}, \ldots, x_n)$ in the kernel of R such that $x_i = 0$ if $i \neq j$ and R_i does not contain a leading one. Moreover, since R is in row echelon form, $x_i = 0$ for $j < i \leq n$. This proves (5). □

PROPOSITION 6.2.16. *Let k be a field and $A \in M_{mn}(k)$.*

(1) There is an invertible matrix Q in $M_m(k)$ such that QA is in reduced row echelon form.

(2) The reduced row echelon form of A is unique in the sense that if Q_1 is another invertible matrix in $M_m(k)$ and Q_1A is in reduced row echelon form, then $QA = Q_1A$.

PROOF. (1): Let $X = \{A_1, A_2, \ldots, A_n\}$ be the columns of A. The column space of A is equal to the span of X in $k^{(m)}$. By Corollary 4.3.6 there exists a subset of X that is a basis for the column space of A. Let $U \subseteq X$ be a basis for the column space of A such that U is minimal with respect to the ordering on 2^X defined in Exercise 1.2.24. Then $U \subseteq X$ has the property that if $A_j \in X - U$, then A_j is a linear combination of $\{A_i \in U \mid i < j\}$. By Theorem 4.3.4, we can extend U to a basis for $k^{(m)}$. Call the resulting basis B. Let Q be the change of basis matrix. Then Q is an invertible matrix in $M_m(k)$. Let $QA = R$. We show that R is a matrix in reduced row echelon form. Let $\operatorname{Rank}(A) = r$ and $M_U = (u_1, \ldots, u_r)$ the m-by-r matrix with columns the r vectors in U. Then QM_U is the m-by-r matrix equal to the first r columns of the identity matrix I_m in $M_m(k)$. Therefore, the columns of A in U correspond to the standard basis vectors $e_1, \ldots, e_r$ in R. The column space of R is spanned by $e_1, \ldots, e_r$, hence rows $r+1, \ldots, m$ of R are zeros. As mentioned above, if $A_j \in X - U$, then A_j is a linear combination of those columns of A that are in U and to the left of A_j. This says that every nonzero row of R has a leading one.

(2): Since Q is invertible, the kernel of ℓ_{QA} is equal to the kernel of ℓ_A. Suppose $Q_1A = R_1$ and $Q_2A = R_2$ are two reduced row echelon forms for A. For sake of contradiction, suppose there is a difference in the columns containing leading ones. Say there is a leading 1 in column i of R_1 but not in column i of R_2. Then this contradicts Lemma 6.2.15 (5) because a column containing a leading 1 is not linearly dependent on the columns to its left. The uniqueness of those columns that do not contain leading ones follows from Lemma 6.2.15 (5) and the fact that the kernels of ℓ_{R_1} and ℓ_{R_2} are equal. □

PROPOSITION 6.2.17. *Let k be a field, A a matrix in $M_{mn}(k)$, and Q an invertible matrix in $M_m(k)$ such that QA is in reduced row echelon form.*

(1) The columns of QA containing leading ones correspond to a set of columns of A that make up a basis for the column space of A.
(2) If A has rank r, then the $n - r$ vectors defined by applying Lemma 6.2.15 (5) to QA make up a basis for the kernel of A.

PROOF. (1): Since Q is invertible, left multiplication by Q maps the column space of A isomorphically onto the column space of QA. A basis for the column space of QA corresponds to a basis for the column space of A. By Lemma 6.2.15 (4), the columns of QA that contain a leading one make up a basis for the column space of QA.

(2): The kernel of QA is equal to the kernel of A, since Q is invertible. □

EXAMPLE 6.2.18. Consider the matrix $A = \begin{bmatrix} 1 & 2 & -1 & 0 \\ 2 & 1 & 1 & 3 \\ 1 & -1 & 2 & 3 \end{bmatrix}$ over a field k, where we assume $\operatorname{char} k \neq 3$. Notice that $Q = \begin{bmatrix} -1/3 & 2/3 & 0 \\ 2/3 & -1/3 & 0 \\ 1 & -1 & 1 \end{bmatrix}$ is invertible

and the inverse is $Q^{-1} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix}$. Multiplying, $QA = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ is in reduced row echelon form. The rank of A is 2, the nullity of A is 2. By Proposition 6.2.17 (1), the first two columns of A make up a basis for the column space of A. By Proposition 6.2.17 (2), we obtain a basis for the kernel of A by writing columns 3 and 4 of QA as linear combinations of columns 1 and 2:

$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

A basis for the kernel of A is $(-1, 1, 1, 0)^T, (-2, 1, 0, 1)^T$.

2.4.1. *A System of Linear Equations.* Let k be a field. Consider a system of m linear equations in n variables over k:

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &=& b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &=& b_2 \\ \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \qquad\quad \vdots & & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &=& b_m \end{array} \tag{2.3}$$

Then the matrix of coefficients $A = (a_{ij})$ is in $M_{mn}(k)$ and the vector $b = (b_1, \ldots, b_m)^T$ on the right-hand side is in $k^{(m)}$. If $x = (x_1, \ldots, x_n)^T$, then (2.3) can be expressed in matrix form: $Ax = b$. With respect to the standard bases on $k^{(n)}$ and $k^{(m)}$, left multiplication by A defines a linear transformation ℓ_A in $\mathrm{Hom}_k(k^{(n)}, k^{(m)})$. The image of ℓ_A is the column space of A. The rank of A is the dimension of the column space of A. The nullity of A is the dimension of the kernel of ℓ_A.

PROPOSITION 6.2.19. *In the above context,*

(1) If b is in the image of ℓ_A, then the system of linear equations (2.3) *has a solution and we say the system of equations is* consistent. *Let $c = (c_1, \ldots, c_n)^T$ be a particular solution. Then the general solution to* (2.3) *is $x = c + z$, where $z = (z_1, \ldots, z_n)^T$ represents a typical element in the kernel of ℓ_A. The nullity of A is equal to the number of degrees of freedom in the solution. The solution x is unique if and only if the nullity of A is zero. If the nullity of A is positive, then we say the system of equations is* underdetermined.

(2) If b is not in the image of ℓ_A, then there is no solution to (2.3). *In this case, we say the system of equations is* inconsistent, *or* overdetermined.

PROOF. (1): The preimage $\ell_A^{-1}(b)$ is equal to the left coset $c + \ker(\ell_A)$. The rest of the proof is left to the reader. □

Let $A \in M_{mn}(k)$ and $B \in M_{mp}(k)$. Consider the matrix equation $AX = B$, where X is a variable. A solution to $AX = B$ is therefore a matrix X in $M_{np}(k)$. The *augmented matrix* is the block matrix $\begin{pmatrix} A & B \end{pmatrix}$, which is in $M_{m,n+p}(k)$. Then $\begin{pmatrix} A & B \end{pmatrix}$ is the matrix whose first n columns are the columns of A and whose last p columns are those of B.

PROPOSITION 6.2.20. *Let k be a field, $A \in M_{mn}(k)$, $B \in M_{mp}(k)$ and $\begin{pmatrix} A & B \end{pmatrix}$ the augmented matrix. If Q is an invertible matrix in $M_m(k)$ such that $Q \begin{pmatrix} A & B \end{pmatrix} = \begin{pmatrix} QA & QB \end{pmatrix}$ is in reduced row echelon form, then the equation $AX = B$ has a*

solution $X \in M_{np}$ if and only if every column of $\begin{pmatrix} QA & QB \end{pmatrix}$ with a leading one is in QA.

PROOF. First assume X exists. Then $(QA)X = QB$ says every column of QB in the column space of QA. By Lemma 6.2.15, QB has no column with a leading one. If $p = 1$, the converse follows from Propositions 6.2.17 (1) and 6.2.19 (1). The general case of the converse can be proved by induction on p. □

EXAMPLE 6.2.21. This is a continuation of Example 6.2.18. Consider the system of 3 linear equations in 4 variables:

$$\begin{aligned} x_1 + 2x_2 - x_3 &= 2 \\ 2x_1 + x_2 + x_3 + 3x_4 &= 7 \\ x_1 - x_2 + 2x_3 + 3x_4 &= 5 \end{aligned}$$

Then the matrix of coefficients is $A = \begin{bmatrix} 1 & 2 & -1 & 0 \\ 2 & 1 & 1 & 3 \\ 1 & -1 & 2 & 3 \end{bmatrix}$ and the right-hand side vector is $b = (2, 7, 5)^T$. From Example 6.2.18, the reduced row echelon form of A is obtained by multiplying by $Q = \begin{bmatrix} -1/3 & 2/3 & 0 \\ 2/3 & -1/3 & 0 \\ 1 & -1 & 1 \end{bmatrix}$. Let $x = (x_1, x_2, x_3, x_4)^T$. A basis for the kernel of A is $(-1, 1, 1, 0)^T, (-2, 1, 0, 1)^T$. Multiply both sides of the matrix equation $Ax = b$ by Q:

$$QAx = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix}$$

Then the general solution is:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \\ 0 \\ 0 \end{bmatrix} + a \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

where a and b represent arbitrary elements of k.

2.5. Exercises.

EXERCISE 6.2.22. Let k be a field, V a k-vector space of dimension n, and $\phi \in \mathrm{Hom}_k(V, V)$. Suppose $B = \{x_1, \ldots, x_n\}$ is a k-basis for V and $a_0, \ldots, a_{n-1}$ are elements of k such that $\phi x_1 = x_2$, $\phi x_2 = x_3$, $\ldots$, $\phi x_{n-1} = x_n$, and $\phi x_n = -a_0 x_1 - a_1 x_2 - \cdots - a_{n-1} x_n$. Prove:

(1) $V_\phi = k[\phi]x_1$. In other words, V_ϕ is a cyclic $k[\phi]$-module and is generated by x_1.
(2) $\mathrm{min.poly}_k(\phi) = x^n + a_{n-1}x_{n-1} + \cdots + a_1 x + a_0$.

EXERCISE 6.2.23. Assume A is an n-by-n matrix over the field $\mathbb{Q}$ such that the minimum polynomial of A in $\mathbb{Q}[x]$ is equal to $(x^2 + 1)(x + 2)$. If $n = 7$, exhibit all possible rational canonical forms for A.

EXERCISE 6.2.24. Let k be a field. Let q and ℓ be monic polynomials in $k[x]$, where q is an irreducible quadratic and ℓ is linear. If A is a 7-by-7 matrix over k such that the minimum polynomial of A in $k[x]$ is $q\ell$, exhibit all possible rational canonical forms for A.

EXERCISE 6.2.25. Let k be a field. Let q and ℓ be monic polynomials in $k[x]$, where q is an irreducible quadratic and ℓ is linear. Let A be a 6-by-6 matrix over k. Exhibit all possible rational canonical forms for A, if the minimum polynomial of A in $k[x]$ is $q^2\ell$. Do the same if the minimum polynomial of A in $k[x]$ is $\ell^2 q$.

EXERCISE 6.2.26. Let k be a field. Let q and t be irreducible monic polynomials in $k[x]$, where $\deg q = 2$ and $\deg t = 3$. Let A be a 15-by-15 matrix over k. Exhibit all possible rational canonical forms for A, if the minimum polynomial of A in $k[x]$ is q^2t^2. Do the same if the minimum polynomial of A in $k[x]$ is q^3t.

EXERCISE 6.2.27. Let k be a field. Let q_1, q_2 and ℓ be distinct irreducible monic polynomials in $k[x]$, where q_1 and q_2 are quadratics and ℓ is linear. Let A be a 10-by-10 matrix over k. Exhibit all possible rational canonical forms for A, if the minimum polynomial of A in $k[x]$ is $\ell q_1^2 q_2$.

EXERCISE 6.2.28. Let k be a field. Let ℓ_1, ℓ_2 be distinct monic polynomials in $k[x]$, where $\deg \ell_1 = \deg \ell_2 = 1$. Let A be an 8-by-8 matrix over k. Exhibit all possible rational canonical forms for A, if the minimum polynomial of A in $k[x]$ is $\ell_1^2\ell_2^3$.

EXERCISE 6.2.29. Let F/k be an extension of fields. Prove the following.

(1) If $A \in M_n(k)$, then the invariant factors of A in $k[x]$ are the same as the invariant factors of A in $F[x]$.
(2) Let $A, B \in M_n(k)$. Then A is similar to B in $M_n(k)$ if and only if A is similar to B in $M_n(F)$.

EXERCISE 6.2.30. Let k be a field and $b \in k$. Let $B \in M_n(k)$ be the Jordan block corresponding to $(x-b)^n$. That is, B is the matrix which has main diagonal entries all equal to b, first lower subdiagonal entries all equal to 1 and 0 elsewhere. Prove that the transpose of B is similar to B. For a continuation of this exercise, see Exercise 6.2.32.

EXERCISE 6.2.31. Let k be a field, V a finite dimensional k-vector space of dimension $n > 1$. Let $\alpha \in \mathrm{Hom}_k(V, V)$ be a nilpotent linear transformation on V. Prove:

(1) There exist unique positive integers $n_1 \leq n_2 \leq \cdots \leq n_r$ such that the invariant factors of α are $\{q_i = x^{n_i} \mid 1 \leq i \leq r\}$.
(2) The rational canonical form and the Jordan canonical form of α are both equal to the block diagonal matrix $\mathrm{diag}(J_{n_1}(0), \ldots, J_{n_r}(0))$.

EXERCISE 6.2.32. Let k be a field and A a matrix in $M_n(k)$. Prove that A is similar to the transpose of A.

3. The Determinant

Throughout this section, R is a commutative ring and n is a fixed positive integer. First we prove that the determinant function $\det : M_n(R) \to R$ exists and is the unique alternating multilinear form (on the columns) such that if I_n is the

identity matrix, then $\det(I_n) = 1$. Then we show that the determinant function is multiplicative and constant on similarity classes. The formulas for computing determinants by cofactor expansion are derived. The formula for the inverse of a matrix involving the adjoint matrix is derived. Using the isomorphism of Proposition 4.5.7, the determinant function extends to a function $\det : \mathrm{Hom}_R(M, M) \to R$ for any free R-module M of finite rank.

3.1. The Determinant of a Matrix. Let $J = \{1, \ldots, n\}$ and $J^n = J \times \cdots \times J$ (n times). We view the symmetric group S_n as the subset of J^n consisting of n-tuples $\vec{j} = (j_1, \ldots, j_n)$ that are permutations of J. The sign of a permutation $\sigma \in S_n$ is denoted $\mathrm{sign}(\sigma)$.

DEFINITION 6.3.1. Let R be a commutative ring, $n \geq 1$, and $(R^n)^n = \bigoplus_{i=1}^n R^n$. Consider a function $f : (R^n)^n \to R$. We say that f is a *multilinear form* if for each i,

$$f(x_1, \ldots, x_{i-1}, \alpha u + \beta v, x_{i+1}, \ldots, x_n) = \\ \alpha f(x_1, \ldots, x_{i-1}, u, x_{i+1}, \ldots, x_n) + \beta f(x_1, \ldots, x_{i-1}, v, x_{i+1}, \ldots, x_n).$$

We say that f is an *alternating form* if $f(x_1, , \ldots, x_n) = 0$ whenever $x_i = x_j$ for some pair $i \neq j$.

LEMMA 6.3.2. *If $f : (R^n)^n \to R$ is an alternating multilinear form and $\sigma \in S_n$ is a permutation on the set $\{1, \ldots, n\}$, then*

$$f(x_{\sigma 1}, \ldots, x_{\sigma n}) = \mathrm{sign}(\sigma) f(x_1, , \ldots, x_n).$$

We say that f is skew symmetric.

PROOF. Because σ factors into a product of transpositions, it is enough to show that acting on the variables by a transposition changes the sign of f. For simplicity's sake, assume $\sigma = (i, j) = (1, 2)$. Look at

$$\begin{aligned} 0 &= f(x_1 + x_2, x_1 + x_2, x_3, \ldots, x_n) \\ &= f(x_1, x_1, x_3, \ldots, x_n) + f(x_1, x_2, x_3, \ldots, x_n) + \\ &\quad f(x_2, x_1, x_3, \ldots, x_n) + f(x_2, x_2, x_3, \ldots, x_n) \\ &= f(x_1, x_2, x_3, \ldots, x_n) + f(x_2, x_1, x_3, \ldots, x_n). \end{aligned}$$

This shows $f(x_1, x_2, x_3, \ldots, x_n) = -f(x_2, x_1, x_3, \ldots, x_n)$. □

LEMMA 6.3.3. *If R is a commutative ring and $r \in R$, there is a unique alternating multilinear form $f : (R^n)^n \to R$ such that $f(e_1, \ldots, e_n) = r$, where $(e_1, \ldots, e_n)$ is the standard basis for R^n.*

PROOF. (Uniqueness) Given $(x_1, \dots, x_n) \in (R^n)^n$, for each i we can write $x_i = a_{1i}e_1 + \cdots + a_{ni}e_n$. Since f is multilinear,

$$\begin{aligned} f(x_1, \dots, x_n) &= f\Big(\sum_{j\in J} a_{j1}e_j, \dots, \sum_{j\in J} a_{jn}e_j\Big) \\ &= \sum_{j_1\in J}\left(a_{j_1 1} f\Big(e_{j_1}, \sum_{j\in J} a_{j2}e_j, \dots, \sum_{j\in J} a_{jn}e_j\Big)\right) \\ &= \sum_{j_1\in J}\sum_{j_2\in J}\left(a_{j_1 1}a_{j_2 2} f\Big(e_{j_1}, e_{j_2}, \dots, \sum_{j\in J} a_{jn}e_j\Big)\right) \\ &\vdots \\ &= \sum_{(j_1,\dots,j_n)\in J^n} a_{j_1 1}\cdots a_{j_n n} f\Big(e_{j_1}, \dots, e_{j_n}\Big). \end{aligned} \tag{3.1}$$

If $\vec{j} = (j_1, \dots, j_n) \in J^n$ is not a permutation, then $f(e_{j_1}, \dots, e_{j_n}) = 0$ since f is alternating. We can restrict the last summation in Eq. (3.1) to those $\vec{j}$ in S_n. In this case, since f is skew symmetric, $f(e_{j_1}, \dots, e_{j_n}) = \operatorname{sign}(\vec{j}) f(e_1, \dots, e_n) = \operatorname{sign}(\vec{j}) r$. This proves that

$$f(x_1, \dots, x_n) = r \sum_{\vec{j}\in S_n} \operatorname{sign}(\vec{j}) a_{j_1 1}\cdots a_{j_n n} \tag{3.2}$$

is completely determined by r and $(x_1, \dots, x_n)$.

(Existence) The formula in (3.2) defines a function $f : (R^n)^n \to R$. Notice that

$$f(e_1, \dots, e_n) = r$$

since only for $\vec{j} = (1, 2, \dots, n)$ is the product formula in the summation (3.2) nonzero. We need to prove f is an alternating multilinear form. Let $\alpha, \beta \in R$, $u, v \in R^n$ and $(x_1, \dots, x_n) \in (R^n)^n$ where $x_k = \alpha u + \beta v$. Write $x_j = \sum_{i=1}^n a_{ij}e_i$. If we write $u = \sum u_i e_i$ and $v = \sum v_i e_i$, then $a_{ik} = \alpha u_i + \beta v_i$, so that $x_k = \sum a_{ik}e_i = \sum(\alpha u_i + \beta v_i)e_i = \alpha u + \beta v$. Then

$$\begin{aligned} f(x_1, \dots, \alpha u + \beta v, \dots, x_n) &= r \sum_{\vec{j}\in S_n} \operatorname{sign}(\vec{j}) a_{j_1 1}\cdots a_{j_k k}\cdots a_{j_n n} \\ &= r \sum_{\vec{j}\in S_n} \operatorname{sign}(\vec{j}) a_{j_1 1}\cdots(\alpha u_{j_k} + \beta v_{j_k})\cdots a_{j_n n} \\ &= r\alpha \sum_{\vec{j}\in S_n} \operatorname{sign}(\vec{j}) a_{j_1 1}\cdots u_{j_k}\cdots a_{j_n n} + \\ &\qquad r\beta \sum_{\vec{j}\in S_n} \operatorname{sign}(\vec{j}) a_{j_1 1}\cdots v_{j_k}\cdots a_{j_n n} \\ &= \alpha f(x_1, \dots, u, \dots, x_n) + \beta f(x_1, \dots, v, \dots, x_n) \end{aligned}$$

shows f is multilinear.

