Lectures on Abstract Algebra

Preliminary Version

Richard Elman

Department of Mathematics,
University of California,
Los Angeles, CA 90095-1555, USA

Contents

Part 1

Preliminaries

1. Introduction

In this introduction, we introduce some of the notations and definitions that we shall use throughout by means of investigating a few mathematical statements. Consider the following statements:

Statements 1.1.

(1) The integer $2^{43112609}-1$, which has 12 978 189 digits, is a prime.
(2) There exist infinitely many (rational) primes.
(3) Let

$$\pi(x) := \text{the number of positive primes} \leq x,$$

then

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log x} = 1.$$

(4) $\sqrt{2} \notin \mathbb{Q}$, where $\mathbb{Q}$ is the set of rational numbers.
(5) The real number π is not "algebraic" over $\mathbb{Q}$.
(6) There exist infinitely many real numbers [even "many more" than the number of elements in $\mathbb{Q}$] not algebraic over $\mathbb{Q}$.

To look at these statements, we need some definitions. Let $a, b \in \mathbb{Z}$, where $\mathbb{Z}$ is the set of integers. We say that a *divides* b (in $\mathbb{Z}$) if

$$\text{there exists an integer } n \text{ such that } b = an, \textit{i.e.,}\ \frac{b}{a} \in \mathbb{Z}.$$

We write

$$a \mid b\,(\text{ in } \mathbb{Z}).$$

For example, $3 \mid 12$ as $12 = 4 \cdot 3$ but $5 \nmid 12$ in $\mathbb{Z}$ where $\nmid$, means does not divide.

An integer p is called a *prime* if $p \neq 0, \pm 1$ and

$$n \mid p \text{ in } \mathbb{Z} \text{ implies that } n \in \{1, -1, p, -p\},\ \textit{i.e.,}\ n = \pm 1, \pm p.$$

For example, $2, 3, 7, 11, \ldots$ are prime. [The prime 2 is actually the "oddest prime of all"!] We shall use that every integer $n > 1$ is divisible by some prime (to be proven later). A complex number x is called *irrational* if x is not a rational number, i.e., $x \in \mathbb{C} \setminus \mathbb{Q} := \{z \in \mathbb{C} \mid z \notin \mathbb{Q}\}$, where $\mathbb{C}$ is the set of complex numbers.

A complex number x is called *algebraic* (over $\mathbb{Q}$) if there exists a nonzero *polynomial* $f(t) = a_n t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$ with $a_0, \ldots, a_n \in \mathbb{Q}$ (some n) not all zero satisfying $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$. We shall write f for $f(t)$ where t *always represents a variable* and $f(x)$ means plug x into f. We let

$$\mathbb{Q}[t] := \text{ the set of polynomials with rational coefficients.}$$

So x is algebraic over $\mathbb{Q}$, if there exists $0 \neq f \in \mathbb{Q}[t]$ such that $f(x) = 0$. A complex number x that is not algebraic (over $\mathbb{Q}$) is called *transcendental* (over $\mathbb{Q}$), so $x \in \mathbb{C}$ is transcendental if there exists no nonzero polynomial $f \in \mathbb{Q}[t]$ satisfying $f(x) = 0$.

With the above definitions, we can look at our statements to see which ones are interesting, deep, etc.

We start with Statement 1 above. This was shown by a UCLA team, using a primality test of Lucas on Mersenne numbers, to be a prime, the first known prime having ten million digits. It is, on the face of it, not very interesting. After all, it is analogous to saying 97 is a prime. However, it is interesting historically. Call an integer $M_n := 2^n - 1$, n a positive integer, a *Mersenne number* and a *Mersenne prime* if it is a prime. In 1644, Mersenne conjectured that for $n \leq 257$, the Mersenne number M_n is prime if and only if $n = 2, 3, 5, 7, 13, 17, 19, 31, 67, 127, 257$. [If M_n is a prime, then n must be a prime as $2^{ab} - 1$ factors if $a, b > 1$.] It was shown in the 1880's that M_{61} was a prime. In 1903 Frank Cole showed that $M_{67} = 193707721 \cdot 761838257287$. [He silently multiplied out these two numbers on a blackboard at a meeting of the American Mathematical Society.] In 1931 it was shown that M_{257} was *composite*, i.e., not $0, \pm 1$, or a prime, finally finishing the original Mersenne's conjecture with the correct solution. Mersenne was interested in these numbers because of its connection to a study in antiquity.

An integer $n > 1$ is called *perfect* if

$$n = \sum_{\substack{d|n \\ 0<d<n}} d \quad \text{or} \quad 2n = \sum_{\substack{d|n \\ 0<d\leq n}} d,$$

e.g., $6 = 1 + 2 + 3$. The first equation says that n is the sum of its *aloquot divisors*. A theorem in elementary number theory says:

Theorem 1.2. (Euclid/Euler) *An even number N is perfect if and only if $N = \frac{1}{2}p(p+1)$ with p a Mersenne prime (so $N = 2^{n-1}(2^n - 1)$ with $p = 2^n - 1$).*

It is still an open problem whether there exists infinitely many even perfect numbers, equivalently, infinitely many Mersenne primes and an open problem whether there exists any odd perfect numbers.

Statement 2 is very interesting and is due to Euclid. It may be the first deep mathematical fact that one learns. The proof is quite simple. If the result is false, let $p_1, \ldots, p_n$ be a complete list of (positive) primes and set $N = p_1 \cdots p_n + 1$. We use (cf. Exercise 4.17), which says:

If $x \mid y$ and $x \mid z$ in $\mathbb{Z}$, then for all $a, b \in \mathbb{Z}$, we have $x \mid ay + bz$,

i.e., if an integer divides two integers, then it divides any $\mathbb{Z}$-*linear combination* of those two numbers. From this, we conclude that if $p_i \mid N$ then $p_i \mid N - p_1 \cdots p_n$, so $p_i \mid 1$ which is impossible. This means that N is a prime different than any of $p_1, \ldots, p_n$, a contradiction.

Statement 3 is a deep theorem called the *Prime Number Theorem* or *PNT*. It quantifies Statement 2. It was first conjectured by Gauss and proven independently in 1896 by de la Valle Poussin and Hadamard using complex analysis and fundamental work of Riemann. Indeed one shows that the *zeta function*, $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$ where s is a complex variables, has no root on the line $\mathrm{Re}(s) = 1$, i.e., $\zeta(1 + \sqrt{-1}t)$ is never zero and that this is equivalent to the PNT. (The most famous problem in mathematics is the Riemann Hypothesis that says the analytic continuation of the zeta function for $\mathrm{Re}(s) > 0$ has all its roots on the line $\mathrm{Re}(s) = 1/2$.) An "elementary" proof in number theory means that it does not use complex analysis. An elementary proof may be very difficult. An elementary proof of PNT was discovered in 1949 by Selberg and Erdös. It is much harder than the non-elementary proof.

Statement 4 is not very interesting, although the Pythagoreans knew it but were afraid to leak the knowledge out, fearing catastrophe if it got out. It did. They were apparently right. To prove this, we need:

Theorem 1.3. (The Fundamental Theorem of Arithmetic) *Every integer $n > 1$ is a product of positive primes unique up to order, i.e., there exist unique primes*

$$(*) \qquad \begin{array}{c} 1 < p_1 < \cdots < p_r \\ \text{and unique integers } e_1, \ldots, e_r > 0 \text{ such that } n = p_1^{e_1} \cdots p_r^{e_r}. \end{array}$$

[We call (*) the *standard representation* or *standard factorization* of n. By the theorem, this representation is unique]. Assuming this theorem is true (Euclid knew its proof and we shall prove in Theorem 4.16) below, we show Statement 4.

If $\sqrt{2}$ is rational, then $\sqrt{2} = \frac{m}{n}$ for some integers m, n with $n \neq 0$. This means that we have an equation of integers $2n^2 = m^2$. But the prime two occurs on the left hand side to an odd power and on the right hand side to an even power, contradicting the uniqueness part of the Fundamental Theorem of Arithmetic. Thus $\sqrt{2}$ is *irrational*, i.e., a complex number that is not rational.

Statement 5 was proven by Lindemann in 1882. This is historically interesting as it solves the famous Greek construction problem, "Squaring the Circle Problem", which asks whether one can construct a circle

with the area the same as the area of a given square using only straightedge and compass (according to specific rules). Its proof is not easy. One must reduce this geometric problem to an algebraic one in Field Theory which is then proven using analysis. A proof is given in Section §67. It is easier to show that the real number e is transcendental. (Cf. Section §65.) The first real number shown to be transcendental was the number $\sum_{n=1}^{\infty} \frac{1}{10^{n!}}$ shown to be transcendental by Liouiville. (Cf. Section §64.) (That it is transcendental is essentially due to the fact that this series converges much too rapidly.) An even deeper result, solved independently by Gelfand and Schneider (which we do not prove), is

Theorem 1.4. (Hilbert's Seventh Problem) *Let α and β be algebraic over $\mathbb{Q}$ with $\alpha \neq 0, 1$ and β irrational, then α^β is transcendental. In particular $\sqrt{2}^{\sqrt{2}}$ is transcendental.*

Lindemann actually proved that if α is algebraic and nonzero, then e^α is not algebraic. Since Euler showed that $-1 = e^{\sqrt{-1}\pi}$, it follows that $\sqrt{-1}\pi$ is not algebraic and from this it follows that π is transcendental, using the fact that the product of algebraic numbers is algebraic, which we prove in Theorem 45.19.

Statement 6 is not as interesting as Statement 5 for the word "transcendental" is a misnomer. However, the proof of this by Cantor changed mathematics and caused much philosophical distress. Indeed Cantor showed that there were "many more" reals than rationals. What does this mean?

Let A and B be two sets. (What is a set?) We say that A and B have the same *cardinality* and write $|A| = |B|$ if there exists a one-to-one and onto function, i.e., a *bijection* $f : A \to B$. We call such an f a *bijective function*. Recall that a function or *map* $f : A \to B$ is one-to-one *or injective* if

$$f(a_1) = f(a_2) \quad \Longrightarrow \quad a_1 = a_2$$

(where $\Longrightarrow$ means *implies*). Let $f(A) := \{f(a) \mid a \in A\}$ denote the *image* of A, then f being one-to-one is equivalent to

$$f^{-1} : f(A) \to A \text{ given by } f(a) \mapsto a$$

is a function. We say f is onto or *surjective* if $B = f(A)$. [We also write the function by $A \xrightarrow{f} B$.]

Review 1.5. You should review the definitions of functions, inverses, etc. that we assume you have learned.

Examples 1.6.

1. $\{1, 2, \ldots, 26\}$ and $\{a, b, \ldots, z\}$ have the same cardinality.
2. $\mathbb{Z}$ and $2\mathbb{Z} := \{2n \mid n \in \mathbb{Z}\}$ have the same cardinality as $f : \mathbb{Z} \to 2\mathbb{Z}$ given by $n \mapsto 2n$ is a bijection.
3. Let $\mathbb{Z}^+$ denote the set of positive integers and $\mathbb{N} = \mathbb{Z}^+ \cup \{0\}$ the set of non-negative integers. Then $\mathbb{Z}$, $\mathbb{N}$, and $\mathbb{Z}^+$ all have the same cardinality. Indeed, the maps
$$f : \mathbb{Z} \to \mathbb{N} \text{ by } n \mapsto \begin{cases} 2n - 1 \text{ if } n > 0 \\ -2n \text{ if } n \leq 0 \end{cases}$$
and $g : \mathbb{N} \to \mathbb{Z}^+$ by $n \mapsto n + 1$ are both bijections.

If A is a set, we call the symbol $|A|$ the *cardinality* of A and interpret it to mean the "number of elements" in A. For example if A is *finite*, i.e., the set A has finitely many elements, write $|A| < \infty$, then $|A|$ is the number of elements in A, e.g., we have $|\{1, \ldots, 26\}| = 26$. If A is not finite, we of course say that A is *infinite*. If a set A satisfies $|A| = |\mathbb{Z}|$, i.e., there exists a bijection $f : \mathbb{Z} \to A$ (equivalently, there exists a bijection $g : A \to \mathbb{Z}$), we say A is *countable*. (Some authors call such a set *countably infinite or denumerable*.)

The following facts can be shown (and we leave their proofs as exercises):

Facts 1.7.

1. *A subset of a countable set is either finite or countable.*
2. $\mathbb{Q}$ *is countable.*
3. *Let* $\mathbb{C}$ *be the set of complex numbers. Then* $\{\alpha \in \mathbb{C} \mid \alpha \text{ algebraic }\}$, *the set of complex algebraic numbers, is countable.*

Theorem 1.8. (Cantor) *The set* $\mathbb{R}$ *of real numbers is not countable.*

Proof. Suppose that $\mathbb{R}$ is countable. As $\mathbb{R}$ and the closed interval $[0, 1]$ have the same cardinality by Exercise 1.12(9), the interval $[0, 1]$ must also be countable by the first fact above. As $\mathbb{Z}^+ \subset \mathbb{Z}$ is not finite, it is also countable by the first fact (or by Example 1.7(3)). Therefore, there exists a bijection $f : \mathbb{Z}^+ \to [0, 1]$, i.e., the closed interval $[0, 1]$ is covered by a sequence. Write each $f(n)$ in (base ten) decimal form, say
$$\begin{aligned} f(1) &= .a_{11}a_{12}a_{13} \cdots a_{1n} \cdots \\ f(2) &= .a_{21}a_{22}a_{23} \cdots a_{2n} \cdots \\ &\vdots \\ f(n) &= .a_{n1}a_{n2}a_{n3} \cdots a_{nn} \cdots \\ &\vdots . \end{aligned}$$

For each n choose $b_n \in \{1, \ldots, 8\}$ with $b_n \neq a_{nn}$. [We omit 0 and 9 to prevent round off problems.] Let $b = b_1 b_2 \cdots b_n \cdots$. Then b and $f(n)$ differ in the nth digit as $b_n \neq a_{nn}$. Moreover, as $b_n \neq 0, 9$, the number b and $f(n)$ cannot be the same real number. Thus $b \neq f(n)$ for all $n \in \mathbb{Z}^+$, so f is not onto, a contradiction. (We have really used the fact that $\mathbb{R}$ is *complete*, i.e., that b is a real number.) $\square$

Warning 1.9. Sets can be tricky – again we ask what is a set?

Indeed, consider the statement

Statement 1.10. There is a universe, i.e., a set that contains all other sets.

But

Contemplate 1.11. (Russell's Paradox) Let

$$A := \{B \text{ is a set} \mid B \notin B\}.$$

Is $A \in A$? Is $A \notin A$?

Exercises 1.12.

1. Show if a nonzero integer a divides integers b and c, it divides $bx+cy$ for any integers x and y.
2. If $a^n - 1$ is a prime, then $a = 2$ and n is a prime. If $2^n + 1$ is a prime, what can you say about n? [The exercise is not sufficient as $23 \mid M_{11}$.]
3. Let $a, b, n \in \mathbb{Z}$ with $n > 1$ and $b \neq 0$. Determine when $n^{\frac{a}{b}}$ is a rational number. Prove your determination. (You can use the Fundamental Theorem of Arithmetic.)
4. Assume that the real number $\sqrt{\pi}$ is transcendental. Show that π is transcendental.
5. Let $f : A \to B$ be a map of sets. Prove that f is injective if and only if given any set C and any two set map $g_i : C \to A$, $i = 1, 2$, with compositions $f \circ g_1 = f \circ g_2$, then $g_1 = g_2$.
6. Let $f : A \to B$ be a map of sets. Prove that f is surjective if and only if given any set C and any two set map $h_i : B \to C$, $i = 1, 2$, with compositions $h_1 \circ f = h_2 \circ f$, then $h_1 = h_2$.
7. Show a subset of a countable set is either countable or finite.
8. If the sets A and B are countable show that the *cartesian product*

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

is also countable. Using induction (to be discussed), show if

$$A_1, \ldots, A_n \text{ are countable so is } A_1 \times \cdots \times A_n.$$

[Hint: Show $\mathbb{Z} \times \mathbb{Z}$ (or $\mathbb{Z}^+ \times \mathbb{Z}^+$ where $\mathbb{Z}^+ := \{n \in \mathbb{Z} \mid n > 0\}$) is countable by drawing a big (= infinite matrix). In a similar way one proves Fact 1.7(2). In fact, the cartesian product of countable sets is countable, which is needed to prove the Fact 1.7(3). Can you prove this?]

9. The closed interval $[0, 1]$ and the set of all real numbers $\mathbb{R}$ have the same cardinality.
10. Any two (finite) line segments (so each has more than one point) have the same cardinality. [Hint: Draw the two line segments parallel to each other.]

CHAPTER I

The Integers

In this chapter, we investigate the basic properties of the set of integers. In particular, we show that every integer can be factored into a product of primes, unique up to order. To do so we introduce concepts that shall be generalized. We assume that the reader has seen proofs by induction. We begin the chapter with an equivalence form of induction called the Well-Order Principle and use it to establish facts about division of integers. In particular, we prove the division algorithm. Of great historical importance is the notion of prime and the property discovered by Euclid that characterizes a prime that became a cornerstone of modern algebra.

2. Well-Ordering and Induction

We denote the *empty set* by $\emptyset$. In this section, we are interested in induction which you should have seen and turns out to be equivalent to the following:

The Well-Ordering Principle 2.1. Let $\emptyset \neq S \subset \mathbb{Z}^+$. Then S contains a *least* (also called a *minimal*) element, i.e.,

$$\text{There exists } a \in S \text{ such that if } x \in S \text{ then } a \leq x.$$

This is an axiom that we accept as being true. You can visualize it by drawing the positive real line and tick off the integers. You move to the right until you get to the first element in S.

One can modify the Well-Ordering Principle to the seemingly more general:

The Modified Well-Ordering Principle 2.2. Let $\emptyset \neq T \subset \mathbb{Z}$. Suppose that there exists an $N \in \mathbb{Z}$ such that $N \leq x$ for all $x \in T$, i.e., T is *bounded from below*. Then T contains a least element.

PROOF. Let $f : \mathbb{Z} \to \mathbb{Z}$ be the map defined by $x \mapsto x + |N| + 1$ where here $|N|$ means the absolute value of N. Then f is a bijection. (What is f^{-1}?) Hence the *restriction* of f to the subset T, which we denote by $f|_T : T \to \mathbb{Z}$, is also injective. As $f(T) = f|_T(T) \subset \mathbb{Z}^+$

there exists a least element $f(a) \in f(T)$, $a \in T$, by the Well-Ordering Principle. Then a is the least element of T as f preserves $\leq$. (You should check this.) □

Remark 2.3. In an analogous way, if $\emptyset \neq T \subset \mathbb{Z}$ is *bound from above*, i.e., there exists an integer N satisfying $s \leq N$ for all $s \in T$. Then T contains a *largest* (also called a *maximal*) element, i.e., an element $a \in T$ such that $x \leq a$ for all x in T.

We have the following simple applications of well-ordering.

Application 2.4. There exists no integer N satisfying $0 < N < 1$ (or strictly between any integers n and $n + 1$).

Let $S = \{n \in \mathbb{Z} \mid 0 < n < 1\}$. If $\emptyset \neq S$ then there exists a least element $N \in S$ by well-ordering, so $0 < N < 1$ which implies that $0 < N^2 < N < 1$ – can you prove this? – and $N^2 \in \mathbb{Z}$ contradicting the minimality of N.

Application 2.5. Let $P(n)$ be a statement that is true or false depending on $n \in \mathbb{Z}^+$. If $P(n)$ is not true for some n then by the Well-Ordering Principle, there exists a least element $n \in \mathbb{Z}^+$ such that $P(n)$ is false. We call $P(n)$ a *minimal counterexample*.

Application 2.6. Suppose that $S \subset \mathbb{Z}^+$ and $1 \in S$. If S satisfies the condition that whenever $n \in S$ also $n + 1 \in S$, then $S = \mathbb{Z}^+$.

Proof. Let

$$T = \mathbb{Z}^+ \setminus S = \{n \in \mathbb{Z}^+ \mid n \notin S\},$$

i.e., $\mathbb{Z}^+ = T \cup S$, the union of T and S, and $\emptyset = T \cap S$, the intersection of T and S. (We say that $\mathbb{Z}^+$ is the *disjoint union* of S and T which we write as $\mathbb{Z}^+ = S \vee T$.)

Suppose that $T \neq \emptyset$. Then by well-ordering there exists a least positive element $n \in T$. In particular, $n - 1 \notin T$. As $1 \notin T$, $n > 1$, we have $n - 1 \in \mathbb{Z}^+$. Hence by minimality, $n - 1 \in S$. The hypothesis now implies that $n \in S$, a contradiction. Thus $T = \emptyset$ so $\mathbb{Z}^+ = S$. □

The applications yield the following:

Application 2.7. (First Principle of Finite Induction) For each positive integer n, let $P(n)$ be a statement that is true or false. Suppose that we know that

(1) Statement $P(1)$ is true.
(2) (*Induction Step*) If $P(n)$ is true then $P(n + 1)$ is true.

Then $P(n)$ is true for all positive integers n.
The hypothesis in the induction step is called the *induction hypothesis*.

Often the following equivalent variant of the First Principle of Finite Induction is more useful. That it is equivalent, is left as an exercise.

Application 2.8. (Second Principle of Finite Induction) For each positive integer n, let $P(n)$ be a statement that is true or false. Suppose that the assumption that $P(m)$ is true for all positive integers $m < n$ implies that $P(n)$ is true. Then $P(n)$ is true for all positive integers n.

Remark 2.9. If $n = 1$ then $\{m < 1 \mid m \in \mathbb{Z}^+\}$ is empty, so one still must show that $P(1)$ is true if you wish to use the Second Principle of Finite Induction as your proof would fail for $n = 1$ otherwise. Remember that well-ordering needs a non-empty set.

Remark 2.10. Note that using the Modified Well-Ordering Principle, one can start induction at any integer and prove things from that point on, e.g., you can start at 0.

We assume that you have seen induction proofs in the past, e.g. in your linear algebra course. Induction proofs can be very complicated. We give such a complicated induction proof. Note in the proof the formal steps versus the mathematical ones.

Example 2.11. The product of any $n \geq 1$ consecutive positive integers is divisible by $n!$.

Proof. Let m and n be two positive integers and set

$$(m)_n := m(m+1)\cdots(m+n-1),$$

the product of n consecutive integers starting from m.

Claim 2.12. We have $n! \mid (m)_n$ for all positive integers m and n.

Of course, the claim is exactly what we want to prove. Note that there are two integers in the claim. We have

If $m \in \mathbb{Z}^+$ is arbitrary and $n = 1$ then $(m)_n = m$ and $n! = 1! \mid m$.

We assume

Induction Hypothesis I. The claim holds for fixed $n = N - 1$ and for all m.

This is the induction step on n, using N for clarity. As we know that this holds for $n = 1$ and for all m, we must show that the result holds for N and all m, i.e., we must show

$$(N-1)! \mid (m)_{N-1} \text{ for all } m \quad \Longrightarrow \quad N! \mid (m)_N \text{ for all } m.$$

Let $m = 1$. Then $(1)_N = N!$ and $N! \mid (1)_N$. Note that this is the first step of an induction on m. So we can assume

Induction Hypothesis II. The claim holds for fixed $n = N$ and $m = M$. (Notationally it is easier to use M rather than $M - 1$).

We must show that Induction Hypotheses I and II imply the result for $n = N$ and $m = M + 1$, i.e.,

$$\text{if } N! \mid (M)_N \text{ and } (N-1)! \mid (M+1)_{N-1} \text{ then } N! \mid (M+1)_N.$$

Note that if we show this, then we have completed the induction step for the Induction Hypothesis II hence the result would be true for $n = N$ and for all m. This in turn completes the induction step for Induction Hypothesis I and hence would prove the claim.

Note also, so far even though a lot has been written, everything has been completely formal and no real work has been done except for the facts that $1 \mid m$ and $n! \mid n!$. We finally must do the mathematics. This comes from the equation

$$\begin{aligned}
&(M+1)_N - (M)_N \\
&\qquad = (M+1)(M+2)\cdots(M+N) - M(M+1)\cdots(M+N-1) \\
&\qquad = (M+1)\cdots(M+N-1)[(M+N) - M] \quad \text{(factoring)} \\
&\qquad = N(M+1)\cdots(M+N-1) = N(M+1)_{N-1}.
\end{aligned}$$

By Induction Hypothesis I, we have

$$(N-1)! \mid (M+1)_{N-1} \text{ so } N! \mid N(M+1)_{N-1},$$

so by Induction Hypothesis II, we have $N! \mid (M)_N$, This means that

$$N! \mid N(M+1)_{N-1} + (M)_N, \ \textit{i.e.}, \ N! \mid (M+1)_N$$

by Exercise 1.12(1), completing the the induction step for Induction Hypothesis II, which completes the induction step for Induction Hypothesis I as needed. □

Corollary 2.13. *Let $n \in \mathbb{Z}^+$. Then there exist n consecutive composite (i.e., non-prime and not 0 or ± 1) positive integers.*

Proof. Let $m \in \mathbb{Z}^+$ and $N = (m)_{n+1} = m(m+1)\cdots(m+n)$. Then by the example, we have $(n+1)! \mid N$. If follows that for any integer s satisfying $2 \le s \le n+1$, we have $s \mid N + s$. So the positive integers

$$(*) \qquad N+2,\ N+3,\ \ldots,\ N+n+1$$

are all composite. □

Of course, the same proof works with $m = 1$, so we did not need the example to prove this. However, (*), with N a prime, says there are arbitrarily large gaps between primes. On the other hand, primes are not 'sparse', as Chebyshev proved Bertrand's Hypothesis, which says

If $n \in \mathbb{Z}^+$ then there exists a prime p satisfying $n \leq p \leq 2n$.

Of course, $2n$ in the above can only be prime if $n = 1$. In fact, earlier Euler showed that the infinite sum $\sum_{p \text{ a positive prime}} \frac{1}{p}$ diverges, so there are 'many more' primes than say squares (in some sense).

Corollary 2.14. *Let m and n be positive integers with $m \leq n$. Define the binomial coefficients*

$$\binom{m}{n} := \frac{m!}{(m-n)!n!} = \frac{m(m-1)\cdots(m-n+1)}{n!}$$

and

$$\binom{-m}{n} := (-1)^n \frac{m(m+1)\cdots(m+n-1)}{n!}.$$

Then $\binom{m}{n}$ and $\binom{-m}{n}$ are integers.

Proof. We know that $\binom{m}{n} = \frac{(m-n+1)_n}{n!}$ is an integer by the example. The second statement follows from the identity

$$\binom{-m}{n} = (-1)^n \binom{m+n-1}{n}.$$

□

Corollary 2.15. *Let $p > 1$ be a prime, then*

$$\binom{p}{1}, \binom{p}{2}, \cdots, \binom{p}{p-1}$$

are all divisible by p, i.e, $p \mid \binom{p}{n}$ for $1 \leq n \leq p-1$.

Proof. If $1 \leq n \leq p-1$, then the example says that

$$n! \mid p(p-1)\cdots(p-n+1) = (p-n+1)_n$$

We also know that s and p have *no common non-trivial factors* if $1 < s < p$. (Proof?) We say that s and p are *relatively prime*. (Cf. 4.4.) We shall show below (cf. Theorem 4.10) but assume it for now the following:

Theorem 2.16. (General Form of Euclid's Lemma) *Let a, b, c be nonzero integers with a and b relatively prime integers. If $a \mid bc$ then $a \mid c$.*

Applying Euclid's Lemma, we conclude that

$$n! \mid (p-1)\cdots(p-n+1) \text{ so } pn! \mid p(p-1)\cdots(p-n+1)$$

for all n satisfying $1 \leq n \leq p-1$. (Why?) □

Exercises 2.17.

1. Prove that the number of subsets of a set with n elements is 2^n.
2. When Gauss was ten years old he almost instantly recognized that $1+2+\cdots+n = \frac{n(n+1)}{2}$. [Actually, what he did was a bit harder.] What is a formula for the sum of the first n cubes? Prove your result.
3. The first nine Fibonacci numbers are $1, 1, 2, 3, 5, 8, 13, 21, 34$. What is the nth Fibonacci number F_n. Show that $F_n < 2^n$.
4. Note that Euclid's proof of the infinitude of primes clearly shows that if p_n is the nth prime then $p_{n+1} \leq p_n^n + 1$. Be more careful and show that $p_{n+1} \leq 2^{2^{n+1}}$. Using this, can you then show $\pi(x) \geq \log\log(x)$ where $\pi(x)$ is the number of primes less than x if $x \geq 2$. [This is a bad estimate.]
5. Prove that the Well-Ordering Principle, the First Principle of Finite Induction, and the Second Principle of Finite Induction are all equivalent.
6. State and prove the binomial theorem. What algebraic properties do you need for your proof to work?
7. Fill in the missing details in the proof of Corollary 2.15.

3. Addendum: The Greatest Integer Function

Our (double) induction showing that binomial coefficients are integers was complicated to write, although the mathematics was simple. In this Addendum, we shall give another proof, mathematically more complicated, but of greater interest. We shall use some results that we shall prove in the next section, in particular, the Division Algorithm 4.2 and the Fundamental Theorem of Arithmetic 4.16. This alternate proof uses the following function:

Definition 3.1. The *greatest integer function*

$[\ ] : \mathbb{R} \to \mathbb{Z}$ is given by $[x] :=$ the greatest integer n satisfying $n \leq x$,

i.e., if $x \in \mathbb{R}$, using Application 2.4, we can write

$$x = n + x_0 \text{ with } 0 \leq x_0 < 1,\ n \in \mathbb{Z} \text{ then } [x] = n.$$

This function satisfies:

Properties 3.2. Let m be a positive integer. Then the following are true:

(1) $[x] \leq x < [x] + 1$.
(2) $[x + m] = [x] + m$.
(3) $[\frac{x}{m}] = [\frac{[x]}{m}]$.
(4) $[x] + [y] \leq [x + y] \leq [x] + [y] + 1$.
(5) If $n, a \in \mathbb{Z}^+$, then $[\frac{n}{a}]$ is the number of integers among $1, \ldots, n$ that are divisible by a.

PROOF. (1) and (2) are easy to see.

(3). Write $x = n + x_0$ with $0 \leq x_0 < 1$ and $n \in \mathbb{Z}$. Using the Division Algorithm 4.2, write $n = qm + r$ with $q, r \in \mathbb{Z}$ and $0 \leq r < m$. Then we have

$$\big[\frac{x}{m}\big]=\big[\frac{n + x_0}{m}\big]=\big[\frac{qm + r + x_0}{m}\big]=\big[q + \frac{r + x_0}{m}\big]= q+\big[\frac{r + x_0}{m}\big].$$

As $r \in \mathbb{Z}$ and $0 \leq r < m$, we have $0 \leq r \leq m - 1$ (cf. Application 2.4); and as $0 \leq x_0 < 1$, we have $r + x_0 < m$, so

$$\big[\frac{x}{m}\big] = q = \big[q + \frac{r}{m}\big] = \big[\frac{n}{m}\big] = \big[\frac{[x]}{m}\big].$$

(4). Write

$$\begin{array}{llllll} x & = n + x_0 & \text{with} & 0 \leq x_0 < 1 & \text{and} & n \in \mathbb{Z} \\ y & = q + y_0 & \text{with} & 0 \leq y_0 < 1 & \text{and} & q \in \mathbb{Z}. \end{array}$$

Then $0 \leq x_0 + y_0$, so

$$[x+y] = [n+q+x_0+y_0] = n+q+[x_0+y_0] = [x]+[y]+[x_0+y_0] < [x]+[y]+2$$

and the result follows easily.

(5). Let $a, 2a, \ldots, ja$ denote all the positive integers $\leq n$ and divisible by a. We must show $[\frac{n}{a}] = j$. Clearly, $ja \leq n < (j + 1)a$ so $j \leq \frac{n}{a} < j + 1$. The result now follows. □

Let $n \in \mathbb{Z}^+$ and $p > 1$ be a prime. By the Fundamental Theorem of Arithmetic 4.16, we know that there exists a unique integer $e \geq 0$ such that $p^e \mid n$ but $p^{e+1} \nmid n$. We shall write this as $p^e \mid\mid n$.

Theorem 3.3. *Let n be a positive integer and p a positive prime. Suppose that $p^e \mid\mid n!$, then*

$$e = \sum_{i=1}^{\infty} \big[\frac{n}{p^i}\big].$$

PROOF. If $p^i > n$, then $[\frac{n}{p^i}] = 0$; so the sum is really a finite sum. We prove the result by induction of n.

$n = 1$. There is nothing to prove.

We can, therefore, make the following:

Induction Hypothesis. Let $e' = \sum_{i=1}^{\infty} \left[\frac{n-1}{p^i}\right]$, then $p^{e'} \mid\mid (n-1)!$.

Let $p^f \mid\mid n$. As $n! = n(n-1)!$, by the induction hypothesis, we have $p^e = p^{e'}p^f = p^{e'+f} \mid\mid n!$, hence $e = f + e'$ or $f = e - e'$. So it suffices to prove that

$$f = \sum_{i=1}^{\infty} [\frac{n}{p^i}] - \sum_{i=1}^{\infty} [\frac{n-1}{p^i}].$$

Claim 3.4. $[\frac{n}{p^i}] - [\frac{n-1}{p^i}] = \begin{cases} 1 & \text{if } p^i \mid n \\ 0 & \text{if } p^i \nmid n. \end{cases}$

Suppose that $p^i \mid n$, then $n = p^i u$ with $u \in \mathbb{Z}$. As $0 < \frac{1}{p^i} < 1$, we have

$$[\frac{n}{p^i}] = u \text{ and } [\frac{n-1}{p^i}] = [u - \frac{1}{p^i}] = u - 1$$

as needed.

Suppose that $p^i \nmid n$. The result is immediate if $p^i > n$, so suppose not. Then we can write $n = p^i u + r$ with u and r integers with r satisfying $1 \leq r < p^i$ using the Division Algorithm 4.2. Hence

$$\frac{n}{p^i} = u + \frac{r}{p^i} \text{ so } [\frac{n}{p^i}] = u$$

and

$$\frac{n-1}{p^i} = \frac{p^i n + r - 1}{p^i} = u + \frac{r-1}{p^i} \text{ for } 0 \leq r - 1 < p^i,$$

so $[\frac{n}{p^i}] = u$ as needed. This proves the claim. But the claim means that

$$\sum_{i=1}^{\infty} [\frac{n}{p^i}] - \sum_{i=1}^{\infty} [\frac{n-1}{p^i}] = \sum_{p^i \mid n} 1 = f. \qquad \square$$

Example 3.5. Using the properties of [ ], we compute e such that $7^e \mid\mid 1000!$.

$$\begin{aligned}
[\frac{1000}{7}] &= 142 \\
[\frac{1000}{7^2}] &= [\frac{\left[\frac{1000}{7}\right]}{7}] = [\frac{142}{7}] = 20 \\
[\frac{1000}{7^3}] &= [\frac{\left[\frac{1000}{7^2}\right]}{7}] = [\frac{20}{7}] = 2 \\
[\frac{1000}{7^4}] &= [\frac{\left[\frac{1000}{7^3}\right]}{7}] = [\frac{2}{7}] = 0,
\end{aligned}$$

so $e = 142 + 20 + 2 = 164$, i.e., $7^{164} \mid\mid 1000!$.

We can generalize our result about binomial coefficients to show that *multinomial coefficients* are integers.

Corollary 3.6. *Suppose that $a_1, \ldots, a_r$ are non-negative integers satisfing $a_1 + \cdots + a_r = n$, then the multinomial coefficient $\dfrac{n!}{a_1! \cdots a_r!}$ is an integer.*

Proof. By the Fundamental Theorem of Arithmetic 4.16 and the theorem, it suffices to prove for each prime p, we have

$$\sum_{i=1}^{\infty} [\frac{n}{p^i}] \geq \sum_{i=1}^{\infty} [\frac{a_1}{p^i}] + \cdots + \sum_{i=1}^{\infty} [\frac{a_r}{p^i}].$$

Applying Property 3.2(4) shows that

$$[\frac{a_1}{p^i}] + \cdots + [\frac{a_r}{p^i}] \leq [\frac{a_1 + \cdots + a_r}{p^i}] = [\frac{n}{p^i}]$$

for all i. Summing over i yields the result. □

That the product of n consecutive positive integers is divisible by $n!$ and that binomial coefficients are integers is now easy to see as $\dfrac{(m)_n}{n!} = \dbinom{m+n-1}{n}$ for all $m \in \mathbb{Z}^+$ and $\dbinom{a}{b} = \dfrac{a!}{b!(a-b)!}$ for all $a, b \in \mathbb{Z}^+$ with $b \leq a$.

Chebyshev cleverly used the binomial coefficient $\dbinom{2n}{n} = \dfrac{(2n)!}{n!n!}$. For each prime $p \leq 2n$, let $r_p \in \mathbb{Z}$ satisfy

$$p^{r_p} \leq 2n < p^{r_p+1}$$

and $p^e \mid\mid \binom{2n}{n}$. Then

$$e = \sum_{i=1}^{\infty} [\frac{2n}{p^i}] - 2\sum_{i=1}^{\infty} [\frac{n}{p^i}].$$

We have

$$[\frac{2n}{p^i}] = [\frac{n}{p^i} + \frac{n}{p^i}] \leq [\frac{n}{p^i}] + [\frac{n}{p^i}] + 1$$

by Property 4.9(4), so $e \leq \sum_{i=1}^{r_p} 1 = r_p$. This is used to prove

$$\binom{2n}{n} = \frac{(2n)!}{n!n!} \text{ divides } \prod_{\substack{p \leq 2n \\ p \text{ prime}}} p^{r_p} \tag{3.7}$$

and if further $n < p \leq 2n$, then $p \mid (2n)!$ but $p \nmid (n!)^2$. Therefore, we see

$$n^{\pi(2n)-\pi(n)} \leq \prod_{\substack{n<p\leq 2n \\ p \text{ prime}}} p \leq \binom{2n}{n} \leq \prod_{\substack{p \leq 2n \\ p \text{ prime}}} p^{r_p} \leq (2n)^{\pi(2n)}. \tag{3.8}$$

Using this Chebyshev showed that there were positive real numbers c_1 and c_2 satisfying

$$c_1 \frac{x}{\log x} \leq \pi(x) \leq c_2 \frac{x}{\log x}$$

(cf. The Prime Number Theorem). He also showed if $n \geq 3$ and $2n/3 < p < n$, with p a prime, then $p \nmid \binom{2n}{n}$ and used this to prove Bertrand's Hypothesis that there always exists a prime p such that $n \leq p \leq 2n$.

Exercises 3.9. 1. Let a and b be positive integers, then $\dfrac{(ab)!}{a!(b!)^a}$ is an integer.

2. Verify equations 3.7 and 3.8.

4. Division and the Greatest Common Divisor

We turn to further applications of well-ordering. We shall need the following:

Properties 4.1. For integers r, n, and m the following division properties hold:

(1) If $r \mid m$ and $r \mid n$, then $r \mid am + bn$ for all integers a and b.
(2) If $r \mid n$, then $r \mid mn$.

(3) If If $r \mid n$ and $n \neq 0$, then the absolute value $|n| \geq |r| \geq r$.
(4) If $m \mid n$ and if $n \mid m$, then $n = \pm m$.
(5) If $mn = 0$, then $m = 0$ or $n = 0$.
(6) If $mr = nr$, then either $m = n$ or $r = 0$.

PROOF. Exercise, but do not use the Fundamental Theorem of Arithmetic. □

The most important elementary property about the integers is the next result that intertwines multiplication and addition of integers. In the sequel, we shall investigate when it holds in more general situations. As it is one of our first proofs about integers, we shall put in a lot of detail. Notice that one can often assume properties about integers that we have not shown and perhaps are not clear. For example, in the proof below, we assume the transitivity of $>$. Can this be proved or must it be axiomized?

Theorem 4.2. (Division Algorithm) *Let m and n be integers with m positive. Then there exist unique integers q and r satisfying:*

(*i*) $n = qm + r$.
(*ii*) $0 \leq r < m$

[Of course, (ii) is the crux.]

PROOF. We have two things to show: existence and uniqueness. We first show

Uniqueness: Let (q, r) and (q', r') be two pairs of integers satisfying the conclusion. We must show $q = q'$ and $r = r'$. We have

$$(*) \qquad \begin{array}{c} qm + r \;=\; n \;=\; q'm + r' \\ 0 \leq r,\, r' < m \end{array}$$

Without loss of generality, we may assume that $r \leq r'$. By (*), we have

$$0 \leq r' - r = (q - q')m.$$

If $q - q' = 0$, i.e., $q = q'$, then $r' - r = 0$ and $r' = r$ and we are done. So we may assume that $q - q' \neq 0$, i.e., $q \neq q'$, then $r' - r = (q - q')m \neq 0$ by Property (5) as $m \neq 0$. Therefore, we have

$$r' - r > 0 \text{ and } m \mid r' - r.$$

Thus by Property (3), we have

$$m \leq r' - r < r' < m,$$

a contradiction. So $q - q' = 0$ as needed.
[Note that we have used the important Property (5):

$$ab = 0 \quad \Longrightarrow \quad a = 0 \text{ or } b = 0,$$

equivalently, if

$$ab = ac \quad \Longrightarrow \quad a = 0 \text{ or } b = c.]$$

Existence:

Case 1. $n > 0$:

Intuitively we know that the number n must lie in some half open interval $[qm, (q+1)m)$, i.e., there exists an integer q such that $qm \leq n < (q+1)m$. Let us show this rigorously. Let

$$S = \{s \in \mathbb{Z}^+ \mid sm > n\} \subset \mathbb{Z}^+.$$

As $m > 0$, we have $m \geq 1$. (Recall we have shown there is no integer properly between 0 and 1.) So $(n+1)m = mn + m \geq n + m > n$. (Can you show that $mn > n$?) Thus $n + 1 \in S$ so $S \neq \emptyset$. By the Well-Ordering Principle there exists a least integer $q + 1 \geq 0$ which means $qm \leq n < (q+1)m$. Let $r = n - qm \geq 0$. We then have

$$0 \leq r = n - qm < (q+1)m - qm = m,$$

so these q, r work.

Case 2. $n < 0$:

By Case 1, there exist integers q_1 and r_1 satisfying

$$-n = |n| = q_1 m + r_1 \quad \text{and} \quad 0 \leq r_1 < m.$$

If $r_1 = 0$, then $q = -q_1$ and $r = 0$ work. If $r_1 > 0$, then

$$\begin{aligned} n &= -q_1 m - r_1 = -q_1 m - m + m - r_1 \\ &= (-q_1 - 1)m + (m - r_1), \end{aligned}$$

so $0 \leq m - r_1 < m$ and $q = -q_1 - 1$ and $r = m - r_1$ work. □

Question 4.3. What if $m < 0$ in the above?

As mentioned above, later we shall be interested in generalizing the Division Algorithm for integers. You already have used an analogue of it in the case of the division of polynomials with real coefficients in the real numbers. Does this analogue hold if we only allow integer coefficients?

We turn to another important property about the integers.

Definition 4.4. Let n and m be integers with at least one of them nonzero. An integer d is called the *greatest common divisor* or *gcd* of m and n if d satisfies the following:

(i) $d > 0$.
(ii) $d \mid m$ and $d \mid n$.
(iii) If e is an integer satisfying $e \mid m$ and $e \mid n$ then $e \mid d$.

If $d = 1$ is a gcd of n and m, we say that n and m are *relatively prime*.

Theorem 4.5. *Let $n \neq 0$ and m be integers. Then a gcd of m and n exists and is unique.*

PROOF. **Uniqueness**: If both d and d' satisfy (i), (ii), and (iii) then $d \mid d'$ and $d' \mid d$ so $d = \pm d'$ by Property 4.1(4), hence $d = |d| = |d'| = d'$ by (i).

Existence: Let

$$S = \{am + bn \mid am + bn > 0 \text{ with } a, b \in \mathbb{Z}\}.$$

[This is the tricky part of this proof. Where did this come from?]

As $n \neq 0$, we have $|n| \in S$, so $S \neq \emptyset$. By the Well-Ordering Principle, there exists a $d = am + bn$ in S, a least element for some integers a, b.

Claim. This d works, i.e., satisfies (i), (ii), and (iii).

By choice, d satisfies (i). To show (ii), we use the Division Algorithm to produce integers q and r satisfying $n = qd + r$ with $0 \leq r < d$. We show $r = 0$, i.e., $d \mid n$. As

$$n = dq + r = (am + bn)q + r, \quad \text{we have} \quad 0 \leq r = (1 - bq)n + (-aq)m.$$

This means that either $r \in S$ or $r = 0$. Since $r < d$, the minimality of d implies that $r \notin S$, so $r = 0$ and $d \mid n$. Similarly, $d \mid m$.

As for (iii), suppose that the integer e satisfies $e \mid m$ and $e \mid n$. Then by 4.1 (1), we have $e \mid am + bn = d$. □

Notation 4.6. If n and m are integers with at least one of them nonzero, we shall denote their gcd by (n, m).

Note that the proof even gives the following

Bonus. Let m and n be integers, at least one nonzero and $d = (n, m)$. Then there exist integers x and y satisfying

$$d = (m, n) = mx + ny.$$

In particular, the *Diophantine equation*

$$(m, n) = mX + nY.$$

(with X and Y variables) has a solution, where the word Diophantine means that we only want integer solutions, e.g., the Diophantine equation $X^2 = 2$ has no solution.

Warning 4.7. The x and y in the Bonus are not unique, in general, e.g.,

$$(2,4) = 2 = 1 \cdot 4 - 1 \cdot 2 = \quad 0 \cdot 4 + 1 \cdot 2$$
$$(2,3) = 1 = 1 \cdot 3 - 1 \cdot 2 = -1 \cdot 3 + 2 \cdot 2$$

See Exercise 4.17 (7) for the general solution to the Diophantine equation $d = mX + nY$.

Unfortunately, our proof does not show how to find any solution to the Diophantine equation $(m,n) = mX + nY$. Euclid knew how to do so and his proof constructed such a solution.

Theorem 4.8. (Euclidean Algorithm) *Let a and b be positive integers with $b \nmid a$. [If $b \mid a$, we have $(a,b) = b$.] Then there exists an integer $k > 0$ and equations*

$$\begin{aligned}
a &= bq_1 + r_1 && \text{with} && 0 < r_1 < b \\
b &= r_1 q_2 + r_2 && \text{with} && 0 < r_2 < r_1 \\
r_1 &= r_2 q_3 + r_3 && \text{with} && 0 < r_3 < r_2 \\
&\vdots \\
r_{k-2} &= r_{k-1} q_k + r_k && \text{with} && 0 < r_k < r_{k-1} \\
r_{k-1} &= r_k q_{k+1}
\end{aligned}$$

for some integers $q_1, \ldots, q_{k+1}$ and $r_1, \ldots, r_k$.

Moreover, $r_k = (a,b)$ and the Diophantine equation $(a,b) = aX + bY$ has a solution.

Finding the gcd of two integers is much easier and faster than factoring numbers. Indeed, using the internet to make purchases or send confidential information is (relatively) safe because of the difficulty in factoring large numbers.

Properties 4.9. Let a, b, and c be integers with $a \neq 0$ and $d = (a,b)$.

(1) We have $d = 1$ if and only if $1 = ax + by$ for some integers x, y.
(2) $(\frac{a}{d}, \frac{b}{d}) = 1$. [Note that $\frac{a}{d}$ and $\frac{b}{d}$ are integers.]
(3) If $d = 1$ and $a \mid bc$ (in $\mathbb{Z}$), then $a \mid c$.
(4) If $a \mid bc$ (in $\mathbb{Z}$), then $\frac{a}{d} \mid c$.
(5) If $m > 0$, then $(ma, mb) = md$.

We have seen Property (3) above. As it is so important, we write it one more time.

Theorem 4.10. (General Form of Euclid's Lemma) *Let a, b be relatively prime integers with a nonzero. If $a \mid bc$ with c an integer, then $a \mid c$.*

A special case of this theorem is:

Theorem 4.11. (Euclid's Lemma) *Let a, b be integers and p be a prime satisfying $p \mid ab$. Then $p \mid a$ or $p \mid b$.*

Of course this means

Consequence 4.12. If a prime p satisfies $p \mid a_1 \cdots a_r$, with a_i integers for $i = 1, \ldots, r$, then there exists an i such that $p \mid a_i$.

PROOF. of the properties and Euclid's Lemma.

(1). We already know if $d = 1$ then $1 = ax + by$ for some $x, y \in \mathbb{Z}$. Conversely, suppose that $1 = ax + by$ for some $x, y \in \mathbb{Z}$. As $d \mid a$ and $d \mid b$, we have $d \mid 1$ by Property 4.1(1). Since $d > 0$, we have $d = 1$.

(2). By (1), there exist integers x and y satisfying $d = ax + by$ hence $1 = \frac{a}{d}x + \frac{b}{d}y$ so (2) follows by (1).

(3). Whenever you can write 1 or 0 in a non-trivial way do so! It is a very useful technique. So by (1), we have

$$1 = ax + by \text{ for some } x, y \in \mathbb{Z} \qquad \text{(Key Observation)} \tag{4.13}$$

and given such an equation you can always multiply it, say by the integer c to get

$$c = cax + cby. \qquad \text{(Key Trick)} \tag{4.14}$$

As $a \mid a$ and $a \mid cb$, we conclude that $a \mid c$.

(4). The hypothesis means that $bc = an$ for some integer n, so $\frac{a}{d}n = \frac{b}{d}c$. By (2), we know that $(\frac{a}{d}, \frac{b}{d}) = 1$ hence $\frac{a}{d} \mid c$ by (3).

(5). We leave this is an exercise.

(proof of) Euclid's Lemma. If p is a prime and $p \mid ab$ but $p \nmid a$, then $(p, a) = 1$ as only $\pm 1, \pm p$ divide p. Since $p \mid ab$, we conclude that $p \mid b$ by (3). □

The converse of Euclid's Lemma is also true, viz.,

Proposition 4.15. *Let p be an integer. Then p is a prime if and only if whenever $p \mid ab$, with a and b integers, then $p \mid a$ or $p \mid b$.*

This proposition, which we leave as an exercise (cf. Exercise 4.17(8)) is a key to much of (commutative) ring theory. It says that we have two ways to define the notion of a prime element in the integers. For

more general structures, called rings, we can look at both of these conditions. Unfortunately, they need not be equivalent and it turns out that the more useful condition is the condition $p \mid ab$ then $p \mid a$ or $p \mid b$. In fact, its generalization has major repercussions not only in algebra but also in geometry. We shall investigate some of these later on.

We had a further loose end to establish, viz., a proof of the Fundamental Theorem of Arithmetic, which we turn to next. We first state it again.

Theorem 4.16. (The Fundamental Theorem of Arithmetic) *Every integer $n > 1$ is a product of positive primes unique up to order, i.e., there exist unique primes*

$$(*) \qquad \begin{array}{c} 1 < p_1 < \cdots < p_r \\ \textit{and integers } e_1, \ldots, e_r > 0 \textit{ such that } n = p_1^{e_1} \cdots p_r^{e_r} \end{array}$$

[We call (*) the *standard representation* or *standard factorization* of n, and this representation is unique].

PROOF. **Existence**. Let

$$S = \{n > 1 \text{ in } \mathbb{Z} \mid n \text{ is not a product of primes}\}.$$

We must show $S = \emptyset$. Suppose this is false. By the Well-Ordering Principle, there exists a minimal element $n \in S$. Clearly, no prime lies in S, so n is not a prime. Hence there exist integers n_1, n_2 satisfying $n = n_1 n_2$ and $1 < n_i < n$, $i = 1, 2$ (as there exists an integer $1 < n_1 < n$ dividing n). By minimality, $n_1, n_2 \notin S$, so each is a product of primes. Hence so is $n = n_1 n_2$, a contradiction.
[This is a wonderful argument, and we shall see it quite often.]

Uniqueness. Suppose that

$$p_1^{e_1} \cdots p_r^{e_r} \;=\; n \;=\; q_1^{f_1} \cdots q_s^{f_s}$$

with $1 < p_1 < \cdots < p_r$ and $1 < q_1 < \cdots < q_s$ primes and all the e_i, f_j positive integers. We may assume that $p_1 \leq q_1$. As $p_1 \mid n = q_1^{f_1} \cdots q_s^{f_s}$, we must have $p_1 \mid q_i$ for some i by Euclid's Lemma. But $p_1 \leq q_i$ are primes, so we must have $i = 1$ and $p_1 = q_1$. Thus, by cancellation,

$$p_1^{e_1 - 1} \cdots p_r^{e_r} \;=\; q_1^{f_1 - 1} \cdots q_s^{f_s}$$

and we are done by induction. – What is the induction hypothesis and why are we done? □

The proof of the existence statement in the Fundamental Theorem of Arithmetic is used repeatedly in mathematics and its consequence occurs much more frequently than the uniqueness statement. If we are studying a class of objects in which there is a set of basic objects

(atoms), the basic question that arises is whether we can break up an element in the class into a finite number of these atoms (in some way). In the Fundamental Theorem of Arithmetic the atoms are the primes.

Exercises 4.17.

1. Prove all the properties in Properties 4.1 without using the Fundamental Theorem of Arithmetic.
2. Let $F = \mathbb{R}, \mathbb{C}$ or $\mathbb{Q}$ [or any *field*, or even any *ring*, cf. Definition 8.3]. Let $F[t]$ be the set of polynomials with coefficients in F with the usual addition and multiplication. (Which are?) State and prove the analog of the Division Algorithm for integers. (Use your knowledge of such division. Use the degree of a polynomial as a substitute for statement (ii) in the Division Algorithm for integers .) What can you do if you take polynomials with coefficients in $\mathbb{Z}$?
3. Prove the following modification of the Division Algorithm: If m and n are two integers with m nonzero, then there exist unique integers q and r satisfying

$$n = mq + r \text{ with } -\frac{1}{2}|m| < r \leq \frac{1}{2}|m|.$$

   Moreover, if $m > 0$ is odd, then we can find unique integers q and r satisfying

$$n = mq + r \text{ with } 0 \leq |r| < \frac{m}{2}.$$

   [Note that, in this case, $m/2$ is not an integer.]
4. Prove the Euclidean Algorithm 4.8, i.e., show that such a k exists and the Diophantine equation $(m, n) = mX + nY$ has a solution.
5. In the Euclidean Algorithm show each nonzero remainder r_i ($i \geq 2$) satisfies $r_i < \frac{1}{2}r_{i-2}$. Deduce that the number of steps in the algorithm is less than

$$\frac{2\log b}{\log 2}$$

   where b is the larger of the numbers a, b. Using the modified Division Algorithm, show the number of steps there is at most $\frac{\log b}{\log 2}$.
6. Solve the Diophantine equation

$$(39493, 19853) = 39493\,X + 19853\,Y.$$

7. Let m, n, and d be nonzero integers. Show that the Diophantine equation $d = mX + nY$ has a solution if and only if $(m, n) \mid d$. If this is the case and (x, y) is a solution in integers, then the general solution is $(x + \frac{n}{(m,n)}k,\ y - \frac{m}{(m,n)}k)$ with $k \in \mathbb{Z}$.
8. Let p be an integer. Then p is a prime if and only if whenever $p \mid ab$, with a and b integers, then $p \mid a$ or $p \mid b$.
9. Carefully write up the induction step and its proof for the uniqueness part of the Fundamental Theorem of Arithmetic in two different ways, i.e., by inducting on two different things.
10. Let n and m be integers with at least one of them nonzero. An integer l is called the *least common multiple* or *lcm* of m and n if l satisfies the following:
    (*i*) $l \geq 0$.
    (*ii*) $m \mid l$ and $n \mid l$.
    (*iii*) If k is an integer satisfying $m \mid k$ and $n \mid k$ then $l \mid k$.
    Show that an lcm of m and n exists, call it $[m, n]$. In addition, show that $mn = [m, n](m, n)$.
11. Define $\sigma : \mathbb{Z}^+ \to \mathbb{Z}^+$ by $\sigma(n) = \sum_{d|n} d$, the sum of the (positive) divisors of n. Show
    (*i*) If m and n are relatively prime (positive) integers, then $\sigma(mn) = \sigma(m)\sigma(n)$.
    (*ii*) If p is a (positive) prime integer and n an integer, then $\sigma(p^n) = (p^{n+1} - 1)/(p - 1)$.
12. In the notation of the previous result, a positive integer n is called a *perfect number* if $\sigma(n) = 2n$. Prove Theorem 1.2.

CHAPTER II

Equivalence Relations

This chapter is the most important foundational chapter in Part One. It introduces the notion of an equivalence relation and determines equivalent formulations of it. In particular, an equivalence relation is shown to be the same as a surjective map. The importance of equivalence relations is that it leads to 'quotients'. This is probably the hardest concept to understand in Part One and Part Two. It is used to coarsen problems that may more easily be solved, and hopefully leads to the solution of specific problems in which we are interested. The basic difficulty is that it may easy to define an equivalence relation on collection of sets but it probably is not useful if it must reflect additional properties (e.g., algebraic properties) of those maps, i.e., if the surjective map arising from the equivalence relation (or defining it) does not reflect these additional properties. For example, what information does a surjective linear transformation of vector spaces give us?

5. Equivalence Relations

In this section, we study one of the basic concepts in mathematics, equivalence relations. You should have seen this concept. Recall the following:

Definition 5.1. Let A and B be two sets. A *relation* on A, B is a subset $R \subset A \times B$. It is customary to write aRb if $(a, b) \in R$ and we shall always do so. If $A = B$ then we call such a relation a *relation on* A.

Question 5.2. What kind of relation is a function $f : A \to B$?

Besides functions, we are mostly interested in equivalence relations, which we also recall.

Definition 5.3. A relation R on A is called an *equivalence relation* on A if the following holds: For all $a, b, c \in A$, we have the following:

(1) aRa. (*Reflexitivity*)

(2) If aRb then bRa. (*Symmetry*)

(3) If aRb and bRc then aRc. (*Transitivity*)

Remarks 5.4. We often write an equivalence relation by $\sim$ or $\approx$.

Question 5.5. Why do we need (1) in the definition, i.e., why don't (2) and (3) imply (1)?

Whenever a new term is defined, you should try to find examples. It is usually difficult to have an intuitive idea of an abstract concept without examples that you know. We give a few examples of sets with an equivalence relation on it.

Examples 5.6. The following are equivalence relations:

1. Any set A under $=$ where if $a, b \in A$ then $a = b$ in A means that we have equality of sets $\{a\} = \{b\}$.
2. Triangles in $\mathbb{R}^2$ under congruence (respectively, similarity). [Question: Is the mirror image of a triangle congruent to the original triangle?]
3. The set $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ under $\sim$ where
$$(a, b) \sim (c, d) \text{ if } ad = bc \,(\text{in } \mathbb{Z}).$$
4. $\mathbb{Z}$ under $\equiv \mod 2$ where
$$m \equiv n \mod 2 \text{ if } m - n \text{ is even } i.e., 2 \mid m - n.$$
Equivalently, both m and n are odd or both are even.
5. Let R be $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, or any field (or, in fact, any ring, e.g., $\mathbb{Z}$, cf. Definition 8.3). Set
$$\mathbb{M}_n(R) := \{n \times n \text{ matrices with entries in } R\}.$$
Then $\sim$ is an equivalence relation on $\mathbb{M}_n(R)$ where
$$A \sim B \text{ if there exists } C \in \mathbb{M}_n(R) \text{ invertible such that } A = CBC^{-1}.$$
We call this equivalence relation *similarity of matrices*. We call two matrices A and B *similar* if $A \sim B$. [Cf. Change of Basis Theorem in linear algebra.]

6. The Change of Basis Theorem in linear algebra, in fact, leads to the following equivalence relation. Let R be any field (or ring). Define
$$R^{m\times n} := \{m \times n \text{ matrices with entries in } R\}.$$
Then $\approx$ is an equivalence relation on $R^{m\times n}$, where
$$A \approx B \text{ if there exists invertible matrices}$$
$$C \in \mathbb{M}_m(R) \text{ and } D \in \mathbb{M}_n(R) \text{ satisfying } A = CBD.$$
We say two matrices A and B in $R^{m\times n}$ are *equivalent* if $A \approx B$.
7. Let R be a field (or ring), then $\backsim$ is an equivalence relation on $\mathbb{M}_n(R)$, where
$$A \backsim B \text{ if there exists } C \in \mathbb{M}_n(R) \text{ invertible such that } A = CBC^t.$$
where C^t is the transpose of C.
8. Define $\backsim_u$ on $\mathbb{M}_n(\mathbb{C})$ by
$$A \backsim_u B \text{ if there exists } C \in \mathbb{M}_n(\mathbb{C}) \text{ invertible such that } A = CBC^*.$$
where C^* is the adjoint of C. Then $\backsim_u$ is an equivalence relation on $\mathbb{M}_n(\mathbb{C})$

Definition 5.7. Let $\sim$ be an equivalence relation on A. For each a in A, the set
$$[a] = [a]_\sim := \{b \in A \mid a \sim b\}$$
is called the *equivalence class* of a *relative to* $\sim$. We shall usually write
$$\overline{a} \text{ for } [a]$$
and call the following set of subsets of A,
$$\overline{A} = A/\sim := \{\overline{a} \mid a \in A\},$$
the *set of equivalence classes* of $\sim$ on A.

Warning 5.8. We have $\overline{a} \subset A$ not $\overline{a}$ is an element of A.

Let $\sim$ be an equivalence relation on A. Then we have a map
$$^- : A \to \overline{A} \text{ given by } a \mapsto \overline{a}.$$
This map is clearly surjective and is called the *natural* or *canonical surjection*. [It is the "obvious" map as it depends on no choices.] Thus if α is an element in $\overline{A}$, there exists an element a in A such that $\alpha = \overline{a}$.

Examples 5.9. 1. $\sim$ on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ given by
$$(a, b) \sim (c, d) \text{ if } ad = bc \,(\text{in } \mathbb{Z}).$$
Then we have
$$\overline{(a, b)} \text{ is just } \frac{a}{b} \text{ in } \mathbb{Q} = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})/\sim .$$

2. $\mathbb{Z}$ under $\equiv \mod 2$. Then we have

$$\overline{0} = \overline{\pm 2} = \overline{\pm 4} = \cdots = \overline{2n} = \cdots = \{\text{all even integers}\}$$
$$\overline{1} = \overline{\pm 3} = \overline{\pm 5} = \cdots = \overline{2n+1} = \cdots = \{\text{all odd integers}\}.$$

We shall write $\overline{\mathbb{Z}} = \mathbb{Z}/\equiv \mod 2$ as $\mathbb{Z}/2\mathbb{Z}$, so

$$\mathbb{Z}/2\mathbb{Z} = \{\overline{0}, \overline{1}\}.$$

Recall the following:

Definition 5.10. Let A_i, $i \in I$ be sets. [We call I an *indexing set*.] The *union* of the sets A_i is the set

$$\bigcup_{i\in I} A_i := \{x \mid \text{ there exists an } i \in I \text{ such that } x \in A_i\}$$

We shall usually denote this set by $\bigcup_I A_i$. This union is a *disjoint union* if $A_i \cap A_j = \emptyset$ for all $i, j \in I$ with $i \neq j$, and denote it by $\bigvee_I A_i$. Of course, the *intersection* of the sets A_i is the set

$$\bigcap_{i\in I} A_i := \{x \mid x \in A_i \text{ for all } i \in I\}.$$

that we denote by $\bigcap_I A_i$.

Proposition 5.11. *Let $\sim$ be an equivalence relation on A. Then*

$$A = \bigvee_{\overline{A}} \overline{a}.$$

In particular, if $a, b \in A$, then

$$\textit{either } \overline{a} = \overline{b} \quad \textit{or} \quad \overline{a} \cap \overline{b} = \emptyset,$$

hence

$$\overline{a} = \overline{b} \quad \textit{if and only if} \quad a \sim b.$$

Proof. As $a \sim a$ for all $a \in A$ by Reflexitivity and $a \in \overline{a}$ by definition, we have $A = \bigcup_A \overline{a}$. By Symmetry, we have $a \sim b$ if $\overline{a} = \overline{b}$. If $c \in \overline{a} \cap \overline{b}$, then $c \sim a$ and $c \sim b$, so $a \sim c$ by Symmetry hence $a \sim b$ by Transitivity, so $a \in \overline{b}$, i.e., $\overline{a} \subset \overline{b}$. Similarly, we have $\overline{b} \subset \overline{a}$, hence $\overline{a} = \overline{b}$. □

Definition 5.12. Let $\sim$ be an equivalence relation on A. An element $x \in \overline{a}$ is called a *representative* of $\overline{a}$. For example, a is a representative of $\overline{a}$. So by the proposition if x is a representative of $\overline{a}$, then $\overline{x} = \overline{a}$. A *system of representatives* for A relative to $\sim$ is a set

$$\mathcal{S} = \{\text{precisely one element from each equivalence class}\},$$

so $A = \bigvee_{\mathcal{S}} \overline{x}$. For example, if $\sim$ is $\equiv \mod 2$ then $\{-36, 15\}$ is a system of representatives of $\equiv \mod 2$ and $\mathbb{Z} = \overline{-36} \vee \overline{15}$.

In later sections, this shall be very useful, so we give it the name of

Mantra 5.13. of Equivalence Relations. In the above setup, we have

$$A = \bigvee_{\mathcal{S}} \overline{x}.$$

In particular, if $|A| < \infty$, then

$$|A| = \sum_{\mathcal{S}} |\overline{x}|.$$

This is only useful if we can compute the size of the equivalence class $\overline{x}$.

Exercises 5.14.

(1) A *partition* of a set A is a collection $\mathcal{C}$ of subsets of A such that $A = \bigvee_{\mathcal{C}} B$. Let R be an equivalence relation on A. Then $\overline{A}$ partitions A. Conversely, let $\mathcal{C}$ partition A. Define a relation $\sim$ on A by $a \sim b$ if a and b belong to the same set in $\mathcal{C}$. Show $\sim$ is an equivalence relation on A. So an equivalence relation and a partition of a set are essentially the same.

(2) Draw lines in $\mathbb{R}^2$ perpendicular to the X-axis through all integer points and the analogous lines parallel to the Y-axis. Define a partition of the plane that arise. [Be careful about the points on the lines.]
[**Question**. Can you do this in such a way that when viewed in $\mathbb{R}^3$ the corresponding equivalence classes looks like a torus?]

6. Modular Arithmetic

We introduce one of the most important equivalence relations in elementary arithmetic. It generalizes $\equiv \mod 2$.

Definition 6.1. Fix an integer $m > 1$ and let $a, b \in \mathbb{Z}$. We say that a is *congruent* to b *modulo* m and write $a \equiv b \mod m$ if $m \mid a - b$ in $\mathbb{Z}$. The set

$$\begin{aligned} \overline{a} = [a]_m &:= \{x \in \mathbb{Z} \mid x \equiv a \mod m\} \\ &= \{x \in \mathbb{Z} \mid x = a + km \text{ some } k \in \mathbb{Z}\} \end{aligned}$$

is a subset of $\mathbb{Z}$ called the *residue class* of a *modulo* m. We shall denote this set as above or as $a + m\mathbb{Z}$. For example,

$$m\mathbb{Z} = 0 + m\mathbb{Z} = \{km \mid k \in \mathbb{Z}\},$$

all the multiples of m.

Question 6.2. In the above, what happens if we would let $m = 1$? If we let $m = 0$? If we let $m < 0$?

Let a and b be integers and $m > 1$ an integer. We have three ways of saying the same thing. They are:

$$m \mid a - b \quad \text{or} \quad a \equiv b \mod m \quad \text{or} \quad \overline{a} = \overline{b}.$$

One usually uses the one that is most convenient, although algebraically the last is the most interesting.

Proposition 6.3. *Let $m > 1$ in $\mathbb{Z}$. Then $\equiv \mod m$ is an equivalence relation. In particular,*

$$\mathbb{Z} = \overline{0} \vee \overline{1} \vee \cdots \vee \overline{m-1}.$$

Write $\mathbb{Z}/m\mathbb{Z}$ for $\mathbb{Z}/\equiv \mod m$. Then

$$\mathbb{Z}/m\mathbb{Z} = \{\overline{0}, \overline{1}, \cdots, \overline{m-1}\} \text{ and } |\mathbb{Z}/m\mathbb{Z}| = m.$$

Let $a, b, c, d \in \mathbb{Z}$ satisfy

$$a \equiv b \mod m \text{ and } c \equiv d \mod m,$$

then

$$a + c \equiv b + d \mod m \text{ and } a \cdot c \equiv b \cdot d \mod m.$$

Equivalently,

$$\text{if } \overline{a} = \overline{b} \text{ and } \overline{c} = \overline{d}, \text{ then } \overline{a+c} = \overline{b+d} \text{ and } \overline{a \cdot c} = \overline{b \cdot d}.$$

Define

$$+ : \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \text{ by } \overline{a} + \overline{b} := \overline{a+b}$$

and

$$\cdot : \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \text{ by } \overline{a} \cdot \overline{b} := \overline{a \cdot b}.$$

Then $+$ and $\cdot$ are well-defined.

[To be *well-defined* means that they are functions. In this case, this means that if $\overline{a} = \overline{a_1}$ and $\overline{b} = \overline{b_1}$, then $\overline{a} + \overline{b} = \overline{a_1} + \overline{b_1}$ and $\overline{a} \cdot \overline{b} = \overline{a_1} \cdot \overline{b_1}$.]
Moreover, $(\mathbb{Z}/m\mathbb{Z}, +, \cdot)$ satisfies the axioms of a commutative ring, i.e., for all $a, b, c \in \mathbb{Z}$, we have

(1) $(\overline{a} + \overline{b}) + \overline{c} = \overline{a} + (\overline{b} + \overline{c})$.
(2) $\overline{0} + \overline{a} = \overline{a} = \overline{a} + \overline{0}$.
(3) $\overline{a} + (\overline{-a}) = \overline{0} = \overline{-a} + \overline{a}$.
(4) $\overline{a} + \overline{b} = \overline{b} + \overline{a}$.
(5) $(\overline{a} \cdot \overline{b}) \cdot \overline{c} = \overline{a} \cdot (\overline{b} \cdot \overline{c})$.
(6) $\overline{1} \cdot \overline{a} = \overline{a} = \overline{a} \cdot \overline{1}$.
(7) $\overline{a} \cdot \overline{b} = \overline{b} \cdot \overline{a}$.

(8) $\overline{a} \cdot (\overline{b} + \overline{c}) = \overline{a} \cdot \overline{b} + \overline{a} \cdot \overline{c}$.
(9) $(\overline{b} + \overline{c}) \cdot \overline{a} = \overline{b} \cdot \overline{a} + \overline{c} \cdot \overline{a}$.

PROOF. Exercise. □

The element $\overline{0}$ is called the *zero* or the *unity* of $\mathbb{Z}/m\mathbb{Z}$ under $+$ and $\overline{1}$ is called the *one* or the *unity* of $\mathbb{Z}/m\mathbb{Z}$ under $\cdot$.

Notation 6.4. As is customary, we shall usually drop the multiplication symbol $\cdot$ in $a \cdot b$ when convenient.

Remark 6.5. Let $\sim$ be an equivalence relation on A. If B is a set, to show that $f : \overline{A} \to B$ is *well-defined*, i.e., f is a function, one must show that

$$\overline{a} = \overline{a'} \quad \Longrightarrow \quad f(\overline{a}) = f(\overline{a'}),$$

i.e., is independent of the representative for $\overline{a}$.

The proposition then also says the canonical surjection

$$^{-} : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \text{ given by } x \mapsto \overline{x}\,(= [x]_m)$$

satisfies

$$\overline{a+b} = \overline{a} + \overline{b} \text{ and } \overline{a \cdot b} = \overline{a} \cdot \overline{b}$$

and

$$0 \mapsto \overline{0} \text{ and } 1 \mapsto \overline{1}.$$

Definition 6.6. A *commutative ring* is a set R together with two maps

$$+ : R \times R \to R \text{ and } \cdot : R \times R \to R,$$

write $+(a, b)$ as $a+b$ and $\cdot(a, b)$ as $a \cdot b$, called *addition* and *multiplication*, respectively, satisfying for all $a, b, c, \in R$ the axioms $(1)-(9)$ above hold with R replacing $\mathbb{Z}/m\mathbb{Z}$. In (2), there exists an element 0 satisfying (2) and (3) and in (6) an element 1 satisfying (6). In particular, R is not empty. If we do not necessarily assume that R satisfies (7), we call R a *ring*.

Examples 6.7. The sets $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/m\mathbb{Z}$ are commutative rings and if R is a ring then so is $\mathbb{M}_n(R)$ and $R[t]$ under the usual $+$ and $\cdot$ of matrices and polynomials, respectively.

We call a map $f : R \to S$ of rings, a *ring homomorphism*, if it preserves $+$, $\cdot$, and takes unities to unities. For example,

$$^{-} : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \text{ is a surjective ring homomorphism.}$$

[A surjective ring homomorphism is also called an *epimorphism*.]

To investigate the interrelationship of congruences of different moduli is an appropriate study. We turn to the important and useful case

of congruences of pairwise relatively prime moduli. We begin with the following lemma.

Lemma 6.8. *Let m, n, a_i, and $1 \leq i \leq r$ be integers.*

(1) If $(a_i, m) = 1$ for $i = 1, \ldots, r$, then $(a_1 \cdots a_r, m) = 1$.

(2) If $(a_i, a_j) = 1$ for $i \neq j$ and $a_i \mid n$ for $i, j = 1, \ldots, r$ then $a_1 \ldots a_r \mid n$.

PROOF. (1). By induction, it suffices to do the case $r = 2$. (Why?) By Key Observation (4.13), we have equations

$$x_1 a_1 + y_1 m = 1 = x_2 a_2 + y_2 m,$$

for some $x_1, x_2, y_1, y_2 \in \mathbb{Z}$, so

$$1 = (x_1 a_1 + y_1 m)(x_2 a_2 + y_2 m) = x_1 x_2 a_1 a_2 + \text{something} \cdot m$$

and we are done.

[**Note**. We only used induction for the convenience of notation – do you see why?]

(2). The case $r = 1$ is immediate. By induction, we have $a_1 \cdots a_{r-1} \mid n$. By (1), we have $(a_1 \cdots a_{r-1}, a_r) = 1$ (setting $m = a_r$). By Key Observation (4.13), there exists an equation

$$a_1 \cdots a_{r-1} x + a_r y = 1 \text{ for some } x, y \in \mathbb{Z}.$$

By Key Trick (4.14), we have

$$a_1 \cdots a_{r-1} n x + a_r n y = n.$$

As $a_1 \cdots a_{r-1} a_r \mid a_1 \cdots a_{r-1} n$ and $a_1 \cdots a_{r-1} a_r \mid a_r n$, we conclude that $a_1 \cdots a_r \mid n$ as needed □

Theorem 6.9. (Chinese Remainder Theorem) *Let m_i be integers with $(m_i, m_j) = 1$ for $1 \leq i, j \leq r$ and $i \neq j$. Set $m = m_1 \cdots m_r$ and suppose that $c_1, \ldots, c_r$ are integers. Then there exists an integer x satisfying all of the following:*

$$(*) \qquad \begin{aligned} x &\equiv c_1 \mod m_1 \\ &\vdots \\ x &\equiv c_r \mod m_r. \end{aligned}$$

Moreover, the integer x is unique modulo m, i.e., if an integer y also satisfies $y \equiv c_i \mod m_i$ for $1 \leq i \leq r$, then $x \equiv y \mod m$.

PROOF. **Existence**. Let $n_i = \frac{m}{m_i} = m_1 \cdots \widehat{m_i} \cdots m_r$ where $\widehat{\ }$ means omit. By Lemma 6.8 (1), we have $(m_i, n_i) = 1$ for $1 \leq i \leq r$, so by Key Observation 4.13, there exist equations

$$1 = d_i m_i + e_i n_i,$$

for some integers d_i, e_i, $i = 1, \ldots, r$. Set $b_i = e_i n_i$, for $i = 1, \ldots, r$. Then

$$1 = d_i m_i + b_i \text{ and } m_j \mid b_i \text{ if } i \neq j.$$

This means that

$$1 \equiv b_i \mod m_i \text{ and } 0 \equiv b_i \mod m_j \text{ if } i \neq j.$$

Consequently,

$$x_0 := c_1 b_1 + \cdots + c_r b_r \equiv c_i b_i \equiv c_i \mod m_i, \text{ with } i = 1, \ldots, r$$

and x_0 works.

Uniqueness. Suppose that y_0 also works, then $x_0 \equiv y_0 \mod m_i$ for $1 \leq i \leq r$, i.e., $m_i \mid x_0 - y_0$ for $1 \leq i \leq r$. By Lemma 6.8 (2), we have $x_0 \equiv y_0 \mod m$. □

We want to interpret what we did above in the language of "rings" and "ring homomorphisms". Let $m > 1$ in $\mathbb{Z}$ and $(a, m) = 1$. By Key Observation (4.13), there are integers x and y satisfying $1 = ax + my$. Applying the canonical surjection $^- : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ to this yields

$$\overline{1} = \overline{ax + my} = \overline{ax} + \overline{my} = \overline{ax} + \overline{0}\overline{y} = \overline{ax}\,(= \overline{xa})$$

i.e., $\overline{a}$ has a *multiplicative inverse* in $\mathbb{Z}/m\mathbb{Z}$. An element $\overline{a}$ having a multiplicative inverse is called a *unit*. Let

$$(\mathbb{Z}/m\mathbb{Z})^\times := \{\overline{a} \in \mathbb{Z}/m\mathbb{Z} \mid \overline{a} \text{ is a unit}\}.$$

So if $a \in \mathbb{Z}$ and $(a, m) = 1$, then $\overline{a} \in (\mathbb{Z}/m\mathbb{Z})^\times$.

Conversely, suppose $a \in \mathbb{Z}$ satisfies $\overline{a} \in (\mathbb{Z}/m\mathbb{Z})^\times$. Then there exists $b \in \mathbb{Z}$ satisfying $\overline{a}\overline{b} = \overline{1}$. In particular, we have $m \mid ab - 1$, so $ab - 1 = mk$ for some integer k, equivalently, $ab - mk = 1$. We therefore have

Conclusion 6.10. Let a and $m > 1$ be integers, then

$$\overline{a} \in (\mathbb{Z}/m\mathbb{Z})^\times \text{ if and only if } (a, m) = 1.$$

We can now interpret Lemma 6.8 (1) and Properties 4.9 (1) as follows: If x and y are integers then

$$\overline{x}, \overline{y} \in (\mathbb{Z}/m\mathbb{Z})^\times \text{ if and only if } \overline{xy} \in (\mathbb{Z}/m\mathbb{Z})^\times.$$

In particular, the set of units $(\mathbb{Z}/m\mathbb{Z})^\times$ of $\mathbb{Z}/m\mathbb{Z}$ is *closed* under multiplication (but not addition), i.e., we can write

$$\cdot : (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/m\mathbb{Z})^\times \to (\mathbb{Z}/m\mathbb{Z})^\times \text{ taking } (\overline{a}, \overline{b}) \to \overline{a} \cdot \overline{b}.$$

We wish to generalize this.

Definition 6.11. Let R be a ring with $1 \neq 0$ and set

$$R^\times := \{r \in R \mid \text{there exists an } a \in R \text{ such that } ar = 1 = ra\},$$

the *units* of the ring R. Note that this set is closed under multiplication.

Question 6.12. What are $\mathbb{Z}^\times$, $\mathbb{Q}^\times$, $\mathbb{R}^\times$, $\mathbb{C}^\times$, $\mathbb{M}_n(\mathbb{R})^\times$, $\mathbb{M}_n(\mathbb{Z})^\times$?

We now interpret the Chinese Remainder Theorem in this new language. As before let m_i, $1 \leq i \leq r$, be integers with $(m_i, m_j) = 1$ if $i \neq j$, and $m = m_1 \cdots m_r$. Then $m_j \mid m$ for all j. In particular, if $m \mid a - b$, then $m_j \mid a - b$ for all j, i.e., if $a \equiv b \mod m$, then $a \equiv b \mod m_j$ for all j. This means that the map

$$\mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/m_j\mathbb{Z} \text{ given by } [a]_m \mapsto [a]_{m_j}$$

is well-defined (i.e., a function). E.g., $7 \equiv 1 \mod 6$, so $7 \equiv 1 \mod 3$ and $7 \equiv 1 \mod 2$. This leads to a map

$$\begin{aligned} \mathbb{Z}/m\mathbb{Z} &\to \mathbb{Z}/m_1\mathbb{Z} \times \cdots \times \mathbb{Z}/m_r\mathbb{Z} \\ \text{given by } [a]_m &\mapsto ([a]_{m_1}, \ldots, [a]_{m_r}). \end{aligned} \tag{6.13}$$

Giving the right hand set component-wise operations turns it into a commutative ring and we see immediately that this map is a ring homomorphism. (Note that the zero of the right hand side is $([0]_{m_1}, \ldots, [0]_{m_r})$ and the one is $([1]_{m_1}, \ldots, [1]_{m_r})$). The Chinese Remainder Theorem says that this map is bijective. The inverse of this map is also checked to be a ring homomorphism. So the Chinese Remainder Theorem says the map above is a *ring isomorphism* , i.e., a bijective ring homomorphism whose inverse is also a ring homomorphism. (We shall see that this last condition is unnecessary.) This means that the rings

$$\mathbb{Z}/m\mathbb{Z} \text{ and } \mathbb{Z}/m_1\mathbb{Z} \times \cdots \times \mathbb{Z}/m_r\mathbb{Z} \text{ are } \textit{isomorphic},$$

i.e., they look the same algebraically. For example, if m and n are relatively prime integers greater than 1, then $\mathbb{Z}/mn\mathbb{Z}$ and $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ are isomorphic. In particular, if $n = p_1^{e_1} \cdots p_s^{e_s}$ is a standard factorization, then $\mathbb{Z}/n\mathbb{Z}$ and $\mathbb{Z}/p_1^{e_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_s^{e_s}\mathbb{Z}$ are isomorphic. This last is useful as it reduces solving many equations modulo n to congruences modulo the various prime power constituents in its standard factorization.

It is left as an exercise to show the isomorphism in (6.13) above induces a bijection between

$$(\mathbb{Z}/m\mathbb{Z})^\times \text{ and } (\mathbb{Z}/m_1\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/m_r\mathbb{Z})^\times.$$

This means if we define the *Euler phi-function* $\varphi : \mathbb{Z}^+ \to \mathbb{Z}^+$ by $\varphi(1) = 1$ and $\varphi(m) = |(\mathbb{Z}/m\mathbb{Z})^\times|$ for $m > 1$, then $\varphi(mn) = \varphi(m)\varphi(n)$ whenever m and n are relatively prime positive integers.

Exercises 6.14. 1. (Fermat's Little Theorem) Let p be a prime, then for all integers a, we have

$$a^p \equiv a \mod p.$$

[You can use the Binomial Theorem if you state it carefully and then prove the *Children's Binomial Theorem* that says for all integers a and b, we have $(a+b)^p \equiv a^p + b^p \mod p$. We will have another proof of this later.]

2. Prove that there exists infinitely many primes p satisfying $p \equiv 3 \mod 4$. [It is true that there exists infinitely many primes p satisfying $p \equiv 1 \mod 4$ but this is harder.]
[Hint. Look at the proof for the infinitude of primes.]
3. Prove Proposition 6.3.
4. Find the smallest positive integer x satisfying the congruences

$$x \equiv 3 \mod 11 \quad x \equiv 2 \mod 12 \quad \text{and} \quad x \equiv 3 \mod 13.$$

5. What are the units of the rings $\mathbb{Q}$, $\mathbb{Z}$, $\mathbb{R}[t]$, $\mathbb{Z}[t]$, $(\mathbb{Z}/4\mathbb{Z})[t]$, and $\mathbb{M}_n(\mathbb{C})$?
6. Show if m and n are relatively prime positive integers, then the map

$$(\mathbb{Z}/mn\mathbb{Z})^\times \to (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times \text{ given by } [a]_{mn} \mapsto ([a]_m, [a]_n)$$

is a bijection. In particular, $\varphi(mn) = \varphi(m)\varphi(n)$ whenever m and n are relatively prime positive integers.
7. Let n and r be positive integers and p a positive prime. Show that $\varphi(p^e) = p^{e-1}(p-1)$ and $\varphi(n) = n \prod_{\substack{p|n \\ p \text{ prime}}} (1 - \frac{1}{p})$.

7. Surjective maps

In the previous sections, we introduced equivalence relations, using it to construct the integers mod m and the corresponding surjection from the integers to the integers mod n. We want to expand the usefulness of equivalence relations. In mathematics, one tries to reduce the study of functions that preserve a given structure into injective and surjective maps that preserve that structure. The reason for studying injective maps is that if one can *embed*, i.e., find an injection of an object into one with more structure, one can often pull back information. The reason to study surjective maps is that usually the target is simpler, e.g., $\mathbb{Z}/m\mathbb{Z}$ is simpler than $\mathbb{Z}$. If we are just studying sets, i.e., without any additional structure, this reduction can always be accomplished. Equivalence relations give the surjective maps.

Exercise 6.14 (1) showed that equivalence relations were essentially the same as partitioning. Let A be a set with $\sim$ an equivalence relation on it. Then the set of equivalence classes, $\overline{A} := \{\overline{a} \mid a \in A\}$, partitioned A, and the exercise says that partitions of A induce an equivalence relations on A. Let $\mathcal{S}$ be a system of representatives for A relative to $\sim$, so $A = \bigvee_{\mathcal{S}} \overline{a}$. We then have the canonical surjection

$$^{-} : A \to \overline{A} \text{ given by } a \mapsto \overline{a},$$

i.e., every equivalence relation leads to a surjective map. We want a converse. Let

$$f : A \to B \text{ be a surjective map.}$$

We define an equivalence relation $\sim$ on A by

$$a \sim a' \text{ if } f(a) = f(a').$$

This is clearly an equivalence relation on A, as $=$ is an equivalence relation on B. We compute the equivalence class of a to be

$$\overline{a} = \{x \in A \mid x \sim a\} = \{x \in A \mid f(x) = f(a)\},$$

called the *fiber* of f at $f(a)$ and let

$$f^{-1}\big(f(a)\big) = \{x \in A \mid f(x) = f(a)\}.$$

[In general, f^{-1} is not a function, but we use the general notation if $f : A \to B$ is any map and $D \subset B$, then $f^{-1}(D) := \{x \in A \mid f(x) \in D\}$. If $D = \{b\}$ (the *preimage* of D), then the *fiber* $f^{-1}(b)$ of f at b is $f^{-1}(\{b\})$.

We now know, if $\sim$ is the equivalence relation defined by the surjective map f, then we have

(*) $\quad \overline{a} = \overline{a'}$ if and only if $a \sim a'$ if and only if $f(a) = f(a')$.

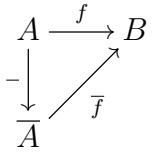
But this means if we define

$$\overline{f} : \overline{A} \to B \text{ by } \overline{a} \mapsto f(a),$$

then

$$\overline{f}(\overline{a}) = f(a),$$

so $\overline{f}$ is a well-defined and an injective map by (*). We say that f *induces* $\overline{f}$. This induced map is also surjective since $\overline{f}(\overline{A}) = f(A)$, hence it is a bijection. So set theoretically we cannot tell the sets $\overline{f}(\overline{A})$ and B apart. Moreover, we have a *commutative diagram*



i.e., $f = \overline{f} \circ {}^{-}$. We call this the *First Isomorphism Theorem of Sets*. [In general, we say a diagram *commutes* if following the composition of any maps (arrows) from the same place to the same target gives equal maps.]

This is so useful that we summarize the above.

Summary 7.1. Let $f : A \to B$ be a surjective map. Define an equivalence relation $\sim$ by $a \sim a'$ if $f(a) = f(a')$. Then

(1) $\sim$ is an equivalence relation.
(2) $\overline{A} = \{f^{-1}\big(f(a)\big) \mid a \in A\}$.
(3) ${}^{-} : A \to \overline{A}$ is given by $a \mapsto \overline{a}$.
(4) $\overline{f} : \overline{A} \to B$ is a well-defined bijection.
(5) The diagram

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle -}\downarrow & \nearrow_{\overline{f}} & \\ \overline{A} & & \end{array} \quad \text{commutes.}$$

We can push this a bit further. Let $g : A \to C$ be a map and set $B = f(A)$. Let $f : A \to B$ be given by $f(a) = g(a)$ for all $a \in A$, i.e., we change the target. Then g is the composition

$$A \xrightarrow{f} B \xrightarrow{\text{inc}} C,$$

with inc the inclusion map.

Notation 7.2. In the future, we shall usually abuse notation and write the same letter for a function and the function arising by changing its target (if it makes sense) when no confusion can arise.

For clarity, we still write g for this f here. Let $\sim$ be the equivalence relation given by $a \sim a'$ if $f(a) = f(a')$. Then we have a commutative diagram

$$\begin{array}{ccc} A & \xrightarrow{g} & C \\ {\scriptstyle -}\downarrow & \searrow^{f} & \uparrow{\scriptstyle \text{inc}} \\ \overline{A} & \xrightarrow[\overline{f}]{} & B. \end{array}$$

If we set $\overline{g} = \text{inc} \circ \overline{f}$, we get a commutative diagram

$$\begin{array}{ccc} A & \xrightarrow{g} & C \\ {\scriptstyle -}\downarrow & \nearrow_{\overline{g}} & \\ \overline{A} & & \end{array} \tag{7.3}$$

with $^-$ a surjection and $\overline{g}$ an injection, the *First Isomorphism Theorem (alternate version)*, showing that every map *factors* as a composition injection $\circ$ surjection. Note that the equivalence classes of $\sim$ are precisely the non-empty fibers of the map g.

If we give our sets additional structure, we would be interested in maps that preserve this additional structure. In general, the map $\overline{g}$ may not preserve the additional structure nor the set $\overline{A}$ have a compatible structure. We shall see in the sequel that in certain cases we can achieve this.

The map $g : A \to C$ is also a composition of surjection $\circ$ injection. Indeed let $\Gamma_g : A \to A \times C$ be given by $a \mapsto \big(a, g(a)\big)$ (called the *graph of* g) and $\pi_C : A \times C \to C$ by $(a, c) \mapsto c$, the *projection* of $A \times C$ onto C, then $g = \pi \circ \Gamma_g$. This shows that to study (set) maps, it suffices to study injections and surjections.

Part 2

Group Theory

CHAPTER III

Groups

In this chapter, we begin our study of abstract algebra. Our basic example is a *group*, a set G with one (binary) operation $\circ$ so that $x \circ y$ lies in G for all x and y in G, subject to three axioms: associativity, existence of a unity, and the existence of inverses. An example that you know from linear algebra is the set $\mathrm{GL}_n(\mathbb{R})$ of all $n \times n$ matrices over the reals of determinant nonzero, called the $n \times n$ *general linear group* over $\mathbb{R}$. . Although easy to define, the structure of groups is far from simple. A primary reason for this is that, in general, we do not assume that $x \circ y = y \circ x$ in G.

In algebra, as other fields of mathematics, one studies sets with additional structures by investigating maps between those sets that preserve the additional structure. In particular, one would like to know when two objects with additional structure are essentially the same. Here the prototypes that you have encountered in linear algebra are a linear transformation between vector spaces and an isomorphism of vector spaces. Indeed one idea to keep in mind is that in modern mathematics maps between objects are more important than the objects themselves. In this chapter, we give many examples of groups and maps between them as well as establishing the basic theorems needed to investigate groups.

It is important that you learn the many examples given below (as well as others), as you can only theorize what is true based on examples that you know – of course, your guess may very well not be true. In this course, we shall mostly apply our theorems to groups having finitely many elements, because this allows us to use the power of equivalence relations in an effective way – we can count. The general theory of arbitrary groups is very difficult, but extremely important not only in algebra but other fields of study, because of the general linear group (and its subgroups).

8. Definitions and Examples

A map $\cdot : G \times G \to G$ is called a *binary operation*. We shall always write $a \cdot b$ for $\cdot(a, b)$.

Definition 8.1. Let G be a set with a binary operation $\cdot : G \times G \to G$. We call $(G, \cdot)$ a *group* if it satisfies the following axioms:

Associativity. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in G$.

Unity. There exists an element $e \in G$ such that for all $a \in G$, we $a \cdot e = a = e \cdot a$. The element e is called a *unity* or an *identity*.

Inverses. There exists a unity e in G; and, for any $x \in G$, there exists a $y \in G$ satisfying $xy = e = yx$.

If, in addition, G satisfies

Commutativity. $a \cdot b = b \cdot a$ for all a, b in G,

we call G an *abelian* group.

We usually write G for $(G, \cdot)$ and ab for $a \cdot b$ if $\cdot$ is clear.

The most common algebraic property on a set with a binary operation is associativity, although there are interesting algebraic objects not satisfying it. However, all the algebraic objects that we shall study will satisfy an analogue of associativity. If our set G under its binary operation only satisfies Associativity and Unity, then it is called a *monoid*.

Remarks 8.2. Let $\cdot : G \times G \to G$ be a binary operation.

1. If G satisfies associativity and $a_1, \ldots, a_n \in G$, then $a_1 \cdots a_n$ makes sense, i.e., is independent of parentheses. In particular, if $n \in \mathbb{Z}^+$ then a^n makes sense with $a = a^1$ and $a^n = a(a^{n-1})$ for $n > 1$. If G is a monoid, we let $a^0 = e$. The nasty uninteresting induction on the complexity of parentheses is left to the diligent reader.
2. If G satisfies Unity, then the unity e is unique. Indeed if e' is another unity then $e = ee' = e'$.
   [The unity of an abstract group will usually be written e_G or e but for many specific types of groups the unity will be written as 0 or 1.]
3. If G is a monoid, then $a \in G$ has at most one inverse. If it has an inverse, then it is denoted by a^{-1} in which case a is the inverse of a^{-1}. Indeed if b and c are inverses of a then
   $$b = b \cdot e = b \cdot (a \cdot c) = (b \cdot a) \cdot c = e \cdot c = c.$$
   [In fact, the proof shows if there exists $b, c \in G$ satisfying $ba = e = ac$ then $b = c$ is the inverse of a.]
4. If G is a monoid and $a, b \in G$ have inverses then so does ab and $(ab)^{-1} = b^{-1}a^{-1}$. In addition, if $n \in \mathbb{Z}^+$ then $a^{-n} = (a^{-1})^n$.
   Question. Let a, b be elements in a monoid such that ab has an inverse. Is it true that a and b have inverses?

5. If G is a group, then the *cancellation laws* hold, i.e., for all $a, b, c, \in G$, we have
$$ab = ac \implies b = c \qquad \text{and} \qquad ba = ca \implies b = c.$$
In particular, if $a = xy$ in G, then $x = ay^{-1}$ and $y = x^{-1}a$
6. If G is a group under a binary operation notated by $+ : G \times G \to G$, then G will always be an abelian group. We then call G an *additive* group and write 0 (or 0_G) for e_G and $-a$ for a^{-1}.

Definition 8.3. Using this new language, it is easy to define a *ring* more carefully. Let R be a set with two binary operations $\cdot : R \times R \to R$ and $+ : R \times R \to R$. Then R is a ring under *addition* $+$ and *multiplication* $\cdot$ if $(R, +)$ is an additive group, $(R, \cdot)$ is a monoid, and R satisfies the *Distributive Laws*
$$a \cdot (b + c) = a \cdot b + a \cdot c \text{ and } (b + c) \cdot a = b \cdot a + c \cdot a,$$
for all $a, b, c \in R$. The distributive laws interrelate the two binary operations. The multiplicative unity is written 1 (or 1_R if there can be confusion). The ring is called a *commutative ring* if $(R, \cdot)$ is a commutative monoid. One checks that $1 = 0$ in R if and only if $R = \{0\}$, the *zero* or *trivial ring*. If R is not the zero ring it is called a *division ring* if $(R \setminus \{0\}, \cdot)$ is a group. A commutative division ring is called a *field*.

It is worth mentioning again that when one has defined a new concept, one should always try to find many examples. Theorems usually arise from knowledge of explicit examples. We present many examples of groups.

Examples 8.4. 1. A *trivial* group is a group consisting of a single element, necessarily the unity.
2. $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/m\mathbb{Z}\,(m > 1)$, or any ring is an additive group under $+$.
3. $\mathbb{R}^+$ the set of positive real numbers, is an abelian group but $\mathbb{Z}^+$ is only an abelian (i.e., commutative) monoid under multiplication with unity $e = 1$, and not a group. Neither are monoids under addition, although they still satisfy associativity.
4. If F is a field, then by definition $F^\times = F \setminus \{0\}$ is an abelian group under multiplication, e.g., if $F = \mathbb{Q}, \mathbb{R}$, or $\mathbb{C}$. More generally, if R is any ring then its set of units $R^\times$, i.e., the set of elements having a multiplicative inverse is a group under $\cdot$, abelian if R is commutative, is called the *group of units of* R. This leads to many examples of groups.
5. Let V be a vector space over a field F, then $(V, +)$ is an additive group.

6. Let S be a non-empty set, then

$$\Sigma(S) := \{f : S \to S \mid f \text{ is a bijection}\}$$

is a group under the composition of functions. The unity is the identity map on S which we shall write as 1_S. A bijection $f : S \to S$ is called a *permutation*. If $f \in \Sigma(S)$, what is f^{-1}? We call $\Sigma(S)$ the *group of all permutations* of S. It is also a *transitive* group on S, that is for all $x, y \in S$, there exists a permutation $f \in \Sigma(S)$ satisfying $f(x) = y$. If $S = \{1, \ldots, n\}$ then $\Sigma(S)$ is denoted S_n and called the *symmetric group on n letters* . Note that $|S_n| = n!$.

Question. Let $S = \{a, b, \ldots, z\}$. Can you tell the difference between $\Sigma(S)$ and S_{26} algebraically?

For the next examples, we need the definition of a subgroup of a group. A subset H of a group G is called a *subgroup* of G if it becomes a group under the restriction of the binary operation on G to H, i.e., $\cdot|_{H\times H} : H \times H \to H$ makes sense (meaning that the image of $\cdot|_{H\times H}$ lies in H) and has the same unity as G (although this last condition will be seen to be unnecessary).

7. Let S be a non-empty set and $x_0 \in S$. The set

$$\Sigma(S)_{x_0} := \{f \in \Sigma(S) \mid f(x_0) = x_0\} \subset \Sigma(S),$$

is a group called the *stabilizer* of x_0 in $\Sigma(S)$. We say elements of $\Sigma(S)_{x_0}$ *fix* x_0. We shall see that groups often arise as functions acting on sets and a major problem is to find *fixed points* of this action, e.g., in the above x_0 is a *fixed point* of the action of $\Sigma(S)_{x_0}$ on S. For example $(S_n)_n$ looks algebraically like S_{n-1}.

$\Sigma(S)_{x_0}$ is also a *subgroup* of $\Sigma(S)$,
We can generalize the stabilizer of x_0 as follows: Let $x_0, \ldots, x_n$ be elements of S, then

$$\Sigma(S)_{x_0} \cap \cdots \cap \Sigma(S)_{x_n} = \{f \in \Sigma(S) \mid f(x_i) = x_i, \text{ for } i = 1, \ldots, n\}$$

is a subgroup of $\Sigma(S)$ (as is $\Sigma(S)_{x_i}$ for all i) stabilizing (fixing) $x_0, \ldots, x_n$.

8. Let G be a group, H_i with $i \in I$ be subgroups of G. Then $\bigcap_I H_i$ is a subgroup of G. In general, $\bigcup_I H_i$ is not a subgroup of G. [Can you find a condition when it is?]
9. Let G be a group and $W \subset G$ a subset. Set

$$\mathcal{W} = \{H \subset G \mid H \text{ is a subgroup of } G \text{ with } W \subset H\}.$$

As $W \subset G$, we have $\mathcal{W} \neq \emptyset$. Let

$$\langle W \rangle := \bigcap_{\mathcal{W}} H = \bigcap_{\substack{W \subset H \subset G \\ H \text{ subgroup of } G}} H.$$

This is the unique smallest subgroup of G containing W. (Can you show this?) We say that W *generates* $\langle W \rangle$ and that W is a *set of generators* for $\langle W \rangle$. A set of generators is not unique. If $W = \{a_1, \ldots, a_n\}$ is finite, we also write $\langle a_1, \ldots, a_n \rangle$ for $\langle \{a_1, \ldots, a_n\} \rangle$. We say G is *finitely generated* if there exists a finite set W such that $G = \langle W \rangle$ and *cyclic* if there exists an element $a \in G$ such that $G = \langle a \rangle$. If this is the case then $G = \{a^n \mid n \in \mathbb{Z}\}$ and is abelian. For example, $(\mathbb{Z}, +)$ and $(\mathbb{Z}/m\mathbb{Z}, +)$, $m > 1$, are cyclic with generators 1 and $\overline{1}$, respectively.

Question. What are the analogues of the above for vector spaces? The analogy will not go too far, as the concept of linear independence almost always fails.

Warning. If G is a group that can be generated by two elements, G need not be abelian. As an example consider a figure eight in the plane and take the group generated by the two counterclockwise rotations of 360^o, each one rotating around its respective circle starting from the intersection point. (Can you see why this group is not abelian?)

10. Let

$$T := \{z \in \mathbb{C} \mid |z| = 1\},$$

where $|z| = \sqrt{z\overline{z}}$ with $\overline{z}$ the complex conjugate of z. This is an abelian group under multiplication, called the *circle group*. It is a subgroup of $\mathbb{C}^\times$. If $n \in \mathbb{Z}^+$ then

$$\mu_n := \{z \in T \mid z^n = 1\} = \langle e^{2\pi\sqrt{-1}/n} \rangle$$

is a cyclic subgroup of T called the *group of nth roots of unity*. Another subgroup of T is

$$\{z \in T \mid z \in \mu_n \text{ for some } n \in \mathbb{Z}^+\} = \bigcup_{\mathbb{Z}^+} \mu_n.$$

Note that a subgroup of an abelian group is abelian.

11. The symmetries of a geometric object often form a group. We look at some special cases. Consider an equilateral triangle in the plane with a side on the X-axis labeling the ordered vertices A, B, C, with line segment AB the base. Let r be a counterclockwise rotation of 120^o. The triangle looks the same but the vertices are now ordered

C, A, B, with base CA. Composing rotations shows $r^2 = r \circ r$ takes the triangle to B, C, A, and r^3 back to A, B, C, i.e., we have the *relation* r^3 is the identity. So $\langle r \rangle$ is a cyclic group of three elements. (Should we be able to compare this group to μ_3?) Now, viewing the plane in $\mathbb{R}^3$ let f denote the flip along the line of the apex of the triangle perpendicular to the base. It takes the ordered vertices A, B, C to A, C, B. Under composition, we have the *relation* f^2 is the identity which we write as 1 and $\langle f \rangle$ is a cyclic group of two elements. (Should we be able to compare this group to μ_2 and to $(\mathbb{Z}/2\mathbb{Z}, +)$ and to $(\mathbb{Z}^\times, \cdot)$?) Composing f and r leads to the *relations* $f^{-1} \circ r \circ f = r^2 = r^{-1}$ and a non-abelian group with six elements, viz.,

$$\{1, r, r^2, f, fr, rf\}.$$

Show this. We say that r and f generate this group subject to the relations $r^3 = 1$, $f^2 = 1$, and $f^{-1}rf = r^{-1}$. Note that it follows that $r^2 = r^{-1}$. It is customary to write this as

$$\langle r, f \mid r^3 = 1,\ f^2 = 1, f^{-1}rf = r^{-1} \rangle,$$

i.e., in the form

$$\langle \text{generators} \mid \text{relations} \rangle.$$

This group is called the symmetries of an equilateral triangle or the *dihedral group of order six* and denoted by D_3. (Cf. D_3 and S_3.) More generally, consider a regular n-gon, $n \geq 3$, with r a counterclockwise rotation of $2\pi/n$ radians and f a flip along the perpendicular at the bisection point of the base. Then under composition, we get a non-abelian group with $2n$ elements. It is a group that is defined by two *generators* r and f satisfying the three *relations*

$$r^n = 1 \qquad f^2 = 1 \qquad f^{-1} \circ r \circ f = r^{-1}.$$

It follows that $f^{-1} \circ r \circ f = r^{n-1}$. It is called the symmetries of the regular n-gon or the *dihedral group of order* $2n$ and denoted by D_n. As above, this is usually written

$$D_n := \langle r, f \mid r^n = 1,\ f^2 = 1, f^{-1}rf = r^{-1} \rangle.$$

Note if $n > 3$ then D_n and S_n have a different number of elements.

12. Let $Q = \{1, -1, i, -i, j, -j, k, -k\}$. This becomes a non-abelian group if we invoke the relations $(-1)^2 = 1$, $k = ij = -ji$ and $i^2 = j^2 = -1$ called the *quaternion group*.
13. Let $F = \mathbb{Q}, \mathbb{R}, \mathbb{C}$, or any field. Then

$$\mathrm{GL}_n(F) := \{A \in \mathbb{M}_n(F) \mid \det A \neq 0\}$$

is a group under matrix multiplication called the *general linear group of degree n*. It and its subgroups are probably the most important groups in mathematics. If $n = 1$ then $\mathrm{GL}_1(F) = F^\times$ but if $n > 1$, this group is not abelian. (Proof?).

More generally, let R be any ring. Then the set of units $\big(\mathbb{M}_n(R)\big)^\times$ is a group under multiplication of matrices. If R is commutative, it turns out that the determinant of a matrix still makes sense and $A \in \big(\mathbb{M}_n(R)\big)^\times$ if and only if $\det(A) \in R^\times$, so we let

$$\mathrm{GL}_n(R) := \big(\mathbb{M}_n(R)\big)^\times$$

also called the *general linear group of degree n*. If $A \in \mathbb{M}_n(R)$, we denote its ijth entry by A_{ij}. Then as usual, its *transpose* is A^t, where $(A^t)_{ij} = A_{ji}$ and if $R = \mathbb{C}$, its *adjoint* or complex conjugate transpose is A^* where $(A^*)_{ij} = \overline{A_{ji}}$. If A is a diagonal matrix, we also write $A = \mathrm{diag}(a_1, \ldots, a_n)$, e.g., the identity is the matrix $I = I_n = \mathrm{diag}(1, \ldots, 1)$.

We define some of the interesting subgroups of the general linear group over a commutative ring R (the same definitions hold for over an arbitrary ring except when the determinant is involved).

$\mathrm{SL}_n(R) = \{A \in \mathrm{GL}_n(R) \mid \det A = 1\}$	*special linear group*
$\mathrm{O}_n(R) = \{A \in \mathrm{GL}_n(R) \mid A^t A = I\}$	*orthogonal group*
$\mathrm{SO}_n(R) = \mathrm{SL}_n(R) \cap \mathrm{O}_n(R)$	*special orthogonal group*
$\mathrm{D}_n(R) = \{A \in \mathrm{GL}_n(R) \mid A \text{ diagonal}\}$	*diagonal group*
$\mathrm{T}_n(R) = \{A \in \mathrm{GL}_n(R) \mid A_{ij} = 0 \text{ if } i < j\}$	*upper triangular group*
$\mathrm{ST}_n(R) = \{A \in \mathrm{T}_n(R) \mid A_{ii} = 1 \text{ all } i\}$	*strictly upper triangular group*
Let $J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$ in $\mathrm{GL}_{2n}(R)$, then	
$\mathrm{Sp}_{2n}(F) = \{A \in \mathrm{GL}_n(R) \mid A^t J A = J\}$	*symplectic group*
Let $I_{3,1} = \mathrm{diag}(1, 1, 1, -1)$ in $\mathrm{GL}_4(\mathbb{R})$, then	
$\mathrm{O}_{3,1}(\mathbb{R}) = \{A \in \mathrm{GL}_4(\mathbb{R}) \mid A^t I_{3,1} A = I_{3,1}\}$	*Lorenz Group*
$\mathrm{U}_n(\mathbb{C}) = \{A \in \mathrm{GL}_n(\mathbb{C}) \mid A^* A = I\}$	*unitary group*
$\mathrm{SU}_n(\mathbb{C}) = \mathrm{SL}_n(\mathbb{C}) \cap \mathrm{U}_n(\mathbb{C})$	*special unitary group.*

14. Let V be a vector space over a field F. Then

$$\mathrm{Aut}_F(V) := \{T : V \to V \mid T \text{ a linear (i.e., vector space) isomorphism}\}$$

is a group under composition called the *automorphism group* of V. An isomorphism of a vector space to itself is called an *automorphism*. This generalizes greatly, and its generalization probably is the most important type of group, e.g., cf. $\mathrm{Aut}_F(V)$ and $\mathrm{GL}_n(F)$ if V is an n-dimensional vector space.

15. Let G_i, $i \in I$, be groups and $G = \bigtimes_I G_i$ the *cartesian product* of the sets G_i, $i \in I$. We write elements in G by $(g_i)_I$. [Technically,

$$\bigtimes_I G_i := \{f : I \to \bigcup_I G_i \mid f(i) \in G_i \text{ for all } i \in I\}.]$$

Then G is a group under *component-wise operation*, e.g., $e_G = (e_{G_i})_I$. This group is called the *external direct product* of the G_i's. If all the G_i are abelian, so is $\bigtimes_I G_i$.

For example, writing $\mathbb{Z}/2\mathbb{Z} = \{\overline{0}, \overline{1}\}$, we have an abelian group

$$V := \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} = \{(\overline{0},\overline{0}), (\overline{0},\overline{1}), (\overline{1},\overline{0}), (\overline{1},\overline{1})\}.$$

It satisfies:

$$\begin{aligned}
(a,b) + (a,b) &= (\overline{0},\overline{0}) \text{ if } a,b \in \mathbb{Z}/2\mathbb{Z} \\
(\overline{1},\overline{0}) + (\overline{0},\overline{1}) &= (\overline{1},\overline{1}) \\
(\overline{1},\overline{1}) + (\overline{0},\overline{1}) &= (\overline{1},\overline{0}) \\
(\overline{1},\overline{1}) + (\overline{1},\overline{0}) &= (\overline{0},\overline{1}) \\
&\textit{etc.}
\end{aligned}$$

This group is called the *Klein Four (Vier) Group*. Note that this is an abelian group that is not cyclic. Also it is worth noting that V is a two-dimensional vector space over the field $\mathbb{Z}/2\mathbb{Z}$.

16. Let $m > 0$. Then we know that $(\mathbb{Z}/m\mathbb{Z}, +)$ is cyclic and $\big((\mathbb{Z}/m\mathbb{Z})^\times, \cdot\big)$ is abelian but the latter may not be cyclic. What is $(\mathbb{Z}/8\mathbb{Z})^\times$ or more generally $(\mathbb{Z}/2^n\mathbb{Z})^\times$ if $n > 2$? The answer is interesting, as it caused a well-known theorem in Number Theory called the Grünwald's Theorem to be in fact false. The correction by Wang had to take the answer to this question into account.
17. If $a, b \in \mathbb{Z}^+$ have d as their gcd, then the additive group $\langle a, b\rangle = \langle d\rangle$.

Exercises 8.5. 1. Let a, b be elements in a monoid such that ab has an inverse. Is it true that a and b have inverses? Prove this if true, give a counterexample if false.

2. If H_i, $i \in I$, are subgroups of a group G, show that $\bigcap_I H_i$ is a subgroup of G, but $\bigcup_I H_i$ is not a subgroup in general (i.e., give a

counterexample). Find a nontrivial condition on the H_i such that $\bigcup_I H_i$ is a subgroup.

3. Show if G is a group in which $(ab)^2 = a^2b^2$ for all $a, b \in G$, then G is abelian.
4. Determine all groups having at most six elements.
5. Show the quaternion group is a group of eight elements.
6. Show that the dihedral group D_n has $2n$ elements.
7. Let G be a group and H a subgroup of G. Define the *core* of H in G by $\mathrm{Core}_G(H) := \bigcap_G xHx^{-1}$. Show $\mathrm{Core}_G(H)$ is the unique largest subgroup N of H satisfying $xNx^{-1} = N$ for all x in G.
8. Let G be a group and H a subgroup of G, Define the *normalizer* of H in G by $N_G(H) := \{x \in G \mid xHx^{-1} = H\}$. Show that $N_G(H)$ is a subgroup of G containing H and is the unique largest subgroup K of G containing H satisfying $H = xHx^{-1}$ for all x in K.
9. Show if G is a group in which $(ab)^n = a^n b^n$ for all $a, b \in G$ and $n = i, i+1, i+2$ for some positive integer i, then G is abelian.
10. Let $W \subset G$. Show that $\langle W \rangle$ is the unique smallest subgroup of G containing W.
11. If G is a group and $\emptyset \neq W \subset G$, a subset, then

$$\langle W \rangle = \{g \in G \mid \text{There exist } w_1, \ldots, w_r \in W (\text{ some } r) \text{ not necessarily distinct such that } g = w_1^{e_1} \cdots w_r^{e_r},\ e_1, \ldots, e_r \in \mathbb{Z}\},$$

i.e., W is an "alphabet" for the "words", i.e., elements in G. Unfortunately, spelling may not be unique in G for any "alphabet".

12. Let F be a field (or any non-trivial ring). Show that $\mathrm{GL}_n(F)$ is not abelian for any $n > 1$.
13. Let p be a prime and $F = \mathbb{Z}/p\mathbb{Z}$. Show that F is a field.
[Hint: First show that F is a *domain*, i.e., a commutative ring satisfying: whenever $a, b \in F$ satisfy $ab = 0$ in F then $a = 0$ or $b = 0$. Then show that any domain with finitely many elements is a field.]
14. Compute $(\mathbb{Z}/8\mathbb{Z})^\times$. Find a different group isomorphic to it. Can you do the same for $(\mathbb{Z}/2^n\mathbb{Z})^\times$ if $n = 4, 5$? if n is any integer at least three?
15. Let $\mathcal{F} := \{f \mid f : \mathbb{Z}^+ \to \mathbb{C}\}$ and I in $\mathcal{F}$ be given by $I(1) = 1$ and $I(n) = 0$ for all $n > 1$. Show that $\mathcal{F}$ is an abelian monoid and $\mathcal{A} := \{f \mid f \in \mathcal{F} \text{ with } f(1) \neq 0\}$ is an abelian group with unity I under the *Dirichlet product* given by

$$(f \star g)(n) := \sum_{d|n} f(d) g(\frac{n}{d}).$$

9. First Properties

Recall that we defined a subgroup of a group as follows:

Definition 9.1. A subset H of a group G is called a *subgroup* of G if it becomes a group under the restriction of the binary operation on G to H, i.e., $\cdot|_{H\times H} : H \times H \to H$ makes sense (meaning that the image of $\cdot|_{H\times H}$ lies in H) and has the same unity as G.

We begin this section with a criterion for a subset of a group to be a subgroup. This can be compared to the analogous criterion for sets to be subspaces of a vector space.

Proposition 9.2. *Let G be a group and H be a non-empty subset of G. Then H is a subgroup of G if and only if the following two conditions hold:*

(*i*) *If a and b are elements of H, then ab is an element of H.*
[We say that H is *closed* under $\cdot$, i.e., this means that the restriction

$$\cdot|_{H\times H} : H \times H \to H$$

is defined. Note we have changed the target from G to H.]

(*ii*) *If $a \in H$ then $a^{-1} \in H$.*
Moreover, these two conditions are equivalent to

(*) $$\text{If } a, b \in H \quad \text{then} \quad ab^{-1} \in H.$$

Proof. ($\Rightarrow$) follows from the definition of subgroup, since by definition, $e_G = e_H$.

($\Leftarrow$): We first note that (*i*) and (*ii*) implies (*) for if $a, b \in H$, then $a, b^{-1} \in H$, hence $ab^{-1} \in H$. Conversely, suppose that (*) holds. As $a \in H$ implies $a, a \in H$, we have $e_G = aa^{-1} \in H$. By definition, $e_G a = a = ae_G$ for all $a \in H$, so $e_G = e_H$. As $e_G, a \in H$, we have $e_G a^{-1} = a^{-1}$ lies in H. We conclude that if $a, b \in H$, then $a, b^{-1} \in H$, hence $ab = a(b^{-1})^{-1} \in H$. Finally, since associativity holds in G, it holds in the subset H. □

Note that property (*) implies that we do not need to assume in the definition of a subgroup that it has the same unity as the group.

Corollary 9.3. *Let G be a group and $H \subset G$ a non-empty subset. If H is a finite set, then H is a subgroup if and only if H is closed under $\cdot$.*

PROOF. We need only check:
($\Leftarrow$): It suffices to show if $a \in H$ then $a^{-1} \in H$. Our hypothesis implies that

$$S := \{a^n \mid n \in \mathbb{Z}^+\} \subset H \text{ is also a finite set.}$$

We need the following, whose proof we leave as an exercise:

Dirichlet Pigeon Hole Principle: If $N \geq n+1$ objects are placed in $\leq n$ boxes, then at least one box contains at least two elements.

In our case, $|S| < \infty$, so there exists integers $n > m \geq 1$ satisfying $a^n = a^m$, i.e., $a^{n-m} = a^n a^{-m} = e_G = a a^{n-m-1}$. Therefore, e_G lies in H and $a^{n-m-1} = a^{-1}$ with $n - m - 1 > 0$ lies in H. □

Note that this proof shows that if G is a group with $a \in G$ and $\langle a \rangle$ finite, then there exists $N \in \mathbb{Z}^+$ such that $a^N = e$ and by well-ordering there exists a minimal such N.

If G is a group, we call $|G|$ the *order* of G, so G is a *finite* group if it has finite order, *infinite* if not. If $a \in G$, we call $|\langle a \rangle|$ the order of a and say a has *finite order* if $|\langle a \rangle|$ is finite.

Examples 9.4. 1. $(\mathbb{Z}, +) = \langle 1 \rangle$ is an infinite cyclic group.
2. If $n \in \mathbb{Z}^+$ then $(\mathbb{Z}/n\mathbb{Z}, +) = \langle \overline{1} \rangle$ is a cyclic group of order n as is $\mu_n = \langle e^{2\pi\sqrt{-1}/n} \rangle$.
3. $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ is an abelian group of order four and not cyclic.
4. If $n > 2$ then D_n is a non-abelian group of order $2n$.
5. The quaternion group is a non-abelian group of order 8. Does it look the same algebraically as D_4?
6. If $n > 2$, then S_n is a non-abelian group of order $n!$.
7. $(\mathbb{Z}/n\mathbb{Z})^\times$ is an abelian group of order $\varphi(n)$.
8. $\mathrm{GL}_n(\mathbb{R})$, $n > 1$, is an infinite non-abelian group.

We need to know how to determine when two groups are the same algebraically.

Let $\varphi : G \to G'$ be a map of groups. We call φ a *group homomorphism* if

$$\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b) \text{ for all } a, b \in G$$
$$\varphi(e_G) = e_{G'}.$$

Note in the first equation $\cdot$ on the left hand side is in G and on the right hand side is in G'. We shall also see that we really do not need the second equation. We put it in because "homomorphisms" of algebraic structures should preserve those structures, and the unity is a special

element of a group. That it is preserved given the first equation is fortuitous but should not *a priori* be assumed.

If φ above is a (group) homomorphism and if, in addition, φ is:

(1) injective, we say φ is a *(group) monomorphism* or *monic*.
(2) surjective, we say φ is a *(group) epimorphism* or *epic*.
(3) bijective and φ^{-1} is a homomorphism, we say φ is a *(group) isomorphism*.

We let

$$\ker \varphi := \{a \in G \mid \varphi(a) = e_{G'}\}$$

called the *kernel* of φ and

$$\operatorname{im} \varphi := \{\varphi(a) \in G' \mid a \in G\}$$

called the *image* of φ.

If G and G' are groups and there exists an isomorphism $\varphi : G \to G'$, which we also write as $\varphi : G \xrightarrow{\sim} G'$,, we say that G and G' are *isomorphic* and write $G \cong G'$.

Remark 9.5. Let $\varphi : G \to G'$ be a group homomorphism. We leave it as an exercise to show that φ is a monomorphism if and only if given any group homomorphisms $\psi_1, \psi_2 : H \to G$ with compositions satisfying $\varphi \circ \psi_1 = \varphi \circ \psi_2$, then $\psi_1 = \psi_2$; and φ is an epimorphism if and only if given any group homomorphisms $\theta_1, \theta_2 : G' \to H$ with compositions satisfying $\theta_1 \circ \varphi = \theta_2 \circ \varphi$, then $\theta_1 = \theta_2$. (Cf. Exercise 1.12(5) and (6).)

Remark 9.6. You should know that a bijective linear transformation of vector spaces is an isomorphism, i.e., the inverse is automatically linear. The same is true for groups, i.e., a bijective homomorphism of groups is an isomorphism. We shall see this is also true in other algebraic cases. The proof is essentially the same as that for bijective linear transformations.

A group homomorphism satisfies the following:

Properties 9.7. Let $\varphi : G \to G'$ be a group homomorphism.

1. $\varphi(e_G) = e_{G'}$.
2. $\varphi(a^{-1}) = \varphi(a)^{-1}$ for all $a \in G$.
3. $\ker \varphi \subset G$ and $\operatorname{im} \varphi \subset G'$ are subgroups.
4. φ is monic if and only if $\ker \varphi = \{e_G\}$.
5. φ is epic if and only if $\operatorname{im} \varphi = G'$.

PROOF. (1). $e_{G'}\varphi(e_G) = \varphi(e_G) = \varphi(e_G e_G) = \varphi(e_G)\varphi(e_G)$.
By Cancellation, $e_{G'} = \varphi(e_G)$.
(2). $e_{G'} = \varphi(e_G) = \varphi(aa^{-1}) = \varphi(a)\varphi(a^{-1})$. Similarly, $e_{G'} = \varphi(a^{-1})\varphi(a)$.
(3), (5) are left as exercises.
(4). We have $\varphi(a) = \varphi(b)$ if and only if

$$e_{G'} = \varphi(a)\varphi(b)^{-1} = \varphi(a)\varphi(b^{-1}) = \varphi(ab^{-1})$$

if and only if $ab^{-1} \in \ker\varphi$. So $\ker\varphi = \{e_G\}$ if and only if $\varphi(a) = \varphi(b)$ implies $a = b$. □

Examples 9.8. 1. Let G and H be groups, then the map $\varphi : G \to H$ given by $x \mapsto e_H$ is a group homomorphism, called the *trivial homomorphism*.

2. Let H be a subgroup of G. Then the inclusion map of H in G is a group homomorphism. Indeed this observation leads to a better way to define subgroup, viz., let $H \subset G$ be a subset with both H and G groups. Then H is a subgroup of G if and only if the inclusion map is a monomorphism.

3. Let F be a field then the map

$$\det : \mathrm{GL}_n(F) \to F^\times \text{ given by } A \to \det A$$

is an epimorphism with $\ker\det = \mathrm{SL}_n(F)$. (The same is true if F is just a commutative ring.)

4. Let $n \in \mathbb{Z}^+$. Then the map $^- : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ given by $x \mapsto \overline{x}$ is an epimorphism of additive groups with $\ker^- = n\mathbb{Z} := \{kn \mid k \in \mathbb{Z}\}$.

5. Let $n \in \mathbb{Z}^+$. Then the map $\varphi : \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}^\times$ given by $\overline{x} \mapsto e^{2\pi\sqrt{-1}x/n}$ is a well-defined (why?) map, and is a monomorphism with $\mathrm{im}\,\varphi = \mu_n$.

6. Let G be an abelian group. Then the map $\varphi : G \to G$ given by $x \mapsto x^{-1}$ is a group homomorphism. If G is not abelian, this map is never a group homomorphism.

Classifying objects up to isomorphism is usually impossible and difficult when possible. Indeed some of the most important theorems arise when this can be accomplished. However, in a few cases it is quite elementary. For example, two finite sets are isomorphic (i.e., bijective) if and only if they have the same number of elements, and two finite dimensional vector spaces over a field F are isomorphic if and only if they have the same dimension. Another is the case for cyclic groups that we now show.

Theorem 9.9. (Classification of Cyclic Groups) *Let $G = \langle a \rangle$ be a cyclic group and $\theta : \mathbb{Z} \to G$ the map given by $m \to a^m$. Then θ is a group epimorphism. It is an isomorphism if and only if G is infinite. If G is finite, then $|G| = n$ if and only if $\ker \theta = n\mathbb{Z}$. If this is the case, then the map $\overline{\theta} : \mathbb{Z}/n\mathbb{Z} \to G$ given by $\overline{m} \mapsto a^m$ induced by θ is a group isomorphism.*

PROOF. As $\theta(i + j) = a^{i+j} = a^i a^j = \theta(i)\theta(j)$, the map is a group homomorphism. It is clearly an epimorphism.

We must show that θ is an isomorphism if and only if $|G|$ is infinite and if not then $\ker \theta \neq \{e\}$ to establish the first part of the theorem.

We have θ is one-to-one if and only if $a^i \neq a^j$ whenever $i \neq j$ if and only if $a^m \neq e_G$ for all $m \neq 0$ if and only if θ is an isomorphism (as it is onto), in which case G is infinite.

So we may assume that there exists $N \in \mathbb{Z}^+$ with $a^N = e_G$, hence θ is not one-to-one. We show that this means that G is finite. By the Well-Ordering Principle, there exists a least $n \in \mathbb{Z}^+$ such that $a^n = e_G$. We show

Claim 9.10. $a^i = a^j$ if and only if $i \equiv j \mod n$. In particular, $a^m = e_G$ if and only if $n \mid m$.

Note that the claim implies that $|G| = n < \infty$ and $\ker \theta = n\mathbb{Z}$. So suppose that the integers i and j are not equal. We may assume that $j > i$. Set $m = j - i \in \mathbb{Z}^+$ and write $m = kn + r$ with $0 \leq r < n$ and k integers using the Division Algorithm. We know that

$$a^i = a^j \text{ if and only if } a^m = a^{j-i} = e_G \text{ if and only if } a^r = (a^n)^k a^r = a^m.$$

By the minimality of n, this can occur if and only if $n \mid m$ if and only if $m \equiv 0 \mod n$ if and only if $i \equiv j \mod n$ as required.

We know by the First Isomorphism of Sets 7.1 that θ induces a bijection $\overline{\theta}$. (Make sure that you understand this.) It is clearly a homomorphism, so a group isomorphism. □

Subgroups of a cyclic group are classified by the following, whose proof we leave as an exercise (that you should do):

Theorem 9.11. (Cyclic Subgroup Theorem) *Let $G = \langle a \rangle$ be a cyclic group, $H \subset G$ a subgroup. Then the following are true:*

1. *$H = \{e\}$ or $= \langle a^m \rangle$ with $m \geq 1$ the least positive integer such that $a^m \in H$. If $|G| = n\,(< \infty)$ then $m \mid n$. If G is infinite then $|H| = 1$ or H is infinite.*
2. *If $|G| = n$ and $m \mid n$ then $\langle a^m \rangle$ is the unique subgroup of G of order $\frac{n}{|m|}$.*

(3) *If* $|G| = n$ *and* $m \not| \, n$ *then* G *has no subgroup of order* m.
(4) *If* $|G| = n$, *the number of subgroups of* G *is equal to the number of positive divisors,* $\sum_{d|n} 1$, *of* n *(where in such a sum, we always assume* $d > 0$*).*
(5) *If* $|G|$ *is a prime then* $\{e\}$ *and* G *are the only subgroups of* G.

Exercises 9.12. 1. Let p be a prime and $F = \mathbb{Z}/p\mathbb{Z}$. Then F is a field by Exercise 8.5 (13). Compute $|G|$ for $G = \mathrm{GL}_n(F)$, $\mathrm{SL}_n(F)$, $\mathrm{T}_n(F)$, $\mathrm{ST}_n(F)$, and $D_n(F)$.

2. Prove the Dirichlet Pigeon Hole Principle.

3. Prove Remark 9.5.

4. Prove Property 9.7 (3) and Remark 9.6, i.e., a bijective group homomorphism is a group isomorphism. In particular, if $\varphi : G \to G'$ is a monomorphism then G is isomorphic to $\varphi(G)$.

5. Let G_i, $i \in I$, be groups. Show that $\bigtimes_I G_i$ satisfies the following property relative to the group homomorphisms $\pi_j : \bigtimes_I G_i \to G_j$ defined by $\{g_i\}_I \mapsto g_j$ for all $j \in I$: Given a group G and group homomorphisms $\phi_j : G \to G_j$, there exist a unique group homomorphism $\psi : G \to \bigtimes_I G_i$ satisfying $\phi_j = \pi_j \circ \psi$ for all $j \in I$.

6. Prove the Cyclic Subgroup Theorem 9.11.

7. Let G be an abelian group. Suppose that the elements a, b of G are of relatively prime orders m and n respectively. Show ab has order mn. Is this true if G is not abelian? Prove or give a counterexample.

8. Find a group isomorphic to $\mathrm{ST}_2(\mathbb{R})$ besides itself. Prove your answer.

9. Show that the quaternion group Q and the dihedral group D_4 are non-isomorphic groups of order 8.

10. Let G be the subgroup of $\mathbb{M}_2(\mathbb{C})$ generated by

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \text{ and } \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}.$$

Show G is isomorphic to the quaternion group Q.