Now we show f is alternating. Suppose $i < j$ and let τ be the transposition that switches i and j. The alternating group A_n has index 2 in S_n, so every odd permutation is of the form $\sigma\tau$ for some $\sigma \in A_n$. Assume $x_i = x_j$ and show

$f(x_1, \ldots, x_n) = 0$. For all k we have $a_{ki} = a_{kj}$. Also, if $\sigma \in A_n$ then $\sigma\tau(i) = \sigma(j)$ and $\sigma\tau(j) = \sigma(i)$.

$$\begin{aligned} f(x_1, \ldots, x_n) &= r \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{\sigma(1)1} \cdots a_{\sigma(n)n} \\ &= r \sum_{\sigma \in A_n} \left(a_{\sigma(1)1} \cdots a_{\sigma(n)n} - a_{\sigma\tau(1)1} \cdots a_{\sigma\tau(n)n} \right) \\ &= r \sum_{\sigma \in A_n} \left(a_{\sigma(1)1} \cdots a_{\sigma(n)n} - a_{\sigma\tau(1)1} \cdots a_{\sigma\tau(i)i} \cdots a_{\sigma\tau(j)j} \cdots a_{\sigma\tau(n)n} \right) \\ &= r \sum_{\sigma \in A_n} \left(a_{\sigma(1)1} \cdots a_{\sigma(n)n} - a_{\sigma(1)1} \cdots a_{\sigma(j)i} \cdots a_{\sigma(i)j} \cdots a_{\sigma(n)n} \right) \\ &= r \sum_{\sigma \in A_n} \left(a_{\sigma(1)1} \cdots a_{\sigma(n)n} - a_{\sigma(1)1} \cdots a_{\sigma(j)j} \cdots a_{\sigma(i)i} \cdots a_{\sigma(n)n} \right) \\ &= 0. \end{aligned}$$

□

DEFINITION 6.3.4. By viewing the columns of a matrix in $M_n(R)$ as vectors in R^n, we identify $M_n(R)$ with $(R^n)^n$. The *determinant* is the unique alternating multilinear form $\det : M_n(R) \to R$ such that $\det(I_n) = 1$. By Lemma 6.3.3,

$$\det(a_{ij}) = \sum_{\vec{j} \in S_n} \operatorname{sign}(\vec{j}) a_{j_1,1} \cdots a_{j_n,n}.$$

LEMMA 6.3.5. *Let $A, B \in M_n(R)$.*

(1) $\det(AB) = \det(A)\det(B)$.
(2) If A and B are similar, then $\det(A) = \det(B)$.
(3) $\det(A) = \det(A^T)$.
(4) The determinant is an alternating multilinear form on the rows of matrices in $M_n(R)$.

PROOF. (1): Fix A. Taking $r = \det(A)$ in (3.2) defines an alternating multilinear form $g : M_n(R) \to R$, where $g(C) = \det(A)\det(C)$. Define another function $f : M_n(R) \to R$ by $f(C) = \det(AC)$. Since $f(I_n) = \det(A)$, by Lemma 6.3.3, it is enough to prove that f is alternating and multilinear. Assume $\alpha, \beta \in R$, $u, v \in R^n$ and $C = (c_1, \ldots, c_n) \in M_n(R)$ where $c_k = \alpha u + \beta v$. Then

$$\begin{aligned} f(c_1, \ldots, \alpha u + \beta v, \ldots, c_n) &= \det\left(A(c_1, \ldots, \alpha u + \beta v, \ldots, c_n)\right) \\ &= \det\left(Ac_1, \ldots, \alpha Au + \beta Av, \ldots, Ac_n\right) \\ &= \alpha \det\left(Ac_1, \ldots, Au, \ldots, Ac_n\right) + \\ &\qquad \beta \det\left(Ac_1, \ldots, Av, \ldots, Ac_n\right) \\ &= \alpha f(c_1, \ldots, u, \ldots, c_n) + \beta f(c_1, \ldots, v, \ldots, c_n). \end{aligned}$$

If two columns of C are equal, then two columns of AC are equal, so f is alternating.

(2): If $A = X^{-1}BX$, then

$$\begin{aligned} \det(A) &= \det(X^{-1})\det(B)\det(X) \\ &= \det(B)\det(X^{-1})\det(X) \\ &= \det(B)\det(X^{-1}X) \\ &= \det(B). \end{aligned}$$

(3): Since R is commutative, for every $\sigma \in S_n$ we have

$$a_{\sigma(1),1} \cdots a_{\sigma(n),n} = a_{1,\sigma^{-1}(1)} \cdots a_{n,\sigma^{-1}(n)}.$$

This together with the fact that $\operatorname{sign}(\sigma) = \operatorname{sign}(\sigma^{-1})$ lead to

$$\begin{aligned}
\det(A) &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n} \\
&= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1,\sigma^{-1}(1)} \cdots a_{n,\sigma^{-1}(n)} \\
&= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} \\
&= \det(A^T).
\end{aligned}$$

(4): Follows from (3). □

DEFINITION 6.3.6. For $A \in M_n(R)$, let A_{ij} be the matrix in $M_{n-1}(R)$ obtained by deleting row i and column j from A. Then $\det(A_{ij})$ is called the *minor* of A in position (i, j) and $(-1)^{i+j} \det(A_{ij})$ is called the *cofactor* of A in position (i, j).

LEMMA 6.3.7. *If A is a matrix in $M_n(R)$, then the following are true.*

(1) For each row i, $\det(A) = \sum_{j=1}^{n} a_{ij}(-1)^{i+j} \det(A_{ij})$, and

(2) For each column j, $\det(A) = \sum_{i=1}^{n} a_{ij}(-1)^{i+j} \det(A_{ij})$.

PROOF. We prove that the determinant can be computed by cofactor expansion of row i. The statement about column expansion follows from Lemma 6.3.5 (3). Define a function $f : M_n(R) \to R$ by the formula $f(A) = \sum_{j=1}^{n} a_{ij}(-1)^{i+j} \det(A_{ij})$. The reader should verify that $f(I_n) = 1$. By Lemma 6.3.3 it is enough to show that f is alternating and multilinear.

Assume the columns of A are $(A_1, \ldots, A_n)$ and assume $A_k = A_\ell$ and $k < \ell$. Therefore $a_{ik} = a_{i\ell}$. If $j \neq k$ and $j \neq \ell$, then A_{ij} has two columns that are equal, so $\det(A_{ij}) = 0$. The formula for f reduces to

$$\begin{aligned}
f(A) &= a_{ik}(-1)^{i+k} \det(A_{ik}) + a_{i\ell}(-1)^{i+\ell} \det(A_{i\ell}) \\
&= a_{ik}(-1)^{i+k} \det(A_{ik}) + a_{ik}(-1)^{i+\ell} \det(A_{i\ell}) \\
&= a_{ik} \left((-1)^{i+k} \det(A_{ik}) + (-1)^{i+\ell} \det(A_{i\ell}) \right).
\end{aligned}$$

But A_{ik} is obtained from $A_{i\ell}$ by permuting the columns. In fact, $\ell - k - 1$ transpositions are sufficient. Since the determinant form is skew symmetric, $\det(A_{ik}) = (-1)^{\ell-k-1} \det(A_{i\ell})$. The reader should verify that $(-1)^{i+k} + (-1)^{i+\ell}(-1)^{\ell-k-1} = 0$, hence

$$\begin{aligned}
f(A) &= a_{ik} \left((-1)^{i+k} \det(A_{ik}) + (-1)^{i+\ell} \det(A_{i\ell}) \right) \\
&= a_{ik} \left((-1)^{i+k} \det(A_{ik}) + (-1)^{i+\ell}(-1)^{\ell-k-1} \det(A_{ik}) \right) \\
&= a_{ik} \det(A_{ik}) \left((-1)^{i+k} + (-1)^{i+\ell}(-1)^{\ell-k-1} \right) \\
&= 0
\end{aligned}$$

which proves f is alternating.

Assume the columns of A are $(A_1, \ldots, A_n)$ where $A_k = \alpha u + \beta v$ for some $u, v \in R^n$. Let $B = (b_{ij})$ be the matrix obtained by replacing column k of A with the vector u. Let $C = (c_{ij})$ be the matrix obtained by replacing column k of A with the vector v. We show that $f(A) = \alpha f(B) + \beta f(C)$. Because they differ only

in column k, we have $A_{ik} = B_{ik} = C_{ik}$. The determinant is multilinear, so if $j \neq k$, then $\det(A_{ij}) = \alpha \det(B_{ij}) + \beta \det(C_{ij})$. Therefore

$$\begin{aligned} f(A) &= \sum_{j=1}^{n} a_{ij}(-1)^{i+j} \det(A_{ij}) \\ &= \sum_{j \neq k} a_{ij}(-1)^{i+j} \left(\alpha \det(B_{ij}) + \beta \det(C_{ij})\right) + (\alpha b_{ik} + \beta c_{ik})(-1)^{i+k} \det(A_{ik}) \\ &= \alpha \sum_{j=1}^{n} b_{ij}(-1)^{i+j} \det(B_{ij}) + \beta \sum_{j=1}^{n} c_{ij}(-1)^{i+j} \det(C_{ij}) \\ &= \alpha f(B) + \beta f(C) \end{aligned}$$

□

DEFINITION 6.3.8. Let $A \in M_n(R)$. The *adjoint* of A, denoted A^a, is the transpose of the matrix of cofactors of A. Therefore, $A^a = \left((-1)^{i+j} \det(A_{ji})\right)$.

LEMMA 6.3.9. $A^a A = A A^a = \det(A) I_n$.

PROOF. Assume $i \neq j$. Let B be the matrix which is equal to A with column i replaced with a copy of column j. Compute $\det(B) = 0$ by column expansion down column i. Use the facts that $B_{ki} = A_{ki}$ and $b_{ki} = b_{kj} = a_{kj}$ for each k.

$$\begin{aligned} 0 &= \sum_{k=1}^{n} b_{ki}(-1)^{i+k} \det(B_{ki}) \\ &= \sum_{k=1}^{n} a_{kj}(-1)^{i+k} \det(A_{ki}) \end{aligned}$$

Let $A^a A = (c_{ij})$. Then

$$c_{ij} = \sum_{k=1}^{n} (-1)^{i+k} \det(A_{ki}) a_{kj} = \begin{cases} \det(A) & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

□

COROLLARY 6.3.10. *Let R be a commutative ring and $A \in M_n(R)$. Then A is invertible if and only if $\det(A)$ is a unit in R.*

PROOF. If $AB = I_n$, then $\det(A)\det(B) = 1$. The converse follows from Lemma 6.3.9 because in this case $A^{-1} = \det(A)^{-1} A^a$. □

DEFINITION 6.3.11. By Lemma 6.3.5 (2), the determinant function is constant on similarity classes. If M is a finitely generated free R-module of rank n and $\phi \in \mathrm{Hom}_R(M, M)$, then the determinant of ϕ is defined to be the determinant of the matrix of ϕ with respect to any basis B of M. Given M and a basis B, using the isomorphism of Proposition 4.5.7, the determinant function extends to a function $\det : \mathrm{Hom}_R(M, M) \to R$ such that the diagram

$$\begin{array}{ccc} \mathrm{Hom}_R(M, M) & \xrightarrow{M(\cdot, B)} & M_n(R) \\ & {\scriptstyle \det} \searrow \quad \swarrow {\scriptstyle \det} & \\ & R & \end{array}$$

commutes.

3.2. The Characteristic Polynomial. Using determinants, the characteristic polynomial is defined for a square matrix over a commutative ring R. As in Definition 6.3.11, the definition of characteristic polynomial extends to any $\phi \in \mathrm{Hom}_R(M, M)$, if M is a finitely generated free R-module of finite rank.

DEFINITION 6.3.12. Let R be a commutative ring and $M \in M_n(R)$. If x is an indeterminate, then we can view M as a matrix in $M_n(R[x])$. The *characteristic polynomial* of M is $\mathrm{char.poly}_R(M) = \det(xI_n - M)$, which is a polynomial in $R[x]$. Computing the determinant using row expansion (Lemma 6.3.7) along row one, it is easy to see that $\mathrm{char.poly}_R(M)$ is monic and has degree n. The characteristic polynomial is constant on similarity classes, by Exercise 6.3.18. If P is a finitely generated free R-module and $\phi \in \mathrm{Hom}_R(P, P)$, then the characteristic polynomial of ϕ is defined to be the characteristic polynomial of the matrix of ϕ with respect to any basis of P.

THEOREM 6.3.13. *Let k be a field and V a finite dimensional vector space over k. Let $\phi \in \mathrm{Hom}_k(V, V)$. As in Theorem 6.2.1, let $q_1, q_2, \ldots, q_r$ be the invariant factors of ϕ.*

(1) $\mathrm{char.poly}_k(\phi) = q_1 q_2 \cdots q_r$.

(2) (Cayley-Hamilton) If $p(x) = \mathrm{char.poly}_k(\phi)$, then $p(\phi) = 0$. The minimal polynomial of ϕ divides the characteristic polynomial of ϕ. That is, $\mathrm{min.poly}_k(\phi) \mid \mathrm{char.poly}_k(\phi)$.

(3) If $f \in k[x]$ is irreducible, then $f \mid \mathrm{char.poly}_k(\phi)$ if and only if $f \mid \mathrm{min.poly}_k(\phi)$. The roots of the minimal polynomial $\mathrm{min.poly}_k(\phi)$ *are precisely the roots of the characteristic polynomial* $\mathrm{char.poly}_k(\phi)$.

PROOF. (1): By Corollary 6.2.5 there is a basis for V such that the matrix of ϕ is the block diagonal matrix $(C(q_1), C(q_2), \ldots, C(q_r))$, where $C(q_i)$ is the companion matrix for q_i. By Exercise 6.3.17, the characteristic polynomial of $C(q_i)$ is q_i. Apply Exercise 6.3.19 iteratively to show that $\mathrm{char.poly}_k(\phi) = q_1 q_2 \cdots q_r$.

(2): By Theorem 6.2.1, $\mathrm{min.poly}_k(\phi) = q_r$.

(3): By Theorem 6.2.1, $q_1 \mid q_2 \mid \cdots \mid q_r$. The irreducible factors of $\mathrm{char.poly}_k(\phi)$ are equal to the irreducible factors of $\mathrm{min.poly}_k(\phi)$. □

EXAMPLE 6.3.14. Consider the matrix $B = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 0 & 1 & 1 \end{bmatrix}$ over the field $\mathbb{Q}$. Then $B^2 = \begin{bmatrix} 0 & 1 & 1 \\ 0 & -1 & -1 \\ -1 & 0 & 0 \end{bmatrix}$, and $B^3 = \begin{bmatrix} -1 & 0 & 0 \\ 1 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$. Using determinants we compute the characteristic polynomial of B:

$$\begin{aligned} \mathrm{char.poly}(B) &= \det(x - B) \\ &= \begin{vmatrix} x-1 & -1 & -1 \\ 1 & x+1 & 1 \\ 0 & -1 & x-1 \end{vmatrix} \\ &= (x-1)(x+1)(x-1) + 1 + (x-1) + (x-1) \\ &= x(x^2 - x + 1). \end{aligned}$$

The roots of the characteristic polynomial are 0, $\alpha = (1-\sqrt{3}i)/2$ and $\beta = (1+\sqrt{3}i)/2$. By Theorem 6.1.10, $0, \alpha, \beta$ are also roots of the minimal polynomial of B. This proves that $\text{min. poly}(B) = x(x^2-x+1)$. The rational canonical form of B over $\mathbb{Q}$ is therefore equal to the companion matrix of $x(x^2-x+1)$, which is $C(x^3-x^2+x) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}$. Let $V = \mathbb{Q}^{(3)}$ and $\psi \in \text{Hom}_{\mathbb{Q}}(V, V)$ the linear transformation corresponding to left multiplication by B. Since $\text{min. poly}(\psi)$ has degree 3, we know V is a cyclic $\mathbb{Q}[\psi]$-module. Let $u_1 = (1,0,0)^T$, $u_2 = Bu_1 = (1,-1,0)^T$, and $u_3 = Bu_2 = (0,0,-1)^T$. Then $U = \{u_1, u_2, u_3\}$ is a basis for V such that $M(\psi, U, U) = C(x^3-x^2+x)$. Set $P = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$. Then we see that $P = P^{-1}$ and $PBP = C(x^3-x^2+x)$. The Jordan canonical form of ψ exists over $F = \mathbb{Q}(\alpha)$, the splitting field of x^2-x+1. Since B has 3 distinct eigenvalues, the Jordan form of ψ is the diagonal matrix $\begin{bmatrix} 0 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \beta \end{bmatrix}$. By Theorem 6.1.12, a Jordan basis for B is a basis of eigenvectors. Using elementary row operations and gaussian elimination, the reduced row echelon form of B is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. Therefore, $v_1 = (0,1,-1)^T$ is an eigenvector for 0. Using the identity $\alpha^2 - \alpha + 1 = 0$, we find the reduced row echelon form of $B - \alpha$ is $\begin{bmatrix} 1 & 0 & \alpha-1 \\ 0 & 1 & 1-\alpha \\ 0 & 0 & 0 \end{bmatrix}$. Therefore, $v_2 = (1-\alpha, \alpha-1, 1)^T$ is an eigenvector for α. Likewise, $v_3 = (1-\beta, \beta-1, 1)^T$ is an eigenvector for β. Then $V = \{v_1, v_2, v_3\}$ is a Jordan basis for ψ. Let P be the matrix with columns v_1, v_2, v_3. Using a symbolic calculator such as [**28**], for instance, one can show that $P^{-1}BP$ is equal to the matrix with diagonal $(0, \alpha, \beta)$.

EXAMPLE 6.3.15. Consider the matrix $A = \begin{bmatrix} 2 & 3 & 1 \\ -1 & 2 & 1 \\ 4 & -1 & -1 \end{bmatrix}$ over the field $\mathbb{Q}$. Using determinants we compute the characteristic polynomial of A:

$$\begin{aligned} \text{char. poly}(A) &= \det(x - A) \\ &= \begin{vmatrix} x-2 & -3 & -1 \\ 1 & x-2 & -1 \\ -4 & 1 & x+1 \end{vmatrix} \\ &= (x-2)^2(x+1) - 12 - 1 + (x-2) + 3(x+1) - 4(x-2) \\ &= x^2(x-3). \end{aligned}$$

The roots of the characteristic polynomial are 0, and 3. Since the matrix

$$A(A-3) = \begin{bmatrix} -1 & 2 & 1 \\ 3 & -6 & -3 \\ -7 & 14 & 7 \end{bmatrix}$$

has rank 1, we see from Theorem 6.1.10, that the minimal polynomial of A is $\text{min. poly}(A) = x^2(x-3)$. The rational canonical form of A over $\mathbb{Q}$ is therefore

equal to the companion matrix of $x^3 - 3x^2$, which is $C(x^3 - 3x^2) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 3 \end{bmatrix}$. Let $V = \mathbb{Q}^{(3)}$ and $\phi \in \mathrm{Hom}_{\mathbb{Q}}(V, V)$ the linear transformation corresponding to left multiplication by A. Since $\min.\,\mathrm{poly}(\phi)$ has degree 3, we know V is a cyclic $\mathbb{Q}[\phi]$-module. Let $u_1 = (1, 0, 0)^T$, $u_2 = Au_1 = (2, -1, 4)^T$, and $u_3 = Au_2 = (5, 0, 5)^T$. Then $U = \{u_1, u_2, u_3\}$ is a basis for V such that $M(\phi, U, U) = C(x^3 - 3x^2)$. Set $Q = \begin{bmatrix} 1 & 2 & 5 \\ 0 & -1 & 0 \\ 0 & 4 & 5 \end{bmatrix}$. Then we see that $AQ = QC(x^3 - 3x^2)$. The Jordan canonical form of ψ exists over $\mathbb{Q}$. By Theorem 4.6.12, the elementary divisors of ϕ are $x^2, x - 3$. The Jordan canonical form for ϕ is $J(\phi) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. The cyclic submodule of V corresponding to the eigenvalue 0 has dimension 2. The matrix $A - 3 = \begin{bmatrix} -1 & 3 & 1 \\ -1 & -1 & 1 \\ 4 & -1 & -4 \end{bmatrix}$ has rank 2 and $A^2(A - 3) = 0$. Set $w_1 = (1, 1, -4)^T$ and $w_2 = Aw_1 = (1, -3, 7)^T$. Then $A^2 w_1 = 0$ and $Aw_2 = 0$. Set $w_3 = (1, 0, 1)^T$. Then $(A - 3)w_3 = 0$, so w_3 is an eigenvector for 3. Let P be the matrix with columns w_1, w_2, w_3. The reader should verify that P is invertible and $AP = PJ(\phi)$. So w_1, w_2, w_3 is a Jordan basis for ϕ.

EXAMPLE 6.3.16. Let k be a field and $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. The characteristic polynomial of A is $(x - 1)^2 - 1 = x^2 - 2x = x(x - 2)$. If $\mathrm{char}\, k \neq 2$, then A has two distinct eigenvalues, hence the Jordan form of A is diagonal: $J(A) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$. A Jordan basis for A is a basis of eigenvectors, $(1, -1)^T$, $(1, 1)^T$. If $\mathrm{char}\, k = 2$, then 0 is the only eigenvalue of A. The Jordan form of A is therefore $J(A) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ and a Jordan basis for A is $(1, 0)^T$, $(1, 1)^T$.

3.3. Exercises.

EXERCISE 6.3.17. Suppose k is a field and

$$M = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 & -a_0 \\ 1 & 0 & 0 & \dots & 0 & 0 & -a_1 \\ 0 & 1 & 0 & \dots & 0 & 0 & -a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \\ 0 & 0 & 0 & \dots & 0 & 0 & -a_{n-3} \\ 0 & 0 & 0 & \dots & 1 & 0 & -a_{n-2} \\ 0 & 0 & 0 & \dots & 0 & 1 & -a_{n-1} \end{bmatrix}$$

is a matrix in $M_n(k)$.

(1) Prove that $\min.\,\mathrm{poly}_k(M) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$.
(2) Prove that $\mathrm{char}.\,\mathrm{poly}_k(M) = \min.\,\mathrm{poly}_k(M)$.
(3) Prove that the rank of M is equal to the rank of the transpose of M.

EXERCISE 6.3.18. Let R be a commutative ring and A and B similar matrices in $M_n(R)$. Prove that $\mathrm{char}.\,\mathrm{poly}_R(A) = \mathrm{char}.\,\mathrm{poly}_R(B)$.

EXERCISE 6.3.19. Let R be a commutative ring, $A \in M_m(R)$, $B \in M_n(R)$. Define the *direct sum of A and B* by

$$A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

which is a matrix in $M_{m+n}(R)$. The direct sum $A \oplus B$ is sometimes called a *block diagonal matrix* and is denoted $\text{diag}(A, B)$. Prove:

(1) $\det(A \oplus B) = \det(A)\det(B)$.
(2) $\text{char. poly}_R(A \oplus B) = \text{char. poly}_R(A)\,\text{char. poly}_R(B)$.
(3) $\text{Rank}(A \oplus B) = \text{Rank}(A) + \text{Rank}(B)$.

EXERCISE 6.3.20. (Cramer's Rule) Let R be a commutative ring. Suppose $A \in M_n(R)$, $x, b \in R^n$ such that $Ax = b$. Prove that $x_i \det(A) = \det(B_i)$, where $B_i = (a_1, \ldots, b, \ldots, a_n)$ is the matrix obtained by replacing column i of A with the column vector b.

EXERCISE 6.3.21. Let $\theta : R \to S$ be a homomorphism of commutative rings.

(1) Show that θ induces a homomorphism of rings $\theta : M_n(R) \to M_n(S)$.
(2) Show that $\theta(\det(M)) = \det(\theta(M))$, for every M in $M_n(R)$.
(3) We know from Theorem 3.6.2 that θ induces a homomorphism of rings $R[x] \to S[x]$. Show that $\theta(\text{char. poly}_R(M)) = \text{char. poly}_S(\theta(M))$.

EXERCISE 6.3.22. Let $A = \begin{bmatrix} 0 & 1 & 1 \\ -4 & -4 & -1 \\ 0 & 0 & -2 \end{bmatrix}$ in the ring of 3-by-3 matrices over the field $\mathbb{Q}$.

(1) Find $\text{char. poly}(A)$, the characteristic polynomial.
(2) Find $\text{min. poly}(A)$, the minimal polynomial.
(3) Find the invariant factors of A in $\mathbb{Q}[x]$.
(4) Find the elementary divisors of A in $\mathbb{Q}[x]$.
(5) Find the rational canonical form of A.
(6) Find the Jordan canonical form of A.
(7) Find an invertible matrix P such that $P^{-1}AP$ is equal to the Jordan canonical form of A. In other words, find a Jordan basis for the linear transformation on $\mathbb{Q}^{(3)}$ defined by A.

EXERCISE 6.3.23. Let R be a commutative ring and $A \in M_n(R)$. For each i, let A_i denote column i. Assume $1 \leq i < j \leq m$ and $\alpha \in R$. If B is the matrix obtained by replacing A_j with $\alpha A_i + A_j$, show that $\det(B) = \det(A)$.

EXERCISE 6.3.24. This exercise is a generalization of Example 6.3.16. Let k be a field and $A = (a_{ij})$ the n-by-n matrix in $M_n(k)$ with $a_{ij} = 1$ for every pair (i, j).

(1) Assume the characteristic of k does not divide n. Prove the following:
   (a) $\text{min. poly}_k(A) = x(x - n)$.
   (b) $\text{char. poly}_k(A) = \pm x^{n-1}(n - x)$.

(c) The set

$$v_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, v_{n-1} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, v_n = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}$$

is a Jordan basis for A.

(2) Assume the characteristic of k divides n. Prove the following:
   (a) $\text{min. poly}_k(A) = x^2$.
   (b) $\text{char. poly}_k(A) = \pm x^n$.
   (c) The set $v_1, v_2, \ldots, v_{n-2}, v_{n-1} = (0, 0, \ldots, 0, 1)^T$, v_n is a Jordan basis for A, where $v_1, \ldots, v_{n-2}$ and v_n are the vectors from Part (1) (c).

EXERCISE 6.3.25. Let R be an integral domain and M a finitely generated R-module. Let $\phi \in \text{Hom}_R(M, M)$. Show that there exists a monic polynomial $p(x) \in R[x]$ such that $p(\phi) = 0$.

EXERCISE 6.3.26. Let R be a commutative ring and $n \geq 1$. Define the *trace* of a matrix $\alpha = (\alpha_{ij}) \in M_n(R)$ by $\text{trace}(\alpha) = \sum_{i=1}^{n} \alpha_{ii}$.

(1) Prove that the trace mapping is an R-module homomorphism from $M_n(R)$ to R.
(2) Prove that $\text{trace}(\alpha\beta) = \text{trace}(\beta\alpha)$.
(3) Prove that if α and β are similar, then $\text{trace}(\alpha) = \text{trace}(\beta)$.

EXERCISE 6.3.27. Let R be a commutative ring, M a finitely generated free R-module, and X a basis for M over R. Define the trace of $\phi \in \text{Hom}_R(M, M)$ to be $\text{trace}(\phi) = \text{trace}(M(\phi, X))$. Show that this definition is independent of the choice for X. Show that the trace mapping is an R-module homomorphism from $\text{Hom}_R(M, M)$ to R.

EXERCISE 6.3.28. Let k be a field, $n \geq 1$, $f = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in k[x]$ and $M = C(f)$ the companion matrix of f. Prove the following.

(1) $\det(M) = (-1)^n a_0$.
(2) $\text{trace}(M) = -a_{n-1}$.

EXERCISE 6.3.29. Let R be a commutative ring and M a finitely generated free R-module of rank n. Let $\phi \in \text{Hom}_R(M, M)$. Show that if $\text{char. poly}_R(\phi) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$, then $\text{trace}(\phi) = -a_{n-1}$ and $\det(\phi) = (-1)^n a_0$.

EXERCISE 6.3.30. Let k be a field, V a finitely generated vector space over k, and $\phi \in \text{Hom}_k(V, V)$. Suppose $q = \text{min. poly}_k(\phi) = x^m + a_{m-1}x^{m-1} + \cdots + a_0$ is irreducible in $k[x]$. Prove the following.

(1) $\text{char. poly}_k(\phi) = q^r$ for some integer r.
(2) $\det(\phi) = (-1)^{mr} a_0^r$.
(3) $\text{trace}(\phi) = -r a_{m-1}$.

EXERCISE 6.3.31. Let k be a field and A a matrix in $M_n(k)$ such that $\text{Rank}(A) = r < n$. Prove:

(1) $\det(A) = 0$.

(2) If B is an $r+1$-by-$r+1$ submatrix of A, then $\det(B) = 0$.
(3) A contains an r-by-r submatrix of rank r.

EXERCISE 6.3.32. Let k be a field and f an irreducible polynomial with coefficients in k. Show that if M is an n-by-n matrix over k such that $f(M) = 0$, then $\deg(f) \leq n$.

EXERCISE 6.3.33. Let R be a commutative ring and $n \geq 1$. If $A \in M_n(R)$, show that the trace of A (see Exercise 6.3.26) satisfies:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} e_{ij}Ae_{ji} = \text{trace}(A)I_n$$

where e_{ij} denotes the elementary matrix and $I_n = e_{11} + \cdots + e_{nn}$ the identity matrix (see Section 1.5).

EXERCISE 6.3.34. Let R be a commutative ring and $A = M_n(R)$ the ring of n-by-n matrices over R. The so-called *trace pairing* $\tau : A \times A \to R$ is defined by $\tau(\alpha, \beta) = \text{trace}(\alpha\beta)$, where the trace map is defined in Exercise 6.3.26. Show that τ satisfies these properties:

(1) $\tau(\alpha, \beta) = \tau(\beta, \alpha)$.
(2) $\tau(a_1\alpha_1 + a_2\alpha_2, \beta) = a_1\tau(\alpha_1, \beta) + a_2\tau(\alpha_2, \beta)$ for $a_1, a_2 \in R$.
(3) $\tau(\alpha, b_1\beta_1 + b_2\beta_2) = b_1\tau(\alpha, \beta_1) + b_2\tau(\alpha, \beta_2)$ for $b_1, b_2 \in R$.
(4) If $\alpha \neq 0$ is fixed, then $\tau(\alpha, \ ) : A \to R$ is nonzero. That is, there exists β such that $\tau(\alpha, \beta) \neq 0$.

We say that τ is a *symmetric nondegenerate bilinear form.*

EXERCISE 6.3.35. Let $n \geq 2$. As in Section 1.5, if $\sigma \in S_n$, then P_σ denotes the n-by-n permutation matrix associated to σ. Show that $\det P_\sigma = \text{sign}(\sigma)$.

EXERCISE 6.3.36. As in Exercise 6.3.35, let P_σ be the n-by-n permutation matrix in $M_n(\mathbb{Q})$ associated to $\sigma \in S_n$.

(1) If σ is the n-cycle $(1, 2, 3, \ldots, n)$, show that $\text{char. poly}(P_\sigma)$ is equal to $x^n - 1$.
(2) If σ is an arbitrary n-cycle in S_n, show that $\text{char. poly}(P_\sigma)$ is equal to $x^n - 1$.
(3) If $\sigma = \sigma_1\sigma_2\cdots\sigma_k$ is a product of disjoint cycles where $|\sigma_i| = s_i$ and $n = s_1 + s_2 + \cdots + s_k$, show that $\text{char. poly}(P_\sigma)$ is equal to $\prod_{i=1}^{k}(x^{s_i} - 1)$.

EXERCISE 6.3.37. As in Exercise 6.3.36, let P_σ be the n-by-n permutation matrix in $M_n(\mathbb{Q})$ associated to $\sigma \in S_n$. If $\sigma = \sigma_1\sigma_2\cdots\sigma_k$ is a product of disjoint cycles where $|\sigma_i| = s_i$ and $n = s_1 + s_2 + \cdots + s_k$, show that $\mu(x) = \text{min. poly}(P_\sigma)$ is equal to the square free part of $\chi(x) = \text{char. poly}(P_\sigma)$. That is, if $\delta(x) = \gcd(\chi, \chi')$, then $\mu(x) = \chi(x)/\delta(x)$.

EXERCISE 6.3.38. Let R be a commutative ring and A a square matrix in $M_n(R)$. Show that the following are equivalent: (1) A is invertible. (2) A is left invertible. (3) A is right invertible. (For a generalization see [**11**, Exercise 2.4.19].)

4. The Normal Basis Theorem

Let F/k be a Galois extension of fields with a cyclic Galois group $\text{Aut}_k(F) = \langle\sigma\rangle$. Let $n = \dim_k(F)$. By Lemma 5.3.6, we can view σ as an element of

$\mathrm{Hom}_k(F,F)$. As in Section 6.2, the k-linear transformation σ turns F into a $k[\sigma]$-module. In Theorem 6.4.1, we apply results from Sections 6.2 and 6.3.2 to show that F is a cyclic $k[\sigma]$-module. This implies that F has a so-called normal basis.

THEOREM 6.4.1. *(The Normal Basis Theorem) Let F/k be a Galois extension of degree n with finite cyclic group $\mathrm{Aut}_k(F) = \langle\sigma\rangle$. Then there exists $\alpha \in F$ such that the set $\{\alpha, \sigma(\alpha), \sigma^2(\alpha), \ldots, \sigma^{n-1}(\alpha)\}$ is a basis for F as a k-vector space. We call the basis $\{\alpha, \sigma(\alpha), \sigma^2(\alpha), \ldots, \sigma^{n-1}(\alpha)\}$ a* normal basis for F/k.

PROOF. We have $\dim_k(F) = n$. View $1, \sigma, \sigma^2, \ldots, \sigma^{n-1}$ as elements of the ring of endomorphisms $\mathrm{Hom}_k(F,F)$. Then $\mathrm{char.poly}_k(\sigma)$ has degree n (see Definition 6.3.12). Since $\mathrm{Aut}_k(F) = \langle\sigma\rangle$ has order n, the minimal polynomial of σ divides $x^n - 1$. By Theorem 5.3.7, the automorphisms $1, \sigma, \sigma^2, \ldots, \sigma^{n-1}$ are linearly independent over k, so the degree of $\mathrm{min.poly}_k(\sigma)$ is at least n. Therefore, $\mathrm{min.poly}_k(\sigma) = x^n - 1$. Since the minimal polynomial and the characteristic polynomial of σ both have degree n, this implies they are equal. By Theorem 6.3.13, F is a cyclic $k[\sigma]$-module. By Theorem 6.2.1, there exists $\alpha \in F$ such that the set $\{\alpha, \sigma(\alpha), \sigma^2(\alpha), \ldots, \sigma^{n-1}(\alpha)\}$ is a k-basis for F. □

CHAPTER 7

Ideal Class Groups

The goal of this chapter is to construct the group of ideal classes of a special type of commutative ring. We focus our attention on the class of all integrally closed noetherian integral domains S such that every nonzero prime ideal of S is maximal. These rings are commonly called Dedekind domains. An algebraic number field is a finite algebraic extension L of $\mathbb{Q}$. The integral closure of $\mathbb{Z}$ in an algebraic number field L is a Dedekind domain. Another important class of examples are the affine coordinate rings of nonsingular algebraic curves.

There is a brief introduction to the notion of integral extensions of commutative rings presented in Section 7.1. Included is a proof that the integral closure $\bar{R}$ of a principal ideal domain R in a finite separable extension L of its quotient field K is finitely generated and free as an R-module and the rank $\text{Rank}_R(\bar{R})$ is equal to the dimension $\dim_K(L)$. Fractional ideals of an integral domain in its field of fractions are defined in Section 7.2. In Section 7.3 we restrict our attention to Dedekind domains. An integral domain S with field of fractions L is called a Dedekind domain if $S \neq L$, S is integrally closed in L, S satisfies the ascending chain condition on ideals, and every nonzero prime ideal in S is maximal. Let S be the integral closure of a principal ideal domain R in a finite separable extension L/K of the quotient field K of R. We show that the ring S is a Dedekind domain, a very useful existence theorem for constructing a Dedekind domain as an extension of a principal ideal domain. We show that for a Dedekind domain S, every fractional ideal is invertible, the set of all fractional ideals is a free $\mathbb{Z}$-module and the set of nonzero prime ideals is a free basis. The ideal class group of S is defined, and a special case of Nagata's Theorem is proved.

Section 7.4 is a short introduction to the theory of Algebraic Curves over a field. Two elementary examples are studied. The first is an irreducible affine conic, the second a nonsingular affine elliptic curve. These are examples of rings which are not unique factorization domains. As an application of Nagata's Theorem, we show that the class group of the ring $k[x,y]/(x^2+y^2-1)$ is a cyclic group of order 2. Using the ideal class group of the affine coordinate ring $k[x,y]/(y^2-x(x^2-1))$, we describe the group law on the set of rational points of the affine elliptic curve defined by $y^2 = x(x^2-1)$. Our treatment of the group law on the elliptic curve is not self-contained. For the proof that the cubic curve is integrally closed, we refer the reader to a treatment of nonsingular curves. For the proof that two distinct prime divisors represent different ideal classes, the reader is referred to a book on algebraic curves.

1. Integral Extensions

Let A be a ring with center $Z(A)$ and assume R is a subring of $Z(A)$. Since $Z(A)$ is a commutative ring, R is a commutative ring. Since every $r \in R$ is central,

we have $rx = xr$ for all $x \in A$. In the terminology of Definition 4.4.1, A is an R-algebra. If $\alpha \in A$, then by $R[\alpha]$ we denote the subring of A generated by R and α. As in Theorem 4.4.8, the evaluation homomorphism $\tau : R[x] \to A$ defined by $x \mapsto \alpha$ is an R-algebra homomorphism. By Exercise 3.6.36, the image of τ is $R[\alpha]$, and $R[\alpha]$ is a commutative ring. If there is a monic polynomial $p(x)$ in the kernel of τ, then we say α is *integral over* R. In this case, $p(\alpha) = 0$. Note that if R is a field, then α is algebraic over R if and only if α is integral over R. Let $\bar{R}$ be the set of all $\alpha \in A$ such that α is integral over R. Then $\bar{R}$ is called the *integral closure of* R *in* A. Notice that every element of R is integral over R. In fact, if $r \in R$, then r is a root of the monic polynomial $x - r$. Hence we have $R \subseteq \bar{R} \subseteq A$. If $\bar{R} = A$, then every element of A is integral over R and we say A is an *integral extension of* R. If $\bar{R} = R$, then no element of $A - R$ is integral over R and in this case we say R is *integrally closed in* A.

PROPOSITION 7.1.1. *If R is a unique factorization domain with quotient field K, then R is integrally closed in K.*

PROOF. By Exercise 3.7.20, if $p(x)$ is a monic polynomial in $R[x]$, $u \in K$, and u is a root of $p(x)$, then u is in R. □

EXAMPLE 7.1.2. The following examples follow immediately from of Proposition 7.1.1.

(1) The ring of integers $\mathbb{Z}$ is integrally closed in $\mathbb{Q}$.
(2) If k is a field, then the ring of polynomials $k[x]$ is integrally closed in $k(x)$.
(3) If R is a UFD, then by Theorem 3.7.5, the polynomial ring $R[x_1, \ldots, x_n]$ is a UFD and is integrally closed in its quotient field.

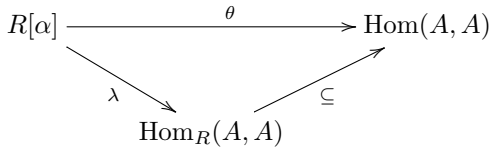
EXAMPLE 7.1.3. Let D be a square free integer such that $D \equiv 1 \pmod 4$. If $u = \sqrt{D}$, then we saw in Example 3.7.9 that the ring $S = \mathbb{Z}[u]$ is an integral domain, the quotient field of S is $L = \mathbb{Q}[u]$, and $\alpha = (1 + u)/2$ is an element of $L - S$ that is integral over S. Hence S is not integrally closed in L. Question: Is $\mathbb{Z}[\alpha]$ integrally closed in L?

EXAMPLE 7.1.4. Let R be an integral domain with field of fractions K. Let $M_n(R)$ denote the ring of n-by-n matrices over R. Viewing R as a subring of K, we can therefore view $M_n(R)$ as a subring of $M_n(K)$. Let $\alpha \in M_n(R)$ be an n-by-n matrix over R. By Definition 6.3.12 we see that the characteristic polynomial $p(x) = \text{char. poly}_R(\alpha)$ is the determinant of the matrix $xI_n - \alpha$. Therefore, $p(x)$ is a monic polynomial of degree n with coefficients in R. This implies $\text{char. poly}_R(\alpha) = \text{char. poly}_K(\alpha)$. The Cayley-Hamilton Theorem (Theorem 6.3.13 (2)) implies that $p(\alpha) = 0$. Hence, α is both algebraic over K and integral over R. This proves $M_n(R)$ is an integral extension of R.

EXAMPLE 7.1.5. Let R be an integral domain and M a finitely generated R-module. Let $\phi \in \text{Hom}_R(M, M)$. By Exercise 6.3.25, there exists a monic polynomial $p(x) \in R[x]$ such that $p(\phi) = 0$. If we assume M is a faithful R-module, the left regular representation map $\theta : R \to \text{Hom}_R(M, M)$ is one-to-one. Hence, $\text{Hom}_R(M, M)$ is an integral extension of R.