10. Cosets

We now define an equivalence on a group G mimicking congruence modulo m for integers. Although the similarity is precise, the reason for doing so is not so clear. Later we shall see how this equivalence arises in a natural way.

Let H be a subgroup of G. Define $\equiv \mod H$ by

$$\text{if } a, b \in G \text{ then } a \equiv b \mod H \text{ if and only if } b^{-1}a \in H.$$

Then $\equiv \mod H$ is an equivalence relation. (Show this.) If $a \in G$, the equivalence class $\overline{a}$ of a is called the *left coset* of a relative to H. Analogous to $\equiv \mod m$ for integers, we shall usually write aH for $\overline{a}$, the left coset of $a \in G$ relative to H, — remember the binary operation for $\mathbb{Z}$ is $+$ — and shall write G/H for $\overline{G} = G/\equiv \mod H$. So we have the natural set surjection

$$^{-} : G \to G/H \text{ given by } a \mapsto \overline{a} = aH.$$

Warning 10.1. In general, G/H is not a group, so the natural surjection above is not a group homomorphism. (Can you give an example?)

Observation 10.2. Let $H \subset G$ be a subgroup and $H = eH$. ($e = e_G = e_H$.) Then

$$aH = H \text{ if and only if } aH = H = eH \text{ if and only if } a = e^{-1}a \in H,$$

i.e., in our other notation

$$\overline{a} = \overline{e} \text{ if and only if } e \in \overline{a} \text{ if and only if } a \in \overline{e}.$$

This means our new notation is sensible, for if $H \subset G$ is a subgroup with $a \in G$, then

$$\begin{aligned} aH &= \{b \in G \mid b \equiv a \mod H\} = \{b \in G \mid a^{-1}b \in H\} \\ &= \{b \in G \mid a^{-1}b = h \text{ some } h \in H\} \\ &= \{b \in G \mid b = ah \text{ some } h \in H\} = \{ah \mid h \in H\}. \end{aligned}$$

(Cf. $a + m\mathbb{Z}$ for the equivalence $\equiv \mod m$.)

We now use the Mantra of Equivalence Relations 5.13. Let H be a subgroup of G and $\mathcal{H}$ be a system of representatives for $\equiv \mod H$. We call $\mathcal{H}$ a *left transversal* of H in G. Then we have

Mantra 10.3. for Cosets. In the above setup, we have

$$G = \bigvee_{\mathcal{H}} aH.$$

In particular, if $|G| < \infty$, then

$$|G| = \sum_{\mathcal{H}} |aH|.$$

As mentioned before, this is only useful if we can compute the size of the equivalence class aH. We call $|\mathcal{H}|$ the *index* of H in G and denote it by $[G : H]$. The Mantra allows us to establish our first important counting result:

Theorem 10.4. (Lagrange's Theorem) *Let G be a finite group and H a subgroup of G. Then*

$$|G| = [G : H]|H|. \text{ In particular, } |H| \mid |G| \text{ and } [G : H] \mid |G|.$$

Proof. We begin with the following:

Claim 10.5. *If G is an arbitrary group (i.e., without assuming it is finite) and H a subgroup, then $|aH| = |H|$ for all $a \in G$.*

Define the *left translation map*

$$\lambda_a : H \to aH \text{ by } h \mapsto ah.$$

By our computation above, we know that λ_a is onto. But it is one-to-one also, as $ah = ah'$ in G with $h, h' \in G$ implies that $h = h'$ by cancellation. Thus λ_a is a bijection, establishing the claim.

If $\mathcal{H}$ is a left transversal for H in G, the mantra together with the claim imply that

$$|G| = \sum_{\mathcal{H}} |aH| = \sum_{\mathcal{H}} |H| = |\mathcal{H}||H|$$

as needed. □

Notation 10.6. If G is a group and $H = \{e_G\}$, we shall write 1 for H unless G is additive in which case we shall write 0 for H. The subgroup $\{e_G\}$ is called the *trivial group* in G.

With this notation, $|G| = [G : 1]$, so if G is finite, Lagrange's Theorem can be written

$$[G : 1] = [G : H][H : 1].$$

We, of course, have an analogous Lagrange's Theorem for *right cosets*. (Definition?) It follows by both Lagrange's Theorems (i.e., for left and right cosets) that if G is a finite group and H a subgroup that $|G|/|H|$ is equal to the (left) index of H in G which is equal to the right index of H in G, i.e., the number of right cosets of H in G is the same as the number of left cosets of H in G. This means that $[G : H]$ makes sense without prescribing right or left cosets. It does not mean, however, that if $a \in G$, then $aH = Ha$. This is usually false. We shall find an important condition on H for this to be true for all a in G.

Warning 10.7. In general, the converse to Lagrange's Theorem is false, i.e., if G is a finite group with $m \in \mathbb{Z}^+$ satisfying $m \mid |G|$, then there may not be a subgroup H in G such that $|H| = m$, if $m \neq 1$ or $|G|$. Of course, if G is a finite cyclic group then the converse does hold by the Cyclic Subgroup Theorem 9.11.

Lagrange's Theorem has many immediate consequences.

Corollary 10.8. *Let G be a finite group and $a \in G$. Then the order of a divides $|G|$.*

Corollary 10.9. *Let G be a group and K and H two finite subgroups of G of relatively prime degree, then $K \cap H = \{e_G\}$.*

PROOF. We leave this as an exercise. □

Corollary 10.10. *Let G be a finite group of prime order p. Then $G \cong \mathbb{Z}/p\mathbb{Z}$. In particular, 1 and G are the only subgroups of G.*

Notation 10.11. If A is a subset of B but $A \neq B$, we shall write $A < B$.

The corollary says if G is a finite group of prime order, then G contains no *proper* subgroups , i.e., G contains no subgroup H satisfying $1 < H < G$.

Corollary 10.12. *Let G be a finite group and $a \in G$. Then $a^{|G|} = e$.*

PROOF. If n is the order of $\langle a \rangle$, then $|G| = nm$ for some $m \in \mathbb{Z}^+$ so $e = (a^n)^m = a^{|G|}$. □

Corollary 10.13. (Euler's Theorem) *Let m, n be relatively prime integers with $m > 1$. Then $n^{\varphi(m)} \equiv 1 \mod m$, where φ is the Euler φ-function.*

PROOF. We have shown that the unit group $(\mathbb{Z}/mZ)^\times$ is given by

$$(\mathbb{Z}/mZ)^\times = \{\overline{a} \mid a \in \mathbb{Z}, (a, m) = 1\}$$

and by definition, its cardinality is $\varphi(m)$. □

Corollary 10.14. (Fermat's Little Theorem) *Let p be a positive prime integer. Then $n^p \equiv n \mod p$ for all integers n. If $p \nmid n$, then $n^{p-1} \equiv 1 \mod p$.*

PROOF. As $\varphi(p) = p - 1$, this follows from Euler's Theorem together with the observation that $0^p = 0$. □

We give a simple application of Euler's Theorem that shows the usefulness of elementary number theory in coding theory. All messages to be encoded are assumed to have been translated into (sequences of) integers. This application is called the *RSA code* that allows encoding and decoding using a public key, i.e., an integer m publicly known, together with a secret key, a number not publicly known. It and its variants are a major way of sending important data, e.g., credit card numbers over the internet. To be effective the secret key is based on

the fact that the public key cannot be factored (in a reasonable amount of time). To set up the code one chooses the public key m to be an integer that is the product of two large distinct primes p and q. We know that

$$\begin{aligned}\varphi(m) &= \varphi(p)\varphi(q) = (p-1)(q-1)\\ &= pq - (p+q) + 1 = m - (p+q) + 1.\end{aligned}$$

So to know $\varphi(m)$, we must know p and q, hence must be able to factor m. A public encoder e is chosen and fixed. This is an integer satisfying $\big(e, \varphi(m)\big) = 1$. Then d is chosen (using the knowledge of $m = pq$, hence $\varphi(m)$) to be an integer satisfying $de \equiv 1 \mod \varphi(m)$, i.e., the inverse of e modulo $\varphi(m)$. Write $de = 1 + k\varphi(m)$ with $k \in \mathbb{Z}$. d is the (non-public) decoder. Messages are written in numbers $x < m$. So RSA code works as follows:

$$\begin{aligned}&\textbf{Encode}: \quad x \mapsto x^e \mod m \quad [e \text{ is publically known}]\\ &\textbf{Decode}: \quad y \mapsto y^d \mod m \quad [d \text{ is secret}].\end{aligned}$$

Claim 10.15. $x^{ed} \equiv x \mod m$, i.e., for all positive integers $x < m$, we get x back.

If $(x, m) = 1$, then by Euler's Theorem, we have

$$x^{de} = x \cdot x^{k\varphi(m)} \equiv x \mod m$$

as needed, so we may assume that $(x, m) \neq 1$. As x satisfies $1 < x < m$, either $p \mid x$ or $q \mid x$ but not both. We may assume that $p \mid x$. Therefore, $x = pn$ for some integer n and $(x, q) = 1$. Since $p \mid x$,

$$x^{ed} \equiv 0 \equiv x \mod p.$$

By Fermat's Theorem,

$$x^{\varphi(m)} = (x^{q-1})^{p-1} \equiv 1^{p-1} \equiv 1 \mod q.$$

Since p nd q are relatively prime and both p and q divide $x^{ed} - x$, we have $m = pq \mid x^{ed} - x$, so $x^{ed} \equiv x \mod m$. This establishes the claim.

Some care must be taken in the choice of the large primes p and q. For example, if all primes less than N are known and either $p - 1$ or $q - 1$ factors into products of such primes, than m can be factored using what is called the Pollard $(p-1)$- Factoring Algorithm. Note that the idea of the RSA code is also based on group theory. Other public codes have been developed using group theory.

Exercises 10.16. In the following exercises, let G be a group with H and K subgroups.

1. Show that $\equiv \mod H$ is an equivalence relation.
2. Let G be a group of order p^n where p is a prime and $n \geq 1$. Prove that there exists an element of order p in G.
3. Let $HK := \{hk \mid h \in H, k \in K\}$. Then (clearly) $H/(H \cap K)$ is a subset of $G/(H \cap K)$ and HK/K is a subset of G/K. Show that $f : H/(H \cap K) \to HK/K$ by $h(H \cap K) \mapsto hK$ is a well-defined bijection.
4. (Poincaré) Suppose both H and K have finite index in G. Show that $H \cap K$ has finite index in G.
5. If $K \subset H \subset G$, show that $[G : K] = [G : H][H : K]$ (even if any are infinite if read correctly).
6. If K and H are finite subgroups of G of relatively prime degree, then $K \cap H = \{e_G\}$.
7. Let $a \in (\mathbb{Z}/n\mathbb{Z})^\times$ have order m and $b \in (\mathbb{Z}/n\mathbb{Z})^\times$ have order n.
   (*i*) Show that the order of a^r is $r/(m, r)$.
   (*ii*) Show if $(m, n) = 1$, then the order of ab is mn.
8. Let p be an odd prime. Prove that $(\mathbb{Z}/p\mathbb{Z})^\times$ is a cyclic group of order $p - 1$. [You may use the fact, that the polynomial $t^n - 1$ has at most n roots in $(\mathbb{Z}/p\mathbb{Z})[t]$ for any $n > 0$.]
9. Let m and n be positive integers that are not relatively prime. Show that $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ is abelian but not cyclic.
10. Prove that a group of order 30 can have at most 7 subgroups of order 5.
11. (Wilson's Theorem) Let p be a positive integer. Then p is a prime if and only if $(p-1)! \equiv -1 \mod p$. (This is not a practical criterion.) [Hint: If $1 \leq j \leq p - 1$, when can $j^2 \equiv 1 \mod p$?]

11. Homomorphisms

We begin with a useful fact that we leave as an exercise:

Observation 11.1. If $\varphi : G \to G'$ is a group homomorphism and H is a subgroup of G, then $\varphi(H) \subset G'$ is a subgroup.

If G is a group and $x \in G$, then the map

$$\theta_x : G \to G \text{ given by } g \mapsto xgx^{-1}$$

is called *conjugation* by x. If $H \subset G$ is a subgroup, then we let

$$xHx^{-1} := \theta_x|_H(H) = \theta_x(H) = \{xhx^{-1} \mid h \in H\}.$$

Lemma 11.2. *Let G be a group and $x \in G$. Then $\theta_x : G \to G$ is an isomorphism. In particular, if $H \subset G$ is a subgroup, so is $\theta_x(H)$ and $H \cong \theta_x(H) = xHx^{-1}$. In particular, $|H| = |xHx^{-1}|$.*

PROOF. Let $g, g' \in G$. As $xgx^{-1} = xg'x^{-1}$ implies $g = g'$, the map θ_x is one-to-one. The equation

$$\theta_x(gg') = xgg'x^{-1} = xgeg'x^{-1} = xgx^{-1}xg'x^{-1} = \theta_x(g)\theta_x(g') \tag{11.3}$$

shows that θ_x is a homomorphism.

This idea of writing an identity in a clever way is a most useful device, that we shall call *Great Trick*. You have undoubtedly used it quite often for the elements 1 and 0 in $\mathbb{R}$.

By the Observation 11.1, we know that $\theta_x(H) \subset G$ is a subgroup, so we can view $\theta_x : H \to \theta_x(H)$. As it is a bijective homomorphism, it is an isomorphism. □

A group isomorphism $\varphi : G \to G$ is called a *group automorphism*. So θ_x above is a group automorphism for all $x \in G$. If G is an abelian group, then θ_x is the identity on G for all $x \in G$. If G is not abelian, then there must exist elements x and y in G satisfying $xy \neq yx$, so $\theta_x \neq 1_G$. In general, if H is a subgroup, $xHx^{-1} \neq H$, i.e., the isomorphism $\theta_x|_H : H \to xHx^{-1}$ (replacing the target with $\theta_x(H)$) is not an automorphism. Can you come up with a counterexample? Of course, if $x \in H$ then $xHx^{-1} = \theta_x(H) = H$.

We say that a subgroup H of a group G is *normal* and write $H \triangleleft G$ if $xHx^{-1} = \theta_x(H) = H$ for all $x \in G$.

Remarks 11.4. Let H be a subgroup of G.

1. $H \triangleleft G$ does not mean that $\theta_x|_H = 1_H : H \to H$ for all $x \in G$ (or even all $x \in H$). In fact, it usually is not. For example, (see below)
$$H := \langle r \rangle \triangleleft D_3 \text{ but } frf^{-1} = r^{-1}, \text{ so } \theta_f|_H \neq 1_H.$$
2. Let
$$\mathrm{Aut}(G) := \{\sigma : G \to G \mid \sigma \text{ is an automorphism}\},$$
then $\mathrm{Aut}(G)$ is a group under composition. (Check.) A conjugation $\theta_x : G \to G$, with $x \in G$, is also called an *inner automorphism* of G. Set
$$\mathrm{Inn}(G) := \{\theta_x \mid x \in G\}.$$
Check that this is a subgroup of $\mathrm{Aut}(G)$, called the *inner automorphism group* of G. Moreover, the following are equivalent:
(a) $H \triangleleft G$.

(b) $\theta(H) = H$ for all $\theta \in \mathrm{Inn}(G)$.
(c) $\theta|_H \in \mathrm{Aut}(H)$ for all $\theta \in \mathrm{Inn}(G)$.
(d) The restriction map $|_H : \mathrm{Inn}(G) \to \mathrm{Aut}(H)$ is defined (i.e., the image lies in $\mathrm{Aut}(H)$).

3. $H \lhd G$ if and only if $xH = Hx$ for all $x \in G$. (Why?)
4. If $xHx^{-1} \subset H$ for all $x \in G$, then $y^{-1}Hy \subset H$ for all $y \in G$ (let $y = x^{-1}$), so $H \subset yHy^{-1}$ for all $y \in G$. Hence

$$H \lhd G \text{ if and only if } xHx^{-1} \subset H \text{ for all } x \in G.$$

Examples 11.5. Let G be a group.

1. $1 \lhd G$ and $G \lhd G$. If G is a non-trivial group and these are the only normal subgroups of G, then we say G is a *simple group*. We have seen that $\mathbb{Z}/p\mathbb{Z}$ is a simple group for all primes p. One of the great theorems in mathematics is the classification of all finite simple groups. The proof is thousands of pages long.
2. If G is abelian, every subgroup is normal. The converse is false, i.e, there exist nonabelian groups in which every subgroup is normal.
3. Let
$$Z(G) := \{x \in G \mid xg = gx \text{ for all } g \in G\}.$$
This set is called the *center of* G. It is a normal subgroup of G. In fact, any subgroup of $Z(G)$ is a normal subgroup of G. Moreover, G is abelian if and only if $G = Z(G)$.
4. If F is a field (even a commutative ring), then $\mathrm{SL}_n(F) \lhd \mathrm{GL}_n(F)$
5. Let $\theta : G \to G'$ be a group homomorphism. Then $\ker\theta \lhd G$.
[In general, $\mathrm{im}\,\theta$ is not a normal subgroup of G. (Can you give an example?)]
6. If G is an abelian group, then $\varphi : G \to G$ given by $x \mapsto x^{-1}$ is an automorphism. It is the identity map if and only if $x^2 = e$ for all $x \in G$, otherwise it is of order two. In particular, if G is a finite abelian group of order at least three, then $2 \mid |\mathrm{Aut}(G)|$. Can you prove this?
7. $\mathrm{Inn}(G) \lhd \mathrm{Aut}(G)$.
8. If H is a subgroup of G of index two then $H \lhd G$. Indeed, if $a \in G \setminus H$ then $G = H \vee aH = H \vee Ha$, so we must have $aH = Ha$, hence $aHa^{-1} = H$.

For the next example, we shall need matrix representations of linear transformations of finite dimensional vector spaces. In Appendix §103, we review this notation. (This appendix does a more general case but is applicable here.) In particular, if $T : V \to W$ is a linear transformation of vector spaces over a field F with bases $\mathcal{B}$ and $\mathcal{C}$

respectively, then we let $[T]_{\mathcal{B},\mathcal{C}}$ denote the matrix representation of this linear transformation relative to these bases. If $V = V'$ and $\mathcal{B} = \mathcal{C}$, we write this as $[T]_{\mathcal{B}}$.

9. If $\sigma \in S_n$, we can write σ by

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}$$

where the top row is the elements of the domain and the bottom row the corresponding values. Let $\mathcal{S}_n := \{e_1, \dots, e_n\}$ be the *standard basis* for $V = \mathbb{R}^n$. For each σ, define the map $T_\sigma : V \to V$ by $\Sigma_i \alpha_i e_i \mapsto \Sigma_i \alpha_i e_{\sigma(i)}$, a linear transformation. T_σ is a vector space isomorphism with inverse $T_{\sigma^{-1}}$. Define θ by

$$\theta : S_n \to \mathrm{GL}_n(\mathbb{R}) \text{ is given by } \sigma \mapsto [T_\sigma]_{\mathcal{S}_n}. \tag{11.6}$$

It is easy to see that θ is a group monomorphism. Each $[T_\sigma]_{\mathcal{S}_n}$ is a *permutation matrix*, i.e., a matrix with the element 1 in precisely one row and one column and all other entries 0. So it is just a permutation of the rows (or columns) of the identity matrix. Let $\mathrm{Perm}_n(\mathbb{R})$ be the set of permutation matrices. It is the image of θ, so a group. Changing the target, we view $\theta : S_n \to \mathrm{Perm}_n(\mathbb{R})$, an isomorphism. If $A \in \mathrm{Perm}_n(\mathbb{R})$, then $\det A = \pm 1$, so we have a composition of maps

$$S_n \xrightarrow{\theta} \mathrm{Perm}_n(\mathbb{R}) \xrightarrow{\det} \{\pm 1\}.$$

It is a homomorphism as both θ and det are. If $n > 1$ then the kernel of this map is called the *alternating group* on n *letters* and denoted by A_n. Elements of A_n are called *even permutations* and elements of $S_n \setminus A_n$ are called *odd permutations*. We have $A_n \lhd S_n$, and, in fact, $[S_n : A_n] = 2$ if $n \geq 2$. We shall study this group in Section §22.

Remarks 11.7. 1. We have seen that a group of prime order is a simple group, i.e., a group without any proper normal subgroups. These turn out to be the only abelian simple groups. The group A_5 is the non-abelian simple group of smallest order. This was shown by Abel. In fact,

Theorem 11.8. (Abel's Theorem) *For every $n \geq 5$, the group A_n is simple.*

The group A_2 is the trivial group and A_3 is the cyclic group of order three so simple. The group A_4 is not simple.

Abel's Theorem is important as it is the key to proving the Abel-Ruffini Theorem that states that there is no formula for the roots of a general fifth degree polynomial with rational coefficients involving only the extraction of nth roots and $+$ and $\cdot$. We shall prove Abel's Theorem in Theorem 22.11 below.

2. If K and H are subgroups of G satisfying $K \subset H \subset G$ with $K \lhd H$ and $H \lhd G$, it is not necessarily true that $K \lhd G$, i.e., being normal is not transitive.
3. If K and H are subgroups of G satisfying $K \subset H \subset G$ with $K \lhd G$, then it is true that $K \lhd H$.
4. There is a useful stronger property that a subgroup of a group can satisfy than being normal. We say that a subgroup H of a group G is a *characteristic subgroup* of G if for every $\sigma \in \mathrm{Aut}(G)$, we have $\sigma|_H$ lies in $Aut(H)$. Equivalently, the restriction map res : $\mathrm{Aut}(G) \to \mathrm{Aut}(H)$ given by $\theta \mapsto \theta|_H$ is well-defined, i.e., the target is correct. If H is a characteristic subgroup of G, we write $H \lhd\lhd G$. Clearly, unlike being normal, being characteristic is transitive, i.e., if $K \lhd\lhd H \lhd\lhd G$, then $K \lhd\lhd G$.

Exercises 11.9. 1. Prove Observation 11.1.

2. Find all subgroups of S_3 and determine which ones are normal.
3. Show that a subgroup $H \subset G$ is normal if and only if $gH = Hg$ for all $g \in G$. If H is not normal is it still true that for each $g \in G$ there is an $a \in G$ such that $gH = Ha$?
4. Determine all subgroups of the quaternion group of order 8 and determine which ones are normal.
5. Let G be a group. Show
   (*i*) The center $Z(G)$ of a group G is a subgroup.
   (*ii*) Any subgroup of the center of G is a normal subgroup of G.
   (*iii*) G is abelian if and only if it equals its center.
6. Determine the center of the dihedral group D_n and prove your assertion.
7. Let H be a subgroup of G. In Exercises 8.5 (7) and (8), we defined the core of H in G to be $\mathrm{Cor}_G(H) := \bigcap xHx^{-1}$ and the *normalizer* of H in G to be $N_G(H) : \{x \in G \mid xHx^{-1} = H\}$, respectively. Show
   (*i*) $\mathrm{Cor}_G(H)$ is the unique largest normal subgroup of G contained in H.
   (*ii*) $N_G(H)$ is the unique largest subgroup of G containing H as a normal subgroup.
8. Show $\mathrm{Inn}(G) \lhd \mathrm{Aut}(G)$.

9. Prove the assertions in Example 11.5(6).
10. Let G be a cyclic group. Determine $\mathrm{Aut}(G)$ and $\mathrm{Inn}(G)$ up to isomorphism as groups that we know. Prove your result.
[Hint. Where do generators go?]
11. Let G and H be finite cyclic groups of order m and n, respectively. Show the following:
   (*i*) If m and n are relatively prime, then $\mathrm{Aut}(G \times H) \cong \mathrm{Aut}(G) \times \mathrm{Aut}\, H$ and is abelian.
   (*ii*) If m and n are not relatively prime, then $\mathrm{Aut}(G \times H)$ is never abelian. (Cf. Exercise 11.9(12).)
12. Compute $\mathrm{Aut}\big((\mathbf{Z}/p\mathbf{Z})^n\big)$ up to isomorphism and its order when p is a prime.
13. Let $\varphi : G \to \mathrm{Aut}(H)$, write $x \mapsto \varphi_x$, be a group homomorphism where G and H are groups. Define $H \rtimes_\varphi G$ to be the cartesian product $H \times G$ under the binary operation given by
$$(h, g) \cdot (h', g') = (h\varphi_g(h'), gg')$$
for all $g, g' \in G$ and $h, h' \in H$. Show that $H \rtimes_\varphi G$ is a group, called the *semidirect product* of H and G, with $H \times 1$ a normal subgroup. Note that the product $G \times H$ is the case that φ is the trivial homomorphism.
14. Let $\sigma \in \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ be the automorphism defined by $x \mapsto -x$ and $\varphi : \mathbb{Z}/2\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ be the group homomorphism defined by $\overline{1} \mapsto \sigma$. Show that $\mathbb{Z}/n\mathbb{Z} \rtimes_\varphi \mathbb{Z}/2\mathbb{Z} \cong D_n$.
15. Find all subgroups of A_4 and determine which ones are normal. In particular, show that A_4 is not simple.
16. Let G be a group in which every subgroup is normal. Is G an abelian group? Prove or provide a counterexample.
17. A *commutator* in G is an element of the form $xyx^{-1}y^{-1}$ where $x, y \in G$. Let G' be the subgroup of G generated by all commutators, i.e., every element of G' is the product of commutators. We call G' the *commutator* or *derived subgroup* of G. It is also denoted $[G, -G]$. Show
   (*i*) $G' \triangleleft G$.
   (*ii*) $G' \triangleleft\triangleleft G$.
18. Let $K \subset H \subset G$ be subgroups of G. Show all of the following:
   (*i*) If $K \triangleleft\triangleleft H$ and $H \triangleleft G$ then $K \triangleleft G$.
   (*ii*) $Z(G) \triangleleft\triangleleft G$.
   (*iii*) $G' \triangleleft\triangleleft G$.

(*iv*) Inductively define $G^{(n)}$ as follows: $G^{(1)} = G'$. Having defined $G^{(n)}$ define $G^{(n+1)} := (G^{(n)})'$. Then $G^{(n+1)} \lhd\lhd G$.

12. The First Isomorphism Theorem

Given an equivalence relation, we saw that we could always form the set of equivalence classes. If G is a group and H is a subgroup, we can form the set of equivalence classes G/H. Unfortunately, in general this does not have a group structure. We also know that equivalence relations arise from surjective maps. We begin by looking at the case when $\varphi : G \to G'$ is a group epimorphism. This defines an equivalence relation $\sim$ on G by $a \sim a'$ if $\varphi(a) = \varphi(a')$ and with our usual notation, we have

$$\overline{a} = \overline{a'} \text{ if and only if } a \sim a' \text{ if and only if } \varphi(a) = \varphi(a')$$

with

$$\overline{a} = \varphi^{-1}\big(\varphi(a)\big) = \{x \in A \mid \varphi(x) = \varphi(a)\}.$$

But now, we have the additional information that φ is a homomorphism, so $\varphi(aa') = \varphi(a)\varphi(a')$. This means that

$\varphi(a) = \varphi(a')$ if and only if

$$e_{G'} = \varphi(a)^{-1}\varphi(a') = \varphi(a^{-1}a') \text{ and } e_{G'} = \varphi(a')\varphi(a)^{-1} = \varphi(a'a^{-1})$$

if and only if both $a^{-1}a'$ and $a'a^{-1}$ lie in $\ker\varphi$.

Let $K = \ker\varphi$. So we have

$$\overline{a} = \overline{a'} \text{ if and only if } \varphi(a) = \varphi(a') \text{ if and only if } a^{-1}a', a'a^{-1} \in K.$$

We know that if $a^{-1}a' \in K$, then $\overline{a'} = aK$, so

$$\overline{a} = \varphi^{-1}\big(\varphi(a)\big) = \{a' \mid a' = ak \text{ some } k \in K\} = aK,$$

the left coset of a in G relative to K. Hence $\overline{G} = G/K$. Similarly,

$$\overline{a} = \varphi^{-1}\big(\varphi(a)\big) = \{a' \mid a' = ka \text{ some } k \in K\} = Ka,$$

the right coset of a in G relative to K. So we now have

$$\begin{aligned}\overline{a} = \overline{a'} &\text{ if and only if } \varphi(a) = \varphi(a') \\ &\text{if and only if } aK = a'K = \overline{a} = Ka = Ka'.\end{aligned}$$

In particular, we must have $aK = Ka$, equivalently, $K = aKa^{-1}$ for all $a \in G$ as $a \sim a$, i.e., $K \lhd G$. Of course, we knew this before, but it now indicates where cosets are coming from, at least in a special case, and ties up various mathematical threads that we are developing. Note that $\overline{e_G} = \varphi^{-1}\big(\varphi(e_G)\big) = \varphi^{-1}(e_{G'}) = \ker\varphi$.

We now apply the First Isomorphism of Sets 7.1 to obtain a commutative diagram:

(†)
$$\begin{array}{ccc} G & \xrightarrow{\varphi} & G' \\ {\scriptstyle -}\downarrow & \nearrow_{\overline{\varphi}} & \\ \overline{G} & & \end{array}$$

where $\overline{\varphi} : \overline{G} \to G'$ is given by $\overline{a} \mapsto \varphi(a)$, a well defined bijection. We also know that

$$\overline{a} = \overline{a'} \text{ if and only if } a \sim a' \text{ if and only if } \overline{\varphi}(\overline{a}) = \overline{\varphi}(\overline{a'}),$$

so the fact that φ is a homomorphism implies that

(*)
$$\overline{\varphi}(\overline{aa'}) = \varphi(aa') = \varphi(a)\varphi(a') = \overline{\varphi}(\overline{a})\overline{\varphi}(\overline{a'}).$$

We can use this and the fact that $\overline{\varphi}$ is a bijection to put a group structure on $\overline{G}$. Define

$$\cdot : \overline{G} \times \overline{G} \to \overline{G} \text{ by } \overline{a} \cdot \overline{a'} := \overline{aa'}.$$

It is well-defined for if $\overline{a} = \overline{x}$ and $\overline{x'} = \overline{a'}$, then

$$\overline{\varphi}(\overline{xx'}) = \overline{\varphi}(\overline{x}) \cdot \overline{\varphi}(\overline{x'}) = \overline{\varphi}(\overline{a}) \cdot \overline{\varphi}(\overline{a'}) = \overline{\varphi}(\overline{aa'})$$

by (*), so $\overline{xx'} = \overline{aa'}$ as $\overline{\varphi}$ is a bijection. For all $a \in G$, we have $\overline{e_G} \cdot \overline{a} = \overline{a} = \overline{a} \cdot \overline{e_G}$ and $\overline{a} \cdot \overline{a^{-1}} = \overline{e_G} = \overline{a^{-1}} \cdot \overline{a}$, so $\overline{G}$ is a group with identity $\ker \varphi = \overline{e_G}$ and inverse of $\overline{a}$ given by $\overline{a}^{-1} = \overline{a^{-1}}$. Furthermore, (*) implies that

$$\overline{\varphi}(\overline{a} \cdot \overline{a'}) = \overline{\varphi}(\overline{aa'}) = \overline{\varphi}(\overline{a}) \cdot \overline{\varphi}(\overline{a'}),$$

so $\overline{\varphi}$ is a bijective group homomorphism, hence an isomorphism. Our diagram (†) becomes

$$\begin{array}{ccc} G & \xrightarrow{\varphi} & G' \\ {\scriptstyle -}\downarrow & \nearrow_{\overline{\varphi}} & \\ G/\ker\varphi & & \end{array}$$

with $G/\ker\varphi$ a group, $^-$ a group epimorphism and $\overline{\varphi}$ an isomorphism.

We can generalize this to any group homomorphism $\psi : G \to G''$, by letting $G' = \operatorname{im}\psi$ and φ the map $\varphi : G \to G'$ given by $\varphi(a) = \psi(a)$. As the inclusion map $\text{inc} : G' \to G''$ is a group homomorphism, we are in a similar situation as in the Diagram 7.3, but with all maps homomorphisms. We write this, after changing notation as:

Theorem 12.1. (First Isomorphism Theorem) *Let $\varphi : G \to G'$ be a group homomorphism. Then we have a commutative diagram of groups and group homomorphisms*

$$\begin{array}{ccc} G & \xrightarrow{\varphi} & G' \\ {\scriptstyle -}\downarrow & & \uparrow{\scriptstyle \mathrm{inc}} \\ G/\ker\varphi & \xrightarrow[\overline{\varphi}]{} & \operatorname{im}\varphi \end{array}$$

with $^-$ a group epimorphism, $\overline{\varphi}$ a group isomorphism between $G/\ker\varphi$ and $\operatorname{im}\varphi$ and inc *a group monomorphism.*

Example 12.2. Let G be a group. The map $\theta : G \to \mathrm{Aut}(G)$ given by $x \mapsto (\theta_x : g \mapsto xgx^{-1})$ is a group homomorphism. If $\theta_x = 1_G$, the identity map on G, then $xgx^{-1} = g$, i.e., $xg = gx$ for all $g \in g$. Therefore, $\ker\theta = Z(G)$, the center of G. By the First Isomorphism Theorem, θ induces an isomorphism $\overline{\theta} : G/Z(G) \to \mathrm{Inn}(G)$.

The only missing piece is: When is G/H a group for a subgroup H of G such that the canonical surjection $^- : G \to G/H$ is a group epimorphism? If this is the case, then the above analysis says that $H = \ker{}^-$. This is answered by the following:

Theorem 12.3. *Let G be a group and H a normal subgroup of G. Then G/H is a group under the binary operation $\cdot : G/H \times G/H \to G/H$ given by $(aH, bH) \mapsto abH$. If this is the case then $^- : G \to G/H$ is a group epimorphism with kernel H.*

We call G/H the *quotient* or *factor group* of G by (the normal subgroup) H.

Proof. We first show that the map $\cdot$ is well-defined. Suppose that $aH = a'H$ and $bH = b'H$. We must show that $abH = a'b'H$. Equivalently, we must show that

$$\text{if } a'^{-1}a, b'^{-1}b \in H \text{ then } x = (a'b')^{-1}ab \in H.$$

But, using Great Trick and the definition of normality, we have

$$x = b'^{-1}a'^{-1}ab = b'^{-1}a'^{-1}a(b'b'^{-1})b = (b'^{-1}a'^{-1}ab')(b'^{-1}b)$$

lies in H as needed.

Let $a, b, c \in G$. Associativity follow by $aH \cdot (bH \cdot cH) = aH \cdot bcH = abcH = abH \cdot cH = (aH \cdot bH) \cdot cH$. The unity is easily seen to be $e_{G/H} = H = e_G H$ and $(aH)^{-1} = a^{-1}H$. Thus G/H is a group. □

The theorem shows that normal subgroups are equivalent to the kernels of group homomorphisms. Of course, it does not explain how

cosets arise in a natural way, if the subgroup is not a normal subgroup. We will come back to this later.

We are now in a position where we can prove some important results. Especially significant is the following:

Theorem 12.4. (General Cayley Theorem) *Let H be an arbitrary subgroup of a group G. Let $S = G/H$, the set of left cosets of H in G. For each $x \in G$, let $\lambda_x : S \to S$ be defined by $gH \mapsto xgH$. Then λ_x is a permutation and the map $\lambda : G \to \Sigma(S)$ given by $x \mapsto \lambda_x$ is a group homomorphism satisfying the following two properties:*

(1) $\ker \lambda \subset H$.

(2) $\ker \lambda$ *is the unique largest normal subgroup of G contained in H, i.e.,*

$$\textit{If } N \triangleleft G \textit{ and } N \subset H \textit{ then } N \subset \ker \lambda.$$

PROOF. We first must show that the map $\lambda_x : S \to S$, called *left multiplication* by x, is a permutation. But this is clear since $\lambda_{x^{-1}}$ is easily checked to be its inverse.

For all $x, y \in G$, we have $\lambda_{xy}(gH) = xygH = \lambda_x(ygH) = \lambda_x\lambda_y(gH)$ for all $g \in G$, so λ is a group homomorphism and $\ker \lambda \triangleleft G$. If $x \in \ker \lambda$, then $xgH = \lambda_x(gH) = gH$ for all $g \in G$. In particular, $xH = H$ so $x \in H$. Finally, suppose that $N \triangleleft G$ with $N \subset H$. To show $N \subset \ker \lambda$, we must show if $x \in N \subset H$, then λ_x is 1_S, the identity map on S. So let $x \in N$ and $g \in G$. As $g^{-1}xg \in N$, we have $xg \in gN \subset gH$. Therefore, $xgH = gH$, i.e., $\lambda_x(gH) = gH$ for all $g \in G$. Consequently, $\lambda_x = 1_S$ as required. □

In the above, the special case that $H = 1$ yields:

Corollary 12.5. (Cayley's Theorem) *Let G be a group, then the map*

$$\lambda : G \to \Sigma(G) \textit{ given by } x \mapsto (\lambda_x : g \mapsto xg)$$

is a group monomorphism. In particular, if G is a finite group of order n then there exists a monomorphism $G \to S_n$.

If S is a non-empty set, then a subgroup of $\Sigma(S)$ is called a *permuation group*. Cayley's Theorem shows that every abstract group is isomorphic to a permutation group. The above monomorphism is called the (*left*) *regular representation* of G. It shows that every abstract group can be realized by a 'concrete' group, e.g., a permutation group.

Corollary 12.6. *Let H be a subgroup of a group G with $H < G$. If there exists no normal subgroup N of G satisfying $1 < N \subset H$ then $\lambda : G \to \Sigma(G/H)$ by $x \mapsto (\lambda_x : aH \mapsto xaH)$ is a monomorphism.*

PROOF. $\ker \lambda$ is the maximal such normal subgroup. □

Corollary 12.7. (Useful Counting Result) *Let G be a finite group, H a subgroup of G satisfying $|G| \not| [G : H]!$. Then there exists a normal subgroup N of G satisfying $1 < N \subset H$. In particular, G is not a simple group.*

PROOF. Exercise. □

We give two applications of the Useful Counting Result, but to do so we must assume for now the following to be true that we shall prove later.

Theorem 12.8. (First Sylow Theorem) *Let G be a finite group, p a prime such that $p^s \,||\, |G|$ (i.e., $p^s \mid |G|$ but $p^{s+1} \not| |G|$). Then G contains a subgroup of order p^s.*

Examples 12.9. Let G be a group.

1. If $|G| = 24$ then G is not simple, as G contains a group of order 8 by the First Sylow Theorem and $|G| \not| 3!$.
2. Let $p < q$ be positive primes such that $G = pq$. Then G contains a normal group of order q, so is not simple.

Another useful corollary of the General Cayley Theorem is the following whose proof we also leave as an exercise:

Corollary 12.10. *Let G be a finite group. Suppose that p is the smallest prime dividing the order of G. Suppose also that there exists a subgroup H of G of index p. Then $H \lhd G$.*

Exercises 12.11. 1. Let $T : V \to W$ be a surjective linear transformation of vector spaces over F. Do an analogous analysis before The First Isomorphism Theorem 12.1 for $T : V \to W$ and state and prove the analogues of The First Isomorphism Theorem 12.1 and Theorem 12.3 for it.

2. Let G be a finite group and $n > 1$ an integer such that $(xy)^n = x^n y^n$ for all $x, y \in G$. Let
$$G_n := \{z \in G \,|\, z^n = e\} \quad \text{and} \quad G^n := \{x^n \,|\, x \in G\}.$$
Show both G_n and G^n are normal subgroups of G and satisfy $|\,G^n| = [G : G_n]$.
3. Let G be a finite group of order n, F a field (or even any ring with $1 \neq 0$). Show there exists a monomorphism $\varphi : G \to \mathrm{GL}_n(F)$, i.e., any finite group can be realized as a subgroup of matrices.
4. Prove the Useful Counting Result 12.7.

5. Prove Corollary 12.10.
6. Suppose that G is a group containing a subgroup of finite index greater than one. Show that G contains a normal subgroup of finite index greater than one.
7. Let H be a subgroup of G. Define the *centralizer* of H in G to be
$$Z_G(H) := \{x \in G \mid xh = hx \text{ for all } h \in H\}.$$
Show that it is a normal subgroup of $N_G(H)$ and the map given by $x \mapsto (\theta_x : g \mapsto xgx^{-1})$ induces $\tilde{\theta} : N_G(H)/Z_G(H) \to \mathrm{Aut}(H)$, a monomorphism defined by $xZ_G(H) \to \theta_x|_H$.

13. The Correspondence Principle

Recall if $f : A \to B$ is a set map and $D \subset B$ is a subset, then the *preimage* of D in A is the set $f^{-1}(D) := \{a \in A \mid f(a) \in D\}$. We shall need the following that is left as an exercise.

Properties 13.1. Let $f : A \to B$ be a set map, $C \subset A$, and $D \subset B$ subsets, then:

1. $C \subset f^{-1}\big(f(C)\big)$ with equality if f is one-to-one.
2. $f\big(f^{-1}(D)\big) \subset D$ with equality if f is onto.

We apply the First Isomorphism Theorem to establish the following important result:

Theorem 13.2. (Correspondence Principle) *Let $\varphi : G \to G'$ be a group epimorphism.. Then*

(1) If A is a subgroup of G (respectively, a normal subgroup), then $\varphi(A)$ is a subgroup of G' (respectively, a normal subgroup). In particular, group epimorphisms preserve normality.
(2) If A is a subgroup of G containing $\ker\varphi$, then $\varphi^{-1}\big(\varphi(A)\big) = A$.
(3) If B is a subgroup of G' (respectively, a normal subgroup), then $\varphi^{-1}(B)$ is a subgroup of G (respectively, a normal subgroup) containing $\ker\varphi$ and satisfies $B = \varphi\big(\varphi^{-1}(B)\big)$.

In particular, if

$$\mathcal{G}_K := \{A \mid A \subset G \text{ a subgroup with } \ker\varphi \subset A\}$$
$$\mathcal{G}' := \{B \mid B \subset G' \text{ a subgroup}\},$$

then

$$\mathcal{G}_{\ker\varphi} \to \mathcal{G}' \text{ given by } A \mapsto \varphi(A)$$

is a bijection of sets preserving inclusions with inverse $B \mapsto \varphi^{-1}(B)$ and restricting to a bijection on normal subgroups.

PROOF. Let $K = \ker \varphi$.

(1). Let $A \subset G$ be a subgroup, then $\varphi|_A : A \to G'$ is a group homomorphism (why?), so $\varphi(A) = \operatorname{im} \varphi|_A \subset G'$ is a subgroup. Next suppose that $A \lhd G$ and $y \in G'$. As φ is surjective, $y = \varphi(x)$ for some $x \in G$. Hence $y\varphi(A)y^{-1} = \varphi(x)\varphi(A)\varphi(x)^{-1} = \varphi(xAx^{-1}) = \varphi(A)$. Therefore, $\varphi(A) \lhd G'$.

(2). By the properties of preimages, $A \subset \varphi^{-1}\big(\varphi(A)\big)$, so it suffices to show if $K \subset A$, then $\varphi^{-1}\big(\varphi(A)\big) \subset A$. Suppose that $x \in \varphi^{-1}\big(\varphi(A)\big)$. By definition, $\varphi(x) = \varphi(a)$ for some $a \in A$, hence $e_{G'} = \varphi(a)^{-1}\varphi(x) = \varphi(a^{-1}x)$. Thus $a^{-1}x \in K$, and so $x \in aK \subset A$ as needed.

(3). Let B be a subgroup of G'. If $x, y \in \varphi^{-1}(B)$, we have $\varphi(xy^{-1}) = \varphi(x)\varphi(y)^{-1} \in B$, so $\varphi^{-1}(B)$ is a subgroup of G. If $k \in K$ then $\varphi(k) = e_{G'} \in B$, so $K \subset \varphi^{-1}(B)$. As φ is onto, $B = \varphi\big(\varphi^{-1}(B)\big)$ by the properties of preimages. Finally, suppose that $B \lhd G'$. Let $x \in A$, then $\varphi\big(x\varphi^{-1}(B)x^{-1}\big) = \varphi(x)\varphi\big(\varphi^{-1}(B)\big)\varphi(x)^{-1} = \varphi(x)B\varphi(x)^{-1} = B$. Hence $x\varphi^{-1}(B)x^{-1} \subset \varphi^{-1}(B)$, so $\varphi^{-1}(B) \lhd G$. □

We leave the following alternate form of the Correspondence Principle as an exercise. (Note we did not include the analogue of the last statement in the above.)

Theorem 13.3. (Correspondence Principle, Alternate Form) *Let G be a group and K a normal subgroup. Let $^- : G \to G/K$ be the canonical epimorphism given by $x \mapsto xK$ and L a subgroup of G/K. Then*

(1) There exists a subgroup H of G containing K with $L = H/K$.
(2) Let H be as in (1), then $H \lhd G$ if and only if $L \lhd G/K$.
(3) Suppose that H_1, H_2 are two subgroups of G containing K. If $H_1/K = H_2/K$ then $H_1 = H_2$.
(4) If G is a finite group and H is as in (1), then $[G : H] = [G/K : H/K] = [G/K : L]$ and $|H| = |K| \cdot |L|$.

Theorem 13.4. (Third Isomorphism Theorem) *Let G be a group with normal subgroups K and H satisfying $K \subset H$. Then the map*

$$\varphi : G/K \to G/H \text{ defined by } xK \to xH$$

is a group epimorphism with kernel H/K and induces an isomorphism

$$\overline{\varphi} : (G/K)/(H/K) \to G/H.$$

PROOF. As K and H are normal subgroups of G, we know that G/H and G/K are groups. If $xK = yK$ then $y^{-1}x \in K \subset H$, hence $xH = yH$. Therefore, φ is well-defined and clearly surjective. As $\varphi(xKyK) = \varphi(xyK) = xyH = xHyH = \varphi(xK)\varphi(yK)$, the map φ is a group homomorphism. If $xK \in \ker \varphi$ then $H = e_{G/H} = \varphi(xK) = xH$,

so $x \in H$, i.e., $xK \in H/K$. Conversely, if $x \in H$ then $\varphi(xK) = xH = H = e_{G/H}$, so $xK \in \ker \varphi$. The result now follows by the First Isomorphism Theorem. □

The next isomorphism theorem is more useful, but harder to visualize.

If H_1 and H_2 are subgroups of a group G, we let

$$H_1H_2 := \{h_1h_2 \mid h_i \in H_i,\ i = 1, 2\}.$$

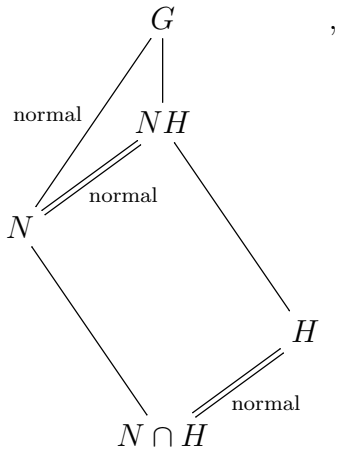
In general, this is not a group.

Theorem 13.5. (Second Isomorphism Theorem) *Let G be a group and H and N be subgroups with N normal in G. Then*

(1) $H \cap N \triangleleft H$.
(2) $HN = NH$ is a subgroup of G.
(3) $N \triangleleft HN$.
(4) $H/H \cap N \cong HN/N$.

In particular, if H_1 and H_2 are finite subgroups of G, then $|H_1H_2| = |H_1||H_2|/|(H_1 \cap H_2|$.

We have the following picture



,

where a group below another group connected by a line is a subgroup and the quotient of the groups defined by the double lines are isomorphic.

Proof. (1). We know that $H \cap N$ is a subgroup of H and it is a normal subgroup in H, since N is normal in G and H is normal in H.
(2). As $aN = Na$ for all $a \in G$ the sets HN and NH are equal, so we need only show it is a group. Let $h_1, h_2 \in H$ and $n_1, n_2 \in N$. We must show $(h_1n_1)(h_2n_2)^{-1}$ lies in HN. But, by Great Trick

$$(h_1n_1)(h_2n_2)^{-1} = h_1n_1{n_2}^{-1}{h_2}^{-1} = (h_1n_1{n_2}^{-1}{h_1}^{-1})(h_1{h_2}^{-1})$$

lies in $NH = HN$.

(3) is immediate.

(4). Define

$$\varphi : H \to HN/N \text{ by } x \mapsto xN.$$

As $\varphi(xy) = xyN = xNyN = \varphi(x)\varphi(y)$ for all x, y in N, the map is a homomorphism. We show $\ker\varphi = H \cap N$. If $x \in \ker\varphi \subset H$, then $N = e_{HN/N} = \varphi(x)$ so $x \in N$. Conversely, if $x \in H \cap N$, then $N = xN = \varphi(x)$ and $x \in \ker\varphi$. By definition

$$HN/N = \{hnN \mid h \in H,\, n \in N\} = \{hN \mid h \in H\},$$

so φ is surjective with kernel $H \cap N$. By the First Isomorphism Theorem φ induces an isomorphism $\overline{\varphi} : H/H \cap N \to HN/N$. □

Exercises 13.6. 1. Prove the properties of preimages in 13.1.

2. Prove the alternate form of the Correspondence Principle 13.3.

3. Let G be a finite group with subgroups H and K. Recall that $HK := \{hk \mid h \in H,\, k \in K\}$. Show that $|HK| = |H||K|/|H \cap K|$.

4. Let N and H be two normal subgroups of G such that $G = HN$. Show that there is an isomorphism $G/(H \cap N) \cong G/H \times G/N$.

5. Let G be a finite group with normal subgroups H and K of relatively prime order. Show that the group HK is cyclic if H and K are cyclic and abelian if H and K are abelian.

6. Let G be a group. Recall the *commutator* G' of G is the subgroup generated by the *commutators* of elements of G, i.e., elements of the form $[x, y] := xyx^{-1}y^{-1}$. (Cf. Exercise 11.9(17).) Show
   - (*i*) G/G' is abelian. G/G' is called the *abelianization* of G and denoted by G^{ab}.
   - (*ii*) If $N \triangleleft G$ and G/N is abelian then $G' \subset N$. In particular, the abelianization $G^{ab} = G/G'$ of G is the maximal abelian quotient of G.
   - (*iii*) If $G' \subset H \subset G$ then $H \triangleleft G$.

7. Let G be a finite abelian group and p a (positive) prime dividing the order of G. Show that there exists an element of order p in G.

8. In the previous exercise, prove the same result without assuming that G is abelian.

14. Finite Abelian Groups

In this section, we prove special cases of results to be significantly generalized in subsequent sections. We do so in order to expose some of the ideas that shall be used later that apply in this much simpler situation.

Lemma 14.1. *Let G be an abelian group and H_1, H_2 finite subgroups of relatively prime order. Then $H_1H_2 = H_2H_1$ is a group and if further $H_1 \cap H_2 = 1$, then $|H_1H_2| = |H_1||H_2|$. Moreover, if the groups H_1 and H_2 are both cyclic, then so is H_1H_2.*

We leave the proof of this as an exercise. The next result is a special case of Cauchy's Theorem (cf. Theorem 20.23 below) when G is abelian.

Proposition 14.2. *Let G be a finite abelian group and $p > 0$ a prime dividing the order of G. Then there exists an element of G of order p.*

Proof. We prove this by induction on $|G|$. We may assume that $G \neq 1$. As G has no nontrivial subgroups if and only if $G \cong \mathbb{Z}/p\mathbb{Z}$ for some prime p, we may assume that $|G|$ is not a prime. In particular, G has a subgroup $1 < H < G$. If $p \mid |H|$, we are done by induction on $|G|$, so we may assume $p \nmid |H|$. Therefore, there exists a prime $q \neq p$ such that $q \mid |H|$. By induction on $|G|$, there exists an element y in H of order q. As G is abelian, $\overline{G} = G/\langle y \rangle$ is a group. Let $^- : G \to \overline{G}$ be the canonical epimorphism. Since $|\overline{G}| < |G|$ and $p \mid |\overline{G}|$ by Lagrange's Theorem, there exists an element z in G such that $\overline{z}$ has order p in $\overline{G}$. It follows that z^p lies in the kernel of $^- : G \to \overline{G}$, so $z^p = y^i$ for some i, hence z^q has order p in G. □

We now use the proposition above to prove Sylow's First Theorem for the case when a group is abelian. (Cf. Theorem 21.1 below.)

Theorem 14.3. *Let G be a finite abelian group and $p > 0$ a prime dividing $|G|$, say $|G| = p^n m$ with p and m relatively prime. Then*

$$G(p) := \{x \in G \mid x^{p^r} = e \text{ some integer } r\} \lhd G \text{ and } |G(p)| = p^n.$$

Moreover, $G(p)$ is the unique subgroup of G of order p^n.

Remark 14.4. In the theorem, if G is finite and not abelian, there may be more than one group of order p^n. Indeed, we shall see there is only one if and only if there is a normal subgroup of order p^n.

Proof. The set $G(p)$ is a subgroup of G (why?) and normal as G is abelian. As every element in $G(p)$ has order a power of p, it follows by the previous proposition that $|G(p)| = p^r$ some $r \geq 1$ and by Lagrange's

Theorem that $r \leq n$. In particular, if x lies in $G(p)$, then $x^{p^n} = e$. We must show that $r = n$. Suppose to the contrary that we have $r < n$. Set $\overline{G} = G/G(p)$ and let $^- : G \to \overline{G}$ be the canonical epimorphism. By the Correspondence Principle (Alternative Form), $p \mid |\overline{G}|$, hence by the previous proposition, there exists an element $x \in G \setminus G(p)$ and an integer $m > 0$ such that

$$\overline{x}^{p^m} = \overline{e}, \text{ i.e., } x^{p^m} \in G(p).$$

As $x^{p^{n+m}} = (x^{p^m})^{p^n} = e$, we have $x \in G(p)$, a contradiction. Thus $|G(p)| = p^n$ as desired.

If H is subgroup of G of order p^n, then $HG(p)$ is a subgroup of G of order $|H||G(p)|/|H \cap G(p)| \geq p^n$ by the Second Isomorphism Theorem. It follows that $|H| = |H \cap G(p)| = |G(p)| = p^n$, hence $H = H \cap G(p) = G(p)$. □

Corollary 14.5. *Let G be a finite abelian group of order $n = p_1^{m_1} \cdots p_r^{m_r}$ with positive primes $p_1 < \cdots < p_r$ and positive integers $m_1, \ldots, m_r$. Then $G = G(p_1) \cdots G(p_r)$. Moreover, $G \cong G(p_1) \times \cdots \times G(p_r)$.*

Proof. As $G(p_1) \cdots G(p_r)$ is a group and we know that

$$|G(p_1) \cdots G(p_r)| = |G(p_1)| \cdots |G(p_r)|$$

by Lemma 14.1, the first statement follows by the theorem. The map

$$G(p_1) \times \cdots \times G(p_r) \to G \text{ given by } (x_1, \ldots, x_r) \mapsto x_1 \cdots x_r$$

is a group homomorphism as $G(p_i)G(p_j) = G(p_j)G(p_i)$ for all i and j. Suppose that $x_1 \cdots x_r = e$ with $x_i \in G(p_i)$ for $i = 1, \ldots, r$. As $x_j^{n/p_i^{m_i}} = e$ for all $j \neq i$ and p divides the order of x_i, it follows that $x_i = e$ for all i. Thus the map is a monomorphism and by counting an isomorphism. □

It follows, to determine all finite abelian groups up to isomorphism, we have reduced to the study of groups of order a power of a prime. If $p > 1$ is a prime, a nontrival group of order a power of p is called a *p-group*.

For notational reasons, it is easier to write abelian groups additively, i.e., use additive groups. If H and K are subgroups of an additive group, we know that $H + K$ is a subgroup. If H and K also satisfy $H \cap K = 0$, we shall write $H \oplus K$ for the group $H + K$. Of course, if H and K are finite groups of relatively prime order, we know that $H + K = H \oplus K$. The interesting case is when this is not true.

Lemma 14.6. *Let G be a finite additive p-group and suppose that the element x in G has maximal order. Then there exists a subgroup H of G satisfying $G = \langle x \rangle \oplus H$.*

PROOF. Let p^n be the order of x. By the Well-ordering Principle, there exists a maximal subgroup H of G satisfying $H \cap \langle x \rangle = \{0\}$. Therefore, $\langle x \rangle + H = \langle x \rangle \oplus H$, and we are done if $\langle x \rangle + H = G$. So suppose not. Let $\overline{G} = G/(\langle x \rangle + H)$ and $^- : G \to \overline{G}$ the canonical epimorphism. By Proposition 14.2, there exists an element $y \in G$ satisfying $y \notin \langle x \rangle + H$ and $\overline{y}$ has order p in $\overline{G}$. In particular, py lies in $\langle x \rangle + H$. Write

$$(*) \qquad py = lx + h \text{ with } l \geq 0 \text{ in } \mathbb{Z} \text{ and } h \in H.$$

It follows that $p^n y = p^{n-1} lx + p^{n-1} h$. Since x has maximal order p^n among all elements in G, we have $0 = p^n y = p^{n-1} lx + p^{n-1} h$. Therefore, $p^{n-1} lx = -p^{n-1} h$ and it lies in $\langle x \rangle \cap H = 0$. It follows that $p \mid l$, say $l = l'p$. Thus (*) becomes

$$(\dagger) \qquad py = l'px + h \text{ equivalently } p(y - l'x) = h \text{ in } H.$$

As $y \notin \langle x \rangle + H$, we must have $y - l'x \notin \langle x \rangle + H$. The maximality of H now yields

$$\langle y - l'x, H \rangle \cap \langle x \rangle \neq 0.$$

Therefore, there exist integers r, s with r nonzero satisfying

$$0 \neq rx = s(y - l'x) + h' \text{ for some } h' \in H.$$

Hence $sy = rx + sl'x - h'$ lies in $\langle x \rangle + H$. We look at the integer s.

If $p \mid s$, then by (†), we have $s(y - l'x)$ lies in H. This implies that rx lies in $\langle x \rangle \cap H = 0$, a contradiction.

If $p \nmid s$, then p and s are relatively prime, so there exists an equation $1 = pa + sb$ for some integers a and b. Since both py and sy lie in $\langle x \rangle + H$, we have $y = apy + bsy$ lies in $\langle x \rangle + H$, a contradiction. The lemma now follows. □

Corollary 14.7. *Let G be a finite p-group. Then G is a product of cyclic p-groups.*

PROOF. We may assume that G is additive and not cyclic. In particular $|G| > p$. By the lemma, $G = \langle x \rangle \oplus H$ for some x in G and subgroup H of G. As G is not cyclic, $H \neq G$, and the result follows by induction on $|G|$. □

Remark 14.8. Let $p > 0$ be a prime, N a positive integer. Suppose that G is an additive group in which every element a in G satisfies $p^N a = 0$. Let x be an element of G of maximal order. Then there exists

a subgroup H of G satisfying $G = \langle x \rangle \oplus H$. The proof is essentially the same as that above except that one replaces the Well-Ordering Principle by Zorn's Lemma, an axiom that we shall study later. (Cf. §25.)

Proposition 14.9. *Every finite abelian group is a product of cyclic groups.*

PROOF. By 14.5 every finite abelian group is a product of finite abelian p-groups. By Corollary 14.5, every finite abelian p-group is a product of cyclic p-groups. □

Theorem 14.10. (Fundamental Theorem of Finite Abelian Groups) *Let G be a finite additive group and for each prime $p > 0$ dividing G, let $G(p)$ be the unique p-subgroup of G of maximal order. Then*

$$G = \bigoplus_{p||G|} G(p)$$

Moreover, if $p \mid |G|$, then

$$G(p) \cong \bigtimes_{i=1}^{r} \mathbb{Z}/p^{n_i}\mathbb{Z}$$

with r unique and $1 \leq n_1 \leq \cdots \leq n_r$ also unique relative to this ordering. In particular, any finite abelian group is a product of cyclic p-groups for various primes p.

PROOF. Let $p \mid |G|$. By Corollary 14.5 and Proposition 14.9, it suffices to show $G(p) \cong \bigtimes_{i=1}^{r} \mathbb{Z}/p^{n_i}\mathbb{Z}$ and uniquely up to isomorphism. As every abelian p-group is isomorphic to a product of cyclic p-groups by Lemma 14.6 and every cyclic p-group must be isomorphic to $\mathbb{Z}/p^a\mathbb{Z}$ for some integer a, to finish, it suffices to show that if

$$(*) \qquad \bigtimes_{i=1}^{r} \mathbb{Z}/p^{n_i}\mathbb{Z} \cong \bigtimes_{j=1}^{s} \mathbb{Z}/p^{m_j}\mathbb{Z}$$

with $n_1 \geq \cdots \geq n_r$ and $m_1 \geq \cdots \geq m_s$ (we changed notation for convenience), then $r = s$ and $n_i = m_i$ for all i. We may assume that the number N of the n_i that are equal to 1 is less than the number M of the m_j that equal 1. In particular, $p^N \leq p^M$. Multiplying (*) by p, we see that

$$\bigtimes_{i=N+1}^{r} \mathbb{Z}/p^{n_i-1}\mathbb{Z} \cong \bigtimes_{j=M+1}^{s} \mathbb{Z}/p^{m_j-1}\mathbb{Z}$$

as $p(\mathbb{Z}/p^k\mathbb{Z}) \cong Z/p^{k-1}\mathbb{Z}$ for all k. By induction, $n_i - 1 = m_i - 1$ for all $i > N$. As $|G(p)| = \prod_{i=1}^{r} p^{n_i} = \prod_{i=1}^{s} p^{m_j}$, we conclude that $r = s$ and $n_i = m_i$ for all i. □

Remarks 14.11. The theorem generalizes to finitely generated abelian groups. We indicate how this would proceed without a full proof. An abelian group G is called

(*i*) *torsion-free* if every non-identity element in G has infinite order.
(*ii*) *torsion* if every non-identity element in G has finite order.

Let G be an abelian group and set

$$G_t := \{x \in G \mid x \text{ has finite order}\}.$$

By the Binomial Theorem, $G_t \subset G$ is a subgroup and by the Correspondence Principle, G/G_t is torsion-free. If G is finitely generated, then one can show that G_t is a finitely generate torsion abelian group hence finite, so decomposes uniquely as in the theorem and, moreover, that $G \cong G_t \times G/G_t$. One also shows if G is a finitely generated torsion-free abelian group, then $G \cong \mathbb{Z}^r$ for a unique integer r. Therefore, any finitely generated abelian group G is isomorphic to $\bigtimes_{p||G_t|} G(p) \bigtimes \mathbb{Z}^r$ for some unique r and the $G(p)$ have unique decompositions given by Theorem 14.10. We shall prove this and an important generalization of this that has significant application to the theory of matrices in §41.

Exercises 14.12. 1. Prove Lemma 14.1. If we assume that G is an arbitrary group, what conditions on H_1 and H_2 will still guarantee the result. Prove this.

2. Let G be an abelian group. Show that $G(p) := \{x \in G \mid x^{p^r} = e \text{ some } r\}$ is a subgroup of G.
3. Show that $p(\mathbb{Z}/p^k\mathbb{Z}) \cong Z/p^{k-1}\mathbb{Z}$ for all primes p and positive integers k.
4. Determine up to isomorphism all finite abelian groups up to order 30.
5. Determine all $n > 1$ such that all abelian groups of order n are cyclic and prove your determination.

15. Addendum: Finitely Generated Groups

Most of what we do in the study of group theory in these pages involves finite groups. In this short section, we show how the General Cayley Theorem 12.4 can be used to obtain a few results about finitely generated groups, i.e., a group G containing a finite set S of elements satisfying $G = \langle S \rangle$. Of course, finite groups are finitely generated, but the results that we obtain here are trivial for finite groups, as the questions we ask are about finitely generated groups that contain a proper subgroup of finite index.

We have seen that if G is an arbitrary group and H a subgroup of finite index n in G, then the General Cayley Theorem gives a group homomorphism

$$\lambda : G \to \Sigma(G/H) \cong S_n \text{ given by } x \mapsto \lambda_x : gH \mapsto xgH,$$

with $\ker\lambda \subset H \subset G$ (and the largest normal subgroup of G in H). Hence, by the First Isomorphism Theorem, λ induces a group monomorphism

$$\overline{\lambda} : G/\ker\lambda \to \Sigma(G/H) \cong S_n \text{ given by } x \mapsto \lambda_x : g \mapsto gH.$$

Since $\Sigma(G/H)$ is a finite group, so is $G/\ker\lambda$, i.e., $\ker\lambda \subset G$ has finite index. Therefore, G contains a normal subgroup of finite index. (Cf. Exercise 12.11(6).) We want to strengthen this result if G is finitely generated.

Proposition 15.1. *Let G be a finitely generated group and n a positive integer. Then there exist finitely many subgroups (if any) of G of index n.*

PROOF. Let $G = \langle a_1, \ldots, a_r\rangle$ and

$$(*) \qquad \varphi : G \to S_n$$

be a group homomorphism. Then the map φ is completely determined by

$$\varphi(a_1), \ldots, \varphi(a_r)$$

in the finite set S_n. Thus there exist only finitely many possible group homomorphisms φ in (*), hence finitely many normal subgroups N of G of finite index with $N = \ker\varphi$ for such φ. Fix a φ. By the First Isomorphism Theorem, we have $G/\ker\varphi\lambda \cong \varphi(G) \subset S_n$ is a subgroup. As S_n is finite, the Correspondence Principle implies that there exist finitely many subgroups H of G satisfying $\ker\varphi \subset H \subset G$, in particular, there can be only finitely many H (if any) with $[G : H] = n$. As there are finitely many φ, the result follows. □

Our strengthening of the goal above can now be stated and established. Recall that a subgroup K of a group G is called *characteristic* if $\sigma|_K$ is an automorphism of K (equivalently, $\sigma(K) = K$) for every automorphism σ of G.

Corollary 15.2. *Let G be a finitely generated group. Suppose G contains a subgroup H of finite index. Then there exists a characteristic subgroup K of H of finite index in G.*

PROOF. By the proposition, there exist finitely many subgroups

$$H = H_1, \ldots, H_m$$

of finite index $[G : H]$. It follows, given an automorphism φ of G, that $\varphi(H) = H_i$ for some i, i.e., φ permutes the H_i. Let $K := \cap_{I=1}^m H_i$. Then we have $\varphi(K) = K$, so K is a characteristic subgroup of H. By Poincaré's Lemma (cf. Exercise 10.16(4)), K is of finite index in G. □

The second result is much harder to prove.

Theorem 15.3. *Let G be a finitely generated group and H be a subgroup of finite index. Then H is a finitely generated group.*

PROOF. Let $G = \langle a_1, \ldots, a_r \rangle$ and

$$y_1 = e, \ldots, y_n$$

be a transversal (i.e., a system of representatives for the cosets) of H in G. Let

$$\lambda : G \to \Sigma(G/H) \text{ be given by } a \mapsto \lambda_x : aH \mapsto xaH.$$

Therefore, λ_x permutes $y_1H, \ldots, y_nH$. Fix i, $1 \leq i \leq r$. Then for each j, $1 \leq j \leq n$, there exists a $k = k(i,j)$ with $1 \leq k \leq n$ satisfying

$$a_i y_j H = y_k H.$$

Therefore, there exist h_{ij} in H, $1 \leq i \leq r$ and $1 \leq j \leq n$ satisfying

$$y_k = a_i y_j h_{ij}.$$

Let $H_0 := \langle h_{ij} \mid 1 \leq i \leq r, 1 \leq j \leq n \rangle \subset G$. Then H_0 is clearly a finitely generated subgroup of G contained in H. So it suffices to show that $H \subset H_0$. Set

$$W := \bigcup y_j H_0.$$

For each i, $1 \leq i \leq r$, we have

$$a_i W = \bigcup_{j=1}^{n} a_i y_j H_0 = \bigcup_{j=1}^{n} y_{k(i,j)} h_{ij} H_0 = \bigcup_{k=1}^{n} y_k H_0 = W,$$

since

$$\{a_i y_1 H, \ldots, a_i y_n\} = \{y_1, \ldots, y_n\}.$$

Consequently, $a_i W = W$ for each $i = 1, \ldots, r$. As $G = \langle a_1, \ldots, a_r \rangle$, we have

$$GW = W = \bigcup_{j=1}^{n} H_0.$$

Since $e \in W$, we conclude that

$$G = W = \bigcup_{j=1}^{n} y_j H_0 = \bigvee_{j=1}^{n} y_j H_0.$$

In particular,

$$H \subset G = \bigvee_{j=1}^{n} y_j H_0.$$

But $H = y_1 H$ is disjoint from $\bigcup_{j=2}^{n} y_j H_0$ as it is disjoint from $\bigcup_{j=2}^{n} y_j H$, so we conclude that $H = H_0$ as needed.

□

16. Series

One way of studying algebraic objects is to break them down into simpler pieces. For example, if G is a group, it may be the direct product of groups, e.g., $\mathbb{Z}/6\mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$. Unfortunately, this does not happen very often for groups. An alternative approach would be to take a given non-trivial group G and find the largest normal subgroup N_1 in G satisfying $N_1 < G$. By the Correspondence Principle, G/N_1 then would be a simple group, and we have, in some sense reduced the study of G to the groups N_1 and G/N_1. We then proceed with the same analysis on N_1, etc. Another approach would be to find a normal subgroup N_1 of G such that G/N_1 is abelian (or cyclic), then continue on N_1, etc. In any of these cases, we produce a chain of subgroups $G > N_1 > N_2 > \cdots$ (where $A > B$ means $A \supset B$, $A \neq B$). Then various possibilities open:

(1) This process never stops.
(2) There exists an i such that N_i is nice, e.g., in our cases simple, abelian, cyclic, respectively.
(3) At some point, there does not exist such an N_i.

Of course, we are interested in the second possibility.

Definition 16.1. Let G be a non-trivial group. A sequence of groups

$$(*) \qquad N_0 \subset N_1 \subset N_2 \subset \cdots \subset N_n = G$$

is called a *subnormal* (respectively, *normal*) *series* if $N_i \triangleleft N_{i+1}$ (respectively, $N_i \triangleleft G$) for all i, and *proper* if $N_i < N_{i+1}$ for all i. The quotients $(G/N_{n-1} =)N_n/N_{n-1}, \ldots, N_1/N_0$ are called the *factors* of the series. If (*) is a subnormal series for which $N_0 = 1$ and $N_n = G$, then it is called

(i) a *cyclic series* if N_{i+1}/N_i is cyclic for all i.
(ii) an *abelian series* if N_{i+1}/N_i is abelian for all i.
(iii) a *composition series* if N_{i+1}/N_i is simple for all i.

A group is called *solvable* if it has an abelian series and *polycyclic* if it has a cyclic series. (Since the trivial group is cyclic, we also say that it has a cyclic series.)

Example 16.2. Every abelian group is solvable and every cyclic group is polycyclic.

One of the major results in finite group theory is the Feit-Thompson Theorem that every finite group of odd order is solvable. It is the first

major theorem in the classification of finite simple groups. Its proof takes over 250 journal pages.

Theorem 16.3. *The following are true:*

(1) A subgroup of a solvable group is solvable.
(2) The homomorphic image (i.e., the image of a group under a group homomorphism) of a solvable group is solvable.
(3) If $N \triangleleft G$ and both N and G/N are solvable then so is G.

Proof. We leave this as an exercise. It is important that you do this exercise, as it will teach you how to use the isomorphism theorems. □

Let G be a group. The *commutator* of x and y in G is $[x, y] := xyx^{-1}y^{-1}$. The group generated by all commutators is called the *derived subgroup* of G and denoted either by $[G, G]$ or by G'. By recursion, we define, the *nth derived subgroup* of G by $G^{(n)} = [G^{(n-1)}, G^{(n-1)}]$. We leave the following easy computations as an exercise.

Properties 16.4. Let G be a group. If a, b, c are elements of G and $\sigma : G \to G_1$ a group homomorphism. Then

(1) $[a, b]^{-1} = [b, a]$.
(2) $\sigma([a, b]) = [\sigma(a), \sigma(b)]$.
(3) $[a, bc] = [a, b]b[a, c]b^{-1} = [a, b][bab^{-1}, bcb^{-1}]$.

Proposition 16.5. *Let G be a group. Then $G^{(n)}$ is a characteristic subgroup of $G^{(n-1)}$ for all n. In particular, $G^{(n)}$ is characteristic in G, and*

$$G^{(n)} \subset G^{(n-1)} \subset \cdots \subset G^{(1)} \subset G^{(0)} = G$$

is a normal series (even a characteristic series — obvious definition).

Proof. Being characteristic is transitive, so this follows immediately, using Properties 16.4 □

Theorem 16.6. *Let G be a group. Then G is solvable if and only if there exist an integer n such that $G^{(n)} = 1$.*

Proof. By Exercise 11.9(17), each factor group $G^{(i)}/G^{(i+1)}$ is abelian, hence if $G^{(n)} = 1$ for some n, then G is solvable. Conversely, suppose that

$$(\dagger) \qquad 1 = N_n \subset N_{n-1} \subset N_{n-2} \subset \cdots \subset N_1 \subset G$$

is an abelian series. It suffices to show that $G^{(i)} \subset N_i$ for all i. By Exercise 13.6(6), we know that $G' \subset N_1$. By induction, we may assume that $G^{(i-1)} \subset N_{i-1}$. But then we have $G^{(i)} = [G^{(i-1)}, G^{(i-1)}] \subset [N_{i-1}, N_{i-1}] \subset N_i$, again using Exercise 13.6(6). □

We turn to composition series. We need the following easy lemma.

Lemma 16.7. (Dedekind's Modular Law) *Let A, B, and C be three subgroups of G. If A is a subset of C, then $A(B \cap C) = (AB) \cap C$.*

PROOF. $\subseteq$: If $a \in A$, $x \in B \cap C$, then $ax \in AB \cap AC \subset (AB) \cap C$.

$\supseteq$: If $a \in A$, $b \in B$ satisfy $ab \in C$, then $b = a^{-1}C \subset C$. □

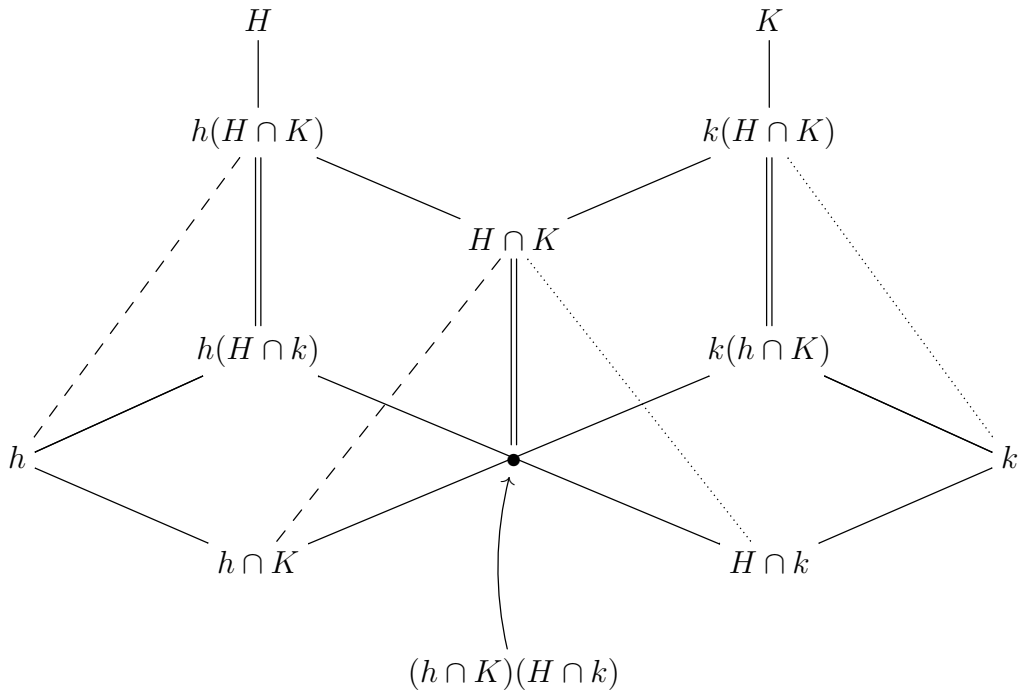
We next prove an elaborate form of the Second Isomorphism Theorem. For notational simplicity, we use right cosets.

Lemma 16.8. (Zassenhaus Butterfly Lemma) *Let G be a group containing four subgroups h, H, k, and K satisfying $h \triangleleft H$ and $k \triangleleft K$. Then we have the following:*

(1) $(H \cap k)(h \cap K) \triangleleft H \cap K$.
(2) $k(h \cap K) \triangleleft k(H \cap K)$.
(3) $h(H \cap k) \triangleleft h(H \cap K)$.
(4) There exist isomorphisms:

$$\begin{aligned} k(H \cap K)/k(h \cap K) &\cong H \cap K/(H \cap k)(h \cap K) \\ &\cong h(H \cap K)/h(H \cap k). \end{aligned}$$

We have the following picture illustrating the subgroups, where a group at the bottom of a line means that it is a subgroup of the group at the other end. We shall see the quotients (top/bottom) of like dotted, double, dashed lines, respectively, are isomorphic.



PROOF. We use the Second and Third Isomorphism Theorems. As $k \lhd K$ in the lemma and $H \cap K \subset K$ is a subgroup, we have $k(H \cap K)$ is a group satisfying $k \lhd k(H \cap K)$, and $H \cap k = (H \cap K) \cap k \lhd H \cap K$ together with an isomorphism

$$(*) \qquad k(H \cap K)/k \xrightarrow{\sim} H \cap K/H \cap k \text{ given by } ka \mapsto (H \cap k)a.$$

Similarly, as $h \lhd H$ and $H \cap K \subset H$ is a subgroup, we have $h(H \cap K)$ is a group satisfying $h \lhd h(H \cap K)$, and $h \cap K \lhd H \cap K$ together with an isomorphism

$$h(H \cap K)/h \xrightarrow{\sim} H \cap K/h \cap K \text{ given by } ha \mapsto (h \cap K)a.$$

As the product of normal subgroups is normal, we also have

$$(\dagger) \qquad (H \cap k)(h \cap K) \lhd H \cap K.$$

By restriction, the isomorphism in (*) induces an isomorphism

$$(\star) \qquad k(H \cap k)(h \cap K)/k \xrightarrow{\sim} (H \cap k)(h \cap K)/H \cap k.$$

By the Modular Law, we have $k(H \cap k) = kH \cap k = k$. Consequently, $k(H \cap k)(h \cap K) = k(h \cap K)$. Applying the Correspondence Principle to the isomorphism in (*), we know that $(H \cap k)(h \cap K) \lhd H \cap K$ corresponds to $k(h \cap K) = k(H \cap k)(h \cap K) \lhd k(H \cap K)$.

Using the Third Isomorphism Theorem, we conclude by (*) and ($\star$) that

$$\begin{aligned} k(H \cap K)/k(h \cap K) &\cong \frac{k(H \cap K)/k}{k(h \cap K)/k} \\ &\cong \frac{H \cap K/H \cap k}{(H \cap k)(h \cap K)/H \cap k} \\ &\cong H \cap K/(H \cap k)(h \cap K). \end{aligned}$$

The other isomorphism is obtained similarly. □

Next we wish to show if a group has a composition series, then the number of terms in a minimal, proper composition series is constant with factors unique up to isomorphism.

Definition 16.9. Given a proper subnormal series of a group G, we say another subnormal series is a *refinement* of the given series, if we add in new subgroups, and it is called a *proper refinement* if the resulting subnormal series is proper. Two proper subnormal series for G are called *equivalent* if they have the same number of terms, say n, and the n factors of each are isomorphic up to a permutation.

If a group has a composition series, we want to prove that the number of factors in a composition series is unique as well as these factors are unique up to isomorphism and order. The next theorem is the key to this. We use the Butterfly Lemma to prove it.

Theorem 16.10. (Schreier Refinement Theorem) *Any two proper subnormal series for a group G have equivalent refinements.*

Proof. Suppose

$$\begin{aligned} 1 = N_0 \lhd N_1 \lhd \cdots \lhd N_r = G \ \text{ and} \\ 1 = H_0 \lhd H_1 \lhd \cdots \lhd H_s = G \end{aligned}$$

are two proper subnormal series for G. We have groups:

$$N_{i,j} = N_i(H_j \cap N_{i+1})$$

that form subnormal series

$$N_i = N_{i,0} \lhd \cdots \lhd N_{i,s} = N_{i+1} \ \text{ for all } i.$$

by the Butterfly Lemma. It follows that we have a subnormal series

$$1 = N_{0,0} \lhd \cdots \lhd N_{r,s} = G. \tag{1}$$

Similarly, we have another subnormal series

$$1 = H_{0,0} \lhd \cdots \lhd H_{s,r} = G \tag{2}$$

where

$$H_{j,i} = H_j(N_i \cap H_{j+1}).$$

By the Butterfly Lemma, we have

$$N_{i,j+1}/N_{i,j} \cong H_{j,i+1}/H_{j,i} \text{ for all } i \text{ and } j,$$

hence the series (1) and (2) are equivalent. In particular, if a factor isomorphic to 1 occurs in one, it occurs in the other. It follows that from these we can obtain common proper refinements. □

If $1 < N_1 \lhd \cdots \lhd N_{r-1} \lhd G$ is a proper subnormal series, we say the series has *length* r (which equals the number of *links*). As the trivial group is not considered a simple group, any composition series for a group must be proper. Moreover, by the Correspondence Principle, no composition series can have a proper refinement.

Theorem 16.11. (Jordan-Holder Theorem) *Let G be a group having a composition series of length r. Then every composition series for G has length r and any proper subnormal series for G can be refined to a composition series for G. Moreover, all composition series of G are equivalent.*

PROOF. Let $1 = N_0 < \cdots < N_n < G$ be a proper subnormal series for G. By the Schreier Refinement Theorem, this subnormal series and a composition series for G have a common proper refinement. The result follows by the above remark. □

Example 16.12. Let $m = a_1 \cdots a_r$ in $\mathbb{Z}$ with each integer $a_i > 1$. Then we have a subnormal series:

$$0 = a_1 \cdots a_r\mathbb{Z}/m\mathbb{Z} < \cdots < a_1\mathbb{Z}/m\mathbb{Z} < \mathbb{Z}/m\mathbb{Z}.$$

As $\mathbb{Z}/m\mathbb{Z}$ is a finite group, it has a composition series by Exercise 16.14(10) that is essentially unique. This is really the Fundamental Theorem of Arithmetic.

Remark 16.13. In an analogous way, one can define composition series for vector spaces. A "simple" vector space is a line (proof?). Finite dimensional vector spaces have composition series (and infinite dimensional ones do not). Let V be a nonzero finite dimensional F-vector space. Say V has a composition series of length d, i.e, we get a a series $0 < V_1 < \cdots < V_d = V$, with simple factors V_i/V_{i-1} (after generalizing the concept of quotients to vector spaces). Thus d is the *dimension* of V, and the Jordan-Holder Theorem gives an alternate proof of the invariance of the number of vectors in the basis for a finite dimensional vector space.

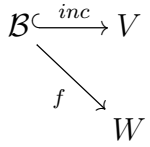
Exercises 16.14. 1. Prove Properties 16.4.

2. Prove that every finite abelian group is polycyclic. Use this to prove that every finite solvable group is polycyclic. Give an example of a solvable group that is not polycyclic.
3. Prove Theorem 16.3 using the isomorphism theorems. (Do not use Theorem 16.6.)
4. Prove the analog of Theorem 16.3 replacing the word solvable with the word polycyclic using the isomorphism theorems.
5. Let H and K be subgroups of a group G. Define
$$[H, K] := \langle [h, k] \mid h \in H, k \in K \rangle.$$
Show both of the following:
   (*i*) If $K \triangleleft G$ and $K \subset H$, then $[H, G] \subset K$ if and only if $H/K \subset Z(G/K)$.
   (*ii*) If $\varphi : G \to L$ is a group epimorphism and $A \subset Z(G)$, then $\varphi(A) \subset Z(L)$.
6. Let G be group.
   (*i*) Set $\Gamma_1(G) = G$ and inductively define $\Gamma_{n+1}(G) := [\Gamma_n(G), G]$ for $n > 1$. Show $\Gamma_{n+1}(G) \subset \Gamma_n(G)$ and $\Gamma_n(G) \triangleleft\triangleleft G$ for all n.
   (*ii*) Set $Z^0(G) = 1$ and inductively define $Z^{n+1}(G)$ by
   $$Z^{n+1}(G)/Z^n(G) = Z(G/Z^n(G)),$$
   i.e., the preimage of $G/Z^n(G)$ in G under the canonical epimorphism. Show $Z^{n+1}(G) \triangleleft Z^n(G)$ and $Z^n(G) \triangleleft\triangleleft G$ for all $n > 0$.
7. Define the *descending central series* of G by
$$G = \Gamma_1(G) \supset \Gamma_2(G) \supset \cdots \supset \Gamma_n(G) \supset \cdots$$
and the *ascending central series* of G by
$$1 = Z^0(G) \subset Z^1(G) \subset \cdots \subset \cdots .$$
Prove that $Z^n(G) = G$ if and only if $\Gamma_n(G) = 1$. Moreover, if this is the case, then $\Gamma_{i+1} \subset Z^{n-i}$ for $i = 0, \ldots, n$.
8. A group G is called *nilpotent* if there exists an n such that $\Gamma_n(G) = 1$ (and n is called the class of G if n is the least such integer). Show the following
   (*i*) If G is nilpotent, then every subgroup of G is nilpotent.
   (*ii*) If G is nilpotent and $H \triangleleft G$, then G/H is nilpotent.
   (*iii*) A direct product of finitely many nilpotent groups is nilpotent.
   (*iv*) Give an example of a group G with $H \triangleleft G$ with H and G/H nilpotent, but G not nilpotent.

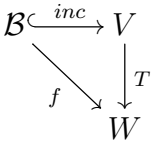
9. Show that a group G is nilpotent if and only if there exists a normal series
$$G = G_0 \supset G_1 \supset \cdots \supset G_n = 1$$
in which each $G_i \triangleleft G$ and satisfies $G_i/G_{+1} \subset Z(G/G_{i+1}$ for all i. Such a series is called a *central series* for G.
10. Prove that every finite group has a composition series. Give an example of an infinite group that does not have a composition series.
11. Let G be a group and N a normal subgroup of G. Prove that G has a composition series if and only if N and G/N have composition series. (It is not true that if G has a composition series that any subgroup does. Cf. Exercise 22.22(11).)

17. Free Groups

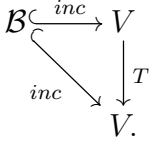
Let V be a vector space over F with basis $\mathcal{B}$. Then $V, \mathcal{B}$ satisfies the following *universal property of vector spaces*: Given any vector space W over F and diagram of set maps



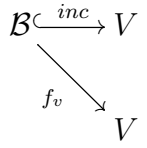
with *inc* the inclusion of $\mathcal{B}$ into V, then there exists a unique linear transformation $T : V \to W$ such that the diagram



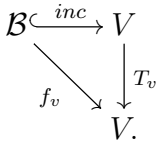
commutes. This follows from the theorem in linear algebra that says a linear transformation $T : V \to W$ is completely determined by where a basis for V is mapped. Conversely, suppose that we know that $\mathcal{B}$ is a subset of the vector space V and $V, \mathcal{B}$ satisfies the universal property above. Then there exists a unique linear transformation $T : V \to V$ satisfying



commutes. By uniqueness, T must be the identity map, and it follows that $\mathcal{B}$ spans V. If $v \in \mathcal{B}$, let



with $f_v : \mathcal{B} \to V$ the map sending $v \mapsto v$ and $v' \mapsto 0$ for all $v' \in \mathcal{B} \setminus \{v\}$. Then there exists a unique linear transformation $T_v : V \to V$ such that

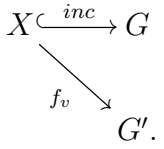


commutes. It follows (check) that $\mathcal{B}$ is linearly independent. In particular, V is completely determined by the universal property of vector spaces. Note that this says that V is completely determined by a set of generators satisfying no non-trivial relations in V.

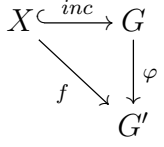
We can, in fact, modify this universal property with the inclusion map replaced by a set injection $i_V : \mathcal{B}_0 \to V$ with $i_V(\mathcal{B}_0) = \mathcal{B}$. Indeed, we can even drop the requirement that i_V be an injection and this still shows that V is a vector space with basis $\mathcal{B}$. We leave this as an exercise.

We now mimic the above for groups.

Definition 17.1. Let G be a group with X a non-empty set in G. We say that G is a *free group* on *basis* X if it satisfies the following *universal property of free groups*: Given any group G' and diagram



of set maps with *inc* the inclusion of X into G, then there exists a unique group homomorphism $\varphi : V \to W$ such that the diagram



commutes.

Remark 17.2. As in the vector space case, we can replace the inclusion map of X in G with a set injection $i_G : X_0 \to G$ with $X = i_A(X_0)$; and, in fact, the map i_G need not be assumed to be an injection.

Remark 17.3. Let X_0 and Y_0 be sets, A and B groups with set maps $i_A : X_0 \to A$ and $i_B : Y_0 \to B$ such that A is a free group on $X = i_A(X_0)$ and B is a free group of $Y = i_B(Y_0)$. Suppose that there exists a bijection $j : X_0 \to Y_0$. Then there exists unique group homomorphisms $\varphi : A \to B$ and $\psi : B \to A$ such that

$$\begin{array}{ccc} X_0 & \xrightarrow{j} & Y_0 \\ {\scriptstyle i_A}\downarrow & & \downarrow{\scriptstyle i_B} \\ A & \underset{\psi}{\overset{\varphi}{\rightleftarrows}} & B \end{array}$$

commutes. But it then follows that $\psi \circ \varphi = 1_A$ and $\varphi \circ \psi = 1_B$, so φ and ψ are inverse isomorphisms. Thus if A and B are free groups on bases of the same cardinality then they are isomorphic.

Theorem 17.4. *Let X be a non-empty set. Then there exists a free group G on X.*

Proof. Choose a set X' disjoint from X of the same cardinality, so we have a bijection $X \to X'$ which we denote by $x \mapsto x^{-1}$. Let X'' be a set disjoint from $X \cup X'$ having precisely one element. Let $X'' = \{e\}$. We call $A = X \cup X' \cup X''$ the *alphabet* and the elements of A the *letters*. If $x \in X$, we also write x^1 for x and if a is a letter, we also write $(a^{-1})^{-1}$ for a. Call a sequence $w = (a_1, \dots, a_n, \dots)$, $a_i \in A$, a *word* if $a_i \neq e$ for only finitely many i. Let $\widetilde{W}$ be the set of words in $\bigtimes_{i=1}^{\infty} A$. The word $(e, \dots, e, \dots)$ is called the *empty word*. A word $w = (a_1, \dots, a_n, \dots)$ in $\widetilde{W}$ is called a *reduced word* if w satisfies

(i) $a_i \neq a_{i+1}^{-1}$ if $a_i \neq e$.
(ii) If $a_i = e$, then $a_k = e$ for all $k \geq i$.

If w is a reduced word, but not the empty word, we can write it as

$$w = x_1^{n_1} \cdots x_m^{n_m},\ x_i \in X,\ n_i \in \{\pm 1\} \text{ for all } i \text{ and some } m \text{ with } x_i^{n_i} \neq x_{i+1}^{-n_i} \text{ for all } i.$$

We write e for the empty word. Let W be the set of reduced words.

Note. *Spelling* in W is unique by the definition of $\bigtimes_{i=1}^{\infty} A$.

Define $\cdot : W \times W \to W$ as follows: $w = e \cdot w = w \cdot e$ for any $w \in W$ and if $w = x_1^{n_1} \cdots x_r^{n_r}$ and $w' = y_1^{m_1} \cdots y_s^{m_s}$ with $x_i, y_j \in X$ and $n_i, m_j \in \{\pm 1\}$ for all i and j, let $w \cdot w' = x_1^{n_1} \cdots x_r^{n_r} y_1^{m_1} \cdots y_s^{m_s}$ by *juxtaposition* and

then *reduce*, i.e., if $x_r^{n_r} = y_1^{-m_1}$ delete the pair unless reduced to the empty word and continue in this way to get a reduced word.

Claim 17.5. The map $\cdot : W \times W \to W$ is well-defined and makes W into a group such that the inclusion $X \subset W$ makes X into a free group on basis X, so $W = \langle X \rangle$.

The details that $\cdot : W \times W \to W$ is a well defined map is left as an exercise (cf. the above). If we show that W satisfies associativity under this map, it is then clear that W is a group. To do this we use a trick of Van der Waerden. For each $x \in X$ and $n \in \{\pm\}$, define a map

$$|x^n| : W \to W \text{ by } e \mapsto x^n \text{ and}$$
$$x_1^{n_1} \cdots x_r^{n_r} \mapsto \begin{cases} x^n x_1^{n_1} \cdots x_r^{n_r} \text{ if } x^n \neq x^{-n_1} \\ x_2^{n_2} \cdots x_r^{n_r} \text{ otherwise.} \end{cases}$$

Clearly,

$$|x^n||x^{-n}| = 1_W = |x^{-n}||x^n|,$$

so $|x^n|$ is a permutation of W, i.e., $|x^n| \in \Sigma(W)$ for all $x \in X$ and $n \in \{\pm 1\}$. Let

$$W_0 := \langle |x| \mid x \in X \rangle \subset \Sigma(W), \text{ a subgroup.}$$

As spelling is unique, we have a surjection

$$\varphi : W \to W_0 \text{ given by } e \mapsto 1_W \text{ and } x_1^{n_1} \cdots x_r^{n_r} \mapsto |x_1^{n_1}| \cdots |x_r^{n_r}|.$$

Let $z \in W_0$. Then there exist $x_i \in X$ and $n_i \in \{\pm 1\}$ such that $z = |x_1^{n_1}| \cdots |x_r^{n_r}|$. If $x_i^{n_i} = x_{i+1}^{-n_i}$, then $|x_i^{n_1}||x_{i+1}^{n_{i+1}}| = 1_W$ and we can cancel them (or reduce to 1_W), which we may assume has been done. But then

$$\psi : W_0 \to W \text{ given by } 1_W \mapsto e \text{ and } |x_1^{n_1}| \cdots |x_r^{n_r}| \mapsto |x_1^{n_1}| \cdots |x_r^{n_r}|(e)$$

is well-defined as $|x_1^{n_1}| \cdots |x_r^{n_r}| = |x_1^{n_1} \cdots x_r^{n_r}|$ and spelling is unique in W. In particular, $\psi \circ \varphi = e$ and $\varphi \circ \psi = 1_W$. It follows that φ is a bijection. By construction

$$\varphi(w_1 \cdot w_2) = \varphi(w_1)\varphi(w_2) \text{ for all } c_1, w_2 \in W.$$

As W_0 is a group, associativity holds in W_0, hence associativity holds in W.

We leave it as an exercise to show that W is a free group on basis X, i.e., satisfies the universal property of free groups. □

Definition 17.6. Let H be a subgroup of a group G. The *normal closure* of H in G is the smallest normal subgroup of G containing H, i.e., the group $\langle xHx^{-1} \mid x \in G \rangle$.

Definition 17.7. A group G is said be defined by *generators* $X = \{x_i \mid i \in I\}$ and *relations* $Y = \{y_j \mid j \in J\}$ if $G \cong E/N$ where E is a free group on basis X and N is the normal closure of $\langle Y \rangle$. If this is the case, we say (X, Y) is a *presentation* and also (by abuse of notation) write $G = \langle X \mid Y \rangle$.

Corollary 17.8. *Every group has a presentation, unique up to isomorphism.*

PROOF. Let G be a group and X a subset of G that generates G (e.g., we can take G itself). Let E be the free group on basis X. By the universal property of free groups, there exists a unique group homomorphism $\varphi : E \to G$ such that

$$\begin{array}{ccc} X & \xhookrightarrow{inc} & E \\ & \searrow^{inc} & \downarrow{\scriptstyle \varphi} \\ & & G \end{array}$$

commutes. As $G =< X >$, the homomorphism φ is onto. Let Y be any subset of $\ker \varphi$ such that $\ker \varphi$ is the normal closure of $< Y >$. Then (X, Y) is a presentation of G. We leave the proof of uniqueness as an exercise. □

Corollary 17.9. (Van Dyck's Theorem) *Let X be a non-empty set and Y a set of reduced words based on X (obvious definition). Let $G = \langle X \mid Y \rangle$ be a presentation of G. Suppose that H is a group satisfying $H = \langle X \rangle$ and H satisfies all the relations in Y. Then there exists a (group) epimorphism $\psi : G \to H$.*

PROOF. Let E be the free group of X and N the normal closure of $< Y >$ in E. As

$$\begin{array}{ccc} X & \xhookrightarrow{inc} & E \\ & \searrow^{inc} & \\ & & H, \end{array}$$

by the universal property of free groups, there exist unique group homomorphisms

$$\begin{array}{ccc} E & \xrightarrow{\theta} & G \\ & \searrow^{\varphi} & \\ & & H. \end{array}$$

We define $\psi : G \to H$ as follows: For each $g \in G$ choose x_g in the fiber $\theta^{-1}(g) := \{y \in E \mid \theta(y) = g\}$ and set $\psi(g) := \varphi(x_g)$. We must show

that ψ is well-defined, i.e., independent of the choice of $x_g \in \theta^{-1}(g)$. Suppose that x, y in E satisfies $\theta(x) = g = \theta(y)$. Then $\theta(y^{-1}x) = e_G$, so $y^{-1}x$ lies in $\ker\theta = N$, the normal closure of $< Y >$ in E. If $H =< X|Y' >$, then N lies in the normal closure $\ker\varphi = N'$ of $< Y' >$ in E by hypothesis. Consequently, $\ker\theta \subset \ker\varphi$, so $\varphi(x) = \varphi(y)$ and ψ if well defined. Clearly ψ is a homomorphism and surjective as both θ and φ are. □

Examples 17.10. Let $Q =< a, b \mid a^2 = b^2,\ a^4 = e,\ bab^{-1} = a^{-1} >$ be the quaternion group and G be the subgroup of $\mathbb{M}_2(\mathbb{C})$ generated by

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \text{ and } \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}.$$

We show $Q \cong G$. Using the Van Dyck's Theorem, we get a group epimorphism $Q \to G$ given by

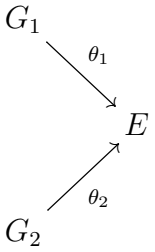
$$a \mapsto \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix} \quad \text{and} \quad b \mapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Check that $|G| \geq 8$. The relations for G imply that $ba = a^3b$ and $b^3 = a^2b$, so we must have $|Q| \leq 8$. It follows that $|Q| = 8$ and $G \cong Q$.

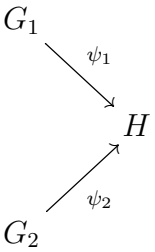
We end this section with a further useful example of universal properties.

Definition 17.11. Let G_1 and G_2 be two groups. A *free product* of G_1 and G_2 is a group E together with group monomorphisms

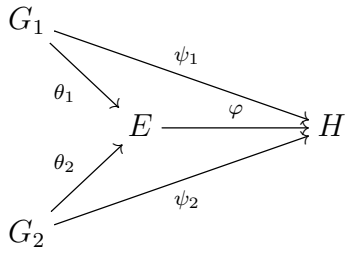
(*)



satisfying the following universal property: If H is a group and we group homomorphisms



then there exists a unique group homomorphism $\varphi : E \to H$ such that the diagram



commutes.

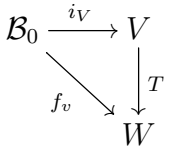
Theorem 17.12. *If G_1 and G_2 are groups, then a fiber product of G_1 and G_2 exists.*

In fact if we fix the monomorphisms in (*), then the fiber product is unique up to a unique isomorphism and it is usually denoted by $G_1 * G_2$.

We leave the proof as an exercise. It is quite similar to the existence of a free group defining an alphabet in this case as follows: Assume that $G_1 \setminus \{e_{G_1}\}$ and $G_2 \setminus \{e_{G_2}\}$ are disjoint. Let $\{e\}$ be a set disjoint from $(G_1 \setminus \{e_{G_1}\}) \cup (G_2 \setminus \{e_{G_2}\})$ and set $A = (G_1 \setminus \{e_{G_1}\}) \cup (G_2 \setminus \{e_{G_2}\}) \cup \{e\}$. Then define words in the obvious way, calling a non-empty word w reduced if $w = a_1 \cdots a_r$ with $a_i \in A \setminus \{e\}$ for $i = 1, \ldots, r$, some r and no adjacent letters lie in the same $G_i,\ i = 1, 2$. Now proceed as before.

Exercises 17.13. 1. Show that if a vector space V over F satisfies the universal property of vector spaces for a subset $\mathcal{B}$, then $\mathcal{B}$ is linearly independent. In particular, 0 cannot lie in $\mathcal{B}$ unless $V = 0$

2. Let V be a vector space over F. Suppose that $\mathcal{B}_0$ is a set and $i_V : \mathcal{B}_0 \to V$ a set map. Show if for every vector space W over F and set map $f : \mathcal{B}_0 \to W$ there exists a unique linear transformation $T : V \to W$ such that

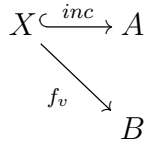


commutes, then i_V is an injective map and $i_V(\mathcal{B}_0)$ is a basis for V.

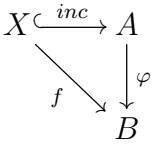
3. Prove the part of Claim 17.5 left undone.

4. Prove that a presentation for a group is unique up to isomorphism.

5. Define a *free abelian group* A on basis X to be an abelian group satisfying the following universal property of free abelian groups: Given any abelian group B and diagram



of set maps with *inc* the inclusion of X in A, then there exists a unique group homomorphism $\varphi : A \to B$ such that



commutes. Prove that a free abelian group on any non-empty set X exists and is unique up to isomorphism.

6. Show that $\mathbb{Z}$ is a free group on basis $\{1\}$.
7. Show that if G is a free group on basis $\mathcal{B}$ and $\mathcal{B}$ has more than two elements that G is not abelian.
8. Prove that free products exist.
9. Define the free product of a set of groups $\{G_i \mid i \in I\}$ and prove it exists. It is denoted by $*_I G_i$.
10. Show that a free group on a set X is isomorphic to $*_X \mathbb{Z}$.
11. Show that a free abelian group on a finite set X is isomorphic to $\bigtimes_X \mathbb{Z}$. (What if X is not finite?)

CHAPTER IV

Group Actions

In general, groups arise in mathematics not abstractly, but in a very concrete way. The simplest geometric example is given an object, what are all the symmetries of that object, e.g., if the object is a circle, a sphere, a tetrahedron or cube in $\mathbb{R}^3$? Of course, we also have already seen examples in the previous chapter. The dihedral group acts on a regular n-gon, the symmetric group acts on a set. These are all examples of a given set upon which a collection of maps acts on it bijectively. This set of maps of the object, can give us much deeper knowledge about that object. As our examples above are maps, the composition of bijections satisfies associativity and the identity map acts as the identity. This is already sufficient to give us our desired abstract formulation. In this chapter, we give many examples of this, and show how powerful an idea it is. Again we concentrate on applications to finite groups, since using the natural equivalence relation that arises allows us to count, resulting in many deep theorems about the structure of such groups.

18. The Orbit Decomposition Theorem

Groups are important because they act on objects. For example,

$\Sigma(S)$ acts on S
$\mathrm{GL}_n(F)$ acts on F^n
D_n acts on the vertices of a regular n-gon,

and these actions give information about the groups and the objects on which they act. We begin by defining how a group G acts on a nonempty set S. We then see how this leads to an equivalence relation $\sim_G$ on S and form the set of equivalence classes $S/\sim_G$, followed by the application of the Mantra of Equivalence Relations. To be useful, especially when $|S|$ is finite, we must compute the cardinality of each equivalence class. Unlike the case of cosets, we cannot expect the cardinality of each equivalence class to be the same.

We begin with the definition of an action.

Definition 18.1. Let G be a group and S a non-empty set. We say that S is a *(left) G-set* under $* : G \times S \to S$, writing $g * s$ for $*(g, s)$, if $*$ satisfies the following: For all $g_1, g_2 \in G$ and for all $s \in S$,

(1) $(g_1 \cdot g_2) * s = g_1 * (g_2 * s)$.
(2) $e * s = s$.

The map $* : G \times S \to S$ is then called a *G-action* on S. Note that the first property is just a generalization of associativity, the second that the unity acts like an identity.

Note. If S is a G-set and H a subgroup of G, then we can also view S as an H-set by *restriction*, i.e., by the H-action $\star : H \times S \to S$ given by $h \star s := h * s$.

We now define the equivalence relation. Let S be a G-set under $*$. Define $\sim_G$ on S by

$$s_1 \sim_G s_2 \text{ if there exists } g \in G \text{ satisfying } s_1 = g * s_2. \tag{18.2}$$

Lemma 18.3. *Let S be a G-set under $*$, then $\sim_G$ is an equivalence relation on S.*

Proof. **Reflexitivity**. For all $s \in S$, we have $s = e * s$, so $s \sim_G s$.

Symmetry. If $s_1 \sim_G s_2$, then there exists a $g \in G$ such that $s_1 = g * s_2$, hence

$$g^{-1} * s_1 = g^{-1} * (g * s_2) = (g^{-1} \cdot g) * s_2 = e * s_2 = s_2.$$

Note this shows

$$s_1 = g * s_2 \text{ if and only if } g^{-1} * s_1 = s_2. \tag{18.4}$$

Transitivity. Suppose that $s_1 \sim_G s_2$ and $s_2 \sim_G s_3$. Then there exist $g, g' \in G$ satisfying $s_1 = g * s_2$ and $s_2 = g' * s_3$. Hence $s_1 = g * s_2 = g * (g' * s_3) = (g \cdot g') * s_3$, so $s_1 \sim_G s_3$. □

Next, let S be a G-set, $s \in S$. The equivalence class of s via $\sim_G$ is the set

$$\begin{aligned} \overline{s} &:= \{s' \in S \mid \text{there exists } g \in G \text{ such that } s' = g * s\} \\ &= \{g * s \mid g \in G\}. \end{aligned}$$

This equivalence class is called the *orbit* of s under $*$ and denoted by $G * s$. We always let

$$\mathcal{O} \text{ be a system of representatives for the equivalence classes under } \sim_G .$$

(So $\mathcal{O}$ consists of precisely one element from each orbit — equivalence class.) We write $G\backslash S$ for $S/\sim_G = \{G * s \mid s \in \mathcal{O}\}$, the set of orbits

for the action $\sim_G$. Applying the Mantra of Equivalence Relations, we have

Mantra of G-actions. Let S be a G-set, $\mathcal{O}$ a system of representatives. Then

$$S = \bigvee_{\mathcal{O}} G * s \text{ and if } |S| \text{ is finite, then } |S| = \sum_{\mathcal{O}} |G * s|.$$

To make this useful, we must be able to compute the size of an orbit if finite. In general, orbits may have different cardinalities. We do this by attaching a group to each element of S in the following way. If S is a G-set, $s \in S$, define the *stabilizer* or *isotropy subgroup* of s by

$$G_s := \{x \in G \mid x * s = s\}.$$

Lemma 18.5. *Let S be a G-set, $s \in S$. Then G_s is a subgroup of G.*

PROOF. By definition of a G-action, $e \in G_s$, so G_s is non-empty. If $x, y \in G$, then $(xy) * s = x * (y * s) = x * s = s$ and if $y * s = s$, then $s = y^{-1} * s$, so G_s is a subgroup. □

We can now compute the cardinality of an orbit.

Proposition 18.6. *Let S be a G-set, $s \in S$. Define*

$$f_s : G/G_s \to G * s \text{ by } xG_s \mapsto x * s.$$

Then f_s is a well-defined bijection. In particular, if $[G : G_s]$ is finite, then

$$|G * s| = [G : G_s] \text{ and } |G * s| \text{ divides } |G|.$$

PROOF. Let $x, y \in G$, then

$$\begin{aligned} x * s = y * s &\text{ if and only if } y^{-1} * (x * s) = s \\ &\text{if and only if } (y^{-1}x) * s = s \\ &\text{if and only if } y^{-1}x \in G_s \\ &\text{if and only if } xG_s = yG_s. \end{aligned}$$

This shows that f_s is well-defined and one-to-one. As f_s is clearly surjective, the result follows. □

Example 18.7. Let S be the faces of a cube and G be the group of rotations of S. So G acts on S. Given any two faces s_1, s_2, there is an element of G taking s_1 to s_2. So there is one orbit under this action. We say that G acts *transitively* on S. If s is a face, then the isotropy subgroup G_s of s is the cyclic group C_4 of rotations of the face s about its center. By the proposition, we have

$$|G|/|C_4| = [G : G_s] = |G * s| = |S| = 6,$$

so $|G| = 24$. More generally, let S be the regular solid having n faces, each which has k edges (or vertices) and G the group of all rotations of S. Then the G-action is transitive and $|G| = nk$ by an analogous argument. It can be shown that there are only five such regular solids: the tetrahedron ($n = 4$, $k = 3$), the *cube*, the *octahedron* ($n = 8$, $k = 3$), the *dodecahedron* ($n = 12, k = 5$), and the *icosahedron* ($n = 20$, $k = 3$). So the corresponding rotation groups have 12, 24, 24, 60, 60 elements and are isomorphic to A_4, S_4, S_4, A_5, A_5 respectively. We give further details in Section §19.

Suppose that S is a G-set, $s \in S$. We say that $G * s$ is a *one point orbit* of S and s *is a fixed point* (under the action of G) if $G * s = \{s\}$. We set

$$F_G(S) := \{s \in S \mid |G * s| = 1\} \subset S,$$

the *set of fixed points* of S under the action of G.

If S is a G-set, a major problem is to determine if $F_G(S)$ is nonempty, and if it is, to compute $|F_G(S)|$.

The following equivalent conditions characterize the one point orbits. The proof is left as an easy exercise.

Lemma 18.8. *Let S be a G-set, $s \in S$. Then the following are equivalent:*

(1) $s \in F_G(S)$.
(2) $G_s = G$.
(3) $G * s = \{s\}$.

In particular, if $\mathcal{O}$ is a system of representative, then $F_G(S) \subset \mathcal{O}$.

If $\mathcal{O}$ is a system of representatives for a G-action on a set S, we always set

$$\mathcal{O}^* = \mathcal{O} \setminus F_G(S).$$

In particular, if $s \in \mathcal{O}^*$, we have $|G * s| = [G : G_s] > 1$, i.e., $G_s < G$. (Recall that the symbol $<$ when used for sets means a subset, but not the whole set.)

Putting this all together, we get the desired result:

Theorem 18.9. (Orbit Decomposition Theorem) *Let S be a G-set. Then*

$$S = F_G(S) \vee \bigvee_{\mathcal{O}^*} G * s.$$

In particular, if S is a finite set, then

$$|S| = \sum_{\mathcal{O}} |G * s| = |F_G(S)| + \sum_{\mathcal{O}^*} [G : G_s].$$

[Note *if $s \in \mathcal{O}^*$, then $1 < [G : G_s]$ and if G is finite then $[G : G_s] \mid |G|$.*]

PROOF. This follows from the Mantra as $|G * s| = [G : G_s]$. □

Exercises 18.10. 1. Let S be a G-set, $s_1, s_2 \in S$. Suppose that there exists an $x \in G$ such that $s_2 = x * s_1$, i.e., $s_2 \in G * s_1$. Show that $G_{s_2} = xG_{s_1}x^{-1}$.

2. Let S be a G-set. We say that the G-action is *transitive* if for all $s_1, s_2 \in S$, there exists a $g \in G$ satisfying $g * s_1 = s_2$, equivalently, $S = G * s$; and is *doubly transitive* if for all pairs of elements (s_1, s_1') and (s_2, s_2') in S with $s_1 \neq s_1'$ and $s_2 \neq s_2'$, there exists a $g \in G$ satisfying $g * s_1 = s_2$ and $g * s_1' = s_2'$. Suppose that S and G are finite. Show if the action is transitive then $|G| \geq |S|$ and if the action is doubly transitive then $|G| \geq |S|^2 - |S|$.

19. Addendum: Finite Rotation Groups in $\mathbb{R}^3$.

In this addendum, we give further details to Example 18.7. Recall if R is a commutative ring, then the *orthogonal group* is $\mathrm{O}_n(R) := \{A \in \mathrm{GL}_n(R) \mid AA^t = I\}$ and the *special orthogonal group* is $\mathrm{SO}_n(R) = \mathrm{O}_n(R) \cap \mathrm{SL}_n(R)$. In this section, we shall determine the finite subgroups of $\mathrm{SO}_3(\mathbb{R})$. Elements of $\mathrm{SO}_3(\mathbb{R})$ are very special as any element is, in fact, a rotation about some axis through the origin in $\mathbb{R}^3$. This is easy to see using linear algebra. Indeed let A be a nonidentity element of $\mathrm{SO}_3(\mathbb{R})$. Viewing $\mathbb{R}^3$ as an inner product space under the dot product $\cdot$ the matrix A is an isometry, i.e., A preserves dot products as $Av \cdot Aw = v \cdot A^tAw = v \cdot w$ for all v and w in $\mathbb{R}^3$ and also in $\mathbb{C}^3$. It follows that every eigenvalue ε of A in $\mathbb{C}$ satisfies $|\varepsilon| = 1$. In particular, ε is a root of unity, so is ± 1 if real. As the characteristic polynomial f_A of A is of odd degree, it has a real root, hence an eigenvalue 1 or -1. If all the roots of f_A are real, $\det A = 1$ forces A to be similar to $\mathrm{diag}(1, -1, -1)$, a rotation of angle π around some axis. If the roots of f_A are not all real, they must be $1, e^{2\pi\sqrt{-1}/n}, e^{-2\pi\sqrt{-1}/n}$, some n, and A is similar to the rotation $\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos 2\pi/n & \sin 2\pi/n \\ 0 & -\sin 2\pi/n & \cos 2\pi/n \end{pmatrix}$ about some axis.

We view $\mathrm{SO}_3(\mathbb{R})$ acting on the unit sphere S^2. So if A is a nonidentity element in $\mathrm{SO}_3(\mathbb{R})$, it is the rotation by some angle about some axis through the origin. This axis intersects S^2 in two points, that we call the *poles* of A. If A is of finite order, then the angle of rotation must be $2\pi/n$ for some integer n. Let G be a finite subgroup of $\mathrm{SO}_3(\mathbb{R})$ of order n, and set $P := \{p \mid p \text{ is a pole of some element } A \text{ in } G\}$, called the *set of poles* of G.

Lemma 19.1. *Let G be a finite subgroup of $\mathrm{SO}_3(\mathbb{R})$ of order n, and P the set of poles of G. Then P is a G-set under the natural action. In particular, if $p \in P$, $|G| = [G : G_p]|G_p|$ and $|G * p| = [G : G_p]$, where G_p is the isotropy subgroup of p and $G * p$ the orbit of p.*

Proof. Let p be a pole of A in G. If B is a matrix in G, then $(BAB^{-1})Bp = Bp$, so Bp is a pole of BAB^{-1}. □

Let G be a finite subgroup of $\mathrm{SO}_3(\mathbb{R})$ of order n and P the set of poles. If p is a pole, let $n_p = |G_p|$. As every nonidentity element of G has two poles and each nonidentity element in G_p has p as a pole, we have

$$2(n-1) = \sum_P (n_p - 1).$$

Let $G * p_1, \ldots, G * p_r$ be the (distinct) orbits of P and $n_i = n_{p_i}$, for $i = 1, \ldots, r$. As the order of stabilizers of elements in a fixed orbit are all the same, we have

$$2n - 2 = \sum_{i=1}^{r} (n_i - 1)|G * p_i|,$$

so dividing by $n = |G|$ implies that

$$2 - \frac{2}{n} = \sum_{i=1}^{r} (n_i - 1)\frac{|G * p_i|}{n} = \sum_{i=1}^{r} (n_i - 1)\frac{1}{|G_{p_i}|} = \sum_{i=1}^{r} (1 - \frac{1}{n_i}).$$

As $n > 1$ and $1 - (1/n_i) \geq 1/2$, we have $1 \leq 2 - (2/n) \leq 2$. It follows that we can only have $r = 1, 2$, or 3, i.e., there are at most three orbits. We investigate each possibility separately.

Case. $r = 1$:

This cannot occur, as $1 \leq 2 = (2/n) = 1 - (1/n_1)$ is impossible.

Case. $r = 2$:

As

$$2 - \frac{2}{n} = (1 - \frac{1}{n_1}) + (1 - \frac{1}{n_2}),$$

we have

$$\frac{2}{n} = \frac{1}{n_1} + \frac{1}{n_2}.$$

It follows that $n = n_1 = n_2$ and $|G * p_1| = |G * p_2| = 1$ as $n = n_i|G * p_i|$. In particular, $|P| = 2$. Since every nonzero element in G has two poles, we must have G is a cyclic group of finite order with axis connecting the two poles.

Case. $r = 3$:

In this case, we have

$$(*) \qquad \frac{2}{n} = \frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3},$$

and we may assume that $n_1 \leq n_2 \leq n_3$. We must find the groups with data (n_1, n_2, n_3). Let $[|G * p_1|, |G * p_2|, |G * p_3|] = [n/n_1, n/n_2, n/n_3]$ be the tuple of the sizes of the corresponding orbits. If $n_1 \geq 3$, then $2/n \leq 0$, which is impossible, so $n_1 = 2$, and

$$\frac{1}{2} + \frac{2}{n} = \frac{1}{n_2} + \frac{1}{n_2}.$$

This is impossible if $n_2 \geq 4$, so $n_2 = 2$ or 3.

Subcase. $r = 3, n_2 = 2$:

By (*), we must have $n_3 = n/2$. Therefore, G_{p_3} has two poles so is cyclic of order $n/2$. As P is a G-set and the nonzero element g in G_{p_1} must be a rotation of angle π, it follows that the poles of g must be perpendicular to the poles of the nonzero elements in $G * p_3$ (and reverse pole pairs). It follows that G must be the dihedral group $D_{\frac{n}{2}}$.

Therefore we may assume that $n_2 = 3$. If $n_3 \geq 6$, then $(1/2) + (1/3) + (1/n_3) \leq 0$, so $n_3 = 3, 4$, or 5.

Subcase. G has data $(2, 3, 3)$:

It follows by (*) that $n = 12$, so G also has orbit data $[6, 4, 4]$. As P is a G-group, if p lies in $G*p_3$ and q is the closest point to p in $G*p_2$, then $G_{p_3}q$ consists of at least three points equidistance from p, which form an equilateral triangle. It follows that there are four such triangles with the poles forming a regular tetrahedron. Therefore, we see that G is the rotation group of a regular tetrahedron and the orbit data $[6, 4, 4]$ is the number of edges, vertices, and faces of the tetrahedron, respectively. (Cf. Example 18.7.) This group can be shown to be isomorphic to A_4 using Exercise 22.22(5) below.

Subcase. G has data $(2, 3, 3)$:

It follows by (*) that $n = 24$, so G also has orbit data $[12, 8, 6]$. In this case, if p lies in $G * p_3$ and q is the closest point to p in $G * p_2$, then $G_{p_3}q$ consists of at least four points equidistance from p, which form a square, and we see that G is the group of rotations of a cube with orbit data $[12, 8, 6]$, the number of edges, vertices, and faces of a cube, respectively. (Cf. Example 18.7.) This group can be shown to be isomorphic to S_4 using Exercise 22.22(5) below.

Subcase. G has data $(2, 3, 5)$:

It follows by (*) that $n = 60$, so G also has orbit data $[30, 20, 12]$. In this case, if p lies in $G * p_3$ and q is the closest point to p in $G * p_2$,

then $G_{p_3}q$ consists of at least five points equidistance from p, which form a pentagon, and we see that G is the group of rotations of a regular icosahedron with orbit data $[30, 20, 12]$, the number of edges, vertices, and faces of a regular icosahedron, respectively. (Cf. Example 18.7.) This group can be shown to be isomorphic to A_5 using Exercise 22.22(5) below.

20. Examples of Group Actions

In this section, to demonstrate the usefulness of this concept, we give many examples of group actions. As this is the most important concept in our study of groups, there are many exercises at the end of this section. To understand this topic, one should do many of them. In the following two sections, we shall give further specific applications of some of examples given in this section.

Example 20.1. Conjugation on Elements.

In this example, we let the G-set S be G itself. The (left) action is given by

$$* : G \times S \to S \quad \text{by} \quad g * s = gsg^{-1}$$

called *conjugation* by G. The orbit of an element $a \in S = G$ is

$$C(a) := G * a = \{xax^{-1} \mid x \in G\},$$

called the *conjugacy class* of a, and the stabilizer of a is

$$Z_G(a) := G_a = \{x \in G \mid xax^{-1} = a\} = \{x \in G \mid xa = ax\}.$$

This subgroup of G is called the *centralizer* of a. So $Z_G(a)$ is the set of elements of G commuting with a. In particular, $\langle a \rangle \subset Z_G(a)$. The set of fixed points of this action is

$$\begin{aligned} F_G(S) &= \{a \in S \mid xax^{-1} = a \text{ for all } x \in G\} \\ &= \{a \in G \mid xa = ax \text{ for all } x \in G\}, \end{aligned} \tag{20.2}$$

the *center* $Z(G)$ of G. Recall it is a normal subgroup of G. Of course, in general, a G-set is not a group, so one cannot expect the set of fixed points to have an algebraic structure.

Let $\mathcal{C}$ be a system of representatives for the conjugation action of G on G. So $\mathcal{C}^* = \mathcal{C} \setminus Z(G)$, and the Mantra of G-actions becomes

$$G = Z(G) \vee \bigvee_{\mathcal{C}^*} C(a),$$

with

$$|G| = |Z(G)| + \sum_{\mathcal{C}^*} [G : Z_G(a)], \tag{20.3}$$

if G is finite. Equation 20.3 is called the *class equation*. We give an interesting application of the class equation. If G is a finite group of order p^n, for some (positive) prime p and $n > 0$, we call G a *p-group*. So a subgroup of a p-group is either a p-group or the trivial group.

Application 20.4. If G is a p-group, then $1 < Z(G) \subset G$.
(Recall this means $1 \neq Z(G)$.) In particular, if G is a p-group of order p^n, with $n > 1$, then G is not simple:

We know if $a \in \mathcal{C}^*$, then $p \mid [G : Z_G(a)]$, so

$$0 \equiv |G| = |Z(G)| + \sum_{\mathcal{O}^*} [G : Z_G(a)] \equiv |Z(G)| \mod p.$$

Since $e \in Z(G)$, we must have $|Z(G)| \geq 1$, so $p \mid |Z(G)|$ and $|Z(G)| \geq p$. In particular, $1 < |Z(G| \subset G$. If G is not abelian, then $1 < Z(G) < G$, and G is not simple, as $Z(G) \triangleleft G$. If G is abelian, then $G = Z(G)$ and there exists an element a in G of order p by Exercise 9.12(2). As $|G| > p$, we have $1 < \langle a \rangle < G$ with $\langle a \rangle \triangleleft G$.

Useful Observation. If we are trying to prove a property about groups and G is a finite non-abelian group with $Z(G) > 1$ (which in general may not occur), then $G/Z(G)$ is a group of order less than that of G and the canonical epimorphism $^- : G \to G/Z(G)$ may allow us to apply induction.

Computation 20.5. Let $G = D_3 = \{e, r, r^2, f, fr, rf\}$, with $|G| = 6$ and satisfying $r^3 = e = f^2$ and $frf^{-1} = r^{-1} = r^2$. We have $fr = r^2 f$ and $rf = fr^2$, so

$$\begin{aligned} C(e) &= \{e\} && \text{and} \quad 1 \mid 6 \\ C(r) &= \{r, r^2\} && \text{and} \quad 2 \mid 6 \\ C(f) &= \{f, fr, rf\} && \text{and} \quad 3 \mid 6 \end{aligned}$$

and

$$\begin{aligned} Z_G(e) &= G && \text{and} \quad |C(e)| = [G : Z_G(e)] = 1 \\ Z_G(r) &= \{e, r, r^2\} && \text{and} \quad |C(r)| = [G : Z_G(r)] = 2 \\ Z_G(f) &= \{e, f\} && \text{and} \quad |C(f)| = [G : Z_G(r)] = 3 \end{aligned}$$

with fixed points

$$Z(G) = \{e\},$$

so

$$\begin{array}{ccccccc} |G| & = & |Z(G)| & + & |C(r)| & + & |C(f)| \\ 6 & = & 1 & + & 2 & + & 3. \end{array}$$

Note that the conjugacy classes (equivalence classes), $C(r) = C(r^2)$ and $C(f) = C(fr) = C(rf)$. This application leads to the following theorem.

Theorem 20.6. *Let G be a group of order p^2 with p a prime. Then G is abelian.*

PROOF. Suppose that G is not abelian, then $Z(G) < G$ by Exercise 20.26(1). By Application 20.4, we have $1 < Z(G)$; so by Lagrange's Theorem, we must have $|Z(G)| = p$. Let $a \in G \setminus Z(G)$. Then we have $Z(G) < Z_G(a)$, as $a \in Z_G(a)$. But then the subgroup $Z_G(a)$ must satisfy $|Z_G(a)| = p^2 = |G|$ by Lagrange's Theorem, hence $Z_G(a) = G$. This implies that $a \in Z(G)$, a contradiction. □

Remarks 20.7. Let G be a p-group.

1. If $G = p^2$, the group G is abelian, but may not be cyclic. If G is cyclic, then $G \cong \mathbb{Z}/p^2\mathbb{Z}$. However, the group $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ is not cyclic.
2. D_4 is a 2-group of order 2^3, but not abelian.

Later we shall prove a major theorem. It will imply that every finitely generated abelian group, which includes all finite abelian groups, is a direct product of cyclic groups. There is also a uniqueness statement. At present, we only can show that abelian groups of order pq, with p, q (not necessarily distinct) primes are of this form.

We now jazz up Example 20.1 by letting G act on a subgroup H by conjugation. For this to work, however, we must have $xHx^{-1} \subset H$ for all $x \in G$, i.e, we must have $H \triangleleft G$. So suppose that $H \triangleleft G$. Let G act on H by conjugation, i.e.,

$$* : G \times H \to H \text{ is given by } g * h = ghg^{-1}.$$

We still have $C(a) = G * a$ for all $a \in H$, so if H is a finite group, we have

$$|H| = |F_G(H)| + \sum_{\mathcal{O}^*} [G : Z_G(a)],$$

as $Z_G(a) = G_a := \{x \in G \mid xax^{-1} = a\}$, with $a \in \mathcal{O}$, where $\mathcal{O}$ is a system of representatives for the G-action on H. An application of this is:

Proposition 20.8. *Let G be a p-group and N a non-trivial normal subgroup of G. Then there exists an non-identity element $x \in N$ such that $xy = yx$ for all $y \in G$, i.e., $1 < N \cap Z(G)$.*
[Cf. the case $N = G$ with Application 20.4.]

PROOF. Let G act on N by conjugation. Every subgroup of G is either 1 or a p-group, hence

$$0 \equiv |N| = |F_G(N)| + \sum_{\mathcal{O}^*} [G : Z_G(a)] \equiv |F_G(N)| \mod p$$

by the Orbit Decomposition Theorem. As $xex^{-1} = e$ for all $x \in G$, the unity e is a fixed point, so $p \mid |F_G(N)| \geq 1$. Therefore, $|F_G(N)| \geq p$. The result now follows from

$$\begin{aligned} F_G(N) &= \{z \in N \mid xzx^{-1} = z \text{ for all } x \in G\} \\ &= \{z \in N \mid xz = zx \text{ for all } x \in G\} = N \cap Z(G). \end{aligned}$$ □

We can even generalize this conjugation type action a bit further as follows: Let H be a subgroup of G. Define the *normalizer* of H in G by

$$N_G(H) := \{x \in G \mid xHx^{-1} = H\}.$$

The normalizer has the following properties:

Properties 20.9. Let H be a subgroup of G, then the following are true:

1. $N_G(H)$ is a subgroup of G.
2. $H \lhd N_G(H)$.
3. If $H \lhd K$ with K a subgroup of G, then $K \subset N_G(H)$.
4. $N_G(H)$ is the unique largest subgroup of G containing H as a normal subgroup.

Replacing G by $N_G(H)$, we can let $N_G(H)$ act on H by conjugation, by the above, i.e., we have an $N_G(H)$-action

$$* : N_G(H) \times H \to H \text{ given by } x * h = xhx^{-1}.$$

In fact, we can go one step further. Let K be a subgroup of G and set $A = K \cap N_G(H)$. Then the action

$$* : A \times H \to H \text{ given by } x * h = xhx^{-1}$$

makes H into an A-set. Instead of pursuing this, we generalize conjugation in another direction.

Example 20.10. Conjugation on Sets.