PROPOSITION 7.1.6. *Let A be a ring and R a subring of the center of A. If R is an integral domain and A is finitely generated as an R-module, then A is an integral extension of R.*

PROOF. Let $\alpha \in A$. Then $R[\alpha]$ is a commutative subring of A, and as in Example 4.1.4 (4), A is a left $R[\alpha]$-module. Let $\theta : R[\alpha] \to \mathrm{Hom}(A, A)$ be the left regular representation of Lemma 4.1.2. Since $R[\alpha]$ is a subring of A, as a left $R[\alpha]$-module, A is faithful. Hence θ is one-to-one. As in Example 4.1.17, the image of θ is a subring of $\mathrm{Hom}_R(A, A)$ and the diagram of ring homomorphisms



commutes. Therefore, λ is one-to-one. By Example 7.1.5, every $\phi \in \mathrm{Hom}_R(A, A)$ is integral over R. Therefore, every element of $R[\alpha]$ is integral over R. □

EXAMPLE 7.1.7. Let k be a field and $R = k[x^2, x^3]$ the subring of $k[x]$ consisting of all polynomials such that the coefficient of x is zero (see Exercise 3.6.21). Then R is an integral domain and the quotient field of R is equal to $k(x)$. Since x^2 is an element of R, this means x is integral over R. Thus, R is not integrally closed in $k(x)$. A typical polynomial f(x) in $k[x]$ can be written in the form $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_dx^d$. If we set $g(x) = f(x) - a_1x$, then $g(x)$ is in R. This shows $f(x) = a_1x + g(x)$, which is in the R-submodule of $k[x]$ generated by 1 and x. Thus, $k[x]$ is a finitely generated R-module. By Proposition 7.1.6, $k[x]$ is an integral extension of R. By Proposition 7.1.1, $k[x]$ is integrally closed in $k(x)$. This proves that $k[x]$ is equal to the integral closure of R in $k(x)$.

EXAMPLE 7.1.8. Let k be a field and $R = k[x^2, x + x^3]$ the ring of Exercise 3.7.21. Then R is a subring of the principal ideal domain $k[x]$, R is an integral domain, and the quotient field of R is equal to $k(x)$. Since x^2 is an element of R, this means x is integral over R. Thus, R is not integrally closed in $k(x)$. As in Example 7.1.7, the reader should verify that $k[x]$ is generated as an R-module by 1 and x, and the integral closure of R in $k(x)$ is equal to $k[x]$.

PROPOSITION 7.1.9. *Let A be a ring and R a subring of $Z(A)$. For $\alpha \in A$, the following are equivalent.*

(1) α is integral over R.
(2) $R[\alpha]$ is a finitely generated R-module.

PROOF. (1) implies (2): Assume α is integral over R. Then there exists a monic polynomial $p(x) = x^n + r_{n-1}x^{n-1} + \cdots + r_1x + r_0$ in $R[x]$ such that $n \geq 1$ and $p(\alpha) = 0$. Let $B = R + R\alpha + \cdots + R\alpha^{n-1}$ be the R-submodule of A generated by the set $\{1, \alpha, \ldots, \alpha^{n-1}\}$. Then B is a finitely generated R-submodule of A and $B \subseteq R[\alpha]$. To finish the proof, we show $R[\alpha] \subseteq B$. It suffices to show $\alpha^m \in B$ for all $m \geq n$. Since $p(\alpha) = 0$, we have $\alpha^n = -(r_{n-1}\alpha^{n-1} + \cdots + r_1\alpha + r_0)$ is in B. Since $\alpha^{n+k} = -(r_{n-1}\alpha^{n+k-1} + \cdots + r_1\alpha^{k+1} + r_0\alpha^k)$, a routine induction argument shows $\alpha^{n+k} \in B$ for $k \geq 0$.

(2) implies (1): Suppose there exist $u_1, \ldots, u_n$ in $R[\alpha]$ such that $R[\alpha] = Ru_1 + \cdots + Ru_n$. Since $R[x] \to R[\alpha]$ is onto, there exist $p_1, \ldots, p_n$ in $R[x]$ such that $u_i = p_i(\alpha)$ for $1 \leq i \leq n$. Pick $N \in \mathbb{N}$ such that N is greater than $\deg p_i$ for each i. Since $\alpha^N \in Ru_1 + \cdots + Ru_n$, there exist $r_1, \ldots, r_n$ in R such that $\alpha^N = r_1p_1(\alpha) + \cdots + r_np_n(\alpha)$. This proves α is integral over R. □

THEOREM 7.1.10. *Let A/R be an extension of commutative rings. If R is an integral domain, then the following are true.*

(1) If $a_1, \ldots, a_n \in A$ are integral over R, then $R[a_1, \ldots, a_n]$ is a finitely generated R-module.
(2) If $\bar{R}$ is the integral closure of R in A, then $\bar{R}$ is an R-subalgebra of A.
(3) (Integral over Integral is Integral) Let $R \subseteq S \subseteq A$ be three rings such that A is integral over S and S is integral over R. Then A is integral over R.
(4) Let $\bar{R}$ be the integral closure of R in A. Then $\bar{R}$ is integrally closed in A.

PROOF. (1): By Proposition 7.1.9, $R[a_1]$ is a finitely generated R-module. Since A is commutative, $R[a_1, a_2] = R[a_1][a_2]$. Since a_2 is integral over R, a_2 is integral over $R[a_1]$. By Proposition 7.1.9, $R[a_1, a_2]$ is a finitely generated $R[a_1]$-module. By Exercise 4.1.24, $R[a_1, a_2]$ is a finitely generated R-module. A routine induction argument proves Part (1).

(2): We have $R \subseteq \bar{R} \subseteq A$. Since R is a subring of A, it suffices to show $\bar{R}$ is a ring. For this, it suffices to show addition and multiplication are binary operations on $\bar{R}$. Let x and y be arbitrary elements of $\bar{R}$. We show $x + y$ and xy are in $\bar{R}$. By Part (1), $R[x, y]$ is a finitely generated R-module. By Proposition 7.1.6, $R[x, y]$ is an integral extension of R. Therefore, $x + y$ and xy are integral over R and belong to $\bar{R}$.

(3): Let $a \in A$ and $p \in S[x]$ a monic polynomial such that $p(a) = 0$. Suppose $p = s_0 + s_1 x + \cdots + s_{n-1}x^{n-1} + x^n$. Set $T = R[s_0, \ldots, s_{n-1}]$. By Part (1), T is a finitely generated R-module. Since $p \in T[x]$, it follows that a is integral over T. By Proposition 7.1.9, that $T[a]$ is finitely generated over T. By Exercise 4.1.24, $T[a] = R[s_0, \ldots, s_{n-1}, a]$ is a finitely generated R-module. Proposition 7.1.6 implies a is integral over R.

(4): By the proof of Part (3), if $a \in A$ is integral over $\bar{R}$, then a is integral over R. □

Theorem 7.1.11 is a version of Theorem 3.7.4 for an integral domain that is integrally closed in its quotient field.

THEOREM 7.1.11. *(Gauss' Lemma) Let R be an integral domain which is integrally closed in its quotient field K. Let $f \in R[x]$ be a monic polynomial. Then f is irreducible in $R[x]$ if and only if f is irreducible in $K[x]$.*

PROOF. If f is reducible in $R[x]$, then f is reducible in $K[x]$. Suppose there is a nontrivial factorization $f = gh$ in $K[x]$. Since f is monic, we factor out the leading coefficients and assume g and h are both monic polynomials in $K[x]$. By Corollary 5.2.6, let L/K be an extension of fields such that L is a splitting field for f over K. By Theorem 7.1.10, let S be the integral closure of R in L. Since f splits in $L[x]$, so does g. Let $g = (x - \alpha_1) \cdots (x - \alpha_n)$ be the factorization of g in $L[x]$. Each α_i is a root of f, hence is integral over R, hence lies in S. Thus $g = (x - \alpha_1) \cdots (x - \alpha_n)$ is in the ring $S[x]$. So each coefficient of g is in $S \cap K$ which is equal to R since R is integrally closed in K. So $g \in R[x]$. The same argument applies to h. □

COROLLARY 7.1.12. *Let R be an integral domain which is integrally closed in its quotient field K. Let A be a K-algebra and α an element of A which is algebraic over K. Then α is integral over R if and only if* $\min.\,\mathrm{poly}_K(\alpha)$ *is in $R[x]$.*

PROOF. Suppose α is integral over R. Then there exists a monic polynomial $p(x)$ in $R[x]$ such that $p(\alpha) = 0$. Let $f = \min.\,\mathrm{poly}_K(\alpha) \in K[x]$. By Theorem 4.4.8 (5), there exists $g \in K[x]$ such that $p = fg$. By Theorem 7.1.11, this implies $f \in R[x]$. □

COROLLARY 7.1.13. *Let R be an integral domain which is integrally closed in its quotient field K. Let L/K be an extension of fields and $\bar{R}$ the integral closure of R in L. If I is a proper ideal in $\bar{R}$, then $I \cap R \neq (0)$.*

PROOF. Let $\alpha \in I - (0)$. Let $f(x) = \min.\,\mathrm{poly}_K(\alpha)$. By Corollary 7.1.12, $f \in R[x]$. There exist $n \geq 1$ and $r_i \in R$ such that $f = x^n + r_{n-1}x^{n-1} + \cdots + r_1 x + r_0$. Since α is invertible in the field L, by Corollary 4.4.14, $r_0 \neq 0$. Then $r_0 = -(\alpha^n + r_{n-1}\alpha^{n-1} + \cdots + r_1\alpha)$ is a nonzero element of $I \cap R$. □

PROPOSITION 7.1.14. *Let R be an integral domain which is integrally closed in its quotient field K. Let L/K be a Galois extension with finite group G. Let S be the integral closure of R in L. For each $\sigma \in G$, $\sigma : S \to S$. That is, σ restricts to an automorphism of S and G acts as a group of automorphisms of S. We have $S^G = R$, $T_K^L : S \to R$, and $N_K^L : S \to R$.*

PROOF. Let $\alpha \in S$ and $f = \min.\,\mathrm{poly}_K(\alpha)$. By Corollary 7.1.12, $f \in R[x]$. If $\sigma \in G$, then $f(\sigma(\alpha)) = 0$, by Proposition 5.3.2. Therefore, $\sigma(\alpha) \in S$. If $\alpha \in S^G$, then α is in $S \cap K$, which is equal to R since R is integrally closed in K. As in Section 5.5.1, the trace and the norm are functions from S to R. □

THEOREM 7.1.15. *Let R be an integral domain which is integrally closed in its quotient field K. Let L/K be a Galois extension with finite group G and let S be the integral closure of R in L. There exist bases $\{\lambda_1, \ldots, \lambda_n\}$ and $\{\mu_1, \ldots, \mu_n\}$ for L/K such that $R\lambda_1 + \cdots + R\lambda_n \subseteq S \subseteq R\mu_1 + \cdots + R\mu_n$.*

PROOF. Our proof is based on [**5**, Theorem 5.17]. Let $n = \dim_K(L)$. Given $\lambda \in L$, by Exercise 7.1.20, there is a nonzero $r \in R$ such that $r\lambda \in S$. Let $\lambda_1, \ldots, \lambda_n$ be a basis for L as a K-vector space. Without loss of generality, we can assume each λ_i is in S. By Proposition 5.5.2, there is a K-basis $\mu_1, \ldots, \mu_n$ for L such that $T_K^L(\lambda_i\mu_j) = \delta_{ij}$ (the Kronecker delta function). Let s be an arbitrary element of S. View s as an element of L and write $s = \alpha_1\mu_1 + \cdots + \alpha_n\mu_n$, where each $\alpha_i \in K$. By Proposition 7.1.14, $T_K^L : S \to R$. For each i, $s\lambda_i \in S$. Then

$$T_K^L(s\lambda_i) = T_K^L\Big(\sum_{j=1}^{n} \alpha_j\mu_j\lambda_i\Big) = \sum_{j=1}^{n} T_K^L\left(\alpha_j\lambda_i\mu_j\right) = \sum_{j=1}^{n} \alpha_j T_K^L\left(\lambda_i\mu_j\right) = \alpha_i$$

shows that each α_i is in R. It follows that $S \subseteq R\mu_1 + \cdots + R\mu_n$. □

COROLLARY 7.1.16. *Let R be a principal ideal domain with quotient field K. If L/K is a finite dimensional separable extension of fields and $\bar{R}$ is the integral closure of R in L, then $\bar{R}$ is a finitely generated free R-module of rank $\dim_K(L)$.*

PROOF. By Corollary 5.3.20 there exists a Galois extension F/K which contains L/K as an intermediate field. Let S be the integral closure of R in F. By Theorem 7.1.15, there is a K-basis $\mu_1, \ldots, \mu_n$ for F such that $S \subseteq R\mu_1 + \cdots + R\mu_n$. The set $\mu_1, \ldots, \mu_n$ is linearly independent over R, because it is linearly independent over K. The R-module $R\mu_1 + \cdots + R\mu_n$ is free of rank n. Since $\bar{R} \subseteq S$ this means

$\bar{R}$ and S are both submodules of a finitely generated free R-module. By Theorem 4.6.1, $\bar{R}$ and S are both finitely generated free R-modules. By Exercise 7.1.20, L is the quotient field of $\bar{R}$. By Exercise 7.1.23, the rank of $\bar{R}$ over R is equal to $m = \dim_K(L)$. □

REMARK 7.1.17. Most of the results of this section are true under more general hypotheses. For example, Proposition 7.1.6 and Theorem 7.1.10 hold without the assumption that R is an integral domain. A version of Theorem 7.1.15 is true when L/K is finite dimensional and separable but not necessarily Galois. In Theorem 7.1.15, if R is a noetherian integral domain which is integrally closed in its quotient field, then S is a finitely generated R-module. For these and more results, see for example, [**9**, Chapter 10].

EXAMPLE 7.1.18. Let $n > 2$ be an integer and ζ a primitive nth root of unity in $\mathbb{C}$. By Corollary 5.5.9, the irreducible polynomial for ζ is the nth cyclotomic polynomial $\Phi_n(x)$ which has degree $\phi(n)$. The cyclotomic extension of order n over $\mathbb{Q}$ is $\mathbb{Q}[\zeta] \cong \mathbb{Q}[x]/(\Phi_n)$. The Galois group of Φ_n is isomorphic to the group of units in the ring $\mathbb{Z}/n$. Although we do not prove it here, the integral closure of $\mathbb{Z}$ in $\mathbb{Q}[\zeta]$ is the ring $\mathbb{Z}[\zeta]$. By Corollary 7.1.16, as a $\mathbb{Z}$-module, $\mathbb{Z}[\zeta]$ is free with basis $\{\zeta^j \mid j = 1, \ldots, \phi(n) - 1\}$. The interested reader is referred to a book on Algebraic Number Theory, for example, [**22**, Section 13.4].

LEMMA 7.1.19. *Let S/R be an integral extension of commutative rings. Let P be a prime ideal in S. Then P is a maximal ideal of S if and only if $P \cap R$ is a maximal ideal of R.*

PROOF. By Proposition 3.2.25, $P \cap R$ is a prime ideal in R. Both rings S/P and $R/(P \cap R)$ are integral domains. By Exercise 7.1.21, S/P is an integral extension of $R/(P \cap R)$. If $P \cap R$ is a maximal ideal, then $R/(P \cap R)$ is a field and by Theorem 4.4.15, S/P is a field. Therefore, P is a maximal ideal in S. For the converse, assume S/P is a field. Then S/P is integral over $R/(P \cap R)$ and by Exercise 7.1.27, every nonzero element of $R/(P \cap R)$ is invertible. □

1.1. Exercises.

EXERCISE 7.1.20. Let R be an integral domain with quotient field K. Let L/K be an extension of fields. If $\lambda \in L$ is algebraic over K, then there exists $r \in R - (0)$ such that $r\lambda$ is integral over R. If L/K is an algebraic extension and S is the integral closure of R in L, then L is equal to the quotient field of S.

EXERCISE 7.1.21. Let A be a ring and R a subring of the center of A. If A is integral over R and I is a two-sided ideal in A, prove that A/I is an integral $R/(I \cap R)$-algebra.

EXERCISE 7.1.22. Let R be a commutative ring and $A = R[x]$ the polynomial ring in one variable over R. Show that R is integrally closed in A if and only if $\mathrm{Rad}_R(0) = (0)$.

EXERCISE 7.1.23. Let S/R be an extension of commutative rings. Assume R and S are both integral domains. Denote by K the quotient field of R and by L the quotient field of S.

(1) Using Exercise 3.5.2, show that L can be viewed as a field extension of K.
(2) Prove that if S is integral over R, then L is algebraic over K.

(3) If S is a free R-module of finite rank n, prove that $\dim_K(L) = n$.

EXERCISE 7.1.24. Let k be a field, x an indeterminate, and $n > 1$ an integer. For the extension of fields $F = k(x^n) \subseteq K = k(x)$, prove the following.

(1) $y^n - x$ is an irreducible polynomial in $K[y]$.
(2) $y^n - x^n$ is an irreducible polynomial in $F[y]$.
(3) $\dim_F(K) = n$.
(4) $\text{Irr. poly}_F(x^{n+1})$ has degree n.
(5) $y^n - x^{n+1}$ is an irreducible polynomial in $K[y]$.

EXERCISE 7.1.25. Let k be a field, x an indeterminate, and $n > 1$ an integer. Let $T = k[x]$, $S = k[x^n, x^{n+1}]$, and $R = k[x^n]$. For the tower of subrings $R \subseteq S \subseteq T$, prove:

(1) T is free over R of rank n.
(2) S is free over R of rank n.
(3) T is not free over S.

For a continuation of this example, see Exercise 7.1.26.

EXERCISE 7.1.26. This exercise is a generalization of Example 7.1.7. Let k be a field, x an indeterminate, and $n > 1$ an integer. In the notation of Exercise 7.1.25, let $T = k[x]$ and $S = k[x^n, x^{n+1}]$. Let $K = k(x)$. Prove that T is equal to the integral closure of S in K.

EXERCISE 7.1.27. Let A be a ring, R a subring of the center of A. If A/R is an integral extension and $\alpha \in R - (0)$, prove that α is invertible in A if and only if α is invertible in R.

2. Fractional Ideals

LEMMA 7.2.1. *Let R be an integral domain with field of fractions K. If F is a nonzero R-submodule of K, then the following are equivalent.*

(1) There exists a finitely generated R-submodule N such that $F \subseteq N \subseteq K$.
(2) There are nonzero elements a, b in K such that $aR \subseteq F \subseteq bR$.
(3) There exists a nonzero c in R such that $cF \subseteq R$.
(4) There exists a nonzero d in K such that $dF \subseteq R$.

PROOF. (1) implies (3): Write $N = Rx_1 + \cdots + Rx_n$ where $x_1, \ldots, x_n$ are elements of K. If c is the product of the denominators of $x_1, \ldots, x_n$, then for each i we have $cx_i \in R$. Therefore $cF \subseteq cN \subseteq Rcx_1 + \cdots + Rcx_n \subseteq R$.

(3) implies (4): Is trivial.

(4) implies (2): Suppose $dF \subseteq R$ and $d \in K - (0)$. If $b = d^{-1}$ and $a \in F - (0)$, then we have $aR \subseteq F = bdF \subseteq bR$.

(2) implies (1): Take $N = bR$. □

DEFINITION 7.2.2. Let R be an integral domain with field of fractions K. A *fractional ideal* of R is a nonzero R-submodule F of K satisfying any of the four equivalent conditions of Lemma 7.2.1. If E and F are two fractional ideals of R, then we define the sum and product as in Definition 3.2.4. The sum, denoted $E + F$ is the R-submodule of K generated by E and F. The product, EF, is the R-submodule generated by the set $\{xy \mid x \in E \text{ and } y \in F\}$. The *colon ideal* or *ideal quotient* of E over F is $E : F = \{x \in K \mid xF \subseteq E$. We sometimes write F^{-1} for $R : F$.

LEMMA 7.2.3. *Let R be an integral domain with quotient field K. If E and F are fractional ideals of R, then the following are true.*

(1) $E+F$, $E\cap F$, EF, $E:F$, F^{-1} are fractional ideals of R.
(2) $F^{-1}F\subseteq R$ and $F^{-1}F$ is an ideal of R.
(3) $F:F$ is a subring of K that contains R.

PROOF. By Lemma 7.2.1 there exist c,d nonzero elements of R such that $cE\subseteq R$ and $cF\subseteq R$. The proof is left as an exercise for the reader. □

DEFINITION 7.2.4. In the context of Lemma 7.2.3, let $\mathrm{Frac}(R)$ denote the set of all fractional ideals of R in K. Then multiplication is a binary operation on $\mathrm{Frac}\,R$ which makes $\mathrm{Frac}\,R$ into an abelian monoid, with R being the identity element. A fractional ideal F is said to be an *invertible fractional ideal*, if $F^{-1}F=R$.

PROPOSITION 7.2.5. *Let R be an integral domain with field of fractions K. If F is an invertible fractional ideal of R in F, then F is a finitely generated projective R-module.*

PROOF. Since $F^{-1}F$ is equal to the unit ideal R, there exists a presentation of 1 in the form $1=\sum_{i=1}^{n}x_iy_i$, where each x_i is in F^{-1} and each y_i is in F. Let $\lambda_i:F\to R$ be the "left multiplication by x_i" function defined by: $\lambda_i(y)=x_iy$. The reader should verify that $\{(y_i,\lambda_i)\mid 1\le i\le n\}$ is a dual basis for F over R. By Exercise 4.5.29, F is a finitely generated projective R-module. □

3. The Ideal Class Group of a Dedekind Domain

The subject of this section is an important class of commutative rings called Dedekind domains, named for the nineteenth century algebraist R. Dedekind.

DEFINITION 7.3.1. Let S be an integral domain with quotient field L. Then we say S is a *Dedekind domain* if the following properties are satisfied.

(1) S is integrally closed in L.
(2) S satisfies the ascending chain condition on ideals. Equivalently, every ideal of S is finitely generated and S satisfies the maximum condition on ideals (Proposition 4.6.2).
(3) S is not a field and every nonzero prime ideal of S is maximal.

In Definition 7.3.1, an integral domain that satisfies (1) is said to be integrally closed. A commutative ring that satisfies (2) is said to be noetherian (after E. Noether). A commutative ring that satisfies (3) is said to have Krull dimension one (after W. Krull). Hence, a Dedekind domain is an integrally closed noetherian integral domain with Krull dimension one.

The purpose of this section is to prove a unique factorization theorem for ideals of a Dedekind domain S. If A is a proper ideal of S, then there exist prime ideals $P_1,\dots,P_n$ in S such that $A=P_1\cdots P_n$. The number n is uniquely determined by A and the sequence of primes $P_1,\dots,P_n$ is also unique up to permuting the list entries. The proof given below is based on ideas found in [**7**].

In Theorem 7.3.2 below, we show that the integral closure of a principal ideal domain in a finite separable extension of its quotient field is a Dedekind domain. This is a very common scenario in Algebraic Geometry as well as Algebraic Number Theory. A Dedekind domain is not necessarily a unique factorization domain. For some examples, see Section 7.4.

THEOREM 7.3.2. *Let R be a principal ideal domain with quotient field K. If L/K is a finite dimensional separable extension of fields and S is the integral closure of R in L, then the following are true.*

1. *S is a finitely generated free R-module and $\mathrm{Rank}_R(S)$ is equal to $\dim_K(L)$.*
2. *If P is a nonzero prime ideal of S, then P is a maximal ideal.*
3. *If F is a fractional ideal of S in L, then F is a finitely generated S-module.*
4. *S satisfies the ascending chain condition (ACC) on ideals. That is, given a chain of ideals $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots \subseteq I_n \subseteq \cdots$, there exists $N \geq 1$ such that $I_N = I_{N+1} = \cdots$.*
5. *S satisfies the maximum condition on ideals. That is, any nonempty family of ideals in S has a supremum.*

PROOF. (1): This is Corollary 7.1.16.

(2): By Corollary 7.1.13, $P \cap R$ is a nonzero prime ideal of R. By Exercise 3.4.30, $P \cap R$ is a maximal ideal of R. By Lemma 7.1.19, P is a maximal ideal of S.

(3): By Part (1), S is a finitely generated R-module. If F is a fractional ideal of S, then by Lemma 7.2.1, F is an R-submodule of bS, for some $b \in L$. By Theorem 4.6.1, F is a finitely generated free R-module. A generating set for F over R is definitely a generating set for F over S. This proves (3).

(4) and (5): By (3), every ideal of S is finitely generated. Parts (4) and (5) follow from (3) and Proposition 4.6.2. □

For the remainder of this section, S is a Dedekind domain. An ideal I of S will be called a *proper ideal* if $(0) \neq I \neq S$. As stated above, our goal in this section is to prove that a proper ideal A in a Dedekind domain S has a unique factorization as a product of prime ideals. Lemma 7.3.3 is the first step in proving the existence of a prime factorization. In the following, a product of prime ideals implicitly means a product of nonzero prime ideals.

LEMMA 7.3.3. *If A is a proper ideal in S, then the following are true.*

1. *A contains a product of prime ideals. That is, either A is a prime ideal, or there exist nonzero primes $P_1, \ldots, P_n$ such that $P_1 \cdots P_n \subseteq A$.*
2. *In the context of (1), if $\mathfrak{m}$ is a maximal ideal of S such that $P_1 \cdots P_n \subseteq A \subseteq \mathfrak{m}$, then there exists $1 \leq i \leq n$ such that $P_i = \mathfrak{m}$.*

PROOF. (1): For contradiction's sake, assume there exists a proper ideal A of S which does not contain a product of prime ideals. By Definition 7.3.1 (2), we can assume A is a maximal counterexample. Then A is not a prime ideal. By Exercise 7.3.12, there exist proper ideals I and J of S such that $A \subsetneq I$, $A \subsetneq J$, and $IJ \subseteq A$. Since A is a maximal counterexample, the ideals I and J both contain a product of prime ideals. Then so does A, a contradiction.

(2): This is Exercise 7.3.13. □

LEMMA 7.3.4. *If A is a proper ideal of S, then $S \subsetneq A^{-1}$.*

PROOF. By Definition 7.2.2, $A^{-1} = S : A = \{x \in L \mid xA \subseteq A$. Therefore, $S \subseteq A^{-1}$. We show $S \neq A^{-1}$. Pick an element $a \in A - (0)$. By Lemma 7.3.3 (1), there exist prime ideals P_i such that $P_1 \cdots P_n \subseteq aS \subseteq A$. Out of all products of prime ideals contained in aS, pick one such that n is minimal. Let $\mathfrak{m}$ be a maximal ideal containing A. By Lemma 7.3.3 (2), $\mathfrak{m}$ is equal to one of the primes in the factorization. Without loss of generality, say $P_1 = \mathfrak{m}$. We have $\mathfrak{m}P_2 \cdots P_n \subseteq aS \subseteq$

A, where it is understood that $P_2 \cdots P_n = S$, if $n = 1$. By the minimal choice of n, $P_2 \cdots P_n$ is not contained in aS. Let $b \in P_2 \cdots P_n$ such that a does not divide b. Then in the quotient field L we have $a^{-1}b \in L - S$. It follows from

$$\begin{aligned} a^{-1}bA &\subseteq a^{-1}b\mathfrak{m} \\ &\subseteq a^{-1}\mathfrak{m}P_2 \cdots P_n \\ &\subseteq S \end{aligned}$$

that $a^{-1}b \in A^{-1}$. □

LEMMA 7.3.5. *If P is a proper prime ideal of S, then $P^{-1}P = S$. That is, P is an invertible ideal.*

PROOF. By Lemma 7.2.3, $P^{-1}P$ is an ideal in S. By Lemma 7.3.4, $S \subsetneq P^{-1}$. Therefore, $P \subseteq P^{-1}P \subseteq S$. By Definition 7.3.1 (3), P is a maximal ideal. Therefore, either $P = P^{-1}P$ or $P^{-1}P = S$. If $P = P^{-1}P$, then $P^{-1} \subseteq P : P$. By Exercise 7.3.14 (3), this implies $P^{-1} \subseteq S$, a contradiction to Lemma 7.3.4. □

THEOREM 7.3.6. *If A is a proper ideal in S, then there are prime ideals $P_1, \ldots, P_n$ in S such that $A = P_1 \cdots P_n$. This factorization is unique in the sense that if $A = Q_1 \cdots Q_m$ is another prime factorization of A, then $n = m$ and after a suitable permutation of the prime factors we have $P_i = Q_i$ for each i.*

PROOF. First we prove the existence claim. By Lemma 7.3.3, there exist prime ideals $P_1, \ldots, P_n$ such that $P_1 \cdots P_n \subseteq A$. Define $n(A)$ to be the minimal number such that A contains a product of $n(A)$ primes. Then $n(A) = 1$ if and only if A is a prime ideal of S. The proof is by induction on $n(A)$. Suppose A is a proper ideal of S such that $n(A) \geq 2$ and that a prime factorization exists for any ideal B such that $n(B) < n(A)$. Start with $P_1 \cdots P_{n(A)} \subseteq A$. Let $\mathfrak{m}$ be a maximal ideal such that $A \subseteq \mathfrak{m}$. By Lemma 7.3.3, after permuting the factors if necessary, we have $P_1 = \mathfrak{m}$. Then $\mathfrak{m}P_2 \cdots P_{n(A)} \subseteq A \subseteq \mathfrak{m}$. By Lemma 7.3.5, $\mathfrak{m}^{-1}\mathfrak{m} = S$. Therefore, $P_2 \cdots P_{n(A)} \subseteq \mathfrak{m}^{-1}A \subseteq S$. Since A is not prime, we know $A \neq \mathfrak{m}$, which implies $\mathfrak{m}^{-1}A \neq S$. By the induction hypothesis, $\mathfrak{m}^{-1}A$ is equal to a product $Q_1 \cdots Q_m$ for certain prime ideals Q_i in S. Then $A = \mathfrak{m}Q_1 \cdots Q_m$ is a prime factorization of A.

Now we prove the uniqueness claim. Let A be a proper ideal of S and assume $P_1, \ldots, P_n, Q_1, \ldots, Q_m$ are prime ideals such that $A = P_1 \cdots P_n = Q_1 \cdots Q_m$. Since $P_1 \cdots P_n \subseteq P_1$, by Lemma 7.3.3, there is some $1 \leq j \leq m$ such that $Q_j = P_1$. After a suitable permutation of the factors, assume $Q_1 = P_1$. By Lemma 7.3.5, P_1 is invertible, so after multiplying both sides by P_1^{-1} we have the identity $P_2 \cdots P_n = Q_2 \cdots Q_m$. A familiar induction argument completes the proof. □

COROLLARY 7.3.7. *Every fractional ideal of S in L is invertible. The set* $\mathrm{Frac}(S)$ *of all fractional ideals of S is an abelian group, where the binary operation is multiplication. The set* $\mathrm{Prin}(S)$ *of all principal fractional ideals is a subgroup. The abelian group* $\mathrm{Frac}(S)$ *is a free $\mathbb{Z}$-module, the set of nonzero prime ideals of S is a free basis.*

PROOF. A nonzero prime ideal of S is invertible by Lemma 7.3.5. If $\alpha \in L^*$, then $(\alpha S)(\alpha^{-1}S) = S$, so a principal ideal is invertible. If F is a fractional ideal, then there is some $\alpha \in L^*$ such that $\alpha F \subseteq S$, by Lemma 7.2.1. By Theorem 7.3.6, $\alpha F \subseteq S$ is a product of invertible fractional ideals. It follows that F is invertible. The rest is left to the reader. □

DEFINITION 7.3.8. Let S be a Dedekind domain with quotient field L. Let $\mathrm{Frac}(S)$ be the set of all fractional ideals of S in L and $\mathrm{Prin}(S)$ the set of all principal fractional ideals. The quotient group $\mathrm{Frac}(S)/\mathrm{Prin}(S)$ is called the *class group* of S and is denoted $\mathrm{Cl}(S)$.

As shown in Corollary 7.3.9 below, the class group of S measures how close S is to being a principal ideal domain.

COROLLARY 7.3.9. *Let S be a Dedekind domain. Then S is a principal ideal domain if and only if $\mathrm{Cl}(S)$ is the trivial group.*

PROOF. If S is a principal ideal domain, then every ideal in S is principal. By Corollary 7.3.7, every fractional ideal of S is principal, hence the class group of S is trivial. Conversely, if S is a Dedekind domain which is not a principal ideal domain, then $\mathrm{Prin}(S)$ is not equal to $\mathrm{Frac}(S)$ and the class group of S is nontrivial. □

We end this section with a proof of a special case of Nagata's Theorem. The interested reader is referred to [**9**, Theorem 15.4.16] for a more general version. For computing the class group of an integrally closed integral domain R, localization methods are particularly useful. If K is the quotient field of R and $f \in R - \{0\}$, then by $R[f^{-1}]$ we denote the subring of K generated by R and f^{-1}. A typical element of $R[f^{-1}]$ is a quotient of the form r/f^k, where $k \geq 0$. If F is a fractional ideal of R in K, then the product $FR[f^{-1}]$ is a fractional ideal of $R[f^{-1}]$ in K. By Lemma 7.3.10 and Exercise 7.3.17, if S is a Dedekind domain, then so is $S[f^{-1}]$ for any $f \in S - \{0\}$.

LEMMA 7.3.10. *Let R be an integral domain with quotient field K and $f \in R - \{0\}$.*

(1) If J is an ideal in $R[f^{-1}]$, then $J = (J \cap R)[f^{-1}]$.
(2) If P is a prime ideal in R and $f \notin P$, then $P[f^{-1}]$ is a prime ideal in $R[f^{-1}]$.
(3) There is a one-to-one correspondence between prime ideals of $R[f^{-1}]$ and the set of all prime ideals of R that do not contain f.

PROOF. The proof is left as an exercise for the reader. □

THEOREM 7.3.11. *(Nagata's Theorem) Let S be a Dedekind domain with quotient field L. If $f \in S - \{0\}$, then there is an onto homomorphism of abelian groups $\gamma : \mathrm{Cl}(S) \to \mathrm{Cl}(S[f^{-1}])$. Let $Sf = P_1^{e_1} \cdots P_n^{e_n}$ be the prime factorization of the principal ideal Sf, where the P_i are pairwise distinct primes and $e_i \geq 1$ for each i. Then the kernel of γ is generated by the ideal classes of $P_1, \ldots, P_n$.*

PROOF. The abelian group $\mathrm{Frac}(S)$ is a free $\mathbb{Z}$-module with basis the set of nonzero prime ideals of S. By Lemma 7.3.10, we can view $\mathrm{Frac}(S[f^{-1}])$ as the submodule of $\mathrm{Frac}(S)$ generated by those prime ideals of S which do not contain f. Let $\pi : \mathrm{Frac}(S) \to \mathrm{Frac}(S[f^{-1}])$ be the projection map. Then π is an onto homomorphism. The kernel of π is the free $\mathbb{Z}$-submodule $\mathbb{Z}P_1 \oplus \cdots \oplus \mathbb{Z}P_n$. It is clear that

$$\pi : \mathrm{Prin}(S) \to \mathrm{Prin}(S[f^{-1}]) \tag{3.1}$$

is a homomorphism. A principal fractional ideal of $S[f^{-1}]$ in L is equal to $S[f^{-1}] = (Su)[f^{-1}]$ for some $u \in L$. Therefore, the map in (3.1) is onto. By Theorem 2.3.12

there is a commutative diagram

$$\begin{array}{ccc} \mathrm{Frac}(S) & \longrightarrow & \mathrm{Cl}(S) \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \gamma} \\ \mathrm{Frac}(S[f^{-1}]) & \longrightarrow & \mathrm{Cl}(S[f^{-1}]) \end{array}$$

and γ is onto. By Exercise 7.3.16, the kernel of γ is generated by the ideal classes of the prime ideals $P_1, \ldots, P_n$. □

3.1. Exercises.

EXERCISE 7.3.12. Let R be a commutative ring in which $0 \neq 1$. Prove that R is not an integral domain if and only if there exist ideals I and J such that $I \neq (0)$, $J \neq (0)$ and $IJ = (0)$.

EXERCISE 7.3.13. Let R be a commutative ring, $n > 0$, and $\mathfrak{m}_1, \ldots, \mathfrak{m}_n$ maximal ideals in R. If $\mathfrak{m}$ is a maximal ideal in R such that $\mathfrak{m}_1 \cdots \mathfrak{m}_n \subseteq \mathfrak{m}$, then there exists $1 \leq i \leq n$ such that $\mathfrak{m}_i = \mathfrak{m}$.

EXERCISE 7.3.14. Let S be a Dedekind domain with quotient field L and F a fractional ideal of S in L. Prove the following:

(1) F is isomorphic as an S-module to a proper ideal of S.
(2) F is finitely generated as an S-module.
(3) $F : F = S$.
(4) F is a free S-module if and only if F is a principal fractional ideal.

EXERCISE 7.3.15. Prove Lemma 7.3.10.

EXERCISE 7.3.16. The purpose of this exercise is to prove a special case of the so-called Snake Lemma ([**9**, Theorem 6.6.2]). Let R be a ring, A an R-module with a submodule A_0, B an R-module with a submodule B_0, and $\pi : A \to B$ an epimorphism such that $\pi(A_0) = B_0$. Prove:

(1) There is an epimorphism $\gamma : A/A_0 \to B/B_0$.
(2) If $\eta : A \to A/A_0$ is the natural map, then $\eta(\ker(\pi)) = \ker(\gamma)$.

EXERCISE 7.3.17. Let R be an integral domain with quotient field K. Let $f \in R - \{0\}$. Prove:

(1) If R is integrally closed in K, then $R[f^{-1}]$ is integrally closed in K.
(2) If R is a Dedekind domain, then $R[f^{-1}]$ is a Dedekind domain.
(3) If R is a principal ideal domain, then $R[f^{-1}]$ is a principal ideal domain.

EXERCISE 7.3.18. Let S be a Dedekind domain, P a nonzero prime ideal in S, $e > 0$ and $A = S/(P^e)$. Prove the following generalization of Exercise 4.6.18:

(1) Every ideal in A is principal.
(2) A is a field if and only if $e = 1$.
(3) A is a local ring with maximal ideal P/P^e.
(4) A has exactly $e + 1$ ideals, namely: $(0) \subseteq P^{e-1}/P^e \subseteq \cdots \subseteq P^2/P^e \subseteq P/P^e \subseteq A$.

EXERCISE 7.3.19. Let S be a Dedekind domain, $P_1, \ldots, P_n$ distinct nonzero prime ideals of S, $e_1, \ldots, e_n$ positive integers, $I = P_1^{e_1} P_2^{e_2} \cdots P_n^{e_n}$, and $A = S/I$. Prove the following generalization of Exercise 4.6.19:

(1) A is isomorphic to the direct sum of the local rings $\bigoplus_i S/P_i^{e_i}$.
(2) Every ideal in A is principal.
(3) Including the two trivial ideals (0) and A, there are exactly $(e_1+1)(e_2+1)\cdots(e_n+1)$ ideals in A.
(4) A has exactly n maximal ideals, namely $P_1/I, \ldots, P_n/I$.

EXERCISE 7.3.20. Let S be a Dedekind domain, I a proper ideal in S, and $\alpha \in I - (0)$. Show that there exists $\beta \in I$ such that $I = S\alpha + S\beta$.

EXERCISE 7.3.21. Let S be a Dedekind domain and $P_1, \ldots, P_n$ distinct nonzero prime ideals of S. Let $e_1, \ldots, e_n, f_1, \ldots, f_n$ be nonnegative integers and set $I = \prod P_i^{e_i}$ and $J = \prod P_i^{f_i}$. Prove the following generalization of Exercise 3.4.28.

(1) $I \subseteq J$ if and only if $e_i \geq f_i$ for each i.
(2) If $m_i = \min(e_i, f_i)$ for each i, then $I + J = \prod P_i^{m_i}$.
(3) If $M_i = \max(e_i, f_i)$ for each i, then $I \cap J = \prod P_i^{M_i}$.

EXERCISE 7.3.22. Let S be a Dedekind domain. Show that if A and B are ideals in S such that $A \subseteq B$, then there exists an ideal C in S such that $A = BC$.

EXERCISE 7.3.23. Let S be a Dedekind domain. Show that if I and J are proper ideals in S, then there exists an element β in S and an ideal C in S satisfying $J + C = S$ and $IC = S\beta$.

EXERCISE 7.3.24. Let S be a Dedekind domain. Show that if I and J are proper ideals in S such that $I + J = S$, then there exists an isomorphism of S-modules $I \oplus J \cong S \oplus IJ$.

EXERCISE 7.3.25. Let S be a Dedekind domain with quotient field L. Suppose E and F are fractional ideals of S in L. Prove:

(1) For any $\alpha \in L^*$, there is an S-module isomorphism $E \cong \alpha E$ defined by $x \mapsto \alpha x$.
(2) There exist α, β in L such that αE and βF are proper ideals of S and $\alpha E + \beta F = S$.
(3) There exists an isomorphism of S-modules $E \oplus F \cong S \oplus EF$.