Let $S = \mathcal{P}(G) := \{A \mid A \subset G, \text{ a subset}\}$, the *power set* of G. Then S becomes a G-set by

$$* : G \times S \to S \text{ given by } x * A = xAx^{-1}$$

where $xAx^{-1} = \{xax^{-1} \mid a \in A\}$. Note that $|A| = |xAx^{-1}|$ for all $x \in G$. We also call this action *conjugation*. The orbit of $A \in \mathcal{P}(G)$ is the *conjugacy class*

$$C(A) = G * A = \{xAx^{-1} \mid x \in G\},$$

and the stabilizer of A is

$$N_G(A) := G_A = \{x \in G \mid xAx^{-1} = A\},$$

called the *normalizer* of A in G. The Orbit Decomposition Theorem yields

$$S = \mathcal{P}(G) = \bigvee_{\mathcal{O}} C(A) = F_G(S) \vee \bigvee_{\mathcal{O}^*} C(A).$$

If G is finite so is $\mathcal{P}(G)$, and we would then have

$$2^{|G|} = |S| = |F_G(S)| + \sum_{\mathcal{O}^*} [G : N_G(A)].$$

This is not so useful, so we try to cut S down. We use the following:

Observation 20.11. Let S be a non-empty set and $\sim$ an equivalence relation of S. Suppose a subset $\emptyset \neq T \subset S$ satisfies the following:

$$\text{For all } \ a \in T, \ \text{ we have } [a]_\sim \subset T.$$

Then $\sim|_T$ (actually $\sim|_{T\times T}$) defines an equivalence relation on T and

$$T/\sim_T = \{[a]_\sim \mid a \in T\}.$$

[We are just saying: If $\mathcal{C}$ partitions S and $\emptyset \neq \mathcal{D} \subset \mathcal{C}$, then $\mathcal{D}$ partitions $\bigcup_{\mathcal{D}} \overline{a}$.]

There are many interesting subsets of $\mathcal{P}(G)$. Recall that if H is a subgroup of G, then so is xHx^{-1} for all $x \in G$.

Specific Examples 20.12. Let G be a group. Then the following subsets of $\mathcal{P}(G)$ are G-sets by conjugation (i.e., restricting the action above) if non-empty:

(1) $\mathcal{P}(G)$.
(2) $\mathcal{G} := \{H \mid H \subset G \text{ a subgroup }\}$.
(3) $\mathcal{T}_n := \{A \mid A \subset G \text{ with } |A| = n\}$.
(4) $\mathcal{G} \cap \mathcal{T}_n$.
(5) If G is a finite group, p a (positive) prime satisfying $p^r || |G|$, with $r > 0$, so $G = p^r m$ with p and m relatively prime, then
$$Syl_p(G) := \mathcal{G} \cap \mathcal{T}_{p^r} = \{H \mid H \text{ a subgroup of } G \text{ of order } p^r\}.$$
(6) If $H \in \mathcal{G}$ then $C(H)$.
(7) If $A \in \mathcal{P}(G)$ then $C(A)$.

We can also push this further by restricting the action of G on one of these sets to the action by a subgroup. We shall later study this action in greater detail, especially (5) and (6).

Example 20.13. Translation Action.

Let G be a group, then the power set $\mathcal{P}(G)$ is a G-set via

$$* : G \times \mathcal{P}(G) \to \mathcal{P}(G) \text{ defined by } g * T = gT := \{gx \mid x \in T\}$$

called *translation*. As before, this is more interesting when we replace $\mathcal{P}(G)$ by appropriate subsets. For example, using the notation of (20.12), G acts on $\mathcal{T}_n$ by translation.

Warning 20.14. G does not act on $\mathcal{G}$ by translation as $H \in \mathcal{G}$ does not mean that $x * H \in \mathcal{G}$. In fact, $x * H$ would lie in $\mathcal{G}$ if and only if $x \in H$.

One also can translate other sets by G. For example, if H is a subgroup of G, we can let G act on the cosets G/H by translation, i.e., let

$$* : G \times G/H \to G/H \text{ be given by } x * aH = xaH.$$

We have used this action before, viz., in the proof of the General Cayley Theorem. The stabilizer of aH in G/H under translation by G is given by:

$$\begin{aligned} G_{aH} &= \{x \in G \mid xaH = aH\} = \{x \in G \mid a^{-1}xaH = H\} \\ &\{x \in G \mid a^{-1}xa \in H\} = \{x \in G \mid x \in aHa^{-1}\} \\ &= aHa^{-1}. \end{aligned}$$

Note that if $H < G$ in the above then $F_G(G/H) = \emptyset$ as $G = G_{aH}$ would imply that $G = aHa^{-1}$.

The following proposition is a useful computing devise. We leave its proof as an exercise.

Proposition 20.15. *Let G be a finite group, H a subgroup of G. Suppose that H is a p-group and $p \mid [G : H]$. Then $p \mid [N_G(H) : H]$. In particular, $H < N_G(H)$. Consequently, if G is a finite p-group and $H < G$ a subgroup, then $H < N_G(H)$.*

We finish the discussion of translation actions to answer our outstanding question on how cosets arise naturally. Indeed, if H is a subgroup of G, and G is a *right H-set* by the *right H-action* by translation on G, i.e.

$$* : G \times H \to G \text{ given by } g * h = gh,$$

then the orbit $g * H$ is the left coset gH and the Orbit Decomposition Theorem in this case is just Lagrange's Theorem.

Example 20.16. Endomorphic Action.

Let V be a vector space over a field F. A linear transformation $T : V \to V$ is called an *endomorphism* of V. The set

$$\mathrm{End}_F(V) := \{T \mid T : V \to V \text{ a linear transformation}\}$$

is a ring under the usual addition and composition of functions. (From linear algebra, you should know the relationship between $\mathrm{End}_F(V)$ and $\mathbb{M}_n(F)$ if V is n-dimensional.) Let G be $\mathrm{End}_F(V)$ viewed as an additive group. Then we have V a G-set by

$$* : G \times V \to V \text{ given by } T * v = T(v)$$

called *evaluation*. If $V = F^n$, we can replace $\mathrm{End}_F(V)$ by $\mathbb{M}_n(F)$.

Much more interesting is the next example.

Example 20.17. Automorphic Action.

Let V be a vector space over a field F and $G = \mathrm{Aut}_F(V) = (\mathrm{End}_F(V))^\times$. Then V is a G-set via the *evaluation*

$$* : G \times V \to V \text{ given by } T * v = T(v).$$

Example 20.18. Addendum to Example 20.1.

Let R be any ring, $S = \mathbb{M}_n(R)$, and $G = \mathrm{GL}_n(R)$. Then S is a G-set under conjugation, i.e.,

$$* : G \times S \to S \text{ is given by } A * B = ABA^{-1}.$$

This is an important action and leads to the following:

Problem 20.19. Let F be a field and $*$ the action above. Find a nice system of representatives for this action.

If every non-constant polynomial with coefficients in the field F has a root in F, we say that F is *algebraically closed*, e.g., the complex numbers are algebraically closed by the *Fundamental Theorem of Algebra* (to be proven later). The above sought after system of representatives is called the set of *Jordan canonical forms*. For a general field F, the sought after set is the set of *rational canonical forms*. Proving this will be a major goal later.
[We have another action by

$$* : G \times S \to S \text{ given by } A * B = ABA^t.$$

and ask the same questions. This is especially interesting if $S = \mathbb{M}_n(F)$ is replaced by the subset of symmetric matrices or the subset of skew symmetric matrices.]

Example 20.20. Evaluation.

We have previously seen another example of an action given by evaluation, viz., if S is a non-empty set, then S is a $\Sigma(S)$-set by the evaluation map

$$* : \Sigma(S) \times S \to S \text{ given by } \gamma * s = \gamma(s).$$

If H is a subgroup, we can restrict the action to H. For example, let $\gamma \in \Sigma(S)$ and $\Gamma = \langle \gamma \rangle$. Then S becomes a Γ-set by the action $* : \Gamma \times S \to S$ given by the *evaluation*

$$\gamma^i * s = \gamma^i(s) \text{ for all integers } i.$$

We shall use this action later.

It should be mentioned that evaluation actions are used all the time. If you have an object, then its *automorphism group* (definition?) acts on it by evaluation. For example, such objects can be algebraic as above, topological, or geometric.

Example 20.21. Pullback Action.

As mentioned before a G-action on a set S can be restricted to the action of any subgroup of G. In the examples above, we have used this. We now generalize this. Recall that a subgroup of a group is just a group in which the inclusion map is a monomorphism. More generally, suppose that we have a group homomorphism $\theta : G \to G'$ and a G'-set S defined by $* : G' \times S \to S$. Then S becomes a G-set by the *pullback action* defined by

$$\star : G \times S \to S \text{ is given by } g \star s = \theta(g) * s.$$

Example 20.22. Shift Action.

Our last example, is an action that leads to a nice result called Cauchy's Theorem, which is the first step toward showing that finite groups have subgroups of the order of the power of a prime for any power of that prime dividing the order of the group. Let G be a finite group and p a (positive) prime dividing the order of G. Set

$$S = \{(g_1, \ldots, g_p) \in G^p \mid g_1 \cdots g_p = e \text{ in } G\} \subset G^p,$$

where $G^p = G \times \cdots \times G$ (p times). As $(e, \ldots, e) \in S$, the set S is non-empty. Moreover, it follows immediately that

$$(g_1, \ldots, g_p) \in S \qquad \Longrightarrow \qquad g_p = (g_1 \cdots g_{p-1})^{-1} \text{ in } G.$$

Therefore, for all ordered $(p-1)$-tuples $g_1, \ldots, g_{p-1}$, with $g_i \in G$, we have a unique g_p such that $(g_1, \ldots, g_p)$ lies in S. Therefore, $|S| = |G^{p-1}|$, so $p \mid |S|$. If $y \in \mathbb{Z}$, let $\tilde{y} \in [y]_p = \overline{y}$ denote the smallest positive integer in $\overline{y}$. With this notation, if $x \in \mathbb{Z}$, the ordered tuple

$$\widetilde{1+x}, \ldots, \widetilde{p+x}$$

will be a (cyclic) permutation of $1, \ldots, p$. Let $(\mathbb{Z}/p\mathbb{Z}, +)$ act on S by

$$(*) \qquad \overline{x} * (g_1, \ldots, g_p) = (g_{\widetilde{1+x}}, \ldots, g_{\widetilde{p+x}})$$

i.e., if $1 \leq i, x \leq p$, then the action of $\overline{x}$ shifts i by x slots to the number $x+i$ until it hits p which it then maps to 1, etc. For example, if $p = 5$, then $\overline{3} * (g_1, g_2, g_3, g_4, g_5) = (g_4, g_5, g_1, g_2, g_3)$. The right hand side of $(*)$ still lies in S, since $g_1 \cdots g_p = e$ implies $g_{i+1} \cdots g_p g_1 \cdot g_i = e$ by conjugating successively by $g_1^{-1}, \ldots, g_i^{-1}$ $(1 \leq i < p)$. It follows easily that S is a $\mathbb{Z}/p\mathbb{Z}$-set under this action. With this action, we can now prove:

Theorem 20.23. (Cauchy's Theorem) *Let p be a (positive) prime dividing the order of the finite group G. Then there exists an element of G of order p.*

Proof. Let S be the $\mathbb{Z}/p\mathbb{Z}$-group under the shift action of Example 20.22. Apply the Orbit Decomposition Theorem 18.9 to get

$$|G^p| = |S| = |F_{\mathbb{Z}/p\mathbb{Z}}(S)| + \sum_{\mathcal{O}^*} [\mathbb{Z}/p\mathbb{Z} : (\mathbb{Z}/p\mathbb{Z})_x].$$

If $x \in \mathcal{O}^*$, then $(\mathbb{Z}/p\mathbb{Z})_x = 1$ as $\mathbb{Z}/p\mathbb{Z}$ has only the trivial subgroups. So

$$0 \equiv |S| \equiv |F_{\mathbb{Z}/p\mathbb{Z}}(S)| \mod p.$$

As $(e, \ldots, e) \in F_{\mathbb{Z}/p\mathbb{Z}}(S)$, we have $|F_{\mathbb{Z}/p\mathbb{Z}}(S)| \geq p > 1$. Thus there exists an element $(e, \ldots, e) \neq (g_1, \ldots, g_p) \in F_{\mathbb{Z}/p\mathbb{Z}}(S)$. But this means that

$$\begin{aligned}(g_1, \ldots, g_p) = (g_2, \ldots, g_p, g_1) = \cdots \\ = (g_{i+1}, \ldots, g_p, \ldots, g_i) = \cdots = (g_p, g_1, \ldots, g_{p-1})\end{aligned}$$

lies in G^p. It follows that $e \neq g_1 = \cdots = g_p$. Consequently, if $g = g_1$, then $(g, \ldots, g) \in S$, which means that $g^p = e$ as needed. □

Corollary 20.24. *Let p be a prime and G a non-trivial finite group. If every non-identity element in G has order a power of p, then G is a p-group.*

Corollary 20.25. *Let p and q be primes and G of order pq. Then G is not simple.*

PROOF. If $p = q$, then G is abelian hence not simple as G is not of prime order. If $p < q$, then G contains a normal group of order q by Useful Counting 12.7.

□

Exercises 20.26. 1. Let G be a group. Show all of the following:
(a) $Z(G) = \bigcap_{a \in G} Z_G(a)$.
(b) $a \in Z(G)$ if and only if $C(a) = \{a\}$ if and only if $|C(a)| = 1$.
(c) $a \in Z(G)$ if and only if $G = Z_G(a)$.
(d) If G is finite, then $a \in Z(G)$ if and only if $|Z_G(a)| = |G|$.

2. Let G be a group and $k(G)$ the number of conjugacy classes in G. Suppose that G is finite. Show that $k(G) = 3$ if and only if G is isomorphic to the cyclic group of order three or the symmetric group on three letters.
3. Let G be a group of order p^2 with p a prime. Show that either $G \cong \mathbb{Z}/p^2\mathbb{Z}$ or $G \cong \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$.
4. Show that Properties 20.9 are valid.
5. Compute all the conjugacy classes and isotropy subgroups in A_4.
6. Let G be a group of order p^n, p a prime. Suppose the center of G has order at least p^{n-1}. Show that G is abelian.
7. Let G be a p-group. Using Exercise 16.14(8), show that G is a nilpotent group. In particular a finite product of p-groups for various p is nilpotent.
8. Let G be a nilpotent group (cf. 16.14(8) and H a proper subgroup pf G. Show that $H \neq N_G(H)$.
9. Suppose that H is a proper subgroup of a finite group G. Show that $G \neq \bigcup_{g \in G} gHg^{-1}$. [This may not be true if G is infinite.]
10. Let H be a subgroup of G. Let H act on G/H by translation. Compute the orbits, stabilizers, and fixed points of this action.
11. Prove Proposition 20.15. [Hint use the previous exercise.]
12. Let G be a group and S a non-empty set. Show
(a) If $* : G \times S \to S$ is a G-action, then $\varphi : G \to \Sigma(S)$ given by $\varphi(x)(s) = x * s$ (i.e., if we let $\varphi_x = \varphi(x)$ then $\varphi_x(s) = x * s$) is a group homomorphism (called the *permutation representation*).
(b) If $\varphi : G \to \Sigma(S)$ is a group homomorphism then $* : G \times S \to S$ given by $x * s = \varphi_x(s)$ where $\varphi_x = \varphi(x)$ is a G-action.
13. Let G be a group of order pq with $q \leq p$ primes. Show that G is not simple and has a normal subgroup of order p.
14. Let G be a finite p-group. Show if $p^n \,|\, |G|$ then G has a normal subgroup of order p^n.

15. Let p be an odd prime. Classify all groups G of order $2p$ up to isomorphism.
16. Let S be a G-set. Suppose that both S and G are finite. If $x \in G$ define the *fixed point set* of x in G by

$$F_S(x) := \{s \in S \mid x \cdot s = s\}.$$

Show the number of orbits N of this action satisfies

$$N = \frac{1}{|G|} \sum_G |F_S(x)|.$$

[So N is the average of the size of the fixed point sets of elements of G.]

21. Sylow Theorems

Let G be a finite group of order of order $p^r m$ with p a prime relatively prime to m and $r > 0$. A subgroup H of G is called a *Sylow p-subgroup* of G if H has order p^r. Let

$$Syl_p(G) := \{H \mid H \text{ a Sylow } p\text{-subgroup of } G\}.$$

The natural question is answered by the next result. Its proof uses much of what we have done.

Theorem 21.1. (First Sylow Theorem) *Let p be a (positive) prime dividing the order of a finite group G. Then $Syl_p(G)$ is non-empty, i.e., there exists a Sylow p-subgroup.*

We prove the stronger

Theorem 21.2. (Generalized First Sylow Theorem) *Let p be a (positive) prime such that p^s, with $s \geq 0$, divides the order of a finite group G. Then there exists a subgroup of G of order p^s.*

Proof. We induct on the order of G. If $|G| \leq p$, the result is trivial, so we may assume that $s > 1$. We make the following:

Induction Hypothesis. If T is any finite group satisfying $|T| < |G|$ and $p^s \mid |T|$, then T contains a subgroup of order p^s.

We must show that G contains a subgroup of order p^s. If $H < G$ is a subgroup of G satisfying $p \not| [G : H]$, then $p^s \mid |H|$. By the Induction Hypothesis, H, hence G, contains a subgroup of order p^s. Therefore, we are done unless we make the following:

Assumption. If $1 < H < G$ is a subgroup, then $p \mid [G : H]$.

For example, if $a \in G$ satisfies $Z_G(a) < G$, then $p \mid [G : Z_G(a)]$. In particular, this is true for any $a \notin Z(G)$, i.e., $a \in \mathcal{C}^* = \mathcal{C} \setminus Z(G)$,

where $\mathcal{C}$ is a system of representatives for conjugate action of G on G in Example 20.1. Applying the Class Equation 20.3, we have

$$0 \equiv |G| = |Z(G)| + \sum_{\mathcal{C}^*} [G : Z_G(a)] \equiv |Z(G)| \mod p,$$

so $p \mid |Z(G)|$. By Cauchy's Theorem 20.23, there exists an element $a \in Z(G)$ of order p. Let $H = \langle a \rangle$, a subgroup of G of order p. Since $a \in Z(G)$, we have $xa^i x^{-1} = a^i$ for all $x \in G$ and $i \in \mathbb{Z}$. In particular, $H \triangleleft G$ (cf. Example 11.5(3)) and G/H is a group of order $[G : H] = |G|/|H| = |G|/p < |G|$. Consequently, $p^{s-1} \mid |G/H|$; and, by the Induction Hypothesis, there exists a subgroup $T \subset G/H$ of order p^{s-1}. Let $^- : G \to G/H$ be the canonical epimorphism. Then H is the kernel and by the Correspondence Principle, there exists a subgroup $\widetilde{T}$ of G containing H and satisfying $T = \widetilde{T}/H$. Hence $|\widetilde{T}| = |T||H| = p^s$, and $\widetilde{T}$ works. □

To investigate the set of Sylow p-subgroups, $Syl_p(G)$ of G, we need three lemmas.

Lemma 21.3. *Let H and K be subgroups of a finite group G. Then $|HK| = |H||K|/|H \cap K|$.*

Proof. This is Exercise 13.6(3). □

The main mathematical reason for the validity of the further Sylow theorems below is the following:

Lemma 21.4. *Let G be a finite group, P a Sylow p-subgroup, and H a subgroup of G that is also a p-group. If H is a subgroup of $N_G(P)$, i.e., $hPh^{-1} = P$ for all $h \in H$, then H is a subgroup of P. In particular, if, in addition, H is also a Sylow p-subgroup of G, then $H = P$.*

Proof. The key to the proof of this lemma is the Second Isomorphism Theorem together with the counting version of the Second Isomorphism Theorem of Lemma 21.3. By definition, $P \triangleleft N_G(P)$ and by hypothesis $H \subset N_G(P)$. Hence, by the Second Isomorphism Theorem, we have HP is a subgroup of $N_G(P)$, with $P \triangleleft HP$ and $H \cap P \triangleleft H$, and satisfying $HP/P \cong H/H \cap P$. In particular, by Lemma 21.3, we have

$$(*) \qquad |HP| = |P| \frac{|H|}{|H \cap P|}.$$

As P, HP, and $H/H \cap P$ are all p-groups, (*) implies that HP is a p-group satisfying $P \subset HP \subset G$. Thus HP is a Sylow p-subgroup of G, i.e., $|P| = |HP|$, so $P = HP$, hence $H \subset P$. If, in addition, $|H| = |P|$ then $H = P$. □

We convert the lemma above into the language of group actions with the action given by conjugation on sets. Explicitly, we use Specific Examples 20.12(6) after restricting the action of G to a subgroup.

Lemma 21.5. *Let P be a Sylow p-subgroup of a finite group G and H a subgroup of G that is also a p-group. Then the following hold:*

(1) $C(P)$ consists of Sylow p-subgroups of G.

(2) The conjugacy class $C(P)$ is an H-set by conjugation, i.e., via the action $* : H \times C(P) \to C(P)$ by conjugation.

(3) Suppose that T is a fixed point under the H-action in (2), i.e.,

$$T \in F_H\big(C(P)\big) = \{W \in C(P) \mid xWx^{-1} = W \text{ for all } x \in H\}.$$

Then $H \subset T$.

(4) If H in (3) is a Sylow p-subgroup, then, under the H-action in (2), H is the only possible fixed point, i.e., $F_H\big(C(P)\big) \subset \{H\}$.

Proof. (1): As $|P| = |xPx^{-1}|$ for all x in G, this is clear.

(2) is immediate.

(3): As T is an H-fixed point, the orbit $H * T = \{T\}$, so $xTx^{-1} = T$ for all $x \in H$. By definition, this means that $H \subset N_G(T)$. By Lemma 21.4 and (1), we conclude that $H \subset T$.

(4): If $T \in F_H\big(C(P)\big)$, then $H \subset T$ by (3). If further, H is a Sylow p-subgroup, then by (3), we have $H = T$. □

Note. Under the hypothesis of Lemma 21.5, we have the isotropy subgroup of P is given by $H_P = N_G(P) \cap H$ for any Sylow p-subgroup of G. But P is then also a Sylow p-subgroup of $N_G(P)$, so by restricting the H-action to H_P the proof of Lemma 21.5 shows that $N_G(P) \cap H \subset P$.

Theorem 21.6. (Sylow Theorems). *Let G be a finite group, p a (positive) prime dividing the order of G. Then the following are true:*

First Sylow Theorem. *There exists a Sylow p-subgroup of G.*

Second Sylow Theorem. *All Sylow p-subgroups are conjugate, i.e., if P is a Sylow p-subgroup of G, then $C(P) = Syl_p(G)$.*

Third Sylow Theorem. *Let P be a Sylow p-subgroup of G. Then the following are true:*

(i) $|Syl_p(G)| = [G : N_G(P)]$.

(ii) $|Syl_p(G)| \mid |G|$.

(iii) $|Syl_p(G)| \equiv 1 \mod p$.

(iv) $|Syl_p(G)| \mid [G : P]$, *i.e., if* $|G| = p^n m$ *with* $p \nmid m$, *then* $|Syl_p(G)| \mid m$.

Fourth Theorem. *Let H be a subgroup of G that is a p-group. Then H lies in some Sylow p-subgroup of G.*

PROOF. Let H be a subgroup of G that is a p-group. We have already proven the First Sylow Theorem, so let P be a Sylow p-subgroup of G. By Lemma 21.5, we have $C(P) \subset Syl_p(G)$ is an H-set under conjugation. By the Orbit Decomposition Theorem, we have

$$|C(P)| = |F_H\big(C(P)\big)| + \sum_{\mathcal{O}^*}[H : H_T].$$

If $T \in \mathcal{O}^*$, then $H_T < H$, so $p \mid [H : H_T]$, hence

$$(\dagger) \qquad |C(P)| \equiv |F_H\big(C(P)\big)| \mod p.$$

Note. Equation (†) is interesting, since H does not appear at all on the left hand side, i.e., (†) is true for all p-subgroups of H. To evaluate a constant function (of such p-subgroups of H) given by (†), it suffices to evaluate at any such H.

Case 1. Let $H = P$:

By Lemma 21.5, we conclude that $F_P\big(C(P)\big) \subset \{P\}$. As $xPx^{-1} = P$ for all $x \in P$, we must have $F_P\big(C(P)\big) = \{P\}$.

We use Case 1, to evaluate the constant in (†) as

$$(*) \qquad |C(P)| \equiv |F_P\big(C(P)\big)| = 1 \mod p.$$

Case 2. Let H be a Sylow p-subgroup (not necessarily P):

Applying (†) and (*), we have

$$1 \equiv |C(P)| \equiv |F_H\big(C(P)\big)| \mod p.$$

Hence $F_H\big(C(P)\big)$ is non-empty so $F_H\big(C(P)\big) = \{H\}$ by Lemma 21.5. In particular, $H \in F_H\big(C(P)\big) \subset C(P)$ proving that $Syl_p(G) = C(P)$, which is the Second Sylow Theorem. Plugging this information into (*), we have

$$|Syl_p(G)| = |C(P)| \equiv 1 \mod p.$$

Since $|C(P)| = [G : N_G(P)] = |G|/|N_G(P)|$ and $P \subset N_G(P)$ with $p \nmid [G : N_G(P)]$, we also have proven the Third Sylow Theorem.

Case 3. Let H be an arbitrary p-subgroup of G:

By (*), we have

$$|F_H\big(C(P)\big)| \equiv |C(P)| \equiv 1 \mod p.$$

Thus there exists a $T \in F_H\big(C(P)\big)$, hence $H \subset T \in Syl_p(G)$ by Lemma 21.5. This proves the Fourth Theorem.

□

Recall that a subgroup H of a group G is called a *characteristic subgroup* of G if the restriction map $res : \mathrm{Aut}(G) \to \mathrm{Aut}(H)$ given by $\varphi \mapsto \varphi|_H$ is well-defined, i.e., the target is correct. We write $H \triangleleft\triangleleft G$. We know that characteristic subgroups are normal, and being characteristic is a transitive property.

Proposition 21.7. *Let G be a finite group, p a (positive) prime dividing the order of G, and P a Sylow p-subgroup. Then the following are equivalent:*

(1) $Syl_p(G) = \{P\}$.
(2) $C(P) = \{P\}$.
(3) $P \triangleleft G$.
(4) $P \triangleleft\triangleleft G$.

Proof. The equivalence of (1), (2), and (3) follows from the Second Sylow Theorem. Clearly, (1) implies (4) as all isomorphic images of a group have the same cardinality. Finally (4) implies (3) by the comment above.

□

Corollary 21.8. *Let G be a finite group, p a (positive) prime dividing the order of G, and P a Sylow p-subgroup. Then $N_G(P) = N_G\big(N_G(P)\big)$.*

Proof. As P is a Sylow p-subgroup of $N_G(P)$ and normal in it, P is a characteristic subgroup of $N_G(P)$ by the proposition. As $N_G(P)$ is a normal subgroup of $N_G\big(N_G(P)\big)$, we conclude by Exercise 11.9(18) that P is normal in $N_G\big(N_G(P)\big)$, i.e., $N_G\big(N_G(P)\big) \subset N_G(P)$. So we must have equality. □

A stronger result is left as Exercise 21.11(1) below.

Examples 21.9. Let G be a finite group, p a prime dividing the order of G. In the following examples, we shall use the following notation:

P_p will denote an arbitrary Sylow p-subgroup of G.
n_p will denote the number of Sylow p-subgroups.

So there exists an integer k (depending on p) with

$n_p = |Syl_p(G)| = [G : N_G(P)] = 1 + pk$
P_p is normal if and only if $k = 0$.

1. Every group of order 15 is cyclic (hence isomorphic to $\mathbb{Z}/15\mathbb{Z}$) and every group of order 45 is abelian (hence isomorphic to $\mathbb{Z}/45\mathbb{Z}$ or $\mathbb{Z}/15\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$):

In both cases, the Third Sylow Theorem implies the Sylow p-subgroups are normal for $p = 3, 5$. As groups of order p are cyclic and of order p^2 are abelian, the result follows by Exercise 13.6(5). (We leave the statement about isomorphism as an exercise.)

2. If G is a group of order $p^r q$ with $p > q$ primes, then P_p is normal by the Third Sylow Theorem.
3. If G is a group of order $315 = 3^2 \cdot 5 \cdot 7$, then G is not simple:

   Suppose this is false. Then no Sylow p-subgroup is normal. We have
   $n_3 = 1 + 3k \mid 5 \cdot 7$ some k, so $[G : N_G(P_3)] = 7$.
   $n_5 = 1 + 5k \mid 3^2 \cdot 7$ some k, so $[G : N_G(P_5)] = 21$.
   By Lagrange's Theorem,
   $$N_G(P_3) \text{ is of order } 45 \text{ and } N_G(P_5) \text{ is of order } 15,$$
   so both are abelian, and $Syl_5\big(N_G(P_3)\big) \subset Syl_5(G)$. In particular, if $T \in Syl_5\big(N_G(P_3)\big)$, we have $T \lhd N_G(P_3)$ (or $1 + 5k \mid 3^2$ implies $k = 0$). Thus $N_G(P_3) \subset N_G(T)$. We conclude that
   $$45 = |N_G\big(P_3)\big)| \mid |N_G(T)| = 15$$
   by Lagrange's Theorem, which is impossible.
4. If G is a group of order $525 = 3 \cdot 5^2 \cdot 7$, then G is not simple:

   Suppose this is false. Then no Sylow p-subgroup is normal. Let P and Q be distinct Sylow 5-subgroups and set $I = P \cap Q$. We must have $|I| = 1$ or 5. Although PQ is not a group, we know by Exercise 13.6(3) that
   $$|PQ| = |P|\,|Q|/|P \cap Q| = |P|\,|Q|/|I|.$$
   Suppose that $I = 1$. Then $|PQ| = 5^4 > |G|$, so this is impossible. Hence $|I| = 5$. Both P and Q are abelian as of order 5^2, so $P, Q \subset N_G(I)$. Therefore,
   $$|N_G(I)| \geq |PQ| = 5^3.$$
   As $|N_G(I)| \mid |G|$, it follows that
   $$|N_G(I)| = 3 \cdot 5^2 \cdot 7 \text{ or } 5^2 \cdot 7.$$
   In the first case, $I \lhd N_G(I) = G$; and in the second case $[G : N_G(I)] = 3$ and $|G| \not| 3!$, so $N_G(I)$ contains a non-trivial normal subgroup of G by Useful Counting 12.7. In fact, $N_G(I) \lhd G$, as 3 is the smallest prime dividing $|G|$ (cf. Corollary 12.10).
5. If G is a group of order $945 = 3^3 \cdot 5 \cdot 7$, then G is not simple:

   Suppose this is false. Then no Sylow p-subgroup is normal. As $n_3 = 1 + 3k \mid 5 \cdot 7$, we have $[G : N_G(P_3)] = n_3 = 7$. Since $|G| \not| 7!$,

the subgroup $N_G(P_3)$ contains a non-trivial normal subgroup of G by Useful Counting 12.7.

6. If G is a group of order $1785 = 3 \cdot 5 \cdot 7 \cdot 17$, then G is not simple:
   Suppose this is false. Then no Sylow p-subgroup is normal. We have
   $$n_{17} = 1 + 17k \mid 3 \cdot 5 \cdot 7, \quad \text{so } n_{17} = 5 \cdot 7.$$
   $$n_3 = 1 + 3k \mid 5 \cdot 7 \cdot 17, \quad \text{so } n_3 = 7,\ 5 \cdot 17, \quad \text{or } 5 \cdot 7 \cdot 17.$$
   As $|G| = 1785 \not| 7!$, we can eliminate $n_3 = 7$ by Useful Counting 12.7. Thus
   $$|N_G(P_{17})| = 3 \cdot 17 \text{ and } |N_G(P_3)| = 3 \cdot 7 \text{ or } 3.$$
   As $3 \,||N_G(P_{17})|$ and $3 \,||G|$, we may assume that $P_3 \subset N_G(P_{17})$. But $1 + 3k \mid 17$ implies $k = 0$, so $P_3 \lhd N_G(P_{17})$. Consequently, $N_G(P_{17}) \subset N_G(P_3)$, implying that
   $$3 \cdot 17 = |N_G(P_{17})| \leq |N_G(P_3)| = 3 \cdot 7 \text{ or } 3,$$
   which is impossible.
7. Here is a fact useful in counting. Suppose that $|G| = pm$ with $p \not| m$, p a prime, then
   $$G \text{ contains } n_p(p-1) \text{ elements of order } p, \text{ so } |G| \geq n_p(p-1) + 1,$$
   adding in the identity. Similarly, if $|G| = pqm$ with $p \not| m,\ q \not| m$, and p and q distinct primes, then
   $$|G| \geq n_p(p-1) + n_q(q-1) + 1,$$
   adding in the identity. This last remark does not work if $p = q$, because two Sylow p-groups can intersect non-trivially, and it is difficult to estimate the number of elements in the union of all the Sylow p-subgroups.

For some other results, see the exercises below.

The proof of the following theorem illustrates many of the ideas that we have used in the examples.

Theorem 21.10. *Let p and q be primes, G a group of order $p^s q$ with $s \geq 1$. Then G is not simple.*

Proof. We may assume that $p \neq q$ and there exist no normal Sylow subgroups. In particular, if $n_p = |Syl_p(G)|$, then $1 < n_p = 1 + kp \mid q$ with $k \geq 1$, an integer. In particular, $n_p \leq q$. Among all the Sylow p-subgroups of G, choose two distinct ones, P_1 and P_2, satisfying $I = P_1 \cap P_2$ has maximal order. If $I = 1$, then all Sylow p-subgroups of G intersect pairwise in the identity, so the number of non-identity elements in the Sylow p-subgroups is $(p^s - 1)q$. As $|G| = p^s q$,

we conclude that the Sylow q-group is normal, a contradiction. So we may assume that $1 < I$. As I is a p-group but not a Sylow p-subgroup, we have $1 < I < N_{P_i}(I)$ for $i = 1, 2$ by Proposition 20.15. Let $H = \langle N_{P_1}(I), N_{P_2}(I)\rangle$, the group generated by $N_{P_1}(I)$ and $N_{P_2}(I)$. Then we must have $I \lhd H$.

Claim. $q \mid |H|$.

Suppose not. Then H is a p-group, so by the Fourth Theorem, there exists a Sylow p-group P_3 of G such that $H \subset P_3$. Hence for $i = 1, 2$, we have

$$P_1 \cap P_2 = I < N_{P_i}(I) \subset P_i \cap H \subset P_i \cap P_3.$$

This contradicts the maximality of $|I|$, lest $P_1 = P_1 \cap P_3 = P_2 \cap P_3 = P_2$. This proves the claim, i.e., $q \mid |H|$. Let Q be a Sylow q-group of H, hence also of G as $q \,||\, |G|$. Consequently, by Exercise 13.6(3), $|P_1Q| = |P_1|\,|Q|/|P_1 \cap Q| = |P_1|\,|Q| = |G|$ as p and q are relatively prime. This means that

$$(*) \qquad G = P_1Q := \{xy \mid x \in P_1,\, y \in Q\}.$$

If $g \in G$, set $I^g := gIg^{-1}$ and $N = \langle I^x \mid x \in G\rangle$. By Great Trick 11.3. we have $I \subset N \lhd G$. Let $g \in G$. By (*), we can write $g = xy$, for some $x \in P_1$ and $y \in Q$. In particular, as $y \in Q \subset H \subset N_G(I)$, we have

$$I^g = I^{xy} = xyIy^{-1}x^{-1} = xIx^{-1} = I^x \subset P_1,$$

as $I \subset P_1$. Thus $N \subset P_1 < G$, so $1 < N \lhd G$, and the theorem follows. □

Exercises 21.11. 1. Let G be a finite group, p a (positive) prime dividing the order of G, and P a Sylow p-subgroup. Suppose that H is a subgroup of G satisfying $N_G(P) \subset H \subset G$, then $H = N_G(H)$.

2. Let G be a finite group and $K \lhd G$. If P is a p-Sylow subgroup of K, show that $G = KN_G(P)$.

3. Let G be a finite group of order p^2q with p and q primes. Show that G is not simple without using Theorem 21.10.

4. Let G be a finite abelian group of order p^2q with p and q distinct primes. Show that G is isomorphic to either $\mathbb{Z}/p^2q\mathbb{Z}$ or $\mathbb{Z}/p\mathbb{Z}\times\mathbb{Z}/pq\mathbb{Z}$ and the second group is isomorphic to $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}$.

5. Let G be a finite group of order p^2q^2 with p and q primes. Show that G is not simple.

6. Let G be a group of order 56. Using Example 21.9(7) but not Theorem 21.10, show that G is not simple.

7. Let G be a finite group of order pqr with p, q, and r primes. Show that G is not simple.

8. Let G be a finite group, p the smallest prime dividing the order of G. Suppose that G contains a cyclic Sylow p-subgroup. Show that $N_G(P) = Z_G(P)$. (Cf. Exercise 12.11(7)).
9. Let p be a prime dividing the order of a finite group G. Let $P \in \mathrm{Syl}_p(G)$. If $N \triangleleft G$ show that $N \cap P \in \mathrm{Syl}_p(N)$.
10. Show that no group of order 144 is simple.
11. Show that no group of order 2000 is simple.
12. Using Exercise 16.14(8) and Exercise 20.26(7) show that a finite group is nilpotent if and only if it is a product of its Sylow subgroups.
13. A normal subgroup $N \neq 1$ of G is called a *minimal normal* subgroup if it contains no normal subgroups of G other than itself and 1. Show if G is a nontrivial finite solvable group, then any minimal normal subgroup of G is an *elementary p-group*, i.e., a group isomorphic to $(\mathbb{Z}/p\mathbb{Z})^n$ for some prime p and positive integer n.

22. The Symmetric and Alternating Groups

Recall if $\sigma \in S_n$, we write σ as

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}$$

where the top row is the elements of the domain and the bottom row the corresponding values. For example, if

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} \quad \text{and} \quad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix}$$

then

$$\tau\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix} \quad \text{and} \quad \sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}.$$

We want to write permutations in other ways.

Definition 22.1. A permutation $\alpha \in S_n$ is called a *cycle* of *length* r, or simply an *r-cycle*, if there exist distinct elements $a_1, \dots, a_r$ in $\{1, \dots, n\}$ (so $r \leq n$) satisfying

(1) $\alpha(a_i) = a_{i+1}$ for $i = 1, \dots, r-1$.
(2) $\alpha(a_r) = a_1$.
(3) $\alpha(a) = a$ for all $a \in \{1, \dots, n\} \setminus \{a_1, \dots, a_r\}$.

We denote the above r-cycle by $(a_1 \cdots a_r)$. If $r > 1$, we say that α is *non-trivial*.

Examples 22.2. (*i*) $(a_1) = 1$ the identity on $\{1, \dots, n\}$.
(*ii*) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix} = (23)$.

(*iii*) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} = (34)(12) = (12)(34).$

(*iv*) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix} = (1342).$

(*v*) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix} = (1243).$

The basic properties of cycles are given by:

Properties 22.3. Let $\alpha = (a_1 \cdots a_r)$ in S_n with $n \geq 2$. Then

(1) $\alpha = (a_i a_{i+1} \cdots a_r a_1 \cdots a_{i-1})$ if $1 \leq i \leq r$.
(2) α has order r.
(3) $\alpha^{-1} = (a_r \cdots a_1)$.
(4) If $\sigma \in S_n$, then $\sigma\alpha\sigma^{-1} = \big(\sigma(a_1) \cdots \sigma(a_r)\big)$.

PROOF. We leave the proof of (1) and (3) as an exercise.
(2): $\alpha^i(a_1) = a_{i+1} \neq a_1$ for $1 \leq i < r$, so the order of α is at least r. It is easy to check that $\alpha^r(a_i) = a_i$ for $1 \leq i \leq r$, so the order of $\alpha = r$.
(4): we have

$$\sigma\alpha\sigma^{-1}(j) = \begin{cases} \sigma(a_{i+1}) & \text{if } \sigma^{-1}(j) = a_i \text{ , i.e., } j = \sigma(a_i) \text{ for } i < r. \\ \sigma(a_1) & \text{if } \sigma^{-1}(j) = a_r \text{ , i.e., } j = \sigma(a_r). \\ j \ (= \sigma\sigma^{-1}\big(j)\big) & \text{otherwise.} \end{cases}$$

□

We say that two cycles $\alpha = (a_1 \cdots a_r)$ and $\beta = (b_1 \cdots b_l)$ are *disjoint* if $\{a_1, \ldots, a_r\} \cap \{b_1, \ldots, b_l\} = \emptyset$.

Lemma 22.4. *If α and β are disjoint cycles, then α and β commute, i.e., $\alpha\beta = \beta\alpha$.*

PROOF. $\beta(a_i) = a_i$ for all i and $\alpha(b_j) = b_j$ for all j. □

We now use the evaluation action in Example 20.20, which we recall.

Construction 22.5. Suppose that $S = \{1, \ldots, n\}$ and $\gamma \in S_n$ with $n > 1$. Let $\Gamma = \langle\gamma\rangle \subset S_n$. Then S is a Γ-set via the evaluation

$$* : \Gamma \times S \to S \text{ defined by } \gamma^j * a = \gamma^j(a).$$

The orbit of $a \in S$ is $\Gamma * a = \{\gamma^j(a) \mid j \in \mathbb{Z}\}$. Since $|\Gamma| \leq |S_n| = n! < \infty$, just as in the proof of the Classification of Cyclic Groups 9.9, there exists a least positive integer $m = m(a)$ satisfying $\gamma^m(a) = a$. Therefore, the orbit of a is $\Gamma * a = \{a, \gamma(a), \ldots, \gamma^{m-1}(a)\} \subset S$. Associate to this orbit, the m-cycle $\gamma_a = \big(a\gamma(a) \cdots \gamma^{m-1}(a)\big)$ in S_n.

Let $\mathcal{O}$ be a system of representatives for the equivalence relation $\sim_\Gamma$ arising from the Γ-action on S and let $b \in \{1, \ldots, n\}$. There exists an $a \in \mathcal{O}$ satisfying $b \in a * \Gamma$ and $\gamma(b) = \gamma_a(b)$. By the Mantra of Gp-actions, $S = \bigvee_{\mathcal{O}} \Gamma * a$. Since disjoint cycles commute, we have $\gamma = \prod_{\mathcal{O}} \gamma_a$, i.e., γ is a product of disjoint cycles with the product unique up to order. We call this the (*full*) *cycle decomposition* of γ.
For example,

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 5 & 4 & 1 & 6 \end{pmatrix} = (135)(2)(4)(6)$$

and

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 5 & 3 & 4 & 6 \end{pmatrix} = (12)(354)(6).$$

Note (again) that a 1-cycle (a) is the identity 1_S. Therefore, the *fixed points* of the Γ-action (i.e., elements a in S satisfying $\gamma(a) = a$) are precisely the elements of S occurring in the 1-cycles (if any) in a cycle decomposition of γ. We usually delete 1-cycles in the cycle decomposition of a permutation, e.g., $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 5 & 4 & 1 & 6 \end{pmatrix}$ is written (135). We shall call the cycle decomposition *full* if we do not delete the 1-cycles.

We turn to another representation of permutations. We call a 2-cycle a *transposition*. Note if $i \neq j$ then $(ij)^{-1} = (ij)$ has order 2 as needed. We first show that transpositions generate S_n (for $n > 1$).

Proposition 22.6. *Let $n > 1$, then every element in S_n is a product of the transpositions $(12), (13), \ldots, (1n)$.*

Proof. We first show that S_n is generated by transpositions. If $\sigma = (a_1 \ldots a_r)$, then

$$\sigma = (a_{r-1}a_r) \cdots (a_2a_r)(a_1a_r) \tag{22.7}$$

is a product of $r - 1$ transpositions.
To show that S_n is generated by the specific transpositions in the proposition, we use Properties 22.3(4) and (5), respectively, to see that

$(1i) = (i1) = (i1)^{-1}$ if $i \neq 1$

$(ij) = (1i)(1j)(1i)^{-1} = (1i)(1j)(1i)$ if 1, i, and j are distinct.

The result follows using the cycle decomposition of permutations. □

Remarks 22.8. If $\sigma \in S_n$, with $n > 1$, then σ is not a product of a unique number of transpositions.. However, by Equation 11.6, we have group homomorphisms

$$S_n \xrightarrow{\theta} \mathrm{Perm}_n(\mathbb{R}) \xrightarrow{\det} \{\pm 1\},$$

given by $\sigma \mapsto [T_\sigma]_{\mathcal{S}_n} \mapsto \det[T_\sigma]_{\mathcal{S}_n}$, where $\mathcal{S}_n$ is the standard basis for $\mathbb{R}^n$. The first map is an isomorphism, the second an epimorphism. So we have the alternating group $A_n := \ker \det \circ \theta \lhd S_n$ and $S_n/A_n \cong \{\pm 1\} \cong \mathbb{Z}/2\mathbb{Z}$ by the First Isomorphism Theorem. In particular, we have $[S_n : A_n] = 2$. If σ is a transposition, $\det[T_\sigma]_{\mathcal{S}_n} = -1$ (as it corresponds to interchanging two columns of the identity matrix). It follows that if $\tau \in S_n$ is a product of r transpositions and a product of s transpositions, then $(-1)^r = (-1)^s$, i.e., $r \equiv s \mod 2$. It follows by Equation 22.7, if σ is an r-cycle with r odd, then $\sigma \in A_n$ and if r is even, then $\sigma \in S_n \setminus A_n$. In particular, if $n \geq 3$, then every 3-cycle lies in A_n. Using the cycle decomposition of any permutation, we conclude:

$$A_n = \{\sigma \in S_n \mid \sigma \text{ is a product of an even number of transpositions}\}$$

and

$$S_n \setminus A_n = \{\sigma \in S_n \mid \sigma \text{ is a product of an odd number of transpositions}\}.$$

Elements in A_n are called *even permutation* and elements in $S_n \setminus A_n$ are called *odd permutations*. So if τ is an odd permutation, we have $S_n = A_n \vee \tau A_n$.

Actually, one must be a bit careful in the above so that we do not get a circular argument. This would depend on the proof of the determinant, whose existence we have not proven. One can prove the properties of the determinant so that one does not have a circular argument, the problem being that interchanging two rows of a square matrix changes the sign of its determinant. (Cf. Section §102 for a sophisticated proof of the determinant and its properties.)

The possible place for the circular reasoning arises from the definition of the *signum* map. For completeness, we shall give a proof for this. The standard proof uses polynomials, and although simple seems out of place, so we give a group theoretic proof. Let σ be a permutation in S_n and suppose that $\sigma = \gamma_1 \cdots \gamma_r$ is its full cycle decomposition. Define the *signum* of σ by $\operatorname{sgn}(\sigma) = (-1)^{n-r}$. As a cycle decomposition is unique, this defines a map $\operatorname{sgn} : S_n \to \{\pm 1\}$.

Note if σ is an s-cycle, then in its full cycle decomposition, the number of 1-cycles is $n - s$, so $\operatorname{sgn} \sigma = (-1)^{s-1}$. In particular, if σ is a transposition, then $\operatorname{sgn}(\sigma) = -1$. The result that we need is that sgn is a group homomorphism. Clearly, the sgn of disjoint cycles is the product of the sgn of those cycles. As every permutation is a product of transpositions, the result now follows easily from the following:

Let $a, b, c_1, \ldots, c_r, d_1, \ldots, d_s$ be distinct elements in $\{1, \ldots, n\}$ and $\tau = (ab)$. If $\sigma = (ac_1 \cdots c_r)$, then $\tau\sigma = (c_1 \cdots c_r ba)$, hence

$$\operatorname{sgn}(\tau\sigma) = (-1)^{r+1} = -\operatorname{sgn}(\sigma);$$

and if $\sigma = (ac_1 \cdots c_r bd_1 \cdots d_s)$, then $\tau\sigma = (ac_1 \cdots c_r)(bd_1 \cdots d_s)$, hence

$$\operatorname{sgn}(\tau\sigma) = (-1)^{r+s} = -(-1)^{r+s+1} = -\operatorname{sgn}(\sigma).$$

In particular, the parity of the number of transpositions in a decomposition of a permutation is invariant.

We turn to the study the alternating group. We know that 3-cycles are even. We start by showing that 3-cycles generate A_n. As A_n, $n = 1, 2$ is the trivial group, it is generated by the empty set of 3-cycles, so we may assume that $n \geq 3$.

Proposition 22.9. *The alternating group A_n is a normal subgroup of index two in S_n, and, if $n \geq 3$, then A_n is generated by the 3-cycles:*

$$(123),\ (124), \ldots, (12n).$$

Proof. We need only show the last statement. We first show that A_n is generated by 3-cycles. We have shown that every element in A_n is a product of an even number of transpositions, so it suffices to show a product of two distinct transpositions is a product of 3-cycles. Suppose that i, j, k, l are distinct elements in $\{1, \ldots, n\}$. Then this follows by the equations

$$(ij)(ik) = (ikj)$$
$$(ij)(kl) = (kil)(ijk).$$

To finish, we show that the alleged 3-cycles generate. Let i, j, k be distinct elements in $\{1, \ldots, n\}$, and different from 1 and 2. Then, using Property 22.3, this follows from the equations:

$$(ijk) = (12i)(2jk)(12i)^{-1}$$
$$(2jk) = (12i)(12k)(12j)^{-1}$$
$$(1jk) = (12k)^{-1}(12j)(12k). \qquad \square$$

The key to Abel's Theorem on the simplicity of the alternating group A_n for $n \geq 5$, is the following:

Lemma 22.10. *Let K be a normal subgroup of A_n. If K contains a 3-cycle, then $K = A_n$.*

Proof. As K contains a 3-cycle, we have $n \geq 3$. By the proposition, it suffices to show that $(12k)$ lies in K for $k = 3, \ldots, n$. Changing notation, we may assume that (123) lies in K, hence $(213) = (123)^{-1}$

lies in K. Let $\sigma = (12)(3k)$ with $k > 3$, a product of an even number of transpositions, so an element of A_n. As $K \lhd A_n$, we have

$$(12k) = \sigma(213)\sigma^{-1} \text{ lies in } K, \text{ for all } k \geq 3$$

as needed. □

We now prove Abel's Theorem. This gives our first examples of non-abelian simple groups.

Theorem 22.11. (Abel's Theorem) *Let $n \neq 4$, then the alternating group A_n is simple.*

Proof. We know that A_2 and A_3 are simple, so assume that $n \geq 5$, and $1 < K \lhd A_n$. We must show $K = A_n$. By the lemma, it suffices to show that K contains a (single) 3-cycle.
Let $e \neq \alpha \in K$ be chosen such that

$$\alpha \text{ moves precisely } m \text{ elements in } \{1, \ldots, n\}$$

and

$$\begin{gathered}\text{no other non-identity element in } K \\ \text{moves a fewer number of elements,}\end{gathered}$$

where we say α *moves* i, if $\alpha(i) \neq i$, i.e.,

$$\begin{gathered}m \text{ is the minimal number of elements moved} \\ \text{by any non-identity element in } K.\end{gathered}$$

Let

$$\alpha = \gamma_1 \cdots \gamma_r$$

be the decomposition of α into a product of disjoint cycles. We may assume that γ_1 is a k-cycle with k maximal among the lengths of $\gamma_1, \ldots, \gamma_r$.
As α is not the identity, we must have $m \geq 2$. If $m = 2$, then we must have $\alpha = \gamma_1$ and be a transposition, hence it is an odd permutation, so not an element of A_n, a contradiction. If $m = 3$, then we must have $\alpha = \gamma_1$ and be a 3-cycle, and we are done by the lemma. So we may assume that $m \geq 4$. For clarity, it is convenient to changing notation if necessary. So we may assume that

$$\begin{aligned}\alpha &\text{ moves } 1, 2, 3, 4. \\ \alpha &\text{ moves } 5 \text{ if } m \geq 5. \\ \gamma_1 &= (12 \cdots k).\end{aligned}$$

We have two cases:
Case 1. $k \geq 3$: If this is the case, then $m \geq 5$:

For if not, then we have $m = 4$, and we must have $\alpha = \gamma_1 = (1234)$ and be a 4-cycle which is not in A_n, a contradiction.

Case 2. $k = 2$: Then each γ_i is a transposition or a 1-cycle, and we may assume that $\alpha = (12)(34)\cdots$:
Set $\beta = (345)$ in A_n. As $K \triangleleft A_n$, we have $\alpha_1 = \beta\alpha\beta^{-1}$ lies in K and

$$\alpha_1 = \beta\alpha\beta^{-1} = \begin{cases} (124\cdots)\cdots & \text{in Case 1.} \\ (12)(45)\cdots & \text{in Case 2.} \end{cases}$$

In either case, we check that $\alpha(3) \neq \alpha_1(3)$, so $e \neq \alpha^{-1}\alpha_1$ lies in K; and $\alpha(1) = \alpha_1(1)$, so $\alpha^{-1}\alpha_1$ fixes 1. Since β fixes every $i > 5$, we have $\alpha^{-1}\alpha_1$ fixes every $i > 5$ that α fixes. As $m \geq 5$ in Case 1 and in this case $\alpha^{-1}\alpha_1$ fixes 1, it follows by the minimality of m that this can happen only if α satisfies Case 2. But then $\alpha(1) = \alpha_1(1)$ and $\alpha(2) = \alpha_1(2)$, so $\alpha^{-1}\alpha_1$ fixes 1 and 2. As $m \geq 4$ (with α fixing 5 if $m = 4$), it follows that $\alpha^{-1}\alpha_1$ moves at most $m - 1 = m - 2 + 1$ elements contradicting the minimality of m. The result follows. □

Remark 22.12. If G is a group containing a non-abelian simple group, then G is not solvable by Theorem 16.3. Therefore, S_n is not solvable for any $n \geq 5$.

We wish to show that A_5 is the only simple group of order 60 up to isomorphism (non-abelian as its order is not a prime). It is the first non-abelian simple group by computation, using the Sylow Theorems. We make the general observation:

Lemma 22.13. *Let $n \geq 5$, then A_n is the only subgroup of S_n of index two.*

Proof. Suppose that K is a normal subgroup of G of index two and $K \neq A_n$. By the Second Isomorphism Theorem, we conclude A_nK is a group with $K < A_nK$ normal and $A_nK/K \cong A_n/(A_n \cap K)$. As $A_n < A_nK \subset S_n$, we must have $S_n = A_nK$, so $2 = [A_n : A_n \cap K]$. In particular, $1 < A_n \cap K < A_n$. It follows that $A_n \cap K$ is a proper normal subgroup of the simple group A_n, a contradiction. □

Theorem 22.14. *Up to isomorphism, the alternating group A_5 is the only simple group of order* 60.

Proof. Let G be a simple group of order 60. We prove this in three steps.

Step 1. If G contains a subgroup H of index 5, then $G \cong A_5$:
Let $\lambda : G \to \Sigma(G/H)$ given by $x \mapsto (\lambda_x : gH \mapsto xgH)$ be the homomorphism defined in the General Cayley Theorem 12.4. Since $H < G$,

we have $\ker(\lambda) \subset H < G$. Consequently, λ must be monic (since not the trivial map), as G is simple. Therefore, G is isomorphic to $\lambda(G) \subset \Sigma(G/H)$, a subgroup of index two. By the lemma $G \cong A_5$.

Step 2. G does not contain a subgroup of index three:

Suppose that H is a subgroup of index three in G. Then $|G| \nmid [G : H]!$, which implies that G is not simple by Useful Counting 12.7.

Step 3. Finish:

By the first two steps, we may assume that G has no subgroup of index 3 or index 5. Let P be a Sylow 2-subgroup of G. Then P is not a normal subgroup of G by assumption, hence by the Third Sylow Theorem, we have $[G : N_G(P)] = 1 + 2k \mid 3 \cdot 5$, for some integer $k \geq 1$. By the first two steps, we must have $[G : N_G(P)] = 15$, so $|N_G(P)| = 4$. (Note that this means that $P = N_G(P)$.) Let $Q \neq P$ be another Sylow 2-subgroup of G, and set $I = P \cap Q$. We may also assume that Q and P were chosen such that $I = P \cap Q$ has maximal cardinality.

Case 1. $1 < I$, so $|I| = 2$:

In this case, we have, by Exercise 13.6(3),

$$|PQ| = |P|\,|Q|/|P \cap Q| = |P|\,|Q|/|I| = 2^3.$$

As both P and Q are groups of order 2^2, they are both abelian, so normalize I. Consequently, $P, Q \subset N_G(I)$. It follows that

$$2^3 = |PQ| \leq |N_G(I)| \mid 2^2 \cdot 3 \cdot 5 \quad \text{and} \quad 2^2 \mid |N_G(I)|.$$

Therefore, $|N_G(I)| = 2^2 \cdot 3, 2^2 \cdot 5$, or $2^2 \cdot 3 \cdot 5$. By the first two steps, we conclude that $|N_G(I)| = 2^2 \cdot 3 \cdot 5 = |G|$. Hence $I \lhd N_G(I) = G$, contradicting the simplicity of G.

Case 2. $I = 1$:

As Q and P were chosen with $I = P \cap Q$ of maximal cardinality, and each Sylow 2-subgroup contains $(4 - 1) = 3$ non-identity elements, the Sylow 2-subgroups of G contain $45 = 15(4 - 1)$ distinct non-identity elements. By the simplicity of G, the number of the Sylow 5-subgroups, which divides $2^2 \cdot 3$, must be at least six. Each Sylow 5-subgroup of G is cyclic of order 5, so G must contain at least $6 \cdot (5 - 1) = 24$ distinct elements of order 5, which implies that $|G| \geq 45 + 24 > 60$, a contradiction. □

Remarks 22.15. Let F be a field, then the group

$$\mathrm{PSL}_n(F) := \mathrm{SL}_n(F)/Z\big(\mathrm{SL}_n(\mathrm{F})\big)$$

is called a *projective unimodular group* of F. It can be shown that $\mathrm{PSL}_n(F)$ is simple for all $n \geq 3$.

We shall see later that for all primes p and $q = p^n$, with $n \in \mathbb{Z}^+$, there exists a field $\mathbb{F}_q$ with q elements, and unique up to isomorphism of fields (definition?). Computation shows that

$$|\mathrm{PSL}_2(\mathbb{F}_q)| = \begin{cases} (q+1)(q^2-q) & \text{if } q = 2^n \\ \frac{1}{2}(q+1)(q^2-q) & \text{if } q = p^n \text{ with } p \text{ and odd prime} \end{cases}$$

(cf. Exercise 9.12(1)). It can be shown that $\mathrm{PSL}_2(\mathbb{F}_q)$ is simple if and only if $q > 3$. The computation shows that $\mathrm{PSL}_2(\mathbb{F}_4)$ and $\mathrm{PSL}_2(\mathbb{F}_5)$ both have order 60, so are isomorphic to A_5. The simple group $\mathrm{PSL}_2(\mathbb{F}_7)$ of order 168 is the next non-abelian simple group up to isomorphism. Computation shows that the groups A_8 and $\mathrm{PSL}_3(\mathbb{F}_4)$ both have $8!/2$ elements. It can also be shown that they are not isomorphic, i.e., there exist non-isomorphic simple groups of the same order.

Our results and exercises have shown that groups of orders $p^s q$ ($s \geq 1$), p^2q^2, and pqr are not simple when p, q, and r are primes. It follows that these groups are therefore solvable, by induction and reduction to simpler cases. Burnside showed that groups of order $p^a q^b$ are solvable for any primes p and q for over a hundred years ago. The original proof did not use abstract group theory (but group representation theory). A group theoretic proof was finally found, but is much more difficult than this original proof. More spectacularly, Feit-Thompson proved that any group of odd order is solvable. The proof is very difficult, and was the first major step in classifying finite simple groups. Groups of even order are, therefore, much harder to understand. We give two examples when we can say something more.

The first example, is left as an easy exercise (cf. the proof of Lemma 22.13):

Proposition 22.16. *Let $n \geq 5$. If H is a normal subgroup of S_n, then either $H = 1$, $H = A_n$, or $H = S_n$.*

The order of A_5 is divisible by 4. We show groups of even order larger than two but not divisible by 4 are never simple, i.e., if G is a group of order $2n$, with $n > 1$ an odd integer, then G is not simple.

Review 22.17. Let G be a nontrivial finite group. We have seen that Cayley's Theorem gives a group monomorphism

$$\lambda : G \to \Sigma(G) \text{ defined by } x \mapsto (\lambda_x : a \mapsto xa)$$

called the (*left*) *regular representation* of G. If $xa = \lambda_x(a) = a$, then $x = e$, i.e., the permutation λ_a has a fixed point if and only if $a = e$. Note also that the regular representation corresponds to the translation action of G on $S = G$ which is transitive, i.e., S is the only orbit under

this action. In particular, this action has no fixed points (one-point orbits).

We need two lemmas.

Lemma 22.18. *Let G be a group of order mn and $\lambda : G \to \Sigma(G)$ the regular representation with $m > 1$. If the element a in G has order m, then*

(1) λ_a is a product of n disjoint m-cycles.

(2) λ_a is an odd permutation if and only if m is even and n is odd.

[Note. This characterizes elements in $\lambda(G)$, but not in $\Sigma(G)$.]

PROOF. (1): By the review, we know if $1 \leq r < m$, then the permutation $\lambda_{a^r} = (\lambda_a)^r$ has no fixed points. Let $\lambda_a = \gamma_1 \cdots \gamma_s$ be a full cycle decomposition of λ_a. As λ_{a^r} has no fixed points for $1 \leq r < m$, each γ_i must be a cycle of order at least m. But $\lambda_{a^m} = 1_G$, so each γ_i must be a cycle of length m. Since each element of G occurs precisely once in the cycle decomposition for λ_a, we must have $s = n$.

By (1), we have a full cycle decomposition of λ_a given by $\lambda_a = \gamma_1 \cdots \gamma_n$, with each γ_i an m-cycle. Therefore, λ_a is an odd permutation if and only if $n(m-1)$ is odd, i.e., if and only if n is odd and m is even. □

Lemma 22.19. *Let G be a finite group and $\lambda : G \to \Sigma(G)$ the regular representation. If $\lambda(G)$ contains an odd permutation, then there exists a normal subgroup of G of index two.*

PROOF. Let $|G| = n$ and identify $\Sigma(G)$ and S_n. As $\lambda(G)$ contains an odd permutation, $H := A_n \cap \lambda(G) < \lambda(G)$ and $\lambda(G)A_n = \Sigma(G)$. The regular representation is monic, so, using the Second Isomorphism Theorem, we must have

$$\begin{aligned}[G : \lambda^{-1}(H)] &= [\lambda(G) : H] = [\lambda(G) : A_n \cap \lambda(G)] \\ &= [A_n\lambda(G) : A_n] = [S_n : A_n] = 2.\end{aligned}$$

Therefore, $\lambda^{-1}(H) < G$ is a subgroup of index two hence is also normal. □

Application 22.20. Let G be a finite group of order $2n$ with n odd. Then G contains a normal subgroup of index two. In particular, if $n > 1$, G is not simple.

PROOF. By Cauchy's Theorem, there exists an element a in G of order two. Let $\lambda : G \to \Sigma(G)$ be the regular representation. By Lemma 22.18, the permutation λ_a is a product of n disjoint transpositions, hence is an odd permutation. The result now follows by Lemma 22.19. □

This application can be pushed a bit further. Since all Sylow p-subgroups of a group are conjugate, if one is abelian (respectively, cyclic) all are.

Theorem 22.21. *Let G be a finite group of order $2^r m$ with m odd. If G contains a cyclic Sylow 2-subgroup then there exists a normal subgroup of G of index 2^r. In particular, if $m > 1$ or $r > 1$, then G is not simple.*

PROOF. We may assume that $m > 1$ and $r \geq 1$. Let $\lambda : G \to \Sigma(G)$ be the regular representation. By Lemma 22.18, the group $\lambda(G)$ contains an odd permutation, so by Lemma 22.19, there exists a normal subgroup N of G of index two. Hence $|N| = 2^{r-1}m$, and we may assume that $r > 1$. Let $P_0 \in Syl_2(N)$. By the Fourth Theorem, there exists $P \in Syl_2(G)$ containing P_0. By the remark above, P is cyclic, say $P = \langle x \rangle$, so $P_0 = \langle x^2 \rangle$ by the Cyclic Subgroup Theorem. By induction on r, there exists a normal subgroup H in N of index 2^{r-1}. By the Second Isomorphism Theorem, $P_0 H \subset N$ is a subgroup, and P_0 and H have relatively prime orders, so $P_0 \cap H = 1$ (why?). It follows by counting that $N = P_0 H$. Let $H^x := xHx^{-1}$. We have $G = N \vee xN$, since $x \notin N$. Therefore, $H \lhd G$ if and only if $H^x = H$, as $H \lhd N$. If this is the case, we are done, so we may assume that $H \neq H^x$. The group N being normal in G implies that $N^x = xNx^{-1} = N$. Since $N = P_0 H$ and $P_0^x = xP_0x^{-1} = P_0$ in cyclic P, we have $N = P_0 H = N^x = P_0^x H^x$, so $H^x \subset N$. As $|H| = |H^x|$, we know that $2 \not| \, |H^x|$. By the Second Isomorphism Theorem, as $H \lhd N$ and $H^x \subset N$, we have HH^x is a subgroup of N. In particular, $|H| \, | \, |HH^x|$. But we also have $2 \not| \, |H||H^x|/|H \cap H^x| \, | \, |HH^x|$, so we must have $|HH^x| \leq |N|/2^e = |H|$, hence $|HH^x| = |H| = |H^x|$. We conclude that $H = H \cap H^x = H^x$, a contradiction. Therefore, $H \lhd G$. □

Exercises 22.22. 1. Let α be a product of disjoint cycles $\gamma_1, \ldots, \gamma_m$. Show that the order of α is the least common multiple of the orders of the γ_i, $1 \leq i \leq m$. In particular, show that this means that the order of each γ_i divides the order of α and further, that if $p > 1$ is a prime, then every power of a p-cycle is either a p-cycle of the identity.

2. We say that an element $\alpha \in S_n$ has *cycle type* $(r_1, \ldots, r_m)$ if α is a product of disjoint cycles $\gamma_1, \ldots, \gamma_m$ with γ_i an r_i-cycle for $1 \leq i \leq m$ and $r_1 \leq \cdots \leq r_m$. Prove that two elements of S_n are conjugate in S_n if and only if they have the same cycle type.

3. A permutation in S_n is called *regular* if it is the identity or has no fixed points and is the product of disjoint cycles of the same length.

Show that a permutation is regular if and only if it is a power of an n-cycle.

4. Let α be an r-cycle. If $k > 0$ is an integer and (r, k) the gcd of r and k, show that α^k is a product of (n, k) disjoint cycles, each of length $r/(r, k)$.
5. Show the following:
   (a) A_4 can be generated by two elements x and y satisfying $x^2 = y^3 = (xy)^3 = 1$.
   (b) S_4 can be generated by two elements x and y satisfying $x^2 = y^3 = (xy)^4 = 1$.
   (c) A_5 can be generated by two elements x and y satisfying $x^2 = y^3 = (xy)^5 = 1$.
6. Let $n \geq 5$. If H is a normal subgroup of S_n, then either $H = 1$, $H = A_n$, or $H = S_n$.
7. Let G be a finite group and N a normal subgroup of G of prime index p. Prove that $[G, G] \subset N$ and use it to prove that S_n, $n \geq 5$ is not polycyclic, hence not solvable without using Abel's Theorem.
8. Let $n \geq 5$. Then the only proper subgroup of S_n of index at most $n - 1$ is A_n.
9. Let G be a free group on basis $\mathcal{B}$ with $\mathcal{B}$ having at least two elements. Show that G is not solvable.
10. Show that the fiber $\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/5\mathbb{Z}$ (cf. Definition 17.11 and Theorem 17.12) is not solvable.
11. Let $A_\infty := \bigcup_{i=1}^{\infty} A_n$ and $S_\infty := \bigcup_{i=1}^{\infty} S_n$ be groups defined in the obvious way (with A_n and S_n respective subgroups for all n). Show that A_∞ is a normal subgroup of index two in S_∞ and $1 < A_\infty < S_\infty$ is a composition series. Also show the infinite abelian subgroup of S_∞ generated by the transpositions $(2k-1\,2k)$, $k \geq 1$, does not have a composition series.(Cf. Exercise 16.14(11).)

Part 3

Ring Theory

CHAPTER V

General Properties of Rings

In this chapter, we begin our formal study of rings, especially that for commutative rings. We find the analogues in ring theory for the notions and basic theorems we learned in group theory. The analogue in ring theory of a normal subgroup is called an ideal. As in the group theory, this turns out to be those subsets of a ring that are kernels of ring homomorphisms. This leads to analogous isomorphism theorems and an analogous correspondence principle.

When studying the integers, we looked at integers themselves. In this chapter, we switch our perspective from studying elements to studying ideals. We discover the type of ideal corresponding to a prime integer in commutative rings. This type of ideal, called a *prime ideal*, is the major object of study in commutative ring theory whose definition is based upon Euclid's Lemma. (For elements the analogue of a prime integer is called an *irreducible element*.) We also introduce an important generalization of the ring of integers called a principle ideal domain or PID. Although this ring is a very special type of domain (a commutative ring in which $ab = 0$ implies $a = 0$ or $b = 0$), it will be the main type of domain that we shall study. However, we shall show that even for a general domain, there exists a *field of fractions* or *quotient field*, thus constructing, in particular, the field of rational numbers. We shall also find the ring analogue of the Chinese Remainder Theorem that we have previously proved for the integers. Finally, we shall define the set theoretic extension of finite induction used in algebra, called Zorn's Lemma. This axiom, equivalent to the Axiom of Choice that you may have seen, is a powerful tool. We shall then give some simple applications of it that are basic to ring theory and linear algebra.

23. Definitions and Examples

We begin by recalling the definition of a ring, together with specific types of rings.

Definition 23.1. Let R be a set with two binary operations

$$+ : R \times R \to R \qquad \text{and} \qquad \cdot : R \times R \to R.$$

We call R a *ring* (under $+$ and $\cdot$) if:

(1) $(R, +)$ is an additive group (with identity written 0 or 0_R).
(2) $(R, \cdot)$ is a monoid (with identity written 1 or 1_R).
(3) R satisfies the *distributive laws*, i.e., for all $a, b, c \in R$, we have:
  (a) $a \cdot (b + c) = a \cdot b + a \cdot c$.
  (b) $(b + c) \cdot a = b \cdot a + c \cdot a$.

If (2) is replaced by

(2′) $(R, \cdot)$ satisfies associativity (i.e., possibly no 1),

then (following Jacobson), we call R (under $+$ and $\cdot$) a *rng*. [Take out the condition of having an identity, take out the i from ring.]
A ring satisfying

(4) $(R \setminus \{0\}, \cdot)$ is a group

is called a *division ring*.
As usual if $a, b \in R$, we usually write ab for $a \cdot b$.

Also recall that if R is a ring with $1 \neq 0$, then

$$\begin{aligned} R^\times &:= \{x \in R \mid \text{There exists } x^{-1} \in R \text{ such that } xx^{-1} = 1 = x^{-1}x\} \\ &= \{x \in R \mid \text{There exists } a, b \in R \text{ such that } ax = 1 = xb\} \end{aligned}$$

is called the *group of units* of R. (Why is it a group?) In particular, a ring is a division ring if and only if $R^\times = R \setminus \{0\}$ (cf. Property 23.4(3c) below).

Definition 23.2. A ring R is called a *commutative ring* if for all $a, b \in R$, we have

$$ab = ba.$$

A commutative ring R is called an *(integral) domain* if

(i) $0 \neq 1$.
(ii) Let $a, b \in R$. If $ab = 0$, then either $a = 0$ or $b = 0$.

A commutative division ring is called a *field*.

Remark 23.3. Check that a field is a domain.

Rings satisfy the following basic properties (which we leave as an exercise):

Properties 23.4. Let R be a rng and $a, b, c \in R$. Then the following are true:

(1) $a \cdot 0 = 0 = 0 \cdot a$.
(2) $a \cdot (-b) = (-a) \cdot b = -(a \cdot b)$.
(3) If R is a ring, then for all $a \in R$:
  (a) $(-1) \cdot a = -a$.

(b) $(-1)\cdot(-1) = 1$.
(c) $|R| > 1$ if and only if $1 \neq 0$.

(4) If R is a commutative ring with $1 \neq 0$, then R is a domain if and only if R satisfies the *Cancellation Law*: If a, b, c lies in R, then

$$a \cdot b = a \cdot c \text{ with } a \neq 0 \text{ implies } b = c.$$

We already know many examples of rings. We recall some of these.

Examples 23.5. 1. A *trivial ring* is a ring with one element. By the above, this is the case if and only if $1 = 0$ in R.

2. $\mathbb{Z}$ is a domain.

3. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields.

4. The subset of $\mathbb{C}$ given by

$$\mathbb{Z}[\sqrt{-1}] = \{a + b\sqrt{-1} \mid a, b \in \mathbb{Z}\}$$

is a domain under the restriction of the $+$ and $\cdot$ of $\mathbb{C}$ called the *Gaussian integers*. If $z \in \mathbb{Z}[\sqrt{-1}]$, there exist unique integers a and b satisfying $z = a + b\sqrt{-1}$. (Cf. the next example.)

5. The subset of $\mathbb{C}$ given by

$$\mathbb{Q}[\sqrt{-1}] = \{a + b\sqrt{-1} \mid a, b \in \mathbb{Q}\}$$

is a field under the restriction of the $+$ and $\cdot$ of $\mathbb{C}$. (Can you show this?) Note that if $z \in \mathbb{Q}[\sqrt{-1}]$, there exist unique integers a and b satisfying $z = a + b\sqrt{-1}$. Why?

6. $\mathbb{Z}/n\mathbb{Z}$, with $n \in \mathbb{Z}^+$, is a commutative ring. It is a domain if and only if n is a prime if and only if it is a field.

7. If R is a ring so is $\mathbb{M}_n(R)$. If R is a commutative ring with $|R| > 1$, then $\mathbb{M}_n(R)$ is commutative if and only if $n = 1$. (If R is a rng then so is $\mathbb{M}_n(R)$ — obvious definition.)

8. If R is a ring, then $R[t]$, the set of polynomials with coefficients in R, is a ring under the usual $+$ and $\cdot$ of polynomials.
[What are the usual $+$ and $\cdot$ of polynomials? What is $1_{R[t]}$?]

9. Let R be a ring. Define the *ring of (formal) power series* over R to be the set

$$R[[t]] := \{\sum_{i=0}^{\infty} a_i t^i \mid a_i \in R \text{ for all } i\}$$

with properties and operations:

$$0_{R[[t]]} = 0_R \text{ and } 1_{R[[t]]} = 1_R, \text{ with } t^0 = 1_R$$
$$\sum_{i=0}^{\infty} a_i t^i = \sum_{i=0}^{\infty} b_i t^i \text{ if and only if } a_i = b_i \text{ for all } i$$
$$\sum_{i=0}^{\infty} a_i t^i + \sum_{i=0}^{\infty} b_i t^i := \sum_{i=0}^{\infty} (a_i + b_i) t^i$$
$$\sum_{i=0}^{\infty} a_i t^i \cdot \sum_{i=0}^{\infty} b_i t^i := \sum_{i=0}^{\infty} c_i t^i \text{ with } c_i = \sum_{j=0}^{i} a_j b_{i-j} \text{ for all } i.$$

Just as for groups, we need to define the maps of interest between rings.

Definition 23.6. A map $\varphi : R \to S$ of rings is called a *ring homomorphism* if for all $a, b \in R$, the map φ satisfies:

(1) $\varphi(0_R) = 0_S$.
(2) $\varphi(a + b) = \varphi(a) + \varphi(b)$.
(3) $\varphi(1_R) = 1_S$.
(4) $\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$.

I.e., φ is both a group homomorphism $\varphi : (R, +) \to (S, +)$ and a 'monoid homomorphism' $\varphi : (R, \cdot) \to (S, \cdot)$. [If we omit (3), we call φ a *rng homomorphism*.].
A ring homomorphism $\varphi : R \to S$ is called

i. a *ring monomorphism* or *monic* or *mono* if φ is injective.
ii. a *ring epimorphism* or *epic* or *epi* if φ is surjective.
iii. a *ring isomorphism* if φ is bijective with inverse a ring homomorphism.
iv. a *ring automorphism* if $R = S$ and φ is a ring isomorphism.

Of course, if R and S are fields, etc., then φ is called a *field homomorphism*, etc.
If $\varphi : R \to S$ is a ring homomorphism, we let

$$\ker \varphi := \{x \in R \mid \varphi(x) = 0_S\} \subset R$$

called the *kernel* of φ, and

$$\operatorname{im} \varphi := \{\varphi(x) \mid x \in R\} \subset S$$

called the *image* of φ.

Remarks 23.7. 1. A ring homomorphism $\varphi : R \to S$ is a ring isomorphism if and only if φ is bijective. (Proof?)

2. We say two rings R and S are *isomorphic* if there exists an isomorphism $\varphi : R \to S$. If this is the case, we write $R \cong S$ and often write the isomorphism as $\varphi : R \xrightarrow{\sim} S$.
3. A ring homomorphism $\varphi : R \to S$ is monic if and only if $\ker \varphi = 0 := \{0\}$.
4. Let $n \in \mathbb{Z}^+$, then the canonical map $^- : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ is a ring epimorphism with $\ker ^- = n\mathbb{Z} = \{ny \mid y \in \mathbb{Z}\}$.

Remark 23.8. Let $\varphi : R \to S$ be a ring homomorphism. We leave it as an exercise to show that φ is a monomorphism if and only if given any ring homomorphisms $\psi_1, \psi_2 : T \to R$ with compositions satisfying $\varphi \circ \psi_1 = \varphi \circ \psi_2$, then $\psi_1 = \psi_2$. (Cf. Exercise 1.12(5).) However, the analogue of Exercise 1.12(6) is false. In particular, a ring homomorphism $\varphi : R \to S$ with the property that whenever there exists ring homomorphisms $\theta_1, \theta_2 : S \to T$ with compositions satisfying $\theta_1 \circ \varphi = \theta_2 \circ \varphi$, then $\theta_1 = \theta_2$ does not necessarily imply that φ is surjective. For example the inclusion map of $\mathbb{Z}$ into $\mathbb{Q}$ is a ring homomorphism that satisfies this condition, but is not a surjective ring homomorphism. (Cf. Exercise 24.20(5)). In category theory, being epi is defined by this condition. For this reason, it is no longer common to define a surjective ring homomorphism as a ring epimorphism. However, for consistency of notation, we shall use this older definition that an epimorphism is a surjective homomorphism.

We also need the analogs from group theory of subgroup and normal subgroup.

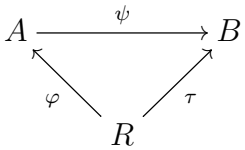
Definition 23.9. Let R be a ring, $S \subset R$ a subset with S a ring. We say S is a *subring* of R if the inclusion map, $inc : S \subset R$ is a ring monomorphism, i.e., $1_S = 1_R$ and S is *closed* under the ring operations $+$ and $\cdot$ on R, which means that the restriction of the operations on R to S have image in S, viz., $+|_{S\times S} : S \times S \to S$ and $\cdot|_{S\times S} : S \times S \to S$. [We also have the obvious notion of a *subrng* of a rng, if we ignore, in this case, the condition on the 1's, if any.]

Examples 23.10. 1. $\mathbb{Z} \subset \mathbb{Q}$ and $\mathbb{Z}, \mathbb{Q}, \mathbb{R} \subset \mathbb{C}$ are subrings.
2. If R is a domain, e.g., a field, and $S \subset R$ a subring, then S is a domain.
3. Let $\mathbb{Z}\times\mathbb{Z}$ be a ring under componentwise $+$ and $\cdot$, so $1_{\mathbb{Z}\times\mathbb{Z}} = (1_{\mathbb{Z}}, 1_{\mathbb{Z}})$ and $\mathbb{Z} \times \mathbb{Z}$ is commutative ring, but not a domain. We also have $\mathbb{Z} \times 0 := \mathbb{Z} \times \{0\} \subset \mathbb{Z} \times \mathbb{Z}$ is not a subring but is a subrng.
4. Let R be a ring. The *center* of R is the set

$$Z(R) := \{a \in R \mid ax = xa \text{ for all } x \in R\}.$$

It is a commutative subring of R. For example, the center of a division ring is a field.

Remark 23.11. The last example allows us to generalize our terminology. Let R be a commutative ring, A a ring and $\varphi : R \to A$ a ring homomorphism. We call A an *R-algebra* via φ if $\varphi(R) \subset Z(A)$. If B is another R-algebra via τ, then a ring homomorphism $\psi : A \to B$ is called an *R-algebra homomorphism* if



commutes. Such a τ is called an *R-algebra monomorphism* (respectively, an *R-algebra epimorphism*, *R-algebra isomorphism*) if it is a ring monomorphism (respectively, epimorphism, isomorphism).

We need the analog of normal subgroups in ring theory, i.e., those subobjects of a ring that are kernels of ring homomorphisms. If a ring homomorphism $\varphi : R \to S$ satisfies $\varphi(1_R) = 0_S$, then $S = 0$ as $\varphi(1_R) = 1_S$. In particular, $\ker \varphi$ cannot be a subring of R unless $\ker \varphi = R$; so in general kernels are not subrings. They are, however, always subrngs. Just as in the case when being a subgroup of a group is not sufficient to be the kernel of a group homomorphism, neither is being a subrng sufficient. Rather we need the following objects:

Definition 23.12. Let R be a ring and $\mathfrak{A}$ a non-empty subset of R. We call $\mathfrak{A}$ a (2-*sided*) *ideal* of R if $(\mathfrak{A}, +) \subset (R, +)$ is a subgroup and for all r in R and a in $\mathfrak{A}$, we have both ar and ra lie in $\mathfrak{A}$, i.e., $\cdot : R \times \mathfrak{A} \to \mathfrak{A}$ and $\cdot : \mathfrak{A} \times R \to \mathfrak{A}$. So an ideal of R is a subrng of R closed under multiplication by R on both the right and left. If $\mathfrak{A}$ only satisfies being a subrng and closed under multiplication by R on the left (respectively, on the right), we call $\mathfrak{A}$ a *left ideal* (respectively, *right ideal*). Of course, if R is a commutative ring, the notions of (2-sided) ideal, right ideal, and left ideal coincide, making life much simpler.

Examples 23.13. Let R be a nontrivial ring.

1. $0 = \{0\}$ is an ideal of R called the *trivial ideal* and R is an ideal of R called the *unit ideal*. If these are the only ideals of R, then we call R a *simple ring*. (Cf. simple groups.)

2. Let $a \in R$. Define

$$(a) = RaR := \{\sum_{i=1}^{n} x_i a y_i \mid n \in \mathbb{Z}^+,\ x_i, y_i \in R,\ 1 \leq i \leq n\},$$

an ideal of R called the *principal ideal generated* by a. It is the smallest ideal in R containing a. [Cf. this with a cyclic subgroup of a group.] If R is commutative, this is much simpler, as (using the distributive laws), it is just

$$(a) = Ra := \{ra \mid r \in R\}.$$

3. Let $a_1, \ldots, a_n \in R$. The smallest ideal *generated* by $a_1, \ldots, a_n$ is

$$(a_1, \ldots, a_n) = Ra_1R + \cdots + Ra_nR := \{\sum_{i=1}^{n} \sum_{j=1}^{m_i} x_{ij} a_i y_{ij} \mid x_{ij}, y_{ij} \in R \text{ all } i, j \text{ some } n, m_i \in \mathbb{Z}^+\}.$$

If R is commutative, this is simply

$$(a_1, \ldots, a_n) = Ra_1 + \cdots + Ra_n := \{\sum_{i=1}^{n} r_i a_i \mid r_i \in R,\ 1 \leq i \leq n\},$$

i.e., it is all *R-linear combinations* of $a_1, \ldots, a_n$. [Cf. this with the span of finitely many vectors in a vector space.]

4. Let I be an (indexing) set, $A = \{a_i \mid i \in I\}$. Then the smallest ideal in R generated by A is

$$\langle A \rangle = (a_i)_I := \{a \in R \mid \text{There exist } a_{i_1}, \ldots, a_{i_n} \in A, \text{ some } n \text{ and } i_j \in I \text{ satisfying } a \in (a_{i_1}, \ldots, a_{i_n})\}.$$

[So $\langle A \rangle$ is the union of all the ideals generated by $a_{i_1}, \ldots, a_{i_n}$ in A with $i_1, \ldots, i_n \in I$, for some n.]

5. Let $\mathfrak{A}$ and $\mathfrak{B}$ be ideals in R. Then

$$\mathfrak{A} + \mathfrak{B} := (\mathfrak{A} + \mathfrak{B}) = \{a + b \mid a \in \mathfrak{A} \text{ and } b \in \mathfrak{B}\}$$
$$\mathfrak{A} \cap \mathfrak{B} := \{x \mid x \in \mathfrak{A} \text{ and } x \in \mathfrak{B}\}$$
$$\mathfrak{A}\mathfrak{B} := (\{ab \mid a \in \mathfrak{A},\ b \in \mathfrak{B}\}) = \{\sum_{i=1}^{n} a_i b_i \mid a_i \in \mathfrak{A},\ b_i \in \mathfrak{B},\ 1 \leq i \leq n, \text{ some } n\}$$

are ideals in R. Moreover, we have

$$\mathfrak{A}\mathfrak{B} \subset \mathfrak{A} \cap \mathfrak{B} \subset \mathfrak{A} + \mathfrak{B}$$

(and they are usually all different).

Note the formal difference between the definition of $\mathfrak{AB}$ for rings and HK, with H, K subgroups of G, although this difference vanishes if H or K is a normal subgroup (the analogue of an ideal). So $\mathfrak{AB}$ is the ideal in R generated by ab with $a \in \mathfrak{A}$, $b \in \mathfrak{B}$. More generally, we can recursively define the product of finitely many ideals. (Note that infinite products do not make sense in a ring.)

Note $\mathfrak{A} \cup \mathfrak{B}$ is usually not an ideal (cf. groups and vector spaces). Can you give a condition that will guarantee that it is?

6. If $\varphi : R \to S$ is a ring homomorphism, then $\ker \varphi$ is an ideal of R. But in general $\operatorname{im} \varphi$ is not an ideal of S. Indeed, it is an ideal if and only if φ is surjective. It is, however, a subring of S
7. If $\mathfrak{A}$ is an ideal of R and $\mathfrak{A} \cap R^\times \neq \emptyset$, then $\mathfrak{A} = R$, the unit ideal (and conversely). Indeed, if $x \in \mathfrak{A} \cap R^\times$ and $r \in R$, then $r = r1 = rx^{-1}x \in \mathfrak{A}$.

We now give some basic examples.

Specific Examples 23.14. Let R be a ring.

1. If R is a division ring, then R is simple. Indeed, if $0 < \mathfrak{A} \subset R$ is an ideal, then $\mathfrak{A} \cap R^\times \neq \emptyset$.
2. Suppose R is not the trivial ring. If R is commutative, then R is simple if and only if R is a field. Indeed if R is simple, $0 \neq a \in R$, then $(a) = R$; so there exists an $r \in R$ satisfying $1 = ar = ra$ showing a is a unit. The converse is the previous example.
3. $\mathbb{M}_n(\mathbb{R})$, for $n > 1$, is simple and not a division ring. In fact, if R is simple, so is $\mathbb{M}_n(R)$ which is never a division ring if $n > 1$. (We leave this as an exercise.)
4. Let $R = \mathbb{Z}$ and $0 < \mathfrak{A} \subset \mathbb{Z}$ an ideal. Then $(\mathfrak{A}, +) \subset (\mathbb{Z}, +)$ is a subgroup, so $\mathfrak{A} = \mathbb{Z}n$ for some $n \in \mathbb{Z}^+$ as an additive group. Clearly, $(n) = \mathbb{Z}n \subset \mathbb{Z}$ is an ideal. Therefore, every ideal in $\mathbb{Z}$ is of the form (n) for some $n \geq 0$ in $\mathbb{Z}$. In particular, every ideal in $\mathbb{Z}$ is principal, where an ideal $\mathfrak{A}$ is called *principal* if there exists an $a \in \mathfrak{A}$ such that $\mathfrak{A} = (a)$.
5. A domain R is called a *principal ideal domain* or a *PID* if every ideal in R is principal, e.g., $\mathbb{Z}$ is a PID.
6. We give further examples of PID's (without proof):
   (*i*) Any field.
   (*ii*) $F[t]$, if F is a field. (Can you show this?)

(*iii*) Let

$$\mathbb{Z}[\theta] := \{a + b\theta \mid a, b \in \mathbb{Z}\} \subset \mathbb{C}$$

with

$$\theta = \sqrt{-1},\ \sqrt{-2},\ \frac{-1-\sqrt{-3}}{2},\ \frac{-1-\sqrt{-7}}{2},\ \frac{-1-\sqrt{-11}}{2},$$
$$\frac{-1-\sqrt{-19}}{2},\ \frac{-1-\sqrt{-43}}{2},\ \frac{-1-\sqrt{-67}}{2},\ \text{or}\ \frac{-1-\sqrt{-163}}{2}.$$

Under the $+$ and $\cdot$ in $\mathbb{C}$, these are rings and each is a PID.

Gauss had conjectured that these were the only PID's of this form (with negative square roots). It was proved by Harold Stark.

7. $\mathbb{Z}[\sqrt{-5}] := \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\} \subset \mathbb{C}$ is a domain but not a PID.
8. $\mathbb{Z}[t]$ is a domain but not a PID.

We begin our study by generalizing concepts that we investigated when studying the integers.

Definition 23.15. Let R be a commutative ring, a, b in R with $a \neq 0$. We say that a *divides* b and write $a \mid b$, if there exists an x in R satisfying $b = ax$.

Properties 23.16. Let R be a commutative ring and a, b, c elements in R.

(1) If $a \mid b$ and $a \mid c$, then $a \mid bx + cy$ for all x, y in R.
(2) If R is a domain, then $a \mid b$ and $b \mid a$ if and only if there exists a unit u in R satisfying $a = ub$.

Note the second property generalizes the corresponding result for the integers, as ± 1 are the only units in $\mathbb{Z}$.

Proof. The proof of the first property is the same as for the integers, so we need only prove the second. If $a = ub$ for a unit u then $b = u^{-1}a$. Conversely, if $a \mid b$ and $b \mid a$ then there exist $x, y \in R$ satisfying $b = ax$ and $a = by$. As $a \neq 0$ in the domain R and $a = by = axy$, we have $1 = xy$ by the Cancellation Law, so $x, y \in R^\times$. □

Property (2) motivates the following definition: If R is a domain, a and b nonzero elements in R, we say that a is an *associate* of b and write $a \approx b$ if there exists a unit u in R such that $a = ub$. Thus a is an associate of b if and only if $a \mid b$ and $b \mid a$. (Check that $\approx$ is an equivalence relation.) For example, if $R = \mathbb{Z}$, then the only associates of n in $\mathbb{Z}$ are $\pm n$. (We also let 0 be the associate of 0.)

Remarks 23.17. Let R be a commutative ring. Then the relationship between division and principal ideal is given by the following: Let a, b, and c be nonzero elements in R.

1. $a \in (b)$ if and only if $b \mid a$ if and only if $(a) \subset (b)$.
2. $(a) = (b)$ if and only if $a \mid b$ and $b \mid a$.
3. $a \mid b$ and $a \mid c$ if and only if $(b, c) \subset (a)$.
4. $(b, c) = (a)$ if and only if $a \mid b$ and $a \mid c$ and there exist x and y in R satisfying $a = bx + cy$.
5. Suppose that R is a domain, then
   (a) $a \approx b$ if and only if $(a) = (b)$.
   (b) If $a = rb$ in R, then $a \approx b$ if and only if $r \in R^\times$. In particular, if $a \not\approx b$ and $a = rb$ in R, then $(a) < (b)$. Indeed, if $ub = a = rb$ with $u \in R^\times$, then $u = r$ as R is a domain.

If $R = \mathbb{Z}$ and $p \neq \pm 1$ is an integer, we know that p is a prime if and only if $p \mid ab$ implies $p \mid a$ or $p \mid b$. This motivates the most important definition concerning ideals.

Definition 23.18. Let R be a commutative ring and $\mathfrak{A} < R$ an ideal. We call $\mathfrak{A}$ a *prime ideal* in R if

$$ab \in \mathfrak{A} \text{ implies that } a \in \mathfrak{A} \text{ or } b \in \mathfrak{A}.$$

It is called a *maximal ideal* in R if

$$\mathfrak{A} < \mathfrak{B} \subset R \text{ is an ideal, then } \mathfrak{B} = R.$$

[Note this definition of maximal ideal (even left maximal ideal — definition?) also makes sense in any ring, commutative or not.]

Examples 23.19. Let R be a nontrivial commutative ring.

1. R is a domain if and only if $0 = (0)$ is a prime ideal in R:
   We have 0 is a prime ideal if and only if $ab \in (0)$ implies $a \in (0)$ or $b \in (0)$ if and only if $ab = 0$ implies $a = 0$ or $b = 0$.
2. Every maximal ideal $\mathfrak{m} < R$ is a prime ideal:
   Let $ab \in \mathfrak{m}$ with $a \notin \mathfrak{m}$. Then $\mathfrak{m} < \mathfrak{m} + Ra \subset R$ is an ideal, hence $R = \mathfrak{m} + Ra$. Therefore, there exist $m \in \mathfrak{m}$ and $r \in R$ satisfying $1 = m + ra$. Hence $b = bm + rab \in \mathfrak{m} + \mathfrak{m} \subset \mathfrak{m}$.
3. Suppose that R is a PID. Then every nonzero prime ideal in R is maximal:
   [Warning: This is not true in general, e.g., in $\mathbb{Z}[t]$ – why?.]
   Let $0 < \mathfrak{p} < R$ be a prime ideal and suppose that $\mathfrak{p} < \mathfrak{A} \subset R$ is an ideal. By the definition of a PID, there exist $p \in \mathfrak{p}$ and $a \in \mathfrak{A}$ satisfying $\mathfrak{p} = (p)$ and $\mathfrak{A} = (a)$. Therefore, we have $p = ra$ for some

$r \in R$ and $a \notin (p)$. As $ra = p \in \mathfrak{p}$, we have $r \in \mathfrak{p}$. Write $r = sp$, with $s \in R$. Then $p = ra = psa$ in the domain R, so $1 = sa$. Hence $a \in R^\times$ and $\mathfrak{A} = R$. This shows that $\mathfrak{p}$ is maximal.

4. The prime ideals in $\mathbb{Z}$ are:

$$\begin{aligned} 0 = (0) &\qquad \text{prime, not maximal} \\ p\mathbb{Z} = (p) &\qquad p \text{ a prime, maximal.} \end{aligned}$$

We want to view prime ideals as an ideal theoretical property rather than a property about elements in the ideal. This is easily done and is the direct analogue of Euclid's Lemma for ideals.

Lemma 23.20. *Let R be a commutative ring and $\mathfrak{p} < R$ an ideal. Then $\mathfrak{p}$ is a prime ideal if and only if for all ideals $\mathfrak{A}$ and $\mathfrak{B}$ in R satisfying $\mathfrak{AB} \subset \mathfrak{p}$, either $\mathfrak{A} \subset \mathfrak{p}$ or $\mathfrak{B} \subset \mathfrak{p}$. Moreover, if $\mathfrak{A} < R$ is not a prime ideal, then there exists ideals $\mathfrak{A} < \mathfrak{B}_i$ in R with $i = 1, 2$ satisfying $\mathfrak{B}_1\mathfrak{B}_2 \subset \mathfrak{A}$.*

PROOF. ($\Leftarrow$): If $xy \in \mathfrak{p}$ with $x, y \in R$ then $(x)(y) \subset \mathfrak{p}$, so either $x \in (x) \subset \mathfrak{p}$ or $y \in (y) \subset \mathfrak{p}$.

($\Rightarrow$): Suppose that the ideal $\mathfrak{A} \not\subset \mathfrak{p}$ but $\mathfrak{AB} \subset \mathfrak{p}$ with $\mathfrak{B}$ an ideal. Then there exists an element $a \in \mathfrak{A} \setminus \mathfrak{p}$. As $ab \in \mathfrak{p}$ for all $b \in \mathfrak{B}$, it follows that $\mathfrak{B} \subset \mathfrak{p}$.

As for the last statement, suppose that $\mathfrak{A} < R$ is not a prime ideal. Then there exist ideals $\mathfrak{B}_i \not\subset \mathfrak{A}$, $i = 1, 2$ but $\mathfrak{B}_1\mathfrak{B}_2 \subset \mathfrak{A}$. Let $\mathfrak{B}_i{}' = \mathfrak{B}_i + \mathfrak{A} > \mathfrak{A}$, then $\mathfrak{B}_1{}'\mathfrak{B}_2{}' \subset \mathfrak{B}_1\mathfrak{B}_2 + \mathfrak{A} \subset \mathfrak{A}$. □

Exercises 23.21. 1. Prove if R is a domain so is the ring of formal power series $R[[t]]$.

2. Let R be a commutative ring. Show that if $f = 1 + \sum_{i=1}^{\infty} a_i$ is a formal power series in $R[[t]]$, then one can determine $b_1, \ldots, b_n, \ldots$ such that $g = 1 + \sum_{i=1}^{\infty} b_i$ is the multiplicative inverse of f in $R[[t]]$. In particular,

$$R[[t]]^\times = \{a_0 + \sum_{i=1}^{\infty} a_i \in R[[t]] \mid a_0 \in R^\times\}.$$

3. Let R be a commutative ring and $A = R[[t]]$. Show that

$$\{\mathfrak{p}A \mid \mathfrak{p} \text{ is a prime ideal in R}\} \cup \{\mathfrak{p}A + tA \mid \mathfrak{p} \text{ is a prime ideal in R}\}$$

is the set of prime ideals in A and

$$\{\mathfrak{m}A + tA \mid \mathfrak{m} \text{ is a maximal ideal in R}\}$$

is the set of maximal ideals in A. In particular, if R has precisely one maximal ideal, then so does A. [A commutative ring having precisely one maximal ideal is called a *local ring*.]

4. Let A be an additive group and let $\mathrm{End}(A)$ denote the set of group homomorphisms of A to A. Prove that $\mathrm{End}(A)$ is a ring under addition and composition of elements in $\mathrm{End}(A)$. Also show the unit group of $\mathrm{End}(A)$ is the group of group automorphisms of A.
5. Let F be a field, then $\mathbb{M}_n(F)$ is a simple ring. If $n > 1$ then $\mathbb{M}_n(F)$ is not a division ring.
6. Show that a ring homomorphism $\varphi : R \to S$ is a monomorphism if and only if given any ring homomorphisms $\psi_1, \psi_2 : T \to R$ with compositions satisfying $\varphi \circ \psi_1 = \varphi \circ \psi_2$, then $\psi_1 = \psi_2$.
7. Show if R is a commutative ring, then, for any ideal $\mathfrak{B}$ in $\mathbb{M}_n(R)$, there exists an ideal $\mathfrak{A}$ in R satisfying $\mathfrak{B} = \mathbb{M}_n(\mathfrak{A})$ (obvious definition). In particular, if R is simple, so is $\mathbb{M}_n(R)$. Show that $\mathbb{M}_n(R)$ is never a division ring if $n > 1$.
8. Let R be a commutative ring. Prove the *Binomial Theorem*: Let a and b be elements of R and n a positive integer, then
$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}.$$
9. If R is a ring satisfying $x^2 = x$ for all x in R, then R is commutative.
10. If R is a rng satisfying $x^3 = x$ for all x in R, then R is commutative.
11. Let R be a commutative ring and $\mathfrak{A}$ be an ideal in R satisfying
$$\mathfrak{A} = \mathfrak{m}_1 \cdots \mathfrak{m}_r = \mathfrak{n}_1 \cdots \mathfrak{n}_s$$
with all the $\mathfrak{m}_i$ distinct maximal ideals and all the $\mathfrak{n}_j$ distinct maximal ideals. Show that $r = s$ and there exists a $\sigma \in S_r$ satisfying $\mathfrak{m}_i = \mathfrak{n}_{\sigma(i)}$ for all i.
12. Let R be a commutative ring and $\mathfrak{A}$ an ideal of R. Suppose that every element in $R \setminus \mathfrak{A}$ is a unit of R. Show that $\mathfrak{A}$ is a maximal ideal of R and that, moreover, it is the only maximal ideal of R.
13. Let R be the set of all continuous functions $f : [0,1] \to \mathbb{R}$, then R is a commutative ring under $+$ and $\cdot$ of functions. Show that any maximal ideal of R has the form $\{f \in R \mid f(a) = 0\}$ for some fixed a in $[0,1]$.
14. Let $\varphi : R \to S$ be a ring homomorphism of commutative rings. Show that if $\mathfrak{B}$ is an ideal (respectively, prime ideal) of S then $\varphi^{-1}(\mathfrak{B})$ is an ideal (respectively, prime ideal) of R. Give an example, where $\mathfrak{B}$ is a maximal ideal of S but $\varphi^{-1}(\mathfrak{B})$ is not a maximal ideal of R.

15. Show $F[t]$, with F a field, is a PID.
16. Show that $\mathbb{Z}[t]$, the ring of polynomials with coefficients in $\mathbb{Z}$, is not a PID. Show this by finding a non-principal maximal ideal in $\mathbb{Z}[t]$. Also find a nonzero prime ideal in $\mathbb{Z}[t]$ that is not maximal and prove it is such.
17. Let $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$ be ideals in R, at least $n-2$ of which are prime. Let $S \subset R$ be a subrng (it does not have to have a 1) contained in $\mathfrak{A}_1 \cup \cdots \cup \mathfrak{A}_n$. Then one of the $\mathfrak{A}_j$'s contains S. In particular, if $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ are prime ideals in R and $\mathfrak{B}$ is an ideal properly contained in S satisfying $S \setminus \mathfrak{B} \subset \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n$, then S lies in one of the $\mathfrak{p}_i$'s.

24. Factor Rings and Rings of Quotients

We begin by generalizing the construction of the integers modulo n.

Definition 24.1. Let R be a ring, $\mathfrak{A}$ an ideal in R. If a and b are elements in R, write $a \equiv b \mod \mathfrak{A}$ if $a - b$ is an element of $\mathfrak{A}$.

A proof analogous proof to that for the integers modulo n, establishes the following:

Proposition 24.2. *Let R be a ring and $\mathfrak{A}$ an ideal in R. Then $\equiv \mod \mathfrak{A}$ is an equivalence relation. Suppose that the elements a, a', b, b' in R satisfy:*

$$a \equiv a' \mod \mathfrak{A} \qquad \text{and} \qquad b \equiv b' \mod \mathfrak{A},$$

then

$$\begin{aligned} a + b &\equiv a' + b' \mod \mathfrak{A}, \\ a \cdot b &\equiv a' \cdot b' \mod \mathfrak{A}. \end{aligned}$$

Let $\overline{R} = R/\mathfrak{A} := R/(\equiv \mod \mathfrak{A})$. If $a \in R$, set

$$\overline{a} = \{b \in R \mid b \equiv a \mod \mathfrak{A}\} = a + \mathfrak{A}.$$

If

$$^{-} : R \to R/\mathfrak{A} \text{ given by } a \mapsto \overline{a}$$

denotes the canonical surjection and a, b lie in R, define

(i) $$\overline{a} + \overline{b} := \overline{a + b}$$

(ii) $$\overline{a} \cdot \overline{b} := \overline{a \cdot b}.$$

Then under these operations, $\overline{R}$ is a ring with $0_{\overline{R}} = \mathfrak{A}$ and $1_{\overline{R}} = 1_R + \mathfrak{A}$ with $\overline{R}$ commutative if R is. The canonical surjection is an epimorphism with kernel $\mathfrak{A}$. We call $\overline{R}$ the factor or quotient ring of R by $\mathfrak{A}$.

Proofs analogous to those in Group Theory show:

Theorem 24.3. (First Isomorphism Theorem) *Let $\varphi : R \to S$ be a ring homomorphism. Then we have a commutative diagram of rings and ring homomorphisms*

$$\begin{array}{ccc} R & \xrightarrow{\varphi} & S \\ {\scriptstyle -}\downarrow & & \uparrow{\scriptstyle \mathrm{inc}} \\ R/\ker\varphi & \xrightarrow[\overline{\varphi}]{} & \operatorname{im}\varphi \end{array}$$

with $^-$ a ring epimorphism, $\overline{\varphi}$ a ring isomorphism between $R/\ker\varphi$ and $\operatorname{im}\varphi$, and inc *a ring monomorphism.*

and

Theorem 24.4. (Correspondence Principle) *Let $\varphi : R \to S$ be a ring epimorphism. Then*

$$\{\mathfrak{B} \mid \mathfrak{B} \text{ an ideal in } R \text{ with } \ker\varphi \subset \mathfrak{B}\} \longrightarrow \{\mathfrak{C} \mid \mathfrak{C} \text{ an ideal in } S\}$$

given by $\mathfrak{B} \mapsto \varphi(\mathfrak{B})$ is an order preserving bijection.

We give some applications.

Proposition 24.5. *Let R be a nontrivial commutative ring and $\mathfrak{A}$ an ideal of R. Then*

(1) $\mathfrak{A}$ is a prime ideal of R if and only if $R/\mathfrak{A}$ is a domain.
(2) $\mathfrak{A}$ is a maximal ideal of R if and only if $R/\mathfrak{A}$ is a field.

Proof. Let $^- : R \to R/\mathfrak{A}$ be the canonical epimorphism, so $\ker{}^- = \mathfrak{A}$.

(1): $\overline{R}$ is a domain if and only if $(\overline{0})$ is a prime ideal in $\overline{R}$ if and only if whenever $a, b \in R$ satisfy $\overline{a} \cdot \overline{b} = \overline{0}$, then $\overline{a} \in (\overline{0})$ or $\overline{b} \in (\overline{0})$ if and only if whenever $a, b \in R$ satisfy $a \cdot b \in \mathfrak{A}$, then $a \in \mathfrak{A}$ or $b \in \mathfrak{A}$ if and only if $\mathfrak{A}$ is a prime ideal.

(2): $(\overline{0}) \neq \overline{R}$ is a field if and only if $\overline{R}$ is simple if and only if $(\overline{0})$ and $\overline{R}$ are the only ideals of $\overline{R}$ if and only if $\{\mathfrak{B} \mid \mathfrak{A} \subset \mathfrak{B} \subset R \text{ is an ideal}\} = \{\mathfrak{A}, R\}$ if and only if $\mathfrak{A} < R$ is maximal by the Correspondence Principle. □

Question Is (2) true if the ring R is not necessarily commutative and we replace field by division ring?

Compare the next application with the Third Isomorphism Theorem of Groups.

Proposition 24.6. *Let R be a ring and $\mathfrak{A} \subset \mathfrak{B} \subset R$ ideals. Then*

$$\mathfrak{B}/\mathfrak{A} := \{b + \mathfrak{A} \mid b \in \mathfrak{B}\} \subset R/\mathfrak{A}$$

is an ideal in $R/\mathfrak{A}$ and

$$R/\mathfrak{B} \cong (R/\mathfrak{A})/(\mathfrak{B}/\mathfrak{A}).$$

PROOF. It is easy to check that $R/\mathfrak{A} \to R/\mathfrak{B}$ given by $r + \mathfrak{A} \mapsto r + \mathfrak{B}$ is a well-defined epimorphism with kernel $\mathfrak{B}/\mathfrak{A}$. □

Proposition 24.7. *Let $\varphi : R \to S$ be a ring epimorphism of commutative rings with kernel $\mathfrak{A}$. Then the map*

$$\{\mathfrak{p} \mid \mathfrak{p} < R \text{ a prime ideal with } \mathfrak{A} \subset \mathfrak{p}\} \to \{\mathfrak{P} \mid \mathfrak{P} < S \text{ a prime ideal}\}$$

given by $\mathfrak{p} \mapsto \varphi(\mathfrak{p})$ is a bijection.

PROOF. We may assume that S is not the trivial ring for if not then both sets of prime ideals would be empty. By the First Isomorphism Theorem, we may assume that $S = R/\mathfrak{A}$ and φ is the canonical epimorphism $^- : R \to R/\mathfrak{A}$. Let $\mathfrak{A} \subset \mathfrak{B} \subset R$ be an ideal. Then $\mathfrak{B}$ is a prime ideal in R if and only if $R/\mathfrak{B}$ is a domain if and only if $(R/\mathfrak{A})/(\mathfrak{B}/\mathfrak{A})$ is a domain if and only if $\mathfrak{B}/\mathfrak{A}$ is a prime ideal in $R/\mathfrak{A}$. □

Warning 24.8. If $\varphi : R \to S$ is a ring homomorphism of rings with $\mathfrak{p}$ a prime ideal in R, it does not follow that the ideal generated by image of $\mathfrak{p}$, $S\varphi(\mathfrak{p})$, is a prime ideal in S. For example, the inclusion $\mathbb{Z} \subset \mathbb{Q}$ is a ring homomorphism and (2) is a prime ideal in $\mathbb{Z}$, but its image (2) is the unit ideal in $\mathbb{Q}$.

Construction 24.9. Let R be a nontrivial ring. As any ring homomorphism must take the multiplicative identity to the multiplicative identity, there exists a unique ring homomorphism from $\mathbb{Z}$ to R. It is given by

$$\iota : \mathbb{Z} \to R \text{ given by } m \mapsto m1_R = \begin{cases} \underbrace{1_R + \cdots + 1_R}_{m} & \text{if } m \geq 0 \\ \underbrace{-1_R + \cdots + -1_R}_{-m} & \text{if } m < 0. \end{cases}$$

Since $\mathbb{Z}$ is a PID, there exists a unique integer $n \geq 0$ such that $\ker \iota = (n) [= (-n)]$. We call n the *characteristic* of the ring R and write $\operatorname{char}(R)$ (or just $\operatorname{char} R$) for this integer. We shall only be interested in the characteristic of commutative rings.

Remarks 24.10. Let R be a commutative ring and $\iota : \mathbb{Z} \to R$ the unique ring homomorphism.

1. $\operatorname{char}(R) = 0$ if and only if $m1_R \neq 0$ for all nonzero integers m if and only if ι is monic.
2. If $\operatorname{char}(R) \neq 0$ then, by the Division Algorithm, $\operatorname{char}(R)$ is the least positive integer n such that $n1_R = 0$.
3. If $\varphi : \mathbb{Z}/n\mathbb{Z} \to R$ and $\psi : \mathbb{Z}/m\mathbb{Z} \to R$ are ring monomorphisms with n, m positive integers, then $n = \operatorname{char}(R) = m$ and $\varphi = \psi$ are induced by ι.

Examples 24.11. 1. $\mathbb{Z}$ is a domain, and $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields of characteristic zero.

2. If $n \in \mathbb{Z}^+$ then $\operatorname{char}(\mathbb{Z}/n\mathbb{Z}) = n$.
3. If $\varphi : R \to S$ is a ring monomorphism, then $\operatorname{char}(R) = \operatorname{char}(S)$.
4. The canonical epimorphism $^- : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ shows that, in general, the characteristic of a ring is not preserved under a ring homomorphism.

Proposition 24.12. *Let R be a domain. Then either*

(1) $\operatorname{char}(R) = 0$ *and there exists a ring monomorphism $\mathbb{Z} \to R$ or*
(2) $\operatorname{char}(R) = p$*, p a prime, and there exists a ring monomorphism $\mathbb{Z}/p\mathbb{Z} \to R$.*

PROOF. Let $\iota : \mathbb{Z} \to R$ be the unique ring homomorphism, then $\ker \varphi = (n)$ for some non-negative integer n, so $\operatorname{char}(R) = n$. As R is a domain so is $\operatorname{im} \iota \cong \mathbb{Z}/n\mathbb{Z}$. Therefore, $n = 0$ or n is a prime and the induced map $\bar{\iota} : \mathbb{Z}/n\mathbb{Z} \to R$ is a monomorphism. □

The proposition says that if R is a domain, we may view $\mathbb{Z} \subset R$ if $\operatorname{char}(R) = 0$ and $\mathbb{Z}/p\mathbb{Z} \subset R$ if $\operatorname{char}(R) = p$.

We have now generalized the construction of $\mathbb{Z}/n\mathbb{Z}$ to $R/\mathfrak{A}$, where $\mathfrak{A}$ is an ideal in R. Our next goal is to construct $\mathbb{Q}$ from $\mathbb{Z}$. Our construction will work for any domain.

Construction 24.13. Let R be a domain. We wish to construct fractions from R, i.e., if $a, b, c, d \in R$ with $b \neq 0$ and $d \neq 0$, we want to have the following:

(i) $$\frac{a}{b}+\frac{c}{d}=\frac{ad+bc}{bd}.$$ [Note $bd \neq 0$ in the domain R.]

(ii) $$\frac{a}{b}\cdot\frac{c}{d}=\frac{ac}{bd},$$

and we know that we must have

(iii) $$\frac{a}{b}=\frac{c}{d} \text{ if and only if } ad=bc \text{ in } R.$$

We do this construction as follows: Let

$$\mathcal{S}=\{(a,b) \mid a\in R,\, 0\neq b\in R\}=R\times(R\setminus\{0\})$$

and

$$(a,b)\sim(c,d) \text{ in } \mathcal{S} \text{ if } ad=bc.$$

Check that $\sim$ is an equivalence relation.

Let

$$\frac{a}{b}:=[(a,b)]_{\sim}, \text{ the equivalence class of } (a,b) \text{ under } \sim \text{ and}$$

$$K:=\mathcal{S}/\!\sim=\{\frac{a}{b} \mid (a,b)\in\mathcal{S}\}=\{\frac{a}{b} \mid a,b\in R,\, b\neq 0\}.$$

Then K satisfies (iii) above. Define $+ : K\times K\to K$ and $\cdot : K\times K\to K$ by (i) and (ii) above respectively.

Claim. $+$ and $\cdot$ are well-defined:

Suppose that $(a,b)\sim(a',b')$ and $(c,d)\sim(c',d')$ in $\mathcal{S}$. We must show

$$(ad+bc)b'd'=(a'd'+b'c')bd$$
$$(ac)(b'd')=(a'c')(bd)$$

knowing that

$$ab'=a'b \text{ and } cd'=c'd,$$

which is easily checked. Thus K has a $+$ and $\cdot$. It is routine to check that K is a commutative ring with $0_K=\frac{0}{b}$, and $1_K=\frac{b}{b}$ with $0\neq b\in R$. If $\frac{a}{b}\neq 0_K$ then $a\neq 0$ in R, hence $\frac{b}{a}$ is defined and is the inverse of $\frac{a}{b}$, so K is a field called the *field of quotients* or the *quotient field* of R. We write $qf(R)$ for K.

Example. $\mathbb{Q}=qf(\mathbb{Z})$.

This construction of $qf(R)$ is essentially unique as $qf(R)$ has the following *universal property*:

Theorem 24.14. (The Universal Property of the Field of Quotients) *Let R be a domain. Then there exists a field K, and a ring monomorphism $i : R \to K$ satisfying the following Universal Property: If F is a field, then for any ring monomorphism $\varphi : R \to F$, there exists a unique ring monomorphism $\psi : K \to F$ such that the diagram*

(†)
$$\begin{array}{ccc} R & \xrightarrow{\ i\ } & K \\ & \searrow^{\varphi} & \downarrow{\scriptstyle \psi} \\ & & F \end{array}$$

commutes.

In particular, if K' is a field and $j : R \to K'$ a ring monomorphism also satisfying the universal property (†), then there exists a unique (field) isomorphism $\sigma : K \to K'$ such that the diagram

$$\begin{array}{ccc} R & \xrightarrow{\ i\ } & K \\ & \searrow^{j} & \downarrow{\scriptstyle \sigma} \\ & & K' \end{array}$$

commutes.

Proof. Let $K = qf(R)$, then $i : R \to K$ given by $r \mapsto \frac{r}{1}$ is a ring homomorphism. Let b be a nonzero element in R. Then $\frac{r}{1} = \frac{0}{b}$ in K if and only if $rb = 1 \cdot 0 = 0$ in the domain R if and only if $r = 0$, so this homomorphism is injective.

Now suppose that we are given a ring monomorphism $\varphi : R \to F$ with F a field. Define

$$\psi : K \to F \text{ by } \psi(\frac{a}{b}) = \varphi(a)\varphi(b)^{-1}.$$

This makes sense, as $\varphi(b) \neq 0$, since φ is monic and F is a field.

Claim 1. ψ is well-defined and monic.

$\frac{a}{b} = \frac{a'}{b'}$ in K if and only if $ab' = a'b$ in R if and only if $\varphi(a)\varphi(b') = \varphi(a')\varphi(b)$ in F as φ is monic if and only if $\varphi(a)\varphi(b)^{-1} = \varphi(a')\varphi(b')^{-1}$ in the field F if and only if $\psi(\frac{a}{b}) = \psi(\frac{a'}{b'})$ in F. It is easily seen that ψ is a homomorphism. The Claim follows.

Claim 2. ψ is unique.

If ψ' is another monomorphism making the diagram (†) commute, then

$$\psi'(\frac{a}{1}) = \varphi(a) = \psi(\frac{a}{1_K}) \text{ for all } a \in R$$

in the field F. In particular, if $b \neq 0$ in R, we have

$$\psi'(\frac{b}{1})\psi'(\frac{1}{b}) = \psi'(1) = 1_F = \psi(\frac{b}{1})\psi(\frac{1}{b})$$

in F, so

$$\psi'(\frac{1}{b}) = \psi(\frac{1}{b}),$$

hence, for all a and b in R with $b \neq 0$,

$$\psi'(\frac{a}{b}) = \psi'(\frac{a}{1})\psi'(\frac{1}{b}) = \psi(\frac{a}{1})\psi(\frac{1}{b}) = \psi(\frac{a}{b}).$$

This proves the second claim. Finally, we show:

Claim 3. The map $i : R \to K$ satisfies the uniqueness property relative to the diagram (†).

Suppose that $j : R \to K'$ also satisfies (†). Then there exist unique ring monomorphisms $\psi : K \to K'$ and $\psi' : K' \to K$ such that we have commutative diagrams

$$\begin{array}{ccc} R & \xrightarrow{i} & K \\ & \searrow^{j} & \downarrow{\scriptstyle \psi} \\ & & K' \end{array} \qquad \text{and} \qquad \begin{array}{ccc} R & \xrightarrow{j} & K' \\ & \searrow^{i} & \downarrow{\scriptstyle \psi'} \\ & & K. \end{array}$$

By (†), we have $\psi'\psi = 1_K$ and $\psi\psi' = 1_{K'}$ as

$$\begin{array}{ccc} R & \xrightarrow{i} & K \\ & \searrow^{i} & \downarrow{\scriptstyle 1_K} \\ & & K \end{array} \qquad \text{and} \qquad \begin{array}{ccc} R & \xrightarrow{j} & K' \\ & \searrow^{j} & \downarrow{\scriptstyle 1_{K'}} \\ & & K' \end{array} \qquad \text{commute.}$$

It follows that ψ and ψ' are inverse isomorphisms. □

Notation 24.15. If R is a domain, we shall always view $R \subset qf(R)$, i.e., we shall identify R with its image $\{\frac{r}{1} \mid r \in R\}$ in $qf(R)$, e.g., $\mathbb{Z} \subset \mathbb{Q}$.

Remarks 24.16. In general, it is very difficult to define a map between objects. We may have an idea where generators of the domain object must go, but to show the relations of the domain are preserved usually presents problems, i.e., to show the putative map is well-defined. Universal properties are methods to address this in certain cases. One tries to construct an object with given properties together with a map, so that there is automatically a map (that preserves the desired structure) between it and any other other object that satisfies the given properties. Moreover, that map is unique. Therefore, if you have such

a universal property, you automatically obtain unique maps to other appropriate objects. The above is an example of such. Another, that you have seen in linear algebra, is given a vector space together with a given basis, any linear transformation from it to another vector space is completely determined where the basis maps.

The following is a useful observation:

Observation 24.17. Let F be a field and R a nontrivial ring. If $\varphi : F \to R$ is a ring homomorphism, then φ is monic as $\ker \varphi < F$ and F is simple.

If F is a field, then we can say a bit more.

Corollary 24.18. *Let F be a field of characteristic n.*

(1) If $n = 0$, then there exists a (unique) monomorphism $\mathbb{Q} \to F$.
(2) If $n > 0$, then $n = p$ is a prime and there exists a (unique) monomorphism $\mathbb{Z}/p\mathbb{Z} \to F$.

If K is a field, there exists a unique smallest subfield Δ_K of K, viz., the intersection of all the subfields of K. This field is called the *prime subfield* of K. By the corollary, we know that

$$\Delta_K \cong \begin{cases} \mathbb{Q} & \text{if } \operatorname{char}(K) = 0. \\ \mathbb{Z}/p\mathbb{Z} & \text{if } \operatorname{char}(K) = p. \end{cases}$$

For example, let K be a finite field with N elements. Then there exists a unique prime p such that $\operatorname{char} K = p$ and a unique monomorphism $\mathbb{Z}/p\mathbb{Z} \to K$. We also have K is a finite dimensional vector space over $\mathbb{Z}/p\mathbb{Z}$, say of dimension n. So $N = p^n$. We know that $K^\times$ is a group of order $N - 1$, so $x^{N-1} = 1$ for all x in $K^\times$ and $x^N = x$ for all $x \in K$. This shows that the analogue of Fermat's Little Theorem holds for all finite fields. We shall show later that $K^\times$ is, in fact, a cyclic group as well as showing that for every prime p and positive integer m there exists a field of cardinality p^m, and that this field is unique up to an isomorphism of fields. We also point out that there are many infinite fields of characteristic $p > 0$. Indeed, the ring $(\mathbb{Z}/p\mathbb{Z})[t]$ is an infinite domain (why?), hence so is its quotient field.

Next we turn to a generalization of the Chinese Remainder Theorem.

If R_i, $i \in I$, are rings, then the cartesian product of the R_i, $R = \bigtimes_I R_i$ becomes a ring via componentwise operations. We also write this as $\prod R_i$. It is commutative if all the R_i are commutative. It is easy to check that $R^\times = \prod_I (R_i)^\times$. [As usual, we are sloppy about writing elements in $\prod_I R_i$.]

Theorem 24.19. (Chinese Remainder Theorem) *Suppose that R is a commutative ring with $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$ ideals in R that are comaximal, i.e., $\mathfrak{A}_i + \mathfrak{A}_j = R$ for all $i, j = 1, \ldots, n$ with $i \neq j$. Let*

$$\varphi : R \to \prod_{i=1}^{n} R/\mathfrak{A}_i \text{ be given by } r \mapsto (r + \mathfrak{A}_1, \ldots, r + \mathfrak{A}_n).$$

Then φ is a ring epimorphism with

$$\ker \varphi = \mathfrak{A}_1 \cap \cdots \cap \mathfrak{A}_n = \mathfrak{A}_1 \cdots \mathfrak{A}_n,$$

hence induces isomorphisms

$$\overline{\varphi} : R/(\mathfrak{A}_1 \cdots \mathfrak{A}_n) \to R/\mathfrak{A}_1 \times \cdots \times R/\mathfrak{A}_n$$
$$\overline{\varphi} : (R/(\mathfrak{A}_1 \cdots \mathfrak{A}_n)^\times \to (R/\mathfrak{A}_1)^\times \times \cdots \times (R/\mathfrak{A}_n)^\times.$$

PROOF. Clearly we have φ is a ring homomorphism with $\ker \varphi = \mathfrak{A}_1 \cap \cdots \cap \mathfrak{A}_n$. By this and the First Isomorphism Theorem, we need only show that $\mathfrak{A}_1 \cap \cdots \cap \mathfrak{A}_n = \mathfrak{A}_1 \cdots \mathfrak{A}_n$ and φ is onto, i.e., if $x_1, \ldots, x_n$ lie in R, there exists an $x \in R$ satisfying $x \equiv x_i \mod \mathfrak{A}_i$ for all i. We prove this by induction on n.

$n = 2$: We have $R = \mathfrak{A}_1 + \mathfrak{A}_2$, so $1 = a_1 + a_2$ for some $a_i \in \mathfrak{A}_i$, $i = 1, 2$. Therefore, $a_1 \equiv 1 \mod \mathfrak{A}_2$ and $a_2 \equiv 1 \mod \mathfrak{A}_2$; hence $x = x_1 a_2 + x_2 a_1 \equiv x_i \mod \mathfrak{A}_i$ for $i = 1, 2$. As $\mathfrak{A}_1\mathfrak{A}_2 \subset \mathfrak{A}_1 \cap \mathfrak{A}_2$, we need only show that $\mathfrak{A}_1 \cap \mathfrak{A}_2 \subset \mathfrak{A}_1\mathfrak{A}_2$ to complete the $n = 2$ case. If $b \in \mathfrak{A}_1 \cap \mathfrak{A}_2$, then $b = b \cdot 1 = ba_1 + ba_2 \in \mathfrak{A}_2\mathfrak{A}_1 + \mathfrak{A}_1\mathfrak{A}_2 \subset \mathfrak{A}_1\mathfrak{A}_2$.

[**Warning**: If $\mathfrak{B}_1$, $\mathfrak{B}_2$ are ideals in a ring, usually $\mathfrak{B}_1\mathfrak{B}_2 < \mathfrak{B}_1 \cap \mathfrak{B}_2$.]

$n > 2$: By hypothesis, we have $R = \mathfrak{A}_1 + \mathfrak{A}_i$ for $i > 1$, so $1 = a_i + b_i$ for some $a_i \in \mathfrak{A}_1$ and $b_i \in \mathfrak{A}_i$ for each $i > 1$. Consequently, we have $1 = \prod_{i=2}^{n}(a_i + b_i)$ lies in $\mathfrak{A}_1 + (\mathfrak{A}_2 \cdots \mathfrak{A}_n)$ (why?), so $R = \mathfrak{A}_1 + (\mathfrak{A}_2 \cdots \mathfrak{A}_n)$. By the $n = 2$ case and induction, we conclude that $\mathfrak{A}_1 \cap \cdots \cap \mathfrak{A}_n = \mathfrak{A}_1 \cap (\mathfrak{A}_2 \cdots \mathfrak{A}_n) = \mathfrak{A}_1 \cdots \mathfrak{A}_n$.

By induction, there exists a y in R satisfying

$$y \equiv x_i \mod \mathfrak{A}_i \text{ for } i = 2, \ldots, n;$$

and by the $n = 2$ case, there exists an x in R satisfying

$$x \equiv x_1 \mod \mathfrak{A}_1$$
$$x \equiv y \mod \mathfrak{A}_2 \cdots \mathfrak{A}_n.$$

As $\mathfrak{A}_2 \cdots \mathfrak{A}_n \subset \mathfrak{A}_i$ for $i = 2, \ldots n$, we have $x \equiv x_i \mod \mathfrak{A}_i$ for $i = 1, \ldots, n$, proving that φ is surjective.

We leave the isomorphism of multiplicative groups as an exercise. □

Exercises 24.20. 1. Let $\varphi : R \to S$ be a ring epimorphism (of rings). Do an analogous analysis before The First Isomorphism Theorem for Groups 12.1 for $\varphi : R \to S$ to see how ideals and factor rings arise naturally.

2. Let $R = (\mathbb{Z}/2\mathbb{Z})[t]$, $f = f(t) = t^2 + t + 1$, and $g = t^2 + 1$. Show all of the following:
   (*i*) $R/(f)$ is a field with four elements.
   (*ii*) $R/(g)$ is not a domain and has four elements.
   (*iii*) Neither $R/(f)$ nor $R/(g)$ is isomorphic to the ring $\mathbb{Z}/4\mathbb{Z}$.
3. Let R be a commutative ring. Suppose for every element x in R there exists an integer $n = n(x) > 1$ such that $x^n = x$. Show that every prime ideal in R is maximal.
4. (Second Isomorphism Theorem) Let $R \subset S$ be a subring, $\mathfrak{A} \subset S$ an ideal. Then $\mathfrak{A} \cap R$ is an ideal in R, $R + \mathfrak{A}$ is a subring of S, and $(R + \mathfrak{A})/\mathfrak{A} \cong R/(R \cap \mathfrak{A})$.
5. Show that the inclusion map $i : \mathbb{Z} \subset \mathbb{Q}$ satisfies $i \circ \psi_1 = i \circ \psi_2$, whenever $\psi_1, \psi_2 : R \to \mathbb{Z}$ are ring homomorphisms. This shows that a surjective ring homomorphism is not equivalent to this property. (Cf. Remark 9.5.)
6. Let R be a commutative ring of characteristic $p > 0$, p a prime. Prove that the map $R \to R$ by $x \mapsto x^p$ is a ring homomorphism. It is called the *Frobenius homomorphism*. In particular, the *Children's Binomial Theorem holds*, i.e., $(x + y)^p = x^p + y^p$ in R for all x and y in R.
7. Let R be a finite domain. Show that $\mathrm{char}(R) = p$ is a prime and R is a field. Moreover, R is a vector space over $\mathbb{Z}/p\mathbb{Z}$.
8. Show that if R is a domain, so is the polynomial ring $R[t]$. In particular, show that there exists fields properly containing the complex numbers. Does the field that you constructed have the property that every non-constant polynomial over it has a root? Prove or disprove this.
9. Let $\varphi : R \to S$ be a ring isomorphism. Show φ induces a group isomorphism $R^\times \to S^\times$ by restriction. We also write this map as φ.
10. Prove the isomorphism statement about the multiplicative groups in the Chinese Remainder Theorem 24.19

25. Zorn's Lemma

One would like to have a method of proof generalizing induction. There are many equivalent ways of doing this: Well-Ordering Principle, Transfinite Induction, Zorn's Lemma. These are also equivalent to the

Axiom of Choice, Tychonoff's Theorem, the theorem that all vector spaces have bases. In algebra, the most useful form is Zorn's Lemma, which generalizes the Well-Ordering Principle. These are independent of the other axioms in Zermelo-Frankel Set Theory, and we accept them as an axiom.

Definition 25.1. Let S be a set with a relation R on S. We call S a *partially ordered set* or *poset* (under R) if the following hold: For all a, b, c in S

(1) $a\mathrm{R}a$. [Reflexitity]

(2) If $a\mathrm{R}b$ and $b\mathrm{R}a$ then $a = b$.

(3) If $a\mathrm{R}b$ and $b\mathrm{R}c$ then $a\mathrm{R}c$. [Transitivity]

We also say (S, R) is a poset, if the relation is not obvious. So the second condition replaces symmetry in the definition of an equivalence relation. If S is a poset (under R), a, b elements in S, then a and b may be *incomparable*, i.e., neither $a\mathrm{R}b$ nor $b\mathrm{R}a$. We say that the poset S is a *chain* or *totally ordered* if for all a and b in S, we have either $a\mathrm{R}b$ or $b\mathrm{R}a$, i.e., all elements are *comparable*.

Examples 25.2. 1. Recall if T is a set, then the *power set* of T is defined by $\mathcal{P}(T) := \{A \mid A \subset T\}$. It becomes a poset under $\subset$ (or $\supset$). It is almost never a chain. (When is it?)

2. $(\mathbb{Z}, \leq)$ and $(\mathbb{R}, \geq)$ are chains.

3. As $\{1, \sqrt{-1}\}$ is a basis for $\mathbb{C}$ as a vector space over $\mathbb{R}$, for each $\alpha \in \mathbb{C}$, there exist unique $x, y \in \mathbb{R}$ such that $\alpha = x + y\sqrt{-1}$. Define the *lexicographic order* on $\mathbb{C}$ by $\alpha = x + y\sqrt{-1} \leq_L \beta = w + z\sqrt{-1}$ if $\alpha = \beta$ or $x < w$ or ($x = w$ and $y < z$) in $\mathbb{R}$. Then $(\mathbb{C}, \leq_L)$ is a chain. **Note:** If $\alpha \leq_L \beta$ and $\gamma \leq_L \delta$, then $\alpha + \gamma \leq_L \beta + \delta$, but $\alpha\gamma >_L \beta\delta$ is possible (where $>_L$ is the obvious relation), e.g., $0 \leq_L \sqrt{-1}$ but $\sqrt{-1}\sqrt{-1} <_L 0$.
[One can generalize the lexicographic order on $\mathbb{C}$ to any finite dimensional vector space V over $\mathbb{R}$ with ordered basis $\{v_1, \ldots, v_n\}$. How?]

Definition 25.3. Let (S, R) be a poset, T a subset of S. An element a in S is called an *upper bound* of T if $x\mathrm{R}a$ for all $x \in T$, and s is called a *maximal element* of S if $s\mathrm{R}y$ with $y \in S$ implies that $s = y$. We say that the poset S is *inductive* if every chain in S has an upper bound in S.

Examples 25.4. 1. 1 is a maximal element in $([0, 1], \leq)$.

2. 0 is a maximal element in $([0,1],\geq)$. (Of course, it is usually called a *minimal element.*)
3. $(\mathbb{Z},\leq)$ is not inductive.
4. $(\mathbb{Z}^+,\geq)$ is inductive. (Note here 'upper bound' is usually called 'lower bound' for obvious reasons.)

We can now state the axiom that we shall use as an extension of finite induction.

Lemma 25.5. (Zorn's Lemma) *Let S be a non-empty inductive poset. Then S contains a maximal element.*

This is indeed an extension of finite induction, for if S is a subset of $\mathbb{Z}^+$, then $(S,\geq)$ has a maximal element, which is exactly the Well-Ordering Principle. As mentioned, we cannot prove this lemma; we accept it as an axiom.

We give some applications of Zorn's Lemma. The first is the result that vector spaces always have bases.

Proposition 25.6. *Let V be a nonzero vector space over a field F and S a linearly independent subset of V. Then S extends to a basis of V, i.e., is part of a basis for V.*

Proof. Let

$$\mathcal{S} = \{T \mid S \subset T \subset V \text{ with } T \text{ linearly independent}\},$$

a nonempty poset under $\subset$. Let $\mathcal{C}$ be a chain in $\mathcal{S}$.

Claim. $A = \bigcup_{\mathcal{C}} T$ is an upper bound for $\mathcal{C}$ in $\mathcal{S}$:

Suppose that A is linearly dependent. By the definition of linear dependence, there exists a finite linearly dependent subset $\{v_1,\ldots,v_n\}$ of A. For each i, $1 \leq i \leq n$, there exists a $T_i \in \mathcal{C}$ with $v_i \in T_i$. Since $\mathcal{C}$ is a chain and n finite, there exists a T in $\mathcal{C}$ containing T_i for $i = 1,\ldots,n$, hence containing $\{v_1,\ldots,v_n\}$, contradicting the fact that T is linearly independent. This shows that A is indeed an upper bound of $\mathcal{C}$ in $\mathcal{S}$.

As the hypotheses of Zorn's Lemma have been fulfilled, there exists a maximal element T in $\mathcal{S}$. We show that T is a basis for V. If the span, $\langle T\rangle$, of T is not V, then there exists $w \in V \setminus \langle T\rangle$. Let $T' = T \cup \{w\}$. As T' must be linearly dependent by maximality of T, there exist β and β_v in F for $v \in T$, with *almost all* $\beta_v = 0$, i.e., $\beta_v = 0$ except for finitely many v in T, but not all β, β_v zero that satisfies $\sum_T \beta_v v + \beta w = 0$. As T is linearly independent, $\beta \neq 0$, so $w = -\sum_T \beta^{-1}\beta_v v$ lies in $\langle T\rangle$, a contradiction. Consequently, T spans V, so is a basis for V. □

Proposition 25.7. *Let V be a nonzero vector space over a field F and S a spanning set for V. Then a subset of S is a basis of V.*

We leave a proof of this as an exercise. You should try the obvious proof. It does not work. Can you figure out what this means? Rather you will have to modify the proof above.

The two proposition imply, just as in the finite dimensional case, the following:

Corollary 25.8. *Let V be a nonzero vector space over a field F and $\mathcal{B}$ a subset of V. Then the following are equivalent:*

(1) $\mathcal{B}$ is a basis for V.
(2) $\mathcal{B}$ s a maximal linearly independent set in V (relative to $\subset$).
(3) $\mathcal{B}$ is a minimal spanning set for V (relative to $\supset$).

We next apply Zorn's Lemma to Ring Theory.

Proposition 25.9. *Let R be a nontrivial ring and $\mathfrak{A} < R$ an ideal. Then there exists a maximal ideal $\mathfrak{m}$ in R such that $\mathfrak{A} \subset \mathfrak{m}$. In particular, any nontrivial ring contains maximal ideals.*

PROOF. Let

$$S = \{\mathfrak{B} \mid \mathfrak{A} \subset \mathfrak{B} < R \text{ an ideal}\}.$$

The set S is a poset under $\subset$, and S is non-empty as $\mathfrak{A} \in S$. Let $\mathcal{C}$ be a chain in S, so

$$\text{for all } \mathfrak{B}' \text{ and } \mathfrak{B}'' \text{ in } \mathcal{C} \text{ either } \mathfrak{B}' \subset \mathfrak{B}'' \text{ or } \mathfrak{B}'' \subset \mathfrak{B}'.$$

Claim. $\bigcup_{\mathfrak{B}\in\mathcal{C}} \mathfrak{B} < R$ is an ideal and hence an upper bound for $\mathcal{C}$ in S: Let $r \in R$ and $x, y \in \bigcup_{\mathcal{C}} \mathfrak{B}$. Then there exist $\mathfrak{B}'$ and $\mathfrak{B}''$ in $\mathcal{C}$ with $x \in \mathfrak{B}' \in \mathcal{C}$ and $y \in \mathfrak{B}'' \in \mathcal{C}$. As $\mathcal{C}$ is a chain, we may assume that $\mathfrak{B}' \subset \mathfrak{B}''$, so $rx, xr, x \pm y \in \mathfrak{B}'' \subset \bigcup_{\mathcal{C}} \mathfrak{B}$, showing $\bigcup_{\mathcal{C}} \mathfrak{B}$ is an ideal. If $\bigcup_{\mathcal{C}} \mathfrak{B}$ is the unit ideal, there exists a $\mathfrak{B}'$ in $\mathcal{C}$ such that $1 \in \mathfrak{B}'$. This means that $\mathfrak{B}'$ is the unit ideal, so not in S, a contradiction. By Zorn's Lemma, there exists an ideal $\mathfrak{m} \in S$ that is a maximal element. So $\mathfrak{m}$ satisfies $\mathfrak{A} \subset \mathfrak{m} < R$ is an ideal, maximal with respect to this property. Thus $\mathfrak{m}$ is a maximal ideal of R containing $\mathfrak{A}$. □

Remark 25.10. The reason the above worked was the claim, i.e.,

(*i*) The union of ideals in a chain is an ideal under the partial order $\subset$.
[The same is true for chains of subgroups, subspaces of a vector space, and other algebraic objects.]
Note. This does not work for arbitrary unions.

(*ii*) There exists an element lying in none of the elements in any chain (in this case the element 1).
In fact, if R is a non-trivial rng (even commutative), it may not have maximal ideals.

We next want to give further application of Zorn's Lemma to commutative ring theory that generalizes the existence of maximal ideals and which is quite useful. We begin with an important definition.

Definition 25.11. Let R be a nontrivial commutative ring, S a nonempty subset of R. We call S a *multiplicative set* if

(i) $1 \in S$.

(ii) If s_1 and s_2 are elements of S, then $s_1 s_2$ lies in S, i.e., S is closed under multiplication.

If T is a subset of R, we say that T *excludes* S if $S \cap T = \emptyset$.

Examples 25.12. Let R be a nontrivial commutative ring, $a \in R$, and $\mathfrak{A} < R$ an ideal.

1. $\mathfrak{A}$ is a prime ideal if and only if $R \setminus \mathfrak{A}$ is a multiplicative set.
2. $\{a^n \mid n \geq 0\}$ is a multiplicative set.
3. Let S be a multiplicative set with $0 \notin S$. Then S contains no *nilpotent elements*, i.e., an element $x \in R$ that satisfies $x^n = 0$ for some $n \in \mathbb{Z}^+$.
4. $R^\times$ is a multiplicative set.
5. An element $x \in R$ is called a *zero divisor* of R if there exists a nonzero element $y \in R$ satisfying $xy = 0$. Let

$$\mathrm{zd}(R) := \{x \in R \mid x \text{ is a zero divisor}\}.$$

Then $R \setminus \mathrm{zd}(R)$ is a multiplicative set.

Using the concept of multiplicative sets, we can now prove the generalization of the existence of maximal ideals that we seek. The proof will be just as the one for the existence of maximal ideals but using the ideal analogue for prime ideals of Euclid's Lemma given in Lemma 23.20.

Theorem 25.13. (Krull) *Let R be a commutative ring and S a multiplicative set. Suppose that $\mathfrak{A} < R$ is an ideal in R excluding S. Then there exists an ideal $\mathfrak{B} \subset R$ containing $\mathfrak{A}$ excluding S and maximal with respect to these two properties. Moreover, any such $\mathfrak{B}$ is a prime ideal.*

Note: If R is a non-trivial commutative ring, applying the above to the multiplicative set $R^\times$ shows the existence of maximal ideals in a commutative ring.

PROOF. Let

$$\mathcal{I} = \{\mathfrak{B} \mid \mathfrak{A} \subset \mathfrak{B} < R \text{ an ideal excluding } S\}$$

We know that $\mathcal{I}$ is not empty as $\mathfrak{A} \in \mathcal{I}$. Partially order $\mathcal{I}$ by $\subset$. Let $\mathcal{C}$ be a chain in $\mathcal{I}$.

Claim. $\bigcup_{\mathcal{C}} \mathfrak{B}$ is an ideal in R excluding S (hence not the unit ideal). In particular, $\bigcup_{\mathcal{C}} \mathfrak{B}$ is an upper bound for $\mathcal{C}$ in $\mathcal{I}$:

$\bigcup_{\mathcal{C}} \mathfrak{B}$ is an ideal exactly as before. If $b \in S \cap \bigcup_{\mathcal{C}} \mathfrak{B}$, then there exists an ideal $\mathfrak{B} \in \mathcal{C}$ such that $b \in S \cap \mathfrak{B}$, a contradiction. This establishes the claim.

By Zorn's Lemma, there exists an ideal $\mathfrak{B}$ in $\mathcal{I}$ satisfying

(*i*) $\mathfrak{B} \cap S = \emptyset$.
(*ii*) $\mathfrak{A} \subset \mathfrak{B}$.
(*iii*) If $\mathfrak{B} < \mathfrak{B}' \subset R$ is an ideal, then $\mathfrak{B}' \cap S \neq \emptyset$.

To finish, we must show that $\mathfrak{B}$ is a prime ideal. Suppose this is not the case. Then by Lemma 23.20, there exist ideals $\mathfrak{B} < \mathfrak{B}_i$, $i = 1, 2$, satisfying $\mathfrak{B}_1\mathfrak{B}_2 \subset \mathfrak{B}$. By (*iii*), there exist $s_i \in \mathfrak{B}_i \cap S$, $i = 1, 2$. As S is a multiplicative set, $s_1 s_2 \in \mathfrak{B} \cap S$, a contradiction. □

Note the last part of the proof is structurally similar to factoring integers into primes. This is an argument that is often used.

Example 25.14. Let $C([0, 1])$ denote the ring of continuous real-valued functions on $[0, 1]$. Exercise 23.21(13) said that the maximal ideals in $C([0, 1])$ are in one to one correspondence with $[0, 1]$, the correspondence given by $a \mapsto \mathfrak{m}_a = \{f \in C([0, 1]) \mid f(a) = 0\}$. (This is true with $[0, 1]$ replaced by any finite closed interval (or any compact subset) of the real line.) Let S denote the set of all nonzero polynomial functions in $C([0, 1])$. (Cf. Remark 31.10(3) for a precise definition of a polynomial function.) Then S is a multiplicative set excluding the zero ideal. Hence by the theorem, there exists a prime ideal $\mathfrak{p}$ excluding S. Consequently, $\mathfrak{p}$ does not contain any linear polynomial function $x - a$, $a \in [0, 1]$. In particular, $\mathfrak{p} \not\subset \mathfrak{m}_a$ for any $a \in [0, 1]$, so $\mathfrak{p}$ is not maximal.

The theorem allows us to begin to see the importance of prime ideals in understanding the structure of a commutative ring.

Definition 25.15. Let R be a non-trivial commutative ring. The *nilradical* of R is the set

$$\text{nil}(R) := \{a \mid a \text{ is a nilpotent element in } R\}.$$

Corollary 25.16. *Let R be a non-trivial commutative ring. Then* $\mathrm{nil}(R)$ *is an ideal in R and*

$$\mathrm{nil}(R) = \bigcap_{\substack{\mathfrak{p}<R \\ \mathfrak{p} \text{ a prime ideal}}} \mathfrak{p}.$$

PROOF. We leave it as an exercise to prove $\mathrm{nil}(R)$ is a ideal.

$\mathrm{nil}(R) \subset \bigcap_{\substack{\mathfrak{p}<R \\ \mathfrak{p} \text{ a prime ideal}}} \mathfrak{p}$: If $a \in \mathrm{nil}(R)$, there exists an $n \in \mathbb{Z}^+$ such that $a^n = 0$ and $0 \in \mathfrak{A}$ for every ideal, in particular for prime ideals. Hence $a \in \mathfrak{p}$ for all prime ideals $\mathfrak{p}$ in R.

$\bigcap_{\substack{\mathfrak{p}<R \\ \mathfrak{p} \text{ a prime ideal}}} \mathfrak{p} \subset \mathrm{nil}(R)$: Suppose this is not true, then there exists an element

$$a \in \bigcap_{\substack{\mathfrak{p}<R \\ \mathfrak{p} \text{ a prime ideal}}} \mathfrak{p} \setminus \mathrm{nil}(R).$$

In particular, the multiplicative set $S = \{a^n \mid n \geq 0\}$ does not contain 0 hence the trivial ideal (0) excludes S. By the theorem of Krull, there exists a prime ideal $\mathfrak{p}$ excluding S. Thus $a \notin \mathfrak{p}$, a contradiction.

Therefore, $\mathrm{nil}(R) = \bigcap_{\substack{\mathfrak{p}<R \\ \mathfrak{p} \text{ a prime ideal}}} \mathfrak{p}$ as needed. □

Note. If R is a domain, then $\mathrm{nil}(R) = 0$. The converse is false, e.g., $\mathbb{Z} \times \mathbb{Z}$ has trivial nilradical, but is not a domain. A commutative ring whose nilradical is trivial is called a *reduced ring*.

Exercises 25.17. 1. Let V be a finite dimensional vector space over $\mathbb{R}$ with ordered basis $\{v_1, \ldots, v_n\}$. Define a lexicographic order of V relative to this ordered basis.

2. Prove Proposition 25.7.

3. Rngs may not have maximal ideals. This problem constructs one. Let p be a prime number. Let $\mathbb{Z}_{p^\infty}$ be the (additive) subgroup of $\mathbb{Q}/\mathbb{Z}$ consisting of all elements having order some power of p, i.e., $x = \alpha + \mathbb{Z}$, $\alpha \in \mathbb{Q}$, lies in $\mathbb{Z}_{p^\infty}$ if and only if $p^r x = 0$ in $\mathbb{Q}/\mathbb{Z}$, i.e., $p^r \alpha$ is an integer, for some positive integer r.
   (*i*) Show that the set $\{\frac{1}{p^r} + \mathbb{Z} \mid r \text{ a non-negative integer }\}$ generates $\mathbb{Z}_{p^\infty}$.
   (*ii*) Show the subgroup $\left\langle \frac{1}{p^r} + \mathbb{Z} \right\rangle$ of $\mathbb{Z}_{p^\infty}$ is isomorphic to $\mathbb{Z}/p^r\mathbb{Z}$.

(*iii*) Show that any subgroup of $\mathbb{Z}_{p^\infty}$ is $\left\langle \frac{1}{p^r} + \mathbb{Z} \right\rangle$ for some non-negative integer r and the subgroups of $\mathbb{Z}_{p^\infty}$ form a chain under set inclusion.

(*iv*) Make $\mathbb{Z}_{p^\infty}$ into a rng by defining $x \cdot y = 0$ for all $x, y \in \mathbb{Z}_{p^\infty}$. Show that $\mathbb{Z}_{p^\infty}$ has no maximal ideals

4. Let R be a commutative ring. Prove the following:
   (*i*) Let $\mathcal{S}$ be the set of non finitely generated ideals in R. Suppose $\mathfrak{A}$ is a maximal element in $\mathcal{S}$. Then $\mathfrak{A}$ is a prime ideal.
   (*ii*) Let $\mathfrak{A}$ be a non finitely generated ideal in R. Suppose an ideal $\mathfrak{B}$ in R has the property that it contains $\mathfrak{A}$ is not finitely generated, and is maximal with respect to this property. Show that there exists a prime ideal in R containing $\mathfrak{A}$ that is not finitely generated.
5. Let R be a commutative ring. Prove
   (*i*) If x is nilpotent in R, then $1 + x$ is a unit in R.
   (*ii*) The nilradical of R is an ideal.
   (*iii*) Compute the nilradical of the rings: $\mathbb{Z}/12\mathbb{Z}$, $\mathbb{Z}/n\mathbb{Z}$, $n > 1$, and $\mathbb{Z}$.
6. Let R be a commutative ring. The *Jacobson radical* of R is defined to be $\mathrm{rad}(R) := \bigcap_{\mathrm{Max}(R)} \mathfrak{m}$, the intersection of all maximal ideals in R. Show that x lies in $\mathrm{rad}(R)$ if and only if $1 - yx$ is a unit in R for all y in R.
7. Let R be a commutative ring and $\mathfrak{A} < R$ an ideal. Define the *radical* radical!of an ideal of $\mathfrak{A}$ to be the set

$$\sqrt{\mathfrak{A}} := \{x \in R \mid x^n \in \mathfrak{A} \text{ for some } n \in Z^+\}.$$

   Show the following:
   (*i*) $\mathfrak{A}$ is an ideal and

$$\sqrt{\mathfrak{A}} = \bigcap_{\substack{\mathfrak{A} \subset \mathfrak{p} < R \\ \mathfrak{p} \text{ a prime ideal}}} \mathfrak{p}.$$

   (*ii*) Let $^- : R \to R/\mathfrak{A}$ be the canonical ring epimorphism, then $\mathrm{nil}(\overline{R}) = \sqrt{\mathfrak{A}}/\mathfrak{A}$.
8. Let R be a commutative ring, $\mathfrak{A} < R$ an ideal, and $^- : R \to R/\mathfrak{A}$ be the canonical epimorphism. We say that $\mathfrak{A}$ is a *primary ideal* if $ab \in \mathfrak{A}$ implies that $a \in \mathfrak{A}$ or $b^n \in \mathfrak{A}$ for some positive integer n. Let $\mathfrak{A} < R$ be an ideal. Show both of the following:

(*i*) $\mathfrak{A}$ is a primary ideal if and only if every zero divisor of $R/\mathfrak{A}$ is nilpotent.

(*ii*) If $\mathfrak{A}$ is primary, then its radical, $\sqrt{\mathfrak{A}}$, (see the previous exercise) is a prime ideal.

9. Determine all primary ideals if R is a PID. (Cf. the previous exercise for the definition of a primary ideal.)

10. Let R be a commutative ring. An element e of R is called an *idempotent* if $e^2 = e$. For example, if S is another commutative ring the element $(1_R, 0_S)$ is an idempotent in the ring $R \times S$. The object of this exercise is to prove a converse. Let e be an idempotent of R. Then prove

(*i*) $e' := 1 - e$ is an idempotent of R.

(*ii*) The principal ideal Re of R is a ring with identity $1_{Re} = e$. [Note that Re is not a subring of R since Re will not have the same identity as R unless $e = 1$.]

(*iii*) R is ring isomorphic to $Re \times Re'$.

26. Addendum: Localization

In section §24, we showed that a field of quotients exists for any domain by defining fractions. The purpose was to create a ring containing the given ring R in which every nonzero element of R has a multiplicative inverse. [Note if 0 has an inverse then the ring must be trivial.] Instead, we can try to invert only some elements. This turns out to be a crucial idea, and the method is to invert elements in a multiplicative set. But to make it effective, we must weaken what we mean when we say two 'fractions' are equal.

Construction 26.1. We shall leave most of the details of this construction as an exercise, an exercise that you should do. Let R be a commutative ring and S a multiplicative set in R. Let

$$\mathcal{S} = \{(a, s) \mid a \in R,\ s \in S\} = R \times S.$$

Define a relation on $\mathcal{S}$ as follows: Write

$$(a, s) \sim (a', s') \text{ in } \mathcal{S}$$

if there exists an $s'' \in \mathcal{S}$ satisfying $s''(as' - a's) = 0$ in R.
Then $\sim$ is a equivalence relation. Let

$$\frac{a}{s} \text{ denote the equivalence class of } (a, s)$$

and set

$$S^{-1}R = \{\frac{a}{s} \mid a \in R,\ s \in S\} = R/\sim .$$

Define

$$\frac{a}{s_1} + \frac{b}{s_2} := \frac{as_2 + bs_1}{s_1 s_2}$$
$$\frac{a}{s_1} \cdot \frac{b}{s_2} := \frac{ab}{s_1 s_2}$$

for all $a, b \in S$ and for all $s_1, s_2 \in S$. Then

$$+ : S^{-1}R \times S^{-1}R \to S^{-1}R \quad \text{and} \quad \cdot : S^{-1}R \times S^{-1}R \to S^{-1}R$$

are well-defined maps making $S^{-1}R$ into a commutative ring, called the *localization* of R at S with $0_{S^{-1}R} = \frac{0}{s}$ and $1_{S^{-1}R} = \frac{s}{s}$, for any $s \in S$. The map

$$\varphi : R \to S^{-1}R \text{ given by } r \mapsto \frac{r}{1}$$

is a ring homomorphism.

Examples 26.2. Let R be a commutative ring and S a multiplicative set in R.

1. If $0 \in S$, then $S^{-1}R$ is the trivial ring.
2. Suppose that R is a domain. Then $T = R \setminus \{0\}$ is a multiplicative set and $T^{-1}R$ is the quotient field of R.
3. Let $\varphi : R \to S^{-1}R$ be the ring homomorphism given by $r \mapsto \frac{r}{1}$. Then
   (*i*) $\varphi(S) \subset (S^{-1}R)^\times$.
   (*ii*) $\ker \varphi = \{x \in R \mid \text{there exists an } s \text{ in } S \text{ such that } sx = 0\}$.
   (*iii*) φ is a ring monomorphism if and only if $S \cap \mathrm{zd}(R) = \emptyset$.
   (*iv*) If R is a domain and S does not contain zero, then φ is a ring monomorphism. Note if this is the case that $\frac{a}{s} = \frac{b}{s}$ in $S^{-1}R$ if and only if $a = b$. We then view φ as the inclusion map, so $R \subset S^{-1}R \subset qf(R)$ (cf. (2)).
4. Let $\mathfrak{p}$ be a prime ideal in R. Then $T = R \setminus \mathfrak{p}$ is a multiplicative set and the localization $T^{-1}R$ is denoted $R_{\mathfrak{p}}$. For example, if p is a prime in $\mathbb{Z}$ and $T = \{n \in \mathbb{Z} \mid p \not| \, n\}$, then $\mathbb{Z}_{(p)} = \{\frac{n}{m} \mid n, m \in \mathbb{Z} \text{ relatively prime with } p \not| \, m\}$.
5. Let f be an element of R. Then $T = \{f^n \mid n \geq 0\}$ is a multiplicative set and the localization $T^{-1}R$ is denoted R_f. For example, if p is a prime in $\mathbb{Z}$, then $\mathbb{Z}_p = \{\frac{n}{m} \mid n, m \in \mathbb{Z} \text{ relatively prime with } p \mid m\}$.

The existence of the localization of a commutative ring at a multiplicative set is a basic construction in commutative algebra whose importance cannot be overestimated. It is especially important when the multiplicative set is the complement of a prime ideal. For example, the ring $\mathbb{Z}_{(p)}$ in Example 26.2(4) has only one nonzero prime ideal and

all prime integers relatively prime to p become units. One can view $\mathbb{Z}_{(p)} \subset \mathbb{Q}$. It turns out that in this ring every element can be written up^n for some unit u in $\mathbb{Z}_{(p)}$ and some positive integer n. In the more general case that R is a commutative ring and $\mathfrak{p}$ a prime ideal in R, let $\mathfrak{p}R_{\mathfrak{p}}$ denote the image of $\mathfrak{p}$ in $R_{\mathfrak{p}}$. Then $\mathfrak{p}R_{\mathfrak{p}}$ is the unique maximal ideal in $R_{\mathfrak{p}}$. One can obtain valuable information for R from $R_{\mathfrak{p}}$ via the natural map $R \to R_{\mathfrak{p}}$ just as one get valuable information for R from the canonical epimorphism $R \to R/\mathfrak{p}$.

Exercises 26.3. 1. Prove all of the claims in Construction 26.1.

2. Let $\mathfrak{p}$ be a prime ideal in a commutative ring R. Show that$R_{\mathfrak{p}}$ has a unique maximal ideal. What is it? A commutative ring with a unique maximal ideal is called a *local ring*.
3. Let $\mathfrak{A}$ be an ideal in a commutative ring R and S a multiplicative set in R. Define
$$S^{-1}\mathfrak{A} := \{\frac{a}{s} \mid a \in \mathfrak{A},\, s \in S\}.$$
Show that $S^{-1}\mathfrak{A}$ is an ideal in $S^{-1}R$ and not the unit ideal if $\mathfrak{A}$ excludes S.
[Note. If $\frac{x}{s}$ lies in $S^{-1}R$, this does not mean that $x \in \mathfrak{A}$, only that there exist an $a \in \mathfrak{A}$ and an $s' \in S$ such that $\frac{x}{s} = \frac{a}{s'}$.]
4. Let R be a commutative ring, S a multiplicative set in R, and $\varphi : R \to S^{-1}R$ the ring homomorphism given by $r \mapsto \frac{r}{1}$. Suppose that $\mathfrak{B}$ is an ideal in $S^{-1}R$. Show that $\mathfrak{A} = \varphi^{-1}(\mathfrak{B})$ is an ideal in R and $S^{-1}\mathfrak{A} = \mathfrak{B}$.
5. Let R be a commutative ring, S a multiplicative set in R. Show if $\mathfrak{p}$ is a prime ideal in R excluding S, then $S^{-1}\mathfrak{p}$ is a prime ideal, and every prime in $S^{-1}R$ is of this form. Use this to show if $0 \notin S$, then there exists a prime ideal in R excluding S. (Cf. this to Theorem 25.13.)
6. Let R be a commutative ring and S a multiplicative set in R. Let $\varphi : R \to S^{-1}R$ be given by $r \mapsto \frac{r}{1}$. Show that this satisfies the following *universal property*: If $\psi : R \to R'$ is a ring homomorphism with R' commutative and $\psi(S)$ a subset of the unit group of R', then there exists a unique ring homomorphism $\theta : S^{-1}R \to R'$ such that
$$\begin{array}{ccc} R & \xrightarrow{\varphi} & S^{-1}R \\ & \searrow^{\psi} & \downarrow{\scriptstyle \theta} \\ & & R' \end{array}$$
commutes.

CHAPTER VI

Domains

In this chapter, we study in greater depth special domains. The main goal is to study those domains that satisfy the analogue of the Fundamental Theorem of Arithmetic called Unique Factorization Domains or UFDs and explicit examples of these domains. The UFDs include PIDs and the historically interesting special class of PIDs that satisfy the division algorithm called *euclidean domains*. The explicit example of a euclidean domain, the Gaussian integers, is used to prove the classical two square theorems in Number Theory. We also prove Lagrange's Four Square Theorem to indicate the method of proof developed by Fermat's called Infinite Descent (a form of induction) to prove this theorem. We also introduce a more general class of commutative rings called Noetherian rings. These are the most important commutative rings in algebra and in algebraic geometry (as well as historically the most important because of the work of Hilbert). We shall study these rings in greater detail in a later chapter. For purposes here they represent a generalization of PIDs. From the viewpoint of induction, they represent rings where finite induction has a direct analogue. Theorems proven using Zorn's Lemma over a commutative ring, usually do not need it if the ring is Noetherian.

27. Special Domains

We have seen that prime elements in $\mathbb{Z}$ are those elements that satisfy one of two equivalent conditions: Let $a, b \in \mathbb{Z}$. Then $p \neq 0$ is a prime if

1. whenever $p = ab$, then either $a = \pm 1$ or $b = \pm 1$, i.e., either a or b is a unit or
2. whenever $p \mid ab$ then $p \mid a$ or $p \mid b$.

The first was the original definition, the second Euclid's Lemma. We have generalized the concept of a prime element in $\mathbb{Z}$ to the concept of a prime ideal, but have not generalized the first equivalence in this way, nor will we, as its importance is related to elements. Elements satisfying the generalization of the first condition will be called *irreducible elements*. That elements in a domain factor into a product of

irreducible elements (up to units) holds for a large class of domains. What does not hold in general and what makes the two conditions not equivalent is an appropriate uniqueness statement. We shall carefully study domains in which we do have such a uniqueness statement. In general the second condition above is the more important as it gives more information about rings.

Definition 27.1. Let R be a domain and a a nonzero nonunit in R. We say a is *irreducible* if whenever $a = bc$ with $b, c \in R$, then either b or c is a unit in R (if a is not irreducible, we say that it is *reducible*) and is a *prime element* if whenever $a \mid bc$ with $b, c \in R$, then $a \mid b$ or $a \mid c$.

Note that any associate of an irreducible element (respectively, prime element) is irreducible (respectively, prime).

Remark 27.2. Let R be a domain and $p \in R$ a prime element. Then p is irreducible i.e., being a prime element is stronger than being an irreducible element.

PROOF. Suppose that $a = bc$ with $b, c \in R$. As a is a prime element, either $a \mid b$ or $a \mid c$, say $a \mid b$. Then $b = ax$ some nonzero element $x \in R$. Therefore, we have $b = ax = bcx$ in the domain R, so cancellation implies that $1 = cx$, i.e., $c \in R^\times$ as needed. □

We want to generalize the Fundamental Theorem of Arithmetic to a wider class of domains. Before, our approach relied on the concept of greatest common divisors. In general, our notion of a gcd of two elements below in a domain may not exist. Moreover, for integers, a gcd was always unique because we could make it positive as $\mathbb{Z}^\times = \{\pm 1\}$. In general, we cannot do this, so must omit this condition that a gcd is unique if it exists.

Definition 27.3. Let R be a domain, a, b and d nonzero elements in R. We call d a *greatest common divisor* or *gcd* of a and b if:

(i) $d \mid a$ and $d \mid b$ in R.

(ii) If c in R satisfies $c \mid a$ and $c \mid b$ in R, then $c \mid d$ in R.

If a gcd of a and b exists and is a unit, then we say that a and b are *relatively prime*.

Remarks 27.4. Let R be a domain and a, b nonzero elements in R.

1. As mentioned above, a gcd of a, b may not exist. For example, there exist elements in the subring $R := \{x + 2y\sqrt{-1} \mid x, y \in \mathbb{Z}\}$ of the Gaussian integers $\mathbb{Z}[\sqrt{-1}] := \{x + y\sqrt{-1} \mid x, y \in \mathbb{Z}\}$ not having a gcd.

2. If a gcd d of a, b exists it may not satisfy $d = ax + by$ for some $x, y \in R$, as it did for $\mathbb{Z}$.
3. If $d \in R$ satisfies $(d) = (a, b)$ then d is a gcd of a and b (and, of course, there exist $x, y \in R$ satisfying $d = ax + by$).
4. If d and d' are two gcd's of a and b in R, then $d \mid d'$ and $d' \mid d$, hence $d \approx d'$, i.e., $(d) = (d')$. Moreover, if this is the case, then any *generator* c of (d), i.e., an element c in R satisfying $(c) = (d)$, is a gcd of a and b. In particular, a gcd of a and b, if it exists, is unique up to units in R.

We codify some of the remarks above in the following:

Proposition 27.5. *Let R be a PID, a and b nonzero elements in R. Let $(d) = (a, b)$. Then d is a gcd of a and b in R, unique up to units. In particular, gcd's of nonzero elements in a PID always exist and if d' is a gcd of a and b, the $d' = ax + by$ for some $x, y \in R$.*

The proposition allows us to extend the analogue of Euclid's Lemma to any PID.

Corollary 27.6. (Euclid's Lemma) *Let R be a PID and $r \in R$. Then r is irreducible if and only if r is a prime element.*

Proof. We have already shown that prime elements are irreducible. Conversely, assume that r is an irreducible element in R and satisfies $r \mid ab$ with $a, b \in R$. As R is a PID, there exists a $c \in R$ satisfying $(c) = (r, a)$. Write $r = xc$ with $x \in R$. As r is irreducible, either c is a unit or x is a unit in R. If c is a unit, then (c) is the unit ideal, hence $1 = ry + az$ in R for some $y, z \in R$. It follows that $r \mid ryb + abz = b$ as needed. If x is a unit then $(r) = (c) = (r, a)$, so $a \in (r)$, i.e., $r \mid a$. □

The Fundamental Theorem of Arithmetic relied on the equivalence of the notion of irreducible and prime elements. We wish to generalize this theorem to a wider class of domains, so we must say what we mean by uniqueness of factorization. For the integers this meant up to ± 1, the units in $\mathbb{Z}$, so we will use the following generalization.

Definition 27.7. Let R be a domain. We call R a *unique factorization domain* or *UFD* if for any nonzero nonunit r in R, we have

(*i*) $r = f_1 \cdots f_n$ for some irreducible elements $f_1, \ldots, f_n$ in R for some $n \in \mathbb{Z}^+$.

(*ii*) If $r = f_1 \cdots f_n = g_1 \cdots g_m$ with $f_1, \ldots, f_n, g_1, \ldots, g_m$ irreducible elements in R for some $n, m \in \mathbb{Z}^+$, then $n = m$ and there exists a permutation $\sigma \in S_n$ satisfying $f_i \approx g_{\sigma(i)}$ for all i.

[Note that if $r = uf_1 \cdots f_n$ with u a unit in R and $f_i \in R$ for all i, then we can 'absorb' the unit by replacing any one of the factors with an associate.]

Remark 27.8. If R is a UFD, then, as desired, an element in R is prime if and only if it is irreducible.

PROOF. We need only show if r is an irreducible element in R that it is a prime. So suppose that $r \mid ab$ with $a, b \in R$ nonunits. Write $a = f_1 \cdots f_n$ and $b = g_1 \cdots g_m$, with $f_1, \ldots, f_n, g_1, \ldots, g_m$ irreducible elements in R for some $n, m \in \mathbb{Z}^+$. Since $ab = rx$ for some $x \in R$ by assumption, the uniqueness property in the definition of UFD's implies $r \approx f_i$ some i or $r \approx g_j$ some j, i.e., $r \mid a$ or $r \mid b$. □

The uniqueness property for a UFD is really just that irreducible elements are prime:

Proposition 27.9. (Euclid's Argument) *Let R be a domain. Suppose that $p_1 \cdots p_n = f_1 \cdots f_m$ with $p_1, \ldots, p_n$ prime elements in R and $f_1, \ldots, f_m$ irreducible elements in R for some $n, m \in \mathbb{Z}^+$. Then $n = m$ and there exists a permutation $\sigma \in S_n$ satisfying $p_i \approx f_{\sigma(i)}$. In particular, each f_i is a prime element in R.*

PROOF. As $p_1 \mid f_1 \cdots f_m$ in R, there exists an i such that $f_i = up_1$ for some element u in R. Since f_i is irreducible and p_1 not a unit, u is a unit in R. Now $p_1 \cdots p_n = up_1 f_2 \cdots f_m$ in the domain R. If $n = 1$ or $m = 1$ we must have $m = n = 1$ by cancellation, since the product of irreducibles is not a unit. So we may assume that $m, n > 1$. Cancellation shows $p_2 \cdots p_n = uf_2 \cdots f_m$, and the result follows by induction. □

An immediate consequence is the following:

Corollary 27.10. *Let R be a domain. Suppose that every nonzero nonunit in R is a product of (finitely many) prime elements, i.e., $a = p_1 \cdots p_n$, with $p_1, \ldots, p_n$ prime elements in R. Then R is a UFD.*

The most important class of commutative rings is the following:

Definition 27.11. A commutative ring R is called a *Noetherian ring* if one of the following equivalent conditions hold:

(i) R satisfies the *ascending chain condition* or *ACC*, viz., given a (countable) chain of ideals

$$\mathfrak{A}_1 \subset \mathfrak{A}_2 \subset \cdots \subset \mathfrak{A}_n \subset \cdots$$

in R, there exists a positive integer N such that $\mathfrak{A}_{N+i} = \mathfrak{A}_N$ for all $i \geq 0$.

[We say that the chain *stabilizes*.]
Equivalently, there exists no infinite chain of ideals in R,

$$\mathfrak{B}_1 < \mathfrak{B}_2 < \cdots < \mathfrak{B}_n < \cdots .$$

(*ii*) Every ideal $\mathfrak{A}$ in R is *finitely generated* (*fg*), i.e., there exist $a_1, \ldots, a_n \in R$, some n, such that $\mathfrak{A} = (a_1, \ldots, a_n)$.

(*iii*) R satisfies the *Maximal Principle*, i.e., any non-empty set of ideals S in R contains a maximal element relative to the relation $\subset$.
[Cf. this condition to the conclusion of Zorn's Lemma.]

Remark 27.12. We leave it as an important exercise (cf. Exercise 27.22(5)) to show these three conditions are equivalent. However, the proof of this really needs the Axiom of Choice to prove ACC implies the Maximal Principle, and it is known that the Axiom of Choice is equivalent to Zorn's Lemma. In fact, it is also known that the Axiom of Choice is also equivalent to the implication that ACC implies the Maximal Principle. For Noetherian rings one usually uses the Maximal Principle instead of directly using Zorn's Lemma in proofs.

Using the second condition we have:

Examples 27.13. 1. Every PID is a Noetherian domain.

2. Let R be a commutative ring. By Exercise 25.17(4), R is Noetherian if and only if every finitely generated prime ideal is finitely generated. Let $A = R[[t]]$. By Exercise 23.21(3), prime ideals in A are of the form $\mathfrak{p}A$ or $\mathfrak{p}A + tA$ for some prime ideal $\mathfrak{p}$ pf R. Hence if R is Noetherian, so is $R[[t]]$.

Noetherian rings are important as they are the rings that arise in classical algebraic geometry. The collection of Noetherian rings also has the properties of being closed under basic ring constructions (and combinations of them): viz.,

(*i*) The homomorphic image of a Noetherian ring is Noetherian.

(*ii*) The localization of a Noetherian ring is Noetherian.

(*iii*) (Hilbert Basis Theorem) If R is a Noetherian ring so is $R[t]$.

We shall leave the first two as exercises and prove the last in Section §38. (Cf. Theorem 38.1). It is not true, however, that a subring of a Noetherian ring is Noetherian. Indeed as any field is a Noetherian domain, one need only find a non-Noetherian domain to provide a counterexample. Can you do so?

We illustrate the importance of the Noetherian condition by the following interesting result.

Theorem 27.14. *Let R be a Noetherian domain and r a nonzero nonunit in R. Then r is a product of (finitely many) irreducible elements.*

PROOF. Let

$$S = \{(a) \,|(0) < (a) < R \text{ with } a \text{ not a product of irreducible elements}\}.$$

Suppose that S is non-empty, i.e., the result is false. Then there exists a principal ideal $(a) \in S$, a maximal element by the Maximal Principle. We call (a) a *maximal counterexample*. Certainly, a cannot be irreducible, so $a = bc$ for some non-units $b, c \in R$. But then $(a) < (b)$ and $(a) < (c)$, so by maximality, both b and c are products of irreducible elements, hence so is $a = bc$. □

The above proof is called proof by *Noetherian induction*. Note the similarity between this proof and that showing positive integers factor into primes. Further application of this type of decomposition into 'smaller pieces' for Noetherian rings is given in the exercises.

Theorem 27.15. *Let R be a PID. Then R is a UFD.*

PROOF. As R is a PID, it is a Noetherian domain. Hence every nonzero nonunit factors into a product of irreducible elements. As every irreducible element in a PID is a prime, factorization is (essentially) unique by Euclid's Argument. □

Corollary 27.16. *$\mathbb{Z}$ is a UFD.*

Next we generalize the Division Algorithm, to study a special type of PID.

Definition 27.17. Let R be a domain. Then R is called a *euclidean domain* if there exists a function

$$\partial : R \setminus \{0\} \to \mathbb{Z}^+ \cup \{0\}$$

satisfying the *Division Algorithm*, i.e., for all $a, b \in R$ with b nonzero, there exist $q, r \in R$ satisfying

(i) $$a = bq + r.$$

(ii) $$r = 0 \text{ or } \partial r < \partial b.$$

We call ∂ a *euclidean function*.

Remark 27.18. If we agree that $-\infty + n = -\infty$ for all integers n and let $\partial 0 = -\infty$, we can write (ii) as

$$\partial r < \partial b.$$

[This is often done.]

Remark 27.19. Historically, a euclidian function was also made to satisfy

(iii) $\partial(ab) \geq \partial(a)$ for all nonzero $a, b \in R$.

If (*iii*) also holds, we shall call this a *strong euclidean domain.*
[If (*iii*) holds, then $\partial(ab) > \partial(a)$ if b is not a unit and $\partial(u) = \partial(1)$ for all units $u \in R^\times$.]

Theorem 27.20. *If R is a euclidean domain, then R is a PID, hence a UFD.*

PROOF. Let $0 < \mathfrak{A} \subset R$ be an ideal and

$$\emptyset \neq \mathcal{S} = \{\partial a \mid 0 \neq a \in \mathfrak{A}\} \subset \mathbb{Z}^+ \cup \{0\}$$

(as $\mathfrak{A} \neq 0$), By the Well-ordering Principle, there exists an element a in $\mathfrak{A}$ with $\partial a \in \mathcal{S}$ minimal.

Claim. $\mathfrak{A} = (a)$ [Cf. the proof that $\mathbb{Z}$ is a PID.]:
Let $b \in \mathfrak{A}$, then $b = aq + r$ in R with $r = 0$ or $\partial r < \partial a$. If $r \neq 0$, then $0 \neq r = b - aq$ lies in $\mathfrak{A}$, contradicting the minimality of ∂a so $r = 0$ and $\mathfrak{A} = (a)$. □

Remarks 27.21. (no justification)

1. $\mathbb{Z}[t]$ is a UFD but not a PID.
2. $\mathbb{Z}[\sqrt{-5}]$ is not a UFD.
3. $\mathbb{Z}[\sqrt{-1}]$, $\mathbb{Z}[\sqrt{-2}]$, $\mathbb{Z}[\frac{-1-\sqrt{-3}}{2}]$, $\mathbb{Z}[\frac{-1-\sqrt{-7}}{2}]$, $\mathbb{Z}[\frac{-1-\sqrt{-11}}{2}]$ are (strong) euclidean domains.
4. $\mathbb{Z}[\frac{-1-\sqrt{-19}}{2}]$ is a PID but not strong euclidean.
5. Any field is euclidean.
6. If F is a field, then the polynomial ring $F[t]$ is euclidean with euclidean function deg.

Exercises 27.22. 1. Let R be a domain and a a nonzero nonunit in R. Show that a is irreducible if and only if the principal ideal (a) is maximal in the set $\{(b) \mid b \text{ a nonzero nonunit in } R\}$. In particular, if R is a PID, then every irreducible element in R is a prime element.

2. Produce elements a and b in the domain $R := \{x+2y\sqrt{-1} \mid x, y \in \mathbb{Z}\}$ having no gcd. Prove your elements do not have a gcd.
3. Let $\mathbb{Z} \subset R$ be a subring with R a UFD. Let d be the gcd of two nonzero integers a and b in $\mathbb{Z}$. Show that d is still a gcd of a and b in R.
4. Show 1 is a gcd for 2 and t in $\mathbb{Z}[t]$, but there are no polynomials $f, g \in \mathbb{Z}[t]$ satisfying $1 = 2f + tg$.

5. Prove that the three conditions defining a Noetherian ring in (27.11) are indeed equivalent.

   (Note: Technically to prove ACC implies the Maximal Principle, one needs the Axiom of Choice as mentioned in Remark 27.12. If you see where, you can invoke it.)

6. Prove Theorem 27.14 using ACC and not the Maximal Principle.

7. A commutative ring R is called *Artinian* if it satisfies the following condition (called the *descending chain condition* or *DCC*): Any chain of ideals

   $$\mathfrak{A}_1 \supset \mathfrak{A}_2 \supset \cdots \supset \mathfrak{A}_n \supset \cdots$$

   (countable) in R *stabilizes*, i.e, there exists an integer N such that $\mathfrak{A}_{N+i} = \mathfrak{A}_N$ for all $i \geq 0$. Equivalently, there exist no infinite chains

   $$\mathfrak{B}_1 > \mathfrak{B}_2 > \cdots > \mathfrak{B}_n > \cdots .$$

   Show that R is Artinian if and only if it satisfies the *Minimal Principle* which says that any non-empty collection of ideals in R has a minimal element (under set inclusion). (Cf. See the Note in Exercise (5) above. It applies here as well.)

8. If R is a domain, show that it is Artinian (cf. previous exercise) if and only if it is a field.

9. Let R be a Noetherian ring. Show if $\varphi : R \to S$ is a ring epimorphism, then S is Noetherian.

10. If R is a Noetherian ring and S a multiplicative set in R, then the localization $S^{-1}R$ is a Noetherian ring. (Cf. Exercise 26.3(4).)

11. Let R be a Noetherian domain. Show that any nontrivial ideal of R contains a finite product of nonzero prime ideals, i.e., if $0 < \mathfrak{A} < R$ is an ideal, then there exist nonzero prime ideals $\mathfrak{p}_1, ..., \mathfrak{p}_n$ in R such that $\mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_n \subset \mathfrak{A}$.

12. Let R be a Noetherian ring. Show that any ideal $\mathfrak{A} < R$ contains a finite product of prime ideals.

13. An ideal $\mathfrak{C}$ in a commutative ring R is called *irreducible* if whenever $\mathfrak{C} = \mathfrak{A} \cap \mathfrak{B}$ for some ideals $\mathfrak{A}$ and $\mathfrak{B}$ in R, then either $\mathfrak{C} = \mathfrak{A}$ or $\mathfrak{C} = \mathfrak{B}$. Show if R is Noetherian, then every ideal $\mathfrak{A} < R$ is a finite intersection of irreducible ideals of R, i.e., $\mathfrak{A} = \mathfrak{C}_1 \cap \cdots \cap \mathfrak{C}_n$, for some irreducible ideals $\mathfrak{C}_i$ in R.

14. Let f, g be polynomials with coefficients in a commutative ring R. Suppose the leading coefficient of f is a unit (i.e., the coefficient of the highest degree term in f is a unit). Show that there are polynomials q and r with coefficients in R such that $g = fq + r$ with either $r = 0$ or the degree of r is less than the degree of f. This says

the Division Algorithm holds when dividing by a polynomial with unit leading term.

15. Let F be a field. Show that $F[t]$, the ring of polynomials with coefficients in F, is a euclidean domain, hence a PID, hence a UFD.
16. Prove that a domain in which every prime ideal is principal is a PID.
17. Let R be a strong euclidean domain, a, b nonzero elements in R, Show that b is a unit if and only if $\delta(b) = \delta(1)$ and if b is a nonunit, then $\delta(ab) > \delta(a)$.
18. Let R be a euclidean domain. Show directly, i.e., without using facts about PIDs or Noetherian domains, the following:
    (a) Every irreducible element in R is a prime element.
    (b) If R is a strong euclidean domain, then R is a UFD.
19. Let R be a UFD with quotient field K. Let S be a multiplicative set in R not containing zero. Show that the localization $S^{-1}R$ of R at S is a UFD. In particular, if $\mathfrak{p}$ is a prime ideal in R, then the localization $R_{\mathfrak{p}} := S^{-1}R$ of R at $S = R \setminus \mathfrak{p}$,

$$R_{\mathfrak{p}} := \{\frac{a}{b} \mid a, b \in R,\ b \notin \mathfrak{p}\} \subset K$$

is also a UFD.

28. Addendum: Characterization of UFDs

We prove a generalization of our result that a PID is a UFD by giving Kaplansky's characterization of UFD's.

Theorem 28.1. (Kaplansky) *Let R be a domain. Then R is a UFD if and only if every nonzero prime ideal contains a prime element.*

PROOF. ($\Rightarrow$): Let $0 < \mathfrak{p} < R$ be a prime ideal and a a nonzero element in $\mathfrak{p}$. Then $a = f_1 \cdots f_r$, with $f_1, \ldots, f_r$ irreducible in R. Since $\mathfrak{p}$ is a prime ideal, there exist an i such that $f_i \in \mathfrak{p}$. As R is a UFD, f_i is a prime element.

($\Leftarrow$): Let

$$S = \{r \in R \mid r \neq 0 \text{ and } r \text{ is a unit or a product of prime elements}\}.$$

Clearly, $S \neq \emptyset$ is a multiplicative set and (0) excludes S. If we show $S = R \setminus \{0\}$, then R is a UFD by Euclid's Argument, i.e., we shall be done.

Claim. S is a *saturated multiplicative set*, i.e., if $a, b \in R$, then

$$a, b \in S \text{ if and only if } ab \in S :$$

As S is a multiplicative set, we need only show if $ab \in S$, then a and b lie in S. Note that $R^\times$ is a saturated multiplicative set, and by

definition a subset of S, so we may assume that neither a nor b is a unit. Write $ab = p_1 \cdots p_r$ with $p_1, \ldots, p_r$ primes hence each p_i lies in S. As $p_1 \mid ab$ and p_1 is a prime element, $p_1 \mid a$ or $p_1 \mid b$, say $p_1 \mid a$. Write $a = p_1 a_1$ with $a_1 \in R$, hence $p_1 a_1 b = p_1 p_2 \cdots p_r$ in the domain R. By cancellation, we conclude that $a_1 b = p_2 \cdots p_r$ in S. By induction on r, we have $a_1, b \in S$. As S is a multiplicative set, $a \in S$. This proves the claim.

As (0) excludes S, by Krull's Theorem, there exists a prime ideal $\mathfrak{p} \in R$ excluding S and maximal with respect to excluding S. If $0 < \mathfrak{p}$, then there exists a prime element $p \in \mathfrak{p}$, hence $p \in S \cap \mathfrak{p}$, a contradiction. Consequently, $(0) = \mathfrak{p}$ is maximal with respect to excluding S. In particular, if $a \in R$ is nonzero and satisfies $(0) < (a)$, we have $(a) \cap S \neq \emptyset$. We conclude that there exists an $r \in R$ such that $ar \in S$. As S is saturated by the claim, $a \in S$ as needed. □

Remarks 28.2. 1. If R is a Noetherian ring, then, *a priori*, we do not need Zorn's Lemma to prove Krull's Theorem, hence in the proof above, if R is a Noetherian domain. However, as mentioned in Remark 27.12 this is really illusory.

2. We shall see that $\mathbb{C}[t_1, \ldots, t_n]$ is a UFD, Using this one can show that the geometric meaning of Kaplansky's Theorem in the case of $\mathbb{C}^n$ is: Any hypersurface in $\mathbb{C}^n$ given by polynomial equations (in several variables) satisfying an irreducible condition can be defined by a single irreducible polynomial.

Exercises 28.3. 1. Prove Krull's Theorem 25.13 if R is a Noetherian ring and Kaplansky's Theorem 28.1 if R is a Noetherian domain without directly using Zorn's Lemma (cf. Remark 27.12).

2. Let $S \subset R$ be a multiplicative set such that $0 \notin S$. As above, we say that S is *saturated* if $ab \in S$ implies that $a, b \in S$. Prove that S is a saturated multiplicative set if and only if $R \setminus S$ is a union of prime ideals. In particular, the set of zero divisors in a commutative ring is a union of prime ideals.

3. Let R be a UFD and $\mathfrak{p}$ a nonzero prime in R. Then $\mathfrak{p}$ properly contains no nonzero prime ideal if and only if $\mathfrak{p}$ is principal and generated by a prime element. Moreover prime elements generate all such prime ideals in R.

29. Gaussian Integers

In this section, we study the Gaussian integers $\mathbb{Z}[\sqrt{-1}]$, a domain as it is a subring of $\mathbb{C}$. The map $\bar{} : \mathbb{C} \to \mathbb{C}$ given by $x + y\sqrt{-1} \mapsto x - y\sqrt{-1}$, $x, y \in \mathbb{R}$, will denote *complex conjugation*. (As $\mathbb{C}$ is a

real vector space on $\{1, \sqrt{-1}\}$, any element z in $\mathbb{C}$ can be written $z = x + y\sqrt{-1}$ for some unique real numbers x and y. This is a field automorphism that *fixes* the reals, i.e., is the identity on the reals. In fact, $\overline{z} = z$ if and only if $z \in \mathbb{R}$. Clearly, it restricts to a ring automorphism of the Gaussian integers. Let $N : \mathbb{C} \to \mathbb{R}^+ \cup \{0\}$ be the map defined by $z \mapsto z\overline{z}$. It is called the *norm map*.

Remarks 29.1. 1. If R is a ring, we let $\mathrm{Aut}_{ring}(R)$ denote the set of ring automorphisms of R. It is a group under composition. As $\overline{\overline{z}} = z$ for all $z \in \mathbb{C}$, complex conjugation has order two in $\mathrm{Aut}_{ring}(\mathbb{C})$. A ring automorphism of commutative rings of order one or two is called an *involution*. So complex conjugation is an involution on $\mathbb{C}$ and on $\mathbb{Z}[\sqrt{-1}]$.

2. Let $x, y \in \mathbb{R}$, then $N(x + y\sqrt{-1}) = x^2 + y^2$, which is positive unless $x = 0 = y$. This indicates that the norm map should be of interest when studying sums of two squares.

3. $N : (\mathbb{C}^\times, \cdot) \to (\mathbb{R}^\times, \cdot)$ is a monoid homomorphism, i.e., $N(1) = 1$ and it is *multiplicative*:

$$N(z_1 z_2) = N(z_1)N(z_2) \quad \text{for all} \quad z_1, z_2 \in \mathbb{C}^\times.$$

In particular, if $x_1, x_2, y_1, y_2 \in \mathbb{C}$, we have the two square identity:

$$(x_1^2 + y_1^2)(x_2^2 + y_2^2) = (x_1x_2 - y_1y_2)^2 + (x_1y_2 + x_2y_1)^2,$$

the product of sums of two squares is a sum of two squares. Of course, once we have this identity, we see that this holds in any commutative ring by multiplying out.

4. $N(z) = N(\overline{z})$ for all $z \in \mathbb{C}$.

Lemma 29.2. *$\mathbb{Z}[\sqrt{-1}]$ is a domain and satisfies the following:*

(1) $N|_{Z[\sqrt{-1}]} : \mathbb{Z}[\sqrt{-1}] \to \mathbb{Z}^+ \cup \{0\}$ and satisfies $N|_{\mathbb{Z}[\sqrt{-1}]}(\alpha) = 0$ if and only if $\alpha = 0$. We shall write N for $N|_{\mathbb{Z}[\sqrt{-1}]}$.

(2) $\mathbb{Z}[\sqrt{-1}]^\times = \{\pm 1, \pm\sqrt{-1}\}$.

(3) Let $a, b \in \mathbb{Z}$. Then $a+b\sqrt{-1}$, is a root of the monic polynomial (i.e., leading coefficient is one), $t^2 - 2at + (a^2 + b^2) \in \mathbb{Z}[t]$.

Proof. We only prove (2) leaving the rest as an easy check. If $uv = 1$ with $u, v \in \mathbb{Z}[t]$, then $1 = N(1) = N(u)N(v)$, so we have $N(u) \in \mathbb{Z}^\times \cap \mathbb{Z}^+ = \{1\}$. As $x^2 + y^2 = 1$ with $x, y \in \mathbb{Z}$ if and only if $x = \pm 1$, $y = 0$ or $x = 0, y = \pm 1$, we conclude that

$$\mathbb{Z}[\sqrt{-1}]^\times \subset \{\beta \mid N(\beta) = 1\} \subset \{\pm 1, \pm\sqrt{-1}\}.$$

Clearly, $\{\pm 1, \pm\sqrt{-1}\} \subset \mathbb{Z}[\sqrt{-1}]^\times$, establishing the lemma. □

Theorem 29.3. *The domain $\mathbb{Z}[\sqrt{-1}]$ is a (strong) euclidean domain, hence a PID and so a UFD.*

PROOF. **Claim**. N is a euclidean function on $\mathbb{Z}[\sqrt{-1}]$. (Consequently, $\mathbb{Z}[\sqrt{-1}]$ is a strong euclidean domain.):

Let α and β be elements in $\mathbb{Z}[\sqrt{-1}]$ with β nonzero. We must show that there exist a γ and ρ in $\mathbb{Z}[\sqrt{-1}]$ satisfying $\alpha = \beta\gamma + \rho$ with $\rho = 0$ or $N(\rho) < N(\beta)$.

In $\mathbb{C}$, we can write

$$\frac{\alpha}{\beta} = \frac{\alpha}{\beta}\frac{\overline{\beta}}{\overline{\beta}} = \frac{\alpha\overline{\beta}}{N(\beta)} = r + s\sqrt{-1}$$

for some $r, s \in \mathbb{Q}$ (as $N(\beta) \in \mathbb{Z}$). Choose $m, n \in \mathbb{Z}$ satisfying

$$|r - m| \leq 1/2 \text{ and } |s - n| \leq 1/2.$$

Set $\gamma = m + n\sqrt{-1}$ and $\rho = \alpha - \beta\gamma$ in $\mathbb{Z}[\sqrt{-1}]$. If $\rho = 0$, we are done, so suppose not. As $N : \mathbb{C}^\times \to \mathbb{R}^+$ and is multiplicative, we have

$$\begin{aligned} N(\rho) &= N(\alpha - \beta\gamma) = N\big(\beta(\frac{\alpha}{\beta} - \gamma)\big) = N(\beta)N(\frac{\alpha}{\beta} - \gamma) \\ &= N(\beta)N\big((r-m) + (s-n)\sqrt{-1}\big) = N(\beta)[(r-m)^2 + (s-n)^2] \\ &\leq N(\beta)(\frac{1}{4} + \frac{1}{4}) < N(\beta). \end{aligned}$$

□

Note in the last line of the proof, our estimate on $N(\rho)$ was $\leq \frac{1}{2}N(\beta)$, so we had some room to work. This explains why the domains in Remark 27.21(3) work — compute the restriction of N to each of these domains.

As mentioned before, we can use the Gaussian integers to study sums of squares of integers. In particular, $m = x^2 + y^2$ with x, y integers if and only if $m = N(x + y\sqrt{-1})$. We wish to determine all the non-negative integers that are sums of two squares. We have seen that a product of two integers each a sum of two squares of integers is itself a sum of two squares. By the Fundamental Theorem of Arithmetic, each non-negative integer factors as $p_1^{e_1} \cdots p_r^{e_r}$ with the p_i distinct positive primes and the $e_i \geq 1$. If e_i is even then $p_i^{e_i}$ is a square (so a sum of two squares). Hence we are reduced to determining which products of distinct primes are sums of two squares of integers. In particular, we must determine which primes are sums of two squares of integers. As 2 is a sum of two squares, we are reduced to odd primes. If $x \in \mathbb{Z}$, then $x^2 \equiv 0, 1 \mod 4$. In particular, a sum of two integer squares can only be congruent to $0, 1$ or 2 modulo four. In particular, no integer n

satisfying $n \equiv 3 \mod 4$ can be a sum of two squares of integers, e.g., 3 and 7 are not sums of two squares.

The key to the solution is the following:

Lemma 29.4. *Let p be a positive prime in $\mathbb{Z}$ satisfying $p \equiv 1 \mod 4$. Then there exists an integer x satisfying $p \mid x^2 + 1$ in $\mathbb{Z}$. In particular, -1 is a square modulo p if $p \equiv 1 \mod 4$.*

PROOF. Let $x = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot \frac{p-1}{2}$ in $\mathbb{Z}$. By Wilson's Theorem (Exercise 10.16(11)), we have

$$\begin{aligned} -1 \equiv (p-1)! &= (1 \cdot 2 \cdot 3 \cdots \frac{p-1}{2})(p - \frac{p-1}{2} \cdots p-3 \cdot p-2 \cdot p-1) \\ &\equiv (1 \cdot 2 \cdot 3 \cdots \frac{p-1}{2})(-\frac{p-1}{2} \cdots -3 \cdot -2 \cdot -1) = (-1)^{\frac{p-1}{2}} x^2 \\ &\equiv x^2 \mod p, \end{aligned}$$

so $p \mid x^2 + 1$. □

Theorem 29.5. (Fermat) *Let p be a positive prime congruent to 1 modulo 4. Then there exist integers x and y such that $p = x^2 + y^2$.*

PROOF. By the lemma there exists an integer x such that $p \mid x^2+1$. Hence $p \mid (x + \sqrt{-1})(x - \sqrt{-1})$ in the UFD $\mathbb{Z}[\sqrt{-1}]$.

Claim. p is not irreducible (i.e., it is *reducible*) in $\mathbb{Z}[\sqrt{-1}]$:

Suppose that p is irreducible in $\mathbb{Z}[\sqrt{-1}]$, then as $\mathbb{Z}[\sqrt{-1}]$ is a UFD, it is a prime. Hence $p \mid x + \sqrt{-1}$ or $p \mid x - \sqrt{-1}$. If $p \mid x + \sqrt{-1}$, then $p(a + b\sqrt{-1}) = x + \sqrt{-1}$ for some integers a and b. It follows that $pb = 1$ in $\mathbb{Z}$ (why?), which is impossible. [Or if $p \mid x + \sqrt{-1}$, then $p = \overline{p} \mid x - \sqrt{-1}$ also. It follows that $p \mid 2x$ hence $p \mid x$ as p is odd.] Similarly, $p \not\mid x - \sqrt{-1}$, so p is not irreducible in $\mathbb{Z}[\sqrt{-1}]$, proving the claim.

Using the claim, we can write $p = \pi\alpha$ for some $\pi, \alpha \in \mathbb{Z}[\sqrt{-1}]$, with π irreducible in $\mathbb{Z}[\sqrt{-1}]$ and α not a unit in $\mathbb{Z}[\sqrt{-1}]$. Taking the norm of this equation yields

$$p^2 = N(p) = N(\pi)N(\alpha).$$

in $\mathbb{Z}$. As $N(\pi)$ and $N(\alpha)$ lie in $\mathbb{Z}^+$ with $\mathbb{Z}$ a UFD, we must have

$$\pi\overline{\pi} = N(\pi) = p = N(\alpha)$$

is a sum of two squares, proving the theorem. In addition, we have also shown, in view of this last equation, that $\overline{\pi} = \alpha$ is also irreducible and $p = \pi\overline{\pi}$ is a factorization of p in $\mathbb{Z}[\sqrt{-1}]$. □

To categorize those integers that are sums of two squares, we need the following lemma:

Lemma 29.6. *Let a and b be two nonzero relatively prime integers and p a (positive) odd prime such that $p \mid a^2 + b^2$. Then $p \equiv 1 \mod 4$.*

PROOF. As $p \mid a^2 + b^2$, we have $p \mid a$ if and only if $p \mid b$. Since a and b are relatively prime, p cannot divide a or b. By Fermat's Little Theorem 10.14, we know that $b^{p-3}b^2 = b^{p-1} \equiv 1 \mod p$. Consequently, $a^2 + b^2 \equiv 0 \mod p$, i.e., $a^2 \equiv -b^2 \mod p$, and $p - 3$ even imply that

$$(b^{\frac{p-3}{2}}a)^2 = b^{p-3}a^2 \equiv -b^{p-3}b^2 \equiv -1 \mod p.$$

Let $x = b^{\frac{p-3}{2}}a$. Then $x^2 \equiv -1 \mod p$ which means that

$$1 \equiv x^{p-1} = (x^2)^{\frac{p-1}{2}} \equiv (-1)^{\frac{p-1}{2}} \mod p.$$

If $p \equiv 3 \mod 4$, then we conclude that $1 \equiv -1 \mod p$, which is impossible. □

The next result solves our stated problem. We leave its proof as an exercise (that you should do).

Theorem 29.7. *Let $n = p_1^{e_1} \cdots p_r^{e_r}$ in $\mathbb{Z}$ with $0 < p_1 < \cdots < p_r$ primes, and $e_i > 0$ for $i = 1, \ldots, r$. Then n is a sum of two squares in $\mathbb{Z}$ if and only if e_i are even integers whenever $p_i \equiv 3 \mod 4$ for $i = 1, \ldots, r$.*

Let p be a (positive) odd prime. By Lemma 29.6, we know that

$$-1 \text{ is } \begin{cases} \text{a square mod } p & \text{if } p \equiv 1 \mod 4. \\ \text{not a square mod } p & \text{if } p \equiv 3 \mod 4. \end{cases}$$

Definition 29.8. If a is an integer not divisible by p, define the *Legendre symbol*

$$\left(\frac{a}{p}\right) := \begin{cases} +1 \text{ if } a \text{ mod } p \text{ is a square.} \\ -1 \text{ if } a \text{ mod } p \text{ is not a square.} \end{cases}$$

It is also convenient to define $\left(\frac{b}{p}\right) = 0$ if $p \mid b$.

Example 29.9. $\left(\frac{0}{7}\right) = 0$, $\left(\frac{1}{7}\right) = \left(\frac{2}{7}\right) = \left(\frac{4}{7}\right) = 1$, and $\left(\frac{3}{7}\right) = \left(\frac{5}{7}\right) = \left(\frac{6}{7}\right) = -1$

We have proven:

Proposition 29.10. (Euler's Formula) *Let p be a (positive) odd prime. Then*

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = \begin{cases} +1 \text{ if } p \equiv 1 \mod 4. \\ -1 \text{ if } p \equiv 3 \mod 4. \end{cases}$$

Corollary 29.11. *The set* $\{4n+1 \mid n \in \mathbb{Z}\}$ *contains infinitely many primes.*

PROOF. We prove the stronger result that the set $\{8n+5 \mid n \in \mathbb{Z}\}$ contains infinitely many primes.

Check. If $n \in \mathbb{Z}$, then $n^2 \equiv 0, 1, 4 \mod 8$. In particular, if n is odd, then $n^2 \equiv 1 \mod 8$.

Let p be an odd prime and set $N = (3 \cdot 5 \cdot 7 \cdots p)^2 + 2^2 \equiv 5 \mod 8$. If q is an odd prime such that $q \mid N$, then by Lemma 29.6, we have $q \equiv 1 \mod 4$, hence $q \equiv 1 \mod 8$ or $q \equiv 5 \mod 8$. If N is a product of primes all congruent to 1 modulo 8, then $N \equiv 1 \mod 8$, a contradiction. Therefore, there exists a prime q satisfying $q \equiv 5 \mod 8$ and $q \mid N$. As none of the primes $3, 5, 7, \ldots, p$ divides N, we must have $q > p$, so there exists a prime larger that p congruent to 5 modulo 8. □

Remark 29.12. It is still an open question whether the set of integers $\{x^2+1 \mid x \in \mathbb{Z}\}$ contains infinitely many primes.

The following is a well-known theorem, although we shall not prove it. The usual proof uses complex analysis, although there is now an elementary (but tricky) proof that does not.

Theorem 29.13. (Dirichlet's Theorem on Primes in an Arithmetic Progression) *Let a and b be two nonzero relatively prime integers. Then there exist infinitely many primes p satisfying $p \equiv a \mod b$.*

We shall prove that this is true for $a = 1$ in Proposition 56.10 below by elementary means. Other questions arise about *arithmetic progressions*, i.e., equally spaced integers, e.g., $a, a+b, a+2b, a+3b, \ldots$ with a and b integers. The distance between two consecutive integers in an arithmetic progression is called the *spacing* of the progression, e.g., b is the spacing in the example. Such a progression can be finite or infinite. The most spectacular theorem is the following proven in 2004:

Theorem 29.14. *(*Green-Tao) *Let N be a positive integer. Then there exists an arithmetic progression consisting of N primes.*

The theorem shows that for any N there exists infinitely many arithmetic progressions consisting of N primes, but it does not indicate anything about the spacing. Turning this theorem around, one can ask does there exist infinitely many primes in an arithmetic progression with a given spacing. The most famous example of this is

Question 29.15. *(*Twin Prime Conjecture) *Do there exist infinitely many twin prime, i.e., primes pairs of the form* $p, p+2$.

Euler had shown that $\sum_{p \text{ a prime}} \frac{1}{p}$ was infinite, so there was some hope that $\sum_{p \text{ a twin prime}} \frac{1}{p}$ was also. Brun showed this to be false in 1915.

We shall also study the Legendre symbol further proving the following wonderful (albeit seemingly strange) theorem of Gauss upon which you should meditate (and we shall prove it in Theorem 56.18 below):

Theorem 29.16. (Quadratic Reciprocity) *Let* p *and* q *be distinct odd primes. Then*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}.$$

Exercises 29.17. 1. Show that the *evaluation map* $e_{\sqrt{-1}} : \mathbb{Z}[t] \to \mathbb{C}$ defined by $f \mapsto f(\sqrt{-1})$ is a ring homomorphism with kernel (t^2+1) and image the Gaussian integers. [Hint: Use Exercise 27.22(14).]

2. Let $R = \mathbb{Z}[\sqrt{-1}]$ and $n = p_1^{e_1} \cdots p_r^{e_r}$ be the standard factorization of the integer $n > 1$. Show that the following are equivalent:
   (*i*) n is a sum of two squares.
   (*ii*) $n = N(\alpha)$ for some $\alpha \in R$.
   (*iii*) If $p_i \equiv 3 \pmod 4$, then e_i is even.
   [Hint: Use Lemma 29.6.]

3. Show the following
   (*i*) Let α be an element in $\mathbb{Z}[\sqrt{-1}]$ such that $N(\alpha)$ is a prime or the square of a prime in $\mathbb{Z}$. Then α is a prime element in $\mathbb{Z}[\sqrt{-1}]$.
   (*ii*) Let π be an prime element in $\mathbb{Z}[\sqrt{-1}]$. Show $N(\pi)$ is a prime or the square of a prime in $\mathbb{Z}$.

4. Determine all prime elements, up to units, in $\mathbb{Z}[\sqrt{-1}]$.

5. Prove that $\mathbb{Z}\left[\frac{-1-\sqrt{-3}}{2}\right]$ and $\mathbb{Z}[\sqrt{3}\,]$ are (strong) euclidean domains.

6. Let $R = \mathbb{Z}[\sqrt{-d}] = \{a + b\sqrt{-d} \,|\, a, b \in \mathbb{Z}\}$, a subring of $\mathbb{C}$ with d a positive square-free integer. Let $N : R \to \mathbb{Z}$ be the norm map, so $\alpha = a + b\sqrt{-d} \mapsto \alpha\overline{\alpha} = a^2 + db^2$. Show all of the following:
   (*i*) The field of quotients of R is $\mathbb{Q}[\sqrt{-d}] = \{a + b\sqrt{-d} \,|\, a, b \in \mathbb{Q}\}$.
   (*ii*) $N : R \setminus \{0\} \to \mathbb{Z}$ is a monoid homomorphism.
   (*iii*) $R^\times = \{\alpha \in R \,|\, N(\alpha) = 1\}$ and compute this group for all d.

(*iv*) The element α is irreducible in R if $N(\alpha)$ is a prime. Is the converse true?

(*v*) Suppose $d \geq 3$, then 2 is irreducible but not prime in R.

7. Show that $\mathbb{Z}[\sqrt{-2}]$ is a euclidean domain.
8. Let $R = \mathbb{Z}[\sqrt{-5}]$. Show the following:
   (*i*) The elements $2, 3, 1 + \sqrt{-5}$, and $1 - \sqrt{-5}$ are all irreducible but no two are associates.
   (*ii*) None of the elements $2, 3, 1 + \sqrt{-5}$, and $1 - \sqrt{-5}$ are prime. In particular, R is not a UFD.
9. Let $R = \mathbb{Z}[\sqrt{-5}]$. Let $\mathfrak{P} = (2, 1 + \sqrt{-5})$. Show
   (*i*) $\mathfrak{P}^2 = (2)$ in R.
   (*ii*) $\mathfrak{P}$ is a maximal ideal.
   (*iii*) $\mathfrak{P}$ is not a principal ideal.

30. Addendum: The Four Square Theorem

In the previous section, we characterized those positive elements that are sums of two squares. In this section, we prove Lagrange's theorem that every positive integer is a sum of four squares. The proof is based on a computation of Euler that shows a product of two integers each a sum four integer squares is a sum of four integer squares. Although he did not know it at the time, this formula comes from the Hamiltonian quaternions, the first known division ring that was not commutative. We begin by constructing it.

Construction 30.1. Let $\mathcal{H}$ be a four dimensional real vector space with basis $\{1, i, j, k\}$. We view 1 as $1_{\mathbb{R}}$, so $\mathbb{R} \subset \mathcal{H}$. We make $\mathcal{H}$ into a ring by defining a multiplication on this basis and extend linearly to the whole space. The multiplication of the basis elements is given by:

$$i^2 = -1, \;\; j^2 = -1, \;\; ij = -ji = k.$$

Extending this linearly give a multiplication on $\mathcal{H}$ as follows:

$$\begin{aligned}(x_0 1 + x_1 i + x_2 j + x_3 k) \cdot (y_0 1 + y_1 i + y_2 j + y_3 k) \\ = (x_0y_0 - x_1y_1 + x_2y_2 - x_3y_3)1 + (x_0y_1 - x_1y_0 + x_2y_3 - x_3y_2)i \\ + (x_0y_2 - x_2y_0 + x_3y_1 - x_1y_3)j + (x_0y_3 - x_3y_0 + x_1y_2 - x_2y_1)k.\end{aligned}$$

By a straight-forward (but arduous) computation, one now checks that this makes $\mathcal{H}$ into a ring with $1_{\mathcal{H}} = 1_{\mathbb{R}}$. Note that the subring generated by $\{1, i\}$ is isomorphic to $\mathbb{C}$ (as are the subrings generated by $\{1, j\}$ and $\{1, k\}$, respectively). The ring $\mathcal{H}$ is not commutative as its *center*, i.e., the subring $\{x \in \mathcal{H} \mid xy = yx \text{ for all } y \text{ in } \mathcal{H}\}$ of $\mathcal{H}$, is precisely

$\mathbb{R}$. (Why is the center a subring?) Note also that $\mathcal{H}^\times$ contains the *quaternion group* $\{1, i, j, ij, -1, -i, -j, -ji\}$.

Analogous to $\mathbb{C}$, there is a *quaternion conjugation* map:

$$^- : \mathcal{H} \to \mathcal{H} \text{ given by } \overline{x_0 1 + x_1 i + x_2 j + x_3 k} := x_0 1 - x_1 i - x_2 j - x_3 k.$$

It is easy to check that for all x, y in $\mathcal{H}$, we have

$$\begin{aligned} \overline{1} &= 1 \\ \overline{x+y} &= \overline{x} + \overline{y} \\ \overline{xy} &= \overline{y}\ \overline{x}. \end{aligned}$$

Therefore, the quaternion conjugation map satisfies the properties of a ring homomorphism except that it reverses multiplication. Such a map of rings is called a *ring antihomomorphism*. Clearly, this map is also a surjective linear transformation of a real vector space. Hence it is bijective, since $\mathcal{H}$ is a finite dimensional vector space over $\mathbb{R}$. Consequently, the map is, in fact, a *ring antiautomorphism*. (Note the composition of two ring antihomomorphisms is a ring homomorphism.) In addition, for all x in $\mathcal{H}$, we have

$$\overline{\overline{x}} = x,$$

so $^-$ is an *involution*, i.e., an antiautomorphism satisfying the composition $^- \circ ^-$ is the identity. (Cf. with complex conjugation.) We also have a *norm map*:

$$N : \mathcal{H} \to \mathbb{R}^+ \cup \{0\} \text{ given by } z \mapsto z\overline{z},$$

that is checked to be multiplicative, i.e., $N(xy) = N(x)N(y)$ for all x, y in $\mathcal{H}$. (Note that if we identify $\mathbb{C}$ as the subring of $\mathcal{H}$ generated by $\{1, i\}$, then $N|_{\mathbb{C}}$ is the usual norm map on $\mathbb{C}$.) This is useful as the norm of $x_0 1 + x_1 i + x_2 j + x_3 k$ is

$$N(x_0 1 + x_1 i + x_2 j + x_3 k) = x_0^2 + x_1^2 + x_2^2 + x_3^2,$$

a sum of four squares. The multiplicativity of this norm map yields *Euler's Equation*: the four square identity

$$\begin{aligned} &(x_0^2 + x_1^2 + x_2^2 + x_3^2) \cdot (y_0^2 + y_1^2 + y_2^2 + y_3^2) \\ &\quad = (x_0 y_0 - x_1 y_1 + x_2 y_2 - x_3 y_3)^2 + (x_0 y_1 - x_1 y_0 + x_2 y_3 - x_3 y_2)^2 \\ &\quad + (x_0 y_2 - x_2 y_0 + x_3 y_1 - x_1 y_3)^2 + (x_0 y_3 - x_3 y_0 + x_1 y_2 - x_2 y_1)^2, \end{aligned}$$

the formula that a product of two sums of fours squares is a sum of four squares, generalizing the norm on the complex numbers giving a formula for the product of sums of two squares. This is the equation that we shall need to prove Lagrange's Theorem that any positive integer is a sum of four squares. Of course, we do not need the quaternions to

establish Euler's Equation — we need only multiply the right hand side out to see that it holds in any commutative ring. Indeed Euler found this equation without knowing about quaternions. It does, however, explain why such an equation exists. It can also be shown that there is a formula that a product of two sums of eight squares is a sum of eight squares. It arises from a generalization of the quaternions called the *octonians.*, a algebraic structure that looks a division ring but does not satisfy the associative law for multiplication. So the two square identity arises from fields, the four square identity from the quaternions (tossed out commutativity). the eight square identity from the octonians (toss out commutativity and associativity). Is there a sixteen square identity arising from a generalization of these algebraic structures by throwing out some other property of rings? Hurwitz showed not.

As with the norm map on $\mathbb{C}$, it follows that $N(z) = 0$ if and only if $z = 0$. In particular, if z is nonzero, then $z\overline{z}/N(z) = 1 = \overline{z}z/N(z)$, so z has a multiplicative inverse. Therefore, $\mathcal{H}$ is a division ring called the *Hamiltonian quaternions*. By definition, $ij \neq ji$, so $\mathcal{H}$ is not a field.

Remark 30.2. Let V be a finite dimensional real vector space that is also a division ring and contains $\mathbb{R}$ in its center. Then it can be shown that V is ring isomorphic to $\mathbb{R}$, $\mathbb{C}$, or $\mathcal{H}$. In particular, $\dim_{\mathbb{R}} V = 1, 2,$ or 4. We shall show this in Section §89.

To show that every positive integer is a sum of four squares, we shall use the following:

Lemma 30.3. *$1 \leq m < p$ and mp is a sum of four integer squares.*

Proof. Let

$$S_1 := \{0^2, 1^2, \ldots, \frac{(p-1)}{2}^2\}$$

$$S_2 := \{-0^2 - 1, -1^2 - 1, \ldots, -\frac{(p-1)}{2}^2 - 1\}.$$

If $x^2 = y^2 \mod p$ with x, y integers, then $p \mid x^2 - y^2 = (x+y)(x-y)$. Consequently, $p \mid x + y$ or $p \mid x - y$. It follows (check) that no two elements in S_1 are congruent modulo p and no two elements in S_2 are congruent modulo p. Since

$$|S_1| + |S_2| = \left(\frac{p-1}{2} + 1\right) + \left(\frac{p-1}{2} + 1\right) = p + 1,$$

there exist $x^2 \in S_1$ and $-y^2 - 1 \in S_2$ satisfying $0 \neq x^2 + y^2 + 1$ and

$$x^2 + y^2 + 1 \equiv 0 \mod p.$$

Hence there exists a positive integer m such that

$$mp = x^2 + y^2 + 1 \leq \left(\frac{p-1}{2}\right)^2 + \left(\frac{p-1}{2}\right)^2 + 1 < \frac{p^2}{2} + 1 < p^2.$$

So $1 \leq m < p$. □

Remarks 30.4. 1. Of course, the proof shows that the element mp is a sum of three squares.

2. The same proof shows that -1 is a sum of two squares in $\mathbb{Z}/p\mathbb{Z}$ for any prime p. In fact, any element a in $\mathbb{Z}/p\mathbb{Z}$ is a sum of two square — replace the -1 in the proof by a in S_1.

3. A similar (but not obvious) proof works to show that any element in a finite field is a sum of two squares.
(Note that if $\text{char}(F) = 2$, then the map $F \to F$ given by $x \mapsto x^2$ is a (field) monomorphism, so an automorphism if F is a finite field.)

Historically, one of the main forms of induction used to prove rather deep theorems was the idea of Fermat, called *Fermat descent*. We prove our theorem about four squares using this method.

Theorem 30.5. (Lagrange). *Every positive integer is a sum of four squares of integers.*

PROOF. We know that $2 = 1 + 1 + 0 + 0$ is a sum of four squares, so by Euler's Formula and the Fundamental Theorem of Arithmetic it suffices to show every positive odd prime is a sum of four squares. (We even know this is true for primes congruent to 1 modulo 4, but will not use this.) Let p be an odd prime. By the lemma and the Well-ordering Principle there exists an element $(x_0, y_0, z_0, w_0, m) \in \mathbb{Z}^5$ with $1 \leq m < p$ and satisfying

$$(*) \qquad x_0^2 + y_0^2 + z_0^2 + w_0^2 = mp$$

with m minimial. To finish, it suffices to show that $m = 1$.

Case 1. m is even:

By (*), an even number of the integers x_0, y_0, z_0, w_0 must be odd (if any are odd). If precisely two of these are odd, we may assume that they are x_0, y_0. So we have $x_0 \equiv y_0 \mod 2$ and $z_0 \equiv w_0 \mod 2$, hence

$$(\frac{x_0 + y_0}{2})^2 + (\frac{x_0 - y_2}{2})^2 + (\frac{z_0 + w_0}{2})^2 + (\frac{z_0 - w_0}{2})^2 = \frac{m}{2}p,$$

contradiction the minimality of m.

Case 2. m is odd:

We use the modification of the Division Algorithm given by Exercise 4.17(3). Using this modified division algorithm, we have equations:

$$\begin{aligned} x_0 &= mx_1 + x_2 \quad \text{with} \ \ |x_2| < \frac{m}{2}. \\ y_0 &= my_1 + y_2 \quad \text{with} \ \ |y_2| < \frac{m}{2}. \\ z_0 &= mz_1 + z_2 \quad \text{with} \ \ |z_2| < \frac{m}{2}. \\ w_0 &= mw_1 + w_2 \quad \text{with} \ \ |w_2| < \frac{m}{2}. \end{aligned}$$

Therefore, we have

$$x_2^2 + y_2^2 + z_2^2 + w_2^2 \equiv x_0^2 + y_0^2 + z_0^2 + w_0^2 \equiv 0 \mod m,$$

so

$$(\dagger) \qquad x_2^2 + y_2^2 + z_2^2 + w_2^2 = Mm$$

for some integer M.

Subcase 1. $M = 0$:
In this subcase, we must have $x_2 = y_2 = z_2 = w_2 = 0$ and $m \mid x_0$, $m \mid y_0$, $m \mid z_0$, $m \mid w_0$, so $m^2 \mid x_0^2 + y_0^2 + z_0^2 + w_0^2 = mp$. Hence $m \mid p$ with $1 \leq m < p$. If follows that $m = 1$, so we are done in this subcase.

Subcase 2. $M > 0$:
We have $(x_1^2 + y_1^2 + z_1^2 + w_1^2) \cdot (x_2^2 + y_2^2 + z_2^2 + w_2^2) = A^2 + B^2 + C^2 + D^2$ by Euler's Equation with $A = x_1x_2 + y_1y_2 + z_1z_2 + w_1w_2$ and for some integers B, C, D. Consequently,

$$\begin{aligned} mp &= x_0^2 + y_0^2 + z_0^2 + w_0^2 \\ &= (mx_1 + x_2)^2 + (my_1 + y_2)^2 + (mz_1 + z_2)^2 + (mw_1 + w_2)^2 \\ &= m^2(x_1^2 + y_1^2 + z_1^2 + w_1^2) + 2mA + mM. \end{aligned}$$

Multiply this equation by M/m and use $(\dagger)$ to obtain the following equation in $\mathbb{Z}$:

$$\begin{aligned} Mp &= mM(x_1^2 + y_1^2 + z_1^2 + w_1^2) + 2MA + M^2 \\ &= (x_2^2 + y_2^2 + z_2^2 + w_2^2)(x_1^2 + y_1^2 + z_1^2 + w_1^2) + 2MA + M^2 \\ &= A^2 + B^2 + C^2 + D^2 + 2MA + M^2 = (A + M)^2 + B^2 + C^2 + D^2 \end{aligned}$$

is a sum of four squares with

$$0 < mM = x_2^2 + y_2^2 + z_2^2 + w_2^2 < 4(\frac{m}{2})^2 = m^2.$$

Therefore, $0 < M < m$, contradicting the choice of m. $\square$

Remark 30.6. It is much harder to determine which positive integers are sums of three squares. This was determined by Gauss, who showed that a positive integer n is a sum of three squares if and only if $n \neq 4^m(8k+7)$ for some non-negative integers m and k.
[Note that $3 = 1+1+1$ and $13 = 2^2+3^2$ but $39 = 3 \cdot 13$ is a sum of four squares (as $39 = 1^2+2^2+3^2+5^2$), but not fewer.]

Hilbert proved a much more general theorem when he solved the *Waring Problem* that given any positive integer n there exists a minimal positive integer $g(n)$ such that every positive integer is a sum of $g(n)$ nth powers. His proof is not constructive, and it is a difficult problem (and open for most n) to determine $g(n)$.

Exercises 30.7. 1. Show $\mathcal{H}$ is a ring and prove quaternion conjugation is multiplicative.

2. Let $z = x_0 1 + x_1 i + x_2 j + x_3 k$. We call z a *pure quaternion* if $x_0 = 0$. Suppose z is nonzero. Show that z is a pure quaternion if and only if z does not lie in $\mathbb{Q}$ but z^2 does lie in $\mathbb{Q}$.

3. Generalize the construction of quaternions as follows: Let F be any field of characteristic different from two and a and b elements of F. Let Q be the 4-dimensional F-vector space on basis $\{1, i, j, k\}$. Make Q into a ring by defining
$$i^2 = a, \;\; j^2 = b, \;\; ij = -ji = k$$
and extend linearly. Define the *conjugate* of z by
$$\overline{x_0 1 + x_1 i + x_2 j + x_3 k} := x_0 1 - x_1 i - x_2 j - x_3 k$$
and the norm map
$$N : Q \to F \;\; \text{by} \;\; z \mapsto z\overline{z}$$
and just as for the Hamiltonian quaternions it is multiplicative. Show that
$$N(x_0 1 + x_1 i + x_2 j + x_3 k) = x_0^2 - a x_1^2 - b x_2^2 + ab x_3^2.$$
We call Q a *general quaternion algebra*. In particular, Q is a division ring if it satisfies the property that $N(z) = 0$ if and only if $z = 0$. (If this does not happen, then $Q \cong \mathbb{M}_2(F)$.) Can you give non-isomorphic examples, of quaternion algebras that are division rings over $\mathbb{Q}$?

4. Show if an integer n satisfies $n = 4^m(8k+7)$, then n is not a sum of three squares. [The converse is the hard part.]

CHAPTER VII

Polynomial Rings

In this chapter we study polynomial rings over rings, concentrating on the case of a polynomial ring in one variable over a domain, e.g., a field. Most importantly, we prove Kronecker's Theorem that a non-constant polynomial over a field F (in one variable) has a root in a field containing F. This will be a basic result when we turn to studying Field Theory. We then investigate polynomial rings over a UFD. We prove that any such ring is itself a UFD using a classical result of Gauss. In the final section of this chapter, we attempt to motivate the commutative algebra needed for algebraic geometry that will be studied more systematically in a later chapter, that is the study of zeros of polynomials in many variables over a field, especially algebraically closed fields.

31. Introduction to Polynomial Rings

We start by explicitly defining polynomial rings. If R is a ring, let $R[t] := \{a_0 + a_1t + \cdots + a_nt^n \mid a_i \in R,\ i = 1, \ldots n,\ \text{ some } n \geq 0\}$ with $a_0+a_1t+\cdots+a_nt^n = 0$ if and only if $a_i = 0$ for $i = 0, \ldots, n$. Therefore, if

$$f = a_0 + a_1t + \cdots + a_nt^n \text{ and } g = b_0 + b_1t + \cdots + b_mt^m,$$

then

$$f = g \text{ if and only if } a_i = b_i \text{ for all } i$$

[where we let $a_i = 0$ for all $i > n$ and $b_j = 0$ for all $j > m$].

With f and g as above, define

$$f + g := (a_0 + b_0) + (a_1 + b_1)t + \cdots = \sum (a_i + b_i)t^i$$

and

$$fg := \sum c_it^i \text{ with } c_i = \sum_{k=0}^{i} a_{i-k}b_k.$$

This makes $R[t]$ into a ring called the *polynomial ring* over R with $0_{R[t]} = 0_R + 0t + \cdots$ and $1_{R[t]} = 1_R + 0_Rt + \cdots$, where we view $R \subset R[t]$ by identifying r and $r1_{R[t]}$ for all $r \in R$, i.e., identifying the ring monomorphism $R \to R[t]$ given by $r \mapsto r1_{\mathbb{R}[t]}$ as the inclusion. So R is

just the set (and subring) of constant polynomials in $R[t]$. Recursively define the *polynomial ring* in n *variables* $t_1, \ldots, t_n$ over R by

$$R[t_1, \ldots, t_n] := (R[t_1, \ldots, t_{n-1}])[t_n].$$

If $f = a_0 + a_1 t + \cdots + a_n t^n$ is a nonzero polynomial with $a_n \neq 0$, we define the *degree* of f by $\deg f := n$ and the *leading coefficient* of f by $\operatorname{lead} f = a_n$. If $\operatorname{lead} f = 1$, we say f is *monic*. If $f = 0$ or $\deg f = 0$, we call f a *constant polynomial*. We let $\operatorname{lead} 0 = 0$ (and one can define $\deg 0 = -\infty$ if we view $-\infty + n = -\infty < n$ for any integer n).

Properties 31.1. Let R be a ring, f, g nonzero polynomials.

1. If R is commutative, then so is $R[t]$.
2. $\deg(f + g) \leq \max\{\deg f, \deg g\}$. (Strict inequality is possible, e.g., if $g = -f$.)
3. $\deg(fg) \leq \deg f + \deg g$ with equality if and only if

$$(*) \qquad \operatorname{lead}(fg) = \operatorname{lead}(f)\operatorname{lead}(g)$$

   is nonzero.
4. If R is a domain, then (*) holds, so $R[t]$ is a domain.

Definition 31.2. Let S be a commutative ring and $a_1, \ldots, a_n$ elements in S. If R a subring of S, then the map

$$e_{a_1,\ldots,a_n} : R[t_1, \ldots, t_n] \to S \text{ given by } f \mapsto f(a_1 \ldots, a_n)$$

is called *evaluation* at $a_1, \ldots, a_n$. We denote the image of $e_{a_1,\ldots,a_n}$ by $R[a_1, \ldots, a_n]$, as elements in this image are sums of the form $r a_1^{i_1} \cdots a_n^{i_n}$ with p$r \in R$ and $i_1, \ldots, i_n$ non-negative integers. The evaluation map $e_{a_1,\ldots,a_n}$ is a ring homomorphism and if $R = S$, it is surjective. If R is not commutative, then this map may not be a homomorphism. (Why?) For this reason, we shall only study polynomial rings over commutative rings. Of course if the subring of S generated by R and $a_1, \ldots, a_n$ is commutative, the evaluation map $e_{a_1,\ldots,a_n}$ is still a ring homomorphism as we may replace S by a commutative subring.

If $\varphi : R \to S$ is a ring homomorphism (respectively, ring monomorphism, ring epimorphism, ring isomorphism) of commutative rings, then φ induces a map

$$\tilde{\varphi} : R[t_1, \ldots, t_n] \to S[t_1, \ldots, t_n]$$
$$\text{given by } \sum_{i_1} \cdots \sum_{i_n} a_{i_1,\ldots,i_n} t_1^{i_1} \cdots t_n^{i_n} \mapsto \sum_{i_1} \cdots \sum_{i_n} \varphi(a_{i_1,\ldots,i_n}) t_1^{i_1} \cdots t_n^{i_n}.$$

and $\tilde{\varphi}$ is a ring homomorphism (respectively, ring monomorphism, ring epimorphism, ring isomorphism). For convenience, we write the iterated sum by $\sum$ or $\sum_{i_1,\ldots,i_n}$. Moreover, if $a_1, \ldots, a_n$ are elements in R,

we have a commutative diagram:

$$\begin{array}{ccc} R[t_1,\dots,t_n] & \xrightarrow{\tilde{\varphi}} & S[t_1,\dots,t_n] \\ {\scriptstyle e_{a_1,\dots,a_n}}\downarrow & & \downarrow{\scriptstyle e_{\varphi(a_1),\dots,\varphi(a_n)}} \\ R & \xrightarrow[\varphi]{} & S. \end{array}$$

The above applies to the case that $S = R/\mathfrak{A}$ with $\mathfrak{A}$ an ideal in R and φ the canonical epimorphism $^- : R \to R/\mathfrak{A}$. Assume that $\mathfrak{A} < R$ (otherwise we would get no further information). It induces a ring epimorphism, $\widetilde{\ } : R[t_1,\dots,t_n] \to (R/\mathfrak{A})[t_1,\dots,t_n]$ given by $\sum_{i_1,\dots,i_n} a_{i_1,\dots,i_n} t_1^{i_1}\cdots t_n^{i_n} \mapsto \sum_{i_1,\dots,i_n} \overline{a_{i_1,\dots,i_n}} t_1^{i_1}\cdots t_n^{i_n}$. In this case, for simplicity of notation, we write $^-$ for $\widetilde{\ }$. Hence if $(a_1,\dots,a_n) \in R^n$, we have a commutative diagram:

$$\begin{array}{ccc} R[t_1,\dots,t_n] & \xrightarrow{-} & \overline{R}[t_1,\dots,t_n] \\ {\scriptstyle e_{a_1,\dots,a_n}}\downarrow & & \downarrow{\scriptstyle e_{\overline{a_1},\dots,\overline{a_n}}} \\ R & \xrightarrow[-]{} & \overline{R}, \end{array}$$

i.e.,

$$\overline{f}(\overline{x_1},\dots,\overline{x_n}) = \overline{f(x_1,\dots,x_n)}.$$

In particular, if no $(a_1,\dots,a_n) \in \overline{R}^n$ satisfies $\overline{f}(a_1,\dots,a_n) = 0$, then no $(x_1,\dots,x_n) \in R^n$ satisfies $f(x_1,\dots,x_n) = 0$. Viewed geometrically, this says $f = 0$ in R^n has no solution if $\overline{f} = 0$ in $\overline{R}^n$ has no solution. This is a very useful idea, that is often used, for instance in number theory, to try to show a given Diophantine equation has no solution.

We return to the one variable case. For further remarks in this chapter about polynomials in many variables, see Addendum §33.

Lemma 31.3. *Let R be a domain, then $R[t]^\times = R^\times$.*

PROOF. Clearly, units in R remain units in $R[t]$. Conversely, let f, g be polynomials in $R[t]$ satisfying $fg = 1$. Then $\deg(fg) = \deg f + \deg g = 0$, so $\deg f = \deg g = 0$ and f, g lie in R. □

Note that $\mathbb{Z}/4\mathbb{Z}$ is not a domain and $(\overline{1} + \overline{2}t)^2 = \overline{1}$ in $\mathbb{Z}/4\mathbb{Z}[t]$.

By Exercise 27.22(31.4), whose solution mimics how you learned how to divide polynomials, we have the following:

Theorem 31.4. (General Division Algorithm) *Let f, g be polynomials with coefficients in a commutative ring R. Suppose the leading coefficient of g is a unit. Then there are polynomials q and r with coefficients*

in R such that $f = gq + r$ with either $r = 0$ or the degree of r is less than the degree of g.

As usual, if R is a commutative ring, we say that α in R is a *root* of a polynomial f in $R[t]$ if $f(\alpha) = e_\alpha(f) = 0$. The General Division Algorithm has many consequences, most that you have encountered in middle school algebra that we now state and prove.

Corollary 31.5. *Let F be a field. Then $F[t]$ is a euclidean domain. In particular, $F[t]$ is a PID, hence also a UFD.*

Corollary 31.6. (Remainder Theorem) *Let R be a commutative ring, α an element in R, and f a nonzero polynomial in $R[t]$. Then there exists a polynomial q in $R[t]$ satisfying $f = (t-\alpha)q + f(\alpha)$. Moreover,*

$$t - \alpha \mid f \text{ in } R[t] \text{ if and only if } f(\alpha) = 0, \text{ i.e., } \alpha \text{ is a root of } f.$$

PROOF. Apply the General Division Algorithm with $g = t - \alpha$ to get $f = (t-\alpha)q + r$ in $R[t]$ for some $q, r \in R[t]$ with $r = 0$ or $\deg r < \deg(t-\alpha) = 1$. It follows that $r \in R$, hence $f(\alpha) = e_\alpha(f) = e_\alpha\big((t-\alpha)q\big) + e_\alpha(r) = r$. The result now follows easily. □

Corollary 31.7. *Let R be a domain, f a nonzero polynomial in $R[t]$, and $x_1, \ldots, x_n$ distinct roots of f in R. Then $\prod_{i=1}^n (t - x_i) \mid f$ in $R[t]$. In particular, $\deg f \geq n$.*

PROOF. If $n = 1$, the result follows by the Remainder Theorem, so assume that $n > 1$. By iterating the Remainder Theorem, we can write $f = (t-x_1)^r h$ in $R[t]$ for some r in $\mathbb{Z}^+$ and h in $R[t]$ with $h(x_1)$ nonzero. As $R[t]$ is a domain, we have $\deg h \leq \deg f - r$. For each $i = 2, \ldots, n$, applying the evaluation homomorphism e_{x_i} shows that $0 = f(x_i) = (x_i - x_1)^r h(x_i)$ in the domain $R[t]$. Hence $h(x_i) = 0$ for $i = 2, \ldots, n$. By induction $\prod_{i=2}^n (t - x_i) \mid h$ in $R[t]$, hence $(t-x_1)^r \prod_{i=2}^n (t-x_i) \mid f$ in $R[t]$. As $\deg f \geq \deg h \geq n - 1$, the result follows. □

Note the proof shows that the degree of f above is greater than the number of roots of f in R "counted with multiplicity."

Corollary 31.8. (Lagrange) *Let R be a domain and f a nonzero polynomial of degree n in $R[t]$. Then f has at most n roots in R.*

Corollary 31.9. *Let R be a domain and f and g nonzero polynomials in $R[t]$. If $f(x_i) = g(x_i)$ for n distinct elements $x_1, \ldots, x_n$ in R with $n > \max(\deg f, \deg g)$, then $f = g$. In particular, if R is infinite and $f(x) = g(x)$ for all $x \in R$, then $f = g$.*

Remarks 31.10. 1. The polynomial $t^2 - 1$ in $(\mathbb{Z}/8\mathbb{Z})[t]$ has four roots: $\pm\overline{1}, \pm\overline{3}$, so the assumption that R be a domain in the Corollary 31.7 is essential.

2. If R is not a commutative ring, as mentioned evaluation is not necessarily a ring homomorphism. This occurs because we assume that the variable t commutes with all elements of R, but the element x at which we wish to evaluate will, in general, not commute with all elements in R. If x does commute with all elements of R, then evaluation at that element will be a ring homomorphism. The theory of polynomials over non-commutate rings is, therefore, quite different. For example the polynomial $t^2 + 1$ in $\mathcal{H}[t[$, with $\mathcal{H}$ the Hamiltonian quaternions, has infinitely many roots.
3. If R is an infinite domain, Corollary 31.9 says that polynomials in $R[t]$ and polynomial functions are essentially the same, where $p : R \to R$ is a *polynomial function* if there exists a polynomial f in $R[t]$ such that $p(x) = f(x)$ for all $x \in R$. We write p_f for this p. So the corollary says that if f and g are polynomials in $R[t]$, with R an infinite domain, then $f = g$ if and only if $p_f = p_g$. (Clearly, if $f = g$ then $p_f = p_g$, and the zero polynomial is the only polynomial one with infinitely many roots.)
4. Let p be a prime in $\mathbb{Z}^+$. Then the polynomials t^p and t in $(\mathbb{Z}/p\mathbb{Z})[t]$ are clearly distinct, but by Fermat's Little Theorem, $p_{t^p} = p_t$ in the notation of the previous remark, so the condition that R be infinite in the previous remark is crucial.

For the next corollary, we set up important notation. Let F be a field and $\mathfrak{A} < F[t]$ be an ideal. We have the canonical ring epimorphism

$$^- : F[t] \to F[t]/\mathfrak{A} \text{ given by } g \mapsto \overline{g} = g + \mathfrak{A}.$$

As $\mathfrak{A} < F[t]$, the ring $F[t]/\mathfrak{A}$ is not the trivial ring. Since a field is simple, the canonical epimorphism induces a ring monomorphism

$$^-|_F : F \to F[t]/\mathfrak{A} \text{ given by } a \mapsto \overline{a}.$$

We shall always view this as an inclusion, i.e., for all $a \in F$, we shall identify

$$a \text{ and } \overline{a}.$$

As $\overline{t} = t + \mathfrak{A}$, this identification means that the canonical epimorphism maps $g = \sum a_i t^i \mapsto \overline{g} = \sum \overline{a_i}\overline{t}^i = \sum a_i \overline{t}^i$, i.e., the canonical epimorphism under this identification is none other than the evaluation map at $\overline{t}$, i.e.,

$$^- : F[t] \to F[t]/\mathfrak{A} \text{ is the map } g \mapsto g(\overline{t}.) \tag{31.11}$$

Corollary 31.12. (Kronecker's Theorem) *Let F be a field and f an irreducible element in $F[t]$. Set $K = F[t]/(f)$. Then K is a field.*

Viewing $F \subset K$, i.e., as a subfield of K as above, we have $\bar{t}$ is a root of f in K.

PROOF. As f is irreducible in the UFD $F[t]$, it is a prime element. As $F[t]$ is a PID, the nonzero prime ideal (f) is maximal, so K is a field. As $0 = \overline{f} = f(\bar{t})$, by the above, the result follows. □

Corollary 31.13. *Let F be a field and f a non-constant polynomial in $F[t]$. Then there exists a field K with $F \subset K$ a subfield such that f has a root in K.*

PROOF. We can write $f = f_1 g$ with f_1 and g polynomials in $F[t]$ and f_1 irreducible. By Kronecker's Theorem, f_1 hence f has a root in $F[t]/(f_1)$, a field containing F. □

Of course, to make Kronecker's Theorem and its corollary really useful, we would want to prove results by induction, the natural choice being the degree of a polynomial. That we can do this follows from the following stronger form of Kronecker's theorem, which we leave as an exercise, but will be essential when we study field theory:

Proposition 31.14. *Let F be a field and f a non-constant polynomial of degree n in $F[t]$. Then $F[t]/(f)$ is a vector space over F of dimension n. In particular, there exists a field K containing F such that f has a root in K and the dimension of K as a vector space over F is at most n.*

We next prove a fundamental fact about finite fields. To do so, we use the First Sylow Theorem. This result is also important in general field theory.

Theorem 31.15. *Let F be a field and G a finite (multiplicative) subgroup of $F^\times$. Then G is cyclic. In particular, if F is a finite field, the group $F^\times$ is cyclic.*

PROOF. Let p be a prime dividing $|G|$ and P a Sylow p-subgroup of G (so the only one as G is abelian). Choose x_P in P such that the order of the cyclic subgroup $\langle x_P \rangle$ generated by x_P in P is maximal. Let N be the order of x_P. Then N is a power of p and we must have $y^N = 1$ for all $y \in P$. (Why?) Therefore, the polynomial $t^N - 1$ has $|P|$ distinct roots in F, so $N = |P|$ by Corollary 31.8. It follows that P is cyclic. Let $x = \prod_{\substack{p \mid |G| \\ p \text{ a prime}}} x_P \in G$. As G is abelian, it follows from Exercise 13.6(5) that $G = \langle x \rangle$ is cyclic. □

We end this section, with a short discussion of irreducible polynomials over a field F. Clearly, every *linear* polynomial, i.e., polynomial of degree one in $F[t]$ is irreducible.

Definition 31.16. The following conditions on a field F are equivalent:

(*i*) Every non-constant polynomial in $F[t]$ has a root in F.
(*ii*) The only irreducible polynomials in $F[t]$ are *linear*.
(*iii*) Every non-constant polynomial in $F[t]$ factors into a product of linear polynomials.

A field F satisfying these equivalent conditions is called *algebraically closed*.

When we study field theory, we shall prove (cf 54.12 below):

Theorem 31.17. (Fundamental Theorem of Algebra) *The field of complex numbers is algebraically closed.*

We shall, however, assume here that it has already been established.

Remark 31.18. The Fundamental Theorem of Algebra allows us to compute all irreducible polynomials over the reals. They are

(*i*) Linear polynomials in $\mathbb{R}[t]$.
(*ii*) *Quadratic polynomomials* (i.e., polynomials of degree two) of the form $at^2 + bt + c$ in $\mathbb{R}[t]$ satisfying, $a \neq 0$ and $b^2 - 4ac < 0$.

In particular, any non-constant polynomial in $\mathbb{R}[t]$ factors into a product of linear and irreducible quadratic polynomials. For example, we have the factorization of the polynomial $t^4 + 1 = (t^2 + \sqrt{2}t + 1)(t^2 - \sqrt{2}t + 1)$. Over any field F, the polynomial ring $F[t]$ contains infinitely many non-associative irreducible polynomials. (Can you prove this?) Finding all irreducible polynomials in $F[t]$, with F or $F[\sqrt{-1}]$ (which is in fact always a field) not algebraically closed is usually very hard, if not impossible, e.g., if $F = \mathbb{Q}$.

One of our goals will be to show that given any field F, there exists an algebraically closed field K containing F and if $F \neq K$, then no field properly between F and K is algebraically closed. Such a field will be called an *algebraic closure* of F. The proof will need Zorn's Lemma. We shall prove this in Section §48 below. In particular, $\mathbb{C}$ is not an algebraic closure of $\mathbb{Q}$.

Exercises 31.19. 1. Let R be a commutative ring. Show that a polynomial $f = a_0 + a_1 t + \cdots + a_n t^n$ in $R[t]$ is a unit in $R[t]$ if and only if a_0 is a unit in R and a_i is nilpotent for every $i > 0$.

2. Let R be a nontrivial commutative ring and f a zero divisor in $R[t]$. Show that there exists a nonzero element b in R so that $bf = 0$.
3. Let R be a nontrivial commutative ring. If $f = a_0 + a_1 t + \cdots + a_n t^n$ is a polynomial in $R[t]$ define the *formal derivative* f' of f to be $f = a_1 + 2a_2 t + \cdots + n a_n t^{n-1}$.
   (*i*) Show the usual rules of differentiation hold.
   (*ii*) Suppose R is a field of characteristic zero. Show that a polynomial $f \in R[t]$ is divisible by the square of a non-constant polynomial in $R[t]$ if and only if f and f' are not relatively prime.
4. Let R be a ring and G a group (or monoid). Define

$$R[G] := \{\sum_G a_g g \mid g \in G \text{ and } a_g \in R \text{ almost all } a_g = 0\}$$

   (where *almost all zero* means that only finitely many are nonzero) with $+$ and $\cdot$ defined by

$$\sum_G a_g g + \sum_G b_g g = \sum_G (a_g + b_g) g \text{ and}$$
$$\sum_G a_g g \cdot \sum_G b_g g = \sum_G c_g g \text{ where } c_g = \sum_{g=hl} a_h b_l$$

   for all a_g, b_g in G. Show this is a ring. It is called the *group* (respectively, *monoid*) ring of R by G. If $\mathbb{N} := \mathbb{Z}^+ \cup \{0\}$, show that $R[\mathbb{N}^n]$ is isomorphic to $R[t_1, \cdots, t_n]$. Describe $R[\mathbb{Z}^n]$.
5. Let F be a subfield of the complex numbers $\mathbb{C}$. Let $f \in F[t]$ be an irreducible polynomial. Show that f has no *multiple root* in $\mathbb{C}$, i.e., a root α of f satisfying $(t - \alpha)^n \mid f$ in $F[t]$ with $n > 1$.
6. Prove Proposition 31.14.
7. Prove that the conditions in Definition 31.16 are equivalent.
8. Show that the irreducible polynomials in $\mathbb{R}[t]$ are those stated in Remark 31.18.
9. Show that over any field F, there exist infinitely many monic irreducible polynomials in $F[t]$. Also show that if F is algebraically closed, then F must have infinitely many elements.
10. Let $F = \mathbb{Z}/p\mathbb{Z}$ with p a prime. Show that $t^4 + 1 \in F[t]$ is reducible.

32. Polynomial Rings over a UFD

In this section, we prove a theorem of Gauss that a polynomial ring over a UFD R is itself a UFD. The idea is to take a polynomial over $R[t]$, with R a UFD, and view it over the UFD $K[t]$ where K is

the quotient field of R. To make this useful, we must derive a way of pulling information back to $R[t]$.

We have defined the gcd of two nonzero elements in a domain R. We can extend this in the obvious way to a finite number of elements, viz., if $a_1, \ldots, a_n$ are elements in R not all zero, then d in R is a *greatest common divisor* or *gcd* of $a_1, \ldots, a_n$ if it satisfies both of the following:

(i) $d \mid a_i$ for $i = 1, \ldots, n$.

(ii) If $e \mid a_i$ for some e in R, $i = 1, \ldots, n$, then $e \mid d$.

[We let the gcd of nonzero a in R and 0 be a.]

We have the following lemma that we leave as an exercise.

Lemma 32.1. *Let R be a UFD and $a_1, \ldots, a_n$ elements in R, not all zero. Then a gcd of $a_1, \ldots, a_n$ exists and is unique up to units.*

Let R be a UFD and $f = a_n t^n + \cdots + a_0$ a nonzero polynomial in $R[t]$. A gcd of $a_0, \ldots, a_n$ is called a *content* of f. It is unique up to units. We call f *primitive* if 1 is a content of f. [So a primitive constant polynomial is just a unit in R.] For notational convenience, we often use the notation $C(f)$ for a choice of a content of f.

Remarks 32.2. Let R be a UFD, b a nonzero element of R and f a nonzero polynomial in $R[t]$.

1. If x is an irreducible element in R, then it remains irreducible when considered as an element of $R[t]$. (Look at degrees if it would factor.)
2. $C(bf) \approx bC(f)$, where as before $\approx$ means is an associate of.
3. There exists a primitive polynomial f_1 in $R[t]$ satisfying $f = C(f)f_1$. (As $R[t]$ is a domain, $\deg f = \deg f_1$.) This follows easily as we can factor out a content.
4. If $\deg f > 0$ and f is irreducible in $R[t]$, then f is primitive. Indeed, we can write $f = C(f)f_1$ for some primitive polynomial f_1 in $R[t]$. If $C(f) \not\approx 1$, then it factors into irreducibles in R hence in $R[t]$. It follows that f cannot be irreducible as $\deg f_1 > 0$, so f is not a unit.
5. $t^2 - 1$ is a primitive polynomial, but not irreducible, so the converse to the previous remark is false.

If R is a UFD, to show $R[t]$ is a UFD, we must show that any nonzero nonunit polynomial in $R[t]$ factors into a product of irreducible polynomials in $R[t]$ and that this factorization is essentially unique. The existence is more elementary and will follow from the following lemma and its corollary.

Lemma 32.3. *Let R be a UFD, K the quotient field of R, and f a nonzero polynomial in $K[t]$. Then there exist a primitive polynomial*

$f_1 \in R[t]$ (of the same degree as f) and an element α in K satisfying $f = \alpha f_1$. Moreover, f_1 and α are unique up to units in R, i.e, up to elements in $R^\times = R[t]^\times$.

PROOF. **Existence**: Write $f = \sum_{i=0}^{n} \frac{a_i}{b_i} t^i$ with a_i, b_i in R and $b_i \neq 0$, $i = 0, \ldots, n$. We clear denominators. Let $b = b_0 \cdots b_n \neq 0$ in the domain R. Then $bf \in R[t]$. Let $c = C(bf)$. Then there exists a primitive polynomial f_1 in $R[t]$ such that $bf = cf_1$, so $f = \frac{c}{b} f_1$ in $K[t]$ as needed.

Uniqueness Suppose that $\frac{c}{b} f_1 = f = \frac{d}{e} f_2$ in $K[t]$, with nonzero $c, b, d, e \in R$ and f_1, f_2 primitive in $R[t]$. Then in $R[t]$, we have $cef_1 = bdf_2$. Therefore, $ce \approx C(cef_1) \approx C(bdf_2) \approx bd$ in R. Consequently, there exists a unit u in $R^\times$ satisfying $ce = ubd$ in R, hence $\frac{c}{b} = u\frac{d}{e}$ in K. Thus $\frac{d}{e} f_2 = \frac{c}{b} f_1 = u\frac{d}{e} f_1$ in the domain $K[t]$. It follows that $f_2 = uf_1$. □

Corollary 32.4. *Let R be a UFD, f, g primitive polynomials in $R[t]$, h a polynomial in $R[t]$, and K the quotient field of R.*

(1) If $g = sf$ in $K[t]$ for some s in K, then s is a unit in R.
(2) If $g = ah$ in $R[t]$ for some a in R, then a is a unit in R and h is primitive in $R[t]$.

In particular, if g is non-constant primitive polynomial in $R[t]$ but not irreducible, then $g = q_1 q_2$ for some q_1, q_2 in $R[t]$ satisfying $0 < \deg q_i < \deg g$, for $i = 1, 2$.

PROOF. (1): We have $1 \cdot g = s \cdot f$ in $K[t]$ with both f and g primitive, so $s = u \cdot 1$ for some unit u in R by Lemma 32.3, hence s is a unit in R.
(2): Write $h = C(h)h_1$ with h_1 primitive in $R[t]$. Then $1 \cdot g = a \cdot C(h)h_1$, so $aC(h) \approx 1$, hence a is a unit in R and h is primitive.
[Note if x is a nonzero element in R then x/x is a unit but x is not necessarily a unit.] □

The key to showing the uniqueness statement for factorization in $R[t]$ when R is a UFD is the following:

Lemma 32.5. (Gauss' Lemma) *Let R be a UFD and f, g non-constant polynomials. Then $C(fg) \approx C(f)C(g)$. In particular, the product of primitive polynomials in $R[t]$ is primitive.*

PROOF. Write $f = C(f)f_1$ and $g = C(g)g_1$ with f_1 and g_1 primitive polynomials in $R[t]$. Then

$$C(fg) \approx C(C(f)f_1 C(g)g_1) \approx C(f)C(g)C(f_1 g_1)$$

and

$$C(f)C(g) \approx C(f)C(f_1)C(g)C(g_1) \approx C(f)C(g)C(f_1)C(g_1),$$

so it suffices to show the last statement, i.e., we may assume that f and g are primitive in $R[t]$ and must show fg is primitive in $R[t]$. Suppose this is false. Then there exists an irreducible element p such that $p \mid C(fg)$ in R. As R is a UFD, p is a prime element, hence $\overline{R} = R/(p)$ is a domain. Let $^- : R[t] \to \big(R/(p)\big)[t]$ be the ring epimorphism given by $\sum a_i t^i \mapsto \sum \overline{a_i} t^i$. By assumption, $\overline{0} = \overline{fg} = \overline{f}\,\overline{g}$ in the domain $\overline{R}[t]$. As f and g are primitive in $R[t]$, the prime p does not divide some coefficient of f and some coefficient of g, i.e., $\overline{f} \neq 0$ and $\overline{g} \neq 0$ in $\overline{R}[t]$, contradicting the fact that $\overline{R}[t]$ is a domain. □

Corollary 32.6. *Suppose that R is a UFD with quotient field K. Let f, g be non-constant primitive polynomials in $R[t]$ and h a non-constant polynomial in $K[t]$. If $f = gh$ in $K[t]$, then h lies in $R[t]$ and is primitive.*

Proof. By Lemma 32.3, we can write $h = \alpha h_1$ with $\alpha \in K$ and h_1 a primitive polynomial in $R[t]$, so $f = \alpha g h_1$. By Gauss' Lemma, gh_1 is primitive, so α lies in $R^\times$ by Corollary 32.4. It follows that h lies in $R[t]$ and is primitive. □

A similar proof yields:

Lemma 32.7. *Let R be a UFD with quotient field K and f a non-constant irreducible polynomial in $R[t]$. Then f remains irreducible in $K[t]$.*

Proof. As f is a non-constant irreducible polynomial in $R[t]$, it is primitive. Suppose that f factors as $f = g_1 g_2$ in $K[t]$, with $g_i \in K[t]$ and $0 < \deg g_i < \deg f$, for $i = 1, 2$. By Lemma 32.3, we can write $g_i = \alpha_i h_i$ with $\alpha_i \in K$ and h_i in $R[t]$ primitive (and $\deg g_i = \deg h_i$), for $i = 1, 2$. Therefore, we have $f = (\alpha_1\alpha_2) h_1 h_2$. By Gauss' Lemma, $h_1 h_2$ is primitive, so by Corollary 32.4, we have $\alpha = \alpha_1\alpha_2$ is a unit in R and $f = \alpha h_1 h_2$ in $R[t]$. It follows that f is reducible in $R[t]$, a contradiction. □

We can now prove our theorem.

Theorem 32.8. *Let R be a UFD, then $R[t]$ is a UFD.*

Proof. **Existence**. Every nonzero element in R is a unit or factors into irreducibles that remain irreducible in $R[t]$, so it suffices to factor a non-constant polynomial in $R[t]$. Let f be such a polynomial. As $f = C(f) f_1$ with f_1 a primitive polynomial in $R[t]$, we may also assume

that f is primitive. If f is not irreducible, then $f = g_1 g_2$ for some polynomials $g_i \in R[t]$ satisfying $0 < \deg g_i < \deg f$ for $i = 1, 2$ by Corollary 32.4. By induction on $\deg f$, we conclude that g_1 and g_2 hence f factor into irreducible polynomials in $R[t]$.

Uniqueness. We know that any nonzero element in R has a unique factorization (up to associates and order), so the content of a polynomial factors uniquely up to units. It follows by Lemma 32.3 that it suffices to show any non-constant primitive polynomial f has a unique factorization (up to associates and order). Suppose that

$$f_1 \cdots f_r = f = g_1 \cdots g_s$$

with all the f_i and g_j irreducible in $R[t]$. By Lemma 32.7, all the f_i and g_j remain irreducible in $K[t]$, where K is the quotient field of R. As $K[t]$ is a UFD, $r = s$, and after relabeling, we may assume that $f_i = c_i g_i$ in $K[t]$ for some $c_i \in K^\times$ for each i. But the f_i and g_j are primitive in $R[t]$ as irreducible, so $c_i \in R^\times$ for all i by Corollary 32.4. We conclude that $f_i \approx g_i$ in $R[t]$ for all i, and the proof is complete. □

Corollary 32.9. *If R is a UFD, then so is $R[t_1, \ldots, t_n]$.*

Examples 32.10. Let R be a UFD and F a field.

1. $\mathbb{Z}[t]$ and $F[t_1, t_2]$ are UFDs but not PIDs. In fact, $R[t]$ is a PID if and only if R is a field.
2. A polynomial ring in any number of variables (even infinitely many — definition of such a polynomial ring?) over R is a UFD.
3. Let $\varphi : R \to S$ be a ring epimorphism of domains. Then S is not necessarily a UFD (even if a domain), e.g. $e_{\sqrt{-5}} : \mathbb{Z}[t] \to \mathbb{Z}[\sqrt{-5}]$ is such an example.

We end this discussion with a few remarks about irreducible elements in $R[t]$, with R a domain.

Remarks 32.11. 1. (Eisenstein's Criterion) Let R be a UFD and K its quotient field. Let $f = \sum_{i=0}^{n} a_i t^i$ be a non-constant polynomial in $R[t]$ of degree n. Suppose that there exists an irreducible element p in R (hence a prime) satisfying:

(i) $p \nmid a_n$.

(ii) $p \mid a_i$, for $i = 0, \ldots, n-1$.

(iii) $p^2 \nmid a_0$.

Then f is irreducible in $K[t]$. In particular, if f is primitive in $R[t]$, then it is irreducible in $R[t]$.

We leave this as an exercise.

2. If R is a domain, f a non-constant polynomial in $R[t]$, let $y = t + a$, for some a in R, and set $g(y) = f(y - a)$. Then f is irreducible in $R[t]$ if and only if g is irreducible in $R[y]$ (as $R[t] \to R[y]$ defined by $\sum a_i t^i \to \sum a_i y^i$ is a ring isomomorphism).
3. Let $p \in \mathbb{Z}^+$ be a prime. Then the polynomial $t^{p-1} + t^{p-2} + \cdots + t + 1$ is irreducible in $\mathbb{Z}[t]$ and $t^p - 1 = (t-1)(t^{p-1} + t^{p-2} + \cdots + t + 1)$ is a factorization into irreducibles in $\mathbb{Z}[t]$ (and $\mathbb{Q}[t]$):
   Let $y = t - 1$, then
$$f = t^{p-1} + t^{p-2} + \cdots + t + 1 = \frac{t^p - 1}{t - 1}$$
$$= \frac{(y+1)^p - 1}{y} = y^{p-1} + py^{p-2} + \binom{p}{2} y^{p-3} + \cdots + p.$$
   As $p \mid \binom{p}{i}$ in $\mathbb{Z}$ for $p = 1, \ldots, p-1$, we conclude that f is irreducible in $\mathbb{Z}[t]$ by Eisenstein's Criterion and the previous remark.
4. Let R be a commmutative ring, $\mathfrak{A} < R$ an ideal, $^- : R \to R/\mathfrak{A}$ the canonical epimorphism. If $f = g_1 \cdot g_2$ in $R[t]$, then $\overline{f} = \overline{g_1} \cdot \overline{g_2}$ in $\overline{R}[t]$. In particular, if R is a domain, $\mathfrak{A}$ a prime ideal, f monic, then $\overline{f}$ irreducible in $\overline{R}[t]$ implies f is irreducible in $R[t]$.

Exercises 32.12. 1. Prove Lemma 32.1.

2. Let R be a domain that is not a field. Show that $R[t]$ is not a PID.
3. Let R be a UFD, K its quotient field. Let f and g be non-constant polynomials in $R[t]$. Write $f = C(f)f_1$ and $g = C(g)g_1$ with f_1 and g_1 primitive polynomials in $R[t]$. Show
   (*i*) If $f|g$ in $K[t]$ then $f_1|g_1$ in $R[t]$. In particular, if f and g are primitive, then $f|g$ in $K[t]$ if and only if $f|g$ in $R[t]$.
   (*ii*) Suppose that f and g are primitive. Then f and g have a common factor over $K[t]$ if and only if they have a common factor over $R[t]$.
4. Let R be a commutative ring, and $\{t_1, \mid i \in I\}$ indeterminants. Define the polynomial ring $R[t_i]_{i \in I}$ in the variables $\{t_1, \mid i \in I\}$. Show that it is a UFD if R is.
5. Let $f = \sum_{i=0}^n a_i t^i$ be a polynomial in $\mathbb{Z}[t]$ with $a_n \neq 0$. Let $r = a/b$ with b nonzero and a and b relatively prime integers. If r is a root of f, show that $b \mid a_n$ and if a is also nonzero, then $a \mid a_0$. In particular, if f is monic, any rational roof of f, if any, is an integer.
6. Prove Eisenstein's Criterion (Remark 32.11(1)).
7. Let $y = t + a$. Show that the map $R[t] \to R[y]$ given by $\sum a_i t^i \mapsto \sum a_i y^i$ is a ring isomorphism.

8. Prove the identity of binomial coefficients

$$\binom{n+1}{n-j} = \sum_{i=j}^{n} \binom{i}{i-j}$$

and use it to prove that the polynomial

$$t^{p^{r-1}(p-1)} + t^{p^{r-1}(p-2)} + \cdots + t^{p^{r-1}} + 1$$

is an irreducible polynomial in $\mathbb{Z}[t]$ (hence $\mathbb{Q}[t]$) for every (positive) prime p in $\mathbb{Z}$.

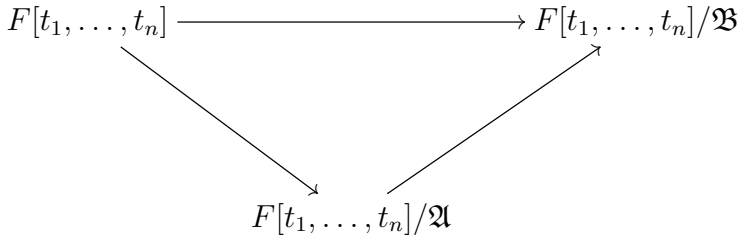
33. Addendum: Polynomial Rings over a Field

In this addendum, we make some motivational remarks that we shall come back to when studying modules. If R is a commutative ring, we have defined the ring $R[t_1, \ldots, t_n]$. We view R as a subring of $R[t_1, \ldots, t_n]$ of constant polynomials. We look at the case, when R is a field.

Let F be a field and S be a nontrivial commutative ring. We have seen that any ring homomorphism $\varphi : F \to S$ is monic. Let $\mathfrak{A} < F[t_1, \ldots, t_n]$ be an ideal. We apply this to $S = F[t_1, \ldots, t_m]/\mathfrak{A}$. Let $^- : F[t_1, \ldots, t_n] \to F[t_1, \ldots, t_n]/\mathfrak{A}$ be the canonical epimorphism. Then the composition

$$(*) \qquad F \subset F[t_1, \ldots, t_n] \xrightarrow{-} F[t_1, \ldots, t_n]/\mathfrak{A}$$

is also monic. As before, we view this as an embedding, i.e., identify a in F with $\overline{a}$ in $F[t_1, \ldots, t_n]/\mathfrak{A}$. Under this identification the map $^- : F[t_1, \ldots, t_n] \to F[t_1, \ldots, t_n]/\mathfrak{A}$ becomes $f \mapsto \overline{f} = f(\overline{t_1}, \ldots, \overline{t_n})$, so it is the evaluation map $e_{(\overline{t_1}, \ldots, \overline{t_n})} : F[t_1, \ldots, t_n] \to F[t_1, \ldots, t_n]/\mathfrak{A}$. Observe that if $\mathfrak{A} \subset \mathfrak{B} < F[t_1, \ldots, t_n]$ are ideals, then we have a commutative diagram



with all the maps the obvious ring epimorphisms. The most interesting case is when $\mathfrak{B}$ is a maximal ideal, hence $F[t_1, \ldots, t_n]/\mathfrak{B}$ is a field. We look at a special case of this.