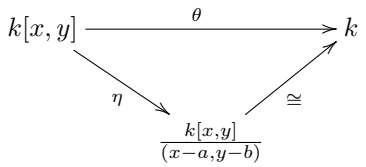
4. Applications to Algebraic Curves

If k is a field, then the affine plane over k is the cartesian product $k^2 = k \times k$. If x and y are indeterminates and $f(x, y) \in k[x, y]$, then

$$Z(f) = \{(a, b) \in k^2 \mid f(a, b) = 0\}$$

is the set of solutions of the equation $f(x, y) = 0$ in k^2. We call $Z(f)$ an affine algebraic curve. This terminology agrees with that of Section 5.1.2 where we discussed lines and circles in the affine plane. The commutative ring $R = k[x, y]/(f)$ is known as the affine coordinate ring of the curve $Z(f)$. There is a correspondence between points (a, b) on the curve $Z(f)$ and maximal ideals in R. For instance, given $(a, b) \in Z(f)$, consider the ideal $(x - a, y - b)$ in $k[x, y]$. Applying Exercise 3.6.37 twice, once for $x - a$ and once for $y - b$, we see that $k[x, y]/(x - a, y - b) \cong k$. Hence $(x - a, y - b)$ is a maximal ideal. Applying Theorem 3.6.2 for x and y, there is a k-algebra homomorphism $\theta : k[x, y] \to k$ defined by $x \mapsto a$ and $y \mapsto b$.

Clearly θ is onto and the maximal ideal $(x-a, y-b)$ is contained in $\ker\theta$. Hence $\ker\theta = (x-a, y-b)$. By Theorem 3.2.15, there is a commutative diagram



of k-algebras where η is the natural map. Since $\theta(f) = f(a,b) = 0$, $f \in \ker\theta = (x-a, y-b)$. If we set $\mathfrak{m} = (x-a, y-b)$ in R, then by Corollary 3.2.17, $R/\mathfrak{m} = k[x,y]/(x-a, y-b) = k$ and $\mathfrak{m}$ is a maximal ideal of R. The correspondence between points of $Z(f)$ and maximal ideals of R is not onto. That is, if M is a maximal ideal of R, then R/M is in general an extension field of k, hence M does not have the form $(x-a, y-b)$ and does not correspond to a point on $Z(f)$. In this case, the maximal ideal M is called an R/M-rational point of the curve $Z(f)$. In the example of Section 7.4.1 below, we show that when k is not algebraically closed, the affine coordinate ring R of the circle $Z(x^2+y^2-1)$ has maximal ideals such that the residue field R/M is strictly greater than k (see Proposition 7.4.1 (4)). Although we do not prove it here, since R/M is a finitely generated k-algebra, by the Hilbert Basis Theorem, R/M is finitely generated and algebraic over k (see [**9**, Proposition 10.2.4] or [**5**, Proposition 7.9], for example). Therefore, if k is algebraically closed, then every point of $Z(f)$ is k-rational.

The example we consider below in Section 7.4.1 is an affine curve C of degree 2 such that the class group $\mathrm{Cl}(R)$ of the affine coordinate ring R is an abelian group of order 2. The example studied in Section 7.4.2 below is an affine cubic curve C such that the class group $\mathrm{Cl}(R)$ of the affine coordinate ring R contains a subgroup corresponding to the set $Z(f)$ of k-rational points on C. This is a very elementary example of a group defined by an elliptic curve. The coordinate rings of the curves studied in the following sections are examples of the type of rings featured in Section 7.3.

4.1. A Nonsingular Affine Conic. If k is a field, then the unit circle C in the plane k^2 is the set of solutions of the equation $x^2+y^2-1=0$. That is,

$$C = \{(x,y) \in k^2 \mid x^2+y^2-1=0\}.$$

This terminology agrees with that of Section 5.1.2. In this section we investigate the commutative ring $R = k[x,y]/(x^2+y^2-1)$ which is known as the affine coordinate ring of the unit circle C over the field k.

First we establish notation that will be in effect throughout this section. Let k be a field such that x^2+1 is irreducible over k. In particular, this implies that the characteristic of the base field k is not 2 (Exercise 3.2.31). Let $k[x]$ be the polynomial ring in one variable over k. Then $k[x]$ is a PID (Corollary 3.6.5) and $x-1$ is a prime in $k[x]$. Let $k(x)$ be the field of rational functions, the quotient field of $k[x]$. Consider the polynomial x^2+y^2-1 in $k[x][y]$. By Eisenstein's Criterion (Theorem 3.7.6) with prime $p = x-1$, $y^2+(x^2-1)$ is irreducible in $k[x][y]$. By Theorem 3.7.5, the polynomial ring $k[x][y]$ is a UFD. Therefore,

$$R = \frac{k[x,y]}{(x^2+y^2-1)}$$

is an integral domain, by Corollary 3.4.14. The ring R is known as the affine coordinate ring of the unit circle C in the affine plane k^2. By Gauss' Lemma (Theorem 3.7.4), x^2+y^2-1 is irreducible in $k(x)[y]$ and

$$F = \frac{k(x)[y]}{(y^2+x^2-1)}$$

is a field. By Exercise 5.3.36, F is a Galois extension of $k(x)$. The Galois group $\mathrm{Aut}_{k(x)} F$ is cyclic of order 2, and generated by the automorphism τ defined by $y \mapsto -y$. Let $K = k(i)$ be the splitting field for x^2+1 over k. The Galois group of K/k is the cyclic group $\langle\sigma\rangle$, where $\sigma(i) = -i$. In the following, cosets in the factor rings R and F are written without brackets or any extra adornment.

PROPOSITION 7.4.1. *In the above context, the following properties hold for R and F:*

(1) F is the quotient field of R.
(2) As a $k[x]$-module, R is free of rank two with basis $1, y$.
(3) There is a norm map $N^R_{k[x]} : R \to k[x]$ defined by $a+by \mapsto (a+by)(a-by) = a^2-b^2y^2 = a^2-b^2(1-x^2)$.
(4) In general, R contains K-rational points.

PROOF. (1): The diagram of ring homomorphisms

$$\begin{array}{ccc} R = \frac{k[x,y]}{(x^2+y^2-1)} & \xrightarrow{\phi} & F = \frac{k(x)[y]}{(x^2+y^2-1)} \\ \uparrow \eta & & \eta \uparrow \\ k[x,y] & \xrightarrow{\alpha} & k(x)[y] \end{array} \tag{4.1}$$

commutes. The vertical maps are the natural maps. The horizontal map α exists by Theorem 3.6.2 applied to $k[x] \to k(x)$. Since $\eta\alpha(x^2+y^2-1) = 0$, ϕ exists by Theorem 3.2.15. Using Gauss' Lemma (Theorem 3.7.4), we see that the kernel of $\eta\alpha$ is the principal ideal (x^2+y^2-1). Therefore, ϕ is one-to-one. By Exercise 3.5.2, we can view the quotient field of R as a subfield of F. In this context, we show that F is equal to the quotient field of R. By Lemma 4.4.5, a $k(x)$-basis for F is $\{1, y\}$. Since $y \in R$ we know y is in the quotient field of R. The quotient field of $k[x]$ is $k(x)$, hence $k(x)$ is in the quotient field of R. A typical element of F is of the form $f(x)+g(x)y$, where $f(x)$ and $g(x)$ are in $k(x)$. Hence a typical element of F is in the quotient field of R.

(2): By Lemma 4.4.5, a $k(x)$-basis for F is $\{1, y\}$. Therefore, $\{1, y\}$ is linearly independent over $k[x]$. Since the image of $\eta\alpha$ is generated by polynomials over k in the variables x and y, $\{1, y\}$ is a generating set for the image of ϕ as a $k[x]$-module. In Diagram (4.1), ϕ is one-to-one. So $\{1, y\}$ is a generating set for R as an A-module.

(3): The norm map $N^F_{k(x)} : F \to k(x)$ restricts to a norm map $R \to k[x]$.

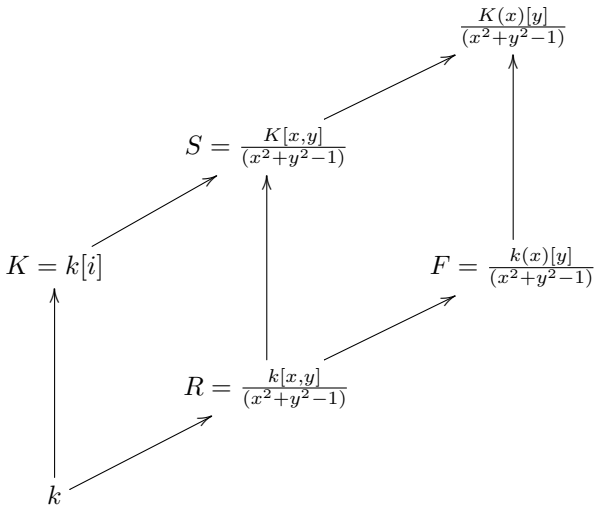
(4): To see that for general k, R has K-rational points, suppose for instance that 2 is a square in k. Then $R/(y-\sqrt{2}) \cong k[x]/(x^2+1) \cong K$. Therefore, the principal ideal $(y-\sqrt{2})$ is a maximal ideal of R with residue field K. □

We retain the notation from Proposition 7.4.1. The affine coordinate ring of the unit circle in the plane K^2 is $S = K[x,y]/(x^2+y^2-1)$. Identifying K with

$k[z]/(z^2+1)$, we see that

$$\begin{aligned} S &= \frac{k[x,y,z]}{(x^2+y^2-1, z^2+1)} \\ &= \frac{R[z]}{(z^2+1)} \\ &= R[i]. \end{aligned}$$

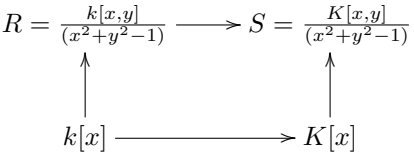
By Exercise 4.2.26, S is a free R-module with rank 2 and basis $1, i$. The diagram



commutes. By Proposition 7.4.1, the quotient field of S is

$$\frac{K(x)[y]}{(x^2+y^2-1)} = \frac{k(x)[y][z]}{(x^2+y^2-1, z^2+1)} = F[i].$$

The field extension $F[i]/F$ is Galois with group $\langle\sigma\rangle$ where $\sigma(i) = -i$. Notice that σ restricts to an R-algebra automorphism of S and the norm $N_R^S : S^* \to R^*$ is a homomorphism of groups. We can also view R as the ramified quadratic extension of $k[x]$ defined by adjoining the square root of $1-x^2$. Likewise, S is the ramified quadratic extension of $K[x]$ defined by adjoining the square root of $1-x^2$. The diagram



commutes, every arrow is the set inclusion map. Geometrically, the extension $R/k[x]$ or $S/K[x]$ corresponds to the projection of the circle C onto the x-axis.

PROPOSITION 7.4.2. *In the above context, the following are true.*

(1) S is a PID. S is integrally closed in $F[i]$.
(2) $S^* = K^* \times \langle x+iy \rangle$.
(3) $R^* = k^*$.
(4) R is not a PID or UFD. In fact, x, $y+1$, $y-1$ are irreducible in R and $x^2 = (1+y)(1-y)$.

(5) R is integrally closed in F.

PROOF. (1) and (2): To show that S is a PID, by Exercise 3.6.30, it suffices to show that S is isomorphic to the ring of Laurent polynomials $K[u, u^{-1}]$. In S we have $x^2 + y^2 - 1 = (x + iy)(x - iy) - 1$. Define K-algebra homomorphisms

$$\frac{K[u,v]}{(uv-1)} \xrightarrow{\phi} \frac{K[x,y]}{(x^2+y^2-1)} \xrightarrow{\theta} \frac{K[u,v]}{(uv-1)}$$

by $\phi(u) = x + iy$, $\phi(v) = x - iy$, $\theta(x) = \frac{u+v}{2}$ and $\theta(y) = \frac{u-v}{2i}$. One can check that ϕ and θ are well defined K-algebra homomorphisms. Since ϕ and θ are inverses of each other, they are isomorphisms. There is an isomorphism of K-algebras

$$\frac{K[u,v]}{(uv-1)} \xrightarrow{\cong} K[u,u^{-1}]$$

induced by $v \mapsto u^{-1}$. By Exercise 3.6.30, $K[u, u^{-1}]$ is a PID and the group of units is equal to the internal direct product $K^* \times \langle u \rangle$. Using the isomorphism θ, this proves S is a PID and

$$S^* = K^* \times \langle x + iy \rangle.$$

Notice that the inverse of $x + iy$ is $x - iy$. By Proposition 7.1.1, a PID is integrally closed in its quotient field. This proves (1) and (2).

(3): We have the homomorphism of groups $N_R^S : S^* \to R^*$ and if $a \in R^*$, then $N_R^S(a) = a^2$. Since $N_R^S(x + iy) = (x + iy)(x - iy) = x^2 + y^2 = 1$, we see that $R^* = (K^*)^{\langle \sigma \rangle} = k^*$.

(4): To prove that x is irreducible in R, we use the norm map $R \to k[x]$ of Proposition 7.4.1 (3). Look at the norm of x from R to $k[x]$:

$$N_{k[x]}^R(x) = x^2.$$

For sake of contradiction, assume x has a nontrivial factorization $x = \alpha\beta$ in R. Since $R^* = k^*$, this means the norm of α is equal to cx for some $c \in k^*$. Suppose $\alpha = a + by$ for some $a, b \in k[x]$. Then the equation $N_{k[x]}^R(a + by) = cx$ becomes

$$a^2 - b^2(1 - x^2) = cx.$$

Substitute $x = 1$ and $x = -1$ to get $c = a(1)^2$ and $-c = a(-1)^2$. Hence

$$-1 = a(1)^2 a(-1)^{-2}$$

which contradicts our assumption that -1 is not a square in k. Therefore, x is irreducible in R. Since

$$N_{k[x]}^R(1 + y) = N_{k[x]}^R(1 - y) = (1 + y)(1 - y) = x^2$$

the same argument proves that $y + 1$ and $y - 1$ are irreducible in R. This proves R is not a UFD, since the identity

$$x^2 = (1 - y)(1 + y)$$

holds in R. Theorem 3.4.15 implies R is not a PID.

(5): Let α be an element of F and assume α is integral over R. We can view α as an element of $F[i]$ which is integral over S. Since S is integrally closed in $F[i]$, this means α is in $S = R[i]$. Since S is free over R with rank 2, there are unique a, b in R such that $\alpha = a + bi$. Since $\alpha \in F$, this implies $b = 0$, hence $\alpha \in R$. □

LEMMA 7.4.3. *In the above context, the ideal $\mathfrak{m} = (x, y - 1)$ of R has the following properties:*

(1) $\mathfrak{m}$ *is a maximal ideal.*
(2) $\mathfrak{m}$ *is not a principal ideal.*
(3) $\mathfrak{m}$ *is a projective* R*-module.*
(4) $\mathfrak{m}$ *is not a free* R*-module.*
(5) $\mathfrak{m}^2$ *is equal to the principal ideal* $(y-1)$*.*

PROOF. (1): Since $R/\mathfrak{m} = k[x,y]/(x,y-1) = k$ is a field, $\mathfrak{m}$ is a maximal ideal.

(2): Assume $\mathfrak{m} = (z)$ is a principal ideal. Then z divides x. Since x is irreducible, this implies z and x are associates. But $R/(x) = k[y]/(y^2-1)$ is not a field. Therefore, $\mathfrak{m} \neq (x)$, a contradiction.

(3) and (4): These follow from (1), (2), Proposition 7.2.5 and Exercise 7.3.14.

(5): Notice that $\mathfrak{m}^2 = (x,y-1)^2$ is generated by the three elements $x^2 = 1-y^2$, $x(y-1)$, and $(y-1)^2$, all of which are in the principal ideal $(y-1)$. Conversely, since $x^2+y^2=1$ in R,

$$\begin{aligned} x^2 + (y-1)^2 &= x^2 + y^2 - 2y + 1 \\ &= 2(1-y) \end{aligned}$$

which shows $y-1$ is in $\mathfrak{m}^2$. This proves $\mathfrak{m}^2 = (y-1)$ is a principal ideal in R. □

COROLLARY 7.4.4. *In the above context, the divisor class group of* R*,* $\mathrm{Cl}(R)$*, is a cyclic group of order* 2*, and is generated by the ideal class of* $\mathfrak{m} = (x, y-1)$*. This ideal corresponds to the point* $(0,1)$ *on the circle* C*.*

PROOF. First we show that the localization $R[x^{-1}]$ is isomorphic to the localization of $k[u,v]/(u^2-v^2-1)$ with u^{-1} inverted. For this, we define k-algebra homomorphisms

$$\frac{k[x,y][x^{-1}]}{(x^2+y^2-1)} \xrightarrow{\phi} \frac{k[u,v][u^{-1}]}{(u^2-v^2-1)} \xrightarrow{\theta} \frac{k[x,y][x^{-1}]}{(x^2+y^2-1)}$$

by $\phi(x) = u^{-1}$, $\phi(y) = vu^{-1}$, $\theta(u) = x^{-1}$, $\theta(v) = yx^{-1}$. One can check that ϕ and θ are well defined k-algebra homomorphisms. Since ϕ and θ are inverses of each other, they are isomorphisms. Likewise, the k-algebra homomorphism

$$\frac{k[u,v]}{(u^2-v^2-1)} \xrightarrow{\psi} \frac{k[s,t]}{(st-1)}$$

defined by $\psi(u) = (t+s)/2$, $\psi(v) = (t-s)/2$ is an isomorphism. As shown in the proof of Proposition 7.4.2, $k[s,t]/(st-1)$ is isomorphic to $k[z,z^{-1}]$. By Exercise 7.3.17 the localization of a PID is a PID. Therefore, $k[z,z^{-1}]$, $k[s,t]/(st-1)$, and $R[x^{-1}]$ are all principal ideal domains. By Corollary 7.3.9, the class group of a PID is trivial. By Theorem 7.3.11, the ideal class group $\mathrm{Cl}(R)$ is generated by $\mathfrak{m} = (x,y-1)$ and $\mathfrak{n} = (x,y+1)$, the primes that contain x. But $\mathfrak{mn} = \mathfrak{m} \cap \mathfrak{n} = (x)$ is principal. Together with Lemma 7.4.3 (5), this proves (6). □

4.2. A Nonsingular Affine Elliptic Curve. This short section is devoted to an example of an algebraic curve that is nonsingular and nonrational. Assume that the characteristic of k, the base field, is not 2. Let $A = k[x]$ be the polynomial ring in one variable over k. Then A is a UFD (Example 3.4.12) and x is a prime in A. Let $K = k(x)$ be the quotient field of A. Consider the polynomial $y^2 - x(x^2-1)$ in $A[y]$. By Eisenstein's Criterion (Theorem 3.7.6) with prime $p = x$, $y^2 - x(x^2-1)$ is irreducible in $A[y]$. By Gauss' Lemma (Theorem 3.7.4), $y^2 - x(x^2-1)$ is irreducible in $K[y]$ and $F = K[y]/(y^2 - x(x^2-1))$ is a field. By Exercise 5.3.36, F/K is a

Galois extension, $\mathrm{Aut}_K(F) = \langle\sigma\rangle$ has order 2, and σ is defined by $y \mapsto -y$. The norm map is $N_K^F : F \to K$.

In the following, cosets in the factor ring F are written without brackets or any extra adornment. By Theorem 3.7.5, the polynomial ring $A[y] = k[x, y]$ is a UFD. Therefore, $R = k[x, y]/(y^2 - x(x^2 - 1))$ is an integral domain, by Corollary 3.4.14. The diagram of ring homomorphisms

$$\begin{array}{ccc} A = k[x] & \longrightarrow & K = k(x) \\ \downarrow & & \downarrow \\ A[y] & \xrightarrow{\alpha} & K[y] \\ \downarrow \eta & & \downarrow \eta \\ R = A[y]/(y^2 - x(x^2-1)) & \xrightarrow{\phi} & F = K[y]/(y^2 - x(x^2-1)) \end{array} \tag{4.2}$$

commutes. The vertical maps are the natural maps. The horizontal map α exists by Theorem 3.6.2 applied to $A \to K$. Since $\eta\alpha(y^2 - x(x^2 - 1)) = 0$, ϕ exists by Theorem 3.2.15. Using Gauss' Lemma (Theorem 3.7.4), we see that the kernel of $\eta\alpha$ is the principal ideal $(y^2 - x(x^2 - 1))$. Therefore, ϕ is one-to-one.

PROPOSITION 7.4.5. *In the above context, the following are true.*

(1) The quotient field of R is F.
(2) As an A-module, R is free of rank 2. The set $\{1, y\}$ is a free basis. The image of ϕ is $\{p(x) + q(x)y \mid$ where $p(x)$ and $q(x)$ are in $A = k[x]\}$.
(3) The homomorphism $A \to R$ defined by sending x to its image in R is one-to-one.
(4) The automorphism $\sigma \in \mathrm{Aut}_K(F)$ defined by $y \mapsto -y$ restricts to an automorphism $\sigma : R \to R$.
(5) For any $a \in R$, define the norm of a to be $N(a) = a\sigma(a)$. Then $N(1) = 1$, $N : R \to A$, and N is multiplicative.
(6) The map on groups of units $k^* \to R^*$ is an isomorphism. That is, the units of R are precisely the units of k.
(7) x and y are irreducible elements of R.
(8) R is not a unique factorization domain.
(9) R is not a principal ideal domain.

PROOF. (1): By Exercise 3.5.2, we can view the quotient field of R as a subfield of F. In this context, we show that F is equal to the quotient field of R. By Lemma 4.4.5, a $k(x)$-basis for F is $\{1, y\}$. Since $y \in R$ we know y is in the quotient field of R. The quotient field of $k[x]$ is $k(x)$, hence $k(x)$ is in the quotient field of R. A typical element of F is of the form $f(x) + g(x)y$, where $f(x)$ and $g(x)$ are in $k(x)$. Hence a typical element of F is in the quotient field of R.

(2): By Lemma 4.4.5, a K-basis for F is $\{1, y\}$. Therefore, $\{1, y\}$ is linearly independent over A. Since the image of $\eta\alpha$ is generated by polynomials over k in the variables x and y, $\{1, y\}$ is a generating set for the image of ϕ as an A-module. As mentioned in the paragraph that precedes the proposition, ϕ is one-to-one. So $\{1, y\}$ is a generating set for R as an A-module.

(3): The composite map $A \to K \to F$ is one-to-one and factors through R.

(4): Using Theorem 3.6.2, we see that the map $\sigma : A[y] \to A[y]$ defined by $y \mapsto -y$ is an automorphism and maps the principal ideal $(y^2 - x(x^2 - 1))$ onto

itself.

$$\begin{array}{ccc} A[y] & \xrightarrow{\sigma} & A[y] \\ {\scriptstyle\eta}\downarrow & & \downarrow{\scriptstyle\eta} \\ R & \longrightarrow & R \end{array} \tag{4.3}$$

The kernel of $\eta\sigma$ is the principal ideal $(y^2 - x(x^2-1))$. Hence $\sigma : R \to R$ is an automorphism.

(5): Let $a \in R$. By (2), a has a unique representation in the form $a = f+gy$, for polynomials f and g in $A = k[x]$. Then $N(a) = a\sigma(a) = f^2-g^2y^2 = f^2-g^2x(x^2-1)$ is in the image of $A \to R$. Notice that $N : R \to A$ is the restriction of $N_K^F : F \to K$, hence $N(1) = 1$ and $N(ab) = N(a)N(b)$ by Section 5.5.1.

(6): The map $k \to R$ is one-to-one because k is a field. We show that $k^* \to R^*$ is onto. Let $a, b \in R$ and assume $ab = 1$. Then $N(a)N(b) = 1$ in A. But $A^* = k^*$. This proves $N(a) \in k$. By (2), a has a unique representation in the form $a = f+gy$, for polynomials f and g in $A = k[x]$. Then $N(a) = f^2 - g^2x(x^2-1) = u$ for some $u \in k^*$. Then $(f(0))^2 = u$. If $g \neq 0$, then the leading term of f^2 which is even is equal to the leading term of $g^2x(x^2-1)$, which is odd, a contradiction. Therefore, $g = 0$ and $a = f = f(0)$ is in k.

(7): If x is not irreducible, then there is a nontrivial factorization $x = ab$. By (5), we have the factorization $N(x) = x^2 = N(a)N(b)$ in $A = k[x]$. Therefore, $N(a) = x$ up to associates. By (2), a has a representation in the form $a = f+gy$, for polynomials f and g in $A = k[x]$. Then up to associates, $N(a) = f^2-g^2x(x^2-1) = x$. Then $f^2 = g^2x(x^2-1)+x$ which is impossible because the degree of the left hand is even and that of the right hand side is odd. This proves x is not in the image of the norm map $N : R \to A$, hence x is irreducible in R.

If y is not irreducible in R, then there is a nontrivial factorization $y = ab$. By (5), we have the factorization $N(y) = x(x^2-1) = N(a)N(b)$ in $A = k[x]$. Therefore, up to associates, one of $N(a)$ or $N(b)$ is in $\{x, x+1, x-1\}$. The same proof from above shows that $x+1$ and $x-1$ are not in the image of $N : R \to A$. Therefore, y is irreducible in R.

(8): In R we have the identity $y^2 = x(x^2-1)$. By the proof of (7), $N(x) = x^2$ and $N(y) = x(x^2-1)$. Therefore, x and y are not associates of each other. So unique factorization does not exist.

(9): Consider the ideal $\mathfrak{m} = (x, y)$. Then $R/\mathfrak{m} = k[x, y]/(x, y) = k$ is a field, hence $\mathfrak{m}$ is a maximal ideal. If $\mathfrak{m} = (a)$ is principal, then $a \mid x$ and $a \mid y$. Since x and y are irreducible, by Lemma 3.4.5, this implies x and y are associates of each other, a contradiction to (8). □

Now we show how the class group of $R = k[x, y]/(y^2 - x(x^2-1))$ induces an abelian group structure on the cubic curve in k^2 defined by $y^2 = x(x^2-1)$. Set $f(x, y) = y^2 - x(x^2-1)$ and denote by C the cubic curve $Z(f)$ in k^2. To make C into a group, it is necessary to add a point which plays the role of the group identity. The point we adjoin is denoted ∞ and is called *the point at infinity* on C. The group identity in $\mathrm{Cl}(R)$ is the ideal class represented by the unit ideal R. A rational point on C corresponds to an ordered pair (a, b) in k^2 such that $f(a, b) = 0$. The ideal $I(a, b) = (x-a, y-b)$ is a maximal ideal in R. The residue field $R/I(a, b)$ is k. To define the group law on the points of C, we embed $C \cup \{\infty\}$ into $\mathrm{Cl}(R)$ by mapping a point $(a, b) \in C$ to the ideal class represented by the maximal ideal

$I(a,b) = (x-a, y-b)$, and map ∞ to the ideal class represented by R. To make $C \cup \{\infty\}$ into a group, we show that the image of the function $C \cup \{\infty\} \to \mathrm{Cl}(R)$ is a subgroup of $\mathrm{Cl}(R)$. Therefore, the group $\mathrm{Cl}(R)$ induces a group law on the cubic curve C. It is customary to write the binary operation on $C \cup \{\infty\}$ using additive notation. The binary operation in the ideal class group $\mathrm{Cl}(R) = \mathrm{Frac}(R)/\mathrm{Prin}(R)$ is written multiplicatively.

This section is not self-contained. References are given for the proof of Proposition 7.4.6.

PROPOSITION 7.4.6. *In the above context, the following are true.*

(1) The ring $R = k[x,y]/(y^2 - x(x^2-1))$ is integrally closed in F.
(2) If (a_1, b_1) and (a_2, b_2) are two distinct points on C, then the maximal ideals $(x-a_1, y-b_1)$ and $(x-a_2, y-b_2)$ represent distinct classes in $\mathrm{Cl}(R)$.

PROOF. (1): A good technique for proving this is to use the so-called jacobian criterion to show that R is regular. For example, see [**9**, Theorem 15.6.5].

(2): For a proof, see [**14**, see Exercise I.6.2 and Example II.6.10.1]. □

PROPOSITION 7.4.7. *In the above context, let $I : C \cup \{\infty\} \to \mathrm{Cl}(R)$ be the function which maps ∞ to the ideal class represented by R and assigns a point (a,b) to the ideal class represented by the maximal ideal $I(a,b) = (x-a, x-b)$. Then the following are true.*

(1) The function $I : C \cup \{\infty\} \to \mathrm{Cl}(R)$ is one-to-one.
(2) The image of I is a subgroup of $\mathrm{Cl}(R)$.

PROOF. (1): This is Proposition 7.4.6 (2).

(2): By Lemma 2.2.2 it suffices to show that if A and B are two elements in the image of I, then A^{-1} and AB are in the image of I.

Let $(a,b) \in C$. We show that $I(a,b)$ and $I(a,-b)$ are inverses of each other in $\mathrm{Cl}(R)$. It suffices to show that the product $(x-a, y-b)(x-a, y+b)$ is equal to the principal ideal $(x-a)$. If $b = 0$, this is Exercise 7.4.10. If $b \neq 0$, then the ideals are comaximal, hence their product is equal to their intersection. Then clearly $(x-a) \subseteq (x-a, y-b)(x-a, y+b)$. To show the reverse inclusion, it suffices to show $y^2 - b^2 \in (x-a)$. This follows from the identities $y^2 = x^3 - x$ and $b^2 = a^3 - a$.

Now let $P_1 = (a_1, b_1)$ and $P_2 = (a_2, b_2)$ be two points in C. We show that in $\mathrm{Cl}(R)$ the product $(x-a_1, y-b_1)(x-a_2, y-b_2)$ is in the image of the function I. If $P_1 \neq P_2$, let L be the line in k^2 through P_1 and P_2. If $P_1 = P_2$, let L be the tangent line to C at P_1. If L is parallel to the $x = 0$ line, then by the above argument we know the classes of the ideals $I(a,b)$ and $I(a,-b)$ are inverses of each other in $\mathrm{Cl}(R)$. So we can assume the equation for L is of the form $y - \ell(x) = 0$. To compute the intersection of the line L and the cubic C, solve the system of equations $y = \ell(x)$ and $y^2 = x^3 - x$. Eliminating y, we get the cubic equation $\ell^2 - x^3 + x = 0$. Two roots are a_1 and a_2. So there is a third root, call it a_3. Set $b_2 = \ell(a_3)$. Then $f(a_3, b_3) = 0$. The only maximal ideals of R that contain $y - \ell(x)$ are $I_1 = (x-a_1, y-b_1)$, $I_2 = (x-a_2, y-b_2)$, $I_3 = (x-a_3, y-b_3)$. By Exercise 4.6.19,

$$\begin{aligned} R/(y-\ell(x)) &\cong k[x]/(x-a_1)(x-a_2)(x-a_3) \\ &\cong R/I_1I_2I_3. \end{aligned}$$

In R the prime factorization of the principal ideal $(y - \ell(x))$ is $I_1 I_2 I_3$. Hence, I_3 is in the image of I. The product of ideal classes in $\mathrm{Cl}(R)$ restricts to a product on the image of I. □

EXAMPLE 7.4.8. In the above context, here is the geometric interpretation of the group law on C. From Proposition 7.4.7 (2), the inverse of a point $P = (a, b)$ on C is the point $(a, -b)$. To multiply two points $P_1 = (a_1, b_1)$ and $P_2 = (a_2, b_2)$, let L be the line through P_1 and P_2. If L is vertical, then $P_1 P_2 = \infty$. Otherwise, let $P_3 = (a_3, b_3)$ be the third point on the intersection $L \cap C$. Then $P_1 P_2$ and P_3 are inverses. Hence $P_1 P_2$ is the point $(a_3, -b_3)$.

4.3. Exercises.

EXERCISE 7.4.9. Let k be a field. Assume the characteristic of k is not 2 or 3 and that k contains a primitive sixth root of unity denoted ζ_6.

(1) Show that $k(x)$ is a cyclic Galois extension of $k(x^6)$ of degree 6 (in other words, a Kummer extension). Let $G = \langle \sigma \rangle$ be the Galois group. Determine the lattice of subfields and lattice of subgroups guaranteed by the Fundamental Theorem of Galois Theory.
(2) Show that G acts on $k[x]$ and the fixed subring is $k[x^6]$. Determine the lattice of fixed subrings of $k[x]$ corresponding to the subgroups of G.
(3) As in Exercise 3.6.21, let $R = k[x^2, x^3]$. Then the quotient field of R is $k(x)$. We say that R is *birational to* $k[x]$. Determine the subgroup of G that fixes R point-wise (that is, the stabilizer of R in G).
(4) True or False?
    (a) $k[x]$ is a free $k[x^2]$-module.
    (b) $k[x]$ is a free $k[x^2, x^3]$-module.
    (c) $k[x^2, x^3]$ is a free $k[x^2]$-module.

EXERCISE 7.4.10. In the context of Proposition 7.4.5, consider the maximal ideals $\mathfrak{m}_0 = (x, y)$, $\mathfrak{m}_{-1} = (x, y + 1)$, and $\mathfrak{m}_1 = (x, y - 1)$. Show that in $\mathrm{Cl}(R)$ each of these ideals represent a class of order two. That is, show that the ideals $\mathfrak{m}_0^2$, $\mathfrak{m}_{-1}^2$, and $\mathfrak{m}_1^2$ are principal.

EXERCISE 7.4.11. Let $k = \mathbb{Z}/5$ be the field of order 5 and C be the cubic curve in k^2 defined by $y^2 - x(x^2 - 1)$. Show that the group $C \cup \infty$ has order 8 and the group invariants are 2, 4.

EXERCISE 7.4.12. Let $k = \mathbb{Z}/17$ be the field of order 17 and C be the cubic curve in k^2 defined by $y^2 - x(x^2 - 1)$. Show that the group $C \cup \infty$ has order 16 and the group invariants are 4, 4. For this exercise, a computer algebra system such as [**28**] will be helpful.

Hints to Selected Exercises

Chapter 1

Exercise 1.1.20. Use Exercise 1.1.14.

Exercise 1.2.17 (3). Use vectors in $\mathbb{Z}^2$. Let $P = (x, y)$, $P_0 = (x_0, y_0)$, $D = (-b/d, a/d)$, $U = (u, v)$. Use (1) to write $P - P_0 = tD + sU$. Take the dot product of both sides with $D^\perp$ to show that $s = 0$.

Exercise 1.2.18. Part (3). Show that the line $ab - a - b = ax + by$ contains the two lattice points $(-1, a - 1)$ and $(b - 1, -a)$. Part (5). For sake of contradiction assume $ab - a - b < n < ab$ and n is not in L. Show that there exists an ordered pair (x_1, y_1) such that $n = ax_1 + by_1$, (x_1, y_1) is in Quadrant IV and $(x_1 - b, y_1 + a)$ is in Quadrant II. Show that (x_1, y_1) is not in the parallelogram with vertices $(b, 0)$, $(0, a)$, $(-1, a - 1)$, $(b - 1, -1)$. Show that this is impossible.

Exercise 1.2.21 (5). Use The Binomial Theorem, Exercise 1.1.18.

Exercise 1.2.25 (2). Start with a solution to the linear diophantine equation $a = bx + cy$. Apply the Division Algorithm to write $y = bq + r$, where $0 \leq r < b$. Take $e = r$ and $f = x + cq$.

Chapter 2

Exercise 2.1.24. Show that the function defined by $x \mapsto x^{-1}$ is an isomorphism from G to G^o.

Exercise 2.3.45. Apply The Fundamental Theorem on Cyclic Groups Theorem 2.3.27.)

Exercise 2.3.47. If $\sigma = (123 \cdots n)$, show that G' is the cyclic group generated by σ^2.

Exercise 2.4.35. Apply Exercise 2.4.28.

Exercise 2.4.37. Apply Exercise 2.4.28.

Exercise 2.4.38. Apply Exercise 2.4.28.

Exercise 2.4.39. If x and y are conjugates, then $|N_G(x)| = |N_G(y)|$.

Exercise 2.5.25. Part (2) (a). See Exercise 2.3.17 (6). Part (4). A homomorphism θ corresponds to an element of the set of ordered pairs $\{(a, b) \in G \times G \mid ab = ba\} = \{(a, b) \in G \times G \mid b \in N_G(a)\}$. Apply Exercise 2.4.39.

Exercise 2.5.30. Let A be an abelian group written additively. Let $G = (A \times A) \rtimes \langle \theta \rangle$ be the nonabelian group of Exercise 2.5.29. Let $\sigma : A \to A$ be the automorphism of A defined by $\sigma(x) = -x$. Show that the quotient $G/Z(G)$ is isomorphic to the semidirect direct product $A \rtimes \langle \sigma \rangle$ of Proposition 2.4.19.

Exercise 2.6.22. Suppose $a \in A$, $b \notin A$, and σ fixes $\mathbb{N}_n - A$. Look at $(ab)\sigma(ab)$.

Exercise 2.6.24. Apply Exercise 2.3.43. Show that every automorphism of S_3 is an inner automorphism.

Exercise 2.6.25. Apply Proposition 1.5.3 to show that the map $\theta : S_n \to \mathrm{GL}_n(\mathbb{Z})$ defined by $\theta(\sigma) = P_\sigma$ is a homomorphism.

Exercise 2.8.12. Apply Theorem 2.3.27, Theorem 2.8.7, Exercise 2.8.11, and Theorem 2.5.2. Prove the following implications: (1) implies (2). (1) implies (4). (2) implies (3). (3) implies (2). (4) is equivalent to (5). (4) implies (3). (2) implies (1).

Exercise 2.8.14. Apply Exercises 2.8.13 and 2.4.24.

Exercise 2.8.16. First prove this if G is abelian. If G is nonabelian use induction on n. Consider the group $G/Z(G)$.

Exercise 2.8.19. Use Proposition 2.7.8 to reduce the case where G is a finite cyclic p-group.

Exercise 2.10.19 (1). Apply Example 2.3.36.

Chapter 3

Exercise 3.1.16 (1). Apply Theorem 2.3.30.

Exercise 3.1.17 (1). Apply Theorem 2.3.30.

Exercise 3.1.20. Find matrices that play the roles of i and j.

Exercise 3.2.31. Apply Exercise 1.2.21.

Exercise 3.2.33. Part (1). Apply Exercise 2.1.23. Part (2). $R-(0)$ is a monoid.

Exercise 3.2.34 (3). Use multiplication by the various e_{ij}.

Exercise 3.2.44 (3). For all p sufficiently large, if $x_1, \ldots, x_p$ are elements of $A \cup B$, show that $x_1 \cdots x_p = 0$.

Exercise 3.2.46. Apply Exercise 2.1.24 and Example 3.2.2 (3).

Exercise 3.2.53. Let $\lambda : R \to \mathrm{Hom}(I, I)$ be the homomorphism of rings from Exercise 3.2.52. Apply Example 3.1.11 and Exercises 3.1.16 and 3.1.17 to show that λ is onto. Lastly, apply Corollary 3.2.16.

Exercise 3.3.15. Show that $1-e$ is a central idempotent. Show that e and $1-e$ are orthogonal idempotents. Take $J = R(1-e)$.

Exercise 3.3.17. Use Exercise 3.3.23.

Exercise 3.3.26. Show that Ring (1) is isomorphic to Ring (2) by the mapping $a + b\sigma \mapsto \begin{pmatrix} a & b \\ b & a \end{pmatrix}$. Show that Ring (1) is isomorphic to Ring (3) by the mapping $a + b\sigma \mapsto (a - b, a + b)$.

Exercise 3.4.30. Given $x \notin P$, apply Corollary 3.4.9 to show that the ideal $Rx + P$ is the unit ideal.

Exercise 3.5.5. Apply Theorem 3.2.21.

Exercise 3.5.7. For (1) implies (2) use Exercise 3.2.43). For (3) implies (1) use Exercise 3.1.26 to show the set $\mathfrak{m} = R - R^*$ is an ideal.

Exercise 3.5.8. Use Exercises 3.3.15 and 3.5.7.

Exercise 3.6.21. Part (2). x^2 and x^3 are both irreducible. Part (3). Neither x^2 nor x^3 is prime.

Exercise 3.6.31. Let $x \in P - (0)$. Show that P contains at least one prime divisor of x.

Exercise 3.6.35 (1). Apply Exercise 3.2.31.

Exercise 3.7.13. For Parts (5) and (6), apply Exercise 3.6.35.

Exercise 3.7.19 (2). Apply Example 3.6.6.

Chapter 4

Exercise 4.1.18. Use Theorem 4.1.12 to show that there is a natural homomorphism of rings $\mathrm{Hom}(M, M) \to \mathrm{Hom}(M/IM, M/IM)$. Use Theorem 3.2.15 to show that there is a homomorphism of rings $R/I \to \mathrm{Hom}(M/IM, M/IM)$. Apply Lemma 4.1.2.

Exercise 4.1.22 (5). Apply Corollary 2.2.21.

Exercise 4.3.9 (1). To show that M is finitely generated, use Lemma 4.3.1, induction, and Theorem 4.3.4.

Exercise 4.3.15. By Proposition 3.2.27, R contains a maximal ideal. Let $\mathfrak{m}$ be a maximal ideal in R and consider $F/\mathfrak{m}F$ as a vector space over $R/\mathfrak{m}$.

Exercise 4.3.19 (2). Apply Exercise 4.3.9 and Corollary 4.1.14 (2).

Exercise 4.3.20. Part (3). Consider the A-submodules Ae_{11} and A.

Part (5). This is a challenge and may involve results not proven in this book. The interested reader is referred to [**9**, Section 7.4.1].