Let $\underline{x} = (x_1, \ldots, x_n)$. For each $\underline{x} = (x_1, \ldots, x_n)$ in F^n set

$$\mathfrak{m}_{\underline{x}} = (t_1 - x_1, \ldots, t_n - x_n),$$

an ideal in $F[t_1, \ldots, t_n]$. Let $\mathfrak{A} = \mathfrak{m}_{\underline{x}}$ in (*). For each $i = 1, \ldots, n$, we have $\overline{t_i} = x_i$ lies in F, so we can view $F[t_1, \ldots, t_n]/\mathfrak{m}_{\underline{x}} \subset F$. By (*), we conclude that $F[t_1, \ldots, t_n]/\mathfrak{m}_{\underline{x}} = F$. In particular, $\mathfrak{m}_{\underline{x}}$ is a maximal ideal in $F[t_1, \ldots, t_n]$. We conclude that the canonical epimorphism $^- : F[t_1, \ldots, t_n] \to F[t_1, \ldots, t_n]/\mathfrak{m}_{\underline{x}}$ takes $f \in F[t_1, \ldots, t_n]$ to $\overline{f} = f(\underline{\overline{t}}) = f(\underline{x}) = e_{\underline{x}}(f)$ in F. So we have

$$f(\underline{x}) = 0 \text{ in } F \text{ if and only if } \overline{f} = 0 \text{ in } F \text{ if and only if } f \in \mathfrak{m}_{\underline{x}}.$$

The question of interest is to solve the following:

Problem 33.1. *Let $f_1, \ldots, f_r$ be polynomials in $F[t_1, \ldots, t_n]$. Does there exist an $\underline{x}$ in F^n such that $f_i(\underline{x}) = 0$ for $i = 1, \ldots r$, i.e., a common point lying on all the hypersurfaces $f_1 = 0, \ldots, f_r = 0$ in F^n?*

Let $\mathfrak{A} = (f_1, \ldots, f_r)$, the ideal generated by the f_i in $F[t_1, \ldots, t_n]$ and $\underline{x} \in F^n$. Then we have $f_i(\underline{x}) = 0$ for $i = 1, \ldots, r$ if and only if $f(\underline{x}) = 0$ for all $f \in \mathfrak{A}$. We set up a useful notation. If $\mathfrak{B}$ is an ideal in $F[t_1, \ldots, t_n]$, let

$$Z_F(\mathfrak{B}) := \{\underline{x} \in F^n \mid f(\underline{x}) = 0 \text{ for all } f \in \mathfrak{B}\}.$$

We call $Z_F(\mathfrak{B})$ the *affine variety* defined by $\mathfrak{B}$ in F^n. If $\mathfrak{B} = (g_1, \ldots, g_s)$, we write $Z_F(g_1, \ldots, g_s)$ for $Z_F(\mathfrak{B})$.

So our problem is: Is $Z_F(\mathfrak{A})$ non-empty?

If $\mathfrak{A}$ is the unit ideal, then $Z_F(\mathfrak{A})$ is empty as $f(\underline{x}) = 1$ if f is the constant function 1. So a necessary condition is that $\mathfrak{A} < F[t_1, \ldots, t_n]$. The equation $t_1^2 + t_2^2 = -1$ has no solution in $\mathbb{R}^2$, so $Z_{\mathbb{R}}(t_1^2 + t_2^2 + 1)$ is empty. In particular, $(t_1^2 + t_2^2 + 1)$ lies in no $\mathfrak{m}_{\underline{x}}$ with $\underline{x}$ in $\mathbb{R}^n$. Therefore, in general, the answer is no, and we would need some stronger condition on the field F.

Next suppose that $\mathfrak{A}$ is an ideal in $F[t_1, \ldots, t_n]$ such that there exists an $\underline{x} \in F^n$ with $\mathfrak{A} \subset \mathfrak{m}_{\underline{x}}$. In particular, if $f \in \mathfrak{A}$, then $f \in \mathfrak{m}_{\underline{x}}$, i.e., if $\mathfrak{A} \subset \mathfrak{m}_{\underline{x}}$, then $\underline{x} \in Z_F(\mathfrak{A})$. Conversely, suppose that $\underline{x} \in Z_F(\mathfrak{A})$. If $f \in \mathfrak{A}$, then we have $f(\underline{x}) = 0$, hence f lies in the kernel of the map $e_{\underline{x}} : F[t_1, \ldots, t_n] \to F[t_1, \ldots, t_n]/\mathfrak{m}_{\underline{x}}$, i.e., $f \in \mathfrak{m}_{\underline{x}}$. Therefore,

$$\mathfrak{A} \subset \mathfrak{m}_{\underline{x}} \text{ if and only if } \underline{x} \in Z_F(\mathfrak{A}). \tag{33.2}$$

This says that the answer to our problem will be yes for every ideal $\mathfrak{A} < F[t_1, \ldots, t_n]$, if for each ideal $\mathfrak{A}$, there exists an $\underline{x} \in F^n$ such that $\mathfrak{A} \subset \mathfrak{m}_{\underline{x}}$. In particular, this will be true if and only if whenever $\mathfrak{m}$ is a maximal ideal in $F[t_1, \ldots, t_n]$, there exists an $\underline{x} \in F^n$ satisfying $\mathfrak{m} = \mathfrak{m}_{\underline{x}}$. (Cf. Exercise 23.21(13).) [Recall we have shown (using Zorn's Lemma) that every nonunit ideal in a nontrivial commutative

ring lies in a maximal ideal (for a Noetherian ring this would follow by the Maximal Principle).]

Hilbert proved the following:

Theorem 33.3. (Hilbert Basis Theorem) *Let R be a Noetherian ring. Then $R[t_1, \ldots, t_n]$ is Noetherian. In particular, if F is a field then $F[t_1, \ldots t_n]$ is Noetherian.*

This means that if $\mathfrak{A}$ is ideal in $F[t_1, \ldots, t_n]$ with F a field, then $Z_F(\mathfrak{A}) = Z_F(f_1, \ldots, f_r)$, for some $f_1, \ldots, f_r$ in $F[t_1, \ldots, t_r]$. He also proved the following wonderful theorem:

Theorem 33.4. (Hilbert Nullstellensatz) *Let F be an algebraically closed field. Then every maximal ideal in $F[t_1; , \ldots, t_n]$ is of the form $\mathfrak{m}_{\underline{x}}$ for some $\underline{x}$ in F^n. In particular, if $\mathfrak{A} < F[t_1, \ldots, t_n]$ is an ideal, then $Z_F(\mathfrak{A})$ is not empty.*

The above is also called the *Weak Hilbert Nullstellensatz*. It says, under the additional requirement that F be algebraically closed, the map

$$F^n \longrightarrow \{\mathfrak{m} \mid \mathfrak{m} < F[t_1, \ldots, t_n] \text{ a maximal ideal }\} \text{ given by } \underline{x} \mapsto \mathfrak{m}_{\underline{x}}$$

is a bijection. Actually, Hilbert proved more. In the above, let

$$\sqrt{\mathfrak{A}} = \{f \in F[t_1, \ldots, t_n] \mid f^n \in \mathfrak{A} \text{ for some } n \in \mathbb{Z}^+\},$$

the *radical* of the ideal $\mathfrak{A}$. Then the *Hilbert Strong Nullstellensatz* says, if F is algebraically closed, $\mathfrak{A} < F[t, \ldots t_n]$ an ideal, then

$$\sqrt{\mathfrak{A}} = \bigcap_{\mathfrak{A} \subset \mathfrak{m}_{\underline{x}}} \mathfrak{m}_{\underline{x}}.$$

We will return to this later and prove these assertions.

Exercise 33.5. Let $R = F[t_1, \ldots, t_n]$, F a field, and $\mathfrak{A}$, $\mathfrak{B}$, and $fA_i, i \in I$, be ideals in R. Show the following:

(*i*) If $\mathfrak{A} \subset \mathfrak{B}$, then $Z_F(\mathfrak{B}) \subset Z_F(\mathfrak{A})$.
(*ii*) $Z_F(\emptyset) = F^n$
(*iii*) $Z_F(R) = \emptyset$.
(*iv*) $Z_F(\sum \mathfrak{A}_i) = \bigcap_I (\mathfrak{A}_i)$.
(*v*) $Z_F(\mathfrak{A}\mathfrak{B}) = Z_F(\mathfrak{A} \cap \mathfrak{B}) = Z_F(\mathfrak{A}) \cup Z_F(\mathfrak{B})$.
(*vi*) $Z_F(\mathfrak{A}) = Z_F(\sqrt{\mathfrak{A}})$.

34. Addendum: Algebraic Weierstraß Preparation Theorem

If R is a commutative ring, we have defined the ring of formal power series $R[[t]]$. Inductively, let $R[[t_1, \ldots, t_n]] := (R[[t_1, \ldots, t_{n-1}]])[[t_n]]$, the formal power series in the variables $t_1, \ldots, t_n$.

Previous exercises imply that $R[[t_1,\dots,t_n]]^\times = R^\times$ (cf. Exercise 23.21(2)) and if R is a domain, then so is $R[[t_1,\dots,t_n]]$ (cf. Exercise 23.21(1)) Moreover, if R is Noetherian, so is $R[[t_1,\dots,t_n]]$ by Example 27.13(2). Further, as with polynomials, if $a_1,\dots,a_n$ lie in R, then the evaluation map $e_{a_1,\dots,a_n} : R[[t_1,\dots,t_n]] \to R$ by $f \mapsto f(a_1,\dots,a_n)$ is a ring epimorphism.

In this section we shall show that if F is a field then $F[[t_1,\dots,t_n]]$ is a UFD. By Exercise 23.21(3), it is also a *local ring*, i.e., commutative ring with a unique maximal ideal. One nice thing about formal power series is that we do not have to worry about convergence. This means that proving the algebraic analogue of the Weierstraß Preparation Theorem in complex analysis becomes a formal proof, hence much easier.

Definition 34.1. Let F be a field and f an element of $F[[t_1,\dots,t_n]]$. We say f is *regular* in t_n if $f(0,\dots,0,t_n) \neq 0$. So if f is regular in t_n, the element f has a term ct_n^i for some nonzero element c of F. If f is regular, write

$$f = g_0 + g_1 t_n + g_2 t^2 + \cdots \quad \text{with } g_i \in F[[t_1,\dots,t_{n-1}]] \text{ for all } i$$

and satisfying

$$0 \neq f(0,\dots,0,t_n) = g_0(0,\dots,0) + g_1(0,\dots,0)t_n + g_2(0,\dots,0)t_n^2 + \cdots$$

If s is the first i such that $g_i(0,\dots,0)$ is nonzero, we say s is the *order of* t_n in $f(0,\dots 0,t_n)$ and write $s = \operatorname{ord}_{t_n} f$.

We need two lemmas.

Lemma 34.2. *Let g be a nonzero element in $F[[t_1,t_2]]$. Then there exists a positive integer r satisfying $g(t_2^r,t_2)$ is nonzero.*

Proof. Order the nonzero monomials $a_{r_1r_2}t_1^{r_1}t_2^{r_2}$ of g lexicographically, i.e., $(r_1,r_2) \geq (s_1,s_2)$ if $r_1 > s_1$ or $r_1 = s_1$ and $r_2 \geq s_2$. Let $a_{s_1s_2}$ be the smallest monomial in g relative to this ordering. Choose $r > s_2$ and let $t_1 \mapsto t_2^r$. If $(r_1,r_r) > (s_1,s_2)$, then $rr_1 > rs_1 + s_2$ since either $r_1 > s_1$ (and $r > s_2$) or $r_1 = s_1$ and $r_2 > s_2$. The only term of order $rs_1 + s_2$ in $g(t_2^r,t_2)$ is $a_{s_1s_2}t_2^{rs_1+s_2}$. □

The next lemma shows that the notion of regularity is rather mild.

Lemma 34.3. *Let $f_1,\dots,f_m$ be nonzero elements in $F[[t_1,\dots,t_n]]$. Then there exists a ring automorphism σ of $F[[t_1,\dots,t_n]]$ fixing $F[[t_n]]$ with $\sigma(f_i)$ regular in t_n for $i = 1,\dots,m$.*

PROOF. As $F[[t_1,\dots,t_n]]$ is a domain, $f = f_1\cdots f_m$ is nonzero. So we are done if we find a σ that works for f, i.e., we may assume that $f = f_1$. By the previous lemma, there exist $r_1,\dots,r_{n-1}$ such that $f(t_n^{r_1}, t_2, \dots t_n) \neq 0$, $f(t_n^{r_1}, t_n^{r_2}, t_3, \dots t_n) \neq 0$, $\dots$, $f(t_n^{r_1},\dots,t_n^{r_n-1}, t_n) \neq 0$. Let σ be the ring automorphism of $F[[t_1,\dots,t_n]]$ fixing $F[[t_n]]$ induced by $t_1 \mapsto t_i + t_n^{r_i}$ for $i = 1,\dots,n-1$. (Why does it exist? Its inverse is $t_1 \mapsto t_i - t_n^{r_i}$.) Then $\sigma\big(f(t_1,\dots,t_n)\big) = f(t_1 + t_n^{r_1},\dots,t_{n-1} + t_n^{r_{n-1}}, t_n)$. Consequently,

$$e_{(0,\dots,0,t_n)}\big(\sigma(f)\big) = f(t_n^{r_1}, t_n^{r_2},\dots,t_n^{r_{n-1}}, t_n) \neq 0. \qquad \square$$

Theorem 34.4. (Algebraic Weierstraß Preparation Theorem) *Let $R = F[[t_1,\dots,t_n]]$ with F a field and $f \in R$ regular in t_n. Suppose that $s = \operatorname{ord}_{t_n} f$ and g lies in R. Then there exist unique elements h and r in R satisfying $g = hf + r$ with $r \in F[[t_1,\dots,t_{n-1}]][t_n]$ and $\deg_{t_n} r < s$.*

PROOF. We induct on n.

Suppose that $n = 1$ and $t = t_1$: Then

$$f = c_s t^s + \sum_{j>s} c_j t^j \quad \text{with} \quad c_s \neq 0$$

$$g = \sum_{i=0}^{s-1} a_i t^i + \sum_{i\geq s} a_i t^i = \sum_{i=0}^{s-1} a_i t^i + t^s\big(\sum_{i\geq s} a_i t^{i-s}\big).$$

Let

$$r = \sum_{i=0}^{s-1} a_i t^i \quad \text{and} \quad h = \big(\sum_{i\geq s} a_i t^{i-s}\big)\big(\sum_{j\geq s} c_j t^{j-s}\big)^{-1}.$$

[Note that $c_s \neq 0$, so the second power series in h is a unit by Exercise 23.21(2).] Then r, h satisfy $g = hf + r$. Since $hf = 0$ or $\operatorname{ord}_t(hf) \geq s$, we see that $r = \big(\sum_{i=0}^{s-1} a_i t^i\big)$ is uniquely determined and hence so is h.

Suppose that $n > 1$: The result holds for $R_0 = F[[t_2,\dots,t_n]]$ by induction. Write

$$f = \sum_{i=0}^{\infty} f_i t_1^i \quad \text{and} \quad g = \sum_{i=0}^{\infty} g_i t_1^i$$

with f_i, g_i in R_0 for all i. Set $h = \sum_{i=0}^{\infty} h_i t_1^i$ and $r = \sum_{i=0}^{\infty} r_i t_1^i$ with all the h_i and r_i in R_0 to be determined by the equation

$$\sum_{i=0}^{\infty} r_i t_1^i = \sum_{i=0}^{\infty} g_i t_1^i - \big(\sum_{i=0}^{\infty} h_i t_1^i\big)\big(\sum_{i=0}^{\infty} f_i t_1^i\big),$$

i.e.,

$$
\begin{aligned}
r_0 &= g_0 - h_0 f_0 \\
r_1 &= g_1 - (h_0 f_1 + h_1 f_0) \\
&\vdots \\
r_i &= g_i - (h_0 f_i + \cdots + h_i g_0) \\
&\vdots
\end{aligned}
\tag{*}
$$

Since $f(0,\dots,0,t_n) = f_0(0,\dots,t_n)$ and f_0 in R_0 is regular in t_n with $s = \operatorname{ord}_{t_n}$, by induction there exist unique h_0, r_0 with the appropriate properties, hence unique $r_1, h_1, \dots$, etc. working from the top of (*) down with $\deg_{t_n} r_i < s$ for each i, where we have replaced f by f_0 and g by $g_0 - h_0 f_1$, etc. As all the h_i, r_i are unique, so are r and h. □

We wish to prove that $F[[t_1,\dots,t_n]]$ is a UFD when F is a field by reducing to the polynomial case where we can use the fact that a polynomial ring over a UFD is a UFD. We begin with the following definition.

Definition 34.5. Let $f \in F[[t_1,\dots,t_n]]$ with F a field. Then f is called a *pseudo-polynomial* of *degree* s, if $f = t_n^s + a_1 t_n^{s-1} + \cdots + a_s$ with a_i in $F[[t_1,\dots,t_{n-1}]]$ and satisfying $a_i(0,\dots,0) = 0$ for $1 \le i \le s$.

In this language, the Weierstraß Preparation Theorem implies:

Corollary 34.6. *Let $R = F[[t_1,\dots,t_n]]$ with F a field and f an element of R regular in t_n with $s = \operatorname{ord}_{t_n} f$. Then there exists a unique pseudo-polynomial of degree s that is an associate of f in $F[[t_1,\dots,t_n]]$.*

Proof. Apply the Weierstraß Preparation Theorem to f and $g = t_n^s$, to obtain unique elements $h \in R$ and $r \in F[[t_1,\dots,t_{n-1}]][t_n]$ with $\deg_{t_n} r < s$ satisfying $t_n^s = hf - r$. Write $r = \sum_{i=0}^{s-1} r_i t^i$ with r_i in $F[[t_1,\dots,t_{n-1}]]$ for all i.

Claim. $f^* = t_n^s + r$ works.

Since $f^* = hf$ with h and r unique, it suffices to show that $r_i(0,\dots,0) = 0$ for all i and $h \in R^\times$. For the latter, by Exercise 23.21(2), it suffices to show that $h(0,\dots,0)$ is nonzero. But if it is, then hf has no monomial of the form at_n^s with $a \in F^\times$, contradicting $f^* = t_n^s + r$. Therefore, h is a unit. Write $f = c_s t^s + \sum_{j>s} f_j t_n^j$ with c_s nonzero in F and each $f_j \in F[[t_1,\dots,t_{n-1}]]$. Then

$$
h(0,\dots,0,t_n)\big(c_s t_n^s + \sum_{j>s} f_j(0,\dots,0) t_n^j\big) = t_n^s + \sum_{i=0}^{s-1} r_i(0,\dots,0) t_n^i.
$$

Comparing t_n^s terms shows that $r_i(0, \dots 0) = 0$ for all i. □

Notation 34.7. We denote the unique pseudo-polynomial of f in Corollary 34.6 by f^*.

Corollary 34.8. *Let $R = F[[t_1, \dots, t_n]]$ with F a field and f and g elements of R both regular in t_n. Then $(fg)^* = f^*g^*$.*

PROOF. Note that fg is regular at t_n as f and g are. By the previous result, there exist unique units $u, v, w \in R^\times$ satisfying $f^* = uf$, $g^* = vg$, and $(fg)^* = wfg$ pseudo-polynomials of the appropriate degree. Thus $f^*g^* = uvfg$ and uniqueness shows that $(fg)^* = f^*g^*$. □

Theorem 34.9. *Let F be a field, then $F[[t_1, \dots, t_n]]$ is a UFD.*

PROOF. We induct on n, the case $n = 0$ being immediate. So we may assume that $n \geq 1$. Let $R_0 = F[[t_1, \dots, t_{n-1}]]$ and $R = R_0[[t_n]] = F[[t_1, \dots, t_n]]$. By induction R_0 is a UFD, hence so is the polynomial ring $R_0[t_n]$ (by Theorem 32.8). As R is Noetherian by Example 27.13(2), every nonzero nonunit in R is a product of irreducibles by Theorem 27.14. Let $f \in R$ be irreducible. Therefore, it suffices to show that f is a prime element of R. Suppose that $f \mid rs$ in R with r and s in R. Write $fg = rs$ with $g \in R$. By Lemma 34.3, there exists a ring automorphism σ of R fixing $R_0[[t_n]]$ and satisfying $\sigma(f)$, $\sigma(g)$, $\sigma(r)$, and $\sigma(s)$ are all regular in t_n. Therefore, we may assume that f, g, r, s are all regular in t_n. By Corollary 34.6, we have $f^*g^* = r^*s^*$, so we may further assume that f, g, r, s are all pseudo-polynomials in $R_0[t_n]$. Since $R_0[t_n]$ is a UFD, we are reduced to showing:

Claim. f is irreducible in $R_0[t_n]$.

If $f \in R_0[t_n]$ is reducible, then $f = f_1f_2$ with $f_1, f_2 \in R_0[t_n]$ satisfying $\deg_{t_n} f_i < \deg_{t_n} f$ for $i = 1, 2$. (Why?) But f is irreducible in R, so either f_1 or f_2 lies in $R^\times$, say f_1. Then

$$0 = f_1^{-1}f - f_2$$
$$0 = 0 \cdot f - 0$$

in R. By the uniqueness statement in the Weierstraß Preparation Theorem, $f_1^{-1} = 0$, which is impossible. This proves the claim and finishes the proof of the theorem. □

Remark 34.10. In general, if R is a UFD, it does not follow that $R[[t]]$ is a UFD. For example, it can be shown that $R = F[t_1, t_2, t_3]/(t_1^2 + t_2^3 + t_3^7)$ is a UFD but $R[[t]]$ is not.

Exercises 34.11. 1. Prove that the map σ in the proof of Lemma 34.3 is well-defined.

2. Let $R = F[[t_1, \ldots, t_n]]$ with F a field. Show that $\mathfrak{m} = (t_1, \ldots, t_n)$ is the unique maximal ideal in R and $\mathfrak{m}/\mathfrak{m}^2$ is a vector space over F of dimension n.

Part 4

Module Theory

CHAPTER VIII

Modules

In this chapter, we study the basic result about modules. A module is a generalization of abelian groups and vector spaces. It is an object satisfying all the axioms of a vector space, except that the scalars are allowed to come from any fixed ring instead of from a field. The basic problem in module theory is given a ring with nice properties, classify modules over this ring up to isomorphism. Of course, this meets with relatively little success for all modules, so one looks at classes of modules over the ring.

In the first section, after giving examples of modules over a given ring, we establish the analog of the isomorphism theorems and correspondence principle. Modules are nicer, in that we do not need a stronger property than submodules to state and prove these results compared to those needing normal subgroups and ideals in group and ring theory respectively. In particular, the quotient module of any given module and a submodule always exists.

We also study the special class of free modules. This class of modules generalizes the notion of vector space. More precisely free module are those that have bases, i.e., linearly independent spanning sets. This leads to the fact that dimension (called the rank) of a free module over a commutative ring makes sense, i.e., we can attach a unique integer to a finitely generated free module over a commutative ring. (Cf. the dimension of a finite dimensional vector space.)

35. Basic Properties of Modules

In this section, we introduce the concept of a module over a ring. As mentioned above, this generalizes the definition of vector spaces, by allowing scalars to come from an arbitrary ring rather than just a field. This difference, however, makes much that one expects in vector space theory to be false in this general case. One of the main reasons for this is that the concept of linear independence becomes a rare phenomenon. We begin with the formal definition.

Definition 35.1. Let R be a ring. A (*left*) *R-module* is an additive group M together with a map, $\cdot : R \times M \to M$ called *scalar multiplication*, which we write as $r \cdot x$ for $\cdot(r, x)$, satisfying for all $r, s \in R$ and $x, y \in M$:

(i) $r \cdot (x + y) = r \cdot x + r \cdot y$.
(ii) $(r + s) \cdot x = r \cdot x + s \cdot x$.
(iii) $r \cdot (s \cdot x) = (r \cdot s) \cdot x$.
(iv) $1 \cdot x = x$.

We call the map $\cdot : R \times M \to M$ an *R-action*. Note that there are two $\cdot$'s in the properties above, the R-action of R on M and the ring multiplication on R. Note that (i) and (ii) are the analogues of the distributive law and (iii) the analogue for the associative law for rings. Condition (iv) says the identity of the ring acts like an identity. You should also compare this to G-sets where G is an additive group (or a monoid). We drop the $\cdot$ of the R-action whenever it is clear.

A subgroup N in M is called a *submodule* if it is an R-module under $\cdot|_{R \times N}$.

Remarks 35.2. 1. One defines *right R-modules* in the obvious way.

2. As usual, a submodule of an R-module M should be a subset of M that is an R-module under the inclusion map. It will be so after we define the appropriate notion of homomorphism. Of course, you should already know what it is, as we are generalizing vector space theory.

3. Some authors call M an R-module without assuming (iv) and call an M also satisfying (iv) a *unitary* R-module. (If R is a rng, we will not assume (iv) in our definition.)

4. Let M be an R-module, x an element in M, and r an element in R. Then
   (a) $0_R x = 0_M$. (We write $0x = 0$.)
   (b) $r0_M = 0_M$. (We write $r0 = 0$.)
   (c) $(-r)x = r(-x) = -(rx)$.
   (d) $m(rx) = r(mx)$ for all integers m.
   We leave this an an exercise.

As usual we start with a number of examples.

Examples 35.3. Let R be a ring and M an R-module.

1. If R is a field (or a division ring), then an R-module is the same as a vector space over R.

2. If R is the ring of integers, then a $\mathbb{Z}$-module is just an additive group. In particular, module theory generalizes both vector space theory and abelian group theory.
3. $0 := \{0\}$ and M are submodules of M.
4. R is an R-module under the addition in R with the R-action given by the multiplication in R.
5. Let $\mathfrak{A}$ be a left ideal in R. Then $\mathfrak{A} + \mathfrak{A} \subset \mathfrak{A}$ and $R\mathfrak{A} \subset \mathfrak{A}$, so $\mathfrak{A}$ is a submodule of R. Indeed, a submodule of R is the same thing as a left ideal. So the concept of an R-module also generalizes the notion of left ideal.
6. $R[t]$ is an R-module. More generally, if R is a subring of S then S is an R-module by restricting the ring multiplication on S to scalar multiplication by R, i.e., restricting $\cdot|_{S\times S}$ to $\cdot|_{R\times S}$.
7. Let S be a nonempty set and X an R-module. Set
$$M := \{f : S \to X \mid f \text{ a map}\},$$
then M is an R-module under the following operations: If f and g lie in M and r in R, then for all s in S define
$$(f+g)(s) := f(s) + g(s)$$
$$r(f(s)) : rf(s)$$
and 0_M defined by $0_M(s) = 0_X$.
8. Let N be a submodule of M, then the factor group M/N ($N \lhd M$ is automatic as M is abelian) becomes an R-module by
$$\cdot : R \times M/N \to M/N \text{ defined by } r(m+N) = rm + N$$
for all $r \in R$ and for all $m \in M$. It is called the *quotient* or *factor module* of M by N.
9. Let N be an abelian group. Write it additively, i.e., N is a $\mathbb{Z}$-module. Then
$$\mathrm{End}(N) = \mathrm{End}_{\mathbb{Z}}(N) := \{\varphi : N \to N \mid \varphi \text{ a group homomorphism}\}$$
is a ring via
   (*i*) $(\varphi+\psi)(x) = \varphi(x) + \psi(x)$ (the usual addition of functions into an additive group)
   (*ii*) $\varphi \circ \psi(x) = \varphi\big(\psi(x)\big)$ (composition of functions).

   for all φ and ψ in $\mathrm{End}(N)$ and for all x in N (as $nx = \underbrace{x + \cdots + x}_{n}$ for all positive integers n); and N is an $\mathrm{End}(N)$-module by *evaluation*, i.e., $\varphi \cdot x := \varphi(x)$ for all $\varphi \in \mathrm{End}(N)$ and $x \in N$.

10. Let F be a field, V a vector space over F. Then
$$\mathrm{End}_F(V) = \{T : V \to V \mid T \text{ a linear operator}\}$$
is a subring of $\mathrm{End}_{\mathbb{Z}}(V)$ and V is an $\mathrm{End}_F(V)$-module via *evaluation*, i.e., $T \cdot x = T(x)$ for all $T \in \mathrm{End}_F(V)$ and $x \in V$.
[Can you define the appropriate notion of homomorphism $f : M \to N$ between modules, then replace $\mathrm{End}_F(V)$ by the appropriate $\mathrm{End}_R(M)$? The ring $\mathrm{End}_R(M)$ is called the *endomorphism ring* of M and elements in it are called R-*endomorphisms*. Then M becomes an $\mathrm{End}_R(M)$-module under evaluation, $f \cdot x := f(x)$ for all $x \in M$ and $f \in \mathrm{End}_R(M)$.]
11. Let $\varphi : R \to S$ be a ring homomorphism and N an S-module. Then N becomes an R-module via the *pullback* defined by
$$\cdot : R \times N \to N \text{ given by } r \cdot x = \varphi(r) \cdot x$$
for all r in R and x in N.
12. Let V be a vector space over the field F and $T \in \mathrm{End}_F(V)$ a fixed linear operator (= endomorphism). Define a ring homomorphism
$$\varphi : F[t] \to \mathrm{End}_F(V) \text{ given by the evaluation } f \to f(T).$$
So we have $\sum a_i t^i \mapsto \sum a_i T^i$. (What is $\ker \varphi$? What can you say if V is a finite dimensional vector space over F?) Then V is an $F[t]$-module via the pullback $f \cdot v := f(T)(v)$ for all $f \in F[t]$ and for all $v \in V$. Let W be a subspace of V. Then W is a submodule of the $F[t]$-module V if and only if W is a T-*invariant subspace*, i.e., $T(W) \subset W$. This is an example that will be used in Section §42.
13. Let $\mathfrak{A} \subset R$ be a left ideal. Then
$$\mathfrak{A}M := \{\sum_{x \in M} a_x x \mid a_x \in \mathfrak{A} \text{ for all } x \in M\} \subset M$$
is an R-module. Note that an infinite sum of elements in an additive group does not make sense, so when we write $\sum_{x \in M} a_x x$, we must have a_x is zero *for almost all* x, i.e., $a_x \neq 0$ for only finitely many x.
14. Let M be an R-module and M_i submodules of M for $i \in I$. Then
   (*i*) $\bigcap_I M_i$ and
   (*ii*) $\sum_I M_i := \{\sum m_i \mid m_i \in M_i,\ m_i = 0 \text{ for almost all } i \in I\}$
   are submodules of M. If the M_i also satisfy $M_i \cap \sum_{I \setminus \{i\}} M_j = 0$ for all $i \in I$, we call $\sum_I M_i$ the *internal direct sum* of the M_i and write it as $\bigoplus_I M_i$.
15. Let M_i, $i \in I$, be R-modules. Recall $(m_i)_I$ is the notation for elements in the cartesian product $\otimes_I M_i$. Set

(*i*) $\coprod_I M_i := \{(m_i)_I \mid m_i \in M_i,\ m_i = 0 \text{ for almost all } i \in I\}$, an R-module under componentwise operations. It is called the *(external) direct sum* of the M_i.

(*ii*) $\prod_I M_i := \{(m_i)_I \mid m_i \in M_i\}$, an R-module under componentwise operations. It is called the *(external) direct product* of the M_i.

We have $\coprod_I M_i$ is a submodule of $\prod_I M_i$ with $\coprod_I M_i = \prod_I M_i$ if and only if I is finite or almost all $M_i = 0$.

16. Let X be a subset of M. Define

$$\langle X\rangle := \sum_X Rx,$$

called the submodule *generated* by X. Note that, as for vector spaces, this is equivalent to $\langle X\rangle = \bigcap_{X\subset N\subset M} N$, where N runs over all submodules of M containing X. If there exists a finite subset Y in M such that $M = \langle Y\rangle$ we say that M is *finitely generated* or simply *fg*. If $Y = \{y_1, \ldots, y_n\}$, we write $\langle y_1, \ldots, y_n\rangle$ for $\langle Y\rangle$. If there exists a $y \in M$ such that $M = \langle y\rangle$, we say that M is *R-cyclic*. (So a cyclic group is the same thing as a cyclic $\mathbb{Z}$-module.)

Definition 35.4. A map $f : M \to N$ of R-modules is called an *R-homomorphism* (respectively, *R-monomorphism, R-epimorphism, R-isomorphism*) if f is *R-linear*, i.e.,

$$f(rx + y) = rf(x) + f(y)$$

for all $r \in R$ and $x, y \in M$ (respectively, and injective, and surjective, and bijective with inverse R-linear). As usual we also use the abbreviations of monic and epic. Also, as one would surmise, f is an R-isomorphism if and only if it is a bijective R-homomorphism.

We say that two R-modules M and N are *R-isomorphic* and write $M \cong N$ if there exists an R-isomorphism $g : M \to N$, which we also write as $g : M \xrightarrow{\sim} N$.

Remark 35.5. As is true with groups and rings an R-homomorphism $f : M \to N$ is an R-monomorphism if and only if given any R-homomorphisms $g_1, g_2 : L \to M$ with compositions satisfying $f \circ g_1 = f \circ g_2$, then $g_1 = g_2$; and as with groups (but not rings) f is an R-epimorphism if and only if given any R-homomorphisms $h_1, h_2 : N \to L$ with compositions satisfying $h_1 \circ f = h_2 \circ f$, then $h_1 = h_2$. (Cf. Exercise 1.12(5) and (6).)

Examples 35.6. 1. A $\mathbb{Z}$-homomorphism is the same thing as an abelian group homomorphism.

2. A linear transformation of F-vector spaces is the same thing as an F-homomorphism of F-modules.
3. Let M and N be R-modules with $N \subset M$. Then N is a submodule of M if and only if the inclusion is an R-homomorphism.
4. Let R be a commutative ring and r an element of R. Then

$$\lambda_r : R \to R \text{ given by } x \mapsto rx$$

is an R-homomorphism (but not a ring homomorphism). More generally, if M is an R-module, then

$$\lambda_r : M \to M \text{ given by } x \mapsto rx$$

is an R-homomorphism.

5. Let N be a submodule of M, then the canonical map

$$^- : M \to M/N \text{ given by } x \mapsto \overline{x} = x + N$$

is an R-epimorphism.

6. Let R be a ring and $\mathfrak{A} < R$ be a (2-sided) ideal, $^- : R \to R/\mathfrak{A}$ the canonical ring epimorphism. If M is an R-module then the R-module $M/\mathfrak{A}M$ is also an $R/\mathfrak{A}$-module by $\overline{r}(x + \mathfrak{A}M) = rx + \mathfrak{A}M$ for all $r \in R$ and $x \in M$, so the R- and $(R/\mathfrak{A})$-actions on $M/\mathfrak{A}M$ are *compatible.*, i.e., $r(x + \mathfrak{A}M) = \overline{r}(x + \mathfrak{A}M)$.
7. Let $f : M \to N$ be an R-homomorphism of R-modules. Then the additive subgroups $\ker f \subset M$ and $\operatorname{im} f \subset N$ are submodules. More generally, if N' is a submodule of N, then $\ker f \subset f^{-1}(N') \subset M$ is a submodule and if M' is a submodule of M, then $f(M')$ is a submodule of N. This makes the study of R-modules and R-homomorphisms very nice. We do not need to worry about some of the special properties needed with groups and rings.
8. If M is an R-module, then

$$\operatorname{End}_R(M) := \{f : M \to M \mid f \text{ an } R\text{-homomorphism}\}$$

is a ring under the $+$ of functions and composition (as asked for before), called the *endomorphism ring* of the R-module M. Elements of $\operatorname{End}_R(M)$ are called *R-endomorphisms.* Its unit group is $\operatorname{Aut}_F(M)$, the group of *R-automorphisms of M*, where an *R-automorphism* is an R-isomorphism $M \to M$. The ring $\operatorname{End}(M) = \operatorname{End}_{\mathbb{Z}}(M)$ contains $\operatorname{End}_R(M)$ as a subring. Note that if R is a field, then a linear operator $T : V \to V$ of a vector space V over R is the same thing as an R-endomorphism of V.

9. If R is a commutative ring and M, N are R-modules, then
$$\mathrm{Hom}_R(M,N) := \{f : M \to N \mid f \text{ an } R\text{-homomorphism}\}$$
is an R-module with the usual $+$ for functions and the R-action $\cdot$ given by $r \cdot f : x \mapsto rf(x)$.

We have analogous isomorphism theorems and a correspondence theorem for modules as we had before with the same proof. We state them for convenience.

Theorem 35.7. (First Isomorphism Theorem) *Let $f : M \to N$ be an R-homomorphism, then we have a commutative diagram of R-modules and R-homomorphisms*
$$\begin{array}{ccc} M & \xrightarrow{f} & N \\ {\scriptstyle -}\downarrow & & \uparrow{\scriptstyle \mathrm{inc}} \\ M/\ker f & \xrightarrow[\overline{f}]{} & \mathrm{im}\, f \end{array}$$
with $^-$, the canonical map, an R-epimorphism, $\overline{f}$, the map given by $\overline{x} \mapsto f(x)$, an R-isomorphism between $M/\ker f$ and $\mathrm{im}\, f$, and the inclusion map inc *an R-monomorphism.*

Theorem 35.8. (Second Isomorphism Theorem) *Let M be an R-module and A and N submodules of M. Then $A/(A \cap N) \cong (A+N)/N$.*

Theorem 35.9. (Third Isomorphism Theorem) *Let M be an R-module with submodules A and N satisfying $A \subset N$. Then we have $M/N \cong (M/A)/(N/A)$.*

Theorem 35.10. (Correspondence Principle) *Let $f : M \to N$ be an R-epimorphism of R-modules. Then*
$$\{A \mid \ker f \subset A \subset M \text{ a submodule}\} \longrightarrow \{B \mid B \subset N \text{ a submodule}\}$$
given by $A \to f(A)$ is an order preserving bijection.

Next we introduce the generalization of a zero divisor in a ring.

Definition 35.11. Let M be an R-module and m an element of M. The *annihilator* of m is defined to be the set
$$\mathrm{ann}_R\, m := \{r \in R \mid rm = 0\}.$$

The annihilator has the following properties:

Lemma 35.12. *Let M be an R-module and m, m' elements in M. Then*

(1) $\mathrm{ann}_R\, m \subset R$ *is a left ideal.*

(2) $\rho_m : R \to M$ *given by* $r \mapsto rm$ *is an* R*-homomorphism and satisfies* $\ker \rho_m = \operatorname{ann}_R m$.
(3) *If* ρ_m *is the* R*-homomorphism in (2), then* ρ_m *induces an* R*-isomorphism* $\overline{\rho_m} : R/\operatorname{ann}_R m \xrightarrow{\sim} Rm$.
(4) *If* R *is a commutative ring and* $Rm \subset Rm'$, *then* $\operatorname{ann}_R m \supset \operatorname{ann}_R m'$. *In particular,* $\operatorname{ann}_R m$ *is independent of the generator for* Rm.

PROOF. (1) and (2) follow immediately.

(3) follows from the First Isomorphism Theorem for modules.

(4): Suppose that we know that $m = am'$ for some $a \in R$. If $rm' = 0$, then $ram' = arm' = 0$. □

We can now characterize cyclic R-modules.

Corollary 35.13. *Let* M *be an* R*-module. Then* M *is a cyclic* R*-module if and only if there exists a left ideal* $\mathfrak{A}$ *in* R *satisfying* $M \cong R/\mathfrak{A}$.

PROOF. $R/\mathfrak{A} = \langle 1 + \mathfrak{A} \rangle = R(1 + \mathfrak{A})$ is cyclic, so this follows by the lemma. □

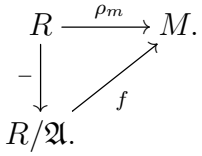
If R is a commutative ring, we can say more.

Proposition 35.14. *Let* R *be a commutative ring,* M *a cyclic* R*-module. Then there exists a unique ideal* $\mathfrak{A}$ *in* R *such that* $M \cong R/\mathfrak{A}$. *Moreover, if* $M = Rm$, *then* $\mathfrak{A} = \operatorname{ann}_R m$. *In particular, if* R *is a PID, there exists an element* a *in* R, *unique up to units, satisfying* $M \cong R/(a)$, *and, if in addition,* $M = Rm$, *then* $(a) = \operatorname{ann}_R m$.

PROOF. We already know if $M = Rm$, then $M \cong R/\operatorname{ann}_R m$. Suppose that we have an R-isomorphism $f : R/\mathfrak{A} \to M$ for some ideal $\mathfrak{A}$ of R. Let $m = f(1 + \mathfrak{A})$. We have

$$f(r + \mathfrak{A}) = f\big(r(1 + \mathfrak{A})\big) = rf(1 + \mathfrak{A}) = rm$$

for all $r \in R$. In particular, $M = Rm$. Let $\rho_m : R \to M$ be the R-epimorphism given by $r \mapsto rm$. Then we have a commutative diagram



Thus

$$\operatorname{ann}_R m = \ker \rho_m = \ker(f \circ {}^-) = \ker {}^- = \mathfrak{A},$$

as f is monic and $f(\overline{r}) = \rho_m(r)$, i.e., $f = \overline{\rho_m}$. By Lemma 35.12(4), $\operatorname{ann}_R m$ is independent of the generator of M when R is commutative. If follows that $\mathfrak{A}$ is unique. Lemma 35.12(4) also implies the statements about PIDs. □

We now set up a lot of useful notation. We write maps $f : A \to B$ as $A \xrightarrow{f} B$ as we have done in commutative diagrams.

Definition 35.15. Let $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$ be R-homomorphisms of R-modules. We say that

$$A \xrightarrow{f} B \xrightarrow{g} C$$

is a *zero sequence* if $g \circ f = 0$, equivalently, $\operatorname{im} f \subset \ker g$ and is an *exact sequence* if $\operatorname{im} f = \ker g$. A longer *sequence*

$$\cdots \xrightarrow{f_{i-1}} A_i \xrightarrow{f_i} A_{i+1} \xrightarrow{f_{i+1}} A_{i+2} \xrightarrow{f_{i+2}} \cdots$$

of R-modules and R-homomorphisms is called a *zero sequence* (respectively, an *exact sequence*) if $f_{j+1} f_j = 0$ (respectively, $\operatorname{im} f_j = \ker f_{j+1}$) for all $j, j+1$ occurring.

Examples 35.16. Let $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$ be R-homomorphisms of R-modules.

1. The R-homomorphism f is monic if and only if $0 \to A \xrightarrow{f} B$ is exact. [Note that $0 \to A$ must take $0 \mapsto 0_A$ as it is an R-homomorphism.]
2. The R-homomorphism g is epic if and only if $B \xrightarrow{g} C \to 0$ is exact.
3. If $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ is exact, it is called a *short exact sequence*. This means that f is monic, $\operatorname{im} f = \ker g$, and g is epic.
4. The sequence $0 \to \ker f \to A \xrightarrow{f} B$ is exact (where the second map is the inclusion) and $0 \to \ker f \to A \xrightarrow{f} \operatorname{im} f \to 0$ is a short exact sequence (writing the same f after changing the target).
5. The R-homomorphism f induces an R-monomorphism $A/\ker f \xrightarrow{\overline{f}} B$ by the First Isomorphism Theorem. Therefore, the First Isomorphism Theorem is just that we have a commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \ker f & \longrightarrow & A & \xrightarrow{\overline{\ }} & A/\ker f & \longrightarrow & 0 \\ & & \| & & \| & & \downarrow{\scriptstyle \overline{f}} & & \\ 0 & \longrightarrow & \ker f & \longrightarrow & A & \xrightarrow{f} & \operatorname{im} f & \longrightarrow & 0. \end{array}$$

with exact rows and with the right hand vertical arrow $\overline{f}$ an R-isomorphism. Can you write the Third Isomorphism Theorem using such sequences?

6. As the factor R-module of M by N exists for any submodule N of an R-module M, besides the image and kernel of the R-homomorphism $B \xrightarrow{g} C$, we can define another R-module, the *cokernel* of g defined by $\operatorname{coker} g := C/\operatorname{im} g$. Then we have a short exact sequence

$$0 \to \operatorname{im} g \to C \xrightarrow{-} \operatorname{coker} g \to 0,$$

where the second map is the inclusion. So we have
   (a) g is injective if and only if $\ker g = 0$.
   (b) g is surjective if and only if $\operatorname{coker} g = 0$.
7. Let A and B be R-modules, then the sequence

$$0 \to A \xrightarrow{\iota_A} A \coprod B \xrightarrow{\pi_B} B \to 0$$

with $\iota_A(a) = (a, 0)$ and $\pi_B(a, b) = b$ is a short exact sequence.
8. Let A_i, B_i, and C_i be R-modules for $i \in I$. If

$$0 \to A_i \xrightarrow{f_i} B_i \xrightarrow{g_i} C_i \to 0$$

is a short exact sequence for all $i \in I$, then

$$0 \to \coprod_I A_i \xrightarrow{\coprod_I f_i} \coprod_I B_i \xrightarrow{\coprod_I g_i} \coprod_I C_i \to 0$$

and

$$0 \to \prod_I A_i \xrightarrow{\prod_I f_i} \prod_I B_i \xrightarrow{\prod_I g_i} \prod_I C_i \to 0$$

are exact. What are the maps $\coprod_I f_i$, $\coprod_I g_i$, and $\prod_I f_i$, $\prod_I g_i$?

Lemma 35.17. (Five Lemma) *Suppose the following is a commutative diagram of R-modules and R-homomorphisms with exact rows:*

$$\begin{array}{ccccccccc}
0 & \longrightarrow & A & \xrightarrow{f} & B & \xrightarrow{g} & C & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle h_A} & & \downarrow{\scriptstyle h_B} & & \downarrow{\scriptstyle h_C} & & \\
0 & \longrightarrow & A' & \xrightarrow[f']{} & B' & \xrightarrow[g']{} & C' & \longrightarrow & 0.
\end{array}$$

If two of h_A, h_B, h_C are R-isomorphisms, then they all are.

The proof of this lemma is called *diagram chasing*, and we leave it as an exercise.

Exercises 35.18. 1. Prove Remark 35.2(4)
2. Prove Remark 35.5.

3. Show an R-homomorphism $f : M \to N$ is an isomorphism if and only if f is bijective.
4. Let M be a *simple* R-module, i.e., M has no proper nonzero submodules. Prove that $\mathrm{End}_R(M)$ is a division ring.
5. Let M be an R-module and M_i submodules of M for $i \in I$. Show that there is always an R-homomorphism $f : \coprod_I M_i \to \sum_I M_i$ and this homomorphism is an R-isomorphism if and only if $\sum_I M_i$ is an internal direct sum.
6. Let M be an R-module and $M_1, \ldots, M_n$ be R-submodules of M. Show that $M = \bigoplus_{i=1}^n M_i$ if and only if there exist R-homomorphisms $\iota_i : M_i \to M$ and $\pi_i : M \to M_i$, $i = 1, \ldots, n$, satisfying all of the following:
   (*i*) $\pi_i \iota_i = 1_{M_i}$ for $i = 1, \ldots, n$.
   (*ii*) $\pi_j \iota_i = 0$ for $i \neq j$.
   (*iii*) $\iota_1 \pi_1 + \cdots + \iota_n \pi_n = 1_M$.
7. Let m be a positive integer. Determine the abelian groups ($\mathbb{Z}$-modules) $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Z}/m\mathbb{Z})$ and $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/m\mathbb{Z}, \mathbb{Z})$ up to isomorphism.
8. Let m and n be positive integers with greatest common divisor d. Show $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}/m\mathbb{Z}, \mathbb{Z}/n\mathbb{Z}) \cong \mathbb{Z}/d\mathbb{Z}$.
9. Let M_i, $i \in I$, be R-modules. Show that the (external) direct sum $\coprod_I M_i$ satisfies the following universal property relative to the R-homomorphisms $\iota_j : M_j \to \coprod_I M_i$ defined by $m \mapsto \{\delta_{ij} m\}_{i \in I}$ for all $j \in I$: Given an R-module M and R-homomorphisms $f_j : M_j \to M$ for all $j \in I$, there exist a unique R-homomorphism $g : \coprod_I M_i \to M$ satisfying $f_j = g \circ \iota_j$ for all $j \in I$.
10. Let M_i, $i \in I$, be R-modules. Show that the (external) direct product $\prod_I M_i$ satisfies the following universal property relative to the R-homomorphisms $\pi_j : \prod_I M_i \to M_j$ defined by $\{m_i\}_I \mapsto m_j$ for all $j \in I$: Given an R-module M and R-homomorphisms $g_j : M \to M_j$ for all $j \in I$, there exist a unique R-homomorphism $h : M \to \prod_I M$ satisfying $g_j = \pi_j \circ h$ for all $j \in I$.
11. Write the Third Isomorphism Theorem using exact sequences as in Example 35.16(5).
12. Prove the Five Lemma 35.17.
13. (Snake Lemma) Let

$$\begin{array}{ccccccccc} & & A & \xrightarrow{f} & B & \xrightarrow{g} & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle \gamma} & & \\ 0 & \longrightarrow & A' & \xrightarrow{f'} & B' & \xrightarrow{g'} & C' & & \end{array}$$

be a commutative diagram of R-modules and R-homomorphisms with exact rows. Prove that

$$\partial : \ker\gamma \to \operatorname{coker}\alpha \text{ defined by } a \mapsto f'^{-1}\beta g^{-1}a + \operatorname{im}\alpha$$

is a well-defined R-homomorphism and the following sequence is exact:

$$\ker\alpha \xrightarrow{f|_{\ker\alpha}} \ker\beta \xrightarrow{g|_{\ker\beta}} \ker\gamma \xrightarrow{\partial} \operatorname{coker}\alpha \xrightarrow{\overline{f}} \operatorname{coker}\beta \xrightarrow{\overline{g}} \operatorname{coker}\gamma$$

Moreover, in this diagram show that $f|_{\ker\alpha}$ is a monomorphism if f is a monomorphism and $\overline{g}$ is an epimorphism if g is an epimorphism.

14. Prove the full version of the Five Lemma: Suppose that

$$\begin{array}{ccccccccc} A & \xrightarrow{\alpha} & B & \xrightarrow{\beta} & C & \xrightarrow{\gamma} & D & \xrightarrow{\delta} & E \\ \downarrow{\scriptstyle h_A} & & \downarrow{\scriptstyle h_B} & & \downarrow{\scriptstyle h_C} & & \downarrow{\scriptstyle h_D} & & \downarrow{\scriptstyle h_E} \\ A' & \xrightarrow{\alpha'} & B' & \xrightarrow{\beta'} & C & \xrightarrow{\gamma'} & D & \xrightarrow{\delta'} & E \end{array}$$

is a commutative diagram of R-modules and R-homomorphisms with exact rows.

(*i*) If h_A is an epimorphism and h_B and h_D are monomorphisms, then h_C is a monomorphism.

(*ii*) If h_E is a monomorphism and h_B and h_D are epimorphisms, then h_C is an epimorphism.

15. A short exact sequence of R-modules

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is called *split* if one of the following three equivalent conditions holds:

(*i*) There exists an R-homomorphism $f' : B \to A$ such that $f'f = 1_A$. We say that f is a *split monomorphism*.

(*ii*) $\operatorname{im} f$ is a *direct summand* of B, i.e., $B = \operatorname{im} f \oplus D$ some R-module D.

(*iii*) There exists an R-homomorphism $g' : C \to B$ such that $gg' = 1_C$. We say that g is a *split epimorphism*.

Prove that these conditions are equivalent.

16. Let R be a commutative ring and M, N R-modules. Recall that $\operatorname{Hom}_R(M,N)$ is an R-module by Example 35.6(9). If $h : A \to B$ is an R-homomorphism of R-modules, define

$$h_* : \operatorname{Hom}_R(N,A) \to \operatorname{Hom}_R(N,B) \text{ by } f \mapsto h \circ f \text{ and}$$
$$h^* : \operatorname{Hom}_R(B,N) \to \operatorname{Hom}_R(A,N) \text{ by } f \mapsto f \circ h.$$

Show that these are R-homomorphisms and if

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is a short exact sequence of R-modules and R-homomorphisms, then

$$0 \to \mathrm{Hom}_R(N, A) \xrightarrow{f_*} \mathrm{Hom}_R(N, B) \xrightarrow{g_*} \mathrm{Hom}_R(N, C) \quad \text{and}$$

$$0 \to \mathrm{Hom}_R(C, N) \xrightarrow{g^*} \mathrm{Hom}_R(B, N) \xrightarrow{h^*} \mathrm{Hom}_R(A, N)$$

are exact. (Note the missing 0 on the right.)

17. Let Q be an R-module. Then Q is called *R-injective* if given any R-monomorphism $f : A \to B$ and R-homomorphism $g : A \to Q$, there exists an R-homomorphism $h : B \to Q$ such that the diagram

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle g}\downarrow & \swarrow_{h} & \\ Q & & \end{array}$$

commutes. Show that Q is an R-injective if and only if, whenever

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is a short exact sequence of R-modules and R-homomorphisms, then

$$0 \to \mathrm{Hom}_R(C, Q) \xrightarrow{g^*} \mathrm{Hom}_R(B, Q) \xrightarrow{f^*} \mathrm{Hom}_R(A, Q) \to 0$$

is exact. (Cf. Exercise (16).)

18. Let Q be an R-module. Show that Q is an R-injective if and only if given any ideal $\mathfrak{A}$ in R and an R-homomorphism $g : \mathfrak{A} \to Q$, there exists an R-homomorphism $h : B \to Q$ such that the diagram

$$\begin{array}{ccc} \mathfrak{A} & \xrightarrow{\text{inc}} & B \\ {\scriptstyle g}\downarrow & \swarrow_{h} & \\ Q & & \end{array}$$

commutes where inc is the inclusion.

19. Show that $\mathbb{Q}$ is an injective $\mathbb{Z}$-module.

36. Free Modules

Vector spaces have nice properties because of the concept of linear independence and the fact that one can divide by any nonzero scalar. In general, R-modules have neither of these two properties. There is little we can do about the lack of units in an arbitrary ring, but we can look at R-modules that are spanned by linearly independent sets. Such

modules are called *free* R-modules. Though sparse, these modules are quite nice, in fact, just as a vector space is isomorphic to a external direct sum of copies of the underlying field, the analogous statement is true for free R-modules. The first result one would then expect is that the notion of dimension makes sense, i.e., any two linearly independent spanning set of a free module should have the same cardinality. Unfortunately, this is false in general. The problem does vanish, however, if the ring R is commutative. We shall leave the proof of this fact as an exercise. This is fortuitous as these are the rings we are studying in detail.

Definition 36.1. A nonzero R-module M is called a *free* R-module (or *R-free*) if there exists a *basis* for M, i.e., a subset $\mathcal{B}$ of M satisfying:

(*i*) $M = \langle \mathcal{B} \rangle$, i.e., $\mathcal{B}$ *generates* or *spans* M.

(*ii*) $\mathcal{B}$ is *linearly independent*, i.e., if $0 = \sum_{\mathcal{B}} r_x x$ (so $r_x = 0$ for almost all $x \in \mathcal{B}$), then $r_x = 0$ for all $x \in \mathcal{B}$. Equivalently, $M = \bigoplus_{\mathcal{B}} Rx$ and if $rx = 0$ with $x \in \mathcal{B}$, $r \in R$, then $r = 0$.

We shall also call the trivial module a free R-module.

Of course, a set $\mathcal{D}$ in M is called *linearly dependent* if it is not linearly independent. Equivalently, there exist $x_1, \ldots, x_n \in \mathcal{D}$, some n, and $r_1, \ldots, r_n$ in R not all zero such that $r_1 x_1 + \cdots + r_n x_n = 0$.

If M is a free R-module on basis $\mathcal{B}$, the definition says that $\mathcal{B}$ generates M and there are no nontrivial relations on these generators. Contrast this with a finite (abelian) group say of order n which says that every element x satisfies the relation $x^n = e$. In particular, this is true of any element in any generating set.

Examples 36.2. Let R be a ring and M be a nontrivial R-module.

1. R is R-free on basis $\{u\}$ with u a unit in R.
2. Suppose that M is R-cyclic. Then M is R-free if there exists an $x \in M$ satisfying $M = Rx$ and $\operatorname{ann}_R x = 0$. It follows that M is R-free if and only if $M \cong R$.
3. If M is a finitely generated R-module, then M is R-free if and only if there exists a finite basis for M if and only if
$$M \cong \underbrace{R \coprod \cdots \coprod R}_{n} = R^n \text{ for some positive integer } n. \text{ (Why?)}$$
4. If R is a field (or a division ring), then every R-module is free and any two bases have the same cardinality.

5. If R is commutative and M is R-cyclic, then M is R-free if and only if $\operatorname{ann}_R m = 0$ whenever $M = Rm$, i.e., for any generator of M.

   To show this one needs to use the first two parts of Exercise 36.12(6) (which establishes that all bases for a finitely generated free module over a commutative ring have the same cardinality). If we know this and M is R-free, then say every basis for M has say n elements. Let $\mathcal{B}$ be a basis for M and $\mathfrak{m}$ be a maximal ideal in R. Then this exercise would allow us to prove the following sequence of isomorphisms:

$$\begin{aligned} Rm/\mathfrak{m}m = M/\mathfrak{m}M &= (\bigoplus_{\mathfrak{B}} Rx)/\big(\mathfrak{m}(\bigoplus_{\mathfrak{B}} Rx)\big) \\ &\cong R^n/\mathfrak{m}R^n \cong (R/\mathfrak{m})^n \end{aligned}$$

   as R- and as $R/\mathfrak{m}$-modules. As $R/\mathfrak{m}$ is a field, this is an isomorphism of vector spaces over $R/\mathfrak{m}$. As $m + \mathfrak{m}m$ generates $Rm/\mathfrak{m}m$, by (3), we must have $n = 1$.

6. If R is not commutative, it can happen that $R^m \cong R^n$ for different positive integers, i.e., different bases can have different cardinalities. For example, let M be a free R-module on a countable (not finite) basis (e.g., $M = R[t]$ (cf (9) below)). Let S be the endomorphism ring $\operatorname{End}_R(M)$. Then it can be shown that S^n is S-free and $S^n \cong S^m$ for all positive integers n and m.

7. Suppose that M is R-free with basis $\mathcal{B}$ and $x \in \mathcal{B}$. Then the map $\rho_x : R \to M$ given by $r \mapsto rx$ is an R-monomorphism. In particular, the cardinality of M is at least that of R. This gives another proof that no nontrivial finite abelian group is $\mathbb{Z}$-free.

8. Let M_i be free R-modules for $i \in I$. Then $\coprod_I M_i$ is R-free. [If $\mathcal{B}_i$ is a basis for M_i for each $i \in I$, what is a basis for $\coprod_I M_i$?] For example $\coprod_{i=1}^{\infty} \mathbb{Z}$ is $\mathbb{Z}$-free on the *standard basis*

$$\mathcal{S} = \{e_i = (0, \dots, \underset{i}{1}, 0, \dots) \mid i \geq 1\}.$$

   [Warning: It turns out that $\prod_{i=0}^{\infty} \mathbb{Z}$ is not $\mathbb{Z}$-free.]

9. $\mathbb{Q}$ is not $\mathbb{Z}$-free.

10. $R[t]$ is R-free on basis $\mathcal{B} = \{t^i \mid i \geq 0\}$. One can now define the ring $R[t]$ to be the free R-module on basis $\mathcal{B}$ with multiplication induced by $t^i \cdot t^j := t^{i+j}$ (and extend linearly). More generally, $R[t_1, \dots, t_n]$ is the free R-module on basis $\{t_1^{i_1} \cdots t_n^{i_n} \mid i_j \geq 0 \text{ for all } i_j\}$.

11. Let R_0 be a ring and $R = R_0[t_1, t_2]$. The R-module $M = Rt_1 + Rt_2$ is a submodule of R as it is a left ideal. Then M is not R-free. (Proof?)

    [This shows that a submodule of a free R-module may not be free.]

12. Let $n \mid m$ in $\mathbb{Z}^+$ and $^- : \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ be the ring epimorphism given by $r + m\mathbb{Z} \mapsto r + n\mathbb{Z}$. Then $\mathbb{Z}/n\mathbb{Z}$ is a $(\mathbb{Z}/m\mathbb{Z})$-module via the pullback, e.g., $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$ are $(\mathbb{Z}/6\mathbb{Z})$-modules. By element count neither $\mathbb{Z}/2\mathbb{Z}$ nor $\mathbb{Z}/3\mathbb{Z}$ are $\mathbb{Z}/6\mathbb{Z}$-free, but $\mathbb{Z}/2\mathbb{Z} \coprod \mathbb{Z}/3\mathbb{Z} \cong \mathbb{Z}/6\mathbb{Z}$ as $(\mathbb{Z}/6\mathbb{Z})$-modules (why?), so a *direct summand* of a free module may not be free. (A submodule N is a *direct summand* of an R-module M if $M = N \oplus M'$ for some submodule M'.)
13. If $M = \bigoplus_I Rx_i$ is a free R-module on basis $\mathcal{B} = \{x_i \mid i \in I\}$ and m is an element in M, then there exist unique $r_i \in R$, $i \in I$, almost all zero such that $m = \sum_I r_i x_i$. The r_i are called the *coordinates* of m *relative to the basis* $\mathcal{B}$ and r_i is called the *coordinate* of m on x_i.
14. The set $\{2\}$ is linearly independent in the free $\mathbb{Z}$-module $\mathbb{Z}$, but it is not a basis as $2\mathbb{Z} < \mathbb{Z}$. Indeed any two distinct elements in $\mathbb{Z}$ form a linearly dependent set. This shows that, in general, not every linearly independent set in a free R-module need be part of a basis.
15. Let R be a domain and M a free R-module. If z is a nonzero element in M satisfying $rz = 0$ for some $r \in R$, then $r = 0$ (for if $\mathcal{B}$ is a basis, then $z = \sum_{\mathcal{B}} r_x x$).

Universal properties usually produce maps satisfying certain properties together with a uniqueness statement. This is very useful, as one of the more difficult problems is to show the existence of a map, or to prove that a potential explicit map is well-defined. Usually the problem arises as we know generators but cannot show that the putative map respects the relations among these generators. As a basis for a free module does not satisfy any nontrivial relations, this causes no problems and free modules satisfy the following universal property:

Theorem 36.3. (Universal Property of Free Modules) *Let $\mathcal{B} = \{x_i\}_I$ be a basis for a free R-module M. If N is an R-module and y_i, $i \in I$, elements in N (not necessarily distinct), then there exists a unique R-homomorphism $f : M \to N$ such that $x_i \mapsto y_i$ for all $i \in I$.*

PROOF. If $z \in M$, there exist unique $r_i \in R$, almost all $r_i = 0$, such that $z = \sum_I r_i x_i$. In particular, the uniqueness of the r_i implies that $f : M \to N$ given by $z \mapsto \sum_I r_i y_i$ is well-defined. Clearly, f is uniquely determined by $x_i \mapsto y_i$ and f is an R-homomorphism. □

In the notation of the theorem, we say that $x_i \mapsto y_i$ *extends linearly* to a homomorphism $f : M \to N$. The theorem says that any R-homomorphism from a free module M to another R-module is completely determined by where a basis for M is sent. Note, of course, that the variant of this statement (and proof) is one that you should know about vector spaces, which are just free modules over a field.

Remark 36.4. A better way of writing the universal property of free modules is the following: Let M be a free R-module on basis $\mathcal{B}$ and N an R-module. Given any set map $g : \mathcal{B} \to N$ there exists a unique R-homomorphism $f : M \to N$ such that the diagram

$$\begin{array}{ccc} \mathcal{B} & \xrightarrow{inc} & M \\ & \searrow^{g} & \downarrow f \\ & & N. \end{array}$$

commutes where *inc* is the inclusion map.

Corollary 36.5. *Let M and N be free R-modules on bases $\mathcal{B}$ and $\mathcal{C}$ respectively. If there exists a bijection $g : \mathcal{B} \to \mathcal{C}$, i.e., $|\mathcal{B}| = |\mathcal{C}|$, then $M \cong N$.*

Proof. The maps g and g^{-1} of sets induce inverse R-isomorphisms $M \to N$ and $N \to M$. □

Corollary 36.6. *Let M be an R-module (respectively, a finitely generated R-module). Then there exists a free R-module (respectively, a finitely generated free R-module) P and an R-epimorphism $g : P \to M$. In particular, we have a short exact sequence*

$$0 \to \ker g \longrightarrow P \xrightarrow{g} M \to 0.$$

Proof. Let $Y = \{y_i\}_I$ generate M. If M is finitely generated, we may assume that I is finite. Let $P = \coprod_I R$ and $\mathcal{S}_I$ the *standard basis* for P. (So $\mathcal{S}_I := \{e_i \mid e_i = (\delta_{ij})_{j\in I})\}_I$, i.e., e_i has 1 in the ith component, 0 elsewhere.) Then the set map $\mathcal{S}_I \to M$ given by $e_i \mapsto y_i$ extends linearly to an R-homomorphism $g : P \to M$. As Y generates M, the map is surjective. □

Unfortunately, the kernel of g in the corollary may not be nice, in particular, it may not be free. If R is a PID, it will be free. We shall show this later under the assumption that M is finitely generated.

Example 36.7. Let R be a commutative ring and M a free R-module on basis $\mathcal{B}$. As R is commutative, $M^* := \mathrm{Hom}_R(M, R)$ is an R-module, called the *dual module* of M. (Warning: It is usually not R-free.) For each $x \in \mathcal{B}$ define the *xth coordinate function* $f_x : M \to R$ by $\sum_{\mathcal{B}} r_x x \mapsto r_x$. It is an R-homomorphism. Set $\mathcal{B}^* := \{f_x \mid x \in \mathcal{B}\}$. The map $\mathcal{B} \mapsto \mathcal{B}^*$ by $x \mapsto f_x$ extends linearly to an R-homomorphism $f : M \mapsto M^*$. This map is monic and its image is R-free on basis $\mathcal{B}^*$. We call $\mathcal{B}^*$ the *dual basis* of $\mathcal{B}$. Unfortunately, this map depends on the basis $\mathcal{B}$, so it is not "natural". However, if we let $M^{**} := \mathrm{Hom}_R(M^*, R)$ and $L : M \to M^{**}$ the evaluation map at each x in M, i.e., the map L

is given by $x \mapsto L_x : f \mapsto f(x)$, then this map is an R-monomorphism and is "natural" as it does not depend on the choice of a basis. If $\mathcal{B}$ is finite, then f is also surjective, hence $M \cong M^*$ (and $M \cong M^{**}$).

We end this section with results left as important exercises.

Definition 36.8. We say that a ring R satisfies the *invariant dimension property* or *IDP* if for every finitely generated free R-module P, every basis for P has the same cardinality (finite by Example 36.2(3). If R satisfies IDP, and P is a finitely generated free R-module, then the cardinality of a basis for P is called the *rank* of P and written $\operatorname{rank} P$.

We know that fields satisfy IDP. In fact, we have shown this (cf. Remark 16.13). It is a main ingredient in the proof of the following theorem that we leave as an exercise (cf. Exercise 36.12(6)).

Theorem 36.9. *Let R be a commutative ring. Then R satisfies IDP, i.e., if M is a finitely generated free R-module then all bases of M have the same cardinality.*

We also leave the proof of the following corollary to this theorem as an exercise (Exercise 36.12(7)).

Corollary 36.10. *Let R be a commutative ring and M and N be finitely generated free R-modules. Then* $\operatorname{rank} M \coprod N = \operatorname{rank} M + \operatorname{rank} N$.

Remark 36.11. It can be shown for any ring R, if M is a free R-module on basis $\mathcal{B}$ with $\mathcal{B}$ not finite, i.e., M is not finitely generated, then every basis of M has the same cardinality.

Exercises 36.12. 1. Show if M is a nonzero finitely generated R-module, then M is R-free if and only if there exists a finite basis for M if and only if $M \cong R^n$ for some positive integer n.

2. Show that $\mathbb{Q}$ is not $\mathbb{Z}$-free. (Hint: Show any two distinct elements of $\mathbb{Q}$ are $\mathbb{Z}$- linearly dependent.)
3. Let R_0 be a commutative ring and $R = R_0[t_1, t_2]$. Let $M = Rt_1 + Rt_2$ Show the R-module M is not R-free.
4. Let R be a nontrivial commutative ring. Show that R is a field if and only if every finitely generated R-module is free.
5. Suppose that
$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$
is a short exact sequence of R-modules with C a free R-module. Show the sequence splits. (Cf. Exercise 35.18(15) for the definition of a split exact sequence.)

6. Let P be a free R-module on basis $\mathcal{B} = \{x_i\}_{i\in I}$. Let $\mathfrak{A} < R$ be a (2-sided) ideal. Show all of the following:
   (*i*) $P/\mathfrak{A}P \cong \coprod_I Rx_i/\mathfrak{A}x_i \cong \coprod_I R/\mathfrak{A}$.
   (*ii*) Let $^- : R \to R/\mathfrak{A}$ be the canonical ring epimorphism. Set $\overline{\mathcal{B}} = \{\overline{x}_i := x_i + \mathfrak{A}P \mid i \in I\}$. Then $P/\mathfrak{A}P$ is a free $\overline{R}$-module on basis $\overline{\mathcal{B}}$ and $|\overline{\mathcal{B}}| = |\mathcal{B}|$.
   (*iii*) Let $\varphi : R \to S$ be a ring epimorphism with $S \neq 0$. If S satisfies IDP so does R.
   (*iv*) Any commutative ring satisfies IDP.
7. Prove Corollary 36.10.
8. Show Example 36.2(6) has the desired properties.
   [Hint: If $\mathcal{B} = \{e_i \mid i \in \mathbb{Z}^+\}$ is a basis for P, show that $\{f_1, f_2\}$ is a basis for S, where for all n, we have $f_1(e_{2n}) = e_n$, $f_1(e_{2n+1}) = 0$ and $f_2(e_{2n}) = 0$, $f_2(e_{2n+1}) = e_n$.]
9. Let P be an R-module. Then P is called *R-projective* if given any R-epimorphism $f : B \to C$ and R-homomorphism $g : P \to C$, there exists an R-homomorphism $h : P \to B$ such that the diagram

$$\begin{array}{ccc} & & P \\ & {\scriptstyle h}\swarrow & \downarrow{\scriptstyle g} \\ B & \xrightarrow[f]{} & C \end{array}$$

   commutes. Show that any free R-module is projective.
10. Show that a direct summand of an R-free module is projective and a direct sum of R-modules is projective if and only if each direct summand of it is R-projective. (Projective modules are nicer than free modules in some ways, but submodules of projective modules may not be projective, e.g., $\mathbb{Z}/2\mathbb{Z}$ is a projective $\mathbb{Z}/6\mathbb{Z}$-module as $\mathbb{Z}/2\mathbb{Z} \coprod \mathbb{Z}/3\mathbb{Z} \cong \mathbb{Z}/6\mathbb{Z}$) as $\mathbb{Z}/6\mathbb{Z}$-modules).
11. Let P be an R-module. Show that P is a projective R-module if and only if, whenever

$$(*) \qquad 0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

   is a short exact sequence of R-modules and R-homomorphisms, then

$$0 \to \mathrm{Hom}_R(P, A) \xrightarrow{f_*} \mathrm{Hom}_R(P, B) \xrightarrow{g_*} \mathrm{Hom}_R(P, C) \to 0$$

   is exact. In particular, if C is R-projective, then (*) is split exact. (Cf. Exercises 35.18(16) and 35.18(17).)

12. Let G_i, $i \in I$, be groups. Show that $\bigtimes_I G_i$ satisfies the following property relative to the group homomorphisms $\pi_j : \bigtimes_I G_i \to G_j$ defined by $\{g_i\}_I \mapsto g_j$ for all $j \in I$: Given a group G and group homomorphisms $\phi_j : G \to G_j$, there exist a unique group homomorphism $\psi : G \to \bigtimes_I G_i$ satisfying $\phi_j = \pi_j \circ \psi$ for all $j \in I$.

13. Let R be a commutative ring and M, N two R-modules. Let P be the free R-module on basis $\{(m,n) \mid m \in M,\, n \in N\}$ and X the submodule of P generated by the elements

    (i) $(m_1+m_2,n)-(m_1,n)-(m_2,n)$ for all $m_1, m_2 \in M$ and $n \in N$.
    (ii) $(m,n_1+n_2)-(m,n_1)-(m,n_2)$ for all $m \in M$ and $n_1, n_2 \in N$.
    (iii) $(rm,n)-r(m,n)$ for all $m \in M$, $n \in N$, and $r \in R$.
    (iv) $(m,rn)-r(m,n)$ for all $m \in M$, $n \in N$, and $r \in R$.

    Let $f : M \times N \to P/X$ be the R-bilinear map induced by $(m,n) \mapsto (m,n)+X$, i.e, an R homomorphism in each variable. Show that $f : M \times N \to P/X$ satisfies the following universal property: If $g : M \times N \to Q$ is an R-bilinear map, then there exists a unique R-homomorphism $h : P/X \mapsto Q$ such that

    $$\begin{array}{ccc} M \times N & \xrightarrow{f} & P/X \\ & \searrow^{g} & \downarrow{\scriptstyle h} \\ & & Q \end{array}$$

    commutes. The R-module P/X is called the *tensor product* of M and N and denoted by $M \otimes_R N$ and the elements $(m,n)+X$ are denoted by $m \otimes n$.

14. Let R be a commutative ring and M and N be R-modules. Show that $M \otimes_R N \cong N \otimes_R M$.

15. Let R be a commutative ring and M an R-module. Show that $M \cong R \otimes_R M$.

16. Let m and n be positive integers with greatest common divisor d. Show $\mathbb{Z}/m\mathbb{Z} \otimes \mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/d\mathbb{Z}$.

17. Let R be a commutative ring and M_i, $i \in I$, and N be R-modules. Show $(\coprod_I M_i) \otimes_R N \cong \coprod_I (M_i \otimes_R N)$. In particular, show if M and N are free R-modules (respectively, and finitely generated), then $M \otimes_R N$ is a free R-module (respectively, of $\operatorname{rank} M \cdot \operatorname{rank} N$).

18. Let R be a commutative ring and $f_1 : M_1 \to N_1$ and $f_2 : M_2 \to N_2$ be R-homomorphisms. Show these induce a unique R-homomorphism $f_1 \otimes f_2 : M_1 \otimes_R M_2 \to N_1 \otimes_R N_2$ satisfying $m_1 \otimes m_2 \mapsto f_1(m_1) \otimes f_2(m_2)$ for all $m_1 \in M_1$, $m_2 \in M_2$.

19. Let R be a commutative ring and
$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$
an exact sequence of R-modules and R-homomorphisms. Show for every R-module M, the sequence
$$A \otimes_R M \xrightarrow{f} B \otimes_R M \xrightarrow{g} C \otimes_R M \to 0$$
is exact.
20. Let R be a commutative ring and A, B, C be R-modules. Show that $\mathrm{Hom}_R(A, \mathrm{Hom}_R(B, C) \cong \mathrm{Hom}_R(A \otimes_R B, C)$.

CHAPTER IX

Noetherian Rings and Modules

Although this chapter is relatively short, it includes some of the most important basic results in commutative algebra, viz. the theorems of Hilbert introduced in Section §33. It is based upon the concept of a Noetherian ring, previously introduced, and its generalization to modules. Hilbert's theorems are necessary to begin the study of algebraic geometry. In particular, they show any set of polynomials not generating the unit ideal in $F[t_1, \ldots, t_n]$, with F an algebraically closed field, have a common zero. A brief introduction to affine plane curves is also discussed.

37. Noetherian Modules

Proposition 37.1. *Let R be a ring and M an R-module. Then the following are equivalent:*

(1) Every submodule of M is finitely generated.

(2) M satisfies the ascending chain condition*, i.e., if $M_i \subset M$ are submodules and*

$$M_1 \subset M_2 \subset \cdots \subset M_n \subset \cdots,$$

then there exists a positive integer N such that $M_N = M_{N+i}$ for all $i \geq 0$. We say every ascending chain of submodules of M stabilizes*. Equivalently, there exists no infinite chain*

$$M_1 < M_2 < \cdots < M_n < \cdots.$$

(3) M satisfies the Maximum Principle*, i.e., if S is a non-empty set of submodules of M, then S contains a maximal element, that is a module $M_0 \in S$ such that if $M_0 \subset N$ with $N \in S$, then $N = M_0$.*

An R-module M satisfying any (hence all) of these equivalent conditions is called a *Noetherian R-module* or *R-Noetherian*.

Proof. (1) $\Rightarrow$ (2): Let

$$\mathcal{C}: \quad M_1 \subset M_2 \subset \cdots \subset M_n \subset \cdots$$

be a chain of submodules of M. It follows that the subset $M' := \bigcup_{i=1}^{\infty} M_i \subset M$ is a submodule. By (1), it is finitely generated, so we can write $M' = \sum_{i=1}^{n} Rx_i$ for some $x_i \in M'$. By definition, $x_i \in M_{j_i}$ some j_i. Let s be the maximum of the finitely many j_i's. Then $M' = M_s$. It follows that $M_s = M' = M_{s+i}$ for all $i \geq 0$.

(2) $\Rightarrow$ (3): Suppose that S is a non-empty set of submodules of M. Let M_1 lie in S. If M_1 is not maximal, there exists an $M_2 \in S$ with $M_1 < M_2$. Inductively, if M_i is not maximal, there exists an M_{i+1} in S with $M_i < M_{i+1}$. By the ascending chain condition, the sequence

$$M_1 < M_2 < \cdots < M_i < \cdots$$

must terminate.

[To prove this statement completely, note that we made a choice of M_2 to contain M_1. This really needs the Axiom of Choice. (Cf. the note in Exercise 27.22(5).)]

(3) $\Rightarrow$ (1): Let $N \subset M$ be a submodule and set

$$S := \{M_i \mid M_i \subset N \text{ is a finitely generated submodule } \}.$$

Then $(0) \in S$ so $S \neq \emptyset$. By assumption, there exists a maximal element $M' \in S$. If $N \neq M'$, then there exists $x \in N \setminus M'$. But M' finitely generated means that $M' + Rx \subset N$ is also finitely generated, so the submodule $M' + Rx$ of N lies in S. This contradicts the maximality of M'. Hence $N = M'$ is finitely generated. □

Remark 37.2. Let $R = F[t_1, \ldots, t_n, \ldots]$ (infinitely many t_i). Let $M = R$ as an R-module. Then M is finitely generated since cyclic but the ideal $(t_1, \ldots, t_n, \ldots)$ is clearly not finitely generated, so R is not a Noetherian R-module. Thus, in general, submodules of finitely generated modules need not be finitely generated.

Definition 37.3. Let R be a commutative ring. We say that R is a *Noetherian ring* if R is a Noetherian R-module.

Note that Definition 37.3 agrees with our previous definition of a Noetherian ring.

We need another Noetherian R-module result. It is a crucial property about Noetherian modules.

Proposition 37.4. *Let M be an R-module and N a submodule of M. Then M is R-Noetherian if and only if N and M/N are R-Noetherian. In particular, if*

$$0 \to M' \to M \to M'' \to 0$$

is an exact sequence of R-modules with two of the modules M, M', M'' being R-Noetherian, then they all are R-Noetherian.

PROOF. $\Rightarrow$: Since $N_0 \subset N$ is a submodule, then $N_0 \subset M$ is a submodule hence finitely generated – or any ascending chain in N is an ascending chain in M. Thus N is R-Noetherian. By the Correspondence Principle, a (countable) chain of submodules in M/N has the form $M_1/N \subset M_2/N \subset \cdots$ where $N \subset M_1 \subset M_2 \subset \cdots$ is a chain of submodules of M. Thus there exists an r such that $M_r = M_{r+j}$ for all $j \geq 0$ and hence $M_r/N = M_{r+j}/N$ for all $j \geq 0$.
$\Leftarrow$ is left as an exercise (cf. the analogous solvable groups result). □

Corollary 37.5. *If M, N are Noetherian R-modules, so is $M \coprod N$.*

PROOF. $(M \coprod N)/N \cong M$ and N are Noetherian. □

The following result shows that the collection of Noetherian R-modules is a good generalization of finite dimensional vector spaces.

Theorem 37.6. *Let R be a Noetherian ring. If M is a finitely generated R-module, then M is R-Noetherian.*

PROOF. Suppose $M = \sum_{i=1}^{n} Rx_i$. Let $f : R^n \to M$ be the R-epimorphism given by $e_i \mapsto x_i$, where $\{e_1, \ldots, e_n\}$ is the standard basis for R^n. Since R is R-Noetherian so is R^n by Corollary 37.5, and hence so is $M \cong R^n/\ker f$ by Propostition 37.4. □

Corollary 37.7. *Let R be a Noetherian ring and M a finitely generated R-module. Then there exists positive integers m and n and an exact sequence*

$$R^m \xrightarrow{g} R^n \xrightarrow{f} M \to 0. \tag{37.8}$$

PROOF. As is the proof of the theorem there exists an integer n and an R-epimorphism $f : R^n \to M$. Since M is finitely generated hence R-noetherian, $\ker f$ is also finitely generated. Therefore, there exists an R-epimorphism $h : R^m \to \ker f$. Setting g to be the composition of f and the inclusion yields the result. □

Remark 37.9. If M is an R-module for which a sequence (37.8) exists, we say that M is a *finitely presented R-module*. The corollary says that any finitely generated module over a Noetherian ring is finitely presented. This is very useful. A commutative ring in which every finitely generated module is finitely presented is called a *coherent* ring. Therefore, every Noetherian ring is coherent.

Proposition 37.10. *Suppose that $f : R \to S$ is a ring epimorphism of commutative rings. If R is a Noetherian ring, then S is also a Noetherian ring.*

PROOF. Let $\mathfrak{A} \subset S$ be an ideal. Then $f^{-1}(\mathfrak{A}) \subset R$ is an ideal hence finitely generated. Thus $\mathfrak{A} = f\big(f^{-1}(\mathfrak{A})\big)$ is finitely generated. □

Exercises 37.11. 1. Complete the proof of Proposition 37.4.

2. Let M be a Noetherian R-module. Show that any surjective R-endomorphism $f : M \to M$ is an isomorphism.

3. Let M be an R-module. Prove the following are equivalent:
   (*i*) M satisfies DCC (the *descending chain condition*), i.e., if $M_i \subset M$ are submodules and
   $$M_1 \supset M_2 \supset \cdots \supset M_n \supset \cdots$$
   then there exists a positive integer N such that $M_N = M_{N+i}$ for all $i \geq 0$.
   (*ii*) M satisfies the *Minimum Principle*, i.e., if S is a non-empty set of submodules of M, then S contains a minimal element, that is a module $M_0 \in S$ such that if $M_0 \supset N$ with $N \in S$, then $N = M_0$.

   If M satisfies one of these two equivalent conditions, we say that M is an *Artinian R-module*.

   [One needs the Axiom of Choice to show DCC implies the Minimum Principle.)

   Also show that if R is a field then an R-module V is Artinian if and only if V is finite dimensional.

4. Let
   $$0 \to M' \to M \to M'' \to 0$$
   be an exact sequence of R-modules. Show that if two of the modules M, M', M'' are R-Artinian, then they all are R-Artinian.

5. Let R be a Noetherian ring, $\mathfrak{A}$ an ideal in R, A a finitely generated R-module, and B a submodule of A. Suppose that C is a submodule of A that contains $\mathfrak{A}B$ and is maximal with respect to the property that $C \cap B = \mathfrak{A}B$. Let x be an element of $\mathfrak{A}$. Show all of the following:
   (*i*) The chain of ideals $\{a \in A \mid x^m a \in C \text{ for all } a \in A\}$, $m \in \mathbb{Z}^+$, stabilizes.
   (*ii*) There exists an integer n such that $(x^n A + C) \cap B = \mathfrak{A}B$.
   (*iii*) $\mathfrak{A}^n A \subset C$ for some n.
   (*iv*) (Krull Intersection Theorem) If $B = \bigcap_{i=0}^{\infty} \mathfrak{A}^n A$, then $\mathfrak{A}B = B$.

6. Let R be a commutative ring, $\mathfrak{A}$ an ideal in R, M an R-module generated by n elements, and x an element of R satisfying $xM \subset \mathfrak{A}M$. Show that $(x^n + y)M = 0$ for some y in $\mathfrak{A}$. In particular, if $\mathfrak{A}M = M$, then $(1 + y)M = 0$ for some $y \in \mathfrak{A}$.

7. Let R be a Noetherian ring. Using the previous two exercises, show the following:
   (*i*) Suppose that R is a domain and $\mathfrak{A}$ an ideal in R. Let M be a finitely generated R-module satisfying $\operatorname{ann}_R(m) = 0$ for all $m \in M$. (We say that M is *R-torsion-free.*) Then $\bigcap_{i=0}^{\infty} \mathfrak{A}^n M = 0$.
   (*ii*) Let $\mathfrak{A} = \bigcap_{\operatorname{Max}(R)} \mathfrak{m}$, the *Jacobson radical* of R (cf. Exercise 25.17(6)). Then $\bigcap_{i=0}^{\infty} \mathfrak{A}^n M = 0$.

38. Hilbert's Theorems

In this section, we prove the theorems that we discussed in Section §33. These theorems form the foundation of affine algebraic geometry.

Theorem 38.1. (Hilbert Basis Theorem) *If R is a Noetherian ring so is the ring $R[t_1, \ldots, t_n]$.*

PROOF. By induction on n, it suffices to show that $R[t]$ is Noetherian. Let $\mathfrak{B} \subset R[t]$ be an ideal. We must show that $\mathfrak{B}$ is finitely generated. Let

$$\mathfrak{A} = \{r \in R \,|\, r = \operatorname{lead} f,\ f \in \mathfrak{B}\}.$$

(Recall $\operatorname{lead} f$ is the leading coefficient of f.)

We first show that $\mathfrak{A}$ is an ideal. (Note that $0 \in \mathfrak{A}$ as $0 \in \mathfrak{B}$.) Let $a, b \in \mathfrak{A}$ and $r \in R$ with $ra + b$ nonzero. Choose $f, g \in \mathfrak{B}$ say of degrees m and n respectively satisfying $\operatorname{lead} f = a$ and $\operatorname{lead} g = b$. Set $h = rt^n f + t^m g$ in $\mathfrak{B}$. Then $ra + b = \operatorname{lead} h$ proving $\mathfrak{A}$ is an ideal. As R is Noetherian, $\mathfrak{A} = (a_1, \ldots, a_n)$ for some $a_1, \ldots, a_n \in \mathfrak{A}$ with $n \in \mathbb{Z}^+$. Choose f_{id_i} in $\mathfrak{B}$ such that $a_i = \operatorname{lead} f_{id_i}$ and $\deg f_{id_i} = d_i$. Let $\mathfrak{B}_0 = (f_{1d_1}, \ldots, f_{nd_n})$, an ideal in $R[t]$, and $N = \max\{d_1, \ldots, d_n\}$.

Let $f \in \mathfrak{B}$ with $\operatorname{lead} f = a$ and $\deg f = d$. Suppose that $d > N$. There exist $r_i \in R$ satisfying $a = \sum_{i=1}^n r_i a_i$, hence $f - \sum_{i=1}^n r_i t^{d-d_i} f_{i,d_i}$ lies in $\mathfrak{B}$ and has degree less than d. It follows by induction that there exists a $g \in \mathfrak{B}_0$ such that $f - g$ lies in $\mathfrak{B}$ with $\deg(f - g) \leq N$. As the R-module $M := \sum_{i=0}^N Rt^i$ is finitely generated and R is Noetherian, M is a Noetherian R-module. In particular, the submodule $M_0 = \{f \in \mathfrak{B} \mid \deg f \leq N\}$ is finitely generated. If $M_0 = \sum_{i=0}^m R[t]g_i$, then $\mathfrak{B} = \mathfrak{B}_0 + \sum_{i=0}^m R[t]g_i$ is finitely generated. □

It is worth noting the similarity of this proof and that for the general division algorithm. Of course, the division algorithm is a much stronger property, but the idea of a proof often generalizes.

Remarks 38.2. 1. Noetherian domains need not be UFD's, as $\mathbb{Z}[\sqrt{-5}]$ is an example of such a domain.

2. If F is a field, $F[t_1, \ldots, t_n, \ldots]$ is a UFD but is not Noetherian.

Definition 38.3. Let $R \subset S$ be commutative rings. We say that S is a *finitely generated commutative R-algebra* (or an *affine R-algebra* when R is a field) if there exist $x_1, \ldots, x_n$ in S satisfying $S = R[x_1, \ldots, x_n]$ as rings. (Recall that $R[x_1, \ldots, x_n]$ is the image of the evaluation map $e_{x_1,\ldots,x_n} : R[t_1, \ldots t_n] \to S$.)

Remarks 38.4. Let R be a commutative ring.

1. $R[t]$ is a finitely generated commutative R-algebra but definitely not a finitely generated R-module. (Why?)
2. Let S be a commutative ring with R a subring. If S is a finitely generated R-module, then S is a finitely generated commutative R-algebra.
3. If S is a finitely generated commutative R-algebra, it is not in general a finitely generated R-module, a much stronger condition. (Cf. (1).)
4. If T is a finitely generated commutative S-algebra and S is a finitely generated commutative R-algebra, then T is a finitely generated commutative R-algebra.
5. If T is a commutative ring with S a subring such that T is a finitely generated S-module and S is a finitely generated commutative R-algebra, then T is a finitely generated commutative R-algebra.

Proposition 38.5. *Let R be a commutative ring and S a finitely generated commutative R-algebra. If R is Noetherian so is S.*

Proof. Let $S = R[x_1, \ldots, x_n]$. Since $R[t_1, \ldots, t_n]$ is Noetherian and we have a ring epimorphism $R[t_1, \ldots, t_n] \to R[x_1, \ldots, x_n]$ via $f(t_1, \ldots, t_n) \mapsto f(x_1, \ldots, x_n)$, all ideals of S are finitely generated by the Correspondence Principle. □

Remark 38.6. Let R be a commutative ring. The proof (or definition) shows that S is a finitely generated commutative R-algebra if and only if there exists an ring epimorphism $R[t_1, \ldots, t_n] \to S$ fixing R. Our definition of a finitely generated commutative R-algebra S is not the usual one. In general, one does not assume that $R \subset S$, hence only that there exists a surjective ring homomorphism $R[t_1, \ldots, t_n] \to S$ for some n. If R is a field, we can always assume that $R \subset S$, as the above map must be monic when restricted to R if R is a field. This is the case of interest here, except for the following lemma.

Lemma 38.7. (Artin-Tate) *Let T be a commutative ring and $R \subset S$ be subrings of T. Suppose that R is Noetherian and T is a finitely generated commutative R-algebra. Suppose that as an S-module T is finitely generated. Then S is a finitely generated commutative R-algebra.*

PROOF. Let $T = R[x_1, \ldots, x_n]$ as a commutative R-algebra for some $x_i \in T$, some n and $T = \sum_{i=1}^{m} Sy_i$ as an S-module for some $y_i \in T$, some m. Then for all $i = 1, \ldots, n$ and $p, q = 1, \ldots, m$, we have equations:

(*i*) $x_i = \sum_{j=1}^{m} a_{ij} y_j$ for some $a_{ij} \in S$ and

(*ii*) $y_p y_q = \sum_{k=1}^{m} b_{pqk} y_k$ for some $b_{pqk} \in S$ (since T is a ring).

Let $S_0 = R[a_{ij}, b_{pqk} \mid i = 1, \ldots, n \text{ and } j, p, q, k = 1, \ldots, m]$, a finitely generated R-algebra. Then $R \subset S_0 \subset S \subset T$.

Claim. T is a finitely generated S_0-module.

Let $f \in T$, so $f = \sum c_{i_1,\ldots,i_n} x_1^{i_1} \cdots x_n^{i_n}$ for some $c_{i_1,\ldots,i_n} \in R$. Applying properties (i) and (ii) repeatedly shows that $f \in \sum_{i=1}^{m} S_0 y_i$. Consequently, $T = S_0 y_1 + \cdots + S_0 y_m$ as claimed.

Since S_0 is a finitely generated commutative R-algebra, it is Noetherian by Corollary 38.5. Thus T, being a finitely generated S_0-module, is a Noetherian S_0-module. Consequently, $S \subset T$ is a finitely generated S_0-module. It follows immediately that S is a finitely generated commutative R-algebra. □

Let R be an affine F-algebra, say $R = F[x_1, \ldots, x_n]$. If R is a domain, we write the quotient field of R by $F(x_1, \ldots, x_n)$. We shall need the following two lemmas about fields that we leave as easy exercises.

Lemma 38.8. *Suppose that $L \subset K \subset M$ are fields, then M is a finite dimensional vector space over L if and only if both M is a finite dimensional vector space over K and K is a finite dimensional vector space over L.*

Lemma 38.9. *Let $L \subset M$ be fields. Suppose that $x \in M$ has the property that $L[x]$ is not a finite dimensional vector space over L. Then $L[x] \cong L[t]$ as rings and $L(x) \cong L(t)$ as fields.*

The clever Artin-Tate Lemma allows us to give an elementary proof of the following result that is essentially a form of the Hilbert Nullstellensatz.

Lemma 38.10. (Zariski's Lemma) *Let F be a field and E a field containing F such that E is an affine F-algebra. Then E is a finite dimensional vector space over F (i.e., a finitely generated F-module).*

PROOF. Suppose that $E = F[x_1, \ldots, x_m]$. Since E is a field, $E = F(x_1, \ldots, x_m)$. Suppose that E is not a finite dimensional vector space over F. Using Lemma 38.8, we see that $F(x_i)$ is an infinite dimensional vector space over F for some i. Relabeling, we may assume $i = 1$. Continuing in this way, we see that after relabeling the x_i, we may assume that $F(x_1, \ldots, x_i)$ is not a finite dimensional vector space over $F(x_1, \ldots x_{i-1})$ for $1 \leq i \leq r$, some r, and E is a finite dimensional fvector space over $F(x_1, \ldots, x_r)$.

Let $K = F(x_1, \ldots, x_r)$. We have $E = K(x_{r+1}, \ldots, x_m)$ is a finitely generated K-module and F is a Noetherian ring (since a field). Thus the Artin-Tate Lemma implies that K is a finitely generated F-algebra. Write $K = F[y_1, \ldots, y_n]$ for some $y_i \in K$. By Lemma 38.9, it follows that and $F[x_1, \ldots, x_r] \cong F[t_1, \ldots, t_r]$ and $K \cong F(t_1, \ldots, t_r)$ as rings. Thus we can write

$$y_i = \frac{f_i(x_1, ..., x_r)}{g_i(x_1, ..., x_r)}$$

for some $f_i, g_i \in F[t_1, \ldots, t_r]$, $g_i \neq 0$, $1 \leq i \leq n$. Let $g = g_1 \cdots g_n$ in $F[t_1, \ldots, t_r]$. Thus $g(x_1, ..., x_r) \in F[x_1, \ldots, x_r]$. We know the UFD $F[t_1, \ldots, t_r]$ contains infinitely many non-associative irreducibles hence prime elements by Exercise 31.19(9). In particular, there exists an irreducible $f \in F[t_1, \ldots, t_r]$ such that $f \nmid g$. Thus $f(x_1, ..., x_r) \nmid g(x_1, ..., x_r)$ in $F[x_1, \ldots, x_r]$. As K is a field,

$$\frac{1}{f(x_1, ..., x_r)} \in K = F[y_1, \ldots, y_n] = F\Big[\frac{f_1(x_1, ..., x_r)}{g_1(x_1, ..., x_r)}, \ldots, \frac{f_n(x_1, ..., x_r)}{g_n(x_1, ..., x_r)}\Big].$$

This leads to an equation for $1/f(x_1, ..., x_r)$. Choosing an appropriate $N \geq 0$ to clear denominators, we see that we have an equation

$$\frac{g(x_1, ..., x_r)^N}{f(x_1, ..., x_r)} \in F[f_1(x_1, ..., x_r), \ldots, f_n(x_1, ..., x_r)].$$

Then $\dfrac{g(x_1, ..., x_r)^N}{f(x_1, ..., x_r)}$ lies in $F[x_1, \ldots, x_r]$, i.e., $f | g^N$ in $F[t_1, \ldots, t_r]$, a contradiction. □

Recall from §33, if $R = F[t_1, \ldots, t_n]$ and $\mathfrak{A} \subset R$ is an ideal, the *affine variety* of $\mathfrak{A}$ in F^n is defined by

$$Z_F(\mathfrak{A}) = \{\underline{a} = (a_1, ..., a_n) \in F^n \mid f(\underline{a}) = 0 \text{ for all } f \in \mathfrak{A}\}.$$

By the Hilbert Basis Theorem, there exist $f_1, \ldots, f_r$ in $F[t_1, \ldots, t_n]$ satisfying $\mathfrak{A} = (f_1, \ldots, f_r)$. We then also write $Z_F(f_1, \ldots, f_r)$ for $Z_F(\mathfrak{A})$. In particular, $\underline{a}$ lies in $Z_F(\mathfrak{A})$ if and only if $f_i(\underline{a}) = 0$ for $i = 1, \ldots, r$, as any element in $\mathfrak{A}$ is an R-linear combination of the f_i's.

Theorem 38.11. (Hilbert Nullstellensatz) (Weak Form) *Suppose that F is an algebraically closed field, $R = F[t_1, \ldots, t_n]$, and $\mathfrak{A} = (f_1, \ldots, f_r)$ is an ideal in R. Then $Z_F(\mathfrak{A})$ is the empty set if and only if $\mathfrak{A}$ is the unit ideal. In particular, if $\mathfrak{A} < R$, then there exists a point $\underline{a} \in F^n$ satisfying $f_1(\underline{a}) = 0, ..., f_r(\underline{a}) = 0$.*

PROOF. Certainly if $\mathfrak{A} = R$, then there exists no $\underline{a} \in F^n$ such that the element 1 in R evaluated at $\underline{a}$ takes the value zero, so we need only show if $\mathfrak{A} < R$, then $Z_F(\mathfrak{A})$ is not empty.

So suppose that $\mathfrak{A} < R$. Then there exists a maximal ideal $\mathfrak{m} < R$ satisfying $\mathfrak{A} \subset \mathfrak{m}$. Let $^- : R \to R/\mathfrak{m}$ be the canonical ring epimorphism. Set $E = R/\mathfrak{m} = F[\bar{t}_1, \ldots, \bar{t}_n]$. As R is an affine F-algebra, so is E. By Zariski's Lemma, E is a finite dimensional F-vector space.

Claim. $E = F$:

Indeed let $x \in E$. Then $F[x]$ is a finite dimensional F-vector space, so $\{1, x, x^2, \ldots, x^n\}$ must be linear dependent over F for some n, i.e., x is a root of some non-zero polynomial $f \in F[t]$. But any such f factors completely over F, as F is algebraically closed. Therefore, $E = F$.

Consequently the point $\underline{t} = (\bar{t}_1, \ldots, \bar{t}_n)$ in $E^n = F^n$ lies in $Z_F(\mathfrak{A})$, since $\mathfrak{A} \subset \mathfrak{m}$, i.e., $\overline{\mathfrak{A}} = \overline{\mathfrak{m}} = 0$ □

Recall if F is a field and $\mathfrak{m}_{\underline{x}}$ is the ideal $(t_1 - x_1, \ldots, t_n - x_n)$ with $\underline{x} = (x_1, \ldots, x_n)$ in F^n, then $\mathfrak{m}_{\underline{x}}$ is a maximal ideal and Equation 33.2 says

$$(*) \qquad \mathfrak{A} \subset \mathfrak{m}_{\underline{x}} \text{ if and only if } \underline{x} \in Z_F(\mathfrak{A}).$$

for an ideal $\mathfrak{A}$ in $F[t_1, \ldots, t_n]$.

The Weak Form of the Hilbert Nullstellensatz says that the analogue of Exercise 23.21(13) that the maximal ideals in the ring of continuous real-valued functions on a finite closed (or compact) set are in one to one correspondence with the points of the set is valid over an algebraically closed field.

Corollary 38.12. *Let F be an algebraically closed field and $\mathfrak{m}$ a maximal ideal in $F[t_1, \ldots, t_n]$. Then there exists an element $\underline{x}$ in F^n such that $\mathfrak{m} = \mathfrak{m}_{\underline{x}}$.*

PROOF. By the Weak Nullstellensatz, there exists a point $\underline{x}$ in $Z_F(\mathfrak{m})$. Let $^- : F[t_1, \ldots, t_n] \to F[t_1, \ldots, t_n]/\mathfrak{m}_{\underline{x}}$ be the canonical epimorphism. It takes $t_i \to \overline{t_i} = x_i$ for each i. By (*), we have $\mathfrak{m} \subset \mathfrak{m}_{\underline{x}}$. So $\mathfrak{m} = \mathfrak{m}_{\underline{x}}$ as $\mathfrak{m}$ is a maximal ideal. □

Note that the corollary says if F is algebraically closed, then there is a bijection between points in F^n and maximal ideals in $F[t_1, \ldots, t_n]$

and the points in $Z_F(\mathfrak{A})$ bijective with the maximal ideals containing $\mathfrak{A}$ in $F[t_1, \ldots, t_n]$. (Cf. Exercise 23.21(13).)

The geometric interpretation of the Weak Hilbert Nullstellensatz is that the intersection of the "hyperplanes"

$$f_1 = 0, \ldots, f_r = 0 \quad \text{in} \quad F^n.$$

always contains a common point over an algebraically closed field F, unless the f_i generate the unit ideal in $F[t_1, \ldots, t_n]$.

Recall also that F being algebraically closed is essential. Indeed $f(t_1, ..., t_n) = t_1^2 + \cdots + t_n^2 + 1$ has no solution in $\mathbb{R}^n$ yet $(f) < \mathbb{R}[t_1, \ldots, t_n]$. Can you state what the above argument shows when F is not algebraically closed?

The Weak Hilbert Nullstellensatz says that $f_1, \ldots, f_r$ determine $Z_F(f_1, \ldots, f_r)$ when F is algebraically closed. The natural question is what polynomials in $F[t_1, \ldots, t_n]$ does $Z_F(f_1, \ldots, f_r)$ determine when F is algebraically closed, i.e., what polynomials in $F[t_1, \ldots, t_n]$ have zeros at every point in $Z_F(f_1, \ldots, f_r)$. Certainly elements in $\mathfrak{A} = (f_1, \ldots, f_r)$ have this property. However, a point in F^n is a zero of a polynomial f in the domain $F[t_1, \ldots t_n]$ if and only if it is a zero of f^m for any $m > 0$ if and only if it is a zero of f^m for some $m > 0$. This means that $Z_F(\mathfrak{A}) = Z_F(\sqrt{\mathfrak{A}})$ for any ideal $\mathfrak{A}$, i.e., at a minimum, every element in the *radical ideal* $\sqrt{\mathfrak{A}} := \{x \mid x^n \in F[t_1, \ldots t_n] \text{ for some positive integer } n\}$ of $\mathfrak{A}$ has every point in $Z_F(\mathfrak{A})$ a zero. The Strong Hilbert Nullstellensatz says, when F is algebraically closed, this is the totality of such polynomials.

Theorem 38.13. (Hilbert Nullstellensatz) (Strong Form) *Suppose that F be an algebraically closed field and $R = F[t_1, \ldots, t_n]$. Let $f, f_1, ..., f_r$ be elements in R and $\mathfrak{A} = (f_1, ..., f_r) \subset R$. Suppose that $f(a) = 0$ for all $a \in Z_F(\mathfrak{A})$. Then there exists an integer m such that $f^m \in \mathfrak{A}$, i.e., $f \in \sqrt{\mathfrak{A}}$. In particular, if $\mathfrak{A}$ is a prime ideal, then $f \in \mathfrak{A}$.*

Proof. (Rabinowitch Trick). We may assume that f is nonzero. Let $S = R[t]$. Define the ideal $\mathfrak{B}$ in S by $\mathfrak{B} = (f_1, ..., f_r, 1 - tf) \subset S$. If $\mathfrak{B} < S$, then there exists a point $(a_1, ..., a_{n+1}) \in Z_F(\mathfrak{B})$. Thus $f_i(a_1, ..., a_n) = 0$ for all i and $1 - a_{n+1} f(a_1, ..., a_n) = 0$. In particular, $(a_1, ..., a_n)$ lies in $Z_F(\mathfrak{A})$. By hypothesis, this means that $f(a_1, ..., a_n) = 0$ which in turn implies that $1 = 0$, a contradiction. Thus $\mathfrak{B} = S$. So we can write

$$1 = \sum_{i=1}^{r} g_i f_i + g \cdot (1 - tf)$$

for some $g, g_i \in S$. Substituting $1/f$ for t and clearing denominators yields the result. □

If R is a commutative ring, and $\mathfrak{A}$ an ideal in R, then applying the Correspondence Principle to the canonical epimorphism $^- : R \to R/\mathfrak{A}$ shows that $\sqrt{\mathfrak{A}}$ is the intersection of all prime ideals containing $\mathfrak{A}$. The strong form of the Nullstellensatz has an algebraic version that says for any field F, if R is an affine F-algebra and $\mathfrak{A}$ an ideal in R, then

$$\sqrt{\mathfrak{A}} = \bigcap_{\substack{\mathfrak{A}\subset\mathfrak{m} \\ \mathfrak{m} \text{ a maximal ideal}}} \mathfrak{m}.$$

We do not prove this here (The proof can be found below in Theorem 85.10 using localization techniques.) Instead we indicate what is going on (also without proof) assuming the reader knows what a topology is.

Let R be a nonzero commutative ring and set

$$\operatorname{Spec}(R) := \{\mathfrak{p} \mid \mathfrak{p} < R \text{ a prime ideal}\},$$

called the *Spectrum* of R. If T is a subset of R, define

$$V_R(T) := \{\mathfrak{p} \mid \mathfrak{p} \in \operatorname{Spec}(R) \text{ with } T \subset \mathfrak{p}\}$$

called a *variety*. . Note that $V_F(T) = V_F(\langle T\rangle)$. We leave the following as an exercise.

Lemma 38.14. *Let R be a commutative ring, $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{A}_i, i \in I$, be ideals in R. Then*

(1) If $\mathfrak{A} \subset \mathfrak{B}$, then $V_R(\mathfrak{B}) \subset V_R(\mathfrak{A})$.
(2) $V_R(\emptyset) = \operatorname{Spec}(R)$
(3) $V_R(R) = \emptyset$.
(4) $V_R(\sum \mathfrak{A}_i) = \bigcap_I V_F(\mathfrak{A}_i)$.
(5) $V_R(\mathfrak{A}\mathfrak{B}) = V_R(\mathfrak{A} \cap \mathfrak{B}) = V_R(\mathfrak{A}) \cup V_R(\mathfrak{B})$.
(6) $V_R(\mathfrak{A}) = V_R(\sqrt{\mathfrak{A}})$.

Lemma (2) − (4) means that the collection

$$\mathcal{C} := \{V_R(T) \mid T \subset R\}$$

forms a system of *closed sets* for a topology on $\operatorname{Spec}(R)$ called the *Zariski topology*.

One shows that the set

$$\operatorname{Max}(R) := \{\mathfrak{m} \mid \mathfrak{m} < R \text{ a maximal ideal}\}$$

of *maximal ideals* in R is precisely the set of *closed points* in $\operatorname{Spec}(R)$, i.e., a closed set with precisely one element. If F is a field, then the collection

$$\mathcal{Z} := \{Z_F(T) \mid T \subset F[t_1, \ldots, t_n]\}$$

also forms a system of closed sets for a topology of F^n by Exercise 33.5 called the *Zariski topology* for F^n. If F is algebraically closed, the weak form of the Nullstellensatz says if $\mathrm{Max}(R)$ is given the induced topology, then

$$\mathrm{Max}(F[t_1, \ldots, t_n]) \to F^n \text{ given by } \mathfrak{m}_{\underline{x}} \mapsto \underline{x}$$

is a homeomorphism and the strong form of the Nullstellensatz says that $\mathrm{Max}(F[t_1, \ldots, t_n])$ is dense in $\mathrm{Spec}(F[t_1, \ldots, t_n])$, i.e., every non-empty open set in $\mathrm{Spec}(R)$ intersects $\mathrm{Max}(R)$ nontrivially. This means that the closed points determine varieties. This last statement generalizes to $\mathrm{Max}(R)$ is dense in $\mathrm{Spec}(R)$ for any affine F-algebra R. (Cf. with (*) above.)

Exercises 38.15. 1. Prove that if R is a Noetherian ring so is the formal power series $R[[t]]$.

2. Prove Lemma 38.8. If M is a finite dimensional vector space over L, then $\dim_L M = \dim_k M \dim_L K$? (Compare this to the generalized version of Lagrange's Theorem for groups.)
3. Prove Lemma 38.9.
4. Prove if R is a commutative ring, then $\sqrt{\mathfrak{A}}$ is the intersection of all prime ideals. containing $\mathfrak{A}$.
5. Prove Lemma 38.14.
6. Let R be a commutative ring. Prove that $\mathrm{Spec}(R)$ contains minimal elements under $\subset$.

39. Addendum: Affine Plane Curves

Let F be a field and $\mathfrak{A}$ an ideal in $F[t_1, \ldots, t_n]$. Then the affine variety $Z_F(\mathfrak{A}) = Z_F(\sqrt{\mathfrak{A}})$ (cf. Exercise 33.5). Note if $\mathfrak{p}$ is a prime ideal then $\mathfrak{p} = \sqrt{\mathfrak{p}}$. One can show:

Fact 39.1. Let F be a field and $\mathfrak{A} < F[t_1, \ldots, t_n]$ an ideal, then

$$(*) \qquad Z_F(\mathfrak{A}) = Z_F(\mathfrak{p}_1) \cup \cdots \cup Z_F(\mathfrak{p}_n)$$

where $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ are the finitely many prime ideals in $F[t_1, \ldots, t_n]$ that *minimally contain* $\mathfrak{A}$ (i.e., if $\mathfrak{P}$ is a prime ideal of $F[t_1, \ldots, t_n]$ satisfying $\mathfrak{A} \subset \mathfrak{P}$, then there exists an i such that $\mathfrak{A} \subset \mathfrak{p}_i \subset \mathfrak{P}$).

We know that $F[t_1, \ldots, t_n]$ is a Noetherian ring. To show that there are finitely many such $\mathfrak{p}_i$ in $F[t_1, \ldots, t_n]$ in Fact 39.1 uses this. (cf. Exercise 27.22(13)). This reduces the study of $Z_F(\mathfrak{A})$ to $Z_F(\mathfrak{p})$ with $\mathfrak{p}$ a prime ideal in $F[t_1, \ldots, t_n]$. Such a variety is called *irreducible*. The decomposition (*) for $Z_F(\mathfrak{A})$ is called the *irreducible decomposition* of $\mathbb{Z}_F(\mathfrak{A})$ (cf. Exercise 27.22(13)) and the $Z_F(\mathfrak{p}_i)$ are called the *irreducible*

components of $Z_F(\mathfrak{A})$. As $F[t_1, \ldots, t_n]$, is a UFD, it also turns out that every irreducible affine variety of "codimension one" in F^n is $Z_F(f)$, for some prime element f. [This essentially follows from the fact that every nonzero prime ideal in a UFD contains a prime element together with Exercise 28.3(3)) and "dimension theory", that we have not and will not develop that shows how this defines codimension algebraically.]

We shall need to use Exercise 32.12(3ii) in the development below. As we are interested in affine varieties in F^2, it is convenient to use X, Y as our variables instead of t_1, t_2.

Example 39.2. Let F be a field and $\mathfrak{A} = (XY)$ in $F[X, Y]$. Then $Z_F(\mathfrak{A}) = Z_F(X) \cup Z_F(Y)$ is an irreducible decomposition. Here $Z_F(X)$ is the Y-axis defined by $X = 0$ and $Z_F(Y)$ is the X-axis defined by $Y = 0$. The origin is the only point in $Z_F(X, Y)$. As $Z_F(X, Y) \subset Z_F(XY)$ (cf. Exercise 33.5), we see that irreducible components can intersect nontrivially. If $f \in F[X, Y]$ is a non-constant polynomial then $Z_F(f)$ is called an *affine (plane) curve* and *irreducible* if f is an irreducible polynomial as $Z_F(f)$ is the locus of $f = 0$ in F^2.

Proposition 39.3. *Let R be a UFD and f and g non-constant polynomials in $R[X, Y]$ having no non-constant common factor in $R[X, Y]$. Then*

$$Z_R(f) \cap Z_R(g) := \{(a, b) \in R^2 \mid f(a, b) = 0 = g(a, b)\}$$

is a finite set in R^2.

Proof. Let F be the quotient field of R and K the quotient field of $F[X]$. We view f and g in $K[Y]$. By Exercise 32.12(3ii), f and g have no common factor in $K[Y]$, i.e., they are relatively prime in the PID $K[Y]$, so there exists an equation, $1 = rf + sg$, with r and s polynomials in $K[Y]$. Clearing the denominators of the K-coefficients leads to an equation $0 \neq h = wf + zg$ in $R[X, Y]$, with w and z in $R[X, Y]$ and h in $R[X]$. Any common solution (a, b) in $Z_F(g) \cap Z_F(g)$ then satisfies $h(a) = h(a, b) = 0$ in R^2. Since h has finitely many roots in the domain R, there exist finitely many a satisfying $(a, b) \in Z_F(g) \cap Z_F(g)$. Applying the same argument with the variables reversed shows that $Z_F(g) \cap Z_F(g)$ is a finite set. □

Remark 39.4. (Cf. Exercise 33.5). Let F be a field, $\mathfrak{A}$ and $\mathfrak{B}$ ideals in $F[X, Y]$, then

$$Z_F(\mathfrak{A} \cap \mathfrak{B}) = Z_F(\mathfrak{A}\mathfrak{B}) = Z_F(\mathfrak{A}) \cup Z_F(\mathfrak{B}).$$

The remark, reduces our study to ideals $\mathfrak{A} = (f_1, \ldots, f_r)$ in $F[X, Y]$ with a gcd of the $f_1, \ldots, f_r$ equal to one, i.e., with no common factors.

If h is a polynomial in $R[X,Y]$, R a ring, define the *total degree* $\deg h$ of h to be

$$\deg h := \max\{i+j \mid aX^iY^j \text{ a monomial in } h \text{ with } a \text{ nonzero}\}.$$

The proposition can be improved to

Fact 39.5. (Bezout's Lemma) Let F be a field and f and g two non-constant polynomials in $F[X,Y]$ having no non-constant common factor in $F[X,Y]$. Then $|Z_F(f) \cap Z_F(g)| \leq \deg f \deg g$.

Our argument does not show this or give any bounds.

The proposition, in some sense, reduces our study of affine plane curves over a field to $Z_F(f)$ with f an irreducible polynomial in $F[X,Y]$ as $Z_F(f) = Z_F(f^r)$ for any positive integer r.

If $f = f(X,Y)$ is a polynomial in two variables, we can view f as a polynomial in Y over $F[X]$ or its quotient field $F(X)$, say the degree in Y of f is n. If we evaluate X at x in F, then $f(x,Y)$ is a polynomial of degree at most n. We want to show if F is algebraically closed of characteristic zero, that for almost all such x, the polynomial in the variable Y, $f(x,Y)$ is precisely of degree n and moreover has no multiple roots in F, where a is a *multiple root* of $f(x,Y)$ in F if $(Y-a)^2 \not| f(x,Y)$ in $F[Y]$ for any root a of $f(x,Y)$ in F.

Definition 39.6. Let f be a polynomial in $F[X,Y]$ of degree $\deg_Y f$ of f in the variable Y satisfying $\deg_Y f = n > 0$. An element x in F is called a *regular value of* f if $f(x,Y)$ has n distinct roots in F.

Remark 39.7. Let F be a field of characteristic zero and x a regular value of f in $F[X,Y]$ of degree n in Y. We can write $f = \sum_{i=0}^{n} q_i(X)Y^i$ with $q_i(X)$ polynomials in $F[X]$ and $q_n(X)$ nonzero. As x is a regular value, $q_n(x)$ is nonzero and $f(x,Y)$ has n distinct roots. In particular none of the roots of $f(x,Y)$ are roots of the formal derivative $\frac{\partial f}{\partial Y}(x,Y)$ by Exercise 31.19(3ii) or by Exercise 39.10(4).

Proposition 39.8. *Let F be an algebraically closed field of characteristic zero, e.g., $F = \mathbb{C}$. Suppose f is an irreducible polynomial in $F[X,Y]$ with the degree $\deg_Y f$ of f in the variable Y satisfying $\deg_Y f = n > 0$. Then*

1. *For each a in F, there exists at most n elements b in F satisfying $(a,b) \in Z_F(f)$.*
2. *There exists a finite subset Δ of F with the property that for any element a in $F \setminus \Delta$, there exist exactly n elements b in F satisfying (a,b) lies in $Z_F(f)$.*

Question 39.9. What is $Z_F(f)$ if f in the proposition is irreducible but $\deg_Y f = 0$?

PROOF. Write $f = \sum_{i=0}^{n} q_i(X)Y^i$ with $q_i(X)$ polynomials in $F[X]$, $q_n(X)$ nonzero for some $n > 0$. If a lies in F, we have $f(a, Y) = \sum_{i=0}^{n} q_i(a)Y^i$ has at most n roots in F unless $q_i(a) = 0$ for all i. But $X - a \mid q_i(X)$ in $F[X]$ for all i if and only if $X - a \mid f$ in $F[X, Y]$. As f is irreducible and $f \notin F[X]$, this possibility cannot occur. The remaining problem is when $f(a, Y)$ has a multiple root in $F[Y]$. Let

$$\Delta = \{a \in F \mid q_n(a) = 0 \text{ or } f(a, Y) \text{ has a multiple root in } F\}.$$

If $a \in \Delta$, then $f(a, Y)$ has at most n roots in F, as $f(a, Y)$ is not zero and if $a \in F \setminus \Delta$ then $f(a, Y)$ is a product of n linear polynomials in $F[Y]$ as F is algebraically closed, i.e., has precisely n distinct roots. So it suffices to show that Δ is finite. Certainly, there exist finitely many a in F with $q_n(a) = 0$, so it suffices to show the following:

Claim. There exist finitely many a in F with the nonzero polynomial $f(a, Y)$ having a multiple root.

As F is algebraically closed, Remark 39.7 says that $f(a, Y)$ has a multiple root if and only if $f(a, Y)$ and $\partial f(a, Y)/\partial Y$ have a common root in F.

We know that

(*i*) $\deg_Y \partial f(X, Y)/\partial Y < \deg_Y f(X, Y)$.

(*ii*) $f(X, Y)$ is irreducible in $F[X, Y]$ (as $f(X, Y) \in F(X)[Y]$ is irreducible by Lemma 32.7, where $F(X)$ is the quotient field of $F[X]$).

(*iii*) $\partial f(X, Y)/\partial Y$ is not zero (as $0 < \deg_Y f(X, Y)$ and the characteristic of F is zero (Why?)).

These conditions mean that f and $\partial f/\partial Y$ have no common factors in $F[X, Y]$. Consequently, by Proposition 39.3, we have $Z_F(f) \cap Z_F(\partial f/\partial Y)$ is a finite set. This proves the claim. □

If $F = \mathbb{C}$ in the above, then the affine plane curve $Z_{\mathbb{C}}(f)$ is called a *Riemann surface*. It is called a surface as $\dim_{\mathbb{R}} \mathbb{C} = 2$ and it is viewed as a surface over $\mathbb{R}$.

Using a little bit of analysis (e.g., the Implicit Function Theorem), one can show that if f is an irreducible polynomial in $\mathbb{C}[X, Y]$ with $\deg_Y f > 0$, then $Z_{\mathbb{C}}(f)$ has the following properties: Let Δ be as in Proposition 39.9 and $Z_{\mathbb{C}}(f)$ the induced topology in $\mathbb{C}^2$. Then there exists a continuous map $\pi : Z_{\mathbb{C}}(f) \to Z_{\mathbb{C}}(Y)$ (so $Z_C(Y)$ is the X-axis in $\mathbb{C}^2$, i.e., the real plane defined by $Y = 0$ in $\mathbb{C}^2$) satisfying

1. $|\pi^{-1}(x)| = n$ for all x in $Z_{\mathbb{C}}(Y) \setminus \Delta$.

2. For every x_0 in $Z_F(Y) \setminus \Delta$, there exists an open neighborhood U of x such that $\pi^{-1}(U)$ is a union of n disconnected sets V_i (so $\pi^{-1}(U) = V_1 \cup \cdots \cup V_n$) with each $V_i \subset Z_{\mathbb{C}}(f)$ open and $\pi|_{V_i} : V_i \to U$ a homeomorphism.

We say that $Z_{\mathbb{C}}(f)$ is an *n-sheeted branched covering* of $Z_F(Y)$.

Exercises 39.10. 1. Prove Remark 39.4.

2. Show if F is a field of characteristic zero and f a non-constant polynomial in $F[t]$, the its derivative f' is never zero. Show that if the characteristic of F is not zero, this may not be true by producing counterexamples.
3. Let F be a field of characteristic zero and f a polynomial in $F[t]$. Let $f = (t-a)^r g$ in $F[t]$ for some positive integer n and polynomial g in $F[t[$ with $g(a)$ nonzero. Show that a is a multiple root of f, i.e., $r > 1$ if and only if a is a root of f'.
4. Show that the Proposition 39.8 still is valid, if we weaken the condition that f be irreducible, to f is irreducible as a polynomial in X.

CHAPTER X

Finitely Generated Modules Over a PID

This is the deepest, and most difficult chapter in this part of the text. We are interested in classifying finitely generated modules over a PID which we succeed in doing. This has applications to abelian group theory and finite dimension vector spaces.

We begin by discussing row reduction for matrices over a ring. In particular, if the ring is a PID, we shall show that there is a unique matrix up to equivalence of matrices called the *Smith Normal Form* of the matrix.. If R is a euclidean domain, we give an algorithm for computing this matrix in §104. We shall show that any finitely generated module over a PID decomposes into a direct sum of two pieces, a free submodule and a *torsion* submodule (a module in which for every element m there exists a nonzero ring element r satisfying $rm = 0$). The torsion submodule also decomposes nicely (in two different ways). There is also a uniqueness statement for each of the two forms. For finitely generated abelian groups, this says that, up to isomorphism, every finitely generated abelian group is isomorphic to $\mathbb{Z}^r \times G$ with r unique and G a unique group that is a product of finite cyclic groups, each of which decomposes by the Chinese Reminder Theorem.

The theory is also applied to linear algebra. If V is a finite dimensional vector space over F and T a linear operator on V, then V becomes and $F[t]$-module by t acting on V by T. This leads to canonical forms of matrices over a field F. If F is arbitrary, the desired canonical form is called Rational canonical form.. If the field is algebraically closed, the desired canonical form is called Jordan canonical form. Further results in linear algebra are also proven, e.g., the Cayley-Hamilton Theorem.

40. Smith Normal Form

If R is a ring, two $m \times n$ matrices A and B in $R^{m\times n}$ are called *equivalent* if there exist invertible matrices $P \in \mathrm{GL}_m(R)$ and $Q \in \mathrm{GL}_n(R)$ such that $B = PAQ$. Compare this to Change of Basis Theorem in linear algebra of matrix representations of linear transformations when R is a field (or division ring). (Cf. Appendix §103.) Certainly, this

relation is an equivalence relation, so one would like to find a good system of representatives. If $m = n$, then we are usually interested in the equivalence relation given by similarity of matrices, i.e., when $Q = P^{-1}$. This last equivalence relation is especially important in linear algebra, i.e., when F is a field. (Cf. this to the Change of Basis Theorem for the matrix representation of a linear operator in one basis to another in Appendix §103.) For a general field this set of representatives will turn out to be the set of matrices in rational canonical form and over an algebraically closed field this set of representatives will turn out to be the set of Jordan canonical forms. In this section, we determine a set of representatives for equivalence of matrices over a PID. As this includes $F[t]$ when F is a field, we shall see that it will aid us solving the equivalence of similarity over matrices. The system of representatives for equivalence of matices will be the set of matrices in Smith Normal Form. In Appendix §104, we will give an algorithm for compute this form of any matrix when R is a euclidean ring. In this section, we do the more general case that R is a PID. Unfortunately, unlike the euclidean case, there is no algorithm to compute it.

Definition 40.1. Let R be a domain and $A \in R^{m\times n}$. We say that A is in *Smith Normal Form* if A is a diagonal matrix of the form

$$A = \operatorname{diag}(q_1, \dots, q_r, 0, \dots, 0) \text{ with } q_1 \mid q_2 \mid q_3 \mid \cdots \mid q_r \text{ in } R \text{ and } q_r \neq 0.$$

[Note that the number of diagonal entries is the minimum of n and m.]

We want to prove that every matrix in $R^{m\times n}$ is equivalent to a matrix in Smith Normal Form if R is a PID. If R is a euclidean domain, using the euclidean function provides a method to induct on the value of remainders in the division algorithm. As mentioned above, this leads to a computable algorithm, one used by computer programs to find the Jordan canonical form of square matrices over the complex numbers. We need a substitute for the euclidean function on a euclidean domain. As any PID is a UFD, we can induct on the number of irreducible factors (counted with multiplicity) for any nonzero nonunit.

Definition 40.2. Let R be a UFD and a a nonzero nonunit in R. Let $a = p_1 \cdots p_r$ be a factorization into irreducibles in R (not necessarily distinct). Define the *length*, $l(a)$ of a to be r. If a is a unit, define its length, $l(a)$ to be zero.

To prove our result, we need the following lemma. It requires R to be a PID.

Lemma 40.3. *Let R be a PID and $A = (a_{ij})$ a nonzero $m \times n$ matrix in $R^{m\times n}$. Fix i, j with $a_{ij} \neq 0$ and $j < n$. Let d in the ideal $(a_{ij}, a_{i,j+1})$ in R satisfy $(d) = (a_{ij}, a_{i,j+1})$. Then we have:*

(1) If $a_{ij} \nmid d$, then $l(d) < l(a_{ij})$. In particular, this is true if $a_{ij} \nmid a_{i,j+1}$.

(2) There exists a matrix B in $\mathrm{SL}_n(R)$ satisfying $AB = (c_{lk})$ with $c_{ij} = d$ and $c_{i,j+1} = 0$.

Remark 40.4. An analogous statement holds for two column entries in A, but in (2), the elment B in $\mathrm{SL}_n(R)$ multiplying A on the left. More generally, a similar statement holds for any two fixed elements in A, at least one not zero, in any fixed row or any fixed column. We prove the lemma as it is stated for convenience of notation.

Proof. (of the lemma). That d satisfies the first condition is immediate. We construct B in (2). For notational convenience, write $v = a_{ij}$ and $w = a_{i,j+1}$, so $(d) = (v, w)$. Then $0 \neq d = xv + yw = (xa + yb)d$ in the domain R, so $1 = xa + yb$ in R, hence

$$P = \begin{pmatrix} x & -b \\ y & a \end{pmatrix} \text{ lies in } \mathrm{SL}_2(R) \text{ [with inverse } P^{-1} = \begin{pmatrix} a & b \\ -y & x \end{pmatrix}].$$

Then

$$\begin{pmatrix} v & w \end{pmatrix} P = \begin{pmatrix} vx + wy & -bv + wa \end{pmatrix} = \begin{pmatrix} d & -bad + bda \end{pmatrix} = \begin{pmatrix} d & 0 \end{pmatrix},$$

so

$$B = \begin{pmatrix} 1 & \cdots & & & & & & 0 \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ \vdots & & & P & & & & \vdots \\ & & & & & & & \\ & & & & 1 & & & \\ & & & & & \ddots & & \\ 0 & \cdots & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ \\ j \\ j+1 \\ \\ \\ \\ \end{matrix}$$
$$\begin{matrix} & & & i & i+1 & & & \end{matrix}$$

works. □

Theorem 40.5. *Let R be a PID and A an $m \times n$ matrix in $R^{m\times n}$. Then A is equivalent to a matrix in Smith Normal Form.*

Proof. We may assume that the matrix $A = (a_{ij})$ is nonzero.

We use elementary row and column operations on matrices. We call an elementary row (respectively, column) operation *Type I* if we add

a multiple of one row (respectively, one column) to another and *Type II* if we permute two rows (respectively, columns). These correspond to multiplying on the right or left by invertible matrices. (For more details, see Appendix §104.) In particular, these elementary row and column operations yield equivalent matrices. Using Type II elementary row and column operations, we may assume that the $(1,1)$ entry in A is a nonzero element a in R such that among all nonzero entries in A, we have $l(a)$ is minimal.

Suppose that a is a unit. Then using Type I elementary row and column operations, we can convert A to the equivalent matrix

$$(*) \qquad \begin{pmatrix} a & \cdots & 0 \\ 0 & & \\ \vdots & A_1 & \\ 0 & & \end{pmatrix}$$

with A_1 an $(m-1) \times (n-1)$ matrix.