Exercise 4.4.25. Part (1). Since x and y are nilpotent, they necessarily belong to any prime ideal. Part (4). A basis for R over k is $1, x, y$.) Part (5). Map x to $\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, and y to $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

Exercise 4.4.26. Part (1). Since x and y are nilpotent, they necessarily belong to any prime ideal. Part (4). A basis for R over k is $1, x, y$.) Part (5). Map x to $\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, and y to $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

Exercise 4.4.27. Apply Corollary 3.3.13 and Exercise 4.3.9.

Exercise 4.5.29. Given a generating set $\{x_1, \ldots, x_n\}$ for M over R, let $\Sigma : R^{(n)} \to M$ be defined by $\Sigma(r_1, \ldots, r_n) = \sum_{i=1}^{n} r_i x_i$. Given a dual basis $\{(x_i, f_i) \mid 1 \leq i \leq n\}$, define $\theta : M \to R^{(n)}$ by $\theta(x) = (f_1(x), \ldots, f_n(x))$. Show that F is isomorphic to $\mathrm{im}\,\theta$ and $\mathrm{im}\,\theta$ is a direct summand of $R^{(n)}$. For the converse, let $\pi_i : R^{(n)} \to R$ be the natural projection map onto coordinate i. If M is projective,

let $\psi : M \to R^{(n)}$ be a map satisfying Proposition 4.2.19 (2). Define $f_i \in M^*$ by $f_i(x) = \pi_i(\psi(x))$.

Exercise 4.6.18. Show that if (a) is a principal ideal in R containing (π^e), then a is an associate of π^k for some integer k such that $0 \leq k \leq e$.

Exercise 4.6.20 (4). Apply Exercise 2.8.12.

Exercise 4.6.22. Assume $\mathrm{char}(R) = n > 0$ and that n is divisible by at least two distint primes. Let C denote the canonical subring of R isomorphic to $\mathbb{Z}/n$. Apply Exercise 4.6.19 to C and use Exercise 3.5.7 (3) to show that R is not a local ring.

Chapter 5

Exercise 5.1.21. Apply Theorem 3.7.10 to show that $y^4 - \alpha(4y^3 - 1)$ is an irreducible polynomial in $K[y]$.

Exercise 5.2.23. Apply Corollary 5.1.7.

Exercise 5.2.25. Apply Corollary 5.1.7 and Lemma 5.2.7.

Exercise 5.3.25. Apply Exercise 3.2.49.

Exercise 5.3.32 (1). Apply Exercise 5.2.25.

Exercise 5.3.48. Show that $f(x + 1)$ and $f(x)$ cannot both be equal to the irreducible polynomial of α over k.

Exercise 5.5.20. Show that $T_k^F D(x) = 0$. Compute the rank and nullity of the linear transformations D and T_k^F.

Chapter 6

Exercise 4.5.28. Show that the left regular representation $\lambda : A \to \mathrm{Hom}_k(M, M)$ is a homomorphism of k-algebras. Apply Proposition 4.5.7 and Exercise 3.2.34.

Exercise 6.1.20 (1). Apply Theorem 4.4.15.

Exercise 6.1.23. Use the function $\tau : L \to D$ of Example 3.2.12.

Exercise 6.1.24. If S is an invertible matrix, show that $S^{-1}f(A)S = f(S^{-1}AS)$. Apply Theorem 6.1.18, Exercise 4.4.20, and Exercise 6.1.23.

Exercise 6.1.26. Apply Proposition 4.5.9 and Theorem 6.1.12.

Exercise 6.3.20. If $A = (a_1, \ldots, a_n)$ is written in columnar form, then $b = x_1a_1 + \cdots + x_na_n$. Use the multilinear and alternating properties when computing $\det(B_i)$.

Exercise 6.3.25. Apply Exercise 4.5.25, Lemma 3.5.1, and Theorem 6.3.13 (2).

Exercise 6.3.26 (2). First show $\mathrm{trace}(\alpha e_{ij}) = \mathrm{trace}(e_{ij}\alpha)$ if e_{ij} is an elementary matrix and α is arbitrary.

Exercise 6.3.36. Part (1). Use Exercise 6.3.17.
Part (2). Use Lemma 2.6.9, Exercise 6.3.18, and Part (1).
Part (3). Use Part (2) and Exercise 6.3.19.

Exercise 6.3.37. If $|\sigma| = d$, then $\mathrm{min.poly}(P_\sigma)$ divides $x^d - 1$ which has no repeated roots. Use Exercise 6.3.36 and Theorem 6.3.13 (3).

Exercise 6.3.38. Use determinants.

Chapter 7

Exercise 7.1.25. Apply Exercise 7.1.24.

Exercise 7.1.26. Apply Exercise 7.1.25.

Exercise 7.1.27. If $\alpha^{-1} \in A$, then there exist $n \geq 1$ and $r_i \in R$ such that $\alpha^{-n} + r_{n-1}\alpha^{1-n} + \cdots + r_1\alpha^{-1} + r_0 = 0$. Use this to prove that $\alpha^{-1} \in R$.

Exercise 7.3.13. Use Proposition 3.2.24 and induction on n.

Exercise 7.3.14. Part (3). Use Lemma 7.2.3. Part (4). Show that if F is free, then an S-basis is linearly independent over L.

Exercise 7.3.17. Part (1). Prove this directly. If $\alpha \in K$ is integral over $R[f^{-1}]$, then for some $r > 0$, αf^r is in R. Part (2). Use Lemma 7.3.10 and Part (1).

Exercise 7.3.18. The only maximal ideal of S that contains P^e is P. Use Theorem 7.3.6 and Corollary 3.2.18.

Exercise 7.3.19. Part (1). Use Corollary 3.3.13 and Exercise 7.3.18. Part (2). Apply Exercises 7.3.18 and 3.3.25. Part (3). Use Theorem 3.3.4.

Exercise 7.3.20. By Exercise 7.3.19, $S/S\alpha$ is a principal ideal ring. Let $\beta \in I - S\alpha$ be a generator for $I/S\alpha$.

Exercise 7.3.23. By Exercises 7.3.19 and 7.3.20, there exists $\beta \in I - IJ$ such that $IJ + S\beta = I$. By Exercise 7.3.22, $S\beta = IC$. From $IJ + IC = I$, conclude that $J + C = S$.

Exercise 7.3.24. Let $\phi : I \oplus J \to S$ be defined by $\phi(x, y) = x - y$. Show that ϕ is onto and the kernel is isomorphic to IJ. Use Propositions 4.2.19 and 4.2.8.

Exercise 7.3.25. Part (2). Pick α such that αE is a proper ideal in S. Pick δ such that δF^{-1} is a proper ideal in S. Apply Exercise 7.3.23 to get an ideal C in S and $\gamma \in S$ such that $S\gamma = \delta F^{-1}C$ and $\alpha E + C = S$. Set $\beta = \gamma\delta^{-1}$. Then $\alpha E + \beta F = S$. Part (3). Use Parts (1) and (2) and Exercise 7.3.24.

Exercise 7.4.9 (4). $k[x^2]$ is a PID, in fact it is a euclidean domain. Section 4.6 applies.

Exercise 7.4.11. Use the geometric method of Example 7.4.8.

Exercise 7.4.12. Use the geometric method of Example 7.4.8.

Acronyms

ACC Ascending Chain Condition
DCC Descending Chain Condition
GCD Greatest Common Divisor
PID Principal Ideal Domain
UFD Unique Factorization Domain

Bibliography

[1] A. A. Albert, *Cyclic fields of degree p^n over F of characteristic p*, Bull. Amer. Math. Soc. **40** (1934), no. 8, 625–631. MR 1562919

[2] Emil Artin, *Galois Theory*, University of Notre Dame, Notre Dame, Ind., 1942. MR 4,66k

[3] Emil Artin and Otto Schreier, *Eine Kennzeichnung der reell abgeschlossenen Körper*, Abh. Math. Sem. Univ. Hamburg **5** (1927), no. 1, 225–231. MR 3069477

[4] Michael Artin, *Algebra*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1991. MR 1129886

[5] M. F. Atiyah and I. G. Macdonald, *Introduction to commutative algebra*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1969. MR 0242802 (39 #4129)

[6] Allan Clark, *Elements of abstract algebra*, Dover Publications, Inc., New York, 1971, unabridged and Corrected republication of the work first published by Wadsworth Publishing Company, Belmont, California, in 1971.

[7] Charles W. Curtis and Irving Reiner, *Representation theory of finite groups and associative algebras*, Pure and Applied Mathematics, Vol. XI, Interscience Publishers, a division of John Wiley & Sons, New York-London, 1962. MR 0144979 (26 #2519)

[8] David S. Dummit and Richard M. Foote, *Abstract algebra*, third ed., John Wiley & Sons Inc., Hoboken, NJ, 2004. MR 2286236 (2007h:00003)

[9] Timothy J. Ford, *Abstract algebra*, pre-preprint. Available at: `https://tim4datfau.github.io/Timothy-Ford-at-FAU/preprints/Algebra_Book.pdf`.

[10] ———, *Separable algebras*, Graduate Studies in Mathematics, vol. 183, American Mathematical Society, Providence, RI, 2017. MR 3618889

[11] ———, *Commutative algebra*, Open Math Notes, American Mathematical Society, Reference # OMN:202409.111443, 2024, available at the URL: `https://www.ams.org/open-math-notes/omn-view-listing?listingId=111443`.

[12] Évariste Galois, *Écrits et mémoires mathématiques*, Les Grands Classiques Gauthier-Villars. [Gauthier-Villars Great Classics], Éditions Jacques Gabay, Paris, 1997, Édition critique intégrale des manuscrits et publications. [Integral critical edition of the manuscripts and publications], With a preface by Jean Dieudonné, Edited, with notes and commentary by Robert Bourgne and Jean-Pierre Azra, Reprint of the second (1976) edition. MR 1452597

[13] Vasily Golyshev and Jan Stienstra, *Fuchsian equations of type DN*, Commun. Number Theory Phys. **1** (2007), no. 2, 323–346. MR 2346574

[14] Robin Hartshorne, *Algebraic geometry*, Springer-Verlag, New York, 1977. MR 0463157 (57 #3116)

[15] I. N. Herstein, *Topics in algebra*, second ed., Xerox College Publishing, Lexington, Mass., 1975. MR 0356988 (50 #9456)

[16] Thomas W. Hungerford, *Algebra*, Graduate Texts in Mathematics, vol. 73, Springer-Verlag, New York, 1980, Reprint of the 1974 original. MR 600654 (82a:00006)

[17] John L. Kelley, *General topology*, Springer-Verlag, New York-Berlin, 1975, Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.], Graduate Texts in Mathematics, No. 27. MR 0370454 (51 #6681)

[18] Serge Lang, *Algebra*, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965. MR 33 #5416

[19] J. H. Maclagan-Wedderburn, *A theorem on finite algebras*, Trans. Amer. Math. Soc. **6** (1905), no. 3, 349–352. MR 1500717

[20] James H. McKay, *Another proof of Cauchy's group theorem*, Amer. Math. Monthly **66** (1959), 119. MR 98777

[21] Eugen Netto, *Ueber die Irreductibilität ganzzahliger ganzer Functionen*, Math. Ann. **48** (1896), no. 1-2, 81–88. MR 1510925

[22] Paulo Ribenboim, *Algebraic numbers*, Pure and Applied Mathematics, vol. Vol. 27, Wiley-Interscience [A Division of John Wiley & Sons, Inc.], New York-London-Sydney, 1972. MR 340212

[23] Joseph J. Rotman, *Advanced modern algebra*, Prentice Hall Inc., Upper Saddle River, NJ, 2002. MR 2043445 (2005b:00002)

[24] ———, *Advanced modern algebra. Part 1*, third ed., Graduate Studies in Mathematics, vol. 165, American Mathematical Society, Providence, RI, 2015. MR 3443588

[25] ———, *Advanced modern algebra. Part 2.*, third ed., Graduate Studies in Mathematics, vol. 180, American Mathematical Society, Providence, RI, 2017, With a foreword by Bruce Reznick. MR 3677125

[26] David Singmaster and D. M. Bloom, *Problems and Solutions: Solutions of Elementary Problems: E1648*, Amer. Math. Monthly **71** (1964), no. 8, 918–920. MR 1532917

[27] Michael Spivak, *Calculus*, fourth ed., Publish or Perish, Inc., PMB 377, 1302 Waugh Drive, Houston, Texas 77019, 2008.

[28] The Sage Development Team, *Sagemath, the Sage Mathematics Software System (Version 9.5)*, The Sage Development Team, 2022-01-30, `http://www.sagemath.org`.

[29] A. R. Wadsworth, *Problems in abstract algebra*, Student Mathematical Library, vol. 82, American Mathematical Society, Providence, RI, 2017. MR 3643210

[30] Helmut Wielandt, *Ein Beweis für die Existenz der Sylowgruppen*, Arch. Math. (Basel) **10** (1959), 401–402. MR 147529

[31] Wikipedia contributors, *Extended euclidean algorithm — Wikipedia, the free encyclopedia*, `https://en.wikipedia.org/w/index.php?title=Extended_Euclidean_algorithm&oldid=1135569411`, 2023, [Online; accessed 30-January-2023].

[32] Ernst Witt, *Über die kommutativität endlicher schiefkörper*, Abh. Math. Sem. Univ. Hamburg **8** (1931), no. 1, 413. MR 3069571

[33] Max Zorn, *A remark on method in transfinite algebra*, Bull. Amer. Math. Soc. **41** (1935), no. 10, 667–670. MR 1563165

Glossary of Notations

(X)	ideal generated by X, 123
(X)	submodule generated by X, 171
1_X	identity map on X, 14
2^X	power set of X, 13
$[G:1]$	order of the group G, 36
$[G:H]$	index of the subgroup H in the group G, 44
$[x]$	congruence class of x modulo m, 21
$\mathrm{annih}_R(M)$	annihilator of M in R, 170
$\mathrm{Aut}(G)$	group of automorphisms of G, 51
$\mathrm{Aut}_R(A)$	automorphism group of an R-algebra A, 190
$\bigcap_{i\in I} X_i$	intersection of a family of sets, 13
$\bigcup_{i\in I} X_i$	union of a family of sets, 13
$\bigoplus_{i\in I} M_i$	direct sum of a family of modules, 177
$\binom{n}{k}$	binomial coefficient, 17
$\mathrm{char}(R)$	characteristic of R, 123
$\mathrm{Cl}(R)$	class group of R, 313
$\deg(f)$	degree of a polynomial f, 153
δ_{ij}	Kronecker δ function, 180
$\mathrm{diag}(a_1,\ldots,a_n)$	diagonal matrix, 31
$\dim_D(V)$	dimension of the D-vector space V, 186
$\ell(G)$	length of a composition series of G, 113
$\emptyset$	empty set, 13
$\mathrm{Frac}\, R$	set of all fractional ideals of R in K, 310
$\gcd(a_1,\ldots,a_n)$	greatest common divisor of $\{a_1,\ldots,a_n\}$, 20
$\mathrm{GL}_n(F)$	general linear group of n-by-n matrices over F, 39
$\mathrm{Hom}(A,A)$	endomorphism ring of an abelian group, 101
$\mathrm{Hom}_R(M,N)$	group of R-module homomorphisms, 174
$\mathrm{im}\,(f)$	image of a homomorphism f, 122
$\mathrm{Inn}(G)$	group of inner automorphisms of G, 62
$\ker(f)$	kernel of a homomorphism f, 122
λ_n	left multiplication by n map, 53
$\langle X \mid Y\rangle$	group defined by generators X and relations Y, 80
$\langle X\rangle$	subgroup generated by X, 43
$\lceil x\rceil$	ceiling of x, 19
$\mathrm{lcm}(a,b)$	least common multiple, 22
$\lfloor x\rfloor$	floor of x, 19
$\lvert a\rvert$	order of the element a, 36
$\lvert X\rvert$	cardinality of the set X, 16
$\mathrm{Map}(X)$	set of all functions from X to X, 17

$\mathbb{C}$	complex numbers, 14
$\mathbb{N}$	natural numbers, 14
$\mathbb{N}_n$	$\{1, 2, \ldots, n\}$, 15
$\mathbb{Q}$	rational numbers, 14
$\mathbb{Q}/\mathbb{Z}$	rational numbers modulo the integers, 54
$\mathbb{Q}/\mathbb{Z}(p)$	p-torsion subgroup of $\mathbb{Q}/\mathbb{Z}$, 209
$\mathbb{R}$	real numbers, 14
$\mathbb{Z}$	integers, 14
$\mathbb{Z}/(m)$	integers modulo m, 21
$\text{min. poly}_k(\alpha)$	minimal polynomial of α over k, 191
μ	the group of all roots of unity in $\mathbb{C}$, 54
μ_n	the group of nth roots of unity, 54
$\text{Perm}(X)$	set of all permutations of X, 34
$\text{PGL}_n(F)$	projective general linear group, 60
$\phi(n)$	Euler ϕ-function, 23
π^n	nth power map, 54
$\text{Prin}\, R$	set of all principal fractional ideals of R in K, 312
$\prod_{i\in I} M_i$	direct product of a family of modules, 177
$\prod_{i\in I} R_i$	direct product of a family of rings, 133
$\prod_{i\in I} X_i$	product of a family of sets, 28
ψ^*	the dual of $\psi \in \text{Hom}_R(M, N)$, 198
$\text{Rad}_R(0)$	nil radical of R, 131
$\text{Rank}(M)$	rank of the module M, 180
$\text{sign}(\sigma)$	sign of a permutation, 85
$\text{SL}_n(F)$	special linear group, 60
$\text{trace}(\alpha)$	trace of a matrix, 299
$\text{trace}(\phi)$	trace of a homomorphism, 299
$\text{tr. deg}_k(F)$	transcendence degree of F/k, 261
$\text{Units}(R)$ or R^*	group of units in the ring R, 116
$a + I$	left coset of I containing a, 126
$a \mid b$	a divides b, 20
A^n	powers of an ideal, 132
A^T	transpose of a matrix, 198
A_n	alternating group on n letters, 84
D_n	dihedral group of order $2n$, 37
e_{ij}	elementary matrix, 118
$F(X)$	free group on the set X, 79
f'	formal derivative of the polynomial f, 157
$G \cong G'$	G is isomorphic to G', 36
G/H	set of all left cosets of G modulo H, 44
G'	commutator subgroup, 62
G^o	opposite group, 40
$H \rtimes K$	semidirect product of H and K, 69
$H\backslash G$	set of all right cosets of G modulo H, 44
$I : J$	ideal quotient, 131
$I_1 + I_2 + \cdots + I_n$	sum of ideals, 134
$I_1 \oplus I_2 \oplus \cdots \oplus I_n$	internal direct sum of ideals, 134
$k(x)$	field of rational functions over k in the variable x, 155

$M(\phi, X, Y)$	matrix of ϕ with respect to the bases X, Y, 196
$M(\pi)$	submodule of M annihilated by powers of π, 205
M/S	factor module of M modulo S, 172
M^*	$\mathrm{Hom}_R(M, R)$, the dual module, 198
$M_1 \oplus M_2 \oplus \cdots \oplus M_n$	direct sum of modules, 178
$M_n(R)$	ring of n-by-n matrices over R, 116
$M_{nm}(R)$	set of all n-by-m matrices over R, 196
$N \trianglelefteq G$	N is a normal subgroup of G, 50
$N_G(X)$	normalizer of X in G, 67
$o(G)$	order of the group G, 36
Q_8	quaternion eight group, 38
$R(G)$	group ring, 116
R/I	residue class ring, 126
$R[x]$	ring of polynomials over R in the variable x, 152
R^I	the free R-module on the index set I, 180
R^o	opposite ring of R, 117
$R^{(n)}$	free R-module of rank n, 180
$R_1 \oplus R_2 \oplus \cdots \oplus R_n$	external direct sum of a finite family of rings, 134
S_n	symmetric group on n letters, 16
U_n	units modulo n, 23
$X = Y$	equality of sets, 13
$X \cap Y$	intersection of sets, 13
$X \cup Y$	union of sets, 13
$x \equiv y \pmod{H}$	x is congruent to y modulo H, 43
$x \equiv y \pmod{m}$	x is congruent to y modulo m, 21
$x \in X$	x is an element of X, 13
$X \subseteq Y$	X is a subset of Y, 13
$X \times Y$	product of sets, 13
$X_1 \cap \cdots \cap X_n$	intersection of a family of sets, 14
$X_1 \cup \cdots \cup X_n$	union of a family of sets, 14
$X_1 \times \cdots \times X_n$	product of a family of sets, 14
xH	left coset of x modulo H, 44
$Y - X$	complement of a set, 13
$Z(A)$	center of a ring A, 117
$Z(G)$	center of a group G, 58

Index