By induction on matrix size, there exist appropriate invertible matrices P_1 and Q_1 such that $P_1 A_1 Q_1$ is in Smith Normal Form. Then

$$\begin{pmatrix} 1 & \cdots & 0 \\ 0 & & \\ \vdots & P_1 & \\ 0 & & \end{pmatrix} \begin{pmatrix} a & \cdots & 0 \\ 0 & & \\ \vdots & A_1 & \\ 0 & & \end{pmatrix} \begin{pmatrix} 1 & \cdots & 0 \\ 0 & & \\ \vdots & Q_1 & \\ 0 & & \end{pmatrix}$$

is in Smith Normal Form and is equivalent to A. This proves the result in the case that a is a unit. Note that if A is equivalent to a matrix in the form of (*) such that a divides every entry in A_1, the argument above shows that we would also be done.

So we may assume that $l(a) > 0$. If $a \nmid a_{ij}$ for some $i = 1$ or $j = 1$, using Lemma 40.3, we can convert A into an equivalent matrix

$$\begin{pmatrix} d & * \cdots & * \\ * & \cdots & * \\ \vdots & & \vdots \\ * & \cdots & * \end{pmatrix}$$

with $l(d) < l(a)$, and we are done by induction on $l(a)$. So we may assume that $a \mid a_{ij}$ if $i = 1$ or $j = 1$. Using Type I row and column operations converts A to a matrix equivalent to A of the form (*). By induction of matrix size, there exist invertible matrices of the appropriate size such that A is equivalent to a diagonal matrix $\operatorname{diag}(a, a_2, \ldots, a_r, 0, \ldots, 0)$ with $a_2 \mid a_3 \mid \cdots \mid a_r$ and $a_r \neq 0$ in R. If

$a \mid a_2$, we are done; if not use an Type I elementary row operation to convert this matrix to

$$\begin{pmatrix} a & a_2 & \cdots & & & & \\ 0 & a_2 & & & & & \\ 0 & 0 & a_3 & & & & \\ \vdots & & & \ddots & & & \\ & & & & a_r & & \\ & & & & & \ddots & \end{pmatrix}.$$

As $a \nmid a_2$, we can apply the Lemma 40.3 to produce an equivalent matrix with $(1,1)$ entry d satisfying $l(d) < l(a)$, and we are done by induction on $l(a)$. □

To prove that the Smith Normal Form of a matrix in $R^{m\times n}$ with R a PID is essentially unique, we shall need facts about the determinant that we shall use but shall not prove here. (However, cf. Section §102 for a sophisticated proof of the determinant and its properties.)

Let R be a commutative ring. Then the *determinant* $\det : \mathbb{M}_n(R) \to R$ is a map that has the following properties:

(1) det is *n-multilinear* (or *n-linear*) as a function of the rows (respectively, columns) of matrices in $\mathbb{M}_n(R)$. This means that it is R-linear (i.e., an R-homomorphism) in each of the n entries fixing the others, i.e.,

$$\begin{aligned}\det(\alpha_1, \ldots, r\alpha_i + \alpha_i', \ldots, \alpha_n) \\ = r\det(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n) + \det(\alpha_1, \ldots, \alpha_i', \ldots, \alpha_n)\end{aligned}$$

for all r in R, and rows (respectively columns) of matrices A in $\mathbb{M}_n(R)$.

(2) det is *alternating* as a function of the rows (respectively, columns) of matrices in $\mathbb{M}_n(R)$, i.e., if A has two identical rows (respectively, columns), then $\det A = 0$.

[Note: This implies that the matrix obtained by interchanging two rows (respectively columns) of A has determinant $-\det A$.

(3) $\det I = 1$

Indeed it can be shown that $\det : \mathbb{M}_n(R) \to R$ is the unique function satisfying (1), (2), and (3); and it is given by

$$\det(a_{ij}) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}$$

where

$$\operatorname{sgn}(\sigma) = \begin{cases} 1 & \text{if } \sigma \in A_n \\ -1 & \text{if } \sigma \notin A_n. \end{cases}$$

Let A be an $m \times n$ matrix in $R^{m\times n}$ and $1 \leq l \leq \min\{m, n\}$. An *$l$-order minor* of A is a determinant of an $l \times l$ submatrix of A, i.e., a matrix obtained from A be deleting $m-l$ rows and $n-l$ columns of A. The key to proving our uniqueness statement is the following lemma:

Lemma 40.6. *Suppose that R is a UFD and A an $m \times n$ matrix in $R^{m\times n}$. Let P and Q be invertible matrices in $\mathrm{GL}_m(R)$ and $\mathrm{GL}_n(R)$ respectively. Set $B = PAQ$. If $1 \leq l \leq \min\{m, n\}$ and*

(1) if the element a in R is a gcd of all the l-order minors of A, and

(2) if the element b in R a gcd of all the l-order minors of B,

then a and b are associates, i.e., the ideals (a) and (b) are the same.

Proof. Let $P = (p_{ij})$, $A = (a_{ij})$, $Q = (q_{ij})$ and c in R a gcd of all the l-order minors of PA. Then the kith entry of PA is $\sum_j p_{kj}a_{ji}$, so the kth row of PA is $\sum_j p_{kj}(a_{j1}\, a_{j2} \cdots a_{jn})$ (with the obvious notation). As the determinant is multilinear as a function of the rows, we have $a \mid c$ in R. As the determinant is multilinear as a function of the columns, an analogous argument shows that $c \mid b$ in R, hence $a \mid b$ in R. But we also have $A = P^{-1}BQ^{-1}$, so arguing in the same way, we conclude that we also have $b \mid a$. Since R is a domain, a and b must be associates. □

Corollary 40.7. *Suppose that R is a PID and A is an $m\times n$ matrix in $R^{m\times n}$ with $B = \operatorname{diag}(d_1, \ldots, d_r, 0 \ldots, 0)$ in $R^{m\times n}$ satisfying $d_1 \mid \cdots \mid d_r$ and $d_r \neq 0$ in R a Smith Normal Form of A. Let*

$$\Delta_l \text{ be a gcd of all the } l\text{-order minors of } A \text{ in } R \text{ for } 1 \leq l \leq r$$

and $\Delta_0 = 1$. Then

$$\Delta_0 \mid \Delta_1 \mid \cdots \mid \Delta_r \quad \text{and} \quad d_l \approx \frac{\Delta_l}{\Delta_{l-1}} \text{ in } R \text{ for all } l > 0.$$

Putting this all together, we obtain the following theorem:

Theorem 40.8. *Let R be a PID and A an $m \times n$ matrix in $R^{m\times n}$. Then A is equivalent to a matrix in Smith Normal Form. Moreover, if $\operatorname{diag}(a_1, \ldots, a_r, 0, \ldots, 0)$ and $\operatorname{diag}(b_1, \ldots, b_s, 0, \ldots, 0)$ are two Smith Normal Forms for A, then $r = s$ and $a_i \approx b_i$ for $1 \leq i \leq r$. In particular, the descending sequence of ideals in R*

$$(a_1) \supset (a_2) \supset \cdots \supset (a_r)$$

completely determine a Smith Normal Form of A with any generator of (a_l) for $1 \leq l \leq r$ being an associate of Δ_l/Δ_{l-1}, where Δ_l is a gcd of all the l-order minors of A in R and $\Delta_0 = 1$.

The elements $a_1 \mid \cdots \mid a_r$ in the theorem are called the *invariants factors* of A. They are unique up to units. If $g : R^m \to R^n$ is an R-homomorphism, a matrix representation $[g]_{\mathcal{B},\mathbb{C}}$ of g in Smith Normal Form will be called *Smith Normal Form* of g and the invariant factors of $[g]_{\mathcal{B},\mathbb{C}}$ will be called the *invariant factors* of g.

Exercise 40.9. Prove Remark 40.4.

41. The Fundamental Theorem

In this section, we completely determine finitely generated modules over a PID up to an isomorphism. The basic idea is to first show, unlike in the general case, that any submodule of a finitely generated free module over a PID is itself free. This means that if we have a finitely generated R-module M with R a PID, that we have a short exact sequence

$$0 \to R^m \xrightarrow{g} R^n \to M \to 0.$$

Therefore, M is isomorphic to $\operatorname{coker} g$. As g is a map of free R-modules, it has a matrix representation A (cf. Appendix §103) relative to the appropriate standard bases. Looking at a Smith Normal Form of A produces a direct sum decomposition for $\operatorname{coker} g$, hence M decomposes into a direct sum of R-cyclic modules. The uniqueness of the Smith Normal Form will show this decomposition is unique up to isomorphism.

We begin by proving that submodules of finitely generated free modules over a PID are free.

Proposition 41.1. *Let R be a PID, N a free R-module of rank n, and M a submodule of N. Then M is a free R-module satisfying* $\operatorname{rank} M \leq \operatorname{rank} N = n$.

Proof. We know that $N \cong R^n$, so we may assume that M is a submodule of R^n; and must show that M is free of rank at most n. We show this by induction on n. If $n = 0$, this is trivial, so we may assume that $n \geq 1$. Let

$$\pi_1 : R^n \to R \text{ given by } (a_1, \ldots, a_n) \mapsto a_1$$

be the projection onto the first coordinate, an R-epimorphism with kernel

$$\ker \pi_1 = \{(0, a_2, \ldots, a_n) \mid a_i \in R\} \cong R^{n-1}$$

free of rank $n-1$. Let $\pi_1|_M : M \to R$ be the restriction. It is an R-homomorphism with $\ker \pi_1|_M \subset \ker \pi_1 \cong R^{n-1}$, so a free R-module

of rank at most $n-1$ by induction. For convenience, let $\pi_1' = \pi_1|_M$. In particular, we are done if π_1' is the zero map. Hence we may assume that π_1' is not trivial, i.e., $0 < \operatorname{im}\pi_1' \subset R$ is a nonzero left ideal in the PID R. It follows that there exists a nonzero $a \in R$ satisfying $(a) = \operatorname{im}\pi_1'$. As $ra = 0$ in the domain R implies that $r = 0$, we conclude that $(a) \cong R$ as R-modules, so (a) is free of rank 1. Choose $m \in M$ such that $\pi_1'(m) = a \neq 0$. If $rm = 0$, then $0 = \pi_1'(rm) = ra$ in the domain R, so $r = 0$ and $Rm \cong R$ is a free R-module of rank one.

Claim. $M = Rm \oplus \ker\pi_1'$:

If we show the claim, we are done. Indeed if Rm is R-free of rank 1 and $\ker\pi_1'$ is R-free of rank at most $n-1$, then M is R-free of rank $\operatorname{rank} Rm + \operatorname{rank}\ker\pi_1' \leq n$ by Corollary 36.10, as needed. So we need only show the claim.

We first show $M = Rm + \ker\pi_1'$:

Let $x \in M$, so $\pi_1'(x) = ra$ for some r in R, hence lies in (a). It follows that $x - rm$ lies in $\ker\pi_1'$, hence x lies in $Rm + \ker\pi_1'$.

Next we show $Rm \cap \ker\pi_1' = \emptyset$:

Suppose that x lies in $Rm \cap \ker\pi_1'$. Then $x = rm$ for some $r \in R$ and $0 = \pi_1'(x) = \pi_1'(rm) = ra$ in the domain R. As a is nonzero, we have $r = 0$, so $x = 0$ as needed.

This establishes the claim and hence the proposition. □

As mentioned before, it is in fact true that any submodule of a free module over a PID is free, finitely generated or not. The fact that a free submodule of a finitely generated free module N has rank bounded by the rank of N is in fact true over any commutative ring, but this is harder (especially the case when the ring is not a domain).

The theorem has an interesting consequence even stronger than that mentioned in the introduction of this section.

Corollary 41.2. *Let R be a PID and M a finitely generated R module. Then there exists an exact sequence of R-modules*

$$0 \to R^m \xrightarrow{g} R^n \xrightarrow{f} M \to 0.$$

with $m \leq n$.

Proof. As M is generated by n elements, for some positive integer n, we have an exact sequence

$$0 \to \ker f \xrightarrow{inc} R^n \xrightarrow{f} M \to 0.$$

As $\ker f \subset R^n$, it is R-free of rank at most n by the proposition. □

If M is a finitely generated R module with R a PID, then the sequence

$$0 \to R^m \xrightarrow{g} R^n \xrightarrow{f} M \to 0.$$

says that M can be generated by n elements and the relations on these generators generates a free R-module of rank at most m, i.e., the minimal number of relations on the generators that generate all relations on these generators is at most m. It is called a *free resolution* of M. We shall also say that M has a *free presentation*. In particular, if n is the minimal number of generators of M, one can show that m is the minimal number of generating relations for M among all free resolutions of M. We can also say something about submodules of finitely generated modules over a PID.

Corollary 41.3. *Let R be a PID and M a finitely generated R-module. Suppose that M can be generated by n elements. Then any submodule of M can be generated by n elements.*

PROOF. Let

$$0 \to R^m \xrightarrow{g} R^n \xrightarrow{f} M \to 0$$

be a free resolution of M and N a submodule of M. By the Correspondence Principle, there exists a submodule B of R^n satisfying $\ker f \subset B \subset R^n$ and $f(B) = N$, i.e., $N \cong B/\ker f$. As B is R-free of rank at most n, it can be generated by $\leq n$ elements, hence the same is true of N. □

Remarks 41.4. 1. If R is a commutative noetherian ring but not a PID, then there exists a non-principal ideal $\mathfrak{A} = (x, y)$ of R. We have an exact sequence

$$0 \to \mathfrak{A} \to R \xrightarrow{-} R/\mathfrak{A} \to 0.$$

R is a cyclic R-module but $\mathfrak{A}$ is not. Nor is $\mathfrak{A}$ R-free as $x \cdot y - y \cdot x = 0$ in R for any generators of $\mathfrak{A}$. In fact, a similar argument shows that an ideal in a ring R is free as an R-module if and only if it is principal. This shows that the last three results do not hold for commutative rings that are not PIDs.

2. The proper $\mathbb{Z}$-free submodule $2\mathbb{Z}$ of $\mathbb{Z}$ is of rank 1, the same as $\mathbb{Z}$.
3. It is, in fact, true that if R is a commutative ring, M a finitely generated free R-module of rank n, then any submodule of M that is also R-free has rank at most n. The proof is not easy (although easier for the case when R is a domain).

Observation 41.5. Let N_i be a submodule of the R-module M_i for $i = 1, 2$. Then we have an R-isomorphism

$$(M_1 \coprod M_2)/(N_1 \coprod N_2) \cong (M_1/N_1) \coprod (M_2/N_2)$$

Theorem 41.6. (Fundamental Theorem of fg Modules over a PID, Form I) *Let R be a PID and M a nontrivial finitely generated R-module. Then M is a direct sum of cyclic R-modules. More precisely, there exist a nonnegative integer r and*

$$(*) \qquad m_1, \dots, m_r \in M \text{ satisfying } M = P \oplus Rm_1 \oplus \cdots \oplus Rm_r$$

with P a free R-module of rank s, some $s \geq 0$, and there exist nonzero nonunits

$$(**) \qquad d_i \in R \text{ satisfying } \begin{cases} \operatorname{ann}_R m_i = (d_i) & i = 1, \dots, r \\ d_1 \mid \cdots \mid d_r. \end{cases}$$

Moreover, r and s are unique in (*) and the d_i in (**) are unique up to units and the order determined by $d_1 \mid \cdots \mid d_r$, i.e., any direct sum decomposition of M satisfying (*) and (**) has the R-free submodule in (*) of the same rank s, the number of non-free R-cyclic modules the same r with the corresponding descending chain of ideals

$$(d_1) \supset (d_2) \supset \cdots \supset (d_r)$$

unique.

The elements $d_1, \dots, d_r$ in the Fundamental Theorem 41.6 are called *invariant factors* of M. Note also that the Rm_i in the Fundamental Theorem 41.6 are unique up to isomorphism with $Rm_i \cong R/(d_i)$. Of course, if $d_i \approx d_j$, then $Rm_i \cong Rm_j$. In particular,

$$M \cong R^s \coprod R/(d_1) \coprod \cdots \coprod R/(d_r),$$

and the right hand side completely determines the isomorphism class of M, i.e., the invariant factors and s completely determine M up to isomorphism.

Proof. **Existence**: We have shown that there exists an exact sequence

$$0 \to R^m \xrightarrow{g} R^n \to M \to 0 \tag{41.7}$$

of R-modules with $m \leq n$, i.e, $M \cong \operatorname{coker} g = R^n / \operatorname{im} g$. Let $\mathcal{S}_m$ and $\mathcal{S}_n$ be the standard bases for R^m and R^n, respectively. By Theorem 40.8 and the Change of Basis Theorem from linear algebra (cf. Theorem 103.7 in Appendix §103), there exist bases $\mathcal{B}$ and $\mathcal{C}$ for R^m and R^n, respectively, such that the $n \times m$ matrix $[g]_{\mathcal{B},\mathcal{C}}$ is a Smith Normal Form of $[g]_{\mathcal{S}_m,\mathcal{S}_n}$.

Let $\mathcal{C} = \{v_1, \dots, v_n\}$ and $[g]_{\mathcal{B},\mathcal{C}} = \operatorname{diag}(d_1, \dots, d_r, 0, \dots, 0)$ with $d_1 \mid \cdots \mid d_r$ in R, $r \le m \le n$, and d_r nonzero. As g is injective, no column of $[g]_{\mathcal{B},\mathcal{C}}$ can be zero. Therefore, $r = m$. We have

$$\operatorname{im} g = \bigoplus_{i=1}^{m} Rd_iv_i = \bigoplus_{i=1}^{m} Rd_iv_i \bigoplus \bigoplus_{i=m+1}^{n} 0.$$

let $d_i = 0$ for $m < i \le n$. As $R \to Rv_i$ by $r \mapsto rv_i$ is an R-isomorphism taking $(d_i) \to Rd_iv_i$ isomorphically, we have $R/(d_i) \cong Rv_i/Rd_iv_i$ for $i = 1, \dots, n$. Hence by Observation 41.5, we have

$$\begin{aligned} M \cong \operatorname{coker} g &\cong \Big(\coprod_{i=1}^{n} Rv_i\Big)/\Big(\coprod_{i=1}^{n} Rd_iv_i\Big) \\ &\cong \coprod_{i=1}^{m}(Rv_i/Rd_iv_i) \coprod \coprod_{i=m+1}^{n} Rv_i \cong \coprod_{i=1}^{m} R/(d_i) \coprod R^{n-m}. \end{aligned}$$

If d_i is a unit, then $R/(d_i) = 0$, so dropping such d_i, if

$$f : M \longrightarrow \coprod_{\substack{i=1 \\ d_i \notin R^\times}}^{m} R/(d_i) \coprod R^{n-m}$$

is the isomorphism, then f^{-1} gives the desired decomposition with $m_i = f^{-1}\big(1_R + (d_i)\big)$ for d_i not a unit or zero and $P = f^{-1}(R^{n-m})$.

For the uniqueness, we need further observations that we leave as exercises.

Observations 41.8. Let R be a PID, e and d nonzero elements in R, g a gcd of e and d, p a prime element in R (so p is nonzero and $R/(p)$ is a field), and N an R-module. Then the following are true:

(1) $dR^m \cong R^m$.

(2) $d\big(R/(e)\big) \cong R/(\frac{e}{g})$.

(3) $R/(p)$ acts on N/pN by

$$\big(r + (p)\big)(x + pN) := rx + pN \ [= r(x + pN)]$$

for all $r \in R$ and $x \in N$, i.e., the $R/(p)$-action and the R-action on N/pN are *compatible* (have the same effect). In particular, N/pN is a vector space over $R/(p)$.

(4) As R-modules and $R/(p)$-vector spaces, we have

$$\big(R/(d)\big)/p\big(R/(d)\big) \cong \begin{cases} R/(p) & \text{if } p \mid d \\ 0 & \text{if } p \nmid d. \end{cases}$$

Using these observations, we show:

Uniqueness: Suppose that

$$R/(d_1) \coprod \cdots \coprod R/(d_r) \coprod R^s \cong M \cong R/(d'_1) \coprod \cdots \coprod R/(d'_{r'}) \coprod R^{s'}$$

with $d_1 \mid \cdots \mid d_r$, the d_i nonzero nonunits and $d'_1 \mid \cdots \mid d'_{r'}$, the $d'_{i'}$ nonzero nonunits, for some integers r, r', s, s'.

We know that a submodule of a free R-module is R-free, hence an isomorphic copy of $R/(d)$ with d a nonzero nonunit in R cannot be a submodule of an R-free. In particular, if there are no d_i or no d'_i, then M is free and there are neither any d_i nor any d'_i. As R is commutative, $s = \operatorname{rank} M = s'$ by Theorem 36.9. So we may assume both $r \geq 1$ and $r' \geq 1$. As $d_r\big(R/(d_i)\big) = 0$ and $d'_{r'}\big(R/(d'_{i'})\big) = 0$ for all i and i'. By Observation 41.8(1), we have

$$R^s \cong d_r d'_{r'} R^s \cong d_r d'_{r'} M \cong d_r d'_{r'} R^{s'} \cong R^{s'}$$

is R-free, hence $s = s'$ by Theorem 36.9. Let $m \in M$ correspond to a generator of $R/(d'_{r'})$ under the given isomorphism. Then $Rd_r m \subset d_r M \cong R^s$ must also be R-free. As $d'_{r'} d_r m = 0$, we have $d_r m = 0$ (Why?), so $d_r \in \operatorname{ann}_R m = (d'_{r'})$. Similarly, $d'_{r'} \in (d_r)$, so $d_r \approx d'_{r'}$, i.e., $(d_r) = (d'_{r'})$.

Assume that $r \geq r' > 0$. Let p be a prime element (so nonzero) satisfying $p \mid d_1$, which exists as d_1 is a nonzero nonunit. Then by Observations 41.8(2) and (4) and Observation 41.5, we have $M/pM \cong \coprod_{i=1}^{r+s} R/(p)$ is an $R/(p)$-vector space, hence of rank $r+s$. Therefore, $\coprod_{i=1}^{r'} R/(d'_i)/p\big(R/(d'_i)\big) \coprod R^s/pR^s$ is free of rank $r+s$. But, by Observation 41.8(4), we also know that it is free of rank at most $r'+s$. It follows that $r = r'$. In particular, we are done if $p \approx d_r \approx d'_{r'}$. By Observations 41.8(1) and (2), we have

$$\coprod R/(\frac{d_i}{p}) \coprod R^s \cong pM \cong \coprod R/(\frac{d'_i}{p}) \coprod R^s.$$

By induction on the length $l(d_r)$ of d_r, we conclude that $\frac{d_i}{p} \approx \frac{d'_i}{p}$ for all i, hence $d_i \approx d'_i$ for all i. □

Note the proof of the uniqueness statement is similar to the corresponding proof of the Fundamental Theorem of Arithmetic.

It should also be noted, that to prove the existence in the theorem, we did not need to use the existence of a free presentation of M. Indeed the same proof works if we have an exact sequence

$$R^m \xrightarrow{g'} R^n \to M \to 0.$$

Since a PID is a Noetherian ring, we know such an exact sequence holds by Corollary 37.7. We discuss this, uniqueness, and its relationship to a free presentation in the following remarks.

Remark 41.9. The proof of Theorem 41.6 used the existence of Smith Normal Form, but not its uniqueness. Indeed if A is an $m \times n$-matrix over a PID with $m \leq n$, then the uniqueness of its Smith Normal Form follows with a proof essentially the same as that for the uniqueness in Theorem 41.6. It is also easy to do the general case using the next remark. In particular, we do not need the determinant theory used to prove the uniqueness of Smith Normal Form previously used. We leave this as an exercise.

Remark 41.10. The proof above of the existence part of the theorem used a free resolution

$$0 \to R^m \xrightarrow{g} R^n \to M \to 0$$

for M. We then added additional $d_i = 0$ for $i = m+1, \ldots, n$ in our computation of the cokernel of g (although it was not really needed). This can be interpreted in the following way: We view R^m as a subspace of R^n given by the first m basis elements in $\mathcal{S}_n$. We then extend the map g to the map $h : R^n \to R^n$ defined by sending the last $n - m$ standard basis elements of R^n to zero. We still have $\operatorname{coker} g = \operatorname{coker} h$. The Smith Normal form G of $[g]_{\mathcal{S}_m,\mathcal{S}_n}$ and the Smith Normal form H of $[h]_{\mathcal{S}_m,\mathcal{S}_n}$ differ only in that the latter has $n - m$ zero columns which gives the rank of the free part of M. The question arises, what if we start with an exact sequence

$$R^{m'} \xrightarrow{g'} R^{n'} \to M \to 0$$

and determine a Smith Normal Form G' for g'. Does this still determine M up to isomorphism? The answer is yes.

First we note that in the proof of the existence part of the Fundamental Theorem, that g was a monomorphism gave the rank of the free part of M as $n - m$. In fact, as shown in the proof, the fact that g is a monomorphism implies that a Smith Normal Form G of g has all its diagonal elements nonzero, say $d_1 \mid \cdots \mid d_m$ and that the nonunit d_i's, say $d_s, \ldots, d_m$ constitute a complete set of invariant factors for M. (Recall that the proof showed that if d_i is a unit, then it gives a zero direct summand for M.)

Let $h : R^n \to R^n$ be any R-homomorphism such that $\operatorname{coker} g = \operatorname{coker} h$. As a Smith Normal Form H of h determines $\operatorname{coker} h$ completely, by the uniqueness statement of the theorem, H must have the same nonzero nonunit diagonal entries as G, up to units, i.e., $d_s, \ldots, d_m$.

So the only difference with that of G is the number of zero columns and the number of unit diagonal entries. The free part of M must be the number of zero columns, so there must be $n-m$ zero diagonal entries in H thus determining the free part of M.

More generally, suppose that $g' : R^{m'} \to R^{n'}$ is an R-homomorphism also satisfying $\operatorname{coker} g' = \operatorname{coker} g$.

First suppose that $m' < n'$. Viewing $R^{m'} \subset R^{n'}$, extend a basis of $R^{m'}$ to $R^{n'}$, and extend the map g' to an R-homomorphism $h' : R^{n'} \to R^{n'}$ taking the added basis elements to zero. This does not change the cokernel of g'. (Note that in the matrix representation relative to this basis and any one for $R^{n'}$, we have added $n'-m'$ zero columns.) As the nonzero nonunit diagonal entries of a Smith Norm Form H' of h' must be $d_s, \dots, d_m$ up to units, with the free part determined by the number of zero columns (which must be $n-m$), the matrix H' determines M up to isomorphism.

If $m' > n'$, then a Smith Normal Form G' of g' has only n' diagonal entries and these determine M just as above, and hence the nonzero nonunit ones must be $d_s, \dots, d_m$ up to units by uniqueness. Therefore, in general, we have: If

$$R^m \xrightarrow{g'} R^n \to M \to 0 \tag{*}$$

is an exact sequence and G' is a Smith Normal Form for g', say G' has r nonzero diagonal entries, then a complete list of invariant factors of M is the nonzero nonunit diagonal entries of G' and the free part of M has rank $n'-r$. Note also, we have shown that if G' has no zero columns, then g' must be monic, and the sequence (*) is a free resolution of M.

Examples 41.11. 1. We determine the abelian group M, up to isomorphism, generated by x_1, x_2, x_3, x_4 subject to the relations

$$\begin{aligned} 2x_1 + 3x_2 + x_4 &= 0 \\ x_1 - 2x_2 + x_3 &= 0. \end{aligned}$$

It is easy to see that the following is an exact sequence:

$$\mathbb{Z}^2 \xrightarrow{g} \mathbb{Z}^4 \xrightarrow{h} M \to 0$$

$$\text{with } g = \begin{pmatrix} 2 & 1 \\ 3 & -2 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } h \text{ takes } e_i \mapsto x_i,\ 1 \le i \le 4,$$

i.e., $M \cong \operatorname{coker} g$. To find the isomorphism type of M we find a Smith Normal Form of g. Using the algorithm in Appendix §104

(essentially the division algorithm), we row and column reduce g to

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Hence by Remark 41.10, M is isomorphic to $\mathbb{Z}/\mathbb{Z} \coprod \mathbb{Z}/\mathbb{Z} \coprod \mathbb{Z} \coprod \mathbb{Z} \cong \mathbb{Z}^2$. We also conclude that

$$0 \to \mathbb{Z}^2 \xrightarrow{g} \mathbb{Z}^4 \xrightarrow{h} M \to 0$$

is a free resolution of M.

2. We determine the abelian group M, up to isomorphism, generated by x_1, x_2, x_3, x_4 subject to the relations

$$\begin{aligned} 4x_1 + 2x_2 + 4x_3 + 3x_4 &= 0 \\ 2x_1 + 2x_2 - 2x_3 + 2x_4 &= 0 \\ -6x_1 + 6x_3 - 6x_4 &= 0 \end{aligned}$$

We have an exact sequence

$$\mathbb{Z}^3 \xrightarrow{g} \mathbb{Z}^4 \xrightarrow{h} M \to 0$$

$$\text{with } g = \begin{pmatrix} 4 & 2 & -6 \\ 2 & 2 & 0 \\ 4 & -2 & 6 \\ 3 & 2 & -6 \end{pmatrix} \text{ and } h \text{ takes } e_i \mapsto x_i,\ 1 \le i \le 4,$$

Row and column reducing g using the division algorithm as in Appendix §104, we see that a Smith Normal Form for g is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{pmatrix} \text{ and } h \text{ takes } e_i \mapsto x_i,\ 1 \le i \le 4,$$

Hence, by Remark 41.10, M is isomorphic to

$$\mathbb{Z}/\mathbb{Z} \coprod \mathbb{Z}/2\mathbb{Z} \coprod \mathbb{Z}/6\mathbb{Z} \coprod \mathbb{Z} \cong \mathbb{Z}/2\mathbb{Z} \coprod \mathbb{Z}/6\mathbb{Z} \coprod \mathbb{Z}.$$

Note that this is also isomorphic to $\mathbb{Z}/2\mathbb{Z} \coprod \mathbb{Z}/2\mathbb{Z} \coprod \mathbb{Z}/3\mathbb{Z} \coprod \mathbb{Z}$. This last isomorphism is what the second form of the Fundamental Theorem will give. We also conclude that

$$0 \to \mathbb{Z}^3 \xrightarrow{g} \mathbb{Z}^4 \xrightarrow{h} M \to 0$$

is a free resolution of M.

We next look at a property of modules that is, in general, weaker than being free but for finitely generated modules over a PID turn out to be equivalent.

Definition 41.12. Let R be a domain an M and R-module. An element m in M is called an (R-) *torsion* element if there exists a nonzero element r in R satisfying $rm = 0$, i.e., $\operatorname{ann}_R m > 0$. Set

$$M_t := \{m \in M \mid m \text{ is an } R\text{-torsion element}\}.$$

We say that M is a *torsion* R-module if $M = M_t$ and a *torsion-free* R-module if $M_t = 0$.

The proof of the following is straight-forward and left as an exercise.

Properties 41.13. Let R be a domain, M an R-module, and m a nonzero element in M. Then

(1) The element m is not an (R-) torsion element if and only if Rm is torsion-free if and only if Rm is R-free.
(2) M_t is a submodule of M.
(3) M/M_t is a torsion-free R-module.
(4) If M is R-free, then it is R-torsion-free.

Remarks 41.14. 1. $\mathbb{Q}$ is $\mathbb{Z}$-torsion-free but not $\mathbb{Z}$-free.

2. Any finite abelian group is a $\mathbb{Z}$-torsion module.

We shall show that over a PID finitely generated torsion-free modules are free. The following computation together with the Fundamental Theorem 41.6 will ensure this.

Example 41.15. Let R be a PID and $d_1 \mid \cdots \mid d_r$ nonzero nonunits in R. Set

$$\begin{aligned} N &= R/(d_1) \coprod \cdots \coprod R/(d_r) \coprod R^s \quad \text{and} \\ N_0 &= R/(d_1) \coprod \cdots \coprod R/(d_r), \end{aligned}$$

so $N = N_0 \coprod R^s$. We have

$$d_r N = d_r N_0 \coprod d_r R^s = d_r R^s \cong R^s$$

is R-free of rank s and $N_0 \subset N_t$. Next, suppose that $(n_0, v) \in N_t$ with $n_0 \in N_0$ and $v \in R^s$. Then there exists a nonzero element a in R satisfying $a(n_0, v) = 0$, i.e., $an_0 = 0$ and $av = 0$. As R^s is R-free, it is R-torsion-free, so $v = 0$ and we conclude that $N_0 = N_t$. As the

diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & N_0 & \longrightarrow & N & \longrightarrow & R^s & \longrightarrow & 0 \\ & & \| & & \| & & \uparrow & & \\ 0 & \longrightarrow & N_t & \longrightarrow & N & \longrightarrow & N/N_t & \longrightarrow & 0. \end{array}$$

has exact rows and is commutative, the vertical right hand arrow exists, (why?), hence is an isomorphism, so $N/N_t \cong R^s$. In particular, N is R-free if and only if $N_0 = 0$ if and only if $N_t = 0$ if and only if N is R-torsion-free.

The example together with the Fundamental Theorem yields:

Corollary 41.16. *Let R be a PID and M a finitely generated R-module. Then*

(1) M is R-torsion-free if and only if M is R-free.
(2) $M = M_t \oplus P$ for some finitely generated free R-module $P \subset M$ with $\operatorname{rank} P$ *unique.*

Remark 41.17. $\mathbb{Q}$ is a torsion-free abelian group but not $\mathbb{Z}$-free, so the assumption in the corollary that the module be finitely generated torsion-free is necessary in general to conclude that it is free. (The abelian group $\prod_{i=1}^{\infty} \mathbb{Z}$ is another example of a torsion-free abelian group that is not $\mathbb{Z}$-free.)

We turn to establishing an alternate form of the Fundamental Theorem. We must do some preliminary work. We begin with another example.

Example 41.18. Let R be a PID and M a nontrivial cyclic R-module. Then $M = Rm \cong R/(d)$, for some $m \in M$ and unique $(d) = \operatorname{ann}_R m < R$. Assume that M is not R-free, then $(d) > 0$ and d is not a unit. Let $d = p_1^{e_1} \cdots p_r^{e_r}$ be a factorization, with $p_1, \ldots, p_r$ non-associative irreducible, hence prime, elements and $e_1, \ldots, e_r$ positive integers. By the Chinese Remainder Theorem, $R/(d) \cong R/(p_1^{e_1}) \times \cdots \times R/(p_r^{e_r})$ as rings. It follows that $R/(d) \cong R/(p_1^{e_1}) \amalg \cdots \amalg R/(p_r^{e_r})$ as R-modules, as the R-action on $R/(a)$ is compatible with the $R/(a)$-action for any $a \in R$. (Cf. Observation 41.8(3).) Therefore, there exist $m_1, \ldots, m_r$ in M satisfying $M = Rm_1 \oplus \cdots \oplus Rm_r$ with $\operatorname{ann}_R m_i = (p_i^{e_i})$ for $i = 1, \ldots, r$.

Although this example suffices to get our alternate form, we first make some further comments. If R is a domain, p an irreducible element in R, and M an R-module, we say that M is *p-primary* if for all x in M, there exists a positive integer n (depending on x) satisfying $p^n x = 0$. If

R is Noetherian, let $\mathcal{P}$ denote a system of representatives of irreducible elements in R under the equivalence relation $\approx$.

Example 41.19. Let R be a Noetherian domain and $p \in \mathcal{P}$. So $R/(p^r)$ is a p-primary R-module for all positive integers r.

Theorem 41.20. (Primary Decomposition Theorem) *Let R be a PID and M a torsion R-module. If $p \in \mathcal{P}$, let*

$$M_p := \{x \in M \mid p^r x = 0 \text{ for some positive integer } r\}.$$

Then M_p is a p-primary submodule of M and $M = \bigoplus_{\mathcal{P}} M_p$. Moreover, if M is finitely generated, then $M_p = 0$ for almost all p in $\mathcal{P}$.

PROOF. Clearly, M_p is a p-primary submodule of M for all $p \in \mathcal{P}$. We may assume that M is not trivial.

$M = \sum_{\mathcal{P}} M_p$: Let x be a nonzero element of M and $(d) = \operatorname{ann}_R x$. Then d is a nonzero nonunit as M is a torsion R-module. It follows by Example 41.18 that x lies in $\sum_{\mathcal{P}} M_p$,

$M = \bigoplus_{\mathcal{P}} M_p$: Suppose that $x \in M_{p_0} \cap \sum_{p_0 \neq p \in \mathcal{P}} M_p$. Let $(d) = \operatorname{ann}_R x$ with d a nonzero nonunit. We have $p_0^{e_0} x = 0$ for some positive integer e_0, as $x \in M_{p_0}$, so $d \mid p_0^{e_0}$. As $x \in \sum_{p_0 \neq p \in \mathcal{P}} M_p$, there exist $p_1, \ldots, p_n$ in $\mathcal{P}$ and positive integers $e_1, \ldots, e_n$ satisfying $x = x_1 + \cdots + x_n$ with $p_i^{e_i} x_i = 0$ for $i = 1, \ldots, n$. It follows that $p_1^{e_1} \cdots p_n^{e_n} x = 0$, so $d \mid p_1^{e_1} \cdots p_n^{e_n}$. By choice p and $p_1^{e_1} \cdots p_n^{e_n}$ are relatively prime. It follows that $x = 0$.

Finally, suppose that M is finitely generated, say $M = Rx_1 + \cdots + Rx_n$ with x_i nonzero for $i = 1, \ldots, n$. Let d_i be a nonzero nonunit satisfying $(d_i) = \operatorname{ann}_R x_i$ for $i = 1, \ldots, n$ and set $d = d_1 \cdots d_n$ a nonzero element in the domain R. It follows that $M_p = 0$ for all $p \in \mathcal{P}$ satisfying $p \nmid d$. Hence $M = \bigoplus_{\substack{p \in \mathcal{P} \\ p \mid d}} M_p$, a finite sum. □

Example 41.21. Suppose that R is a PID and M is a nontrivial cyclic R-module with $\operatorname{ann}_R x = (d)$ and $d = p_1^{e_1} \cdots p_r^{e_n}$, with $p_i \in \mathcal{P}$. Then $M = \bigoplus_{i=1}^n M_{p_i}$ and $M_{p_i} \cong R/(p_i^{e_i})$ for $i = 1, \ldots, n$.

Using this, together with the Fundamental Theorem 41.6, we can reformulate it as

Theorem 41.22. (Fundamental Theorem of fg Modules over a PID, Form II) *Let R be a PID and M a nontrivial finitely generated R-module. Then there exists a finitely generated free submodule P of M*

of unique rank and $p_i \in \mathcal{P}$, $i = 1, \ldots, r$ *some* r *and* x_{ij} *in* M, $1 \leq j \leq n_i$ *and* $1 \leq i \leq r$ *and for some* n_i, *satisfying*

$$M = P \oplus (\bigoplus_{i=1}^{r} \bigoplus_{j=1}^{n_i} Rm_{ij}) \quad \textit{satisfying}$$

and for all i *and* j, *we have*

(1) $Rm_{ij} \cong R/(p_i^{e_{ij}})$ *and*

(2) $1 \leq e_{i1} \leq e_{i2} \leq \cdots \leq e_{in_i}$.

Moreover, this decomposition is unique relative the ordering of the p_i *subject to the order of the* e_{ij} *above and the uniqueness of the rank of* P.

Generators of the ideals $(p_i^{e_{ij}})$ in the theorem are called *elementary divisors* of M. They are unique up to units, and together with the rank of P completely determine M up to isomorphism. The condition on the e_{ij}, of course, orders the set of cyclic modules that are p_i-primary submodules in the decomposition (which are unique up to isomorphism). This data completely determines the isomorphism class of M.

Of course, both forms of the Fundamental Theorem apply to finitely generated abelian groups, although in this case an easier proof will suffice. If we are given any finitely generated abelian group defined by generators and relations, the invariant factors will determine completely the group up to isomorphism. Moreover, given a positive integer n, up to isomorphism, we can write down every abelian group of order n (assuming that we can factor n). We give an example how this works.

Example 41.23. Find all abelian groups up to isomorphism of order $2^3 \cdot 3^2 \cdot 5$. We write these groups in both forms. Given the factorization, it is easier to write down the representatives of isomorphism classes in Form II. To get Form I, we take the biggest p-primary pieces for each prime p and combine them to get one term, then continue with what is left. This is illustrated in our example by the overlines and underlines of various p-primary pieces and what they become in Form I.

$$\mathbb{Z}/2\mathbb{Z} \times \overline{\mathbb{Z}/2\mathbb{Z}} \times \underline{\mathbb{Z}/2\mathbb{Z}} \times \overline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/5\mathbb{Z}} \cong \mathbb{Z}/2\mathbb{Z} \times \overline{\mathbb{Z}/6\mathbb{Z}} \times \underline{\mathbb{Z}/30\mathbb{Z}}$$

$$\mathbb{Z}/2\mathbb{Z} \times \overline{\mathbb{Z}/2\mathbb{Z}} \times \underline{\mathbb{Z}/2\mathbb{Z}} \times \underline{\mathbb{Z}/9\mathbb{Z}} \times \underline{\mathbb{Z}/5\mathbb{Z}} \cong \mathbb{Z}/2\mathbb{Z} \times \overline{\mathbb{Z}/2\mathbb{Z}} \underline{\times \mathbb{Z}/90\mathbb{Z}}$$

$$\overline{\mathbb{Z}/2\mathbb{Z}} \times \underline{\mathbb{Z}/4\mathbb{Z}} \times \times \overline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/5\mathbb{Z}} \cong \overline{\mathbb{Z}/6\mathbb{Z}} \times \underline{\mathbb{Z}/60\mathbb{Z}}$$

$$\overline{\mathbb{Z}/2\mathbb{Z}} \times \underline{\mathbb{Z}/4\mathbb{Z}} \times \underline{\mathbb{Z}/9\mathbb{Z}} \times \underline{\mathbb{Z}/5\mathbb{Z}} \cong \overline{\mathbb{Z}/2\mathbb{Z}} \times \underline{\mathbb{Z}/180\mathbb{Z}}$$

$$\underline{\mathbb{Z}/8\mathbb{Z}} \times \overline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/5\mathbb{Z}} \cong \overline{\mathbb{Z}/3\mathbb{Z}} \times \underline{\mathbb{Z}/120\mathbb{Z}}$$

$$\underline{\mathbb{Z}/8\mathbb{Z}} \times \underline{\mathbb{Z}/9\mathbb{Z}} \times \underline{\mathbb{Z}/5\mathbb{Z}} \cong \underline{\mathbb{Z}/360\mathbb{Z}}.$$

[Note: Abelian direct sums of abelian groups are usually written as products as a finite product of abelian groups is the same as a direct sum of the same groups.]

Exercises 41.24. 1. Let R be a PID. Suppose that M is a finitely generated R-module and

$$0 \to R^m \xrightarrow{g} R^n \to M \to 0$$

is a free presentation of M with n minimal. Without using the Fundamental Theorem (or Smith Normal Forms) show if

$$0 \to R^{m'} \xrightarrow{g'} R^{n'} \to M \to 0$$

is another free presentation of M, then $m' \leq m$.

2. Prove the following generalization of Proposition 41.1 using Exercises 36.12(10), (11)): Let R be a commutative (this is not needed) ring with the property that every (left) ideal of R is R-projective. Let M be an R-module isomorphic to a submodule of R^n. Then M is a direct sum of $m \leq n$ submodules each isomorphic to a (left) ideal of R. In particular M is R-projective.
3. Prove Observation 41.5.
4. Prove all the statements in Observations 41.8.
5. Prove all the properties in 41.13.
6. Prove the assertions in Remark 41.9.
7. Let R be a domain. Show every ideal in R is R-torsion-free, but is free if and only if it is principal. In particular, show that R is a PID if every submodule of a free module is free. (The converse is also true.)
8. Let R be a domain and M a finitely generated R-module. Prove that M is isomorphic to a free submodule of finite rank.

9. Let $f : \mathbb{Z}^n \to \mathbb{Z}^m$ be a $\mathbb{Z}$-module homomorphism. Let $\mathcal{S}_l$ be the standard basis for $\mathbb{Z}^l$. Prove that f is monic if and only if the rank of $[f]_{\mathcal{S}^n,\mathcal{S}^m}$ is n and f is epic if and only if a gcd of the mth ordered minors of $[f]_{\mathcal{S}^n,\mathcal{S}^m}$ is 1.
10. Let R be a commutative ring. Let $\mathrm{E}_n(R)$ be the subgroup of $\mathrm{GL}_n(R)$ generated by all matrices of the form $I + \lambda$ where λ is a matrix with precisely one non zero entry and this entry does not occur on the diagonal. Suppose that R is a euclidean ring. Show that $\mathrm{SL}_n(R) = \mathrm{E}_n(R)$. (Cf. Appendix §104.)
11. Let A be a finite abelian group and let

$$\hat{A} := \{\chi : A \to \mathbb{C}^\times \mid \chi \text{ a group homomorphism}\}.$$

It is easily checked that $\hat{A}$ is a group via $\chi_1\chi_2(x) := \chi_1(x)\chi_2(x)$. Show

(*i*) A and $\hat{A}$ have the same order and, in fact, are isomorphic.

(*ii*) If χ is not the identity element of $\hat{A}$ then $\sum_{a\in A} \chi(a) = 0$.

12. Determine all abelian groups of order 400 up to isomorphism.

42. Canonical Forms for Matrices

In this section, we use the Fundamental Theorem of fg Modules over a PID to determine a good system of representatives for matrices over a field under the equivalence relation of similarity. Recall if V is a finite dimensional vector space over a field F with basis $\mathcal{B}$ and $T : V \to V$ is a linear operator then the *characteristic polynomial* of T is defined to be

$$f_T := \det(tI - [T]_{\mathcal{B}}) \in F[t].$$

This is independent of bases by the Change of Basis Theorem (cf. Appendix §103). Roots of f_T are called *eigenvalues* of T. An element λ is a root of f_T if and only if there exists a nonzero vector v in V such that $T(v) = \lambda v$. Such a v is called a (nonzero) *eigenvector* for T.

Construction 42.1. Let V be a finite dimensional vector space of dimension n over a field F and $T : V \to V$ a linear operator (endomorphism). Define

$$\varphi : F[t] \to \mathrm{End}_F(V) \text{ be given by } f \mapsto f(T)$$

evaluation at T. This is a ring homomorphism with

$$\mathrm{im}\,\varphi = \{\sum_{1=0}^{r} a_i T^i \mid a_i \in F \text{ some } m\},$$

a commutative subring of $\mathrm{End}_F(V)$. By the First Isomorphism Theorem $F[t]/\ker\varphi \cong \mathrm{im}\,\varphi = F[T]$. As $F[t]$ is a vector space over F on basis $\{t^i \mid i \geq 0\}$, we have $\dim_F F[t]$ is infinite. We also know $\dim_F F[T] \leq \dim_F \mathrm{End}_F(V) = \dim_F \mathbb{M}_n(F) = n^2$ is finite. Thus φ cannot be monic. Consequently, $0 < \ker\varphi$. As φ is a ring homomorphism, $\varphi(1) = 1_V$, so φ is not the trivial map. Therefore, $F[t]$ being a PID means:

$$\text{There exists a unique non-constant monic polynomial } q_T \text{ in } F[t] \text{ satisfying } \ker\varphi = (q_T) > 0.$$

We call q_T the *minimal polynomial* of T over F.

Note: By definition, $q_T(T) = 0$. (This polynomial need not be irreducible.)

The vector space V is an $\mathrm{End}_F(V)$-module via evaluation so becomes an $F[t]$-module via the pullback along φ, i.e., $f \cdot v := \varphi(f)(v) = f(T)(v)$ for all $f \in F[t]$ and all $v \in V$. Thus V is an $F[t]$-module via *evaluation* at T. As V is a finite dimensional vector space over F, it is a finitely generated $F[t]$-module. Moreover, $q_T \neq 0$ in $F[t]$ and $q_T \cdot v = q_T(T)(v) = 0v = 0$ for all v in V, so V is a finitely generated torsion $F[t]$-module. Therefore, the two forms of the Fundamental Theorem of fg Modules over a PID are applicable. We shall use both forms. To use them, we need to investigate cyclic submodules of the $F[t]$-module V.

We shall first apply the first form of the Fundamental Theorem 41.6 to the construction above. We shall use it for the case of an arbitrary field. We begin with some preliminaries.

Lemma 42.2. *Suppose V is a finite dimensional vector space over F and $T : V \to V$ a linear operator. Let V be an $F[t]$-module by evaluation at T, then*

$$(q_T) = \mathrm{ann}_{F[t]}\, V := \{f \in F[t] \mid f(v) = 0 \text{ for all } v \in V\} = \bigcap_{v \in V} \mathrm{ann}_{F[t]}\, v.$$

Proof. We know that $q_T v$ is zero for all v in V, so q_T lies in $\mathrm{ann}_{F[t]}\, V$. Let $\varphi : F[t] \to \mathrm{End}_F(V)$ be evaluation at T. Suppose that $f \in \mathrm{ann}_{F[t]}\, V$, then $0 = fv = f(T)(v)$ for all $v \in V$, hence $f(T)$ is the zero endomorphism in $\mathrm{End}_F(V)$, so $f \in \ker\varphi = (q_T)$. The result follows. □

We need another fact from linear algebra.

Definition 42.3. Let F be a field and $h = t^d + a_{d-1}t^{d-1} + \cdots + a_1 t + a_0$ be a monic polynomial in $F[t]$ of positive degree d. Define the

companion matrix of h to be the matrix in $\mathbb{M}_d(F)$ given by

$$C_h := \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & & \cdots & & \vdots \\ 0 & 0 & \cdots & 1 & -a_{d-1} \end{pmatrix}.$$

Remarks 42.4. Let F be a field and g, h be monic non-constant polynomials in $F[t]$.

1. $g = h$ if and only if $C_g = C_h$,
2. $h = f_{C_h}$, the characteristic polynomial of the matrix C_h:
   If $h = t^d + a_{d-1}t^{d-1} + \cdots + a_0$ with $d > 0$, then

$$f_{C_h} = \det \begin{pmatrix} t & 0 & \cdots & 0 & a_0 \\ -1 & t & \cdots & 0 & a_1 \\ 0 & -1 & \cdots & 0 & a_2 \\ \vdots & & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & t + a_{d-1} \end{pmatrix}.$$

Using induction and expanding the determinant by minors along the top row yield

$$f_{C_h} = t(t^{d-1} + a_{d-1}t^{d-2} + \cdots + a_1) + (-1)^{d-1}a_0(-1)^{d-1} = h.$$

To apply the Fundamental Theorem 41.6 to a finite dimensional vector space over a field F and linear operator T with $F[t]$-module structure given by evaluation at T, we must compute $F[t]$-cyclic submodules. The key is to determine an appropriate F-basis, for such a submodule.

Proposition 42.5. *Suppose V is a finite dimensional vector space over F and $T : V \to V$ a linear operator. Let V be a cyclic $F[t]$-module by evaluation at T, i.e., there exists a vector v in V satisfying $V = F[t]v$. Let $\mathcal{B} = \{v, T(v), \ldots, T^{d-1}(v)\}$ with d the degree of the minimal polynomial q_T of T. Then*

(1) $\mathcal{B}$ is an ordered basis for V.
(2) $\operatorname{ann}_{F[t]} v = \operatorname{ann}_{F[t]} V = (q_T)$
(3) The matrix representation $[T]_{\mathcal{B}}$ of T in the $\mathcal{B}$ basis satisfies $[T]_{\mathcal{B}} = C_{q_T}$.
(4) The minimal polynomial q_T satisfies $q_T = f_T$.

PROOF. **$\mathcal{B}$ spans V as a vector space over F**: Let w be a vector in $V = F[t]v$, so there exists a polynomial h in $F[t]$ satisfying $w = hv = h(T)(v)$, As $F[t]$ is a euclidean domain with euclidean

function the degree, we can write $h = q_T s + r$ for some polynomials s and r in $F[t]$ with $r = 0$ or $\deg r < \deg q_T$. Evaluating this polynomial at T yields

$$w = hv = (q_T s + r)v = q_T(T)s(T)(v) + r(T)(v) = r(T)(v)$$

which lies in the span of $\mathcal{B}$.

$\mathcal{B}$ is linearly independent (hence is a basis for V): Suppose that $\sum_{i=0}^{d-1} a_i T^i(v) = 0$ for some a_i in F. Set $g = \sum_{i=0}^{d-1} a_i t^i$ in $F[t]$. Then $gv = g(T)(v) = 0$, so $gfv = fgv = 0$ in $V = F[t]v$ for all polynomials f in $F[t]$, i.e., $gw = 0$ for all w in V. By Lemma 42.2, we have $g \in \operatorname{ann}_{F[t]} V = \operatorname{ann}_{F[t]} v = (q_T)$, so $q_T \mid g$. As $g = 0$ or $\deg g < \deg q_T$, we must have $g = 0$, i.e., $a_i = 0$ for all i. The proof of (3) is simple and (4) follows from Remarks 42.4. □

We now apply the Fundamental Theorem 41.6.

Theorem 42.6. (Rational Canonical Form) *Suppose V is a nontrivial finite dimensional vector space over F and $T : V \to V$ a linear operator. Let V be an $F[t]$-module by evaluation at T. Then there exists a direct sum decomposition of $F[t]$-modules*

$$V = V_1 \oplus \cdots \oplus V_r$$

for some r with each V_i a cyclic $F[t]$-module, $1 \le i \le r$, and satisfying:

1. *Let q_i be the monic non-constant polynomial in $F[t]$ satisfying $(q_i) = \operatorname{ann}_{F[t]} V_i$ for $1 \le i \le r$, then the q_i satisfy $q_1 \mid \cdots \mid q_r$ in $F[t]$ and are unique relative to this ordering and are invariant factors of V as an $F[t]$-module.*
2. *The polynomial q_r in (1) is the minimal polynomial q_T of T, and $(q_T) = \operatorname{ann}_{F[t]} V$.*
3. *Let q_i, $i = 1, \ldots, r$ be as in (1), then $f_T = q_1 \cdots q_r$.*
4. *There exist ordered bases $\mathcal{B}_i$ for V_i, $1 \le i \le r$ satisfying:*
   - *(a) $T|_{V_i}$ lies in $\operatorname{End}_F(V_i)$ for $i = 1, \ldots, r$.*
   - *(b) $[T|_{V_i}]_{\mathcal{B}_i} = C_{q_i}$, for $i = 1, \ldots, r$.*
   - *(c) $q_i = f_{T|_{V_i}} = q_{T|_{V_i}}$.*
   - *(d) Let $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_r$ is an ordered basis for V and*

$$[T]_{\mathcal{B}} = \begin{pmatrix} C_{q_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C_{q_r} \end{pmatrix} \text{ (in block form.)}$$

This matrix is unique relative to the monic polynomials in (1) as ordered there.

The monic polynomials, $q_1, \ldots, q_r$ in (1) in the theorem are called *the invariant factors* of T and the matrix in (4) is called the *rational canonical form* of T.

PROOF. Apply the Fundamental Theorem 41.6 and Proposition 42.5 after noting that V_i being $F[t]$-cyclic implies that $\operatorname{ann}_{F[t]} V_i = (q_i) = (q|_{V_i})$ and $T|_{V_i} : V_i \to V_i$ as V_i is an $F[t]$-module, i.e., V_i is T-invariant as $tv = T(v)$ for all $v \in V_i$ for $i = 1, \ldots, r$. □

An immediate consequence of Theorem 42.6 is the following:

Corollary 42.7. *Suppose V is a nontrivial finite dimensional vector space over F and $T : V \to V$ a linear operator. Let V be an $F[t]$-module by evaluation at T. Then $q_T \mid f_T$ in $F[t]$. In particular, q_T and f_T have the same roots in F and q_T is a product of linear polynomials in $F[t]$ if and only if f_T is a product of linear polynomials in $F[t]$.*

Of course, the multiplicity of a root of q_T and f_T may be different. (Recall from linear algebra that the roots of f_T are the eigenvalues of T.)

An important consequence of Theorem 42.6 is the well-known:

Corollary 42.8. (Cayley-Hamilton Theorem) *Suppose V is a nontrivial finite dimensional vector space over F and $T : V \to V$ a linear operator. Then the characteristic polynomial, f_T in $F[t]$, satisfies $f_T(T) = 0$, i.e., $f_T \in \operatorname{ann}_{F[t]} V = (q_T)$. In particular, $q_T \mid f_T$ in $F[t]$ and q_T and f_T have the same irreducible factors (although not necessarily with the same multiplicities).*

The rational canonical forms of a matrices in $\mathbb{M}_n(F)$ give an excellent system of representatives of similarity classes of such matrices. To establish this, we need the following observation.

Observation 42.9. If R is a domain, A a matrix in $\mathbb{M}_n(R)$ with $\det A$ nonzero, and $AX = 0$ (X and 0 in $R^{n\times 1}$), then $X = 0$, i.e., A is an R-monomorphism when viewed as a map $R^{n\times 1} \to R^{n\times 1}$. [This follows as A is invertible in $\mathbb{M}_n\big(qf(R)\big)$.]

Theorem 42.10. (Classification of Similarity Classes of Matrices over a Field) *Let F be a field and A and B be two $n \times n$ matrices in $\mathbb{M}_n(F)$.*

1. *There exists a unique matrix C in $\mathbb{M}_n(F)$ in rational canonical form with A and C similar.*
2. *The following are equivalent:*
   - *(a) $A \sim B$.*
   - *(b) A and B have the same rational canonical form.*
   - *(c) A and B have the same invariant factors.*

(d) $tI-A$ and $tI-B$ have the same Smith Normal Form when we choose monic invariants (and the nonidentity monic invariant factors are the same as those for A and B in (c)).
(e) $tI-A$ and $tI-B$ are equivalent in $\mathbb{M}_n(F[t])$.

PROOF. Let

$$\mathcal{S}_{n,1} = \{e_1, \ldots, e_n\} \text{ with } e_i = \begin{pmatrix} 0 & 0 & \underset{i}{1} & 0 & \cdots & 0 \end{pmatrix}^t$$

be the standard basis for $F^{n\times 1}$ and also for $F[t]^{n\times 1}$ as free modules over F and $F[t]^{n\times 1}$, respectively. (Aside: $\mathcal{C} = \{t^j e_i \mid 1 \leq i \leq n,\ j \geq 0\}$ is an F-basis for $F[t]^{n\times 1}$.) The linear operator $A : F^{n\times 1} \to F^{n\times 1}$ given by $v \mapsto Av$ satisfies $A = [A]_{\mathcal{S}_{n,1}}$, so there exists a basis $\mathcal{B}$ for $F^{n\times 1}$ such that $[A]_{\mathcal{B}}$ is in rational canonical form by the Change of Basis Theorem (cf. Appendix §103). Therefore, we have (1) and the equivalence of (a), (b), and (c) in (2) follow easily.

By the Universal Property of Free Modules 36.3, there exists a unique $F[t]$-endomorphism (hence also F-linear) $g : F[t]^{n\times 1} \to F[t]^{n\times 1}$ extending $e_j \mapsto te_j - Ae_j$. We, therefore, have a sequence

$$(*) \qquad 0 \to F[t]^{n\times 1} \xrightarrow{g} F[t]^{n\times 1} \xrightarrow{e_A} F^{n\times 1} \to 0$$

with e_A evaluation at A, i.e., defined by $fe_j \to f(A)e_j$ for all $f \in F[t]$ and all j. We shall show that this sequence is an exact sequence.

As $\det(tI_n - A) = f_A$, the characteristic polynomial of the matrix A, hence nonzero, we have g an $F[t]$-monomorphism by Observation 42.9 above. It is also clear that the evaluation map e_A is a surjection. So we must show exactness at the middle term, i.e., $\operatorname{im} g = \ker e_A$.

Clearly, $\operatorname{im} g \subset \ker e_A$. Therefore, $x + \operatorname{im} g \mapsto x + \ker e_A$ defines an an $F[t]$-epimorphism $F[t]^{n\times 1}/\operatorname{im} g \to F[t]^{n\times 1}/\ker e_A$. As we want to show $\operatorname{im} g = \ker e_A$, it suffices to show that this map is an isomorphism. Composing this map with $\overline{e_A} : F[t]^{n\times 1}/\ker e_A \to F^{n\times 1}$, the map induced by e_A given by the First Isomorphism Theorem, defines an $F[t]$-epimorphism $\widetilde{e_A} : \operatorname{coker} g \to F^{n\times 1}$, so it suffices to show that $\widetilde{e_A}$ is an isomorphism, i.e., a bijection.

By Remark 41.10, a Smith Normal Form S of the matrix representation $tI_n - A$ of g determines $\operatorname{coker} g$ as an $F[t]$-module. We may assume that S was chosen with its nonzero diagonal entries monic polynomials. As $\det(tI_n - A) = f_A \neq 0$, S has no nonzero diagonal entries.

Let the non-constant monic diagonal entries of S be $d_1 | \cdots | d_r$. We know that $S = P[g]_{\mathcal{S}_{n,1}}Q$ with $P, Q \in \mathrm{GL}_n(F[t])$, and an invertible matrix in $\mathrm{GL}_n(F[t])$ has determinant in $F[t]^\times = F^\times$. (This also follows

from the algorithm given in Appendix §104, Theorem 104.2, since Type I, II, and III matrices have determinant in $F^\times$). Consequently, we have $f_A = d_1 \cdots d_r$ as f_A and all the d_i are monic. Hence

$$\begin{aligned}\dim_F \operatorname{coker} g &= \sum_{i=1}^{r} \dim_F F[t]/(d_i) \\ &= \sum_{i=1}^{r} \deg d_i = \deg f_A = n = \dim_F F^{n\times 1}\end{aligned}$$

using the isomorphism we showed when proving the Fundamental Theorem 41.6 and Proposition 31.14. It follows that the $F[t]$-epimorphism hence F-epimorphism $\widetilde{e_A}$ is an F-isomorphism, hence $\operatorname{im}(g) = \ker e_A$ and the sequence (*) is exact.

It follows from the exactness of (*) that the non-constant invariant factors of a Smith Normal Form S determines the invariants of $F^{n\times 1}$ as an $F[t]$-module which therefore must be the same as those determined by the invariants of A. The equivalence of (c), (d), and (e) follow. □

Corollary 42.11. *Let F be a field and $A \in \mathbb{M}_n(F)$. Then A is similar to its transpose A^t.*

Proof. It follows Theorem 40.8 (or by the algorithm given in the proof of Theorem 104.2 in Appendix §104) that a matrix and its transpose in $\mathbb{M}_n(F[t])$ have equivalent Smith Normal Forms. In particular, the matrices $tI - A$ and $tI - A^t$ have the same Smith Normal Form when we choose monic invariants. □

Using the proof above, we can give a summary of what is going on. Instead of giving this summary using matrices, we use linear operators.

Summary 42.12. Let V be a finite dimensional vector space over F of dimension n and $T : V \to V$ a linear operator. We view V as an $F[t]$-module by $tv := T(v)$ for all $v \in V$. We then have an exact sequence

$$0 \to F[t]^n \xrightarrow{t1_{F[t]^n} - T} F[t]^n \xrightarrow{e_T} V \to 0 \tag{42.13}$$

with the $F[t]$-homomorphism e_T defined by $\sum f_i e_i \mapsto \sum f_(T) e_i$, where e_i in the ith basis element in the (ordered) standard basis $\mathcal{S}_n$ for $F[t]^n$ (and F^n). We call the sequence 42.13 the *characteristic sequence* of T. The matrix $t1_{F[t]^n} - [T]_{\mathcal{S}_n}$ is called the *characteristic matrix* of $[T]_{\mathcal{S}_n}$. It has nonzero determinant f_T, the characteristic polynomial of T (and is the reason that $t1_{F[t]^n} - [T]_{\mathcal{S}_n}$ is a monomorphism). In particular, the characteristic matrix of $[T]_{\mathcal{S}_n}$ is nonsingular, so has a unique Smith Normal Form of $(tI_n - [T]_{\mathcal{S}_n})$ given by $\operatorname{diag}(1, \ldots, q_1, \ldots, q_r)$ has the q_i

monic non-constant polynomials satisfying $q_1 \mid \cdots \mid q_r$. (The entries of a Smith Normal Form are unique up to units.)

As the sequence 42.13 is exact, we have

$$\begin{aligned}\operatorname{coker}(t1_{F[t]^n} - T) &= F[t]^n / \operatorname{im}(t1_{F[t]^n} - T) \\ &= F[t]^n) / \ker(t1_{F[t]^n} - T) \cong V.\end{aligned}$$

By the Fundamental Theorem, we can decompose the finitely generated torsion $F[t]$- module V into cyclic $F[t]$-submodules, say

$$V = V_1 \coprod \cdots \coprod V_r$$

with $V_i = F[t]v_i$ some v_i for each i and satisfying

$$V_i \cong F[t]/\operatorname{ann}_{F[t]}(v_i) = F[t]/\operatorname{ann}_{F[t]}(V_i) = F[t]/(q_i),\ i = 1, \ldots, r$$

with the descending chain of ideals

$$(q_1) \supset \cdots \supset (q_r)$$

unique. Moreover,

$$q_i = q_{T|_{V_i}},\ \ q_r = q_T,\ \ f_T = q_1 \ldots q_r,$$

and

$$RCF(T) := RCF([T]_{\mathcal{S}_n} = \begin{pmatrix} C_{q_1} & & \\ & \ddots & \\ & & C_{q_r} \end{pmatrix}.$$

In particular, to find the invariant factors of T one computes the Smith Normal Form of the characteristic matrix using the algorithm given in the Appendix §104.

We turn to the second form of the Fundamental Theorem. We shall apply the second form for matrices over an algebraically closed field, i.e., when non-constant polynomials always factor into a product of linear polynomials, as it is this case that the elementary divisors arising from this form are useful. We first establish the cyclic case when the minimal polynomial is a power of a linear polynomial.

Lemma 42.14. *Suppose V is a finite dimensional vector space over F and $T : V \to V$ a linear operator. Let V be a cyclic $F[t]$-module by evaluation at T, i.e., there exists a vector v in V satisfying $V = F[t]v$. Suppose that the minimal polynomial q_T of T is $(t-a)^d$ in $F[t]$. Then*

$\mathcal{B} = \{v, (T-a)(v), \ldots, (T-a)^{d-1}(v)\}$ *is an ordered basis for* V *and*

$$[T]_{\mathcal{B}} = \begin{pmatrix} a & 0 & 0 & \cdots & 0 \\ 1 & a & 0 & \cdots & \\ 0 & 1 & a & & \\ \vdots & & \ddots & \ddots & \\ 0 & & \cdots & 1 & a \end{pmatrix} \text{ in } \mathbb{M}_d(F).$$

PROOF. We know that $\mathcal{C} = \{v, T(v), \ldots, T^{d-1}(v)\}$ is an F-basis for V by our previous work. If $\mathcal{B}$ is linearly dependent, then there exists an equation

$$(*) \qquad 0 = \sum_{i=0}^{d-1} a_i (T-a)^i v \text{ for some } a_i \text{ in } F, \text{ not all zero.}$$

Choose N maximal such that $a_N \neq 0$ and expand (*) to get

$$0 = a_N T^N v + \sum_{i=0}^{N-1} b_i (T-a)^i v \text{ for some } b_i \text{ in } F.$$

As $\mathcal{C}$ is linearly independent, $a_N = 0$, a contradiction. Therefore, $\mathcal{B}$ is linearly independent hence a basis as $|\mathcal{B}| = |\mathcal{C}|$. We have

$$0 = q_r(T) = (T-a)^r \text{ and}$$
$$T(T-a)^i = \big((T-a)+a\big)(T-a)^i = (T-a)^{i+1} + a(T-a)^i$$

for $i = 0, \ldots r-1$. So

$$T(T-a)^i(v) = (T-a)^{i+1}(v) + a(T-a)^i(v) \text{ for } i = 0, \ldots, r-1$$

and the last statement follows. □

When the minimal polynomial of a linear operator on a finite dimensional vector space factors into a product of linear terms, the Fundamental Theorem 41.22 and the last lemma now immediately implies the following:

Theorem 42.15. (Jordan Canonical Form) *Suppose* V *is a nontrivial finite dimensional vector space over* F *and* $T : V \to V$ *a linear operator. Let* V *be an* $F[t]$*-module by evaluation at* T*. Suppose that* q_T *factors into linear terms in* $F[t]$*. Then there exists an ordered basis* $\mathcal{B}$

for V satisfying

$$[T]_{\mathcal{B}} = \begin{pmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_r \end{pmatrix} \quad \text{with} \quad A_i = \begin{pmatrix} a_i & 0 & & \cdots & & 0 \\ 1 & a_i & 0 & & & 0 \\ 0 & 1 & a_i & 0 & & \vdots \\ \vdots & & & \ddots & & \\ & & & 1 & a_i & 0 \\ 0 & & \cdots & & 1 & a_i \end{pmatrix}$$

for $i = 1, \ldots, r$. The a_i, $i = 1, \ldots, r$ are the eigenvalues of T (with $a_i = a_j$, $i \neq j$ possible). This matrix of these blocks is unique up to the order of the blocks. The monic polynomials f_{A_i}, $1 \leq i \leq r$, are the elementary divisors of V with the above $F[t]$-module structure and the characteristic polynomial of T satisfies $f_t = \prod_{i=1}^{r} f_{A_i}$

In the theorem, the blocks A_i are called *Jordan blocks* and, as the monic elementary divisors f_{A_i} are unique, they are called the *elementary divisors* of T. The matrix representing T in 42.15 is called the *Jordan canonical form* of T.

Corollary 42.16. *Let A and B lie in $\mathbb{M}_n(F)$. If each of the characteristic polynomials f_A and f_B factor into a product of linear terms in $F[t]$, then A and B are similar in $\mathbb{M}_n(F)$ if and only if A and B have the same Jordan canonical form (up to block order) if and only if they have the same elementary divisors (counted with multiplicity). In particular, if F is an algebraically closed field, then a matrix in $\mathbb{M}_n(F)$, $n \in \mathbb{Z}^+$, is determined up to similarity by its Jordan Canonical Form.*

Remark 42.17. The rational canonical form of a matrix is quite computable, as the division algorithm is computable, but the Jordan canonical form (when it exists) maybe harder to compute as it depends on factoring the invariant factors of the monic polynomials arising in the rational canonical form.

Corollary 42.18. *Let F be a field and A a matrix in $\mathbb{M}_n(F)$. Then A is similar to a diagonal matrix if an only if the minimal polynomial q_A factors as a product of monic linear factors without repetition, i.e., $q_A = (t - \lambda_1) \cdots (t - \lambda_r)$ for some distinct $\lambda_1, \ldots, \lambda_r$ in F.*

We say that a non-constant polynomial in $F[t]$ *splits* over F if it is a product of linear polynomials in $F[t]$. The corollary says that the matrix is *diagonalizable* if and only if q_A splits over F without any *multiple roots*.

Corollary 42.19. *Let F be a field and A a matrix in $\mathbb{M}_n(F)$. Then A is similar to a triangular matrix if an only if the minimal polynomial q_A splits over F.*

We say that A is *triangularizable* if q_A splits over F.

Examples 42.20. Let F be a field

1. Similarity classes of matrices in $\mathbb{M}_3(F)$ are completely determined by the minimal and characteristic polynomials of a matrix:
   Let A lie in $\mathbb{M}_3(F)$ and $q_1 \mid \cdots \mid q_r$ be the invariant factors of A. So $f_A = q_1 \cdots q_r$ with $q_r = q_A$. As $A \in \mathbb{M}_3(F)$, we know that $r \leq 3$.

   $r = 1$: We must have $q_1 = q_A = f_A$ determines A up to similarity, so A is similar to $C_{q_1} = C_{q_A}$. We can say more if q_A is reducible. First suppose that $q_A = (t - \lambda)h$ in $F[t]$ with h irreducible. By Exercise 42.21(2) below, A is similar to the matrix

   $$\begin{pmatrix} \lambda & 0 \\ 0 & C_h \end{pmatrix},$$

   which is unique up to block order. If q_A splits over F, say $q_A = (t - \lambda_1)(t - \lambda_2)(t - \lambda_3)$, then A is similar to

   $$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \qquad \text{if } \lambda_1, \lambda_2, \lambda_3 \text{ are distinct,}$$

   $$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 1 & \lambda_1 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \qquad \text{if } \lambda_1 = \lambda_2 \neq \lambda_3,$$

   $$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 1 & \lambda_1 & 0 \\ 0 & 1 & \lambda_1 \end{pmatrix} \qquad \text{if } \lambda_1 = \lambda_2 = \lambda_3,$$

   matrices in Jordan Canonical Form, unique up to order of the blocks.

   $r = 2$: We have $q_1 \mid q_2 = q_A$ and $f_A = q_1 q_2$, so $q_1 = f_A / q_A$ in $F[t]$. Therefore, q_A and f_A determine A up to similarity, and A is similar to

   $$\begin{pmatrix} C_{q_1} & 0 \\ 0 & C_{q_2} \end{pmatrix}.$$

   Note that if this is the case, $\deg q_1 = 1$ and $\deg q_2 = 2$, so $q_1 \mid q_2$ implies that q_2 splits so A is triangularizable, i.e., a Jordan canonical

form for A exists. If $q_1 = t - \lambda_1$ and $q_2 = (t - \lambda_1)(t - \lambda_2)$, then A is similar to

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix} \qquad \text{if } \lambda_1 \neq \lambda_2,$$

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 1 & \lambda_1 \end{pmatrix} \qquad \text{if } \lambda_1 = \lambda_2,$$

unique up to the order of the blocks.

$r = 3$: We have $q_1 \mid q_2 \mid q_3 = q_A$, so $q_1 = q_2 = q_3 = q_A$. So in this case q_A and f_A determine A up to similarity, and A is diagonizable with a single eigenvalue, i.e., A is similar to the matrix λI with λ an (the) eigenvalue of A. In particular, A is similar to a matrix in Jordan canonical form.

2. Similarity classes of matrices in $\mathbb{M}_4(F)$ are not determined by minimal and characteristic polynomials. For example the matrices

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

have

$q_A = (t-1)^2$	$q_B = (t-1)^2$
$f_A = (t-1)^4$	$f_B = (t-1)^4$
elementary divisors	elementary divisors
$t-1, t-1, (t-1)^2$	$(t-1)^2, (t-1)^2$,
invariant factors	invariant factors
$t-1 \mid t-1 \mid (t-1)^2$	$(t-1)^2 \mid (t-1)^2$

so they are not similar.

3. A system of representatives for the similarity classes of matrices in $\mathbb{M}_4(F)$ having characteristic polynomial $t(t-1)^3$ are:

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

minimal polynomial	minimal polynomial	minimal polynomial
$t(t-1)$	$t(t-1)^2$	$t(t-1)^3$
elementary divisors	elementary divisors	elementary divisors
$t, t-1, t-1, t-1$	$t, (t-1), (t-1)^2$	$t, (t-1)^3$
invariant factors	invariant factors	invariant factors
$t-1 \mid t-1 \mid t(t-1)^2$	$(t-1) \mid t(t-1)^2$	$t(t-1)^3$

4. A system of representatives for the similarity classes of matrices in $\mathbb{M}_4(F)$ with minimal polynomial $t(t-1)$ are:

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

as these matrices must be diagonalizable with eigenvalues 0 and 1.

5. Let A be a 3×3 matrix in $\mathbb{M}_3(\mathbb{Q})$ satisfying $A^3 = I$. Then $q_A \mid t^3 - 1 = (t-1)(t^2+t+1)$ in $\mathbb{Q}[t]$, so the only possibilities are $t-1$, t^2+t+1, or t^3-1. If $q_A = t^2+t+1$, then $q_1 \mid q_2 = q_A$, but q_A is irreducible, so no linear polynomial divides it, so this case cannot occur. If $q_A = t-1$, A is diagonalizable and $t-1 \mid t-1 \mid t-1$ are the invariants. So the possible rational canonical forms for A are:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

[Note: Using Exercise 42.21(2) below, an alternate matrix similar to

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{is} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}]$$

Exercises 42.21. 1. Prove Observation 42.9.

2. Let F be a field and A a matrix in $\mathbb{M}_n(F)$. Let $f_1, \ldots, f_r$ be the monic elementary divisors of A, i.e., those of $F^{n\times 1}$ as a $F[t]$ module via $tv = Av$ for all v in $F^{n\times 1}$. Show that A is similar to the matrix

$$\begin{pmatrix} C_{f_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C_{f_r} \end{pmatrix}$$

and this matrix is unique up to the order of the blocks.

3. Let A be a matrix in $\mathbb{M}_5(\mathbb{Q})$ satisfying $A^3 = I$. Describe up to similarity all possible rational canonical forms of A and justify your answer.
4. Describe, up to similarity, all 4×4 matrices A over a field F such that $A^5 = I$, in the following three cases and justify your answer.
   (*i*) F is the field of rational numbers.
   (*ii*) F is $\mathbb{Z}/2\mathbb{Z}$
   (*iii*) F is $\mathbb{Z}/5\mathbb{Z}$
5. Compute the Jordan and Rational Canonical Forms of the matrix

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

6. Let V be a real vector space, T an $\mathbb{R}$-endomorphism. Suppose that the minimal polynomial of T factors into linear terms over $\mathbb{R}$ with no repeated and no negative roots. Show that there exists an $\mathbb{R}$-endomorphism S of V such that $S^2 = T$.
7. Recall that a matrix $C \in \mathbb{M}_n\mathbb{C}$ is called *diagonalizable* if there exists a matrix $Q \in \mathrm{GL}_n(\mathbb{C})$ such that QAQ^{-1} is a diagonal matrix. Show if $A, B \in \mathbb{M}_n\mathbb{C}$ are both diagonalizable, then there is a matrix $P \in \mathrm{GL}_n(\mathbb{C})$ satisfying both PAP^{-1} and PBP^{-1} are diagonal matrices if and only if $AB = BA$.
8. Determine a theorem for a canonical form for matrices in $\mathbb{M}_n(F)$ when F is an arbitrary field given by the Fundamental Theorem 41.22.

43. Addendum: Jordan Decomposition

We prove a weaker form of Jordan canonical form that only needs the Primary Decomposition Theorem, not the full Fundamental Theorem 41.6. (The Cayley-Hamilton Theorem can be proven without the full version of the Fundamental Theorem.)

Definition 43.1. Suppose that F is a field, V a vector space over F, and T is a linear operator on V. Let

$$E_T(\lambda) := \{v \in V \mid T(v) = \lambda v\},$$

a subspace of V. If λ is an eigenvalue of T, then $E_T(\lambda)$ is called the *eigenspace* of λ relative to T, i.e., $E_T(\lambda)$ is an eigenspace if and only if it is not the zero subspace. We say that T is *semisimple* if V has a basis consisting of eigenvectors of T, equivalently, $V = \bigoplus_F E_T(\lambda)$. Of course, $T|_{E_T(\lambda)} = \lambda 1_{E_T(\lambda)}$ for all $\lambda \in F$.

Observation 43.2. Let V be a vector space over F and S, T two commuting linear operators on V, i.e., $T \circ S = S \circ T$. If λ is an eigenvalue of T, then $E_T(\lambda)$ is S-invariant, i.e., $S|_{E_T(\lambda)} : E_T(\lambda) \to E_T(\lambda)$:
If $v \in E_T(\lambda)$, then $T\big(S(v)\big) = S\big(T(v)\big) = S(\lambda v) = \lambda S(v)$.

The observation has the following consequence:

Theorem 43.3. *Let V be a finite dimensional vector space over the field F and $\mathcal{S}$ a set of commuting semisimple linear operators in $\mathrm{End}_F(V)$, i.e., if T and S lie in $\mathcal{S}$, they are semisimple and $ST = TS$. Then there exists a basis $\mathcal{B}$ for V consisting of eigenvectors for every T in $\mathcal{S}$ simultaneously. In particular, $T + S$ is semisimple for all $T, S \in \mathcal{S}$.*

Proof. If $T = \rho 1_V$ for some $\rho \in F$ for every $T \in \mathcal{S}$, the result is trivial. So we may assume that there exists a T in $\mathcal{S}$ so that this is not true. As T can have only finitely many eigenvalues, say $\lambda_1, \ldots, \lambda_r$, we have $V = E_T(\lambda_1) \oplus \cdots \oplus E_T(\lambda_r)$ with each $E_T(\lambda_j) < V$. By the observation, $S|_{E_T(\lambda_j)} \in \mathrm{End}_F\big(E_T(\lambda_j)\big)$ for all $S \in \mathcal{S}$. As $q_{S|_{E_T(\lambda_j)}} \mid q_S$ (why?), we know that $q_{S|_{E_T(\lambda_j)}}$ splits over F, so $S|_{E_T(\lambda_j)}$ is semisimple. By induction, the result holds on each $E_T(\lambda_j)$, hence on V. □

Definition 43.4. Let F be a field, V a nontrivial vector space over F, and T a linear operator on V. We say that T is *nilpotent* if there exists a positive integer N such that $T^N = 0$ and *unipotent* if $T - 1_V$ is nilpotent.

Note that a unipotent linear operator is always an isomorphism (why?) and an operator that is both nilpotent and semisimple must be the zero linear operator.

Theorem 43.5. *Let V a finite dimensional vector space over F, and T a linear operator on V. Suppose that f_T splits over F. Then there exist unique linear operators T_s and T_n on V semisimple and nilpotent respectively and if T is a linear automorphism, there exists a unique unipotent operator T_u satisfying the following:*

(1) (Additive Jordan Decomposition) *$T = T_s + T_n$ and $T_s T_n = T_n T_s$.*
(2) (Multiplicative Jordan Decomposition) *If T is a linear automorphism, then $T = T_s T_u$ and $T_s T_u = T_u T_s$.*

(3) *There exist polynomials g and h in $F[t]$ with zero constant term, i.e., $g(0) = 0 = h(0) = 0$ satisfying $T_s = g(T)$ and $T_n = h(T)$. In particular, if a linear operator S on V commutes with T, then it commutes with both T_s and T_n.*

(4) *If T is a linear automorphism and T commutes with a linear operator S on V, then it commutes with T_s and T_u.*

PROOF. Let V be an $F[t]$-module by evaluation at T and suppose that the characteristic polynomial of T is $f_T = \prod_{i=1}^{r}(t-\lambda_i)^{e_i}$ in $F[t]$ with $\lambda_i \neq \lambda_j$ if $i \neq j$, and $e_i > 0$ for all i. Then by the Cayley-Hamilton Theorem, $f_T v = 0$ for all v in V. Let

$$V_i = \ker(T - \lambda_i 1_V)^{e_i} \cong F[t]/\big((t-\lambda_i)^{e_i}\big)$$

for $i = 1, \ldots, r$, so $V = V_1 \oplus \cdots \oplus V_r$ as $F[t]$-modules by the Primary Decomposition Theorem 41.20 (Why?). By the Chinese Remainder Theorem, there exists a polynomial g in $F[t]$ satisfying $g \equiv \lambda_i \mod (t-\lambda_i)^{e_i}$, $1 \leq i \leq r$ and $g \equiv 0 \mod t$. (Of course, if some $\lambda_i = 0$, the last congruence is redundant.). In particular, $g(0) = 0$. Let $T_s = g(T)$. Each V_i is T-invariant, as it is an $F[t]$-module by evaluation at T, so each V_i is also T_s-invariant. We also have $T_s|_{V_i} = \lambda_i 1_{V_i}$ for $i = 1, \ldots, r$, as the λ_i's are eigenvalues of T. Therefore, T_s is semisimple.

Let $h = t - g$ in $F[t]$, so $h(0) = 0$. Set $T_n = h(T)$. We have

$$T = g(T) + h(T) = T_s + T_n \text{ and } T_s T_n = T_n T_s.$$

In addition,

$$0 = (T|_{V_i} - \lambda_i 1_{V_i})_i^e = (T|_{V_i} - T_s|_{V_i})^{e_i} = T_n^{e_i}|_{V_i}$$

for each i, so T_n is nilpotent. This proves (3) and establishes the existence in (1).

If T is a linear automorphism, then 0 is not an eigenvalue of T, so T_s is invertible. Let $T_u = 1_V + T_s^{-1} T_n$. Then T_u is unipotent as $T_s^{-1} T_n = T_n T_s^{-1}$ is nilpotent. Statement (4) and the existence in (2) now follow.

We next show the uniqueness in (1). Suppose that we have a semisimple operator T_s' and a nilpotent operator T_n' satisfying

$$T_s + T_n = T_s' + T_n' \text{ and } T_s' T_n' = T_n' T_s'$$

Then $T_s - T_s' = T_n' - T_n$. By (3), T_s' and T_n' commute with both T_s and T_n, as they commute with $T = T_s' + T_n'$. Therefore, $T_n' - T_n$ is nilpotent and by Theorem 43.3 we have $T_s - T_s'$ is semisimple. Therefore, $T_s - T_s' = T_n' - T_n = 0$, i.e., $T_s = T_s'$ and $T_n = T_n'$.

Finally, we show the uniqueness in (3). Suppose that

$$T_sT_u = T'_sT'_u \text{ and } T'_sT'_u = T'_nT'_u$$

with T_u unipotent. Then $S = T'_u - 1_V$ is nilpotent and commutes with T'_s, hence T'_sS is nilpotent and $T = T'_s + T'_sS$ is a additive Jordan decomposition, so $T'_s = T_s$ (and $T_n = T'_sS$). Hence $T_u = T_s^{-1}T = T'^{-1}_s T = T'_u$, and we are done. □

Our definition of a semisimple operator T on V is equivalent to the minimal polynomial q_T splitting without any repeated factors over F. The theorem generalizes if we can find an alternate condition of a linear operator on V when q_T does not split over F that will agree with our definition of a semisimple operator over a field containing F in which q_T splits. Then we can mimic the proof on a factorization of q_T, if it is not a product of distinct linear terms. To do this, one can redefine an operator T to be a *semisimple* operator on a vector space V over F if it satisfies the following: Let V be an $F[t]$-module by evaluation at T. If W is an $F[t]$-submodule of V then $V = W \oplus W'$ for some $F[t]$-module W', i.e., and exact sequence

$$0 \to W \to V \to W' \to 0$$

of $F[t]$-modules is split exact (cf. Exercise 36.12(15)). With this definition, the theorem still holds if the characteristic of F is zero. [If the characteristic of F is positive, irreducible polynomials can have multiple roots over some field containing F.]

Exercises 43.6. 1. Show a unipotent operator on a vector space is an isomorphism and an operator that is both nilpotent and semisimple on a vector space is the trivial operator.

2. In the proof of Theorem 43.5, show that $V_i = \ker(T - \lambda_i 1_V)^{e_i} \cong F[t]/\big((t-\lambda_i)^{e_i}\big)$ is the $(t-\lambda_i)$-primary submodule of V as an $F[t]$-module.
3. In Theorem 43.5, show if W is an $F[t]$-submodule of V, i.e., is T-invariant, then the Jordan decomposition(s) of T induce those on $T|_W$ and $T_{V/W}$ (where $T_{V/W}(x+W) = T(x)$).

44. Addendum: Cayley-Hamilton Theorem

In this addendum, we generalize the Cayley-Hamilton Theorem. Let R be a commutative ring and M a finitely generated R-module. Let

$$M[t] := \{\sum t^i x_i \mid x_i \in M \text{ almost all } x_i = 0\} = \coprod_{i=0}^{\infty} t^i M.$$

This is clearly an R-module and becomes a finitely generated $R[t]$-module in the obvious way.

For example, if V is a finite dimensional vector space over F on basis $\{v_i\}$, then $V[t]$ is the F-vector space on basis $\{t^i v_j \mid i \geq 0, j\}$ and a free $F[t]$-module on basis $\{v_i\}$. So if $\dim_F V = n$, $V[t] \cong F[t]^n$.

If $f : M \to N$ is an R-homomorphism, define $f[t] : M[t] \to N[t]$ by $\sum t^i x_i \mapsto \sum t^i f(x_i)$, an $R[t]$-homomorphism.

Suppose that N is an $R[t]$-module, then it is also an R-module and N is completely determined by

(i) the R-module structure.

(ii) The R-endomorphism $f : N \to N$ given by $f(x) = tx$.

In particular, an R-module becomes an $R[t]$-module via (i) and (ii). We denote this $R[t]$-module N_f. We then have an $R[t]$-epimorphism $\varphi_f : N[t] \to N_f$ given by $\sum t^i x_i \mapsto \sum f^i(x_i)$.

Theorem 44.1. *Let R be a commutative ring and M a finitely generated R-module. Suppose that f is an R-endomorphism of M. Then, in the notation above, we have an exact sequence*

$$0 \to M[t] \xrightarrow{t1_{M[t]} - f[t]} M[t] \xrightarrow{\varphi_f} M_f \to 0$$

of $R[t]$-modules.

The exact sequence in the theorem is called the *characteristic sequence* for f.

PROOF. **Exactness at the middle** $M[t]$: As

$$\begin{aligned}\big(\varphi_f \circ (t1_{M[t]} - f[t])\big)\big(\sum t^i x_i\big) &= \varphi_f\big(\sum t^{i+1} x_i - t^i f(x_i)\big) \\ &= \sum f^{i+1}(x_i) - f^{i+1}(x_i) = 0,\end{aligned}$$

we have $\operatorname{im}(t1_{M[t]} - f[t]) \subset \ker \varphi_f$. If $x = \sum t^i x_i \in \ker f$, then $\sum f^i(x_i) = 0$, hence

$$x = x - \sum f^i(x_i) = \sum_{i \geq 0} \big(t^i x_i - f^i(x_i)\big) = \sum_{i>0} (t^i 1_{M[t]} - f^i)(x_i).$$

Check if $h_i = \sum_{j=0}^{i} t^j 1_{M[t]} f^{i-j}[t]$, then

$$x = \sum_{i>0} (t1_{M[t]} - f[t])\big(\sum_{i>0} h_i(x_i)\big) \text{ lies in } \operatorname{im}(f(t1_{M[t]} - f[t]).$$

φ_f **is surjective**: This follows as $\varphi_f(x) = x$ for all $x \in M$.

$t1_{M[t]} - f[t]$ is **monic**: We leave this part of the proof as an exercise noting that $t1_{M[t]} - f[t]$ raises the "degree" by one and preserves the "leading coefficients", so induction on degree works. □

Cramer's Rule can be stated in the following way: Let R be a commutative ring, f an R-endomorphism of the free R-module R^n. Set $M = \operatorname{coker} f = R^n/\operatorname{im} f$. Then there exists an R-endomorphism g of R^n such that $fg = gf = (\det f)1_{R^n}$. In particular, $R^n \det f \subset \operatorname{im} f$. If M is a free R=module of rank n on basis $\mathcal{B}$ and $f \in \operatorname{End}_R(M)$, then $M[t]$ is $R[t]$-free on $\mathcal{B}$ and $[f]_{\mathcal{B}} = [f[t]]_{\mathcal{B}}$. If follows that $P(t) := \det(t1_{M[t]} - f[t])$ is that characteristic polynomial of f. Applying the characteristic sequence of f in the theorem implies that $P(f)M_f = 0$. As the R-endomorphism of M_f defined by $P_f(t)$ is just $P_f(f)$, we have $P_f(f) = 0$, i.e., the Cayley-Hamilton Theorem.

Exercise 44.2. Prove that $t1_{M[t]} - f[t]$ in the characteristic sequence is monic.

Part 5

Field Theory

CHAPTER XI

Field Extensions

In this chapter, we begin our study of fields. Given a non-constant polynomial, it is natural to ask what its roots are. To make this more precise, if R is a ring and f is a non-constant polynomial in a single variable over R, one asks does f have any roots in R. If so how many and if not does there exist a ring S with R a subring such that it does. We have seen if R is a domain, then the number of distinct roots f is bounded by the degree of f. The first domain one encounters is the integers. The arithmetic problem, i.e., for f in $\mathbb{Z}[t]$ and monic leads to the study of algebraic number theory that we shall investigate in Chapter XV and is harder than that over fields. If $R = F$ is a field, we already know how to obtain a field K containing F such that f has a root in K. By induction, it is easy to construct a field E containing F for which the polynomial f splits. We shall show the smallest such E is essentially unique. Generalizing this construction, we indicate, using Zorn's Lemma, how to do this for sets of polynomials in $F[t]$. In particular, we construct a field over which every polynomial in $F[t]$ splits (and is the smallest possible). This field turns out to be an algebraically closed field, i.e., one in which every non-constant polynomial over it splits. We also use the theory developed to answer the classical euclidean construction problems using only straight-edge and compass: the trisection of an angle, doubling of a cube, squaring the circle, and construction of regular n-gons, showing how to turn a geometric problem into an algebraic one. The solution of the first two will be done in this chapter as well as the last (assuming a result proven in the next chapter) and the third will be done in §67. Finally, we investigate the difficulty arising from the case of positive characteristic of a field, where irreducible polynomials can have multiple roots over some bigger field.

45. Algebraic Elements

In this section, given a field F, we are mostly interested in roots of polynomials with coefficients in F. We have seen before that if f is a nonzero polynomial in $F[t]$, then there exists a field K containing F

such that f has a root in K. In fact, we explicitly constructed such a field. We shall investigate this situation in depth.

Let K be a field containing F as a subfield, so $F \subset K$. We shall write this as K/F. We call F the *base field* of K/F and K the *extension field* of K/F. If E is a field with $F \subset E \subset K$ (as subfields), written $K/E/F$, we call E an *intermediate field* of K/F. We allow E to be K or F. A sequence of fields $F = F_1 \subset \cdots \subset F_n \subset \cdots$ is called a *tower of fields*. It may be infinite.

Definition 45.1. Let K/F be an extension of fields. Then K is a vector space over F by restriction of scalars. We denote the dimension, $\dim_F K$, of K as an F-vector space by $[K : F]$ and call it the *degree* of K/F. We call K/F a *finite extension* if $[K : F]$ is finite, and an *infinite extension* otherwise.

Examples 45.2. 1. F/F is a finite extension of degree one.

2. $\mathbb{C}/\mathbb{R}$ is a finite extension of degree two.

3. $\mathbb{R}/\mathbb{Q}$ is an infinite extension.

Proposition 45.3. *Let $L/K/F$ be an extension of fields. Then L/F is a finite extension if and only if both L/K and K/F are finite field extensions. Moreover, if L/F is a finite extension, then*

$$[L : F] = [L : K][K : F].$$

Proof. We prove a stronger statement. Let

$$\begin{aligned} \mathcal{B} &= \{x_i\}_I \text{ be an } F\text{-basis for } K, \\ \mathcal{C} &= \{y_j\}_J \text{ be an } K\text{-basis for } L, \text{ and} \\ \mathcal{D} &= \{x_i y_j\}_{I \times J}. \end{aligned}$$

Claim. $\mathcal{D}$ is an F basis for L and $|\mathcal{D}| = |I \times J|$.

If we show this, then we are done.

Note $x_i y_j = x_{i'} y_{j'}$ if and only if $x_i = x_{i'}$ and $y_j = y_{j'}$ as $\mathcal{C}$ is a K-basis of L and x_i, x_i' lie in K, hence $|\mathcal{D}| = |\mathcal{B}||\mathcal{C}| = |I \times J|$.

$\mathcal{D}$ **F-spans** L: Let $z \in L$. Then $z = \sum_J \alpha_j y_j$ for some $\alpha_j \in K$ almost all zero, as $\mathcal{C}$ spans L as a vector space over K. As $\mathcal{B}$ spans K as a vector space over F, each $\alpha_j = \sum_I c_{ij} x_i$ for some $c_{ij} \in F$ almost all zero. Hence we have $z = \sum_I \sum_J c_{ij} x_i y_j$ lies in the F-span of $\mathcal{D}$.

$\mathcal{D}$ **is F-linearly independent**: Suppose that $\sum_{I \times J} c_{ij} x_i y_j = 0$ for some c_{ij} in F, almost all zero. Then $\sum_J (\sum_I c_{ij} x_i) y_j = 0$. As $\mathcal{C}$ is K-linearly independent, $\sum_I c_{ij} x_i = 0$ for all j. As $\mathcal{B}$ is F-linearly independent $c_{ij} = 0$ for all i and j. □

This field theoretic analogue of Lagrange's Theorem has two analogous consequences.

Corollary 45.4. *Let $L/K/F$ be an extension of fields with L/F finite. Then $[K : F] \mid [L : F]$ and $[L : K] \mid [L : F]$. In particular, if $[L : F]$ is prime, then $K = L$ or $K = F$.*

If in the above $[L : F]$ is prime, we say that there exist no *nontrivial intermediate fields* in L/F.

Corollary 45.5. *If $F_1 \subset \cdots \subset F_n$ is a tower of fields with F_n/F_1 finite, then*

$$[F_n : F_1] = [F_n : F_{n-1}] \cdots [F_2 : F_1].$$

Given a field extension, we want to look at intermediate fields.

Lemma 45.6. *Let K be a field and S a non-empty subset of K. Then there exists a unique minimal subfield F_0 of K containing S, i.e., if $S \subset F \subset K$ with F a field, then $F_0 \subset F$.*

Proof. Let

$$\mathcal{F} = \{F \mid F \text{ a field satisfying } S \subset F \subset K\}.$$

As $K \in \mathcal{F}$, the set $\mathcal{F}$ is not empty. Set $F_0 = \bigcap_{\mathcal{F}} F$. Certainly, $S \subset F_0$. We show that F_0 works. We first prove that F_0 is a field. Indeed if $x, y \in F$ with $x \neq 0$, then x and y lie in any $F \in \mathcal{F}$, hence $x \pm y$, xy, x^{-1} also lie in any F in $\mathcal{F}$, establishing F_0 is a field. To finish, we show that F_0 is the unique minimal field containing F_0. But this is clear, for if $E \in \mathcal{F}$, then $F_0 \subset \bigcap_{\mathcal{F}} F \subset E$. □

Notation 45.7. Let K/F be a field extension and S a non-empty subset of K. Let $S = F \cup X$. The field F_0 in Lemma 45.6 will be denoted by $F(X)$. If $X = \{\alpha_1, \ldots, \alpha_n\}$, we shall write $F(\alpha_1, \ldots, \alpha_n)$ for $F(X)$. In particular, if $\alpha \in K$, then

$$F(\alpha) = \text{the unique minimal intermediate field of } K/F \text{ containing } \alpha.$$

Remark 45.8. Let K/F be field extension and X a non-empty subset of K. As usual $F[X]$ will denote the 'polynomials' in the $x \in X$ with coefficients in F, i.e., if we let t_x, with $x \in X$, be independent variables and $T = \{t_x \mid x \in X\}$, then $F[X]$ is the homomorphic image of the evaluation map $e : F[T] \to K$ given by $t_x \mapsto x$ for all x in X. As $F[X] \subset K$, the ring $F[X]$ is a domain. (So $\ker e$ is a prime ideal in $F[t_x]_X$.) By the universal property of quotient fields (Theorem 24.14), we must have $F \subset F[X] \subset qf(F[X]) \subset K$ and $qf(F[X]) \subset F(X)$. It follows by the lemma that $qf(F[X]) = F(X)$. In particular, if $x = \{\alpha\}$, then $F \subset F[\alpha] \subset F(\alpha) = qf(F[\alpha]) \subset K$, with all of these domains and, in fact, except for possibly $F[\alpha]$, all are fields.

Examples 45.9. 1. $\mathbb{R}[\sqrt{-1}] = \mathbb{R}(\sqrt{-1}) = \mathbb{C}$, as all are 2-dimensional real vector spaces.

2. Let $d \in \mathbb{Z}$, then
$$\mathbb{Q}[\sqrt{d}] = \mathbb{Q}(\sqrt{d}) = qf(\mathbb{Z}[\sqrt{d}])$$
a vector space over $\mathbb{Q}$ of dimension at most two and exactly two if $d \notin \{x^2 \mid x \in \mathbb{Z}\}$ in which case $\{1, \sqrt{d}\}$ is a $\mathbb{Q}$-basis.

3. If F is a field, then we have
$$F < F[t] < F(t) := qf(F[t]),\ F[t]^\times = F^\times, \text{ and } F(t)^\times = F(t) \setminus \{0\}.$$
The last follows as $F(t) = \{f/g \mid f, g \in F[t] \text{ with } g \neq 0\}$ is a field.

Definition 45.10. Let K/F be a field extension and α an element of K. We say

(1) α is *algebraic over* F if there exists a nonzero polynomial f in $F[t]$ satisfying $f(\alpha) = 0$, i.e., α is a root of a nonzero polynomial with coefficients in F (hence of a monic polynomial in $F[t]$).

(2) α is called a *transcendental element over* F if it is not algebraic over F.

Examples 45.11. Let F be a field.

1. Every element α in F is algebraic over F as α is a root of the polynomial $t - \alpha$ in $F[t]$.
2. Let a, b be elements in $\mathbb{Q}$ and $\alpha = a + b\sqrt{-1}$ in $\mathbb{C}$. Then α is a root of $t^2 - 2at + (a^2 + b^2)$ in $\mathbb{Q}[t]$, so algebraic over $\mathbb{Q}$.
3. Suppose $\mathbb{Q} \subset F \subset \mathbb{C}$. (It suffices to say $F \subset \mathbb{C}$ – why?) and $d \in \mathbb{Q}$. Then $F[\sqrt{d}] = F(\sqrt{d})$, e.g., $F = \mathbb{Q}$ or $\mathbb{R}$ Let α be an element in $F[\sqrt{d}]$. We can write $\alpha = a + b\sqrt{d}$ for some $a, b \in F$ (why?). Then α is a root of the polynomial $t^2 - 2at + (a^2 - b^2 d)$ in $F[t]$, hence is algebraic over F:

   If $\sqrt{d}$ lies in F, then $F = F[\sqrt{d}] = F(\sqrt{d})$, so suppose not, i.e., suppose that d is not a square in F. If $\alpha = a + b\sqrt{d}$ and $0 \neq \beta = c + e\sqrt{d}$ with $a, b, c, e \in F$, then
   $$\frac{\alpha}{\beta} = \frac{a + b\sqrt{d}}{c + e\sqrt{d}} = \frac{a + b\sqrt{d}}{c + e\sqrt{d}} \cdot \frac{c - e\sqrt{d}}{c - e\sqrt{d}} = \frac{(a + b\sqrt{d})(c + e\sqrt{d})}{c^2 - e^2 d}$$
   lies in $F[\sqrt{d}]$. (Note that $c^2 - e^2 d$ is not zero as d is not a square in F.) In this case, we also have $\dim_F F[\sqrt{d}] = 2$ with $\{1, \sqrt{d}\}$ an F-basis. In addition, the map
   $$\sigma : F(\sqrt{d}) \to F(\sqrt{d}) \text{ given by } a + b\sqrt{d} \mapsto a - b\sqrt{d} \text{ with } a, b \in F$$

is a field automorphism satisfying $\sigma(a) = a$ for all $a \in F$. We call σ an *F-automorphism*. In fact, $\sigma(\alpha) = \alpha$ for $\alpha \in F(\sqrt{d})$ if and only if α lies in F, so the set $\{\alpha \in F(\sqrt{d}) \mid \sigma(\alpha) = \alpha\}$, called the *fixed field* of σ, is precisely F. This means that σ *moves* every element in $F(\alpha) \setminus F$. It is also noteworthy to observe that $\{1, \sigma\}$ (with 1 the identity automorphism $1_{F(\sqrt{d})}$) is a group of order two which is precisely the F-dimension of $F(\sqrt{d})$.

4. Lindemann's Theorem says that the element π is transcendental over $\mathbb{Q}$. This is not easy (cf. Section §67). It is easier to show that the real number e is transcendental over $\mathbb{Q}$ (cf. Section §65), as is the Liouville number $\sum 1/10^{n!}$ (cf. Section §64). As $\mathbb{R}$ is uncountable and the set of roots of all nonzero polynomials over $\mathbb{Q}$ is countable, since there are countably many polynomials over $\mathbb{Q}$ each having finitely many roots, "most" elements in $\mathbb{R}$ are transcendental. However, it is usually very difficult to show that a specific real number is transcendental over $\mathbb{Q}$. It is still unknown if π^e is transcendental over $\mathbb{Q}$.
5. The real numbers π and e are algebraic over the field $\mathbb{R}$.
6. The complex number $e^{\sqrt{-1}\pi}$ is algebraic over $\mathbb{Q}$ as it is equal to -1, but it can be shown that the number e^{π} is transcendental over $\mathbb{Q}$.
7. Let $L/K/F$ be field extensions. If α is an element in L algebraic over F then it is algebraic over K, since $F[t] \subset K[t]$. However, in general, if α is algebraic over K it need not be algebraic over F.
8. Let K/F be a field extension and u an element in K that is transcendental over F. We know that the evaluation map $e_u : F[t] \to F[u]$ given by $f \mapsto f(u)$ is a ring epimorphism and an F-linear transformation. Suppose that f lies in $\ker e_u$. As u is transcendental over F, we must have $f = 0$, hence e_u is a ring monomorphism. As it is also surjective, $e_u : F[t] \xrightarrow{\sim} F[u]$ is an isomorphism. The variable t is not a unit in $F[t]$, since $F[t]^{\times} = F^{\times}$. Therefore, we also have $F[t] < F(t)$, so $F[u] < qf(F[u]) = F(u)$. In addition, as $\{t^i \mid i \geq 0\}$ is an F-basis for $F[t]$, we have $\{u^i \mid i \geq 0\}$ is an F-basis for $F[u]$. In particular, $F[u]$ and $F(u)$ are both infinite dimensional vector spaces over F.

   Note: By (4), this means that $\mathbb{Q}[\pi] \cong \mathbb{Q}[t]$. But $\mathbb{R}[\pi] = \mathbb{R} < \mathbb{R}[t]$, so being transcendental is a relative notion depending on the base field F.

The main theorem about an algebraic element over a field is given by the following foundational result, much of which we have implicitly seen before.

Theorem 45.12. *Let K/F be an extension of fields and α an element in K that is algebraic over F. Then there exists a unique monic irreducible polynomial $m_F(\alpha)$ in $F[t]$ satisfying all of the following:*

(1) α is a root of $m_F(\alpha)$.
(2) If α is a root of a nonzero polynomial g in $F[t]$, then $m_F(\alpha) \mid g$ in $F[t]$ and $\deg g \geq \deg m_F(\alpha)$ with equality if and only if $g = am_F(\alpha)$ for some nonzero element $a \in F$.
(3) Let $n = \deg m_F(\alpha)$. Then $F(\alpha) = F[\alpha]$ and $\{1, \alpha, \ldots, \alpha^{n-1}\}$ is an F-basis for $F[\alpha] = F(\alpha)$. In particular,

$$[F(\alpha) : F] = \dim_F F[\alpha] = \deg m_F(\alpha).$$

The polynomial $m_F(\alpha)$ in the theorem is called the *minimal* or *irreducible polynomial* of α *over* F. The integer $\deg m_F(\alpha) = [F(\alpha) : F]$ is called the *degree* of α *over* F.

Proof. Let $e_\alpha : F[t] \to F[\alpha]$ be given by $f \mapsto f(\alpha)$, evaluation at α, a ring epimorphism and F-linear transformation of vector spaces over F. As α is algebraic over F, α is the root of a nonzero polynomial in $F[t]$ which may be assumed to be monic. Therefore, there exists a positive integer N such that α^N lies in the F-vector space $\sum_{i=0}^{N-1} F\alpha^i$. It follows by induction that α^{N+j} lies in $\sum_{i=0}^{N-1} F\alpha^i$ for all $j \geq 0$, so $F[\alpha] = \sum_{i=0}^{N-1} F\alpha^i$ is a finite dimensional vector space over F. Since $F[t]$ is an infinite dimensional vector space over F, it follows that e_α cannot be a monomorphism, so $0 < \ker e_\alpha < F[t]$ ($e_\alpha(1) = 1$, so e_α is not the trivial map). Since $F[t]$ is euclidean, it is a PID, hence there exists a unique non-constant monic polynomial f in $F[t]$ such that $\ker e_\alpha = (f)$. (It is unique, as generators of $\ker e_\alpha$ only differ by a nonzero element in F.) By the First Isomorphism Theorem, $F[\alpha] = \operatorname{im} e_\alpha \cong F[t]/(f)$. Let $n = \deg f$.

Claim. $f = m_F(\alpha)$ works.

As $F[t]/(f) \cong F[\alpha] \subset K$, it is a domain, so $0 < (f) < F[t]$ is a prime ideal. Since $F[t]$ is a PID, it must be a maximal ideal, hence $F[t]/(f) \cong F[\alpha]$ is a field, so $F[\alpha] = F(\alpha)$.

We next show that $\mathcal{B} = \{1, \alpha, \ldots, \alpha^{n-1}\}$ is an F-basis for $F[\alpha]$. Let $g \in F[t]$. As $F[t]$ is euclidean (under the degree map), $g = fq + r$ for some $q, r \in F[t]$ with $r = 0$ or $\deg r < \deg f$. Since e_α is a ring homomorphism, $g(\alpha) = f(\alpha)q(\alpha) + r(\alpha) = r(\alpha)$, which lies in $\sum_{i=0}^{n-1} F\alpha^i$, the span of $\mathcal{B}$. Moreover, $g(\alpha) = 0$ if and only if $r(\alpha) = 0$. As $r = 0$ or $\deg r < \deg f$, we have $r(\alpha) = 0$ if and only if $r \in \ker e_\alpha$ if and only if $f \mid r$ in $F[t]$ if and only if $r = 0$. We conclude that if $g(\alpha) = \sum_{i=0}^{n-1} a_i\alpha^i = 0$ with $a_i \in F$, then each $a_i = 0$. Consequently, $\mathcal{B}$

is also F-linearly independent, hence an F-basis for $F[\alpha]$. The other needed properties of f now follow easily. □

Remarks 45.13. Let $K/E/F$ be field extensions and α in K an algebraic element over F.

1. $m_E(\alpha) \mid m_F(\alpha)$ in $E[t]$ and $\deg m_E(\alpha) \leq \deg m_F(\alpha)$ (Of course, they can be different.)
2. If $f \in F[t]$ is monic and irreducible with $f(\alpha) = 0$, then $f = m_F(\alpha)$.
3. If $g \in F[t]$ satisfies $\deg g = [F(\alpha) : F]$ and $g(\alpha) = 0$, then g is irreducible.

Proposition 45.14. (Characterization of algebraic elements) *Let K/F be a field extension and $\alpha \in K$. Then α is algebraic over F if and only if $[F(\alpha) : F]$ is finite.*

Proof. ($\Rightarrow$): This follows from the Theorem.

($\Leftarrow$): Suppose that $n = [F(\alpha) : F]$. Then $\{1, \alpha, \ldots, \alpha^n\}$ is F-linearly dependent. Therefore, there exist $a_i \in F$, $1 \leq i \leq n$, not all zero such that $\sum_{i=0}^{n} a_i \alpha^i = 0$, so α is a root of the nonzero polynomial $\sum_{i=0}^{n} a_i t^i$ in $F[t]$. □

Corollary 45.15. *Let K/F be a finite field extension. Then every element in K is algebraic over F.*

Proposition 45.16. (Characterization of transcendental elements) *Let K/F be a field extension and $u \in K$. Then the following are equivalent:*

(1) u is transcendental over F.
(2) $F[u] \cong F[t]$.
(3) $F[u] < F(u)$.
(4) $[F(u) : F]$ is infinite.
(5) $\dim_F F(u)$ is infinite.

Proof. By Example 45.11(8), we know that (1) implies (2), (3), (4), and (5), and if (2), (3), (4), or (5) holds, then u cannot be algebraic and (1) must also hold. □

Corollary 45.17. *Let K/F be a field extension. If $a_1, \ldots, a_r$ are elements of K algebraic over F, then $F[a_1, \ldots, a_r] = F(a_1, \ldots, a_r)$ and $[F(a_1, \ldots, a_r) : F] \leq \prod_{i=1}^{r} [F(a_i) : F]$ (with strict inequality possible).*

Proof. As a_r is algebraic over F, it is algebraic over $F(a_1, \ldots, a_{r-1})$. By induction and the $r = 1$ case (that we have done), $F[a_1, \ldots, a_r] =$

$F[a_1,\dots,a_{r-1}][a_r] = F(a_1,\dots,a_{r-1})[a_r] = F(a_1,\dots,a_r)$ and

$$\begin{aligned}[F(a_1,\dots,a_r):F] &\le [F(a_1,\dots,a_r):F(a_1,\dots,a_{r-1})]\prod_{i=1}^{r-1}[F(a_i):F]\\ &= \deg m_{F(a_1,\dots,a_{r-1})}(a_r)\prod_{i=1}^{r-1}[F(a_i):F]\\ &\le \prod_{i=1}^{r}[F(a_i):F].\end{aligned}$$

□

Corollary 45.18. *Let A be a domain containing a field F. If A is a finite dimensional vector space over F, then A is a field.*

PROOF. If A has an F-basis $\mathcal{B}$, then $A = F[\mathcal{B}]$ and its quotient field $= K = F(\mathcal{B})$ is a finite dimensional vector space over F on basis $\mathcal{B}$. In particular, $A = K$. □

If K/F is a field extension, we arrive at the important conclusion that the set of elements in K algebraic over F form an intermediate field.

Theorem 45.19. *Let K/F be a field extension with a, b elements in K algebraic over F. Then $a \pm b$, ab and b^{-1}, if $b \neq 0$, are algebraic over F. In particular, the set of elements in K algebraic over F forms an intermediate field of K/F.*

PROOF. Let $\gamma = a \pm b$, ab or b^{-1}, if $b \neq 0$. Then $F \subset F(\gamma) \subset F(a,b)$, so

$$[F(\gamma):F] \le [F(a,b):F] \le [F(a):F][F(b):F] < \infty.$$

The result follows. □

Definition 45.20. Let K/F be a field extension. We say that it is an *algebraic extension* (or simply *algebraic*) if every element of K is algebraic over F. If K/F is not algebraic, we say that it is a *transcendental extension*.

We have shown that every finite field extension is algebraic. The converse is false, as we shall see below. First, however, we show that being an algebraic extension is transitive. In fact, we have

Theorem 45.21. *Let $L/K/F$ be field extensions. Then L/F is algebraic if and only if both L/K and K/F are algebraic.*

PROOF. ($\Rightarrow$): Any element in L algebraic over F is clearly algebraic over K and any element in K lies in L so is algebraic over F.

($\Leftarrow$): Let α be an element of L. We must show that α is algebraic over F. As it is algebraic over K, the minimal polynomial $m_K(\alpha) = \sum_{i=0}^{n} c_i t^i$ in $K[t]$ exists. Let $E = F(c_0, \ldots, c_{n-1}) \subset K$. Since the elements $c_0, \ldots, c_{n-1}$ are algebraic over F, we have $[E : F] < \infty$. As $m_K(\alpha) \in E[t]$, we also have $[E(\alpha) : E] < \infty$, so

$$[F(\alpha) : F] \leq [E(\alpha) : F] = [E(\alpha) : E][E : F] < \infty$$

and α is algebraic over F. □

The above trick is very useful. If you want to prove a result and can reduce to a finite amount of data, you can often reduce to checking in a nicer set, e.g., in the above we reduced from an algebraic extension to a finite extension, because we needed to attach only finitely many elements algebraic over the base field. In commutative ring theory, a finite amount of data over a commutative ring, often reduces a proof to a Noetherian ring containing it by the Hilbert Basis Theorem.

Examples 45.22. 1. $\mathbb{C}/\mathbb{R}$ and $\mathbb{Q}(\sqrt{d})/\mathbb{Q}$, with d an integer, are algebraic extensions.

2. $\mathbb{R}/\mathbb{Q}$ is transcendental as $\mathbb{R}$ is uncountable (by Cantor's Theorem) and $\mathbb{Q}$ is countable.

3. Let

$$\Omega := \{\alpha \in \mathbb{C} \mid \alpha \text{ algebraic over } \mathbb{Q}\}.$$

Ω is a countable field as it is the set of roots of countably many nonzero polynomials, each having finitely many roots.

Claim. $\Omega/\mathbb{Q}$ is not finite:

Let p be a (positive) prime in $\mathbb{Z}$ and n a positive integer. By Eisenstein's Criterion (32.11(1)), $t^n - p$ in $\mathbb{Q}[t]$ is irreducible and has a root α in Ω. In particular, $[\Omega : \mathbb{Q}] \geq [\mathbb{Q}(\alpha) : \mathbb{Q}] = \deg m_{\mathbb{Q}}(\alpha) = \deg(t^n - p) = n$, establishing the claim.

4. Ω in (3) is *algebraically closed*, i.e., every non-constant polynomial in $\Omega[t]$ has a root in Ω. To see this, we use the Fundamental Theorem of Algebra (to be proved in 54.12 below) that $\mathbb{C}$ is algebraically closed. If f is a non-constant polynomial in $\Omega[t]$, it has a root α in $\mathbb{C}$, so $\Omega(\alpha)/\Omega$ is algebraic. As $\Omega/\mathbb{Q}$ is algebraic, it follows that so is $\Omega(\alpha)/\mathbb{Q}$, hence α is algebraic over $\mathbb{Q}$ hence is in Ω. As $\mathbb{C}$ is uncountable, we must have $\Omega < \mathbb{C}$.

5. The same argument as in (4) shows if K/F with K algebraically closed and Ω the set of elements of K that are algebraic over F,

then Ω is an algebraically closed field. Moreover, by definition, if $F < E < \Omega$ is a field, then E is not algebraically closed. We call Ω an *algebraic closure* of F. We shall see in §48, that all such algebraic closures of a field F are unique, up to a field isomorphism fixing F.

Calculation 45.23. 1. Let $p \in \mathbb{Z}^+$ be a prime, then the polynomial $f = t^{p-1} + t^{p-2} + \cdots + t + 1$ is irreducible in $\mathbb{Q}[t]$ by the application of Eisenstein's Criterion in 32.11(3). Let ζ_p be a root of f in $\mathbb{C}$. Then $f = m_{\mathbb{Q}}(\zeta_p)$, so $[\mathbb{Q}(\zeta_p) : \mathbb{Q}] = \deg f = p - 1 = \varphi(p)$ (where φ is the Euler phi-function), and $\{1, \zeta_p, \ldots, \zeta_p^{p-2}\}$ is a $\mathbb{Q}$-basis for $\mathbb{Q}(\zeta_p)$.

(Note that $t^p - 1 = (t-1)(t^{p-1} + t^{p-2} + \cdots t + 1)$ is a factorization into irreducibles in $\mathbb{Q}[t]$.)

2. Let α in $\mathbb{C}$ be a root of the irreducible polynomial $f = t^3 - 2t + 2$ in $\mathbb{Q}[t]$ (irreducible by Eisenstein's Criterion). So $[\mathbb{Q}(\alpha) : \mathbb{Q}] = \deg f = \deg m_{\mathbb{Q}}(\alpha) = 3$ and $\mathcal{B} = \{1, \alpha, \alpha^2\}$ is a $\mathbb{Q}$-basis for $\mathbb{Q}(\alpha)$. As $\alpha^3 = 2\alpha - 2$ in $\mathbb{Q}(\alpha)$, we have $-\alpha^{-1} = (\alpha^2/2) - 1$ in $\mathbb{Q}(\alpha)$. Let $\beta \in \mathbb{Q}(\alpha)$, so $\mathbb{Q}(\beta) \subset \mathbb{Q}(\alpha)$, hence $[\mathbb{Q}(\beta) : \mathbb{Q}] \mid [\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$. Therefore, either $\beta \in \mathbb{Q}$ or $\mathbb{Q}(\beta) = \mathbb{Q}(\alpha)$. For example, suppose that $\beta = \alpha^2 - \alpha$. As $\mathcal{B}$ is $\mathbb{Q}$-linearly independent, $\beta \notin \mathbb{Q}$, so $\mathbb{Q}(\beta) = \mathbb{Q}(\alpha)$ and $\mathcal{C} = \{1, \beta, \beta^2\}$ is a $\mathbb{Q}$-basis for $\mathbb{Q}(\beta)$ and $\{1, \beta, \beta^2, \beta^3\}$ is $\mathbb{Q}$-linearly dependent. Computation shows that

$$1 = 1, \quad \beta = \alpha^2 - \alpha, \quad \beta^2 = 3\alpha^2 - 6\alpha + 4$$

and

$$(*) \qquad \beta^3 = 16\alpha^2 - 28\alpha + 18.$$

It follows that the change of basis matrix for the basis $\mathcal{C}$ to the basis $\mathcal{B}$ is

$$[Id]_{\mathcal{C},\mathcal{B}} = \begin{pmatrix} 1 & 0 & 4 \\ 0 & -1 & -6 \\ 0 & 1 & 3 \end{pmatrix}.$$

Inverting this matrix to $[Id]_{\mathcal{B},\mathcal{C}}$ then shows that

$$1 = 1, \quad \alpha = -\frac{\beta^2}{3} + \beta + \frac{4}{3}, \quad \alpha^2 = -\frac{\beta^2}{3} + 2\beta + 4 + \frac{4}{3},$$

and substituting this into (*) yields $\beta^3 - 4\beta^2 - 4\beta - 2 = 0$, i.e., β is a root of monic $g = t^3 - 4t^2 - 4t - 2$ in $\mathbb{Q}[t]$. This must be the irreducible polynomial $m_{\mathbb{Q}}(\beta)$ as g has the same degree and satisfies $g(\beta) = 0$. [Of course, we also know that it is irreducible by Eisenstein's Criterion.]

3. It is easy to check that $[\mathbb{Q}(\sqrt{2},\sqrt{3}) : \mathbb{Q}] = 4$ with $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ a $\mathbb{Q}$-basis for $\mathbb{Q}(\sqrt{2},\sqrt{3})$ (by showing that $\sqrt{3}$ cannot lie in $\mathbb{Q}(\sqrt{2})$). (Cf. Exercise 47.18(12) for the generalization.) Using this we determine $[\mathbb{Q}(\sqrt{2}+\sqrt{3}) : \mathbb{Q}]$ and $m_{\mathbb{Q}}(\sqrt{2}+\sqrt{3})$. Let $K = \mathbb{Q}(\sqrt{2}+\sqrt{3}) \subset \mathbb{Q}(\sqrt{2},\sqrt{3})$. So $[\mathbb{Q}(\sqrt{2}+\sqrt{3}) : \mathbb{Q}] = 1, 2$ or 4. Let $\alpha = \sqrt{2}+\sqrt{3}$. Then
$$1 = 1, \quad \alpha = \sqrt{2}+\sqrt{3}, \quad \alpha^2 = 5 + 2\sqrt{6}.$$
These must be $\mathbb{Q}$-linearly independent (why?). It follows that the set $\{1, \alpha, \alpha^2, \alpha^3\}$ must be a $\mathbb{Q}$-basis for K. Now $\alpha^2 = 5+2\sqrt{6}$, so the polynomial $g = t^2 - (5 + 2\sqrt{6})$ lying in $\mathbb{Q}(\sqrt{6})[t]$ has a root α in K with $\alpha \notin \mathbb{Q}(\sqrt{6})$. Consequently, $[K : \mathbb{Q}(\sqrt{6})] = 2$ and $g = m_{\mathbb{Q}(\sqrt{6})}(\alpha)$. Computation shows that
$$h = (t^2 - \big(5 + 2\sqrt{6}\big)\big(t^2 - (5 - 2\sqrt{6})\big) = t^4 - 10t^2 + 1,$$
so it lies in $\mathbb{Q}[t]$ and has α as a root in K with $\deg h = \deg m_{\mathbb{Q}}(\alpha) = 4$. Thus $h = m_{\mathbb{Q}}(\alpha)$.

We shall show later that if F is a field of characteristic zero and K/F a finite extension, that there always exists a θ in K such that $K = F(\theta)$. (Cf. Theorem 54.9.) In general, such a θ is hard to find, although many different ones work.

Exercises 45.24. 1. Let R be a domain containing a field F and over which it is a finite dimensional vector space. Show that R is a field. In particular, show if D is a division ring containing F and $\alpha \in D$ is a root of a nonzero polynomial in $F[t]$, then $F[\alpha]$ is a field. (Cf. A finite division ring is a field.)

2. (a) Find $u \in \mathbb{R}$ such that $\mathbb{Q}(u) = \mathbb{Q}(\sqrt{2},\sqrt[3]{5})$.
   (b) Describe how you how you would find all $w \in \mathbb{Q}(\sqrt{2},\sqrt[3]{5})$ such that $\mathbb{Q}(w) = \mathbb{Q}(\sqrt{2},\sqrt[3]{5})$.
3. If $a, b \in K$ are algebraic over F are of degree m, n respectively with m, n relatively prime, show that $[F(a, b) : F] = mn$.
4. Show that $[\mathbb{Q}(\sqrt{2},\sqrt{3}) : \mathbb{Q}] = 4$ with $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ a $\mathbb{Q}$-basis for $\mathbb{Q}(\sqrt{2},\sqrt{3})$.
5. If $|F| = q < \infty$ show
   (a) There exists a prime p such that char $F = p$.
   (b) $q = p^n$ some n.
   (c) $a^q = a$ for all $a \in F$.
   (d) If $b \in K$ is algebraic over F then $b^{q^m} = b$ for some $m > 0$.
6. Let u be a root of $f = t^3 - t^2 + t + 2$ in $\mathbb{Q}[t]$ and $K = \mathbb{Q}(u)$.
   (a) Show that $f = m_{\mathbb{Q}}(u)$.

(b) Express $(u^2+u+1)(u^2-u)$, and $(u-1)^{-1}$ in the form au^2+bu+c, for some $a, b, c \in \mathbb{Q}$.

7. Let $\zeta = \cos\frac{\pi}{6} + \sqrt{-1}\sin\frac{\pi}{6}$ in $\mathbb{C}$. Show that $\zeta^{12} = 1$ but $\zeta^r \neq 1$ for $1 \leq r < 12$. Show also that $[\mathbb{Q}(\zeta) : \mathbb{Q}] = 4$ and find $m_{\mathbb{Q}}(\zeta)$.
8. Let $K = F(u)$, u algebraic over F and degree of u odd. Show that $K = F(u^2)$.
9. Let u be transcendental over F and $F < E \subset F(u)$. Show that u is algebraic over E.
10. Let $f, g \in F[t]$ be relatively prime and suppose that $u = f/g$ lies in $F(t) \setminus F$. Show that $F(t)/F(u)$ is finite of degree $d = \max\{\deg(f), \deg(g)\}$
11. If $f = t^n - a \in F[t]$ is irreducible and $u \in K$ is a root of f and $n/m \in \mathbb{Z}$, show that $[F(u^m) : F] = \dfrac{n}{m}$. What is $m_F(u^m)$?
12. If a^n is algebraic over a field F for some $n > 0$, show that a is algebraic over F.

46. Addendum: Transcendental Extensions

What can we say about a field extension of a field that is not algebraic, i.e., is transcendental? In this short section, we answer this question, leaving most of the details to the reader.

Definition 46.1. Let K/F be an extension of fields and suppose that $\mathcal{B} = \{x_i \mid i \in I\}$ a subset of K. We say that the set $\mathcal{B}$ is *algebraically independent* over F if $x_i \neq x_j$ for $i \neq j$ and if $f \in F[\mathcal{B}]$ satisfies $f(x_i)_I = 0$, then $f = 0$. This is equivalent to

$$\{x_{i_1}^{n_{i_1}} \cdots x_{i_r}^{n_{i_r}} \mid n_{i_j} \geq 0 \text{ for all } i_j \text{ in } I, \text{ some } r \geq 0\}$$

is an F-linearly independent set for the F-vector space $F[\mathcal{B}]$.

For example, a set of indeterminants $T = \{t_i \mid i \in I\}$ is an algebraically independent set in the quotient field $F(T)$ of $F[T]$. A maximal algebraically independent subset $\mathcal{B}$ in an extension field K of F is called a *transcendence basis* for K over F.

Remarks 46.2. Let K/F be an extension of fields.

1. There exists a transcendence basis for K over F. If K/F is algebraic, it is the empty set. A proof of this is analogous to the proof of the existence of bases for vector spaces.
2. Is $\mathcal{B}$ is a transcendence basis for K over F then $K/F(\mathcal{B})$ is algebraic.
3. The set $\{t_1, \ldots, t_n\}$ is a transcendence basis for $F(t_1, \ldots, t_n)$ over F.

4. The set $\mathcal{B} = \{x_i\}_I \subset K$ is a transcendence basis for K over F if and only if $\mathcal{B}$ is a maximal subset of K satisfying $F(\mathcal{B}) \cong F(\{t_i\}_I)$.
5. The extension K/F is called *finitely generated* if there exists a finite set $S \subset K$ such that $K = F(S)$. If K/F is finitely generated, then K has a finite transcendence basis over F.

The analogue of dimension of vector spaces holds for the cardinality of transcendence bases. The proof for the case that K/F finitely generated is analogous to the so called Replacement Theorem used to prove that dimension is well-defined for finite dimensional vector spaces. We shall only prove this case in some detail.

Proposition 46.3. *Let K/F be an extension of fields. Then any two transcendence bases of K over F have the same cardinality.*

Proof. Although this follows from the fact that the cardinality of any basis for a fixed vector space is well-defined, we present a proof in the case that K/F is a field extension and K has a finite transcendence basis $\{x_1, \ldots, x_n\}$ over F. Suppose that $\{y_1, \ldots, y_m\}$ is a subset of K algebraically independent over F. We must show $m \leq n$. As $\{x_1, \ldots, x_n\}$ is a maximal algebraically independent set over F, the set $\{y_1, x_1, \ldots, x_n\}$ must be *algebraically dependent*, i.e., is not algebraically independent. Therefore, there exists a nonzero polynomial $f \in F[t_0, t_1, \ldots, t_n]$ satisfying $f(y_1, x_1, \ldots, x_n) = 0$. By assumption, y_1 occurs non-trivially in f as well as some x_i, say x_1. Then x_1 is algebraic over $F(y_1, x_2, \ldots, x_n)$. Hence both $K/F(x_1, \ldots, x_n)$ and $F(y_1, x_1, x_2, \ldots, x_n)$ are algebraic over $F(y_1, x_2, \ldots, x_n)$. It follows that $K/F(y_1, x_2, \ldots x_n)$ is algebraic. A subset $\mathcal{B}$ of $\{y_1, x_2, \ldots, x_n\}$ must be a transcendence basis of K over F and it must include y_1, for if $\mathcal{B} < \{y_1, x_2, \ldots, x_n\}$, reversing the argument, would show that a proper subset of $\{x_1, \ldots, x_n\}$ would be a transcendence basis for K over F. It follows that $\{y_1, x_2, \ldots, x_n\}$ is a transcendence basis for K over F. The argument continues, as in the proof of the Replacement Theorem for vector spaces, by throwing in y_2 and seeing that we have to throw away some x_i in $\{x_2, \ldots, x_n\}$, etc. It follows that $m \leq n$. □

Let K/F be a field extension, then the cardinality of a transcendence basis for K over F is called the *transcendence degree* of K/F and denoted by $\operatorname{tr\,deg}_F K$. It follows that K/F is finitely generated if and only if $\operatorname{tr\,deg}_F K$ is finite.

Proposition 46.4. *Let $L/K/F$ be field extensions. Then* $\operatorname{tr\,deg}_F L$ *is finite if and only if both* $\operatorname{tr\,deg}_F K$ *and* $\operatorname{tr\,deg}_K L$ *are finite. Moreover,*

if tr $\deg_F L$ *is finite, then*

$$\text{tr } \deg_F L = \text{tr } \deg_K L + \text{tr } \deg_F K$$

Exercises 46.5. 1. Prove that if K/F is an extension of fields, then there exists a transcendence basis for K over F.

2. If K/F is an extension of fields, prove that $\mathcal{B} = \{x_i\}_I \subset K$ is a transcendence basis for K/F if and only if $\mathcal{B}$ is a maximal subset of K satisfying $F(\mathcal{B}) \cong F(\{t_i\}_I)$. and $F[\mathcal{B}] \cong F[\{t_i\}_I]$.

3. Prove Prop 46.4.

47. Splitting Fields

We have shown in Kronecker's Theorem 31.12 that an irreducible polynomial over a field F has a root in an extension field. Indeed we construct such as follows:

Let F be a field and f an irreducible polynomial in $F[t]$. As $F[t]$ is a PID, the non-trivial prime ideal (f) in $F[t]$ is maximal, so $K = F[t]/(f)$ is a field. Let $^- : F[t] \to K$ be the canonical epimorphism given by $g \mapsto g + (f)$. The composition $F \subset F[t] \to K$ is monic, since F is simple, so we can, and do, view this as an inclusion of F into K, i.e., we identify a in F and $\overline{a}$ in K, so K/F is a field extension. Under this identification,

$$h := \sum_i a_i t^i \mapsto \overline{h} = \sum \overline{a_i}\overline{t}^i = \sum a_i \overline{t}^i = h(\overline{t})$$

is evaluation at $\overline{t}$. In particular, $\overline{0} = \overline{f} = f(\overline{t})$ in K, so $f \in F[t] \subset K[t]$ has a root in K.

Continuing, if g lies in $F[t]$ then, as $F[t]$ is euclidean, $g = fq + r$ in $F[t]$ for some $q, r \in F[t]$ with $r = 0$ or $\deg r < \deg f$. Therefore, $\overline{g} = \overline{f}\overline{q} + \overline{r} = \overline{r}$. It follows that if $n = \deg f$, then $\mathcal{B} = \{1, \overline{t}, \ldots, \overline{t}^{n-1}\}$ spans the vector space K over F. As $\overline{g} = 0$ in K if and only $f \mid g$ in $K[t]$, we have $g = 0$ or $\deg f \leq \deg g$. It follows that $\mathcal{B}$ is an F-basis for K (why?), $f = (\text{lead } f)m_F(\overline{t})$, and $n = \deg f = \deg m_F(\overline{t}) = [K : F]$.

So we have proved

Proposition 47.1. *Let F be a field and f an irreducible polynomial in $F[t]$. Then there exists an extension field K/F of degree* $\deg f$ *such that f has a root in K.*

Generalizing (as we did before), we have:

Theorem 47.2. (Kronecker's Theorem) *Let F be a field and f a non-constant polynomial in $F[t]$. Then there exists an algebraic extension K/F such that f has a root in K and $[K : F] \leq \deg f$.*

PROOF. Let $f_1 \mid f$ in the UFD $F[t]$ with f_1 irreducible. Then f has a root in the field $K = F[t]/(f_1)$, as f_1 does, with $[K : F] = \deg f_1 \leq \deg f$. □

Definition 47.3. Let K/F be an extension of fields and f a non-constant polynomial in $F[t]$. We say that f *splits over* K if f factors into a product of linear polynomials in $K[t]$ and that K is a *splitting field* of f *over* F or K/F is a *splitting field* of f if K is a minimal field extension of F such that f splits over K, i.e.,

(1) f splits over K.
(2) Either $K = F$ or whenever $K/E/F$ are field extensions with $E < K$, then f does not split over E.

The existence of splitting fields of a non-constant polynomial now will follow by an easy induction using Kronecker's Theorem.

Theorem 47.4. *Let F be a field and f a non-constant polynomial in $F[t]$. Then there exists a splitting field K of f over F satisfying $[K : F] \leq (\deg f)!$.*

PROOF. Let $n = \deg f$. By Kronecker's Theorem, there exists a field extension K_1/F such that $f = (t - \alpha_1)f_1$ in $K_1[t]$ with $f_1 \in K_1[t]$ and $[K_1 : F] \leq n$. By induction, there exists K_2 a splitting field of f_1 over K_1 satisfying $[K_2 : K_1] \leq (\deg f_1)! = (n-1)!$. Thus $[K_2 : F] = [K_2 : K_1][K_1 : F] \leq n!$ and f splits over K_2. By well-ordering, there exists an intermediate field $K_2/K/F$ with K a splitting field of f over F and $[K : F] \leq n!$. (Is $K = K_2$?) □

Remarks 47.5. Let F be a field and f a non-constant polynomial in $F[t]$ of degree n, K a splitting field of f over F, and $\alpha_1, \ldots, \alpha_n$ the roots of f in K. (They may not be distinct.)

1. We have $K = F(\alpha_1, \ldots, \alpha_n)$ and the proof of the theorem shows that $[K : F] \leq (\deg f)! = n! = |S_n|$. One often calls the splitting field K of f over F the *root field* of f *over* F for this reason.
2. Let L/F be a field extension with both K and L lying in some larger field. Then $L(\alpha_1, \ldots, \alpha_n)$ is a splitting field of f over L. Of course, it is possible that $L = L(\alpha_1, \ldots, \alpha_n)$.
3. If $K/L/F$ are field extensions, then K is a splitting field of f over L.
4. If $g \mid f$ in $F[t]$ with g a non-constant polynomial, then g splits over K, hence there exists an intermediate field $K/E/F$ such that E is a splitting field of g over F.

We turn to uniqueness statements for splitting fields. We shall do this in greater generality, so we need some notation.

Definition 47.6. Let $R_1 \subset S_1$ and $R_2 \subset S_2$ be commutative rings and $\varphi : R_1 \to R_2$ and $\tau : S_1 \to S_2$ be ring homomorphisms. We say that τ *lifts* or *extends* φ if $\tau|_{R_1} = \varphi$, i.e, we have a commutative diagram:

$$\begin{array}{ccc} S_1 & \xrightarrow{\tau} & S_2 \\ {\scriptstyle inc}\uparrow & & \uparrow{\scriptstyle inc} \\ R_1 & \xrightarrow{\varphi} & R_2. \end{array}$$

where *inc* is the inclusion. If $R = R_1 = R_2$, $\varphi = 1_R$, and τ lifts φ, we call τ an *R-algebra homomorphism*. If, in addition, τ is an isomorphism (respectively, monomorphism, epimorphism), then we call τ an *R-algebra isomorphism* (respectively, *R-algebra monomorphism, R-algebra epimorphism*). (Cf. Remark 23.11.) If $S = S_1 = S_2$ and τ is an automorphism, we call it an *R-algebra automorphism*. As is standard, if $R = F$ is a field, we call such an F-algebra homomorphism (respectively, F-algebra monomorphism, F-algebra epimorphism, F-algebra isomorphism, F-algebra automorphism (when appropriate) just an *F-homomorphism* (respectively, *F-monomorphism, F-epimorphism, F-isomorphism, F-automorphism*). Of course, an F-algebra homomomorphism is also a F-vector space linear transformation, but the converse is not true in general. Since these standard notations conflict with our usage of a module homomorphism if F is a field, a F-vector space homomorphism will be called an *F-linear homomorphism* or an *F-linear transformation* from now on.

Examples 47.7. 1. Let $\varphi : R \to S$ be a ring homomorphism of commutative rings. Define $\tilde{\varphi} : R[t] \to S[t]$ by $\sum a_i t^i \to \sum \varphi(a_i) t^i$, a ring homomorphism extending φ. *Given such a φ, we shall always let $\tilde{\varphi}$ denote this map.* If φ is a monomorphism, respectively an epimorphism, isomorphism, then $\tilde{\varphi}$ is a monomorphism, respectively an epimorphism, isomorphism.

2. Complex conjugation $\bar{} : \mathbb{C} \to \mathbb{C}$ given by $a + b\sqrt{-1} \mapsto a - b\sqrt{-1}$, for $a, b \in \mathbb{R}$ is an $\mathbb{R}$-automorphism. In fact, it is an $\mathbb{R}$-*involution*, i.e., an automorphism of order two.

3. If F is a field, as we always view $F \subset F[t]/(f)$ for any non-constant polynomial f in $F[t]$, $F[t]/(f)$ is an F-algebra. (By the definition of R-algebra A given in Remark 23.11, we only need a ring homomorphism of a commutative ring R into the center of a ring A.)

To prove a uniqueness statement for splitting fields, we begin with two lemmas.

Lemma 47.8. *Let $\varphi : F \to F'$ be a field isomorphism, f an irreducible polynomial in $F[t]$ and $f' = \tilde{\varphi}(f)$. Then*

(1) f' is an irreducible polynomial in $F'[t]$.
(2) There exists a field isomorphism $\tau : F[t]/(f) \to F'[t]/(f')$ extending φ and taking the root $\overline{t} = t + (f)$ in $F[t]/(f)$ to the root $\overline{t} = t + (f')$ in $F'[t]/(f')$.

PROOF. (1) is clear, so both $F[t]/(f)$ and $F'[t]/(f')$ are fields. Let $^- : F'[t] \to F'[t]/(f')$ be the canonical epimorphism and μ the composition

$$F[t] \xrightarrow{\tilde{\varphi}} F'[t] \xrightarrow{-} F'[t]/(f').$$

As φ is an isomorphism, so is $\tilde{\varphi}$. In particular, μ is epic. By definition, $\ker {}^- = (f')$, so $\tilde{\varphi}$ an isomorphism implies that

$$\begin{aligned}\ker \mu &= \{h \in F[t] \mid \mu(h) = 0\} = \{h \in F[t] \mid \tilde{\varphi}(h) \in (f')\} \\ &= \{h \in F[t] \mid h \in (f)\} = (f).\end{aligned}$$

Therefore, μ induces an isomorphism $\overline{\mu} : F[t]/(f) \to F'[t]/(f')$ given by $g + (f) \mapsto \tilde{\varphi}(g) + (f')$ by the First Isomorphism Theorem. As $\overline{\mu}|_F = \mu|_F = \varphi$ and $\overline{\mu}\big(t + (f)\big) = t + (f')$, the map $\tau = \overline{\mu}$ works. □

Lemma 47.9. *Let K/F be an extension of fields and f an irreducible polynomial in $F[t]$. If α in K is a root of f, then there exists an F-isomorphism*

$$\theta : F[t]/(f) \to F(\alpha) \text{ given by } t + (f) \mapsto \alpha.$$

PROOF. Let $e_\alpha : F[t] \to F[\alpha]$ by $g \mapsto g(\alpha)$ be the evaluation map at α, a ring epimorphism, and, in fact, an F-epimorphism. As $f(\alpha) = 0$, the element α is algebraic over F, so, as before, e_α induces an F-isomorphism $\theta : F[t]/\big(m_F(\alpha)\big) \to F[\alpha] = F(\alpha)$. As f is irreducible with root α, we have $(f) = \big(m_F(\alpha)\big)$, so θ works. □

By combining these two lemmas, we shall easily deduce the following:

Proposition 47.10. *Let $\varphi : F \to F'$ be a field isomorphism, f an irreducible polynomial in $F[t]$ and $f' = \tilde{\varphi}(f) \in F'[t]$. Suppose that K/F is a field extension with α a root of f in K and K'/F' a field extension with α' a root of f' in K'. Then there exists an isomorphism $\sigma : F(\alpha) \to F'(\alpha')$ extending φ and satisfying $\sigma(\alpha) = \alpha'$.*

PROOF. The two lemmas determine isomorphisms

$$F(\alpha) \xleftarrow[\text{lifts } 1_F]{\theta} F[t]/(f) \xrightarrow[\text{lifts } \varphi]{\tau} F'[t]/(f') \xrightarrow[\text{lifts } 1_{F'}]{\theta'} F'(\alpha')$$

with $\theta\big(t+(f)\big)=\alpha$, $\tau\big(t+(f)\big)=t+(f')$, and $\theta'\big(t+(f')\big)=\alpha'$. The map $\theta'\circ\tau\circ\theta^{-1}$ works. □

The following important corollary follows immediately:

Corollary 47.11. *Let K/F be an extension of fields, f an irreducible polynomial in $F[t]$ and α, β two roots of f in K. Then the map*

$$\sigma : F(\alpha) \to F(\beta) \text{ given by } \sum c_i\alpha^i \mapsto \sum c_i\beta^i$$

is an F-isomorphism.

The previous proposition is the first step in the induction proof of the result to which we have been striving.

Theorem 47.12. *Let $\varphi : F \to F'$ be an isomorphism of fields, f a non-constant polynomial in $F[t]$ and $f' = \tilde{\varphi}(f) \in F'[t]$. Suppose that E is a splitting field of f over F and E' is a splitting field of f' over F'. Then there exists an isomorphism $\tau : E \to E'$ extending φ and taking the set of roots of f in E bijectively onto the set of roots of f' in E'.*

An immediate consequence of this theorem is the result that we had wished to prove.

Corollary 47.13. (Uniqueness of Splitting Fields) *Let F be a field and f a non-constant polynomial in $F[t]$. If E and E' are splitting fields of f over F, then there exists an F-isomorphism $E \to E'$ taking the set of roots of f in E bijectively onto the set of roots of f in E'. In particular, a splitting field of f over F is unique up to an F-isomorphism.*

Of course, one would like to say given a polynomial there is a unique splitting field. We could do this if we assume that all the fields in which we work lie in some larger field Ω. Then one could say that a field extension E/F is *the* splitting field of a non-constant polynomial f in $F[t]$, as the splitting field of f is uniquely determined by the roots of f in Ω. For example, we view all splitting fields of rational polynomials to lie in $\mathbb{C}[t]$.

Proof. (of the theorem.) We induct on $n = [E : F]$.

$n = 1$: We have $E = F$, so f splits over F, hence f' splits over F'. In particular, if $f = a\prod(t-\alpha_i)$ in $F[t]$, then $f' = \varphi(a)\prod\big(t-\varphi(\alpha_i)\big)$ in $F'[t]$, and the result follows.

$n > 1$: If f does not split over F, then $f = gh$, for some polynomials g and h in $F[t]$, with g an irreducible polynomial satisfying $\deg g > 1$. As $\tilde{\varphi} : F[t] \to F'[t]$ is an isomorphism, $f' = g'h'$ with $g' = \tilde{\varphi}(g)$

irreducible of degree greater than one, and $h' = \tilde{\varphi}(h)$. Let $\alpha \in E$ be a root of g in E and $\alpha' \in E'$ be a root of g' in E' (which exist as f, f' split over E, E', respectively). By Proposition 47.10, there exists an isomorphism $\tau_1 : F(\alpha) \to F'(\alpha')$ extending φ and mapping α to α'. Let $f = (t - \alpha)f_1$ in $F(\alpha)[t]$, then $\tilde{\varphi}(f) = (t - \alpha')f_1'$ in $F'(\alpha')[t]$ with $f_1' = \tilde{\tau_1}(f_1)$ and $\tilde{\varphi}(f) = \tilde{\tau_1}(f)$.

We know that E is a splitting field of f_1 over $F(\alpha)$ and E' is a splitting field of f_1' over $F(\alpha')$. Since $[E : F(\alpha)] < [E : F]$, induction provides an isomorphism $\tau : E \to E'$ lifting τ_1, hence φ, and taking the set of roots of f_1 bijectively onto the set of roots of f_1'. The result now follows. □

The proof above filled in the maps in the following picture:

$$\begin{array}{ccc} E & \overset{\tau}{\dashrightarrow} & E' \\ | & & | \\ F(\alpha) & \overset{\tau_1}{\dashrightarrow} & F'(\alpha') \\ | & & | \\ F & \xrightarrow{\quad\varphi\quad} & F' \end{array}$$

where we shall always use the vertical lines in such pictures as the appropriate set inclusion.

We now look at what we have accomplished, in the following crucial discussion.

Remark 47.14. Let F be a field, f a non-constant polynomial in $F[t]$ of degree n, and K a splitting field of f over F. Let

$$\mathrm{Aut}_F(K) := \{\tau : K \to K \mid \tau \text{ is an } F\text{-automorphism}\},$$

a group called the *Galois group* of K *over* F. It is also called the *Galois group* of f, although if we omit having a fixed splitting field, then it is only unique up to isomorphism. Now let g be a non-constant polynomial in $F[t]$ and set $S = \{x \in K \mid g(x) = 0\}$, i.e., the set of roots of g in K. Suppose that S is not empty. Let $\tau \in \mathrm{Aut}_F(K)$, then $\tilde{\tau}$ fixes the coefficients of g in $F[t]$, so $\tilde{\tau}(g) = g$. Therefore, if $x \in K$, we have

$$\tau\big(g(x)\big) = \big(\tilde{\tau}(g)\big)(x) = g\big(\tau(x)\big).$$

Thus, if $x \in S$, we have

$$0 = \tau\big(g(x)\big) = g\big(\tau(x)\big), \textit{ i.e.}, \tau(x) \in S.$$

So we have a map,

$$\mathrm{Aut}_F(K) \to \Sigma(S) \text{ given by the restriction } \tau \mapsto \tau|_S.$$

Check that this is, in fact, a group homomorphism. Now let $f = g$ and let $\alpha_1, \ldots, \alpha_n$ be the roots of f in K, not necessarily distinct. We know that $K = F(\alpha_1, \ldots, \alpha_n) = F[\alpha_1, \ldots, \alpha_n]$. Let τ and ψ lie in $\mathrm{Aut}_F(K)$, and suppose that $\tau|_S = \psi|_S$. Then $\tau^{-1} \circ \psi$ fixes F and S, so $\tau^{-1} \circ \psi = 1_K$, as $K = F(S) = F[S]$. We conclude that $\tau|_S = \psi|_S$ if and only if $\tau = \psi$. Therefore, the map $\mathrm{Aut}_F(K) \to \Sigma(S)$ given by $\tau \mapsto \tau|_S$ is injective, so a group monomorphism, and we can view $\mathrm{Aut}_F(K)$ as a subgroup of $\Sigma(S)$ and $\Sigma(S)$ a subgroup of S_n with $\Sigma(S) = S_n$ if and only if $\alpha_1, \ldots, \alpha_n$ are distinct, i.e., f has no multiple roots. Suppose now, in addition, that f is irreducible in $F[t]$. Fix $i, j,\ 1 \leq i, j \leq n$, then K is a splitting field of f over $F(\alpha_i)$ and K is a splitting field of f over $F(\alpha_j)$. We have shown that there exists an F-isomorphism $\sigma : F(\alpha_i) \to F(\alpha_j)$ sending α_i to α_j. The theorem now says that there exists τ in $\mathrm{Aut}_F(K)$ extending σ and taking S onto S bijectively. This means that

$$\mathrm{Aut}_F(K) \subset \Sigma(S) \text{ is a transitive subgroup,}$$

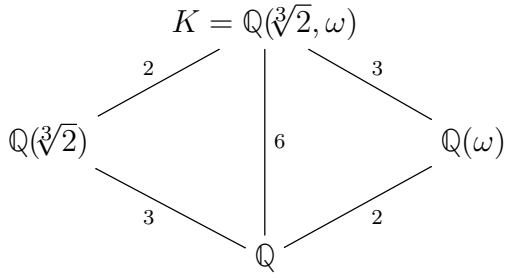
i.e., for every pair of roots α_i and α_j in S, there exists a $\tau \in \mathrm{Aut}_F(K)$ satisfying $\tau(\alpha_i) = \alpha_j$. We shall see in the sequel, that this will essentially reduce the theory of algebraic extensions to group theory. This is what Galois discovered.

We now give a number of examples to illustrate what we have done.

Examples 47.15. Let F be a field.

1. Let $f = t^2 + bt + c$ be a polynomial in $F[t]$ and K/F an extension field with α in K a root of f. If α' is another root of f or $\alpha = \alpha'$ is a *multiple root* of f, then $f = (t - \alpha)(t - \alpha') = t^2 - (\alpha + \alpha')t + \alpha\alpha'$ in $K[t]$. So we have $b = -\alpha - \alpha'$ and $c = \alpha\alpha'$. In particular, $\alpha' = -\alpha - b$ lies in K. If $\alpha = \alpha'$ is a multiple root, then $f = (t - \alpha)^2$ in $K[t]$ and $2\alpha = -b$. If, in addition, $\mathrm{char}\, F \neq 2$, then 2 is a unit and we have $\alpha = -b/2$ and $f = (t - \alpha)^2$ in $F[t]$ splits over F. In the general case, $f = (t - \alpha)(t - \alpha')$ splits over K and $F(\alpha)$ is a splitting field of f over F, so either $F = F(\alpha)$ and f splits over F or $[F(\alpha) : F] = 2$ depending on whether f is irreducible or not. If $\mathrm{char}\, F \neq 2$, this depends on whether $b^2 - 4c$ is a square in F or not, as the quadratic formula can be derived over F if 2 is a unit, just as over the real numbers.
2. The field $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is the splitting field of $(t^2 - 2)(t^2 - 3)$ in $\mathbb{C}$. We already know that $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = 4$ and $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ is a $\mathbb{Q}$-basis of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.

3. Let $f = t^3 - 2$ in $\mathbb{Q}[t]$. It is irreducible by Eisenstein's Criterion. Let $\omega = (-1+\sqrt{-3})/2 = \cos(\frac{2}{3}\pi) + \sqrt{-1}\sin(\frac{2}{3}\pi)$ We have the complex conjugate of ω satisfies $\overline{\omega} = \omega^2$ and $\omega^3 = 1$. The roots of f in $\mathbb{C}$ are: $\sqrt[3]{2}$, $\sqrt[3]{2}\,\omega$, $\sqrt[3]{2}\,\omega^2$, all distinct. Then $K = \mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{2}\,\omega, \sqrt[3]{2}\,\omega^2)$ $= \mathbb{Q}(\sqrt[3]{2}, \omega)$ is a splitting field of f over $\mathbb{Q}$ with $[K : \mathbb{Q}] \leq 3! = 6$. As f is monic irreducible, we have $f = m_{\mathbb{Q}}(\sqrt[3]{2})$ and, therefore, $3 = \deg f = [\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] \mid [K : \mathbb{Q}]$. As $\omega = {}^3\sqrt{2}\,\omega/\sqrt[3]{2}$ is not real, it does not lie in $\mathbb{Q}(\sqrt[3]{2})$. Thus $[K : \mathbb{Q}] > 3$, hence $[K : \mathbb{Q}] = 6$. [Alternatively, $m_{\mathbb{Q}}(\omega) = t^2 + t + 1$ splits in K, so $\mathbb{Q}(\omega) \subset K$ and $2 = \deg m_{\mathbb{Q}}(\omega) \mid [K : \mathbb{Q}]$, so $[K : \mathbb{Q}] = 6$.] We also have



(where the integers in the diagram are the field extension degrees.)

4. Let $p > 1$ be a prime and $f = t^{p-1} + t^{p-2} + \cdots + t + 1$ in $\mathbb{Q}[t]$. We know that f is irreducible by Remark 32.11(3). Let $\zeta = e^{2\pi\sqrt{-1}/p} = \cos(2\pi/p) + \sqrt{-1}\sin(2\pi/p)$, then $t^p - 1 = (t-1)f$ is a factorization into irreducibles in $\mathbb{Q}[t]$. So $\zeta^p = 1$ with $\zeta \neq 1$. As $f(\zeta) = 0$, we have that $[\mathbb{Q}(\zeta) : \mathbb{Q}] = p - 1 = \varphi(p)$. Similarly, $\zeta^2, \ldots, \zeta^{p-1}$ are roots of f, all distinct and different than ζ. It follows that $\mathbb{Q}(\zeta)$ is a splitting field of f (and of $t^p - 1$) over $\mathbb{Q}$. The polynomial f is usually denoted by Φ_p and called the *pth cyclotomic polynomial* in $\mathbb{Q}[t]$ and satisfies $\deg \Phi_p = \varphi(p)$.

5. The polynomial $f = t^4 + t^2 + 1 = (t^2 + t + 1)(t^2 - t + 1)$ in $\mathbb{Q}[t]$ has roots $\zeta = (-1+\sqrt{-3})/2$, $\overline{\zeta} = \zeta^2$, $\eta = (1+\sqrt{-3})/2$, and $\overline{\eta}$, where $^-$ is complex conjugation. We have $K = \mathbb{Q}(\sqrt{-3}) = \mathbb{Q}(\zeta, \eta)$ is a splitting field of f over $\mathbb{Q}$ and $[K : \mathbb{Q}] = 2$. Note that if x is a root of f, then $x^6 = 1$. The polynomial $t^2 - t + 1 = m_{\mathbb{Q}}(\eta)$ is the $6th$ *cyclotomic polynomial* over $\mathbb{Q}$. It is denoted by Φ_6 and satisfies $\deg \Phi_6 = \varphi(6)$.

6. Let p be an odd prime, r a positive integer, and $\zeta = e^{2\pi\sqrt{-1}/p^r}$.

   Claim: $\mathbb{Q}(\zeta)$ is the splitting field of $t^{p^{r-1}(p-1)} + t^{p^{r-1}(p-2)} + \cdots + t^{p^{r-1}} + 1$ and $t^{p^r} - 1 \in \mathbb{Q}[t]$ over $\mathbb{Q}$ and $[\mathbb{Q}(\zeta) : \mathbb{Q}] = \varphi(p^r) = p^{r-1}(p-1)$ (The same is true for $\zeta = e^{2\pi n\sqrt{-1}/p^r}$ for any n not divisible by p satisfying

$1 \leq n < p^r$). In particular, $\Phi_{p^r} = m_{\mathbb{Q}}(\zeta)$, the p^rth *cyclotomic polynomial* over $\mathbb{Q}$ satisfies $\deg \Phi_{p^r} = \varphi(p^r)$:

By Exercise 6.14(7), we know that $\varphi(p^r) = p^{r-1}(p-1)$, so it suffices to show $m_{\mathbb{Q}}(\zeta) = t^{p^{r-1}(p-1)} + t^{p^{r-1}(p-2)} + \cdots + t^{p^{r-1}} + 1$, as ζ^i, $1 \leq i \leq p^r$, are p^r distinct roots of $t^{p^r} - 1$. We need the following identity of binomial coefficients that is easily proven by induction:

$$\sum_{i=j}^{n} \binom{i}{i-j} = \binom{n+1}{n-j}. \tag{47.16}$$

Let $f = t^{p^{r-1}(p-1)} + t^{p^{r-1}(p-2)} + \cdots + t^{p^{r-1}} + 1 \in \mathbb{Z}[t]$ and $t = y + 1$. If we can show that $g(y+1) = f(t)$ satisfies Eisenstein's Criterion for p, we are done by Remark 32.11(2). By the Children's Binomial Theorem (Exercise 24.20(6)),

$$(y+1)^{p^{r-1}i} \equiv (y^{p^{r-1}} + 1)^i = \sum_{j=0}^{i} \binom{i}{j} y^{p^{r-1}j} \mod p\mathbb{Z}[t].$$

We have

$$g(y+1) = \sum_{i=0}^{p-1} \sum_{j=0}^{i} \binom{i}{j} y^{p^{r-1}j} \mod p\mathbb{Z}[t].$$

The coefficient of $y^{p^{r-1}j}$ modulo $p\mathbb{Z}[t]$ for $j \neq 0$ is

$$\sum_{i=j}^{p-1} \binom{i}{i-j} = \binom{p}{p-j} \equiv 0 \mod p.$$

The constant term is precisely p, so the result follows.

7. Let $f = t^3 + t + 1$ in $(\mathbb{Z}/2\mathbb{Z})[t]$. Check that f has no roots in $(\mathbb{Z}/2\mathbb{Z})$.

Observation 47.17. If F is a field and a non-constant polynomial g in $F[t]$ has no roots in F and is of degree at most three, then g is irreducible in $F[t]$.

Therefore, the polynomial f is irreducible in $(\mathbb{Z}/2\mathbb{Z})[t]$. Suppose that $L/(\mathbb{Z}/2\mathbb{Z})$ is an extension field such that f has a root α in L and let $K = (\mathbb{Z}/2\mathbb{Z})(\alpha)$. We have $\alpha^3 = -\alpha - 1 = \alpha + 1$ in K, so $\alpha^{-1} = \alpha^2 + 1$. Let $f = (t - \alpha)(t^2 + at + b)$ in $K[t]$. Multiplying out, yields $\alpha = a$ and $b = 1 + \alpha^2 = \alpha^{-1}$. It follows that $t^2 + at + b = t^2 + \alpha t + (1 + \alpha^2)$ in $K[t]$. If $e_{\alpha^2} : K[t] \to K$ is the evaluation map

at α^2, then

$$\begin{aligned} e_{\alpha^2}(t^2+at+b) &= e_{\alpha^2}\big(t^2+\alpha t+(1+\alpha^2)\big) = \alpha^4+\alpha^3+(1+\alpha^2) \\ &= (\alpha^2+\alpha)+\alpha^3+(1+\alpha^2) \\ &= (\alpha^2+\alpha)+(\alpha+1)+(1+\alpha^2) = 0. \end{aligned}$$

Consequently, α^2 is also a root of f in K with $\alpha \neq \alpha^2$. So we have $f = (t-\alpha)(t-\alpha^2)g$ in $K[t]$ with $\deg g = 1$. Therefore, f splits over $K = (\mathbb{Z}/2\mathbb{Z})(\alpha)$, hence K is a splitting field of f over $(\mathbb{Z}/2\mathbb{Z})$ with $[K : (\mathbb{Z}/2\mathbb{Z})] = \deg f = 3$. Note that K is a field having $2^3 = 8$ elements.

8. Let $p > 0$ be a prime, F a field of characteristic p, and α an element in F. Then $t^p - \alpha^p = (t-\alpha)^p$ in $F[t]$ by the Children's Binomial Theorem (Exercise 24.20(6)), has F as a splitting field of f and f has only one distinct root of multiplicity p. Now let L/F be an extension field, α an element in L satisfying $a = \alpha^p$ lies in F. Then $K = F(\alpha)$ is a splitting field of $t^p - a$ in $F[t]$ over F.

   Claim. If α is not an element in F, then $[K : F] = p$:

   Suppose that $t^p - \alpha = fg$ in $F[t]$ with f and g lying in $F[t]$. We may assume that both f and g are monic and f not a unit. As $fg = t^p - a = (t-\alpha)^p$ in the UFD $K[t]$, we must have $f = (t-\alpha)^r$ in $K[t]$ for some r satisfying $1 \leq r \leq p$. As f lies in $F[t]$, its constant term $\pm\alpha^r$ must lie in F. If $r < p$, then r and p are relatively prime, so α^r, α^p lying in F implies that α lies in F, a contradiction. Therefore, $g = 1$ and $r = p$, so $t^p - a$ is irreducible in $F[t]$.

Exercises 47.18. 1. If $f \in \mathbb{Q}[t]$ and K is a splitting field of f over $\mathbb{Q}$, determine $[K : \mathbb{Q}]$ if f is
   (a) $t^4 + 1$.
   (b) $t^6 + 1$.
   (c) $t^4 - 2$.
   (d) $t^6 - 2$.
   (e) $t^6 + t^3 + 1$.
2. Find the splitting fields K for $f \in \mathbb{Q}[t]$ and $[K : \mathbb{Q}]$ if f is:
   (a) $t^4 - 5t^2 + 6$.
   (b) $t^6 - 1$.
   (c) $t^6 - 8$.
3. Show both of the following:
   (a) If $K/\mathbb{Q}$ and $\sigma \in \operatorname{Aut} K$ then σ fixes $\mathbb{Q}$.
   (b) The fields $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{3})$ are not isomorphic.
4. Prove Identity 47.16.

5. A *primitive nth root of unity* is an element $z \in \mathbb{C}$ such that $z^n = 1$ and $z^r \neq 1$ for $1 \leq r < n$. Show the following:
   (a) There exist $\varphi(n) := |\{d \mid 0 \leq d \leq n, \quad (d, n) = 1\}|$ primitive nth roots of unity.
   (b) If ω is a primitive *nth* root of unity, then $\mathbb{Q}(\omega)$ is a splitting field of $t^n - 1 \in \mathbb{Q}[t]$.
   (c) Let $\omega_1, ..., \omega_{\varphi(n)}$ be the $\varphi(n)$ primitive nth roots of unity of $t^n - 1 \in \mathbb{Q}[t]$ and $\sigma \in \operatorname{Aut} \mathbb{Q}(\omega_1)$ then $\sigma(\omega_1) = \omega_i$ for some i, $1 \leq i \leq \varphi(n)$.
6. Continued from Exercise 5
   (a) For each i, $1 \leq i \leq \varphi(n)$, there exists an $\sigma \in \operatorname{Aut} \mathbb{Q}(\omega_1)$ such that $\sigma(\omega_1) = \omega_i$.
   (b) Let $\Phi_n(t) = (t - \omega_1) \cdots (t - \omega_{\phi(n)})$. Then show $\Phi_n(t) \in \mathbb{Q}[t]$. $\Phi_n(t)$ is called the nth *cyclotomic polynomial*.
   (c) $\Phi_n(t) \in \mathbb{Z}[t]$.
7. Continued from Exercise 6
   (a) $\Phi_n(t) \in \mathbf{Z}[t]$ is irreducible.
   (b) Calculate $\Phi_n(t)$ for $n = 3, 4, 6, 8$ explicitly and show directly that $\Phi_n(t) \in \mathbb{Z}[t]$ is irreducible.
8. Let $F = \mathbb{Z}/p\mathbb{Z}$. Show all of the following:
   (a) There exists a polynomial $f \in F[t]$ satisfying $\deg f = 2$ and f irreducible.
   (b) Use the f in (a) to construct a field with p^2 elements.
   (c) If $f_1, f_2 \in F[t]$ has $\deg f_i = 2$ and f_i irreducible for $i = 1, 2$, show that their splitting fields are isomorphic.
9. Let F be a field of characteristic different from two and $f = at^2 + bt + c$ in $F[t]$. State and prove the quadratic formula.
10. Let K/F be an extension of fields and $f \in F[t]$.
    (a) If $\varphi : K \to K$ is an F-automorphism, then φ takes roots of f in K to roots of f in K.
    (b) If $F \subset \mathbb{R}$ and $\alpha = a + b\sqrt{-1}$ is a root of f with $a, b \in \mathbb{R}$ then $\overline{\alpha} = a - b\sqrt{-1}$ is also a root of f.
    (c) Let $F = \mathbb{Q}$. If $m \in \mathbb{Z}$ is not a square and $a + b\sqrt{m}$ in $\mathbb{C}$ is a root of f with a, b in $\mathbb{Q}$ then $a - b\sqrt{m}$ is also a root of f in $\mathbb{C}$.
11. Any (field) automorphism $\varphi : \mathbb{R} \to \mathbb{R}$ is the identity automorphism.
12. Let $f = (t^2 - p_1) \cdots (t^2 - p_n)$ in $\mathbb{Q}[t]$ with $p_1, \ldots, p_n$ be n distinct prime numbers. Show that $K = \mathbb{Q}[\sqrt{p_1}, \ldots, \sqrt{p_n}]$ is a splitting field of f over $\mathbb{Q}$ and $[K : \mathbb{Q}] = 2^n$.
13. Find a splitting field of $f \in F[t]$ if $F = \mathbb{Z}/p\mathbb{Z}$ and $f = t^{p^e} - t$, $e > 0$.

14. Let K be a splitting field over F of f in $F[t] \setminus F$ and g an irreducible polynomial in $F[t]$. Suppose that g has a root in K. Show that g splits over K.
15. Suppose F is a field and K is a splitting field of $f \in F[t]$ over F. If $K/E/F$ is an intermediate field, $\sigma : E \to K$ be a monomorphism fixing F, show that there exists an F-automorphism $\widehat{\sigma} : K \to K$ lifting σ, i.e., such that $\widehat{\sigma}|_E = \sigma$.
16. Let F be a field of characteristic $p > 0$. Show that $f = t^4 + 1 \in F[t]$ is not irreducible. Let K be a splitting field of f over F. Determine which finite fields F satisfies $K = F$.
17. Let $f = t^6 - 3 \in F[t]$. Construct a splitting field K of f over F and determine $[K : F]$ if $F = \mathbb{Q}, \mathbf{Z}/5\mathbb{Z}$, or $\mathbb{Z}/7\mathbb{Z}$. Do the same thing if f is replaced by $g = t^6 + 3 \in F[t]$ for the same fields F.

48. Algebraically Closed Fields

Recall that a field K is called *algebraically closed* if every non-constant polynomial in $K[t]$ has a root in K, equivalently that every non-constant polynomial in $K[t]$ splits over F. Clearly, this is also equivalent to the condition that if L/K is an algebraic extension of fields, then $L = K$. We call an extension field K of F an *algebraic closure* of F if K/F is algebraic and K is algebraically closed. We shall later prove the Fundamental Theorem of Algebra that $\mathbb{C}$ is an algebraically closed field. We know that $\mathbb{C}$ is not an algebraic closure of $\mathbb{Q}$. In fact, assuming the Fundamental Theorem of Algebra, we showed in Example 45.22(4) that $\Omega := \{x \in \mathbb{C} \mid x \text{ algebraic over } \mathbb{Q}\}$ is an algebraic closure of $\mathbb{Q}$ and is countable. In this section, we shall show that given any field F, there exists an algebraic closure of F, and it is unique up to an F-homomorphism. By the argument in Example 45.22(5), to establish the existence of an algebraic closure of F, it suffices to show that F is a subfield of some algebraically closed field.

We know that if we have a finite set of polynomials, $\{f_1, \ldots, f_n\}$, then we can construct a field K such that each f_i splits in K and K is the smallest such field by letting K be the splitting field of $\prod f_i$ over F. Another way of constructing this K would be to find a splitting field K_1 of f_1 over F, then K_2 a splitting field of f_2 over K_1, and continue in this way. We use both for the case of an infinite set of polynomials.

To use these two ideas, we first set up some notation. Let I be an indexing set and $S = \{t_i \mid i \in I\}$ be a set of distinct indeterminants. Then $F[S]$ is the set of polynomials, coefficients in F in the t_i's, $i \in I$. If $J \subset I$, let $S_J = \{t_j \mid j \in J\}$. As each such polynomial needs only finitely many indeterminants to determine it, i.e., $f \in F[S_J]$, for some

finite subset $J \subset I$, we have

$$F[S] = \bigcup_{\substack{J \subset I \\ J \text{ finite}}} F[S_J].$$

Note that $F[S]$ is a domain, and in fact, a UFD. Using this notation, we give Artin's proof of the following theorem.

Theorem 48.1. *Let F be a field, then there exists an algebraically closed field K containing F.*

PROOF. Let $I = F[t] \setminus F$ be our indexing set and $S = \{t_f \mid f \in I\}$. So if $f \in F[t]$, we have $f(t_f)$ lies in $F[t_f] \subset F[S]$. Set $T = \{f(t_f) \mid f \in F[t] \setminus F\}$ and $\mathfrak{A}$ the ideal in $F[S]$ generated by the set T. So

$$\mathfrak{A} = \{ \sum_{finite} g_i f_i(t_{f_i}) \mid g_i \in F[S]\}.$$

Claim. $\mathfrak{A} < F[S]$:

Suppose that $\mathfrak{A} = F[S]$, then there exist $f_i \in T$ and $g_i \in F[S]$, $i = 1, \ldots, n$, some n, such that

$$1 = g_1 f_1(t_{f_1}) + \cdots + g_n f_n(t_{f_n})$$

and a finite subset $J \subset I$ with $g_1, \ldots, g_n$ lying in $F[S_J]$. Relabel each t_{f_i} as t_i and the t_j's in J such that $J = \{t_1, \ldots, t_N\}$ some N, so $f_i, g_i \in F[t_1, \ldots, t_N]$. We know that there exists a field extension K/F such that each f_i has a root α_i in K for $1 \leq i \leq n$ by the preliminary remarks. Set $\alpha_i = 0$ for $n < i \leq N$. Applying the evaluation map $e_{\alpha_1, \ldots, \alpha_n}$ yields

$$0 = \sum g_i(\alpha_1, \ldots, \alpha_n) f_i(\alpha_i) = 1$$

in K, a contradiction. This establishes the claim.

By the claim, there exists a maximal ideal $\mathfrak{m}$ satisfying $\mathfrak{A} \subset \mathfrak{m} < F[S]$ (which we know exists by Zorn's Lemma). Define $L_1 := F[S]/\mathfrak{m}$, viewed as a field extension of F in the usual way (i.e., identifying F and $\varphi(F)$ where $\varphi : F[S] \to L$ is given by $g \mapsto g + \mathfrak{m}$). By construction, if f is a non-constant polynomial in $F[t]$, then f has a root $t_f + \mathfrak{m}$ in L_1. Inductively, define L_i such that L_i/L_{i-1} and every non-constant polynomial $g \in L_{i-1}[t]$ has a root in L_i by repeating the same construction. Set $L = \cup L_i$. If α, β lie in L with β nonzero, then $\alpha \pm \beta, \alpha\beta, \beta^{-1}$ lie in some subfield L_i of L, hence L is a field. Finally, let f be a non-constant polynomial in $L[t]$. As f has finitely many nonzero coefficients, there exists an i such that f lies in $L_i[t]$ hence has a root in $L_{i+1} \subset L$. Therefore, L is algebraically closed. $\square$

We showed in the argument in Example 45.22(5) that if F is a field and K/F a field extension with K algebraically closed, then the elements in K algebraic over F form an algebraic closure of F. Therefore, we have established the following result.

Corollary 48.2. *Let F be a field. Then an algebraic closure of F exists.*

Algebraic closures are "super" splitting fields, so we expect a uniqueness statement. We prove this.

Theorem 48.3. *Let K/F be an algebraic extension of fields and L an algebraically closed field. Suppose that there exists a homomorphism $\sigma : F \to L$. Then there exists a homomorphism $\tau : K \to L$ that lifts σ. In particular, if K is also algebraically closed and $L/\sigma(F)$ is algebraic, then τ is an isomorphism.*

Proof. We use a Zorn Lemma argument that is often repeated in algebra to define maps. Let

$$\begin{aligned} S := \{(E,\eta) \mid & K/E/F \text{ an intermediate field} \\ & \eta : E \to L \text{ a homomorphism lifting } \sigma\}. \end{aligned}$$

Partially order the set S by $\leq$ defined by

$$(E,\eta) \leq (E',\eta') \text{ if } E \subset E' \text{ and } \eta'|_E = \eta.$$

As (F,σ) lies in S, the set S is nonempty. Let $\mathcal{C} = \{(E_i,\eta_i) \mid i \in I\}$ be a chain in S. We know that $E = \cup_I E_i$ is a field. Define

$$\eta : E \to L$$

as follows: If α lies in E, choose $i \in I$ such that $\alpha \in E_i$ and set $\eta(\alpha) := \eta_i(\alpha)$. By the definition of our partial ordering, $\eta(\alpha)$ is independent of i, i.e., η is well-defined. It follows that (E,η) is an upper bound for $\mathcal{C}$. By Zorn's Lemma, there exists a maximal element (E,η) in S, so $K/E/F$ and $\eta : E \to L$ extends σ.

Claim. $K = E$:

Suppose this is false and $\alpha \in K \setminus E$. As α is algebraic over F, it is algebraic over E. Let β be a root of $\widetilde{\eta}\big(m_E(\alpha)\big)$ in $\eta(E)[t]$ in the algebraically closed field L. Then we know by Proposition 47.10 that there exists an isomorphism

$$\eta' : E(\alpha) \to \eta(E)(\beta) \text{ with } \alpha \mapsto \beta \text{ lifting } \eta.$$

It follows that $(E(\alpha),\eta')$ lies in S with $(E,\eta) < (E(\alpha),\eta')$, contradicting the maximality of (E,η). This proves the claim.

If K is algebraically closed so is the subfield $\tau(K)$ in L. If $L/\sigma(F)$ is algebraic, then $L/\tau(K)$ is algebraic, hence $L = \tau(K)$ and the theorem if proven. □

We immediately see that this implies the desired result.

Corollary 48.4. *Let F be a field. Then any algebraic closure of F is unique up to an F-isomorphism.*

Because of the corollary, one usually fixes an algebraic closure K of a field F and views all roots of all polynomials in $F[t]$ as lying in K. More generally, one fixes some E/F with E algebraic closed and only works with fields K with $E/K/F$. Then E contains a unique algebraic closure of F, hence for each polynomial in $F[t]$ a unique splitting field in E. One then views every algebraic extension of F to be in E. For example, one can let $E = \mathbb{C}$ when $F = \mathbb{Q}$. More generally, if F is a field of characteristic zero and $\widetilde{F}$ an algebraically closed field, we can view $F/\mathbb{Q}$ and the algebraic closure of $\mathbb{Q}$ in $\widetilde{F}$ in $\mathbb{C}$ by replacing it with an isomorphic copy. We shall always do so implicitly.

Exercises 48.5. 1. Let F be a finite field and K an algebraic closure of F. Show that K is countable and infinite. (In all other cases, an algebraic closure of a field has the same cardinality of the given field).

2. Let S be a collection of polynomials over a field F. Define what it means for an extension field K/F to be a *splitting field* of the collection S and prove that one exists and is unique up to an F-isomorphism. Also prove that an algebraic closure of F is a splitting field for the collection of irreducible polynomials in $F[t]$.

49. Constructible Numbers

The ancient Greeks were very interested in constructing geometric objects under certain rules. The most famous ones that they could not solve were the following:

Euclidean Construction Problems. Using a straight-edge and compass, can you:

I. Trisect a given angle?

II. (Delian Problem) Double a cube, i.e., construct the edge of a cube whose volume is twice the volume of a given cube?

III. Square a circle, i.e., construct a square whose area equals the area of a given circle?

IV. Construct a regular n-gon, $n \geq 3$?

Given a specific set of initial points, the following constructions were allowed:

C1. You can draw a line through two initial points.
C2. You can draw a circle with center at an initial point with radius equal to the length between two initial points.

From these constructions on your given set of initial points, you were allowed to enlarge this original set by adding the points of intersection arising from the constructions C1 and C2 and add further points by iterating this process.

We shall solve the Euclidean Problems, by converting them into problems in field theory. In the case of squaring the circle, some analysis will also be needed.

Algebraic Reformulation (Preliminaries). Let $S = \{P_1, \ldots, P_n\}$ be a finite non-empty set of points in the euclidean plane. Recursively define S_r as follows:

$$S_0 = S \text{ and having defined } S_r \text{ define } S_{r+1} := S_r \cup T_r$$

where T_r is the union of the following sets:

1. The set of points of intersections of lines through two distinct points in S_r (that we call S_r-*lines*).
2. The set of points of the intersection of circles having center in S_r and radii equal in length to line segments connecting two distinct points in S_r (that we call S_r-*circles*).
3. The set of intersections of S_r-circles and S_r-lines.

Now set

$$C(P_1, \ldots, P_n) := \bigcup_{r=0}^{\infty} S_r,$$

the *set of points constructible* from $P_1, \ldots, P_n$. We say a point P in the euclidean plane is *constructible from* $P_1, \ldots, P_n$, if $P \in C(P_1, \ldots, P_n)$ and *not constructible* otherwise. If $C = C(P_1, \ldots, P_n)$, call a line segment a C-*line* if an S_r-line for some r and a C-*circle* if an S_r-circle for some r.

Remarks 49.1. Let $P_1, \ldots, P_n$ be our initial given points in the euclidean plane and $C = C(P_1, \ldots, P_n)$ Using elementary euclidean geometry, we deduce the following:

1. $C(P_1) = \{P_1\}$. As this is not interesting, we shall always assume that we begin with at least two given points, i.e., $n \geq 2$.
2. We can begin to define a coordinate system of our euclidean plane by setting $P_1 = (0, 0)$, the origin, and $P_2 = (1, 0)$.

3. Given a line segment of length d between two points in C, we can construct a line segment of length md for all integers m, can bisect the given line segment between two points in C, and can construct a perpendicular at each endpoint of the line segment between two points in C. In particular, we can construct axes.
4. We can drop a perpendicular from a point in C to a given C-line as well as construct a parallel through a point in C to a C-line. In particular, we see that any point in $\mathbb{Z}^2$ is constructible from $P_1, \ldots, P_n$.
5. Given two C-lines of angles θ and φ with the X-axis respectively, we can construct rays (C-lines through the origin) of angles $\theta \pm \varphi$, $-\theta$, $\theta/2$. In fact, this can be done with the X-axis replaced by any allowable C-line.

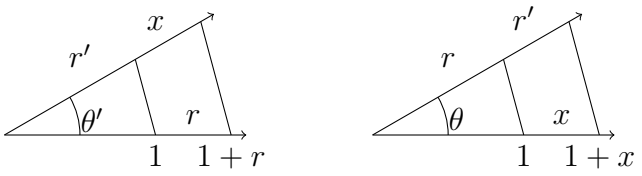
It is now convenient to complexify everything, i.e., to replace the euclidean plane with the complex plane. Let $P_1, \ldots, P_n$, $n \geq 2$, be points in the euclidean plane. As above, we let $P_1 = (0,0)$ and $P_2 = (1,0)$. To complexify, if $P_i = (x_i, y_i)$ lie in $\mathbb{R}^2$, let $z_i = x_i + y_i\sqrt{-1}$, e.g., we have $z_1 = 0$ and $z_2 = 1$, and let $C(z_1, \ldots, z_n)$ denote the complex numbers constructible by straight-edge and compass from $z_1, \ldots, z_n$. We call $C(z_1, \ldots, z_n)$ the *set of complex constructible numbers from* $z_1, \ldots, z_n$. We have $(x, y) \in C(P_1, \ldots, P_n)$ if and only if $x + y\sqrt{-1} \in C(z_1, \ldots, z_n)$. Note that we already know that $C(z_1, \ldots, z_n)$ contains the Gaussian integers. But we have much more.

Theorem 49.2. *Let $n \geq 2$ and $z_1 (= 0), z_2 (= 1), \ldots, z_n$ be complex numbers. Then $C(z_1, \ldots, z_n)$ is a field satisfying $\mathbb{Q} \subset C(z_1, \ldots, z_n) \subset \mathbb{C}$.*

Proof. As $C(z_1, \ldots, z_n) \subset \mathbb{C}$ with $\operatorname{char} \mathbb{C} = 0$, it suffices to show that $C(z_1, \ldots, z_n)$ is a field. Let z, z' lie in $C(z_1, \ldots, z_n)$, with z nonzero. As we can draw parallel lines, we can add given vectors (line segments starting at the origin), so we see that $z \pm z'$ lies in $C(z_1, \ldots, z_n)$. We need to show that zz' and z^{-1} lie in $C(z_1, \ldots, z_n)$. Write z and z' in polar form, say $z = re^{\sqrt{-1}\theta}$ and $z' = r'e^{\sqrt{-1}\theta'}$ with r, r', θ, θ' real numbers, r, r' non-negative. Then $zz' = rr'e^{\sqrt{-1}(\theta+\theta')}$ and $z^{-1} = r^{-1}e^{-\sqrt{-1}\theta}$. We know that we can construct rays of angle $\theta \pm \theta'$ and $-\theta$ and line segments of length $r = |z|$ and $r' = |z'|$. On the X-axis in the complex plane mark off 1 and $1 + r$ and the ray of angle θ' through z', which we may assume to be nonzero. Then the line through $1 + r$ parallel to the line through z' and 1 intersects the ray of an angle θ' with the X-axis at $(r + x)e^{\sqrt{-1}\theta'}$, and x is seen to be rr' by similar triangles. Reversing, the roles of x and r' constructs r/r'. (Cf. Figure 49.3 below.) As

associativity, commutativity, etc. hold in $\mathbb{C}$, we see that $C(z_1, \ldots, z_n)$ is a field. □

(49.3)



The field $C(z_1, \ldots, z_n)$ is easily characterized using computation of the intersection points of lines and circles.

Theorem 49.4. *Let $z_1 (= 0), z_2 (= 1), \ldots, z_n$ be points in $\mathbb{C}$ with $n \geq 2$. Then the field $C(z_1, \ldots, z_n)$ has the following properties:*

(1) The point z_i lies in $C(z_1, \ldots, z_n)$ for $i = 1, \ldots, n$.
(2) If z lies in $C(z_1, \ldots, z_n)$, so does its complex conjugate $\overline{z}$.
(3) if z^2 lies in $C(z_1, \ldots, z_n)$, so does z, i.e., if z lies in $C(z_1, \ldots, z_n)$, so does $\sqrt{z}$.

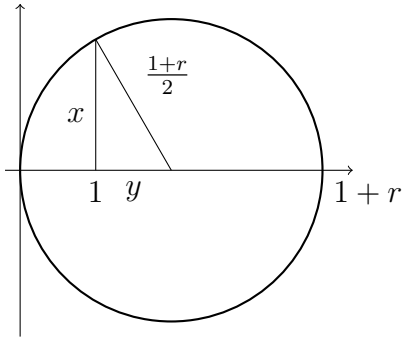
Moreover, $C(z_1, \ldots, z_n)$ is the unique smallest subfield of $\mathbb{C}$ satisfying these properties, i.e., if a subfield C' of $\mathbb{C}$ satisfies (1), (2), and (3), then $C(z_1, \ldots, z_n) \subset C'$.

Proof. (1) is clear.

(2). If $z = re^{\sqrt{-1}\theta}$, then $\overline{z} = re^{-\sqrt{-1}\theta}$, and (2) follows.

(3). Since we can construct rays of angle $\theta/2$ from rays of angle θ, it suffices to show that we can construct $\sqrt{r}$ given a segment of length r. Draw a circle centered at the point $(1 + r)/2$ on the X-axis of radius $(1 + r)/2$. Erect a perpendicular at the point 1 on the X-axis parallel to the Y-axis. This line segment has length $\sqrt{r}$ by the Pythagorean Theorem. (Cf. Figure 49.5.)

(49.5)



This shows that $C(z_1, \ldots, z_n)$ satisfies the desired properties. Suppose that C' also does. As $z_1, \ldots, z_n$ all lie in C', by the definition of $C(z_1, \ldots, z_n)$, it suffices to show that intersection points of two C'-lines, two C'-circles, and a C'-line and a C'-circle, lie in C'. To do this note that $\sqrt{-1}$ lies in C' by (3) and if $z = a + b\sqrt{-1}$ with a, b in $\mathbb{R}$, then z lies in C' if and only if a and b lie in $C' \cap \mathbb{R}$, e.g., if z lies in C', then by (2), so does $b = -\sqrt{-1}(z - \overline{z})/2$ as C' is a field. This means that the equations defining C'-lines and C'-circles have coefficients in $C' \cap \mathbb{R}$, so solving these equations leads to solutions in C' (using (3) for intersections with circles). □

We wish to further characterize the field theoretic properties of fields of constructible numbers. Property (3) in the theorem gives the key. Such fields are closed under taking square roots of elements in the field.

Definition 49.6. Let K/F be an extension of fields. We call K/F a *square root tower over* F if there exist elements $u_1, \ldots, u_n$ in K for some positive integer n, satisfying:

1. $K = F(u_1, \ldots, u_n)$.
2. u_1^2 lies in F.
3. u_i^2 lies in $F(u_1, \ldots, u_{i-1})$ for each $i > 1$.

Remarks 49.7. Let K/F be a square root tower defined by $K = F(u_1, \ldots, u_n)$ satisfying the above properties and E/F an extension field. Assume that both K and E lie in some larger field.

1. We have
$$[F(u_1, \ldots, u_i) : F(u_1, \ldots, u_{i-1})] = \begin{cases} 1 & \text{if } u_i \in F(u_1, \ldots, u_{i-1}) \\ 2 & \text{otherwise.} \end{cases}$$
In particular, $[F(u_1, \ldots, u_i) : F] = 2^e$ for some $e \leq n$.
2. The extension $E(K)/E$ is a square root tower. (Of course, $E(K) = E(u_1, \ldots, u_n)$.)
3. If E/F is also a square root tower, then so is $E(K)/F$.
4. If E/K is a square root tower, then so is E/F
5. Suppose that $K \subset \mathbb{C}$ and $\overline{F} := \{\overline{z} \mid z \in F\}$, where $\overline{z}$ is the complex conjugate of z. Then $\overline{F}$ is a field and $\overline{K}/\overline{F}$ is a square root tower, with $\overline{K} = \overline{F}(\overline{z_1}, \ldots, \overline{z_n})$.
6. Suppose that $\operatorname{char} F \neq 2$ and E/F satisfies $[E : F] \leq 2$. Then E/F is a square root tower:
We may assume that $[E : F] = 2$, so there exists an $\alpha \in E \setminus F$ satisfying $E = F(\alpha)$. Let $m_F(\alpha) = t^2 + bt + a$ in $F[t]$. Then $m_F(\alpha)$

splits over E and has roots α and $\alpha_1 = -b - \alpha$. . Let $\beta = 2\alpha + b$. As $\alpha^2 + b\alpha = -a$ and β does not lie in F, we see that $E = F(\beta)$ and $\beta^2 = -4a + b^2$ lies in F.

In general, it is not true if a field extension E/F has degree a power of two, then it is a square root tower. (Can you find an example?) However, we shall prove the following theorem later (cf. Theorem 54.11):

Theorem 49.8. (Square Root Tower Theorem) *Let F be a field of characteristic zero and K/F an extension of fields that is a splitting field of some non-constant polynomial in $F[t]$. Then K/F is a square root tower if and only if K/F is a field extension of degree a power of two.*

In fact, we shall prove a stronger result, that will be applicable to fields of arbitrary characteristic. We shall use this result when needed later in this section.

The notion of square root tower allows us to give a sufficient condition for a complex number to be constructible from a set of complex numbers.

Theorem 49.9. (Constructibility Criterion) *Given complex numbers $z_1 (= 0), z_2 (= 1), \ldots, z_n$ with $n \geq 2$, set $F = \mathbb{Q}(z_1, z_2, \ldots, z_n, \overline{z_1}, \ldots, \overline{z_n})$. If z is a complex number, then z is constructible from $z_1, z_2, \ldots, z_n$, i.e., z lies in $C(z_1, z_2, \ldots, z_n)$ if and only if there exists a square root tower K/F (with $K \subset \mathbb{C}$) satisfying $z \in K$.*

Proof. Let

$$C' = \{z \in \mathbb{C} \mid \text{There exists a square root tower } K/F \text{ with } z \in K\}.$$

By definition, $F \subset C'$. We must show that $C' \subset C(z_1, \ldots, z_n)$. We know that $C(z_1, \ldots, z_n)$ is a field containing $z_1, \ldots, z_n$ closed under complex conjugation and extracting square roots. In particular, F is a subfield of $C(z_1, \ldots, z_n)$ as is any square root tower K/F. In particular, $C' \subset C(z_1, \ldots, z_n)$. To show that $C(z_1, \ldots, z_n) \subset C'$, it suffices to show by Theorem 49.4 that C' is a field closed under complex conjugation and extracting square roots. Let z and z' be elements in C' satisfying $z \in K$ and $z' \in K'$ with K/F and K'/F square root towers. Then $K(K')/F$ is a square root tower by Remark (3) above containing zz', $z \pm z'$, and z^{-1} (if $z \neq 0$). As $K(K') \subset C'$, we have C' is a field. Moreover, $K(\sqrt{z})/K$ is a square root tower, hence $K(\sqrt{z})/F$ is also a square root tower as K/F is. Since $F = \overline{F}$, we also have $\overline{K}/F$ is a square root tower. Therefore, $\sqrt{z}$ and $\overline{z}$ lie in C' and $C(z_1, \ldots, z_n) \subset C'$ as needed. □

Corollary 49.10. *Let $z_1 (= 0), z_2 (= 1), \dots, z_n$ be complex numbers with $n \geq 2$ and z a complex number in $C(z_1, \dots, z_n)$. Then z is algebraic over $F = \mathbb{Q}(z_1, z_2, \dots, z_n, \overline{z_1}, \dots, \overline{z_n})$ and $[F(z) : F]$ is a power of two.*

PROOF. As z lies in $C(z_1, \dots, z_n)$, there exists a square root tower K/F with $z \in K$, hence $[F(z) : F] \mid [K : F]$, a power of two. □

In many euclidean constructions, one starts with two points, P_1 and P_2, which we may assume correspond to complex numbers $z_1 = 0$ and $z_2 = 1$. In this case, $\mathbb{Q} = \mathbb{Q}(z_1, z_2, \overline{z_1}, \overline{z_2})$ and $C(z_1, z_2)$ is called the field of (*complex euclidean*) *constructible numbers* and any $z \in C(z_1, z_2)$ is called *constructible*. So if z is a constructible number, we have $[\mathbb{Q}(z) : \mathbb{Q}]$ is a power of two.

We now can solve the (Greek) Euclidean Construction Problems. Recall that a cubic (or quadratic) polynomial in $F[t]$, F a field, is irreducible if and only if it does not have a root in F.

Trisection of an angle θ: To solve this construction problem, we need the following lemma.

Lemma 49.11. *Let α be a ray with the X-axis as one side. Then it is constructible form $z_1 (= 0), z_2 (= 1), \dots, z_n$ if and only if $e^{\sqrt{-1}\alpha}$ is constructible from $z_1 (= 0), z_2 (= 1), \dots, z_n$ if and only if $\cos\alpha$ is constructible from $z_1 (= 0), z_2 (= 1), \dots, z_n$.*

PROOF. As $z = e^{\sqrt{-1}\alpha} = \cos\alpha + \sqrt{-1}\sin\alpha$ and we can erect and drop perpendiculars, this is immediate. □

The trisection of an angle θ is, therefore, equivalent to:

Problem. Given $z_1 = 0, z_2 = 1, z_3 = \cos\theta$, can we construct $\cos(\theta/3)$ from $z_1 = 0, z_2 = 1, z_3$?

Let $F = \mathbb{Q}(\cos\theta, \overline{\cos\theta}) = \mathbb{Q}(\cos\theta)$. For any angle α, we have the trigonometric identity

$$\cos 3\alpha = 4\cos^3\alpha - 3\cos\alpha.$$

So $\beta = \cos(\theta/3)$ is a root in $\mathbb{C}$ of $f = 4t^3 - 3t - \cos\theta$ in $F[t]$. The polynomial is *reducible*, i.e., not irreducible in $F[t]$ if and only if f has a root in F if and only if for any root γ of f in $\mathbb{C}$, we have $[F(\gamma) : F] = \deg m_F(\gamma) < 3$ if and only if $F(\beta)/F$ is a square tower if and only if β is constructible from z_1, z_2, z_3. Hence $\beta \notin C(z_1, z_2, z_3)$ if and only if f has no root in F. We, therefore, have proven the following:

Theorem 49.12. *An angle θ can be trisected with straight-edge and compass if and only if the polynomial $f = 4t^3 - 3t - \cos\theta$ in $\mathbb{Q}(\cos\theta)$ is reducible if and only if f has a root in $\mathbb{Q}(\cos\theta)$*

Examples 49.13. 1. Let $\theta = 60^o$, so $\cos\theta = 1/2$, $\mathbb{Q} = \mathbb{Q}(\cos\theta)$. The only possible roots of $f = 4t^3 - 3t - 1/2$ in $\mathbb{Q}$ are ± 1, $\pm 1/2$, $\pm 1/4$, $\pm 1/8$, none of which are, so f is irreducible in $\mathbb{Q}[t]$, hence an angle of 60^0 cannot be trisected with straight-edge and compass.

2. Let $\theta = 90^o$, so $\cos\theta = 0$ and $\mathbb{Q} = \mathbb{Q}(\cos\theta)$. As the polynomial $f = 4t^3 - 3t$ is reducible in $\mathbb{Q}[t]$, the angle 90^o can be trisected with straight-edge and compass. (Of course, we already knew this).

3. Let $\theta = 180^o$, so $\cos\theta = -1$ and $\mathbb{Q} = \mathbb{Q}(\cos\theta)$. As the polynomial $f = 4t^3 - 3t + 1$ has -1 as a root, it is reducible in $\mathbb{Q}[t]$, so the angle 180^o can be trisected with straight-edge and compass. (Of course, this also implies that an angle of 90^o can be trisected.) Note that this means that an equilateral triangle can be constructed from two given points.

Doubling the cube: We are given $z_1 = 0, z_2 = 1$ and we wish to construct a line segment of length a such that $a^3 = 2$ from $z_1 = 0, z_2 = 1$. As the polynomial $m_{\mathbb{Q}}(\sqrt[3]{2}) = t^3 - 2$ in $\mathbb{Q}[t]$ is irreducible (by Eisenstein's Criterion), $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$, and hence this construction cannot be done with straight-edge and compass from the given two points.

Squaring the circle: Given a unit circle, $z_1 = 0$ and $z_2 = 1$, can we construct a segment of length $\sqrt{\pi}$ from these two points? We can construct such a segment of this length from $z_1 = 0$ and $z_2 = 1$ if and only if we can construct a segment of length π from these two points and if this can be done, then $[\mathbb{Q}(\pi) : \mathbb{Q}]$ is a power of two. By Lindemann's Theorem 67.1 below, π is transcendental, so this cannot occur and it is impossible to square a circle with straight-edge and compass.

Construction of regular n-gons: We must determine when it is possible to construct a ray of angle $2\pi/n$ radians from $z_1 = 0$, $z_2 = 1$, equivalently, when it is possible to construct $z = \cos\frac{2\pi}{n} + \sqrt{-1}\sin\frac{2\pi}{n}$ from z_1, z_2. To solve this construction problem, we must use the Square Root Tower Theorem (to be proven in 54.11 below). We begin with two lemmas.

Lemma 49.14. *Any 2^n-gon, $n > 1$, is constructible from two points.*

Proof. As we can erect perpendiculars, i.e., angles $2\pi/4$ radians, we can construct a square. As we can bisect any angle, we can construct a ray of $\frac{1}{2^m}(\pi/2)$ for any m by induction. □

Lemma 49.15. *Let m and n be relatively prime integers each at least three. Then*

(1) *A regular (mn)-gon is constructible from two points if and only if both a regular m-gon and a regular n-gon can be constructible from two points.*
(2) *a regular n-gon can be constructed from two points if and only if a regular $2n$-gon can be constructed from two points.*

PROOF. (2): Angles can be bisected or doubled using straight-edge and compass.

(1): Suppose that a regular (mn)-gon can be constructed, i.e., a ray of angle $\theta = 2\pi/mn$ radians. Then rays of angle $m\theta$ and $n\theta$ can be constructed. Conversely, suppose that $\theta = 2\pi/m$ radians and $\varphi = 2\pi/n$ radians are constructible. As m and n are relatively prime, we have an equation $1 = am + bn$ for some integers a and b. So

$$\frac{1}{mn} = \frac{a}{n} + \frac{b}{m} \quad \text{hence} \quad \frac{2\pi}{mn} = a\frac{2\pi}{n} + b\frac{2\pi}{m}$$

is constructible. □

We call a number $f_n = 2^{2^n} + 1$ a *Fermat number*. If it is a prime, we call it a *Fermat prime*.

Remarks 49.16. 1. $f_0 = 3$, $f_1 = 5$, $f_2 = 17$, $f_3 = 257$, and $f_4 = 65537$ are Fermat primes.
2. (Euler) $f_5 = 4294967297 = 641 \cdot 6700417$.
3. $f_6 = 274177 \cdot 67280421310722$ is composite.
4. f_n for $n = 5\text{—}32$, $36\text{—}39$, $42\text{—}43$, 48, 52, 55, 58, $61\text{—}64$, 66, $71\text{—}73$, 75, 77, 81, 88, $90\text{—}91$, $93\text{—}94$, 99 are known to be composite.

   At the time of this writing $f_{2543548}$ is the largest composite Fermat number known and f_4 is the largest Fermat prime known.

The natural question is whether there exist infinitely many Fermat primes, or even one f_n, with $n > 4$.

Fermat primes are applicable to solving the construction of regular n-gons, because the following theorem solves this Euclidean problem.

Theorem 49.17. *Let n be an integer at least three. Then a regular n-gon can be constructed from two points if and only if $n = 2^k p_1 \cdots p_r$ with k non-negative, and $p_1, \ldots, p_r$ distinct Fermat primes for some r or $n = 2^k$ for some integer k at least two.*

PROOF. By the lemmas, it suffices to determine when a regular p^r-gon can be constructed with p an odd prime, $r \geq 1$, i.e., to determine when $\zeta = \cos\frac{2\pi}{p^r} + \sqrt{-1}\sin\frac{2\pi}{p^r}$ is constructible. By Example 47.15(6),

we know that $[\mathbb{Q}(\zeta) : \mathbb{Q}] = p^{r-1}(p-1)$. By the Square Root Tower Theorem, ζ is constructible if and only if $p^{r-1}(p-1) = 2^e$ for some $e \geq 0$. For this to be true, we must have $r = 1$ and $p = 2^e + 1$ a prime. If $e = ab$, for positive integers a and b with a odd, then

$$p = 2^{ab} + 1 = (2^b + 1)(2^{b(a-1)} - 2^{b(a-2)} + \cdots - 2^b + 1)$$

is a composite number unless $u = 1$, i.e., $p = 2^{2^s} + 1$ is a Fermat prime. This proves the theorem. □

Examples 49.18. 1. A regular 9-gon cannot be constructed either by the n-gon theorem or by our results on trisection of an angle. Indeed, if the 9-gon could be constructed, a ray of 40^o could be constructed, hence an angle of 120^o could be trisected, so an angle of 60^o could be trisected which we know is false.

2. We show that a regular pentagon can be constructed, by producing the necessary square root tower. We must construct $z = e^{2\pi\sqrt{-1}/5}$, a root of $f = t^4+t^3+t^2+t+1$ in $\mathbb{Q}[t]$. We have $z^{-1} = z^4 = e^{8\pi\sqrt{-1}/5} = e^{-2\pi\sqrt{-1}/5} = \overline{z}$ and $z^4 + z^3 + z^2 + z + 1 = 0$. So $z,\ \overline{z}\ z^2,\ \overline{z}^2$ are the roots of f in $\mathbb{C}$. Let

$$x_1 = z + \overline{z} \text{ and } x_2 = z^2 + \overline{z}^2,$$

then

$$x_1 + x_2 = -1$$
$$x_1 x_2 = (z + \overline{z})(z^2 + \overline{z}^2) = -1,$$

so $(t - x_1)(t - x_2) = t^2 + t + 1$. This is not quite what we want, so we complete the square. Let $y_1 = 2x_1 + 1$ and $y_2 = 2x_2 + 1$, then

$$y_1 + y_2 = 0 \text{ and } y_1 y_2 = 5.$$

Hence $t^2 - 5 = (t - y_1)(t - y_2)$ with $t^2 - 5$ irreducible in $\mathbb{Q}[t]$. We may assume that $y_1 = \sqrt{5}$. We have $x_1, x_2 \in \mathbb{Q}(y_1, y_2) = \mathbb{Q}(y_1) = \mathbb{Q}(\sqrt{5})$ and

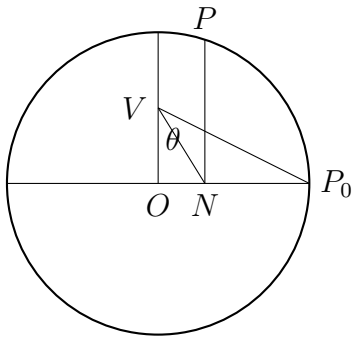
$$x_1^2 = (z + \overline{z})^2 = z^2 + 2z\overline{z} + \overline{z}^2 = x_2 + 2$$
$$(z - \overline{z})^2 = z^2 - 2z\overline{z} + \overline{z}^2 = x_2 - 2.$$

Let $K = \mathbb{Q}(y_1, z - \overline{z})$. As $(z - \overline{z})^2 \in \mathbb{Q}(y_1) = \mathbb{Q}(\sqrt{-5})$, the field extension $K/\mathbb{Q}$ is a square root tower. As $z = (z - \overline{z} + x_1)$ lies in K, a regular pentagon can be constructed.

To actually do the construction in the complex plane, draw a circle at the origin of radius two. Let O be the origin, $V = \sqrt{-1}$, and $P_0 = 2$ in the plane. Let N be the intersection point of the X-axis and the bisector of the angle $\theta = OVP_0$ and P be the intersection

point of the circle (in upper half plane) and the vertical constructed at N. Then the point N is $2z = 2\cos(2\pi/5) + 2\sqrt{-1}\sin(2\pi/5)$. (Cf. Figure 49.19.)

(49.19)



In the figure $\tan\theta = 1/2$, so the length of the line segment $\overline{VP_0}$ is $\sqrt{5}$.

Gauss proved that one could construct the regular 17-gon in 1796 at the age of 19. This was the first important euclidean construction since ancient times. He showed

$$16\cos\frac{2\pi}{17} = -1 + \sqrt{17} + \sqrt{34 - 2\sqrt{17}} + 2\sqrt{17 + 3\sqrt{17} - \sqrt{34 - 2\sqrt{17}} - 2\sqrt{34 + 2\sqrt{17}}}.$$

An actual construction was shown a few years later by Johannes Erchinger. Gauss also proved that if a regular n-gon was constructible then n was of the form in the theorem in 1801. He also stated that this was a necessary condition, but his proof (if he had one) has never been found. Indeed it is doubtful that Gauss actually had a valid proof, since he made this statement prior to the birth of Galois and Galois theory is needed to prove necessity. The first published proof of necessity was given by Pierre Wantzel, who also proved that you could not trisect and angle or double a cube in 1837. Few have heard of Wantzel, so the truism that solving well-known problems will bring you lasting fame is not so well-founded. A regular 257-gon was constructed by Friedrich Richelot (and Schwendenwein) in 1832. The regular 65537-gon was constructed by Johann Hermes in 1895. Apparently Hermes' actual construction rests in a trunk in the attic at the Mathematics Institute at Göttingen (presumably studied by no one). The squaring of the circle was the last of the problems solved when Lindemann proved the transcendence of π in 1882.

Exercises 49.20. 1. Prove that the equations in the last paragraph in the proof of Theorem 49.4 have a solution in $C' \cap \mathbb{R}$.

2. Show the field of complex constructible numbers over $\mathbb{Q}$ is closed under quadratic extensions, but has extensions of every odd prime degree. Does it have extensions of any odd degree? of nonzero even degree? of any degree but two?
3. Show the construction of the regular 5-gon in Figure 49.19 works.
4. Show an angle of $2\pi/n$ radians, n an integer, can be trisected by straight-edge and compass if and only if n is not divisible by 3.

50. Separable Elements

A real polynomial that has multiple roots can cause problems as that means that such a root is also a root of its derivative, hence its graph is more complicated. If the polynomial is irreducible, this cannot happen. However, if the irreducible polynomial has coefficients in a field of positive characteristic, this can occur. In this section, we study this phenomena, formally defining some concepts that we have already used and leaving many of the details as exercises (although some will be proven later). We recall some definitions.

Definition 50.1. Let K/F be an extension of fields, f a non-constant polynomial in $F[t]$ with a root α in K. We say that α is a *multiple root* of g of *multiplicity* r in K if $f = (t-\alpha)^r g$ for some g in $K[t]$ with α not a root of g. If α is a root but not a multiple root, we call it a *simple* root of f. If $f = \sum_{i=0}^{n} a_i t^i$ in $F[t]$, the *derivative* of f is defined to be the polynomial $f' = \sum_{i=1}^{n} i a_i t^{i-1}$ in $F[t]$.

We have, that we (again) leave as an exercise:

Properties 50.2. Let f and g be polynomials in $F[t]$, then

(1) $(f+g)' = f' + g'$.
(2) $(af)' = af'$ for all $a \in F$.
(3) $(fg)' = f'g + fg'$.

As mentioned above, multiple roots are also roots of the derivative. We formally write this as a lemma. This lemma and its proof also says that the notion of multiple root and its muliplicity is independent of the field extension of the base field.

Lemma 50.3. *Let K/F be an extension of fields, f a non-constant polynomial with a root α in K. Then α is a simple root of f if and only if α is not a root of f'.*

PROOF. We can write $f = (t-\alpha)^r g$ for some g in $K[t]$ with $g(\alpha) \neq 0$ for some integer r. Therefore, $f' = r(t-\alpha)^{r-1}g + (t-\alpha)^r g'$ in $K[t]$. As $g(\alpha) \neq 0$ in the domain K, we have $t - \alpha \mid f'$ in $K[t]$ if and only if $t - \alpha \mid r(t-\alpha)^{r-1}g$, i.e., $r - 1 \geq 1$ or $r = 0$ in K. If $r = 0$ in K, then $\operatorname{char} F \mid r$, so $r \geq \operatorname{char} F \geq 2$. The result follows. □

Corollary 50.4. *Let f be a non-constant polynomial in $F[t]$. Then f has a multiple root in some extension field K of F if and only if the ideal (f, f') in $F[t]$ is not the unit ideal.*

PROOF. ($\Rightarrow$): Let α be a multiple root of f in an extension field K of F. Then $m_F(\alpha) \mid f$ and $m_F(\alpha) \mid f'$ in $F[t]$, so $(f, f') \subset \big(m_F(\alpha)\big) < F[t]$.
($\Leftarrow$): $F[t]$ is a PID, so $(f, f') = (g) < F[t]$ for some non-constant polynomial g in $F[t]$. Let h be any non-constant polynomial in $F[t]$ satisfying $h \mid g$. Let K/F be a field extension such that h has a root α in K. Then $f(\alpha) = f'(\alpha) = h(\alpha) = 0$, so f has a multiple root in K. □

This also answers the question of multiple roots for irreducible polynomials.

Corollary 50.5. *Let f be an irreducible polynomial in $F[t]$.*

1. *If $\operatorname{char} F = 0$, then f has no multiple roots in any field extension of F. In particular, f can have only simple roots, if any, in a field extension of F.*
2. *If $\operatorname{char} F = p > 0$ and F has a multiple root in some field extension K of F, then there exists a polynomial g in $F[t]$ satisfying $f = g(t^p)$, i.e., $f = \sum a_i t^{pi}$ in $F[t]$.*

PROOF. We begin the proof, leaving the rest of the proof as exercises. Let K/F be a field extension such that f has a root α in K, so $f = am_F(\alpha)$ for some nonzero a in F. In particular, if α is a multiple root, we must have $f \mid f'$ in $F[t]$. So $f' = 0$ or $\deg f' \geq \deg f$. Consequently, $f' = 0$. The result now follows by Exercises 50.10(2a) and(2b). □

Definition 50.6. An irreducible polynomial f in $F[t]$ is called *separable over F* if it has no multiple roots in any field extension of F. A non-constant polynomial is called *separable over F* if all of its irreducible factors are separable. If f is not separable over F, it is called *inseparable*. Let K/F be a field extension and α an element in K. We say that α is *separable* over F if it is the root of a separable polynomial defined over F and the extension K/F is *separable* if every element of K is separable over F.

Remark 50.7. In reading other books, it is much more common to say that a non-constant polynomial f in $F[t]$ is separable if it has no multiple roots in any extension field of F. It is, however, convenient to use the above definition for a separable polynomial in the sequel.

Remarks 50.8. Let K/F be an extension of fields and f, g non-constant polynomials in $F[t]$.

1. Every element of F is separable over F.
2. An element α in K is separable over F if and only if α is algebraic over F and $m_F(\alpha)$ is separable over F. In particular, if K/F is separable, then it is algebraic.
3. If $\operatorname{char} F = 0$ and K/F is algebraic, then K/F is separable. A field F in which every element algebraic over F is separable is called a *perfect field*. Therefore, any field of characteristic zero is perfect. It turns out that every finite field is also perfect, but there exist fields that are not perfect.
4. The polynomial f is separable over F if and only if every divisor of f is separable over F.
5. If both f and g are separable over F then so is fg.
6. If $L/K/F$ are field extensions with L/F separable, then L/K and K/F are separable.

Theorem 50.9. *Let $K/E/F$ be extensions of field. Then K/F is separable if and only if K/E and E/F are both separable.*

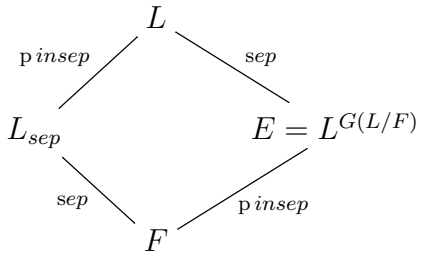
We shall not prove this theorem here. We know that if K/F is separable, then K/E and E/F are separable. The converse is not so easy. To reduce the problem, we can try a trick that we did before. If α in K is separable over E, then $m_E(\alpha)$ in $E[t]$ is separable. If $m_E(\alpha) = t^n + a_{n-1}t^{n-1} + \cdots + a_0$ in $E[t]$, then $m_E(\alpha)$ is separable over $E_0 = F(a_0, \ldots, a_{n-1})$, with a_i, $i = 0, \ldots, n-1$ separable over F. So we are reduced to showing if $\beta_1, \ldots, \beta_n$ are separable over F, then $F(\beta_1, \ldots, \beta_n)/F$ is separable and if, in addition, α is separable over $F(\beta_1, \ldots, \beta_n)$, then α is separable over F. This is still not easy. One way of doing this is given in the exercises. Of course, we may assume that $\operatorname{char} F$ is positive.

Exercises 50.10. 1. Suppose that f be a non-constant polynomial in $F[t]$, $L/K_i/F$ field extensions with $i = 1, 2$, and α an element of $K_1 \cap K_2$. Show that α is a root of f of multiplicity r in K_1 if and only if α is a root of f of multiplicity r in K_2.

2. Let F be a field and f a polynomial in $F[t]$. Show the following:

   (a) If $\operatorname{char} F = 0$, then $f' = 0$ if and only if f is a constant polynomial.
   (b) If $\operatorname{char} F = p$ is positive, then $f(t)^p = f^p(t^p)$ (The Children's Binomial Theorem) and if $f' = 0$, then there exists a polynomial g in $F[t]$ satisfying $f(t) = g(t^p)$.
3. If x is transcendental over a field F, then $t^p - x \in F(x)[t]$ is irreducible for any prime p.
4. Suppose that F is a field of positive characteristic p. Show
   (a) The map $F \to F$ given by $x \mapsto x^p$ is a monomorphism. Denote its image by F^p.
   (b) If K/F is algebraic and $\alpha \in K$ is separable over $F(\alpha^p)$, then $\alpha \in F(\alpha^p)$.
   (c) Every finite field is *perfect*, i.e., every algebraic extension is separable.
5. Suppose that F is a field of positive characteristic p. Show all of the following:
   (a) If K/F is separable field extension, then $K = F(K^p)$, where $K^p := \{x^p \mid x \in K\}$.
   (b) Suppose that K/F is finite and $K = F(K^p)$. If $\{x_1, ..., x_n\} \subset K$ is linearly independent over F then so is $\{x_1^p, ..., x_n^p\}$.
6. Let K/F be an extension of fields. Show all of the following:
   (a) If $\alpha \in K$ is separable over F then $F(\alpha)/F$ is separable.
   (b) If $\alpha_1, ..., \alpha_n \in K$ are separable over F, then $F(\alpha_1, ..., \alpha_n)/F$ is separable.
   (c) Let $F_{\text{sep}} = \{\alpha \in K \mid \alpha \text{ separable over } F\}$. Then F_{sep} is a field.
7. Let $L/K/F$ be field extensions. Show that L/F is separable if and only if L/K and K/F are separable using the previous two exercises.
8. Show that any algebraic extension of a perfect field is perfect.
9. Let F_0 be a field of positive characteristic p. Let $F = F_0(t_1^p, t_2^p)$ and $L = F_0(t_1, t_2)$. Show all of the following:
   (a) If $\theta \in L \setminus F$ then $[F(\theta) : F] = p$.
   (b) There exist infinitely many fields K satisfying $F < K < L$.
10. Let F be a field of positive characteristic p. Show
11. If $F = F^p$, then F is perfect.
12. If F is not perfect, then $F \neq F^{p^r} := \{x^{p^r} \mid x \in F\}$ for any $r \geq 1$.
13. Let K/F be a field extension and $\alpha \in K$. We say that α is *purely inseparable* over F if $m_F(\alpha)$ has only one root in a splitting field of α and K/F is *purely inseparable* if every element of K is purely inseparable over F. Show that α in K is pure inseparable and separable over K if and only if α lies in F.

14. Let K/F be a field extension with $\text{char}(F) = p > 0$. Show that if α in K is algebraic then there exists a nonnegative integer n satisfying α^{p^n} lies in F.
15. Let K/F be an algebraic extension with $\text{char}(F) = p > 0$. Show the following are equivalent:
    (a) K/F is purely inseparable.
    (b) If α in K, then there exists a nonnegative integer n such that $m_F(\alpha) = t^{p^n} - a \in F[t]$.
    (c) If α lies in K, then there exists a nonnegative integer n such that α^{p^n} lies in F.
    (d) $K_{sep} = F$. (Cf. Exercise 6 above.)
    (e) K is generated by purely inseparable elements over F.
16. Let K/F be an algebraic field extension. Show that K/F_{sep} is purely inseparable and $K_{psep}; = \{\alpha \in K | \alpha \text{ is purely inseparable over } F\}$ is a field.
17. Let L/F be a finite normal extension and $E = L^{G(L/F)}$. The
$$E = \{\alpha \mid \alpha \text{ is purely inseparable over } F\}.$$
In particular, L/E is separable and E/f is purely inseparable, i.e., we have



CHAPTER XII

Galois Theory

In the previous chapter, we studied field theory, i.e., fields and field extensions. In the chapter, we interrelate field theory with group theory. The goal will be to interwine the intermediate fields of finite field extension K/F when K is the splitting field of a separable polynomial in $F[t]$ over F and subgroups of the Galois group $\mathrm{Aut}_F(K)$. This will allow us to prove some fundamental results. For example, we shall give an algebraic proof of the Fundamental Theorem of Algebra using only the Intermediate Value Theorem from analysis, prove that there is no formula involving only addition, multiplication, and the extraction of nth roots for various n for the fifth degree polynomial over the rational numbers using the results that were established about solvable groups, and finish the proof for the existence of the regular n-gon for the allowable n. We shall show how the study of roots of unity leads to Quadratic Reciprocity and a special case of Dirichlet's Theorem of Primes in an Arithmetic Progression.

51. Characters

This section is the key to the interaction of field theory and group theory. We follow the approach of E. Artin, who proved the main results using systems of linear equations. The idea is to look at a set S of field homomorphisms from a field K and to determine the subfield of K on which each homomorphism in S acts in the same way. We begin with the following definition:

Definition 51.1. Let F be a field and G be a (multiplicative) group, then a group homomorphism $\sigma : G \to F^\times$ is called an (*linear*) *character*. We say that distinct characters $\sigma_1, \ldots, \sigma_n : G \to F^\times$ are *dependent* (or the set $\{\sigma_1, \ldots, \sigma_n\}$ is *dependent*) if there exist $a_1, \ldots, a_n$ in F, not all zero, satisfying $\sum_i a_i\sigma_i(g) = 0$ for all g in G, i.e., the function $\sum a_i\sigma_i : G \to F$ is the zero map; and *independent* otherwise.

Remarks 51.2. 1. A field homomorphism $\sigma : E \to F$ induces an character by restriction, as a ring homomorphism takes units to units. As field homomorphisms are monic, such a character is always injective.

2. The set
$$\{\sigma \mid \sigma : G \to F^\times \text{ is a character.}\}$$
is not a group under $+$, but is a group under multiplication of functions, i.e., $\sigma\tau(g) := \sigma(g)\tau(g)$ for all g in G, with the identity the map $1 : G \to F^\times$ given by $g \mapsto 1_F$.

The key result is:

Lemma 51.3. (Dedekind's Lemma) *Let $\sigma_1, \ldots, \sigma_n : G \to F^\times$ be distinct characters. Then $\sigma_1, \ldots, \sigma_n : G \to F^\times$ are independent.*

Proof. We prove this by induction on n.

$n = 1$: If $a\sigma_1(g) = 0$ for all g in G, then $a = 0$ as $\sigma_1(g) \neq 0$ for all g in G.

$n > 1$: Suppose that we have an equation

$$(*) \qquad \sum_{i=1}^{n} a_i\sigma_i(g) = 0 \text{ for all } g \text{ in } G \text{ with } a_i \text{ in } F \text{ not all zero.}$$

By induction any proper subset of $\{\sigma_1, \ldots, \sigma_n\}$ is independent, so a_i is nonzero for every i. Multiplying the equation in (*) by a_n^{-1}, we may assume that that $a_n = 1$. So (*) now reads

$$(**) \qquad \sum_{i=1}^{n-1} a_i\sigma_i(g) + \sigma_n(g) = 0 \text{ for all } g \text{ in } G.$$

By assumption, $\sigma_1 \neq \sigma_n$, so there exists an element x in G satisfying $\sigma_1(x) \neq \sigma_n(x)$. As G is a group, $xG = G$; and as the σ_i are group homomorphisms, we have

$$0 = \sum_{i=1}^{n-1} a_i\sigma_i(xg) + \sigma_n(xg) = \sum_{i=1}^{n-1} a_i\sigma_i(x)\sigma_i(g) + \sigma_n(x)\sigma_n(g)$$

for all g in G. Since $\sigma_n(x) \neq 0$, we see that

$$(\dagger) \qquad 0 = \sum_{i=1}^{n-1} \sigma_n(x)^{-1}a_i\sigma_i(x)\sigma_i(g) + \sigma_n(g) \text{ for all } g \text{ in } G.$$

Subtracting the equation in (†) from the equation in (**) yields

$$0 = \sum_{i=1}^{n-1} [a_i - \sigma_n(x)^{-1}a_i\sigma_i(x)]\sigma_i(g) \text{ for all } g \text{ in } G.$$

By induction, we conclude that

$$a_i = \sigma_n(x)^{-1}a_i\sigma_i(x) \text{ for all } i$$

in the domain F with $a_i \neq 0$, hence $\sigma_n(x) = \sigma_i(x)$ for all i, contradicting $\sigma_1(x) \neq \sigma_n(x)$. □

The proof above also shows, with the obvious definitions, that any set of characters $S = \{\sigma_i : G \to F^\times \mid \sigma_i \text{ a character for all } i \in I\}$ is independent.

Corollary 51.4. *Let $\sigma_1, \ldots, \sigma_n : F \to K$ be distinct field homomorphisms, then $\sigma_1, \ldots, \sigma_n : F^\times \to K^\times$ are independent characters.*

Definition 51.5. Let S be a non-empty subset of the full set of all field homomorphisms $F \to K$, $\{\sigma : F \to K \mid \sigma \text{ a field homomorphism}\}$. We call an element a in F a *fixed point under S*, if $\sigma(a) = \tau(a)$ for all σ and τ in S.

Example 51.6. Let S be a non-empty set of field homomorphisms $F \to K$. Assume that both F and K contain the same prime subfield Δ (so $\Delta \cong \mathbb{Q}$ if the characteristic of F is zero and $\Delta \cong \mathbb{Z}/p\mathbb{Z}$, if the characteristic of F is the prime p.) As every homomorphism takes $1_F \to 1_K$ and $1_\Delta = 1_F = 1_K$, we have $\sigma(x) = x$ for all $x \in \Delta$, Therefore, the term 'fixed point' is actually appropriate for elements in Δ.

Usually, sets of fixed points shall arise when we have a field extension K of F and our non-empty set S of homomorphisms $F \to K$ will contain the inclusion map. In this case, a fixed point really shall be an element that is fixed by all the homomorphisms in S.

As should be expected, the set of fixed points has additional structure. We immediately see that we have:

Lemma 51.7. *Suppose that S is a non-empty set of field homomorphisms $F \to K$. Let $E = \{x \in F \mid x \text{ is a fixed point of } S\}$. Then E is a subfield of F.*

If S is a non-empty set of field homomorphisms $F \to K$, we denote the field $\{x \in F \mid x \text{ is a fixed point of } S\}$ by F^S and call it the *fixed field* of S. We are interested in the fixed field of a finite set of field homomorphisms. Following Artin, we use linear algebra to obtain non-trivial solutions to an appropriate set of linear equations. The first main result about such is the following:

Lemma 51.8. (Artin's Lemma) *Suppose that S is a finite non-empty set of field homomorphisms $F \to K$, then $[F : F^S] \geq |S|$.*

Proof. Let $S = \{\sigma_1, \ldots, \sigma_n\}$ with $n \geq 1$. Suppose that $r = [F : F^S] < |S| = n$ and $\{\omega_1, \ldots, \omega_r\}$ is an F^S-basis for F. Consider

the following system of equations over K:

$$(*)\qquad \begin{array}{ccccccc} \sigma_1(\omega_1)x_1 & + & \cdots & + & \sigma_n(\omega_1)x_n & = & 0 \\ \vdots & & & & \vdots & & \\ \sigma_1(\omega_r)x_1 & + & \cdots & + & \sigma_n(\omega_r)x_n & = & 0. \end{array}$$

As $r < n$, there exists a non-trivial solution to the system of equations (*) over K, say

$$x_1, \ldots, x_n \text{ with each } x_i \text{ in } K \text{ and not all zero}$$

is such a solution. Let a be an element of F. We can write $a = \sum a_i\omega_i$ with each $a_i \in F^S$. For each i, multiply the ith equation in (*) by $\sigma_i(a_i)$ to get a new equivalent system of equations

$$\begin{array}{ccccccc} \sigma_1(a_1)\sigma_1(\omega_1)x_1 & + & \cdots & + & \sigma_1(a_1)\sigma_n(\omega_1)x_n & = & 0 \\ \vdots & & & & \vdots & & \\ \sigma_n(a_r)\sigma_1(\omega_r)x_1 & + & \cdots & + & \sigma_n(a_r)\sigma_n(\omega_r)x_n & = & 0. \end{array}$$

Since $\sigma_i(a_i) = \sigma_j(a_i)$ for all i and j and the σ_i are homomorphisms, we have

$$(\dagger)\qquad \begin{array}{ccccccc} \sigma_1(a_1\omega_1)x_1 & + & \cdots & + & \sigma_n(a_1\omega_1)x_n & = & 0 \\ \vdots & & & & \vdots & & \\ \sigma_1(a_r\omega_r)x_1 & + & \cdots & + & \sigma_n(a_r\omega_r)x_n & = & 0. \end{array}$$

Adding all the equations in (†) yields

$$0 = \sum_{i=1}^{r}\sum_{j=1}^{n}\sigma_j(a_iw_i)x_j = \sum_{j=1}^{n}\sum_{i=1}^{r}\sigma_j(a_iw_i)x_j = \sum_{j=1}^{n}\sigma_j(a)x_j$$

for all a in F with x_j not all zero. This means that the characters $\sigma_1, \ldots, \sigma_n$ are dependent, contradicting Dedekind's Lemma. □

A special case of Artin's Lemma, is the one in which we shall be interested.

Corollary 51.9. *Let S be a non-empty finite set of field automorphisms of a field K, then $[K : K^S] \geq |S|$.*

Of course, we should be very interested in the case when we have equality in Artin's Lemma. In general this is not true. We can even have a finite set of field automorphisms S of a field K with $[K : K^S]$ infinite (can you give an example?). What we need is to look at those S with additional structure. For example, let K/F be an extension of fields. Recall the *Galois group* of K/F is defined to be the set

$$\begin{aligned} G(K/F) &= \mathrm{Aut}_F(K) \\ &:= \{\sigma : K \to K \mid \text{an } F\text{-automorphism satisfying } \sigma|_F = 1_F\}. \end{aligned}$$

It is immediate that the following is true.

Lemma 51.10. *Let K/F be an extension of fields, then $G(K/F)$ is a subgroup of the (field) automorphism group of K and F is a subfield of $K^{G(K/F)}$.*

As a consequence, we have:

Corollary 51.11. *Let K/F be a finite extension of fields. Then the Galois group $G(K/F)$ is a finite group and satisfies $[K : F] \geq |G(K/F)|$.*

Proof. As F is a subfield of K^S for any non-empty subset S of $G(K/F)$, we have

$$[K : F] \geq [K : K^S] \geq |S| \text{ for any finite } S.$$

It follows that $G(K/F)$ is finite and satisfies $[K : F] \geq |G(K/F)|$. □

We arrive at the group theoretic condition in which we are interested.

Definition 51.12. Let K/F be a finite extension of fields. We say that the extension is *Galois* if F is the fixed field of $G(K/F)$, i.e., $F = K^{G(K/F)}$.

We have defined a field extension K/F to be *Galois* if K is a finite extension with F the fixed field of $G(K/F)$. It would be more appropriate to call this a *finite Galois* extension and call K/F a *Galois extension* if it is algebraic with F the fixed field of $G(K/F)$.

Remarks 51.13. Let K/F be a field extension.

1. If $F = K^{G(K/F)}$, then for any element $x \in K \setminus F$, there exists an automorphism in $G(K/F)$ that *moves* x, i.e., there exists a σ in $G(K/F)$ satisfying $\sigma(x) \neq x$.
2. If $K/E/F$ is an intermediate field, then $G(K/E) \subset G(K/F)$.
3. Let $E = K^{G(K/F)}$, then by definition of $K/E/F$, we have $G(K/E) \supset G(K/F)$, so by (2), we have
$$G(K/F) = G(K/K^{G(K/F)}).$$
4. If K/F is finite, then a field homomorphism $\sigma : K \to K$ fixing F necessarily is an element of $G(K/F)$ as σ is monic and F-linear.

We give examples of Galois groups.

Examples 51.14. 1. Let $K = F$ be a field. Then $G(K/F) = \{1\}$ and K/F is Galois with $[K : F] = 1 = |G(K/F)|$. Moreover, K is the splitting field of $t - a$ in $F[t]$ over F for any $a \in F$.

2. Let $F = \mathbb{R}$ and $K = \mathbb{C}$. Then $G(K/F) = \{1, {}^{-}\}$ (with $1 = 1_K$ and ${}^{-}$ complex conjugation) and K/F is Galois with $[K : F] = 2 = |G(K/F)|$. Moreover, K is the splitting field of $t^2 + 1$ over F.
3. Suppose that $F = \mathbb{Q}$ and $K = \mathbb{Q}(\sqrt{2})$. Let $\sigma : K \to K$ be defined by $a + b\sqrt{2} \mapsto a - b\sqrt{2}$ for all a and b in $\mathbb{Q}$. Then $G(K/F) = \{1, \sigma\}$ (with $1 = 1_K$) and K/F is Galois with $[K : F] = 2 = |G(K/F)|$. Moreover, K is the splitting field of $t^2 - 2$ over F.
4. Suppose that $F = \mathbb{Q}$ and $K = \mathbb{Q}(\sqrt[3]{2})$. Then $G(K/F) = \{1\}$, so K/F is not Galois. It satisfies $[K : F] = 3 > |G(K/F)|$. Moreover, K is not a splitting field of any polynomial over F:

   We indicate why these facts are true. Any field automorphism of K must take $\sqrt[3]{2}$ to another root of $t^3 - 2$, but $\mathbb{R}$ contains only one root of $t^3 - 2$. If K was the splitting field of some polynomial g in $F[t]$ over F, then it would be the splitting field of an irreducible polynomial of degree three. (Why?) Such a g would have three distinct roots, say $\alpha_1, \alpha_2, \alpha_3$ in K and we would have $K = \mathbb{Q}(\alpha_i)$, $i = 1, 2$, or 3. But then there must be an F-automorphism σ satisfying $\sigma(\alpha_1) = \alpha_2$.
5. Suppose that $F = \mathbb{Q}$ and $K = \mathbb{R}$. Then $G(K/F) = \{1\}$ and $K^{G(K/F)} = K$ with K/F is infinite and not even algebraic. (Cf. Exercise 47.18(11).) However, (by Exercise 53.22(6)) the Galois group $G(\mathbb{C}/\mathbb{Q})$ is uncountable!
6. Suppose that the characteristic of a field F is $p > 0$ and $K = F(\alpha)$ for some α satisfying $\alpha^p = a$ with $a \in F$. If $\alpha \notin F$, then $G(K/F) = \{1\}$ and K/F is not Galois. It satisfies $[K : F] = p > |G(K/F)|$ and K is the splitting field of $t^p - a$ over F.

Artin's Lemma becomes much stronger, when the non-empty finite set of field automorphisms is a group.

Theorem 51.15. (Artin's Theorem) *Let K be a field and G a finite subgroup of field automorphisms of K. Then*

(1) $[K : K^G] = |G|$.
(2) $G = G(K/K^G)$.
(3) K/K^G is Galois.

PROOF. (1): Let $n = |G|$ and $G = \{\sigma_1, \ldots, \sigma_n\}$. By Artin's Lemma, we know that $[K : K^G] \geq |G| = n$. Suppose that $[K : K^G] > n$. Choose $a_1, \ldots, a_{n+1}$ in $K^\times$ that are K^G-linearly independent. The system of n linear equations in $(n+1)$-unknowns

$$(*) \qquad \begin{array}{ccccccc} x_1\sigma_1(a_1) & + & \cdots & + & x_{n+1}\sigma_1(a_{n+1}) & = & 0 \\ \vdots & & & & \vdots & & \\ x_1\sigma_n(a_1) & + & \cdots & + & x_{n+1}\sigma_n(a_{n+1}) & = & 0 \end{array}$$

has a non-trivial solution over K. Among all the non-trivial solutions $x_1, \ldots, x_{n+1}$ over K, choose one with the least number of x_i nonzero. Relabeling, we may assume this solution is

$$x_1, \ldots, x_r, 0, \ldots, 0 \text{ with } x_i \in K^\times,\ 1 \leq i \leq r$$

with r minimimal. Multiplying this solution by x_r^{-1}, we may also assume that $x_r = 1$. As G is a group, we may assume also that $\sigma_1 = 1_K$, so

$$x_1 a_1 + \cdots + x_{r-1} a_{r-1} + a_r = 0, \quad x_i \in K^\times,\ r > 1 \text{ (as } a_1 \neq 0).$$

Since $\{a_1, \ldots, a_r, a_{r+1}, \ldots, a_{n+1}\}$ is K^G-linearly independent, this means that for some integer i, $1 \leq i \leq r$, we have $x_i \notin K^G$. By definition, there exists an integer k such that $\sigma_k(x_i) \neq x_i$. (So $k > 1$.) The system of equations in (*) has the form

$$(**) \qquad 0 = \sum_{i=1}^{r-1} x_i \sigma_j(a_i) + \sigma_j(a_r) = 0 \text{ for } j = 1, \ldots, n.$$

Taking σ_k of each of the equations in (**) yields

$$0 = \sum_{i=1}^{r-1} \sigma_k(x_i)\sigma_k\sigma_j(a_i) + \sigma_k\sigma_j(a_r) = 0 \text{ for } j = 1, \ldots, n.$$

As G is a group, $\sigma_k G = G$, hence these equations are the same as

$$(\dagger) \qquad 0 = \sum_{i=1}^{r-1} \sigma_k(x_i)\sigma_j(a_i) + \sigma_j(a_r) = 0 \text{ for } j = 1, \ldots, n.$$

Subtracting the appropriate equations in (†) from those in (*) yields

$$0 = \sum_{i=1}^{r-1} [x_i - \sigma_k(x_i)]\sigma_j(a_i) = 0 \text{ for } j = 1, \ldots, n.$$

Since $\sigma_k(x_i) \neq x_i$, we have

$$x_1 - \sigma_k(x_1), \ldots, x_{r-1} - \sigma_k(x_{r-1}), 0, \ldots, 0$$

is a non-trivial solution over K to the system of the equations in (*). This contradicts the minimality of r and establishes (1).

(2): By (1), we have $[K : K^G] = |G|$ is finite. By Corollary 51.9, we have $[K : K^G] \geq |G(K/K^G)|$, so $G(K/F)$ is also finite. Since G is a subgroup of $G(K/K^G)$ by definition, we have $|G| = |G(K/K^G)|$ and is finite, so $G = G(K/K^G)$.

(3) is now immediate. □

We want to interrelate Galois groups and field extensions. An application of Artin's Theorem produces one such result.

Corollary 51.16. *Let K be a field, G_1, G_2 two finite subgroups of* $\mathrm{Aut}(K)$*, the group of field automorphisms of K, and $F_i = K^{G_i}$ for $i = 1, 2$. Then $F_1 = F_2$ if and only if $G_1 = G_2$.*

PROOF. ($\Leftarrow$) is clear.

($\Rightarrow$): By Artin's Theorem, $G_i = G(K/F_i)$, so this also follows. □

We now see that for finite field extensions, being Galois only depends on a simple cardinality test.

Corollary 51.17. *Suppose that K/F is a finite extension of fields. Then K/F is Galois if and only if $[K : F] = |G(K/F)|$.*

PROOF. ($\Rightarrow$) follows from Artin's Theorem and the special case of Artin's Lemma, Corollary 51.9.

($\Leftarrow$): Let $E = K^{G(K/F)}$, then we know that $G(K/E) = G(K/F)$. As K/E is Galois, by the proven sufficiency, we have $[K : E] = |G(K/E)|$. Therefore,

$$[K : E] = |G(K/E)| = |G(K/F)| = [K : E][E : F].$$

It follows that $[E : F] = 1$, i.e., $E = F$. □

Exercise 51.18. This exercise computes the Galois group of a non algebraic extension. Using Exercise 45.24(10) show that $G(F(t)/F)$ consists of all F-automorphisms of $F(t)$ mapping t to $(at + b)/(ct + d)$ where $a, b, c, d \in F$ satisfies $ad - bc \neq 0$

52. Computations

Of course, given a finite extension K/F of fields, one would like to determine its Galois group. In general, as expected, this is very difficult. In this section, we do a few computations as well as giving some ideas on how to do such computations. To begin, we discuss the problem by reviewing (and repeating) some of the material that we have already discussed.

Summary 52.1. Let K/F be a finite extension of fields and f a non-constant polynomial in $F[t]$.

1. We know if $K/E/F$ is an intermediate field that
   (a) $G(K/E) \subset G(K/F)$.
   (b) $G(K/F) = G(K/K^{G(K/F)})$.
   (c) $[K : F] \geq |G(K/F)|$ with equality if and only if K/F is Galois, i.e., $F = K^{G(K/F)}$.

2. Let α in K be a root of f and σ an element in $G(K/F)$, then $\sigma(\alpha)$ is also a root of f. In particular, if $f = g_1 \cdots g_r$ in $F[t]$ with g_i non-constant polynomials in $F[t]$, then $G(K/F)$ takes roots of g_i in K to roots of g_i for all i:
   If $f = \sum a_i t^i$, then
$$0 = \sigma\big(f(\alpha)\big) = \sigma(\sum a_i \alpha^i) = \sum \sigma(a_i)\sigma(\alpha^i) = \sum a_i \sigma(\alpha)^i = f\big(\sigma(\alpha)\big).$$
3. If K is a splitting field of f and σ an element in $G(K/F)$, then σ is completely determined by what it does to the roots of f:
   If $\alpha_1, \ldots, \alpha_r$ are the roots of f, then $K = F(\alpha_1, \ldots, \alpha_n)$.
4. Let $S = \{\alpha \mid \ \alpha \text{ is a root of } f\}$. If S is not empty, then the map $G \to \Sigma(S)$ given by $\sigma \mapsto \sigma|_S$ is a group homomorphism. If K is a splitting field of f over F, then this map is monic and we can view $G \subset \Sigma(S)$. (This was Galois' viewpoint.)
5. If α is a root of f in F then $\sigma(\alpha) = \alpha$ for all $\sigma \in G(K/F)$.
6. This is the most useful remark. If K is the splitting field of f over F and g is an irreducible polynomial in $F[t]$ with roots α and β in K, then there exists an F-isomorphism
$$\tau : F(\alpha) \to F(\beta) \text{ satisfying } \alpha \mapsto \beta$$
   and this extends to an F-automorphism τ in $G(K/F)$, so $\tau(\alpha) = \beta$. (Make sure that you can prove this.) In particular,
$$G(K/F) \text{ acts transitively on the roots of } g \text{ in } K.$$
   We shall later see that, in fact,
$$\text{if } g \text{ is irreducible in } F[t] \text{ and}$$
$$\text{has a root in a splitting field } K \text{ of } f, \text{ then } g \text{ splits over } K.$$
   Can you prove this? (You have all the tools to do so.)

To do some of our computations, we shall need a few lemmas.

Lemma 52.2. *Let G be a finite cyclic group of order n and $\mathrm{Aut}(G)$ the automorphism group of G. Then $\mathrm{Aut}(G) \cong (\mathbb{Z}/n\mathbb{Z})^\times$. In particular, $\mathrm{Aut}(G)$ is abelian (and cyclic if n is a prime) of order $\varphi(n)$.*

Proof. Let $G = \langle a \rangle$ and $\sigma \in \mathrm{Aut}(G)$. Then $\sigma(a) = a^i$, for some $1 \leq i \leq n$ and $\langle a \rangle = \langle a^i \rangle$. It follows that i and n are relatively prime. Conversely, if i and n are relatively prime with $1 \leq i \leq n$, then $\sigma_i : G \to G$ given by $a^m \mapsto a^{mi}$ is checked to be a group monomorphism, hence an isomorphism as G is finite. The map $(\mathbb{Z}/n\mathbb{Z})^\times \to \mathrm{Aut}(G)$ given by $i \mod n \mapsto \sigma_i$ is checked to be an isomorphism, so we also have $|\mathrm{Aut}(G)| = |(\mathbb{Z}/n\mathbb{Z})^\times| = \varphi(n)$. □

Remark 52.3. In §105, we shall show the following: If p is an odd prime then $(\mathbb{Z}/p^n\mathbb{Z})^\times$ is cyclic. If $p = 2$, this is not true. Indeed if $m \geq 3$, then $(\mathbb{Z}/2^m\mathbb{Z})^\times \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2^{m-2}\mathbb{Z})$. It then follows, using the Chinese Remainder Theorem, that $(\mathbb{Z}/n\mathbb{Z})^\times$ is cyclic if and only if $n = 2, 4, p^r$, or $2p^r$ where p is an odd prime.

Proposition 52.4. *Let K be a splitting field of $t^n - 1$ in $F[t]$ over F and*

$$U = \{z \in K \mid z^n = 1\} = \{z \mid z \text{ a root of } t^n - 1 \text{ in } K\}.$$

Then U is a cyclic subgroup of $K^\times$. Suppose, in addition, that either $\operatorname{char} F = 0$ *or* $\operatorname{char} F \nmid n$. *Then $|U| = n$ and the Galois group $G(K/F)$ is isomorphic to a subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$. In particular, $G(K/F)$ is abelian and $|G(K/F)| \mid \varphi(n)$.*

Proof. As $(z_i z_j)^n = 1 = (z_i^{-1})^n$, for all $z_i, z_j \in U$, we know that U is a group. Since it is a finite subgroup of $K^\times$, it is cyclic by Theorem 31.15. As $K = F(U)$, every σ in $G(K/F)$ is determined by $\sigma(z)$, with $z \in U$, so

$$\varphi : G(K/F) \to \Sigma(U) \text{ given by } \sigma \mapsto \sigma|_U$$

is a group monomorphism. As U is cyclic, say $U = \langle z \rangle$, and every σ in $G(K/F)$ satisfies and $\sigma(z^i) = \sigma(z)^i$, it follows that every such σ is completely determined by $\sigma(z)$ and $\sigma|_U$ lies in the automorphism group $\operatorname{Aut}(U)$ of U. If $\operatorname{char} F = 0$ or $\operatorname{char} F \nmid n$, we have $(t^n - 1)' = nt^{n-1} \neq 0$ has no root in common with $t^n - 1$ in $F[t]$, i.e., $t^n - 1$ has only simple roots in every field extension of F. In particular, $|U| = n$. □

In the proposition, the element z satisfying $U = \langle z \rangle$ is called a *primitive nth root of unity*. If, in the proposition, $\operatorname{char} F = 0$ or $\operatorname{char} F \nmid n$, then there exists $\varphi(n)$ primitive roots of unity.

Remark 52.5. A (deep) theorem of Kronecker-Weber says: Suppose a finite extension of fields $K/\mathbb{Q}$ is Galois with abelian Galois group. We call such an extension is an *abelian extension* of $\mathbb{Q}$. Then there exists a primitive nth root of unity ζ, e.g., $e^{2\pi\sqrt{-1}/n}$, for some n, such that K is a subfield of $\mathbb{Q}(\zeta)$. One says that all abelian extensions of $\mathbb{Q}$ are *cyclotomic*. We shall show later this is true in the special case that K/F is quadratic.

Remarks 52.6. The following are left as exercises:

1. If $z \in F$, then $|G(K/F)| = 1 = [K : F]$.
2. Suppose that $|F| = q$ and L/F a field extension satisfying $[L : F] = r$. Let $n = q^r - 1$, then $x^n = 1$ for all x in L. The field L is a splitting

field of $t^n - 1$ over F, hence also of $t^{q^r} - t$ in $F[t]$, and $G(K/F)$ is cyclic of order r.

3. If $F = \mathbb{Q}$, then $[K : \mathbb{Q}] = \varphi(n)$ and $G(K/\mathbb{Q}) \cong (\mathbb{Z}/n\mathbb{Z})^\times$.

We did not investigate the case when $\operatorname{char} F \mid n$ in the above. Can you say anything in this case?

Corollary 52.7. *Let F be a field of characteristic zero or $\operatorname{char} F \nmid n$ satisfying the polynomial $t^n - 1$ in $F[t]$ splits over F with U the set of nth roots of units in F. Suppose that K is a splitting field of the polynomial $t^n - a$ in $F[t]$ over F with a nonzero. Then $G(K/F)$ is a cyclic subgroup isomorphic to a subgroup of U. In particular, $|G(K/F)| \mid n$.*

Proof. Let $U = \{z_1, \ldots, z_n\}$, a cyclic subgroup of $F^\times$ of order n. As $(t^n - a)' = nt^{n-1}$ has only zero as a root, $t^n - a$ has n distinct roots in K. Let r in K be a root of $t^n - a$, then $rz_i,\ i = 1, \ldots, n,$ are all its roots. As $z_i \in F$ for all i, we know that $K = F(r)$ and $\sigma(rz_i) = \sigma(r)z_i$ for all $\sigma \in G(K/F)$ and all i. Consequently, $\sigma(r)$ determines σ for each $\sigma \in G(K/F)$. If $\sigma(r) = rz_i$, then $\sigma(r)/r = z_i$ lies in U. Therefore, we have a map

$$\varphi : G(K/F) \to U \text{ defined by } \sigma \mapsto \frac{\sigma(r)}{r}.$$

If σ and τ lie in $G(K/F)$, say $\sigma(r) = rz_i$ and $\tau(r) = rz_j$, then

$$\varphi(\sigma\tau) = \frac{\sigma\tau(r)}{r} = \frac{\sigma(r)z_j}{r} = \frac{\sigma(r)z_j}{r} = z_iz_j = \varphi(\sigma)\varphi(\tau),$$

so φ is a well-defined group homomorphism. If

$$\frac{\sigma(r)}{r} = \frac{\tau(r)}{r}, \text{ then } \sigma(r) = \tau(r),$$

hence $\sigma = \tau$ and φ is monic. As U is cyclic of order n, we have $G(K/F)$ is cyclic of order dividing n. □

With these preliminaries, we can make some explicit computations.

Computation 52.8. Let F be a field. (You should fill in all the omitted details.)

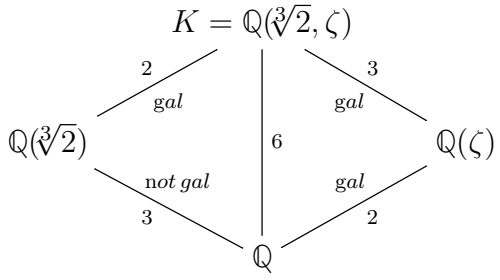
1. $G(F/F) = 1$ and F/F is Galois.
2. Suppose that the characteristic of F is not two, d an element in F that is not a square in F. Then $K = F(\theta)$ is a splitting field of $t^2 - d$ in $F[t]$ over F where θ is a root of $t^2 - d$. Then K/F is Galois and $G(K/F) \cong \mathbb{Z}/2\mathbb{Z}$. (Of course, one often writes $\sqrt{d}$ for θ.)

3. Let $f = (t^2-2)(t^2-3)$ in $\mathbb{Q}[t]$ and $K = \mathbb{Q}(\sqrt{2},\sqrt{3})$. Then K is a splitting field of f over $\mathbb{Q}$. As $\sqrt{2} \notin \mathbb{Q}(\sqrt{3})$, the polynomial t^2-2 is irreducible in $\mathbb{Q}(\sqrt{3})[t]$. Let $G\big(K/\mathbb{Q}\big(\sqrt{3}\big)\big) = \langle\sigma\rangle$ were σ maps $\sqrt{3} \mapsto -\sqrt{3}$. Similarly, $G\big(K/\mathbb{Q}(\sqrt{2})\big) = \langle\tau\rangle$ where τ maps $\sqrt{2} \mapsto -\sqrt{2}$. Then $G(K/\mathbb{Q}) = \{1,\sigma,\tau,\sigma\tau\} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ and $K/\mathbb{Q}$ is Galois.

4. Let $f = t^3-2$ in $\mathbb{Q}[t]$. It is irreducible, so $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$. Let $\zeta = \cos(2\pi/3) + \sqrt{-1}\sin(2\pi/3)$ and $K = \mathbb{Q}(\sqrt[3]{2},\zeta)$. As $m_{\mathbb{Q}}(\zeta) = t^2+t+1$, we have $[Q(\zeta) : \mathbb{Q}] = 2$. The field K is a splitting field of f over $\mathbb{Q}$ so $[K : \mathbb{Q}] = 6$. As K is also the splitting field of f over $\mathbb{Q}(\zeta)$ and of degree three, the Galois group $G\big(K/\mathbb{Q}(\zeta)\big)$ is cyclic and isomorphic to a subgroup of $\mathbb{Z}/3\mathbb{Z}$. Since f remains irreducible in $\mathbb{Q}(\zeta)[t]$ (why?), there exists a $\mathbb{Q}(\zeta)$-automorphism τ of K satisfying $\sqrt[3]{2} \mapsto \zeta\sqrt[3]{2}$. Therefore, $G\big(K/\mathbb{Q}(\zeta)\big)$ is cyclic of order three. The polynomial $m_{\mathbb{Q}}(\zeta)$ has no real roots, so remains irreducible in $\mathbb{Q}(\sqrt[3]{2})[t]$, so K is a splitting field of t^2+t+1 over $\mathbb{Q}(\sqrt[3]{2})$ and there exists a $\mathbb{Q}(\sqrt[3]{2})$-automorphism $\sigma : K \to K$ such that $\zeta \mapsto \zeta^{-1} = \overline{\zeta}$. Therefore, $G\big(K/\mathbb{Q}(\sqrt[3]{2})\big)$ is cyclic of order two. It follows $G(K/\mathbb{Q})$ contains an element of order two and an element of order three, hence $[K : \mathbb{Q}] = 6 = |G(K/\mathbb{Q})|$ and $K/\mathbb{Q}$ is Galois. As

$$\begin{array}{rcccl} \tau\sigma(\sqrt[3]{2}) & = & \tau(\sqrt[3]{2}) & = & \zeta\sqrt[3]{2} \quad \text{and} \\ \sigma\tau(\sqrt[3]{2}) & = & \sigma(\zeta\sqrt[3]{2}) & = & \zeta^{-1}\sqrt[3]{2}, \end{array}$$

$G(K/\mathbb{Q})$ is not abelian, hence isomorphic to $S_3 \cong D_3$. Any element of $G(K/\mathbb{Q})$ must take $\sqrt[3]{2}$ to a root of t^3-2 in K, so $Q(\sqrt[3]{2}) \subset \mathbb{R}$ means that the subgroup $G(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}) = 1$ and $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is not Galois. Note also that $G\big(K/\mathbb{Q}(\zeta)\big)$ is normal in $G(K/\mathbb{Q})$ but $G\big(K/\mathbb{Q}(\sqrt[3]{2})\big)$ is not normal in $G(K/\mathbb{Q})$.
The following picture summarizes this:



5. Let $f = t^5-1$ in $\mathbb{Q}[t]$ and $\zeta = \cos(2\pi/5) + \sqrt{-1}\sin(2\pi/5)$ in $\mathbb{C}$. Then $K = \mathbb{Q}(\zeta)$ is the splitting field of f and $[K : \mathbb{Q}] = 4 = \varphi(5)$,

as $m_{\mathbb{Q}}(\zeta) = t^4 + t^3 + t^2 + t + 1$. We know that group homomorphism $G(K/\mathbb{Q}) \to (\mathbb{Z}/5\mathbb{Z})^\times$ given by $\sigma \mapsto i \mod 5$ if $\sigma \in G(K/\mathbb{Q})$ satisfies $\sigma(\zeta) = \zeta^i$ is monic, so $|G(K/\mathbb{Q})| = 1, 2$, or 4. As K is the splitting field of the irreducible polynomial $m_Q(\zeta)$ in $\mathbb{Q}[t]$, the Galois group $G(K/\mathbb{Q})$ acts transitively on its roots, hence $|G(K/\mathbb{Q})| = 4$. Therefore, $K/\mathbb{Q}$ is Galois and $G(K/\mathbb{Q}) \cong (\mathbb{Z}/5\mathbb{Z})^\times$

Let $\sigma_i : K \to K$ be the $\mathbb{Q}$-automorphism determined by $\zeta \mapsto \zeta^i$, then σ_4 takes ζ to $\zeta^{-1} = \overline{\zeta}$ so has order two. If $H = \langle \sigma_4 \rangle$, a subgroup of index two in $G(K/\mathbb{Q})$, then its fixed field is

$$\begin{aligned} K^H &= \{a + b(\zeta^2 + \zeta^3) \mid a, b \in \mathbb{Q}\} \\ &= \{a + b(\zeta^2 + \overline{\zeta}^2) \mid a, b \in \mathbb{Q}\} = \mathbb{Q}(\sqrt{5}). \end{aligned}$$

(Can you show this?). We have $[K : K^H] = 2 = [K^H : \mathbb{Q}]$, so $K/\mathbb{Q}$ is a square root tower (which leads to the construction of a regular pentagon).

6. Let K be a splitting field of the irreducible polynomial $f = t^4 - 2$ in $\mathbb{Q}[t]$. The roots of f are: $\sqrt[4]{2}$, $\sqrt{-1}\sqrt[4]{2}$, $-\sqrt[4]{2}$, $-\sqrt{-1}\sqrt[4]{2}$; so $K = \mathbb{Q}(\sqrt[4]{2}, \sqrt{-1})$. As $\sqrt{-1} \notin \mathbb{Q}(\sqrt[4]{2}) \subset \mathbb{R}$, the polynomial $t^2 + 1$ is irreducible in $\mathbb{Q}(\sqrt[4]{2})[t]$ and

$$[K : \mathbb{Q}] = [K : \mathbb{Q}(\sqrt[4]{2})][\mathbb{Q}(\sqrt[4]{2}) : \mathbb{Q}] = 8 \geq |G(K/\mathbb{Q})|.$$

Since $t^4 - 1$ splits over $\mathbb{Q}(\sqrt{-1})$, the group $G\big(K/\mathbb{Q}(\sqrt{-1})\big)$ is cyclic of order $1, 2$, or 4 by Corollary 52.7. We show that it is of order 4. We know that $[K : \mathbb{Q}(\sqrt{-1})] = 4$ and K is a splitting field of $m_{\mathbb{Q}}(\sqrt[4]{2})$. It follows that $m_{\mathbb{Q}}(\sqrt[4]{2})$ must remain irreducible in $\mathbb{Q}(\sqrt{-1})[t]$, i.e., $m_{\mathbb{Q}(\sqrt{-1})}(\sqrt[4]{2}) = m_{\mathbb{Q}}(\sqrt[4]{2})$. As $G\big(K/\mathbb{Q}(\sqrt{-1})\big)$ acts transitively on the roots of $m_{\mathbb{Q}(\sqrt{-1})}(\sqrt[4]{2})$, we must have $G\big(K/\mathbb{Q}(\sqrt{-1})\big)$ is a group of order at least 4, hence 4.
Now let

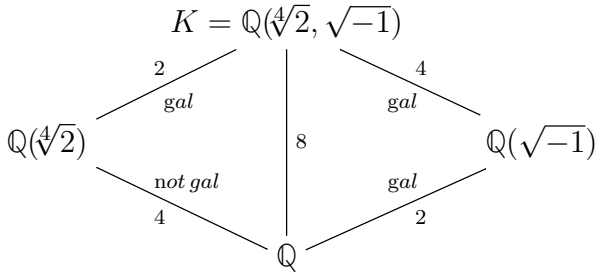
$$\begin{gathered} \tau : K \to K \text{ be the element in } G\big(K/\mathbb{Q}(\sqrt[4]{2})\big) \\ \text{satisfying } \sqrt{-1} \mapsto -\sqrt{-1} \end{gathered}$$

and

$$\begin{gathered} \sigma : K \to K \text{ be the element in in } G\big(K/\mathbb{Q}(\sqrt{-1})\big) \\ \text{satisfying } (\sqrt[4]{2}) \mapsto \sqrt{-1}\sqrt[4]{2}. \end{gathered}$$

Then $\langle \tau, \sigma \rangle \subset G(K/\mathbb{Q})$. It is easy to see this forces $|G(K/\mathbb{Q})| = 8$. Checking $\tau\sigma\tau^{-1} = \sigma^3$ shows that $G(K/\mathbb{Q}) \cong D_4$. Note also that $K/\mathbb{Q}$ is Galois but $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}$ is not as $G(\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}) = \{1, \sigma^2\}$.

The following picture summarizes this:



(Note we have another interesting intermediate field of K/F, viz., $E = \mathbb{Q}(\sqrt{-1}, \sqrt{2})$. As $[K : \mathbb{Q}(\sqrt{-1})] = [K : \mathbb{Q}]/[\mathbb{Q}(\sqrt{-1}) : \mathbb{Q}]$ and $\sqrt{2}$ lies in K with $\sqrt[4]{2}$ not in E. $\mathbb{Q} < \mathbb{Q}(\sqrt{-1}) < E < K$ is a square root tower. Can you say more?)

7. Let p be a prime, $F = (\mathbb{Z}/p\mathbb{Z})(x)$ with x transcendental over $\mathbb{Z}/p\mathbb{Z}$ and $f = t^p - x$ a polynomial in $F[t]$. Let K be a splitting field of f over F and α in K a root of f. Then $t^p - x = t^p - \alpha^p$ in $K[t]$, so f has only one root and $K = F(\alpha)$ with $\alpha \notin F$ (why?). Since $G(K/F)$ takes the roots of f to roots of f, the Galois group $G(K/F) = 1$ and $K = K^{G(K/F)}$. So K/F is not Galois. Check that f is irreducible over F so $[K : F] = p$ and K/F is a splitting field of an irreducible polynomial. Note that K/F is not separable.

Exercises 52.9. 1. Let $K = \mathbb{Q}(r)$ with r a root of $t^3+t^2-2t-1 \in \mathbb{Q}[t]$. Let $r_1 = r^2 - 2$. Show that r_1 is also a root of this polynomial. Find $G(K/\mathbb{Q})$.

2. Suppose the $|K| = p^n$, p a prime, and $F \subset K$. Show that $|F| = p^m$ for some m with $m \mid n$. Moreover, $G(K/F)$ is generated by the *Frobenius automorphism* $\alpha \mapsto \alpha^{p^m}$. In particular, $G(K/F)$ is cyclic.

3. If F is a finite field, n a positive integer, then there exists an irreducible polynomial $f \in F[t]$ of degree n.

4. Let F be a subfield of the real numbers, f an irreducible quartic over F. Suppose that f has exactly two real roots. Show that the Galois group of f is either S_4 or of order 8.

53. Galois Extensions

The object of Galois theory is to connect intermediate fields of a finite field extension extension to subgroups of the Galois group $G(K/F)$. The results and computations of the last section indicated a close relation when K/F was a splitting field of a separable polynomial. We begin by pursuing that here.

Definition 53.1. We call a finite field extension K/F *normal* if K is the splitting field over F of some non-constant polynomial in $F[t]$.

We shall see that the important properties of normal extensions are independent of a polynomial for which it is a splitting field except for whether the polynomial is separable or not.

It is useful to generalize the definition of a normal field extension. If $\{f_i\}_I$ is a set of non-constant polynomials in $F[t]$, we call an algebraic extension K of F a *splitting field* of the set of polynomials $\{f_i\}_I$, if every f_i, $i \in I$, splits in K and K is the smallest algebraic extension with this property. Such a splitting field exists and is unique up to an F-isomorphism. (Cf. Exercise 53.22(1)). For example, the splitting field of the set of all minimal polynomials over F is algebraically closed and called an *algebraic closure of* F. If an algebraic extension K of F is the splitting field of a set of non-constant polynomials in $F[t]$, we say that K/F is *normal*. Note that if $\{f_i\}_I$ is a finite set of non-constant polynomials in $F[t]$, then K is a splitting field of this set of polynomials over F if and only if it is the splitting field of $\prod_I f_i$.

The following is a very useful criterion to check if a finite field extension is normal.

Proposition 53.2. *Let K/F be a finite extension of fields. Then K/F is normal if and only if any irreducible polynomial in $F[t]$ having a root in K splits over K.*

Proof. ($\Rightarrow$): Let K be a splitting field the non-constant polynomial g in $F[t]$ over F and $f \in F[t]$ an irreducible polynomial having a root α in K. Let L/K with $\beta \in L$ a root of f. We must show that β lies in K. As f is irreducible, there exists an F-isomorphism $\tau : F(\alpha) \to F(\beta)$ satisfying $\alpha \mapsto \beta$. As $K(\alpha)$ is a splitting field of g over $F(\alpha)$ and $K(\beta)$ is a splitting field of g over $F(\beta)$, there exists an isomorphism $\sigma : K(\alpha) \to K(\beta)$ lifting τ. In particular, σ is also an F-isomorphism, so we have

$$\begin{aligned}[K(\alpha) : F] &= [K(\alpha) : F(\alpha)][F(\alpha) : F] \\ &= [K(\beta) : F(\beta)][F(\beta) : F] = [K(\beta) : F].\end{aligned}$$

But $\alpha \in K$, so $K(\alpha) = K$, and

$$[K : F] = [K(\alpha) : F] = [K(\beta) : F] = [K(\beta) : K][K : F].$$

Therefore, $[K(\beta) : K] = 1$, i.e., $\beta \in K$, as needed.

($\Leftarrow$): As K/F is finite, there exist $\alpha_1, \ldots, \alpha_r$ in K, some r, with $K = F(\alpha_1, \ldots, \alpha_r)$ and each α_i algebraic over F. Clearly, K is the splitting field of $f = \prod m_F(\alpha_i)$ over F, as each $m_F(\alpha_i)$ has a root in K, hence

splits over K and any splitting field of f over F in K must contain all the α_i's. □

Remark 53.3. 1. $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$ is not normal as the irreducible polynomial $t^3 - 2$ in $\mathbb{Q}[t]$ does not split over $\mathbb{Q}(\sqrt[3]{2})$
[Note: the general proof above is really the same as the special case done before.]

2. Let p be a positive prime and $F = (\mathbb{Z}/p\mathbb{Z})(x)$ with x transcendental over $\mathbb{Z}/p\mathbb{Z}$. Then $f = t^p - x$ in $F[t]$ is irreducible. (Cf. Exercise 50.10 (3).) Let K/F with $\alpha \in K$ a root of f, i.e., $\alpha^p = x$. Set $E = F(\alpha)$, then $f = (t-\alpha)^p$ in $E[t]$, hence E is a splitting field of f over F, therefore, E/F is normal. We saw that $G(E/F) = 1$, hence E/F is not Galois.

Proposition 53.4. *Let K/F be a finite, normal extension and $K/E/F$ an intermediate field. Suppose that $\varphi : E \to K$ is an F-homomorphism. Then there exists an F-automorphism $\sigma \in G(K/F)$ satisfying $\sigma|_E = \varphi$, i.e., every F-homomorphism $E \to K$ arises from an F-automorphism of K by restriction.*

Proof. Let K be a splitting field of the non-constant polynomial f in $F[t]$, hence a splitting field of f over E. In addition, K is a splitting field of $f = \widetilde{\varphi}(f)$ over $\varphi(E)$, so there exists an automorphism $\sigma : K \to K$ lifting $\varphi : E \to \varphi(E)$, so σ lies in $G(K/F)$. □

Using Zorn's Lemma, one can show that the last two propositions hold if we only assume that K/F is algebraic. (Cf. Exercises 53.22(2), (3), respectively.) Alternatively, one can use the existence of an algebraic closure (Corollary 48.2) and its lifting property (Theorem 48.3), together with a restriction argument giving the uniqueness of splitting fields.

Definition 53.5. Let $K/E_i/F$ be finite field extensions, $i = 1, 2$ with K/F normal. We say that E_1 and E_2 are *conjugate over F* if there exists an automorphism $\sigma \in G(K/F)$ satisfying $\sigma|_{E_1} : E_1 \to E_2$ is an isomorphism. Equivalently, by Proposition 53.4, there exists an F-isomorphism $E_1 \to E_2$.

Remarks 53.6. Let L/F be a finite normal extension, $L/K/E_i/F$ field extensions, $i = 1, 2$, with K/F also normal. Then E_1 and E_2 are conjugate relative to the extensions $K/E_i/F$ if and only if they are conjugate relative to the extensions $L/E_i/F$, as any $\sigma \in G(K/F)$ lifts to some $\widehat{\sigma} \in G(L/F)$ and any $\widehat{\sigma} \in G(L/F)$ restricts to $\widehat{\sigma}|_K \in G(K/F)$. That $\widehat{\sigma}$ is an F-automorphism of K follows from the following:

Corollary 53.7. *Let K/F be a finite normal field extension and $K/E/F$ an intermediate field extension. Then K/E is normal. Moreover, the following are equivalent:*

(1) E/F is normal.
(2) $\sigma(E) = E$ for every $\sigma \in G(K/F)$.
(3) $\sigma|_E \in G(E/F)$ for every $\sigma \in G(K/F)$.
(4) The map $\Phi : G(K/F) \to G(E/F)$ given by $\sigma \mapsto \sigma|_E$ is well-defined and an epimorphism.
(5) E is the only conjugate of E.

PROOF. If K is the splitting field of the non-constant polynomial $f \in F[t]$ over F, then it is also the splitting field of f over E, so the first statement follows.

We turn to the equivalences. By Proposition 53.4, any $\varphi \in G(E/F)$ lifts to an element in $G(K/F)$ and if $G(K/F) \to G(E/F)$ by $\sigma \mapsto \sigma|_E$ is well-defined – usually it is not as, in general, $\sigma(E) \not\subset E$ – the equivalences of $(2), (3), (4)$ and (5) become clear. So it suffices to show that (1) and (2) are equivalent.

(1) $\Rightarrow$ (2): Let $x \in E$ and $\sigma \in G(K/F)$. As $x \in E$ and E/F is normal, we know that $m_F(x)$ splits over E. Since $G(K/F)$ takes the roots of $m_F(x)$ to roots of $m_F(x)$, the root $\sigma(x)$ also lies in E. As the argument also works for σ^{-1}, Statement (2) follows.

(2) $\Rightarrow$ (1): Let $f \in F[t]$ be an irreducible polynomial having a root $\alpha \in E$. As K/F is normal, f splits over K. Let $\beta \in K$ be root of f. We must show that $\beta \in E$, i.e., f splits over E. We know that there exists an F-isomorphism $\varphi : F(\alpha) \to F(\beta)$ taking $\alpha \mapsto \beta$. By Proposition 53.4, there exists an automorphism $\sigma \in G(K/F)$ satisfying $\varphi = \sigma|_{F(\alpha)}$, so $\beta = \varphi(\alpha) = \sigma(\alpha)$ lies in E as needed. □

Note in the above if E/F is normal, then $\ker \Phi = G(K/E) \lhd G(K/F)$ and $G(E/F) \cong G(K/F)/G(K/E)$.

Definition 53.8. Let K/F be a finite field extension. Then a field extension L of K is called a *normal closure* of K/F if L/F is normal and $[L : K]$ is minimal with respect to this property. If this is the case, we write L/K is a normal closure of K/F.

Question 53.9. How would you define the normal closure of an algebraic extension of F?

Proposition 53.10. *Let K/F be a finite field extension. Then a normal closure of K/F exists and is unique up to a K-isomorphism.*

PROOF. Let $K = F(\alpha_1, \ldots, \alpha_n)$ with each α_i algebraic over F, $f = \prod m_F(\alpha_i)$ in $F[t]$, and L a splitting field of f over K, hence also a splitting field of f over F. Therefore, L/F is normal. By the uniqueness of splitting fields, L is unique up to a K-isomorphism (relative to f). If E/K is a normal extension, then each $m_F(\alpha_i)$ must split over E, so f must split over E. The result now follows easily. □

If in the above proposition, K/F is also separable, then L/F is also separable by Exercise 50.10(6). We shall also give a proof of this below.

The main technical result needed to establish the field theoretic equivalence of a Galois extension (rather than a group theoretic one) is the following:

Proposition 53.11. *Let K/F be a finite extension of fields of degree n. Suppose that L/K with L/F finite and normal and*

$$\tau_1, \ldots, \tau_m : K \to L$$

are all the distinct F-homomorphisms. Then the following are true:

(1) $m \leq n$.
(2) $m = n$ if and only if K/F is separable.
(3) $\tau_1(K), \ldots, \tau_m(K)$ are all the conjugates of K over F in L (and independent of the normal extension L/F).
(4) Let $E = F\big(\tau_1(K) \cup \cdots \cup \tau_m(K)\big)$. Then E/K is a normal closure of K/F.

Using the exercises in 53.22, the assumption in the proposition that L/K be finite may be dropped.

PROOF. We may assume that τ_1 is the inclusion, so $m \geq 1$. Let $S = \{\tau_1, \ldots, \tau_m\}$.

(1): By assumption, $F \subset K^S$, so $[K : F] \geq |S| = m$ by Artin's Lemma 51.8.

(2): We induct on n. The case of $n = 1$ is immediate, so we may assume that $n > 1$. Let $\alpha \in K \setminus F$ be chosen with α not separable, i.e., $m_F(\alpha)$ not separable, if K/F is not separable. Let

$$r = [F(\alpha) : F] > 1$$

and

$$\alpha = \alpha_1, \ldots, \alpha_{r_0} \text{ the distinct roots of } m_F(\alpha) \text{ in } L.$$

[Note that $m_F(\alpha)$ splits over L as L/F is normal.]
So we have

$$r = r_0 \text{ if and only if } \alpha \text{ is separable over } F$$
$$\text{if and only if } K/F \text{ is separable.}$$

In particular,

$$r_0 < r \text{ if } K/F \text{ is not separable.}$$

Let

$$\varphi_i : F(\alpha) \to F(\alpha_i) \ [\subset L]$$

be the F-isomorphism satisfying $\alpha \mapsto \alpha_i,\ 1 \le i \le r_0$.
Since the roots of $m_F(\alpha)$ go to roots of $m_F(\alpha)$, we have

$$\varphi_i,\ 1 \le i \le r_0, \text{ are all the } F\text{-homomorphisms } F(\alpha) \to L.$$

As L/F is normal, each φ_i lifts to some

$$\widehat{\varphi}_i \in G(L/F),\ 1 \le i \le r_0.$$

[There can be many distinct lifts of each φ_i — chose one.]
Thus

$$\widehat{\varphi}_i|_{F(\alpha)} \ne \widehat{\varphi}_j|_{F(\alpha)} \text{ if } i \ne j.$$

So we have the following:

(*i*) $[K : F(\alpha)] = \dfrac{n}{r}$.

(*ii*) $K/F(\alpha)$ is separable if K/F is separable.

By induction, we have:

There exist m_0 distinct $F(\alpha)$-homomorphisms

$$\psi_1, \dots, \psi_{m_0} : K \to L$$

$$\text{with } m_0 = \frac{n}{r} \text{ if and only if } K/F(\alpha) \text{ is separable.}$$

[Note if K/F is separable so is $K/F(\alpha)$.]
Let

$$\rho_{ij} = \widehat{\varphi}_i \psi_j : K \to L,\ 1 \le i \le r_0,\ 1 \le j \le m_0.$$

and

$$\mathcal{S} = \{\rho_{ij} \mid 1 \le i \le r_0,\ 1 \le j \le m_0\}.$$

Each ρ_{ij} is an F-homomorphism, so $\mathcal{S} \subset \{\tau_1, \dots, \tau_m\}$.

Claim 53.12. *$\mathcal{S}$ contains all the F-homomorphisms $K \to L$ and $|\mathcal{S}| = r_0 m_o$:*

If we establish the claim, then

$$m = r_0 m_0 \le m_0 r \le n.$$

So

$$K/F \text{ is separable if and only if } r = r_0 \text{ and } m_0 = \frac{n}{r}$$
$$\text{if and only if } n = m_0 r_0 = m.$$

In particular, if we prove the claim, (2) follows.

Let $\rho : K \to L$ be an F-homomorphism. Then $\rho(\alpha)$ is a root of $m_F(\alpha)$, so there exists an i satisfying $\rho|_{F(\alpha)} = \varphi_i$, hence $\widehat{\varphi}_i^{-1}\rho : K \to L$ is an $F(\alpha)$-homomorphism. Consequently, there exist a j satisfying $\psi_j = \widehat{\varphi}_i^{-1}\rho$. Therefore, $\rho = \widehat{\varphi}_i\psi_j = \rho_{ij}$ lies in $\mathcal{S}$.

To finish proving the claim, we must show that the ρ_{ij} are distinct. Suppose that $\rho_{ij} = \rho_{kl}$, then

$$\varphi_i = \rho_{ij}|_{F(\alpha)} = \rho_{kl}|_{F(\alpha)} = \varphi_k,$$

hence $\widehat{\varphi}_i = \widehat{\varphi_k}$, so $i = k$. Since

$$\widehat{\varphi}_i\psi_j = \rho_{ij} = \rho_{kl} = \widehat{\varphi_k}\psi_l,$$

it follows that $\psi_j = \psi_l$ as $\widehat{\varphi}_i$ is an automorphism. Consequently, $j = l$. This shows that the ρ_{ij}'s are distinct and establishes the claim.

(3). Each of the $\tau_i : K \to L$ above lifts to some $\widehat{\tau}_i$ in $G(L/F)$ as before. In particular, $\widehat{\tau}_i(K) = \tau_i(K)$. If $\sigma \in G(L/F)$, there exists a j such that $\sigma|_K = \tau_j$. It follows that $\tau_1(K), \ldots, \tau_m(K)$ are all the conjugates of K (repetitions possible).

(4): Suppose that L/F is also a normal closure of K/F. Set $E = F\big(\tau_1(K) \cup \cdots \cup \tau_m(K)\big) \subset L$. Write $K = F(\beta_1, \ldots, \beta_s)$, for some β_i, and set $f = \prod_{i=1}^{s} m_F(\beta_i)$ in $F[t]$. Then L is a splitting field of f over F. By the argument proving (2), we know that for any root of $m_F(\beta_i)$, there exist an F-homomorphism $K \to L$ sending β_i to that root. As $\tau_1, \ldots, \tau_m : K \to L$ are all the F-homomorphisms and each must take roots of $m_K(\beta_i)$ to roots of $m_F(\beta_i)$, the field E contains all the roots of the $m_F(\beta_i)$. As L/F is a splitting field of f, we must have $E = L$. □

Corollary 53.13. *Let K/F be a extension of fields with α an element in K algebraic over F. Then α is separable over F if and only if $F(\alpha)/F$ is separable.*

Proof. Let $L/F(\alpha)$ be a field extension such that L/F is finite and normal. We know that $m_F(\alpha)$ splits over L. Let $\alpha = \alpha_1, \ldots, \alpha_m$ be the distinct roots of $m_F(\alpha)$ in L. We know that there exist F-isomorphisms $\varphi_i : F(\alpha) \to F(\alpha_i)$ satisfying $\alpha \mapsto \alpha_i$ for $i = 1, \ldots, m$ and these are all the distinct F-homomorphisms into L. Thus $F(\alpha)/F$ is separable if and only if $[F(\alpha) : F] = \deg m_F(\alpha) = m$ if and only if α is separable over F. □

Proposition 53.14. *Let K be a splitting field of a separable polynomial in $F[t]$. Then K/F is separable and Galois.*

Proof. Suppose that K is a splitting field of the separable polynomial $f \in F[t]$. We may assume that $F < K$. Let $\alpha \in K$ be a root of f not in F. As α is separable over F, the extension $F(\alpha)/F$

is separable by the corollary. Therefore, there exist $m = [F(\alpha) : F]$ F-homomorphisms $\varphi_i : F(\alpha) \to K$, $1 \leq i \leq m$. As K/F is normal, each φ_i lifts to some $\widehat{\varphi}_i \in G(K/F)$. Since K is a splitting field of the separable polynomial f over $F(\alpha)$, by induction on $[K : F]$, we conclude that $K/F(\alpha)$ is separable. As K/K is the normal closure of $K/F(\alpha)$ (as well as K/F), there exist precisely $[K : F(\alpha)]$ distinct $F(\alpha)$-homomorphisms $\psi_j : K \to K$, $1 \leq j \leq [K : F(\alpha)]$. Moreover, the argument to prove Claim 53.12 shows that

$$\mathcal{S} = \{\widehat{\varphi}_i\psi_j \mid 1 \leq i \leq [F(\alpha) : F],\ 1 \leq j \leq [K : F(\alpha)]\}$$

is the set of all F-homomorphisms $K \to K$ and has $[K : F] = [K : F(\alpha)][F(\alpha) : F]$ elements, hence K/F is separable. Since $\mathcal{S} \subset G(K/F)$, we have $|G(K/F)| = [K : F]$ by Artin's Lemma 51.8, so K/F is also Galois. □

Corollary 53.15. *Let K/F be a finite separable field extension and L/K a normal closure of K/F. Then L/F is separable and Galois.*

PROOF. Let $K = F(\alpha_1, \ldots, \alpha_n)$, then L is a splitting field of the separable polynomial $\prod m_F(\alpha_i)$ in $F[t]$, hence L/F is separable and Galois. □

Corollary 53.16. *Let $K = F(\alpha_1, \ldots, \alpha_n)$ with α_i separable over F for each $i = 1, \ldots, n$. Then K/F is separable.*

PROOF. Let L/K be a normal closure of K/F, hence the splitting field of the separable polynomial $\prod m_F(\alpha_i)$. Therefore, L/F is separable hence so is K/F. □

Corollary 53.17. *If K/F is finite field extension and $K/E/F$, then K/F is separable if and only if K/E and E/F are separable.*

PROOF. ($\Rightarrow$) has already been done.

($\Leftarrow$): Let L/K be a (finite) normal extension and

$$\varphi_1, \ldots, \varphi_m : E \to L \text{ be all the distinct } F\text{-homomorphisms}$$
$$\psi_1, \ldots, \psi_n : K \to L \text{ be all the distinct } E\text{-homomorphisms.}$$

Let $\widehat{\varphi}_i \in G(L/F)$ lift φ_i for $1 \leq i \leq m$. Then as in the proof of Claim 53.12,

$$|\{\widehat{\varphi}_i\psi_j \mid 1 \leq i \leq m,\ 1 \leq j \leq n\}| = mn.$$

As K/E and E/F are separable, $[K : F] = [K : E][E : F] = mn$, hence K/F is separable by Proposition 53.11. □

We arrive at our goal of finding the field theoretic interpretation for a finite field extension K/F to be Galois.

Theorem 53.18. *Let K/F be a finite extension of fields. Then K/F is Galois if and only if K/F is normal and separable.*

PROOF. ($\Leftarrow$) has already been done.

($\Rightarrow$): Let L/K be a normal closure of K/F. As K/F is Galois, the number of F-homomorphisms $K \to L$ is at least $|G(K/F)| = [K : F]$. It follows by Proposition 53.11 there exist exactly $|G(K/F)| = [K : F]$ such F-homomorphisms and, in addition, K/F must be separable. Since each of these homomorphisms lies in $G(K/F)$, Proposition 53.11(4), shows that $L = F\big(\cup_{G(K/F)}\, \sigma(K)\big) = K$. □

Corollary 53.19. *Let K/F be a finite extension and $K/E/F$ an intermediate field. Suppose that K/F is Galois. Then*

(1) K/E is Galois.
(2) E/F is Galois if and only if E/F is normal.

PROOF. As K/F is Galois, it is separable and normal, so K/E and E/F are separable and K/E is normal. Therefore, K/E is Galois and E/F is normal if and only if E/F is Galois. □

Let K/F be a finite field extension and $K/E/F$ an intermediate field. Recall that we have seen, in general, K/F being Galois does not imply that E/F is Galois. For example, $K = \mathbb{Q}(\sqrt[3]{2}, e^{2\pi\sqrt{-1}/3})$, $E = \mathbb{Q}(\sqrt[3]{2})$, and $F = \mathbb{Q}$. A question left to the reader is: If K/E and E/F are Galois, is K/F Galois?

Definition 53.20. A field F is called *perfect* if any algebraic extension K/F is separable.

So we have

Remark 53.21. Let F be a field.

(1) If $\operatorname{char} F = 0$, then F is perfect.
(2) If F is a finite field, then F is perfect.
(3) If F is perfect, K/F a finite extension, then K/F is Galois if and only if K/F is normal.
(4) If x is transcendental over $\mathbb{Z}/p\mathbb{Z}$, with $p > 0$ a prime and $F = (\mathbb{Z}/p\mathbb{Z})(x)$, then F is not perfect as $t^p - x$ in $F[t]$ is not separable.

Exercises 53.22. 1. Let F be a field and $\{f_i\}_I$ be a set of non-constant polynomials in $F[t]$. Prove that a splitting field of $\{f_i\}_I$ over F exists and is unique up to an F-isomorphism.

2. Let K/F be an algebraic extension of fields. Then K/F is normal if and only if any irreducible polynomial in $F[t]$ having a root in K splits over K.
3. Let K/F be a normal (possibly infinite) extension of fields and $K/E/F$ an intermediate field. Suppose that $\varphi : E \to K$ is an F-homomorphism. Then there exists an F-automorphism $\sigma \in G(K/F)$ satisfying $\sigma|_E = \varphi$, i.e., every F-homomorphism $E \to K$ arises from an F-automorphism of K by restriction.
4. Let K/F be an algebraic field extension. Answer Question 53.9 and prove that a normal closure of K/F exists and is unique up to a K-isomorphism.
5. Let F be a field and L and algebraic closure of F. Show that L/F is normal. Is it true that $L^{G(L/F)} = F$, i.e., L/F is (infinite) Galois? If not, give an example when it is not.
6. Show that there exist uncountably many field automorphisms of $\mathbb{C}$.
7. Let K/F be an algebraic extension of fields and $K/E/F$ an intermediate field. Show that K/F is separable if and only if K/E and E/F are separable.
8. Suppose that K/F is Galois and $\alpha \in K$ has precisely r distinct images under $G(K/F)$. Show $[F(\alpha) : F] = r$.
9. Let F be a field of positive characteristic p and $f = t^p - t - a \in F[t]$.
   (a) Show the polynomial f has no multiple roots.
   (b) If α is a root of f, show so is $\alpha + k$ for all $0 \leq k \leq p - 1$.
   (c) Show f is irreducible if and only if f has no root in F.
   (d) Suppose that $a \neq b^p - b$ for any $b \in F$. Find $G(K/F)$ where K is a splitting field of $t^p - t - a \in F[t]$.
10. Let K/F be a finite field extension and $K/E/F$ an intermediate field. Suppose that K/E and E/F are Galois. Is K/F Galois? Either prove or provide a counterexample.
11. Prove if F is not a finite field and u, v are algebraic and separable over F that there exists an element $a \in F$ such that $F(u, v) = F(u + av)$. Is this true if $|F| < \infty$?
12. Let $K = \mathbb{Q}(\sqrt{2}, \sqrt{3}, u)$ where $u^2 = (9 - 5\sqrt{3})(2 - \sqrt{2})$. Show that $K/\mathbb{Q}$ is normal and find $G(K/\mathbb{Q})$.

54. The Fundamental Theorem of Galois Theory

In this section we put together the pieces that we have developed, i.e., the group theoretic and field theoretic interpretations of Galois

extensions. As an application we prove the Fundamental Theorem of Algebra. We begin with the following lemma.

Lemma 54.1. *Let K/F be a finite Galois extension of fields, $H_i \subset G(K/F)$ a subgroup and $E_i = K^{H_i}$ for $i = 1, 2$. If σ is an element of the Galois group $G(K/F)$, then*

$$\sigma|_{E_1} : E_1 \to E_2 \text{ is an isomorphism if and only if } H_2 = \sigma H_1 \sigma^{-1},$$

i.e., E_1 and E_2 are conjugate over F (via σ) if and only if H_1 and H_2 are conjugate in $G(K/F)$ (via σ).

PROOF. We know that K/E_i is Galois and $H_i = G(K/E_i)$ for $i = 1, 2$. In particular, $[K : E_i] = |H_i|$ for $i = 1, 2$, so under either condition, we have

$$|H_1| = |H_2| = [K : E_1] = [K : E_2].$$

($\Rightarrow$): Let $\tau \in H_1$ and $y \in E_2$, then $\sigma^{-1}(y)$ lies in E_1, so $\sigma\tau\sigma^{-1}(y) = \sigma\sigma^{-1}(y) = y$ as $\tau|_{E_1} = 1_{E_1}$. Therefore, we have $\sigma H_1 \sigma^{-1} \subset H_2 = G(K/E_2)$, hence $\sigma H_1 \sigma^{-1} = H_2$ as $|\sigma H_1 \sigma^{-1}| = |H_1| = |H_2| < \infty$.

($\Leftarrow$): Let $x \in E_1$ and $\tau \in H_1$, then $\sigma\tau\sigma^{-1}(\sigma(x) = \sigma\tau(x) = \sigma(x)$, as $\tau|_{E_1} = 1|_{E_1}$. Consequently, $\sigma(x)$ lies in $K^{\sigma H_1 \sigma^{-1}} = K^{H_2} = E_2$, and $\sigma|_{E_1} : E_1 \to E_2$ and is a homomorphism. As $[K : E_1] = [K : E_2]$ and $[E_1 : F] = [E_2 : F]$, the map σ_{E_1} is an isomorphism (since a linear monomorphism of finite dimensional vector spaces of the same dimension). □

Notation 54.2. Let K/F be an algebraic extension of fields. We set

$$\mathcal{F}(K/F) := \{E \mid K/E/F \text{ is an intermediate field}\}$$
$$\mathcal{G}(K/F) := \{H \mid H \subset G(K/F) \text{ is a subgroup}\}.$$

Putting together all our results leads to:

Theorem 54.3. (The Fundamental Theorem of Galois Theory) *Suppose that K/F is a finite Galois extension. Then*

$$i : \mathcal{G}(K/F) \to \mathcal{F}(K/F) \text{ given by } H \mapsto K^H$$

is a bijection of sets with inverse

$$j : \mathcal{F}(K/F) \to \mathcal{G}(K/F) \text{ given by } E \mapsto G(K/E).$$

Moreover,

1. *The bijection i is order-reversing, i.e., $H_1 \subset H_2$ if and only if $K^{H_1} \supset K^{H_2}$.*
2. *If $E \in \mathcal{F}(K/F)$, then E/F is normal (hence Galois) if and only if $G(K/E) \triangleleft G(K/F)$.*

(3) *If $E \in \mathcal{F}(K/F)$ with E/F normal, then the canonical epimorphism*

$$^- : G(K/F) \to G(E/F) \text{ given by } \sigma \mapsto \sigma|_E$$

induces an isomorphism

$$G(K/F)/G(K/E) \to G(E/F) \text{ given by } \sigma G(K/E) \mapsto \sigma|_E.$$

(4) *If $H \in \mathcal{G}(K/F)$, then $|H| = [K : K^H]$ and K/K^H is Galois with $H = G(K/K^H)$.*

(5) *If $H \in \mathcal{G}(K/F)$, then $[G(K/F) : H] = [K^H : F]$.*

We have the following picture:

$$\begin{array}{ccc} K & \text{——} & G(K/K) = 1 \\ | & & | \\ | = \cup \quad E & \text{———} & G(K/E) \quad | = \cap \\ | & & | \\ F & \text{———} & G(K/F) \end{array}$$

with the identified verticals having the same degree (index) and if E/F is normal, $G(K/F)/G(K/E) \cong G(E/F)$.

Proof. We already know that $H_1 = H_2$ in $\mathcal{G}(K/F)$ if and only if $K^{H_1} = K^{H_2}$ in $\mathcal{F}(K/F)$, so $i : \mathcal{G}(K/F) \to \mathcal{F}(K/F)$ is injective. Let $E \in \mathcal{F}(K/F)$, then K/E is Galois, so $G(K/E) \mapsto K^{G(K/E)} = E$, hence $i : \mathcal{G}(K/F) \to \mathcal{F}(K/F)$ is bijective.

(1) is clear.

(2): We know that if $E \in \mathcal{F}(K/F)$, then E/F is normal if and only if $\sigma(E) = E$ for all $\sigma \in G(K/F)$. By Lemma 54.1, this is true if and only if $\sigma G(K/E)\sigma^{-1} = G(K/E)$ for all $\sigma \in G(K/F)$, i.e., if and only if $G(K/E) \triangleleft G(K/F)$.

We have seen if E/F is normal then $\Phi : G(K/F) \to G(E/F)$ by $\sigma \mapsto \sigma|_E$ is a well-defined group epimorphism. As

$$\ker \Phi = \{\sigma \in G(K/F) \mid \sigma|_E = 1_E\} = G(K/E),$$

Φ induces an isomorphism $G(K/F)/G(K/E) \cong G(E/F)$.

(4): As K/K^H is Galois, $H = G(K/K^H)$, so $H = [K : K^H]$ by Artin's Theorem 51.15.

(5): By Artin's Theorem 51.15,

$$[K : K^H][K^H : F] = [K : F] = |G(K/F)| = [G(K/F) : H]|H|,$$

so (4) $\Rightarrow$ (5). $\square$

Corollary 54.4. *Let K/F be a finite Galois extension of fields, then $\mathcal{F}(K/F)$ is finite, i.e., there exist only finitely many intermediate fields E with $K/E/F$.*

PROOF. By the Fundamental Theorem of Galois Theory, we have $|\mathcal{F}(K/F)| = |\mathcal{G}(K/F)|$, hence $\mathcal{F}(K/F)$ is a finite set. □

Corollary 54.5. *Let K/F be a finite separable extension. Then $\mathcal{F}(K/F)$ is finite.*

PROOF. Let L/K be a normal closure of K/F. Then L/F is finite, separable, and normal, so Galois. Thus $|\mathcal{F}(K/F)| \leq |\mathcal{F}(L/F)|$ is finite. □

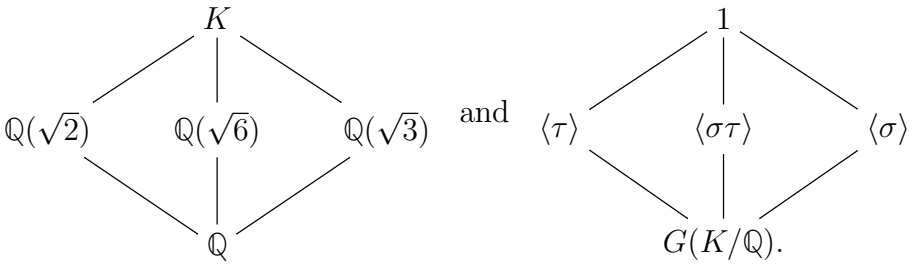
We give a few examples, before deriving significant consequences of this fundamental theorem.

Example 54.6. 1. We know that $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$ is the splitting field of $(t^2-2)(t^2-3)$ in $\mathbb{C}$ over $\mathbb{Q}$. Then we have seen that $G(K/\mathbb{Q}) = \{1, \sigma, \tau, \sigma\tau\} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ with

$\sigma : K \to K$ the $\mathbb{Q}(\sqrt{3})$-automorphism given by $\sqrt{2} \mapsto -\sqrt{2}$,

$\tau : K \to K$ the $\mathbb{Q}(\sqrt{2})$-automorphism given by $\sqrt{3} \mapsto -\sqrt{3}$.

Intermediate fields and corresponding subgroups are given by the following diagrams:

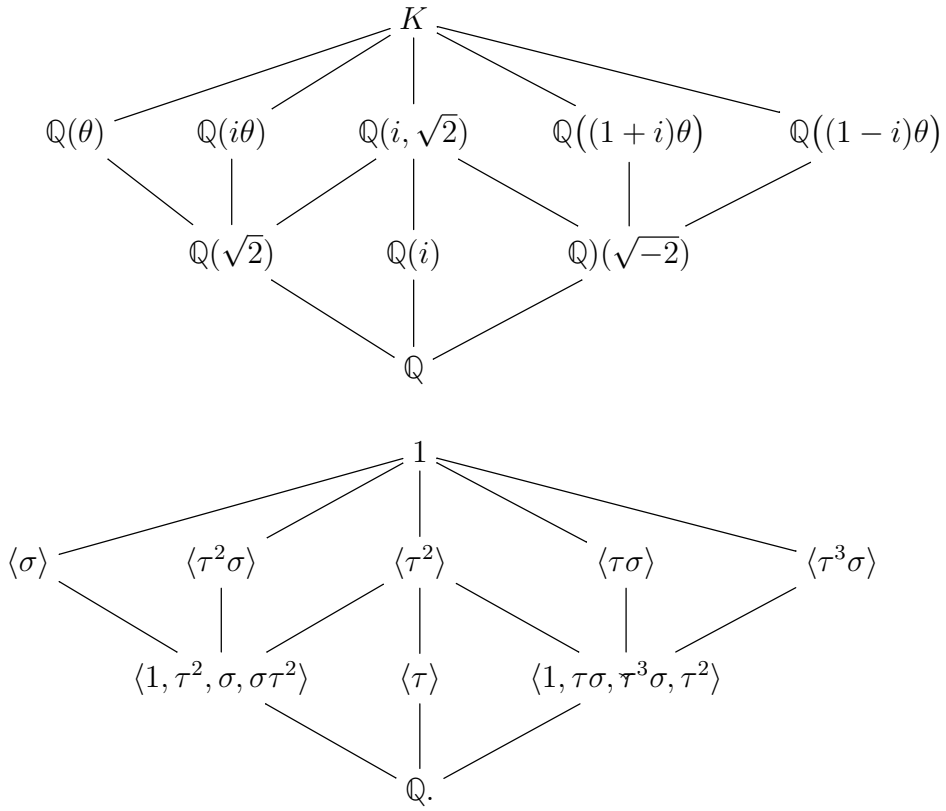


2. Let K be a splitting field of $t^4 - 2$ over $\mathbb{Q}$. Then $K = \mathbb{Q}(\sqrt[4]{2}, \sqrt{-1})$ and $[K : \mathbb{Q}] = 8$. If

$\sigma : K \to K$ the $\mathbb{Q}(\sqrt[4]{2})$-automorphism given by $\sqrt{-1} \mapsto -\sqrt{-1}$,

$\tau : K \to K$ the $\mathbb{Q}(\sqrt{-1})$-automorphism given by $\sqrt[4]{2} \mapsto \sqrt{-1}\sqrt[4]{2}$,

then $|\langle\sigma\rangle| = 2$, $|\langle\tau\rangle| = 4$, $\sigma\tau\sigma^{-1} = \tau^3$, $G(K/\mathbb{Q}) = \langle\sigma, \tau\rangle \cong D_4$. Intermediate fields and corresponding subgroups are given by the following diagrams with $\theta = \sqrt[4]{2}$, $i = \sqrt{-1}$:



3. Let $L = F(t_1, \ldots, t_n)$, the quotient field of $F[t_1, \ldots, t_n]$. The symmetric group S_n acts on L as F-automorphisms by permuting the t_i, i.e.,

$$\sigma(t_i) = t_{\sigma(i)} \text{ for all permutations } \sigma \in S_n \text{ and } i = 1, \ldots, n.$$

We view $S_n \subset G(L/F)$. Let

$$s_j = s_j(t_1, \ldots, t_n) := \sum_{1 \le i_1 < \cdots < i_j \le n} t_{i_1} \cdots t_{i_j}$$

be the jth *elementary symmetric function* of $t_1, \ldots, t_n$ for $j = 1, \ldots, n$ and set $s_0 = 1$. We know that $t_1, \ldots, t_n$ are the roots of the polynomial

$$f := t^n - s_1 t^{n-1} + \cdots + (-1)^n s_n \text{ in } F[s_1, \ldots, s_n][t],$$

as $f = (t - t_1) \cdots (t - t_n)$ in $L[t]$. Let $K = L^{S_n}$, so L/K is Galois. Let $E := F(s_1, \ldots, s_n) \subset K$. Clearly, L is a splitting field of f over E, and by our theory, we know that

$$[L : E] \le (\deg f)! = n! \text{ and } [L : K] = |G(L/K)| = [S_n| = n!.$$

So $K = E = F(s_1, \ldots, s_n)$ and $G(L/K) = G\big(L/F(s_1, \ldots, s_n)\big) \cong S_n$. Note that f must be irreducible in $E[t]$, lest $f = f_1 f_2$ in $E[t]$ with $0 < \deg f_1 < n$ for $i = 1, 2$, and we would have $[L : E] \leq (\deg f_1)!(\deg f_2)! < (\deg f)! = [L : E]$. Note further that $s_1, \ldots, s_n$ are algebraically independent over F as $t_1, \ldots, t_n$ are and K/E is finite. (Cf. Proposition 46.4.)

As an application of this computation, we also get a weak form of the Fundamental Theorem of Symmetric Functions. Let F be a field and S_n act on $F[t_1, \ldots, t_n]$ by restriction of the above action. Then we have with s_j the jth elementary symmetric function in $t_1, \ldots, t_n$:

Corollary 54.7. *Let F be a field and f be a symmetric polynomial in $F[t_1, \ldots, t_n]$, then f is a rational function in the elementary symmetric functions in $t_1, \ldots, t_n$, i.e., $f \in F(s_1, \ldots, s_n)$.*

The Fundamental Theorem of Symmetric Functions implies that a symmetric polynomial in $F[t_1, \ldots, t_n]$ actually lies in $F[s_1, \ldots, s_n]$, even with F replaced by a commutative ring. (cf. Theorem 66.4 for the proof.)

We continue with the Examples 54.6

4. Every finite group G is isomorphic to some Galois group, i.e., there exists a Galois field extension L/K with $G \cong G(L/K)$:

   Let G be a finite group of order n and F be an arbitrary field. Set $L = F(t_1, \ldots, t_n)$. If $E = F(s_1, \ldots, s_n)$ as in previous example, then L/E is Galois with $G(L/E) \cong S_n$. By Cayley's Theorem 12.5, there exists a group monomorphism $\lambda : G \to S_n$, so $G \cong \lambda(G) \subset S_n$. Let $K = L^{\lambda(G)}$, then L/K is Galois with $G(L/K) \cong G$.

Open Problem 54.8. (Inverse Galois Problem) Let H be a finite group. Does there exist a Galois extension L of $\mathbb{Q}$ with $G(L/\mathbb{Q}) \cong H$?

The answer to this question is known for some types of groups, e.g., cyclic groups, abelian groups, solvable groups, S_n, A_n.

We now turn to some significant consequences of the Fundamental Theorem of Galois Theory.

Theorem 54.9. (Primitive Element Theorem) *Let K/F be a finite extension of fields. If $\mathcal{F}(K/F)$ is finite, then there exists an element α in K such that $K = F(\alpha)$. In particular, this is true if K/F is separable.*

Proof. **Case 1**. F is finite:

K must also be a finite field, so $K^\times = \langle \alpha \rangle$ for some $\alpha \in K$. Therefore, $K = F(\alpha)$ and $\mathcal{F}(K/F)$ is finite.

Case 2. F is infinite:
Choose $\alpha \in K$ with $[F(\alpha) : F]$ maximal, so $[F(\alpha) : F] \leq [K : F]$. Suppose that $F(\alpha) < K$, then there exists a $\beta \in K \setminus F(\alpha)$. For each $a \in F$ consider the subfield $F(\alpha + a\beta) < F(\alpha, \beta) \subset K$. As $\mathcal{F}(F(\alpha, \beta)/F)$ is a subset of the finite set $\mathcal{F}(K/F)$ and F is infinite, there exists elements c and d in F with $c \neq d$ satisfying $F(\alpha + c\beta) = F(\alpha + d\beta)$. Since $c - d$ lies in $F^\times$ and $(c-d)\beta$ lies in $F(\alpha + c\beta)$, we must have $\beta \in F(\alpha + c\beta)$ and hence that α also lies in $F(\alpha + c\beta)$. It follows that $F(\alpha, \beta) \subset F(\alpha + c\beta)$, contradicting the maximality of $F(\alpha)$. □

By Exercise 50.10(9b), we know this is not always true if K/F is not separable.

Remark 54.10. Let F be an infinite field and K/F a finite extension with $\mathcal{F}(K/F)$ finite. Suppose that α and β lie in K and S is an infinite subset of F. Then the proof above shows that there exists an element c in S such that $F(\alpha, \beta) = F(\alpha + c\beta)$. In particular, if $K = F(\alpha_1, \ldots, \alpha_n)$ and S is an infinite subset of F, then there exist infinitely many n-tuples $c_1, \ldots, c_n$ in S^n satisfying $K = F(c_1\alpha_1 + \cdots + c_n\alpha_n)$.

We finally prove the loose end in the constructibility of regular n-gons by straight-edge and compass (cf. Theorem 49.8).

Theorem 54.11. (Square Tower Theorem) *Let F be a field of characteristic different from two and K/F a finite normal field extension. Then K/F is a square root tower if and only if $[K : F] = 2^e$ for some positive integer e.*

Proof. ($\Rightarrow$) has previously been established.
($\Leftarrow$): Suppose that $[K : F] = 2^e$, for some integer e. Let $K = F(\alpha_1, \ldots, \alpha_n)$. We know that $\deg m_F(\alpha_i) = [F(\alpha_i : F] \mid [K : F]$ for each i. As $\operatorname{char} F \neq 2$, each $m_F(\alpha_i)$ is separable, hence the extension K/F is separable. Consequently, K/F is normal and separable, hence Galois, so $|G(K/F)| = [K : F] = 2^e$. We know that there exists a subgroup H of the 2-group $G(K/F)$ of index two, therefore, H is a normal subgroup. It follows that K^H/F is normal of degree two. By Remark 49.7(6), K^H/F is a square root tower, as $\operatorname{char} F$ is not two. The field extension K/K^H is Galois of degree 2^{e-1}, so a square root tower by induction on e. Therefore, K/F is a square root tower by Remark 49.7(4). □

Next we give a proof of the Fundamental Theorem of Algebra that uses a minimal amount of analysis. Indeed the only analysis that we need is the Intermediate Value Theorem from Calculus (which is really the completeness of the real line).

Theorem 54.12. (Fundamental Theorem of Algebra) *The field of complex numbers is algebraically closed.*

PROOF. **Claim**. $\mathbb{C} = \mathbb{C}^2 := \{x^2 \mid x \in \mathbb{C}\}$:

Let a be a positive real number. The polynomial $t^2 - a$ in $\mathbb{R}[t]$ has a real root by the Intermediate Value Theorem, since $f(1+a) > 0$ and $f(a) < 0$. Therefore, every positive real number is a square in $\mathbb{R}$ hence in $\mathbb{C}$. Since $-a = (\sqrt{-1})^2 a$, it follows that every real number is a square in $\mathbb{C}$. Let $\alpha = a + b\sqrt{-1}$ with a and b real numbers. We must show α is a square in $\mathbb{C}$. We may assume that $b \neq 0$. Let $x = (a + \sqrt{a^2 + b^2})/2$ in $\mathbb{R}$. Then x is positive (why?), so there exists a real number c such that $x = c^2$. Similarly, $y = (-a + \sqrt{a^2 + b^2})/2$ in $\mathbb{R}$ is positive, so there exists a real number d such that $y = d^2$. Then $\alpha = (c + d\sqrt{-1})^2$, establishing the claim.

[The above arises from $\sqrt{\alpha} = \sqrt{|\alpha|} e^{\sqrt{-1}\,\theta/2}$ with θ the angle satisfying $\tan\theta = b/a$ and the half angle formula in trigonometry – which we do not need!]

Now let f be a non-constant polynomial in $\mathbb{C}[t]$. We must show that f splits over $\mathbb{C}$. Let $K/\mathbb{C}$ be a splitting field of f and L/K a normal closure of $K/\mathbb{R}$. If we show that $L = \mathbb{C}$, we are done. We know that $L/\mathbb{R}$ is Galois and $2 \mid [\mathbb{C} : \mathbb{R}]$. Therefore, there exists a Sylow 2-subgroup H of $G(K/\mathbb{R})$. In particular, $[L^H : \mathbb{R}] = [G(L/\mathbb{R}) : H]$ is odd. Since $\mathbb{R}$ is a field of characteristic zero, it is perfect, so there exists an element $\alpha \in L^H$ satisfying $L^H = \mathbb{R}(\alpha)$ by the Primitive Element Theorem 54.9. The element α in L^H is a root of the irreducible polynomial $m_{\mathbb{R}}(\alpha)$ in $\mathbb{R}[t]$ of odd degree. By the Intermediate Value Theorem, any real polynomial of odd degree has a real root. Therefore, α is real, so $L^H = \mathbb{R}(\alpha) = \mathbb{R}$, Therefore, $G(L/\mathbb{R}) = G(L/L^H)$ is a 2-group. As $L/\mathbb{R}$ is Galois and $\operatorname{char}\mathbb{R} \neq 2$, $L/\mathbb{R}$ is a square root tower by the Square Root Tower Theorem. It follows that $L(\mathbb{C})/\mathbb{C}$ is also a square root tower by Remark 49.7(2). By the claim, $\mathbb{C}$ has no proper square root towers over it, so $L(\mathbb{C}) = \mathbb{C}$ as needed. □

Exercises 54.13. 1. Show that the lattices of subgroups and subfields in Example 54.6(2) are correct.

2. Let F be a perfect field and K/F a finite Galois extension. Show if $G(K/F) \cong S_n$, then K is the splitting field of an irreducible polynomial of degree n. Is this still true if $G(K/F) \cong A_n$?

3. Let K be a splitting field of $f \in \mathbb{Q}[t]$. Find K, $G(K/\mathbb{Q})$, and all intermediate fields if:
   (*i*) $f = t^4 - t^2 - 6$.
   (*ii*) $f = t^3 - 3$.

4. Let K be a splitting field of $t^5 - 2 \in \mathbb{Q}[t]$.
   (*i*) Find $G(K/\mathbb{Q})$.
   (*ii*) Show that there exists a group monomorphism $G(K/\mathbb{Q}) \to S_5$.
   (*iii*) Find all subgroups of $G(K/\mathbb{Q})$ and the corresponding fields.
5. Suppose that L/F is a finite Galois extension and $L/K/F$ an intermediate field. Show that $G(L/F) = N_{G(L/F)}\big(G(L/K)\big)/G(L/K)$, where $N_{G(L/F)}\big(G(L/K)\big)$ is the normalizer of $G(L/K)$ in $G(L/F)$.
6. Suppose that K/F is a finite Galois extension of fields. Let $F \subset E \subset K$ and L the smallest subfield of K containing E such that L/F is normal. Show $G(K/L) = \bigcap_{\sigma \in G(K/F)} \sigma G(K/E)\sigma^{-1}$, the core of $G(K/E)$ in $G(K/F)$.
7. Let L/F be a finite Galois extension, $L/K_i/F$ intermediate fields, and $H_i = G(K_i/F)$ for $i = 1, 2$. Show $H_1 \cap H_2 = G\big(L/K_1(K_2)\big)$ and the fixed field of the smallest group in $G(L/F)$ containing H_1 and H_2 is $K_1(K_2)$
8. Let K/F be a finite Galois extension. Suppose that $p^r \mid [K : F]$ but $p^{r+1} \nmid [K : F]$. Show that there exist fields L_i, $1 \leq i \leq r$, satisfying $F \subseteq L_r < L_{r-1} < \cdots < L_1 < L_0 = K$ with L_i/L_{i+1} is normal, $[L_i : L_{i+1}] = p$ and $p \nmid [L_r : F]$ for each i.
9. Let f be an irreducible quartic over a field F of characteristic zero, G the Galois group of f (i.e., the Galois group of K/F with K a splitting field of f over F), u a root of f. Show that there is no field properly between F and $F(u)$ if and only if $G = A_4$ or $G = S_4$ and that there exist such irreducible polynomials with Galois group A_4 and S_4. (This will explode the myth that there must be an intermediate field when the dimension is not prime.)
10. Let L/F and K/F be finite extensions of fields with L/F Galois and K, L lying in some extension field F. Let $LK = K(L) = L(K)$. Show that LK/K and $L/L \cap K$ are Galois and the restriction map $\varphi : G(LK/K) \to G(L/F)$ given by $\sigma \mapsto \sigma|_L$ a monomorphism with image $G(L/L \cap K)$. In particular, if $L \cap K = F$, then $G(LK/K) \cong G(L/F)$.
11. Suppose that M/F be a finite Galois extension and K and L two intermediate fields of M/F with K/F Galois. Let $KL = L(K)$, $N = G(M/K) \lhd G(M/F)$, and $H = G(M/L)$. Then KL/L is Galois by Exercise 10. Show, in addition, that all of the following are true:
    (a) $G(M/KL) = H \cap N \lhd H$.
    (b) $G(KL/L) \cong H/H \cap N \cong HN/N$.

(c) $G(K/K \cap L) \cong HN/N$.

12. Let E/F be an extension of fields and $E/K_i/F$ be intermediate fields, $i = 1, 2$ with K_i/F finite Galois for $i = 1, 2$. Let $K_1K_2 = K_1(K_2)$. Show K_1K_2/F is a Galois extension and the restriction map $\varphi : G(K_1K_2/F) \to G(K_1/F) \times G(K_2/F)$ given by $\sigma \mapsto (\sigma|_{K_1}, \sigma|_{K_1})$ is a monomorphism. In particular, it is an isomorphism if $K_1 \cap K_2 = F$.
13. Let L/F be a finite Galois extension. Suppose that $G(L/F) = G_1 \times \cdots \times G_n$, a direct product of subgroups. Let K_i be the fixed field of $G_1 \times \cdots \times \{1\} \times \cdots \times G_n$ where the group $\{1\}$ occurs in the ith place. Show that K_i/F is Galois for every i, $K_i \cap F(\cup_{j \neq i} K_j) = F$, and $L = F(\cup_i K_i)$.
14. Prove Remark 54.10.
15. Let F be a field having no nontrivial field extensions of odd degree and K/F a finite field extension. Show if K has no field extensions of degree two, then F is perfect and K is algebraically closed. (Cf. Exercise 50.10(13)).

55. Addendum: Infinite Galois Theory

In this section, we indicate how Galois Theory for finite extensions extends to arbitrary algebraic extensions. In particular, normal extensions need not be finite and an algebraic extension K of F is called *Galois* if $F = K^{G(K/F)}$. We must generalize some of our proofs.

Proposition 55.1. *Let K/F be an algebraic extension. If K/F is normal, and $\sigma : K \to K$ an F-homomorphism, then σ lies in $G(K/F)$, i.e., σ is onto.*

PROOF. Let α lie in K, then the minimal polynomial $m_F(\alpha)$ splits over K (just as in the finite normal case by Exercise 53.22(2)), so there exists an intermediate field $K/E/F$ with E a splitting field of $m_F(\alpha)$ over F. Thus E/F is a finite normal extension, so $\sigma(E) = E$, i.e., $\sigma|_E \in G(E/F)$. Consequently, there exists a β in E, hence K, such that $\sigma(\beta) = \alpha$. □

Proposition 55.2. *Let K/F be an algebraic extension. If K/F is normal and $K/E/F$ is an intermediate field, then any F-homomorphism $\sigma : E \to K$ lifts to an element of $G(K/F)$.*

PROOF. This is Exercise 53.22(3), which follows by a Zorn Lemma argument and the finite extension case. □

Theorem 55.3. *Let K/F be an algebraic extension. If K/F is normal, then K/F is Galois if and only if K/F is separable and normal.*

PROOF. ($\Leftarrow$): Let α lie in $K^{G(K/F)}$. As above there exists an intermediate field $K/E/F$ with E a splitting field of $m_F(\alpha)$ over F. As K/F is separable, so is E/F, hence E/F is finite normal and separable so Galois. Let σ lie in $G(E/F)$, By the last proposition, there exists an extension $\widehat{\sigma}$ of σ in $G(K/F)$, i.e., $\widehat{\sigma}|_E = \sigma$, so by assumption $\sigma(\alpha) = \widehat{\sigma}(\alpha) = \alpha$. Therefore, α lies in $E^{G(E/F)} = F$.

($\Rightarrow$): Let α lie in F. Every element of $S = \{\sigma(\alpha) \mid \sigma \in G(K/F)\}$ is a root of $m_F(\alpha)$ (which a priori may have other and/or multiple roots), so S is a finite set. Let $\alpha = \alpha_1, \alpha_2, \ldots, \alpha_n$ be the finite distinct elements of S and $f = \prod_{1=1}^{n}(t - \alpha_i)$. As $G(K/F)$ permutes $\alpha_1, \ldots, \alpha_n$, we must have $f = \widetilde{\sigma} f$ for every $\sigma \in G(K/F)$, i.e., f lies in $F[t]$, But $f \mid m_f(\alpha)$ in $F[t]$, so $f = m_F(\alpha)$ is separable and $F = m_F(\alpha)$ splits over K. It follows that K/F is separable and normal. □

Remark 55.4. Let K/F be an algebraic Galois extension of fields, H a subgroup of $G(K/F)$, and $E = K^H$. Suppose that σ and τ are elements of $G(K/F)$. If $\tau^{-1}\sigma$ lies in H, then for every $x \in E$, we have $\tau^{-1}\sigma(x) = x$, so $\sigma|_E = \tau|_E$. However, if we only know that $(\tau^{-1}\sigma)|_E = 1_E$, we can only conclude that $\tau^{-1}\sigma$ lies in $G(K/F)$. It does not necessarily follow that $\tau^{-1}\sigma$ lies in H. Therefore, we only know that $H \subset G(K/E) = G(K/K^H)$. In the case that K/F is a finite extension, we get equality by counting using Artin's Theorem, but this does not work in the arbitrary algebraic case.

We do, however, have the following:

Lemma 55.5. *Let K/F be an algebraic and Galois extension of fields and $K/E/F$ an intermediate field. If $[E : F]$ finite, then the index $[G(K/F) : G(K/E)]$ is finite and $[G(K/F) : G(K/E)] = [E : F]$.*

PROOF. Suppose that $\sigma_1, \ldots, \sigma_n : E \to K$ are all the distinct F-homomorphisms. These lift to $\widehat{\sigma}_1, \ldots, \widehat{\sigma}_n$ in $G(K/F)$, i.e., $\widehat{\sigma}_i|_E = \sigma_i$ for each i, by Proposition 55.2 and, as E/F is finite separable, we have $n = [E : F]$ by Proposition 53.11. Suppose that $\widehat{\sigma}$ is an element of $G(K/F)$. Then there exists an i such that $\widehat{\sigma}|_E = \sigma_i$. Therefore, $(\widehat{\sigma}_i)^{-1}\widehat{\sigma}$ lies in $G(K/E)$. It follows that

$$G(K/F) = \bigvee_{i=1}^{n} \widehat{\sigma}_i G(K/E), \tag{55.6}$$

and the result follows. □

We now state without proof the major new result needed. The ideas involve topological knowledge that we have not assumed. We summarize some of the concepts that we need. A *topology* $\mathcal{T}$ on a space

X is a collection of sets, called *open sets* closed under arbitrary unions and finite intersections. X is then called a *topological space* (via $\mathcal{T}$). Let X have a topology $\mathcal{T}$. A *closed set* in X is one that is the complement of an open set in $\mathcal{T}$. If x is a point in X, an open set containing x is called a *neighborhood of* x (e.g., an open interval containing x if x lies in $X = \mathbb{R}$). A subcollection of $\mathcal{T}$ consisting of neighborhoods of x such that every neighborhood of x contains an element of this subcollection is called a *fundamental system of neighborhoods* of x (e.g., all the open intervals with rational end points containing x if $X = \mathbb{R}$). A map $f : X \to Y$ of topological spaces is called *continuous* if the preimage of an open set in Y under f is open in X. Then the following is true:

Theorem 55.7. (Krull) *Suppose that K/F is an algebraic Galois extension of fields. Let*

$$\mathcal{N}(K/F) := \{G(K/E) \mid K/E/F \text{ with } E/F \text{ finite Galois}\}.$$

Then there exists a topology on G compatible with the group structure of G (i.e., the group operation on G is continuous relative to this topology) *and which has $\mathcal{N}(K/F)$ as a fundamental system of neighborhoods of the identity 1_K in $G(K/F)$. With this topology, $G(K/F)$ is a Hausdorff, totally disconnected, topological group.*

Let X be a topological space. It is called *Hausdorff* if given any two distinct points in X, there exist disjoint neighborhoods of these two points and *compact* if every collection of open sets that *cover* X, i.e., whose union is X, contains a finite subcollection that covers X. Finally, X is called *totally disconnected* if there exists a subcollection of the topology consisting of sets that are both open and closed and such that every open set can be obtained from this subcollection by taking arbitrary unions of finite intersections from it. The topology in the theorem is called a *profinite topology*.

In the terminology of the theorem, if σ is an element of $G(K/F)$, then by translation a fundamental system of neighborhoods for σ is just $\{\sigma H \mid H \in \mathcal{N}(K/F)\}$ By equation 55.6, if $H \in \mathcal{N}(K/F)$, then $[G(K/F) : H]$ is finite and every element of $\mathcal{N}(K/F)$ is both open and closed (look at complements). This is the totally disconnected statement. If H is a subgroup of $G(K/F)$, let $\overline{H}$ denote the *closure* of H in $G(K/F)$ in this topology, i.e., the smallest closed set in $G(K/F)$ containing H. This means that if σ in $G(K/F)$ lies in $\overline{H}$, then every open neighborhood of σ intersects H non-trivially. Because of the topology on $G(K/F)$, this means that $H \cap \sigma H'$ is nonempty for every H' in $\mathcal{N}(K/F)$.

Lemma 55.8. *Let K/F be an algebraic Galois extension of fields, H a subgroup of $G(K/F)$, and $E = K^H$. Then $G(K/E) = \overline{H}$, the closure of H in $G(K/F)$.*

PROOF. $G(K/E) \subset \overline{H}$: Let σ lie in $G(K/E)$. We must show that $\sigma \in \overline{H}$. As remarked above, since $\sigma\mathcal{N}(K/F)$ is a fundamental system of neighborhoods for σ, it suffices to show that $H \cap \sigma H'$ is not empty for all H' in $\mathcal{N}(K/F)$. Suppose that $H' \in \mathcal{N}(K/F)$. As $K^{H'}/F$ is finite Galois, it is finite separable, so $K^{H'} = F(\alpha)$ for some α in K by the Primitive Element Theorem 54.9. Let $K/L/E$ be an intermediate field such that L/E is finite normal with $\alpha \in L$. Since L/E is finite, normal, and separable, it is finite Galois. As $H \subset G(K/E)$ and L/E is Galois, $\tau|_E$ lies in $G(K/E)$ for all $\tau \in H$, i.e, we have a group homomorphism $\Psi : H \to G(L/E)$ given by $\tau \mapsto \tau|_L$ and it has $G(L/L^{\operatorname{im}\Psi})$ as (its finite) image. By the definition of H, we have $E = L^{\operatorname{im}\Psi}$, so Ψ is onto. (Note, a priori, we only know that $E \subset K^{G(K/E)}$.) Thus there exists a τ in H with $\tau|_L = \sigma|_L$. In particular, as α lies in L, we have $\sigma(\alpha) = \tau(\alpha)$, so $\sigma^{-1}\tau$ lies in H'. This shows that $\tau \in \sigma H' \cap H$.
$\overline{H} \subset G(K/E)$: Let σ lie in $\overline{H}$. We must show that σ lies in $G(K/E)$, i.e., if α is an element of E, then $\sigma(\alpha) = \alpha$. Let $K/L/F(\alpha)$ with L/F finite, normal, and separable, hence finite Galois. Set $H' = G(K/L)$. By assumption, $H \cap \sigma H'$ is not empty, so there exists an element τ in $H \cap \sigma H'$. In particular, $\tau|_E = 1_E$, since $E = K^H$, and $\sigma(\alpha) = \tau(\alpha)$, since $F(\alpha) \subset K^{H'}$. Therefore, $\sigma(\alpha) = \tau(\alpha) = \alpha$. □

The Fundamental Theorem of Galois Theory in this more general setting becomes:

Theorem 55.9. *Let K/F be an algebraic Galois extension of fields and*

$$\mathcal{G}_c(K/F) := \{H \mid H \text{ is a closed subgroup of } G(K/F)\},$$

then

$$i : \mathcal{G}_c(K/F) \to \mathcal{F}(K/F) \text{ given by } H \mapsto K^H$$

is an order reversing bijection.

PROOF. By the lemma, we know that

$$\mathcal{G}_c(K/F) = \{G(K/E) \mid E \in \mathcal{F}(K/F)\}.$$

i **is injective**: If $K^{H_1} = K^{H_2}$, then by the lemma,

$$H_1 = \overline{H_1} = G(K/K^{H_1}) = G(K/^{H_2}) = \overline{H_2} = H_2.$$

i **is surjective**: Let E lie in $\mathcal{F}(K/F)$, so $E \subset K^{G(K/E)}$. We must show that $G(K/E)$ moves $K \setminus E$. Let $\alpha \in K \setminus E$. As before, let

$K/L/E$ with L/E finite normal, hence Galois, and $\alpha \in L$. As α does not lie in E, there exist an element σ in $G(L/E)$ satisfying $\sigma(\alpha) \neq \alpha$. By Proposition 55.2, there exists a lift $\widehat{\sigma}$ in $G(K/F)$ of σ. So $\widehat{\sigma}(\alpha) = \sigma(\alpha) \neq \alpha$. □

Remark 55.10. Let K/F be an algebraic Galois extension of fields, $E \in \mathcal{F}(K/F)$. Then

(*i*) E/F is normal if and only if $G(K/E) \lhd G(K/F)$.
(*ii*) If E/F is normal, then $G(E/F) \cong G(K/F)/G(K/E)$.

We leave the proofs of these as exercises.

Example 55.11. Let $S = \{\sqrt{a} \mid a > 0 \text{ a prime or } a = -1\} \subset \mathbb{C}$ and K the splitting field of $\{t + a^2 \mid a \in S\}$ in $\mathbb{C}$. Then $K/\mathbb{Q}$ is Galois, and it is easy to see that K is the *compositum* of all quadratic extensions of $\mathbb{Q}$ in $\mathbb{C}$, i.e., $K = \mathbb{Q}\big(\bigcup_{d\in\mathbb{Z}} \mathbb{Q}(\sqrt{d})\big)$, the number of which is countable. [More generally, if L/F is a field extension, $L/K_i/F$, $i \in I$, intermediate extensions, then $F(\bigcup_I K_i)$ is called the *compositum* of the K_i, $i \in I$, in L.] If σ and τ lie in $G(K/\mathbb{Q})$, then we have $\sigma^2 = 1_K$ and $\sigma\tau = \tau\sigma$, i.e., $G(K/\mathbb{Q})$ is abelian and all non-identity elements are of order two. It follows that $G(K/\mathbb{Q})$ is a vector space over $\mathbb{Z}/2\mathbb{Z}$, hence has a $\mathbb{Z}/2\mathbb{Z}$-basis, say $\mathcal{B}$.
Let T be a subset of S and define $\sigma : S \to S$ by

$$\sigma(x) = \begin{cases} x & \text{if } x \in T \\ -x & \text{if } x \notin T. \end{cases}$$

Then σ induces an element in $G(K/\mathbb{Q})$. In particular, $G(K/\mathbb{Q})$ is uncountable, so cannot have a countable $\mathbb{Z}/2\mathbb{Z}$-basis. Therefore, for all but countably many $x \in \mathcal{B}$, the subgroup $H_x := \langle \mathcal{B} \cup \{x\}\rangle$ is a normal subgroup of index two in $G(K/\mathbb{Q})$ that does cannot correspond to a quadratic extension of $\mathbb{Q}$, i.e., only countable many H_x are closed (or open) in $G(K/\mathbb{Q})$.

Exercises 55.12. 1. Let K/F be Galois and

$$\mathcal{N} := \{G(K/E) \mid K/E/F \text{ with } E/F \text{ finite Galois}\}.$$

Show that $\bigcap_{\mathcal{N}} G(K/E) = 1$.
2. Assuming Theorem 55.7, show that the topology is Hausdorff.
3. Show the two items in Remark 55.10 are true.

56. Roots of Unity

An *arithmetic function* $f : \mathbb{Z}^+ \to \mathbb{C}$ is called a *multiplicative function* if $f(1)$ is nonzero and $f(mn) = f(m)f(n)$, whenever m and n

are relatively prime. For example, the Euler phi-function is multiplicative (by the Chinese Remainder Theorem). Note if f is a multiplicative function, then $f(1) = 1$. Let $I : \mathbb{Z}^+ \to \mathbb{C}$ be the multiplicative function defined by $I(1) = 1$ and $I(n) = 0$ for all $n > 1$. The set $\mathcal{A} := \{f \mid f : \mathbb{Z}^+ \to \mathbb{C} \text{ with } f(1) \neq 0\}$ is an abelian monoid with unity I under the *Dirichlet product* $\star : \mathcal{A} \times \mathcal{A} \to \mathbb{C}$ defined by

$$(f \star g)(n) := \sum_{d|n} f(d)g(\frac{n}{d})$$

(cf. Exercise 56.21(1)). This group arises from the study of Dirichlet series, i.e., absolutely convergent series of the form $\sum_n \frac{f(n)}{n^s}$, f a multiplicative function and s a complex variable. It can be shown that $\mathcal{M} := \{f \mid f \in \mathcal{A} \text{ multiplicative}\}$ is a subgroup. Basic to this is Möbius inversion that we now investigate.

Definition 56.1. The *Möbius function* $\mu : \mathbb{Z}^+ \to \mathbb{C}$ is defined by $\mu(1) = 1$ and if $n = p_1^{e_1} \cdots p_r^{e_r}$ is a standard factorization of $n > 1$, then

$$\mu(n) := \begin{cases} (-1)^r \text{ if } n \text{ is square-free, i.e., } e_1 = \cdots = e_r = 1 \\ 0 \text{ otherwise} \end{cases}$$

Let $U : \mathbb{Z}^+ \to \mathbb{C}$ be the multiplicative function defined by $U(n) = 1$ for all n.

Lemma 56.2. *The Möbius function μ is multiplicative and satisfies:*

(1) $I = \mu \star U$ i.e.,

$$\sum_{d|n} \mu(d) = \begin{cases} 1 \text{ if } n = 1 \\ 0 \text{ if } n > 1. \end{cases}$$

(2) If $n = p_1^{e_1} \cdots p_r^{e_r}$ is a standard factorization of $n > 1$, then $\sum_{d|n} |\mu(d)| = 2^r$.

Proof. Let m and n be relatively prime positive integers. Clearly, either m or n is not square free if and only if mn is not square free. It follows easily that μ is multiplicative. Let $\epsilon = \mu$ or $|\mu|$. If $n > 1$ is not square-free, then $\epsilon(n) = 0$. In particular, we may assume that

$n = p_1 \cdots p_r$ is a standard factorization of $n > 1$. Then

$$\begin{aligned}\sum_{d|n} \epsilon(d) &= \epsilon(1) + \sum_i \epsilon(p_i) + \sum_{i_1<i_2} \epsilon(p_{i_1}p_{i_2}) + \\ &\cdots + \sum_{i_1<\cdots<i_j} \epsilon(p_{i_1}\cdots p_{i_j}) + \cdots + \epsilon(p_1\cdots p_r) \\ &= 1 + \epsilon(p_1)\binom{r}{1} + \epsilon(p_1p_2)\binom{r}{2} + \cdots + \epsilon(p_1\cdots p_r)\binom{r}{r}.\end{aligned}$$

If $\epsilon = \mu$, this is $1 - \binom{r}{1} + \binom{r}{2} + \cdots + (-1)^r\binom{r}{r} = (1-1)^r = 0$ and if $\epsilon = |\mu|$, this is $1 + \binom{r}{1} + \binom{r}{2} + \cdots + +\binom{r}{r} = (1+1)^r = 2^r$. □

Proposition 56.3. (Möbius Inversion Formula) *Let f and g be arithmetic functions not zero at 1, then*

$$f(n) = \sum_{d|n} g(d) \text{ if and only if } g(n) = \sum_{d|n} f(d)\mu(\frac{n}{d}).$$

Proof. The proposition is equivalent to $f = g \star U$ if and only if $g = f \star \mu$. By Exercise 8.5(15), the Dirichlet product is associative with I a unity, so $f = g \star U$ if and only if $f \star \mu = (g \star U) \star \mu = g \star (U \star \mu) = g \star I = g$. □

For each n, let ζ_n be a fixed primitive nth root of unity in $\mathbb{C}$. Define the nth *cyclotomic polynomial* $\Phi_n(t)$ in $\mathbb{C}[t]$ by $\Phi_n = \prod_{\langle\zeta_n\rangle}(t - \zeta)$ in $\mathbb{C}[t]$, a polynomial of degree $\varphi(n)$. Clearly,

$$t^n - 1 = \prod_{d|n} \Phi_d,$$

so $\log(t^n - 1) = \sum_{d|n} \log \Phi_d$. By the Möbius Inversion Formula (viewing the cyclotomic polynomials as polynomial functions on the integers or use Exercise 56.21(6)), we see that

$$\Phi_n = \prod_{d|n} (t^d - 1)^{\mu(\frac{n}{d})} = \prod_{d|n} (t^{\frac{n}{d}} - 1)^{\mu(d)} \text{ lies in } \mathbb{C}[t] \cap \mathbb{Q}(t) = \mathbb{Q}[t]$$

and is monic. By induction Φ_d lies in $\mathbb{Z}[t]$ for $d < n$, hence so does $\Phi(n)$ by Corollary 32.6 (the corollary to Gauss's Lemma).

Example 56.4. Using the formula above, we have

$$\begin{aligned}\Phi_{12} &= \prod_{d|12}(t^{\frac{12}{d}}-1)^{\mu(d)}\\ &= (t^{12}-1)^{\mu(1)}(t^6-1)^{\mu(2)}(t^4-1)^{\mu(3)}(t^3-1)^{\mu(4)}\\ &\qquad\qquad (t^2-1)^{\mu(6)}(t-1)^{\mu(12)}\\ &= \frac{(t^{12}-1)(t^2-1)}{(t^6-1)(t^4-1)} = t^4-t^2+1.\end{aligned}$$

Definition 56.5. Let K/F be a finite Galois extension of fields. We say that K/F is an *abelian* (respectively, *cyclic*) *extension* if $G(K/F)$ is abelian (respectively, cyclic).

Theorem 56.6. *Let ζ be a primitive n root of unity in $\mathbb{C}$. Then $\mathbb{Q}(\zeta)/\mathbb{Q}$ is an abelian extension of degree $\varphi(n)$.*

Proof. We must show that Φ_n is irreducible in $\mathbb{Q}[t]$, equivalently in $\mathbb{Z}[t]$. Let ζ be a primitive nth root of unity. As $\mathbb{Z}[t]$ is a UFD, we see that there exist polynomials f and g in $\mathbb{Z}[t]$ satisfying $\Phi_n = fg$ in $\mathbb{Z}[t]$ with f irreducible in $\mathbb{Z}[t]$ and having ζ as a root. As Φ_n is monic, we may assume that both f and g are monic. (Note that this implies that $f = m_{\mathbb{Q}}(\zeta)$.) Let p be a prime such that $p \nmid n$.

Claim: ζ^p is a root of f.

If we prove the claim, it would follow that ζ^m, with m relatively prime to n, is also root of f, hence $f = \Phi_n$ establishing the theorem.

Suppose the claim is false. We can write $t^n - 1 = fh$ in $\mathbb{Z}[t]$ with $h \in \mathbb{Z}[t]$ monic by Corollary 32.6 (the corollary to Gauss's Lemma). As ζ^p is not a root of f, it must be a root of h. This implies that ζ is a root of the monic polynomial $h_p := h(t^p)$ in $\mathbb{Z}[t]$. It follows that $h_p = fg$ in $\mathbb{Z}[t]$ for some monic g in $\mathbb{Z}[t]$ by Corollary 32.6. Let $^- : \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ be the canonical epimorphism; and, if f lies in $\mathbb{Z}[t]$, let $\overline{f}$ be the image of f in $(\mathbb{Z}/p\mathbb{Z})[t]$. Applying the Frobenius homomorphism shows that

$$\overline{h}^p = \overline{h}(t^p) = \overline{h_p} = \overline{f}\,\overline{g}$$

and has $\overline{\zeta}$ as a root, hence so does $\overline{h}$. Therefore, $t^n - 1 = \overline{f}\,\overline{h}$ has a multiple root over $\mathbb{Z}/p\mathbb{Z}$. As $p \nmid n$, this is impossible. This establishes the claim and the theorem. □

Remark 56.7. Let p be an odd prime and ζ a primitive p^rth root of unity. Then $G\big(\mathbb{Q}(\zeta)/\mathbb{Q}\big) \cong (\mathbb{Z}/p^r\mathbb{Z})^\times$ is cyclic of order $p^{r-1}(p-1)$. In particular, $\mathbb{Q}(\zeta)$ contains a unique quadratic extension of $\mathbb{Q}$ using Proposition 105.4.

Corollary 56.8. *Let ζ_m, ζ_n, and ζ_{mn} be primitive mth, nth, and (mn)th roots of unity over $\mathbb{Q}$ with m and n relatively prime positive integers. Then $\mathbb{Q}(\zeta_{mn}) = \mathbb{Q}(\zeta_m, \zeta_n)$ and $\mathbb{Q}(\zeta_m) \cap \mathbb{Q}(\zeta_n) = \mathbb{Q}$.*

PROOF. As m and n are relatively prime, $\zeta_m\zeta_n$ is a primitive (mn)th root of unity. Since the Euler phi-function is multiplicative, the result follows from the theorem. □

By the results in §105, we know $G(\mathbb{Q}(\zeta_n)/\mathbb{Q})$. Indeed if $n = 2^e p_1^{e_1} \cdots p_r^{e_r}$ is a factorization with $2 < p_1 < \cdots < p_r$ primes, with $e \geq 0$ and $e_i > 0$ for all i, then

$$G(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \cong (\mathbb{Z}/2^e\mathbb{Z})^\times \times (\mathbb{Z}/{p_1}^{e_1}\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/{p_r}^{e_r}\mathbb{Z})^\times$$

by the Chinese Remainder Theorem. We know that $(\mathbb{Z}/{p_i}^{e_i}\mathbb{Z})^\times$ is cyclic of order $\varphi(p_i^{e_i}) = p_i^{e_i-1}(p_i - 1)$ for every i and if n is even $(\mathbb{Z}/2^e\mathbb{Z})^\times$ is cyclic only if $e = 1, 2$, otherwise $(\mathbb{Z}/2^e\mathbb{Z})^\times \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2^{e-2}\mathbb{Z})$ if $e \geq 3$.

We next wish to prove a special case of the following well-known result that uses the same idea as the proof of the above as well as our previous idea in proving the infinitude of primes.

Theorem 56.9. (Dirichlet's Theorem on Primes in an Arithmetic Progression) *Let m and n be relatively prime integers. Then there exist infinitely many primes p satisfying $p \equiv m \mod n$.*

We shall not prove this general theorem. The standard proof uses complex analysis, although there is an elementary proof, i.e., one that uses no complex analysis. The special case that we shall prove is the following:

Proposition 56.10. *Let $n > 1$ be an integer. Then there exist infinitely many primes p satisfying $p \equiv 1 \mod n$.*

To prove this, we first establish the following lemma:

Lemma 56.11. *Let a be a positive integer and p a (positive) prime not dividing n. If $p \mid \Phi_n(a)$ in $\mathbb{Z}$, then $p \equiv 1 \mod n$.*

PROOF. Let $^- : \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ be the canonical epimorphism. We know that $t^n - 1 = \prod_{d|n} \Phi_d$ is a factorization of $t^n - 1$ into irreducible polynomials in $\mathbb{Z}[t]$. In particular, $a^n \equiv 1 \mod p$. Let m be the order of $\overline{a}$ in $\mathbb{Z}/p\mathbb{Z}$. So $m \mid n$. If $m < n$, then

$$t^n - 1 = \prod_{d|n} \Phi_d = \Phi_n \prod_{\substack{d|n \\ d<n}} \Phi_d = \Phi_n(t^m - 1) \prod_{\substack{d|n \\ d<n \\ d \nmid m}} \Phi_d.$$

It follows that $\overline{a}$ is a multiple root of $t^n - 1$ in $\mathbb{Z}/p\mathbb{Z}$. But $p \not| \, n$, so $t^n - 1$ has no multiple roots in $\mathbb{Z}/p\mathbb{Z}$. It follows that $m = n$, i.e., n is the order of $\overline{a}$, so $n \mid p - 1$ by Fermat's Little Theorem. The result follows. □

PROOF. (of the Proposition) Suppose that the result is false, and $p_1, \ldots, p_r$ are all the primes p satisfying $p \equiv 1 \mod n$. As Φ_n is monic, we can choose an integer $N > 1$ satisfying $\Phi_n(M) > 1$ for all integers $M > N$ (by calculus). Let $M = p_1 \cdots p_r nN > N$ and p a prime such that $p \mid \Phi_n(p_1 \cdots p_r nN) > 1$. (Replace $p_1 \cdots p_r$ by 1 if $r = 0$.) Since $t^n - 1 = \prod_{d|n} \Phi_d$ in $\mathbb{Z}[t]$, the constant term of Φ_d is ± 1, so

$$\Phi_n(p_1 \cdots p_r nN) \equiv \begin{cases} \pm 1 & \mod p_i \text{ for } 1 \leq i \leq r \\ \pm 1 & \mod n. \end{cases}$$

In particular, $p \neq p_1, \ldots, p_r$ and $p \not| \, n$. It follows that $p \equiv 1 \mod n$ by the lemma, a contradiction. □

The proof of the full theorem shows that primes are "equally distributed" among the integers m relatively prime to n with $1 \leq m \leq n$.

We next turn to a special case of another important theorem whose proof we shall also omit, viz.,

Theorem 56.12. (Kronecker-Weber Theorem) *Let $K/\mathbb{Q}$ be a abelian extension of fields. Then there exists a root of unity ζ in $\mathbb{C}$ such that K is a subfield of $\mathbb{Q}(\zeta)$*

We shall prove this in the case that K is a quadratic extension of $\mathbb{Q}$. The general theorem shows that abelian extensions of $\mathbb{Q}$ are determined by the unit circle. A similar geometric interpretation is true when $\mathbb{Q}$ is replaced by an imaginary quadratic extension of $\mathbb{Q}$, i.e., by $\mathbb{Q}(\sqrt{-d})$, with $d = 1$ or $d > 1$ a square-free integer, and the circle replaced by an appropriate "elliptic curve". Unfortunately, this does not extend further, although a major triumph in number theory was the determination of abelian extensions L of K when $K/\mathbb{Q}$ is finite. There just is no geometric formulation in this general case.

If a is an integer not divisible by p, recall we defined the *Legendre symbol* in Definition 29.8 to be

$$\left(\frac{a}{p}\right) := \begin{cases} +1 \text{ if } a \mod p \text{ is a square.} \\ -1 \text{ if } a \mod p \text{ is not a square.} \end{cases}$$

It is convenient to define $\left(\frac{a}{p}\right) = 0$ if $p \mid a$. Clearly, the Legendre symbol $\left(\frac{a}{p}\right)$ only depends on $a \mod p$. Let p be any prime and consider the squaring map $f : (\mathbb{Z}/p\mathbb{Z})^\times \to (\mathbb{Z}/p\mathbb{Z})^\times$ given by $x \mapsto x^2$, Then f

is a group homomorphism with $\ker f = \{\pm 1\}$ with image $\big((\mathbb{Z}/p\mathbb{Z})^\times\big)^2$. In particular, if p is odd, then $|\operatorname{im} f| = (p-1)/2$, i.e., half of $(\mathbb{Z}/p\mathbb{Z})^\times$ are squares and half non-squares.

Lemma 56.13. *Let p be an odd prime. Then for all integers a and b not divisible by p, we have the following:*

(1) $\big(\frac{a}{p}\big) = \big(\frac{b}{p}\big)$ *if* $a \equiv b \mod p$.

(2) $\big(\frac{ab}{p}\big) = \big(\frac{a}{p}\big)\big(\frac{b}{p}\big)$.

(3) (Euler's Criterion) *If $p \nmid a$, then* $\big(\frac{a}{p}\big) \equiv (a)^{\frac{p-1}{2}} \mod p$. *In particular,* $\big(\frac{-1}{p}\big) = (-1)^{\frac{p-1}{2}}$.

(4) $\sum_{a=1}^{p-1} \big(\frac{a}{p}\big) = 0$.

Proof. We have already observed (1), and (4) is the fact that half of $(\mathbb{Z}/p\mathbb{Z})^\times$ are squares and half not.

(3): We know that $x^{p-1} \equiv 1 \mod p$ if $p \nmid x$, so if a is a square modulo p, then $a^{\frac{p-1}{2}} \equiv 1 \mod p$. Conversely, suppose that a is a square modulo p. As $(\mathbb{Z}/p\mathbb{Z})^\times$ is a cyclic group of order $p-1$, we have $(\mathbb{Z}/p\mathbb{Z})^\times = \langle x \rangle$ for some x. Suppose that $a \equiv x^n \mod p$, then n must be even lest $x^{\frac{p-1}{2}} \equiv 1 \mod p$, hence a is a square modulo p. The case of $a = -1$ follows easily as $\big(\frac{-1}{p}\big) - (-1)^{\frac{p-1}{2}}$ is -2, 0, or 2. (In fact, we have previously shown this case.)

(2) follows from (3). □

Proposition 56.14. *Let p be an odd prime and ζ a primitive pth root of unity. Set*

$$S := \sum_{a=1}^{p-1} \big(\frac{a}{p}\big)\zeta^a$$

in $\mathbb{Q}(\zeta)$ (even in $\mathbb{Z}[\zeta]$). Then

$$S^2 = \big(\frac{-1}{p}\big)p = (-1)^{\frac{p-1}{2}}p \text{ in } \mathbb{Q} \text{ (even in } \mathbb{Z}).$$

In particular,

$$\sqrt{p} \text{ lies in } \mathbb{Q}(\zeta) \text{ if } p \equiv 1 \mod 4$$
$$\sqrt{-p} \text{ lies in } \mathbb{Q}(\zeta) \text{ if } p \equiv 3 \mod 4.$$

Moreover, $\mathbb{Q}(S)$ is the unique quadratic extension of $\mathbb{Q}$ lying in $\mathbb{Q}(\zeta)$.

The sum S in the proposition is called a *Gauss sum*.

PROOF. We have

$$S^2 = \sum_{a=1,\, b=1}^{p-1} \big(\frac{a}{p}\big)\big(\frac{b}{p}\big)\zeta^{a+b} = \sum_{a=1,\, b=1}^{p-1} \big(\frac{ab}{p}\big)\zeta^{a+b}.$$

As a ranges over $1, \dots, p-1 \mod p$, so does ab for $b = 1, \dots, p-1 \mod p$, so upon replacing a by ab, we see that

$$\begin{aligned} S^2 &= \sum_{a=1,\, b=1}^{p-1} \big(\frac{ab^2}{p}\big)\zeta^{b(a+1)} = \sum_{a=1,\, b=1}^{p-1} \big(\frac{a}{p}\big)\zeta^{b(a+1)} \\ &= \sum_{b=1}^{p-1} \big(\frac{-1}{p}\big)\zeta^{bp} + \sum_{a=1}^{p-2} \big(\frac{a}{p}\big) \sum_{b=1}^{p-1} \zeta^{b(a+1)}, \end{aligned} \tag{*}$$

where the first term on the second line arises from the term $a = p-1$. As ζ^r is also a primitive pth root of unity for $r = 1, \dots, p-1$, we have $1 + \zeta^r + (\zeta^r)^2 + \cdots + (\zeta^r)^{p-1} = 0$ for such r. In particular, we have $\sum_{b=1}^{p-1} \zeta^{b(a+1)} = -1$ if $a \neq p-1$ in (*). Therefore, we have, using Lemma 56.13(4),

$$S^2 = \sum_{b=1}^{p-1} \big(\frac{-1}{p}\big) - \sum_{a=1}^{p-2} \big(\frac{a}{p}\big) = (p-1)\big(\frac{-1}{p}\big) + \big(\frac{p-1}{p}\big) = p\big(\frac{-1}{p}\big).$$

Consequently, as $S \subset \mathbb{Q}(\zeta)$, we have $\sqrt{p\big(\frac{-1}{p}\big)}$ lies in $\mathbb{Q}(\zeta)$. The "in particular statement" follows from Euler's Criterion. Finally as $G(\mathbb{Q}(\zeta)/\mathbb{Q}) \cong (\mathbb{Z}/p\mathbb{Z})^\times$ is a cyclic group of order $p-1$, there exists a unique subgroup H of $G(\mathbb{Q}(\zeta)/\mathbb{Q})$ of index two. By the Fundamental Theorem of Galois Theory, we must have $\mathbb{Q}(S) = \mathbb{Q}(\zeta)^H$. □

Remark 56.15. Let p be an odd prime and ζ a primitive p^rth root of unity in $\mathbb{C}$. Then $\mathbb{Q}\big(\sqrt{(-1)^{\frac{p-1}{2}}p}\big)$ is the unique quadratic extension of $\mathbb{Q}$ in $\mathbb{Q}(\zeta)$ by Remark 56.7.

Corollary 56.16. *Let p be a (positive) prime, ζ_p a primitive pth root of unity, and ζ_{4p} a primitive $4p$ th root of unity, then $\sqrt{p}$ and $\sqrt{-p}$ lie in $\mathbb{Q}(\zeta_{4p})$.*

PROOF. Let ζ_n be a fixed primitive nth root of unity in $\mathbb{C}$ for each n.

Case 1. p an odd prime:

If $p \equiv 1 \mod 4$, then by the proposition, we know that $\sqrt{p}$ lies in $\mathbb{Q}(\zeta_p) \subset \mathbb{Q}(\zeta_{4p})$ and $\sqrt{-p}$ lies in $\mathbb{Q}(\sqrt{-1}, \zeta_p) = \mathbb{Q}(\zeta_{4p})$ by Corollary 56.8. The case that $p \equiv 3 \mod 4$ is analogous.

Case 2. $p = 2$:
Let $y = \zeta_8 + \zeta_8^{-1}$. Then ζ_8 is the point $(1+\sqrt{-1})/2$ on the unit circle in the complex plane, i.e., the intersection with the 45^o ray with the real axis and ζ_8^{-1} is its complex conjugate. Computation shows $y^2 = 2$. It follows that $\sqrt{2}$ lies in $\mathbb{Q}(\zeta_8)$. As $\sqrt{-1}$ lies in $\mathbb{Q}(\zeta_4) \subset \mathbb{Q}(\zeta_8)$, it follows that $\sqrt{2}$ and $\sqrt{-2}$ lie in $\mathbb{Q}(\zeta_8)$. □

Theorem 56.17. *Let n be a nonzero integer. Then $\sqrt{n}$ lies in $\mathbb{Q}(\zeta_{4n})$, where ζ_{4n} is a primitive $4n$ th root of unity. In particular, any quadratic extension of $\mathbb{Q}$ lies in $\mathbb{Q}(\zeta)$ for some root of unity ζ in $\mathbb{C}$.*

PROOF. Let ζ_n be a fixed primitive nth root of unity in $\mathbb{C}$ for each n. We may assume that n is square-free. We know the result if $n = -1$, so we may assume that $n = \pm p_1 \cdots p_r$ with $p_1 < \cdots < p_r$ positive primes with $r \geq 1$. Then $\sqrt{n}$ is an element of $\mathbb{Q}(\zeta_4, \zeta_{p_1}, \ldots \zeta_{p_r}) = \mathbb{Q}(\zeta_{4n})$ if n is odd and $\sqrt{n}$ is an element of $\mathbb{Q}(\zeta_8, \zeta_{p_2}, \ldots \zeta_{p_r}) = \mathbb{Q}(\zeta_{4n})$ if n is even. □

We turn to a proof of Quadratic Reciprocity, one of hundreds!

Theorem 56.18. (Law of Quadratic Reciprocity) *Let p and q be distinct odd primes. Then*

$$\big(\frac{p}{q}\big)\big(\frac{q}{p}\big) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}.$$

In addition, we have

(Supplement #1) $$\big(\frac{-1}{p}\big) = (-1)^{\frac{p-1}{2}}.$$

(Supplement #2) $$\big(\frac{2}{p}\big) = (-1)^{\frac{p^2-1}{8}}.$$

PROOF. Let ζ be a primitive pth root of unity in $\mathbb{C}$ and

$$S := \sum_{a=1}^{p-1} \big(\frac{a}{p}\big)\zeta^a \text{ in } \mathbb{Z}[\zeta].$$

By Proposition 56.14 and Euler's Criterion, we have

$$S^q = S(S^2)^{\frac{q-1}{2}} = S\big(\frac{-1}{p}\big)^{\frac{q-1}{2}} p^{\frac{q-1}{2}} = S(-1)^{\frac{p-1}{2}\frac{q-1}{2}} p^{\frac{q-1}{2}}.$$

Multiplying this equation by S, we see that

$$(*) \qquad \begin{aligned} S^{q+1} &= S^2(S^2)^{\frac{q-1}{2}} = S^2(-1)^{\frac{p-1}{2}\frac{q-1}{2}} p^{\frac{q-1}{2}} \\ &= p^{\frac{q-1}{2}}\big(\frac{-1}{p}\big)(-1)^{\frac{p-1}{2}\frac{q-1}{2}} p^{\frac{q-1}{2}}, \end{aligned}$$

hence is an integer. By the Binomial Theorem (actually the Multinomial Theorem), since q is an odd prime, we see upon multiplying out and coalescing terms that

$$S^q = \Big(\sum_{a=1}^{p-1} \big(\frac{a}{p}\big)\zeta^a\Big)^q = \sum_{a=1}^{p-1} \big(\frac{a}{p}\big)^q \zeta^{aq} + qr = \sum_{a=1}^{p-1} \big(\frac{a}{p}\big)\zeta^{aq} + qr$$

for some r in $\mathbb{Z}[\zeta]$. As aq, $a = 1, \ldots, p-1$, runs over all the nonzero residue classes modulo p, we have

$$\begin{aligned} S^q &= \sum_{a=1}^{p-1} \big(\frac{q^2 a}{p}\big)\zeta^{aq} + qr = \big(\frac{q}{p}\big) \sum_{a=1}^{p-1} \big(\frac{aq}{p}\big)\zeta^{aq} + qr \\ &= \big(\frac{q}{p}\big) \sum_{a=1}^{p-1} \big(\frac{a}{p}\big)\zeta^{a} + qr = \big(\frac{q}{p}\big) S + qr. \end{aligned}$$

Multiplying this by S then yields

$$S^{q+1} = \big(\frac{q}{p}\big) S^2 + qr = \big(\frac{q}{p}\big)\big(\frac{-1}{p}\big) p + qrS$$

in $\mathbb{Z}[\zeta]$.

Check 56.19. $q\mathbb{Z}[\zeta_p] \cap \mathbb{Z} = q\mathbb{Z}$.

So we have, using (*),

$$qrS = S^{q+1} - \big(\frac{q}{p}\big)\big(\frac{-1}{p}\big) p \text{ lies in } q\mathbb{Z},$$

which implies

$$S^{q+1} \equiv \big(\frac{q}{p}\big)\big(\frac{-1}{p}\big) p \mod q.$$

Using (*) again, now shows that

$$p\big(\frac{-1}{p}\big)(-1)^{\frac{p-1}{2}\frac{q-1}{2}} p^{\frac{q-1}{2}} \equiv \big(\frac{q}{p}\big)\big(\frac{-1}{p}\big) p \mod q.$$

Since $p \mod q$ is a unit in $\mathbb{Z}/q\mathbb{Z}$ and $p^{\frac{p-1}{2}} \equiv \big(\frac{p}{q}\big) \mod q$ by Euler's Criterion, we have

$$\big(\frac{p}{q}\big) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}} \big(\frac{q}{p}\big) \mod q.$$

Finally, as $\big(\frac{p}{q}\big) - (-1)^{\frac{p-1}{2}\frac{q-1}{2}} \big(\frac{q}{p}\big)$ is $-2, 0$, or 2, the result follows.

Since we have already proven Supplements #1, to finish we need only establish Supplement #2. We work in $\mathbb{Z}/p\mathbb{Z}$ with p an odd prime. Let ζ be a primitive 8th root of unity in an algebraic extension of $\mathbb{Z}/p\mathbb{Z}$, i.e., a root of $t^8 - 1$ over $\mathbb{Z}/p\mathbb{Z}$ generating the 8th roots of unity and

set $y = \zeta + \zeta^{-1}$ in $(\mathbb{Z}/p\mathbb{Z})(\zeta)$. As $\zeta^8 = 1$ and ζ has order 8, we have $\zeta^4 = -1$ in $(\mathbb{Z}/p\mathbb{Z})(\zeta)$ with $1 \neq -1$ since p is odd. Multiplying this last equation by ζ^{-2}, we see that $\zeta^2 + \zeta^{-2} = 0$ in $(\mathbb{Z}/p\mathbb{Z})(\zeta)$, hence $y^2 = (\zeta+\zeta^{-1})^2 = \zeta^2+2+\zeta^{-2} = 2$. (Cf. this to the case in characteristic zero.) Let $^- : \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ be the canonical epimorphism. By Euler's Criterion, $\left(\frac{2}{p}\right) = \overline{2}^{\frac{p-1}{2}} = y^{p-1}$ in $\mathbb{Z}/p\mathbb{Z}$. (Note that $y \notin \mathbb{Z}/p\mathbb{Z}$, so y^{p-1} is not necessarily one.) By the Children's Binomial Theorem, $y^p = \zeta^p + \zeta^{-p}$ in $(\mathbb{Z}/p\mathbb{Z})(\zeta)$.

Case 1. $p \equiv \pm 1 \mod 8$:

As $\zeta^8 = 1$, we have

$$y^p = \zeta^p + \zeta^{-p} = \zeta + \zeta^{-1} = y \text{ in } (\mathbb{Z}/p\mathbb{Z})(\zeta),$$

so $\left(\frac{2}{p}\right) = y^{p-1} = 1$ in $(\mathbb{Z}/p\mathbb{Z})(\zeta)$, hence $\left(\frac{2}{p}\right) = 1$ in $\mathbb{Z}$.

Case 2. $p \equiv \pm 5 \mod 8$:

As $\zeta^4 = -1$, we have

$$y^p = \zeta^p + \zeta^{-p} = \zeta^5 + \zeta^{-5} = -\zeta - \zeta^{-1} = -y \text{ in } (\mathbb{Z}/p\mathbb{Z})(\zeta),$$

so $\left(\frac{2}{p}\right) = y^{p-1} = -1$ in $(\mathbb{Z}/p\mathbb{Z})(\zeta)$, hence $\left(\frac{2}{p}\right) = -1$ in $\mathbb{Z}$.

This completes the proof. □

Examples 56.20. 1. Quadratic Reciprocity and the properties of the Legendre symbol allows computing Legendre symbols quite efficiently. For example,

$$\begin{aligned}\left(\frac{29}{43}\right) &= (-1)^{\frac{29-1}{2}\frac{43-1}{2}}\left(\frac{43}{29}\right) = \left(\frac{43}{29}\right) = \left(\frac{14}{29}\right) = \left(\frac{2}{29}\right)\left(\frac{7}{29}\right)\\ &= (-1)^{\frac{(29-1)(29+1)}{8}}\left(\frac{7}{29}\right) = -\left(\frac{7}{29}\right) = (-1)^{\frac{7-1}{2}\frac{29-1}{2}}\left(\frac{29}{7}\right)\\ &= -\left(\frac{29}{7}\right) = -\left(\frac{1}{7}\right).\end{aligned}$$

So 29 is not a square modulo 43.

2. The Legendre Symbol can also be use to study Diophantine equations. Let x and y be variables. We show that the Diophantine equation $y^2 = x^3 + 45$ has no solution, i.e., no solution in integers. Suppose that $(x_0, y_0) \in \mathbb{Z}^2$ is a solution. We first show that we cannot have $3 \mid x_0$. If $x_0 = 3x_1$, then $y_0 = 3y_1$ for some integers x_1, y_1. It follows that $3^2y_1^2 = 3^3x_1^3 + 3^25$ or $y_1^2 \equiv 5 \equiv 2 \mod 3$ which is impossible. If x_0 is even, then $y_0^2 \equiv 45 \equiv 5 \mod 8$ which is impossible and if $x_0 \equiv 1 \mod 4$, then $y_0^2 \equiv 46 \equiv 2 \mod 4$, which is also

impossible. Therefore, we are reduced to the case $x_0 \equiv 3 \mod 4$, equivalently, $x_0 \equiv 3, 7 \mod 8$.

Case. $x_0 \equiv 7 \equiv -1 \mod 8$.

As $y_0^2 = x_0^3 + 45$, we have $y_0^2 - 2 \cdot 3^2 \equiv x_0^3 + 27 = (x_0 + 3)(x_0^2 - 3x + 9)$ and $x_0^2 - 3x_0 + 9 \equiv -3 \mod 8$. Therefore, there exists a prime p satisfying $p \equiv \pm 3 \mod 8$ and $p \mid x_0^2 - 3x_0 + 9$. This implies that $y_0^2 = 2 \cdot 3^2 \mod p$. If $p \mid y_0$, then $p \mid x_0$ which is impossible. So $p \neq 3$, hence $\left(\frac{2 \cdot 3^2}{p}\right) = \left(\frac{2}{p}\right) = -1$, contradicting the Second Supplement.

Case. $x_0 \equiv 3 \mod 8$.

We have $y_0^2 - 2 \cdot 6^2 = x_0^3 - 27 = (x_0 - 3)(x_0^2 + 3x_0 + 9)$, so there exists a prime p, satisfying $p \equiv \pm 3 \mod 8$ and $p \mid x_0^2 + 3x_0 + 9$, as $x_0^2 + 3x_0 + 9 \equiv 3 \mod 8$. Therefore, $y_0^2 \equiv 2 \cdot 6^2 \mod p$ has a solution. Since $p \neq 3$, hence $\left(\frac{2 \cdot 6^2}{p}\right) = \left(\frac{2}{p}\right) = -1$, contradicting the Second Supplement.

Let $K/\mathbb{Q}$ be a finite field extension. Number theory studies the set $\mathcal{O}_K := \{x \in K \mid x \text{ is a root of a monic polynomial } f \text{ in } \mathbb{Z}[t]\}$. We mention some of the facts about this here, and study this subject in Chapter XV. The set $\mathcal{O}_K$ is a ring, hence a domain. It is called the *ring of algebraic integers* in K. Although it is not a UFD, it has an analogous property about ideals, viz, that every proper ideal in $\mathcal{O}_K$ is a product of prime ideals, unique up to order. Such a domain is called a *Dedekind domain*. It turns out that every nonzero prime ideal $\mathfrak{p}$ in $\mathcal{O}_K$ is maximal, so $\mathcal{O}_K/\mathfrak{p}$ is a field, in fact a finite field so a finite extension of $\mathbb{Z}/p\mathbb{Z}$ where $\mathfrak{p} \cap \mathbb{Z} = p\mathbb{Z}$ for some prime p. Let $f_\mathfrak{p}$ denote the degree of this extension. One studies the factorization of $p\mathbb{Z}$, with p a prime integer, in $\mathcal{O}_K$. Suppose that this factorization is $p\mathcal{O}_K = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$, with $\mathfrak{p}_i$ distinct prime ideals in $\mathcal{O}_K$ and e_i positive integers. A basic result is that $[K : \mathbb{Q}] = \sum_i^r e_i f_{\mathfrak{p}_i}$. If, in addition, $K/\mathbb{Q}$ is Galois then all the e_i's are equal, say equal e, and all the $\mathfrak{p}_i$'s are equal, say equal f, so $[K : \mathbb{Q}] = efr$. The Legendre symbol provides information about the case that $K/\mathbb{Q}$ is a quadratic extension (hence Galois), say $K = \mathbb{Q}(\sqrt{d})$ with d a square-free integer or -1. In this case $\mathcal{O}_K = \mathbb{Z}[D]$ where $D = d$ if $d \equiv 2$ or $3 \mod 4$ and $D = (-1 + \sqrt{d})/2$ if $d \equiv 1 \mod 4$. Let p be a prime integer. Three possibilities can occur (called the *splitting type* of p).

(*i*) $r = 2$, i.e., $p\mathcal{O}_K$ factors as a product of two distinct primes – we say that p *splits completely* in $\mathcal{O}_K$.

(*ii*) $f = 1$, i.e., $p\mathcal{O}_k$ is a prime ideal in $\mathcal{O}_K$ – we say that p is *inert* in $\mathcal{O}_K$.

(*iii*) $e = 1$, i.e., $p\mathcal{O}_K$ is the the square of a prime ideal in $\mathcal{O}_K$ – we say that p *ramifies* in $\mathcal{O}_K$.

A prime p ramifies in $\mathcal{O}_K$ if and only if $p \mid d_K$, where $d_k = 4D$ if $D \equiv 2$ or $3 \mod 4$ and $d_K = D$ if $d \equiv 1 \mod 4$. Let p be an odd prime. Then p ramifies if and only if $p \mid D$, so assume this is not the case. Then $f = 1$ or 2 depending on whether d is a square or not modulo p, i.e., on $(\frac{d}{p})$. If $p\mathbb{Z}$ splits completely and a is a square modulo p, then $p\mathcal{O}_K = (p, a+\sqrt{d})(p, a-\sqrt{d})$. Note that if p and q are odd primes, the Law of Quadratic Reciprocity tells the relation of the splitting type of p in $\mathcal{O}_{\mathbb{Z}[\sqrt{q}]}$ and the splitting type of q in $\mathcal{O}_{\mathbb{Z}[\sqrt{p}]}$. The splitting type of 2 is more complicated.

Exercises 56.21. 1. Show that the set of all arithmetic functions satisfies the associative law under the Dirichlet product. In particular, the set $\mathcal{A}$ of arithmetic functions nonzero at 1 is an abelian monoid.

2. Let N, τ, and σ be the arithmetic function functions defined by

$$\begin{aligned} N(n) &= n, \\ \tau(n) &= \text{ the number of divisors of } n, \text{ i.e., } \tau(n) = \sum_{d|n} 1, \text{ and} \\ \sigma(n) &= \text{ the sum of the divisors of } n, \text{ i.e., } \sigma(n) = \sum_{d|n} d \end{aligned}$$

for all n, respectively. Show that $N = \varphi \star U$, $\varphi = N \star \mu$, $\sigma = \varphi \star \tau$, and $\varphi \star \sigma = N \star N$.

3. Let $\zeta(s) := \sum \frac{1}{n^s}$, s a real (or complex) variable. Show the following:

$$\frac{\zeta(s-1)}{\zeta(s)} = \sum_1^\infty \frac{\varphi(n)}{n^s} \text{ if (the real part of) } s > 2.$$

$$\zeta(s)^2 = \sum_1^\infty \frac{\tau(n)}{n^s} \text{ if (the real part of) } s > 1.$$

$$\zeta(s-1)\zeta(s) = \sum_1^\infty \frac{\sigma(n)}{n^s} \text{ if (the real part of) } s > 2.$$

4. Show all of the following:

(*i*) The Dirichlet product of two multiplicative functions is multiplicative.

(*ii*) If f and g are arithmetic functions not zero at one with g and $f \star g$ multiplicative, then f is multiplicative.

5. The set of multiplicative functions is an abelian group under the Dirichlet product.
6. Prove the following versions of the Möbius Inversion Theorem.
   (a) If $f, g : \mathbb{Z}^+ \to G$ with G an additive group, then
   $$f(n) = \sum_{d|n} g(d) \text{ if and only if } g(n) = \sum_{d|n} f(d)\mu(\frac{n}{d})$$
   for all $n \in \mathbb{Z}^+$, where μ is viewed as a function to G.
   (b) If $f, g : \mathbb{Z}^+ \to G$ with G a multiplicative group, then
   $$f(n) = \prod_{d|n} g(d) \text{ if and only if } g(n) = \prod_{d|n} f(d)\mu(\frac{n}{d})$$
   for all $n \in \mathbb{Z}^+$, where μ is viewed as a function to G.
7. Let F be any field and d a positive integer. Show that $t^d - 1 \mid t^n - 1$ in $F[t]$ if and only if $d \mid n$.
8. Show that every finite cyclic group occurs as a Galois group $G(K/\mathbb{Q})$ for some Galois extension $K/\mathbb{Q}$.
9. Prove that every finite abelian group occurs as a Galois group over the rational numbers. (Hint: Use Proposition 56.10.)
10. Prove Check 56.19.
11. Let $K = \mathbb{Q}(\sqrt{d})$ with d a square-free integer (or -1) and p an odd prime integer. Show directly that $p\mathcal{O}_K = (p, a + \sqrt{d})(p, a - \sqrt{d})$ in $\mathcal{O}_K$ if p does not divide d and a in $\mathbb{Z}$ is a square modulo p.
12. Let $K = \mathbb{Q}(\sqrt{d})$ with d a square-free integer (or -1) and p an odd prime integer. Show directly that $p\mathcal{O}_K$ is a prime ideal if $\mathcal{O}_K$ if a in $\mathbb{Z}$ is not a square modulo p.
13. Let $K = \mathbb{Q}(\sqrt{d})$ with d a square-free integer (or -1) and p an odd prime integer. Show directly that $p\mathcal{O}_K = (p, \sqrt{d})^2$ in $\mathcal{O}_K$ if $p \mid d$.

57. Radical Extensions

In this section, we generalize the notion of square root towers and solve the problem of when a formula for roots of a polynomial involving only addition, multiplication (in which we include taking inverses) and extraction of nth roots of elements for various n over the rational numbers exists. The key is if the Galois group of a polynomial is a solvable group.

Definition 57.1. An extension of fields K/F is called a *radical extension* if there exist elements u_i in K for $1 \leq i \leq m$ (some m) such that $K = F(u_1, \ldots, u_m)$ and for each i there exists a positive integer n_i such that $u_1^{n_1}$ lies in F and $u_i^{n_i} \in F(u_1, \ldots, u_{i-1})$ for each $i > 1$.

As one would expect, radical extensions have properties similar to square root towers, and they do.

Remark 57.2. Let K/F and L/F be field extensions with K and L lying in an extension field of K.

1. If K/F is a square root tower, then K/F is radical.
2. If K/F is the splitting field of $t^n - 1$ in $F[t]$, then K/F is radical.
3. If K/F is the splitting field of $t^n - a$ in $F[t]$, then K/F is radical, as $K = F(\omega, \theta)$ with ω a primitive nth root of unity in K and θ an element in K such that $\theta^n = a$ in K.
4. If K/F is radical, then so is $L(K)/L$.
5. If L/K and K/F are both radical, then so is L/F.
6. If L/F is radical and L/K, then L/K is radical.
7. If $E/L_i/F$ are intermediate fields with each L_i/F radical for $i = 1, \ldots, n$, then $F(L_1 \cup \cdots \cup L_n)/F$ is radical.
8. If K/F is radical and $\sigma : K \to L$ is a (field) homomorphism, then $\sigma(K)/\sigma(F)$ is radical.
9. If K/F is radical and L/K is a normal closure of K/F, then L/F is radical. Indeed this follows from the last two remarks, as $L = F\big(\cup_{G(L/F)} \sigma(K)\big)$.
10. If K/F is a square root tower and L/K is a normal closure of K/F, then L/F is a square root tower.

The last remark allows us to refine the Constructibility Criterion 49.9.

Theorem 57.3. (Constructibility Criterion (Refined Form)) *Let z be a complex number and $z_1(= 0), z_2(= 1), \ldots, z_n$ other complex numbers with $n \geq 2$. Set $F = \mathbb{Q}(z_1, z_2, \ldots, z_n, \overline{z_1}, \ldots, \overline{z_n})$. Then the following are equivalent:*

(1) The complex number z is constructible from $z_1, \ldots, z_n$.
(2) The complex number z is algebraic over F and the normal closure of $F(z)/F$ in $\mathbb{C}$ is a square root tower.
(3) The complex number z is algebraic over F and if $E/F(z)$ is the normal closure of $F(z)/F$ in $\mathbb{C}$, then $[E : F] = 2^e$ for some e.

PROOF. (2) if and only if (3) follows from the Square Root Tower Theorem 54.11.

(2) $\Rightarrow$ (1) follows from the original Constructibility Criterion 49.9.

(1) $\Rightarrow$ (2): By the original Constructibility Criterion 49.9, there exists a square root tower K/F with $z \in K$. By Remark (10) above, we may assume that K/F is normal. Let $K/E/F(z)/F$ with $E/F(z)$ the normal closure of $F(z)/F$ in K. Since $[E : F] \mid [K : F] = 2^e$, some e, we are done by the Square Root Tower Theorem 54.11. □

We wish to find the group theoretic criterion for a finite extension to be a radical extension. For simplicity we shall establish this in the case that the ground field is of characteristic zero.

Let K/F be a finite Galois extension of fields. Recall that we call it an *abelian extension* (respectively, *cyclic extension*) if $G(K/F)$ is abelian (respectively, cyclic).

Example 57.4. 1. Any cyclic extension of fields is abelian.

2. Suppose n is a positive integer and either $\operatorname{char} F = 0$ or $\operatorname{char} F \nmid n$. If K is a splitting field of $t^n - 1$ over F, then K/F is abelian (and, in fact, cyclic if n is $2, 4, p^r, 2p^r$, with p an odd prime, cf. 105.6 below).
3. If K/F is abelian (respectively, cyclic) and $K/E/F$ is an intermediate field, then both K/E and E/F are abelian (respectively, cyclic).
4. Suppose n is a positive integer and either $\operatorname{char} F = 0$ or $\operatorname{char} F \nmid n$. Suppose that $t^n - 1$ splits over F. If K is a splitting field of $t^n - a$ over F, then K/F is cyclic.

We review our study about solvable groups.

Review 57.5. A group G is called *solvable* (respectively, *polycyclic*) if there exist a finite sequence of subgroups of G:

$$1 = N_0 \subset N_1 \subset \cdots \subset N_r = G$$

satisfying $N_i \lhd N_{i+1}$ and N_{i+1}/N_i is abelian (respectively, cyclic) for every $i = 1, \ldots, r-1$, i.e., G has a finite subnormal series with abelian (respectively cyclic) factors. Let G be a group. We know the following facts about solvable and polycyclic groups:

1. Abelian groups are solvable.
2. If G is solvable and H a subgroup of G, then H is solvable.
3. If N is a normal subgroup of G, then G is solvable if and only if both N and G/N are solvable.
4. If G is polycyclic, then G is solvable.
5. If G is finite and solvable, then G is polycyclic.

6. The alternating group A_n with $n \geq 5$ is a nonabelian simple group, hence not solvable.
7. The symmetric group S_n with $n \geq 5$ is not solvable.

Applying the above to field theory we have:

Remarks 57.6. Let K/F be a finite extension of fields with $K/E/F$ an intermediate field.

1. If $G(K/F)$ is solvable, then $G(K/E)$ is solvable.
2. If K/F and E/F are Galois, then $G(K/F)$ is solvable if and only if both $G(K/E)$ and $G(E/F)$ are solvable, since, by the Fundamental Theorem of Galois Theory, $G(E/F) \cong G(K/F)/G(K/E)$.

Given a field extension, it is often useful to extend the base field by adjoining appropriate roots of unity. In investigating radical extensions, this is most useful. Indeed using the last remark we easily establish the following lemma.

Lemma 57.7. *Let n be a positive integer and F a field satisfying $\operatorname{char} F = 0$ or $\operatorname{char} F \nmid n$. Suppose that K/F is a splitting field of a separable polynomial f over F and L/F is a a splitting field of $t^n - 1$ over K. Then $G(L/F)$ is Galois. Moreover, $G(L/F)$ is solvable if and only if $G(K/F)$ is solvable.*

PROOF. We know that L is a splitting field of the separable polynomial $(t^n - 1)f$ over F, so L/F is Galois. As L/K is abelian, by Remark 57.6, the group $G(L/F)$ is solvable if and only if the group $G(K/F)$ is solvable. □

The key result is the following theorem.

Theorem 57.8. *Let F be a field of characteristic zero and K/F a radical extension. If $K/E/F$ is an intermediate field, then $G(E/F)$ is solvable.*

PROOF. We begin with two reductions.

Reduction 1. We may assume that E/F is Galois:

Let $F_0 = E^{G(E/F)} \supset F$. Then K/F_0 is also radical, $G(E/F) = G(E/F_0)$, and E/F_0 is Galois. Replacing F by F_0 establishes the reduction.

Reduction 2. We may assume that K/F is radical:

Let L/K be a normal closure of of K/F. Since K/F is radical, so is L/F. Replacing K by L establishes the second reduction.

So we may now assume that K/F is a Galois and a radical extension. By Remark 57.6, it now suffices to show that $G(K/F)$ is solvable as E/F is Galois. Therefore, we have reduced to proving the following:

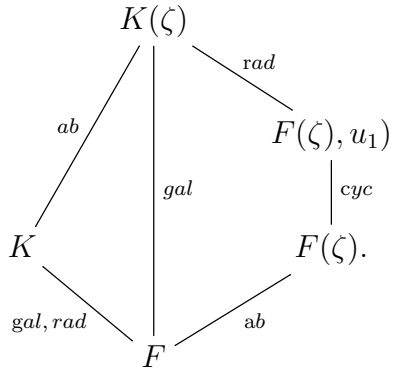
Claim. Let F be a field of characteristic zero and K/F a finite extension that is both radical and Galois. Then $G(K/F)$ is solvable:

Let $K = F(u_1, \ldots, u_m)$ and n_i positive integers, $i = 1, \ldots, m$, satisfying $u_1^{n_1}$ lies in F and $u_i^{n_i} \in F(u_1, \ldots, u_{i-1})$ for each $i > 1$. We prove the claim by induction.

Induction Hypothesis. Let L/E be a finite, radical, Galois extension of fields with E a field of characteristic zero. Suppose that $L = E(v_1, \ldots, v_{m-1})$ and r_i positive integers, $i = 1, \ldots m-1$, satisfying $v_1^{r_1}$ lies in E and $v_i^{r_i} \in E(v_1, \ldots, v_{i-1})$ for each $i > 1$. Then $G(L/E)$ is solvable:

Note: The $m = 0$ case is trivial (and the $m = 1$ case is included in the proof below).

With $K = F(u_1, \ldots, u_m)$ as above, let ζ be a primitive n_1st root of unity, so a primitive root of unity of $t^{n_1} - 1$ over K. We have the following picture with the explanation to follow:



$K(\zeta)/K$ and $F(\zeta)/F$ are Galois and abelian, since a splitting field over $t^{n_1} - 1$ over K and F, respectively. The extension $K(\zeta)/F$ is Galois by the lemma, hence $K(\zeta)/F(\zeta)$ and $K(\zeta)/F(\zeta, u_{n_1})$ are Galois. Since $F(\zeta, u_1)$ is a splitting field of $t^{n_1} - u_1^{n_1}$ over $F(\zeta)$, the extension $F(\zeta, u_1)/F$ is Galois, as it is a splitting field of the separable polynomial $(t^{n_1} - u_1^{n_1})(t^{n_1} - 1)$ over F. We know that $F(\zeta, u_1)/F(\zeta)$ is cyclic and $F(\zeta)/F$ is abelian, so $G(F(\zeta, u_1)/F)$ is solvable by Remark 57.6. In particular, all the extensions in the picture above are Galois. We have $K(\zeta) = F(\zeta, u_1)(u_2, \ldots, u_m)$ with $K(\zeta)/F(\zeta, u_1)$ radical and Galois. Consequently, by the induction hypothesis, $G\big(K(\zeta)/F(\zeta, u_1)\big)$ is solvable. As $K(\zeta)/F$ is Galois and $G\big(K(\zeta)/F(\zeta, u_1)\big)$ is solvable, the

group $G(K(\zeta)/F)$ is solvable by Remark 57.6. As K/F is Galois, the group $G(K/F)$ is solvable again by Remark 57.6. □

In the proof, we can, in fact, arrange for $K(\zeta) = K$ as K/F is normal and $u_1 \in K$, which means that $m_F(u_1)$ splits over K, hence contains ζ for an appropriately chosen root of unity — $t^{n_1} - u_1^{n_1}$ may be reducible over F.

The theorem shows that if f is an irreducible polynomial in $F[t]$, e.g., $F = \mathbb{Q}$, then a formula for a root of f exists that consists of addition, multiplication, and extraction of nth roots for various n, if and only if there exists a radical extension K of F such that f has a root in K. As the conjugate of a radical extension is radical, if such a formula for one of the roots of irreducible f_i, such a formula holds for each root of f_i in the appropriate conjugate.

Let F be a field and f be a non-constant polynomial in $F[t]$. We say that f is *solvable by radicals* if for each irreducible factor f_i of f in $F[t]$, there exists a formula for a root of f_i involving addition, multiplication, and extraction of nth roots for various n, i.e., there exists a radical extension K_i of F such that f_i has a root in K_i for each i.

We analyze what we have shown. Let F be a field of characteristic zero. Suppose that f in $F[t]$ is solvable by radicals and f_i an irreducible factor of f in $F[t]$. Then there exists a radical extension K_i/F such that f_i has a root in K. Assume that each such K_i lies in a common extension field $\widetilde{F}$, e.g., an algebraic closure of F. If there exist m distinct (non-associative) factors f_i of f, let $L = F(K_1 \cup \cdots \cup K_m)$, a radical extension of F. The normal closure $\widetilde{L}/L$ in $\widetilde{F}$ of L/F is also a radical extension, so there exists a radical Galois extension $\widetilde{L}/F$ such that f splits over $\widetilde{L}$. Let $\widetilde{L}/E/F$ with E a splitting field of f over F. Although E/F may not be radical, by the theorem, we know that $G(E/F)$ is solvable. Recall if f is a polynomial in $F[t]$ and K/F is a splitting field of f, then $G(K/F)$ is called the *Galois group of* f. In general it is only unique up to isomorphism, but if we have fixed an algebraic closure of F, as we have done this time, it is unique. Therefore, we have shown the following:

Theorem 57.9. *Let F be a field of characteristic zero and f an non-constant polynomial in $F[t]$. If f is solvable by radicals, then the Galois group of f is solvable.*

We wish to show that there exist polynomials in $\mathbb{Q}[t]$ whose Galois groups are not solvable by radicals. We need the following group theory result and its application to field theory.

Lemma 57.10. *Let p be a prime and G a subgroup of the symmetric group S_p containing a p-cycle and a transposition. Then $G = S_p$.*

This is not true if p is not a prime.

To prove the lemma, we recall the following fact: For any n-cycle $(a_1 \cdots a_n)$, we have

$$\sigma(a_1 \cdots a_n)\sigma^{-1} = \big(\sigma(a_1) \cdots \sigma(a_n)\big)$$

for all permutations σ in S_N with $n \leq N$.

PROOF. (of the lemma) By changing notation, we may assume that $(12) \in G$. Let $\sigma = (a_1 \cdots a_p)$ lie in G. As p is a prime, σ^i, $1 \leq i < p$, is also a p-cycle (cf. Exercise 22.22(1), so we may assume that $\sigma = (12 \cdots p)$. As $\sigma(1\,2)\sigma^{-1} = (2\,3)$, $\sigma(2\,3)\sigma^{-1} = (3\,4)$, etc., we see that $(i\,i+1)$ lies in G for all i. Therefore $(1\,i+1) = (1\,i)(i\,i+1)(1\,i)$ lies in G for all i by induction, hence $G = S_p$ by Proposition 22.6.

□

Proposition 57.11. *Let f be an irreducible polynomial in $\mathbb{Q}[t]$ of degree p with p a prime. Suppose that f has precisely two nonreal roots in the splitting field K in $\mathbb{C}$ of f over $\mathbb{Q}$. Then $G(K/\mathbb{Q}) \cong S_p$.*

PROOF. Let α in K be a root of f, so $[\mathbb{Q}(\alpha) : \mathbb{Q}] = \deg f = p$. As $\operatorname{char} K = 0$, we know that $K/\mathbb{Q}$ is Galois, hence $|G(K/\mathbb{Q}) = [K : \mathbb{Q}] \leq p!$. As $G(K/\mathbb{Q})$ acts as permutations on the roots of f and K is the splitting field of f, we may assume that $G(K/\mathbb{Q}) \subset S_p$. Since f has precisely two nonreal roots, the restriction of complex conjugation to K is a nontrivial ($\mathbb{Q}$-) automorphism of K and it must interchange these two nonreal roots and fix all the other roots of f, i.e., it corresponds to a transposition in $G(K/\mathbb{Q})$. As $p \mid [K : \mathbb{Q}] = |G(K/\mathbb{Q})|$, the group $G(K/\mathbb{Q})$ must contain an element of order p by Cauchy's Theorem 20.23. But the only elements in S_p of order p are p-cycles. Therefore, $G(K/\mathbb{Q}) = S_p$ by the lemma. □

As a consequence, we obtain the following famous theorem:

Theorem 57.12. (Abel-Ruffini) *In general, there is no formula to determine the roots of a general fifth degree polynomial in $\mathbb{Q}[t]$ that involves only addition, multiplication, and extraction of nth roots for various n.*

PROOF. It suffices to give an example of a polynomial f in $\mathbb{Q}[t]$ that has a nonsolvable Galois group. Let $f = t^5 - 6t + 3$. We know that f is irreducible by Eisenstein's Criterion, and by calculus it has three real roots in $\mathbb{R}$, since $f' = 5t^4 - 6$ has only two real roots, $f(-1) > 0$,

and $f(1) < 0$. By the proposition, the Galois group of f is isomorphic to S_5 which is not solvable. □

Corollary 57.13. *In general, there is no formula to determine the roots of a general nth degree polynomial in $\mathbb{Q}[t]$ that involves only addition, multiplication, and extraction of nth roots for various n.*

PROOF. Apply the theorem to any nth degree polynomial with $n \geq 5$ divisible by a fifth degree polynomial not solvable by radicals. □

Remark 57.14. 1. For each positive integer n, we have constructed fields F and K_n/F Galois with $G(K_n/F) \cong S_n$ in Example 54.6(3), so, in general, no formula can exist.

2. Hilbert showed that for each positive integer n, there exists a Galois extension $K_n/\mathbb{Q}$ with $G(K_n/\mathbb{Q}) \cong S_n$. This is not easy. It uses his Irreducibility Theorem. For a proof see Chapter 63. By the Primitive Element Theorem, $K_n = \mathbb{Q}(u)$ for some u, hence the Galois group of $m_{\mathbb{Q}}(u)$ is isomorphic to S_n, so is not solvable for $n \geq 5$.
3. There exist fields of characteristic zero such that every algebraic extension is radical, hence every non-constant polynomial over it is solvable by radicals, e.g., $\mathbb{R}$, $\mathbb{C}$. There are others, e.g., the so called local fields of characteristic zero that arise in number theory.
4. If F is a finite field, then any finite extension of F is cyclic so radical. In particular, the Galois group of any polynomial over f is solvable by radicals.

We want a converse to our result. To do so needs additional concepts.

Definition 57.15. Let K/F be a finite Galois extension of fields and x an element of K. Define the *norm* of x to be

$$\mathrm{N}_{K/F}(x) := \prod_{G(K/F)} \sigma(x)$$

and the *trace* of x to be

$$\mathrm{Tr}_{K/F}(x) := \sum_{G(K/F)} \sigma(x).$$

Examples 57.16. Let F be a field of characteristic different from two and d an element in F that is not a square. Let $K = F(\sqrt{d})$. If $x = a+b\sqrt{d}$ with a, b in F, then $\mathrm{N}_{K/F}(x) = a^2-b^2d$ and $\mathrm{Tr}_{K/F}(x) = 2a$. Both of these lie in F. If, in addition, x does not lie in F, i.e., b is not zero, then $K = F(x)$ and $m_F(x) = t^2 - \mathrm{Tr}_{K/F}(x)t + \mathrm{N}_{K/F}(x) = \big(t - (a + b\sqrt{d})\big)\big(t - (a - b\sqrt{d})\big)$ lies in $F[t]$.

Properties 57.17. Let K/F be a finite Galois extension of fields, $G = G(K/F)$, and x, y elements of K.

1. If τ lies in G, then $\tau G = G = G\tau$, so
$$\tau\big(\mathrm{N}_{K/F}(x)\big) = \mathrm{N}_{K/F}(x) = \mathrm{N}_{K/F}\big(\tau(x)\big)$$
$$\tau\big(\mathrm{Tr}_{K/F}(x)\big) = \mathrm{Tr}_{K/F}(x) = \mathrm{Tr}_{K/F}\big(\tau(x)\big).$$
In particular,
$$\mathrm{N}_{K/F}(x) \text{ and } \mathrm{Tr}_{K/F}(x) \text{ lie in } K^G = F:$$
We have
$$\tau(\prod_G \sigma(x)) = \prod_G \tau\sigma(x) = \prod_G \sigma(x) = \prod_G \sigma\tau(x)$$
$$\tau(\sum_G \sigma(x)) = \sum_G \tau\sigma(x) = \sum_G \sigma(x) = \sum_G \sigma\tau(x).$$
2. Suppose that $G = \{\sigma_1, \dots, \sigma_n\}$. As before let s_j denote the jth elementary symmetric function $\sum_{1\le i_1<\cdots<i_j\le n} t_{i_1}\cdots t_{i_j}$ (and $s_0 = 1$), then
$$m_F(x) = \prod \big(t - \sigma_i(x)\big) = \sum (-1)^{n-i} s_{n-i}\big(\sigma_1(x), \dots, \sigma_n(x)\big)t^i$$
in $F[t]$. In particular, if $K = F(x)$, then
$$\mathrm{N}_{K/F}(x) = s_n\big(\sigma_1(x), \dots, \sigma_n(x)\big)$$
$$\mathrm{Tr}_{K/F}(x) = s_1\big(\sigma_1(x), \dots, \sigma_n(x)\big).$$
What can you say if $F(x) < K$?
3. If x lies in F, then $\sigma(x) = x$ for each σ in G. It follows that if $K/E/F$ then
$$\mathrm{N}_{E/F}(x) = x^{[E:F]} \text{ and } \mathrm{Tr}_{E/F}(x) = [E:F]x,$$
as every F-embedding of E into K lifts to an element of G.
4. The norm is multiplicative, i.e., $\mathrm{N}_{K/F}(xy) = \mathrm{N}_{K/F}(x)\,\mathrm{N}_{K/F}(y)$ and, if x is nonzero, $\mathrm{N}_{K/F}(x^{-1}) = \big(\mathrm{N}_{K/F}(x)\big)^{-1}$ (as $\mathrm{N}_{K/F}(1) = 1$). In particular,
$$\mathrm{N}_{K/F} : K^\times \to F^\times \text{ is a group homomorphism.}$$
5. The map $\mathrm{Tr}_{K/F} : K \to F$ is an F-linear functional of F-vector spaces. By Dedekind's Lemma 51.3, $\mathrm{Tr}_{K/F}$ is not the trivial map as it is the sum of distinct F-homomorphisms. In particular, $\dim \ker \mathrm{Tr}_{K/F} = [K:F] - 1$.

We can generalize the norm and trace as follows:

Remark 57.18. Let E/F be a finite separable extension of fields of degree n with K/E a normal closure of E/F. Hence K/F is Galois. Let

$$\tau_1, \ldots, \tau_m : E \to K$$

be all the distinct F-homomorphisms. We know that $m = n$ as E/F is finite and separable. Let x be an element in E. Define the *norm* of x to be

$$\mathrm{N}_{E/F}(x) := \prod_i \tau_i(x)$$

and the *trace* of x to be

$$\mathrm{Tr}_{E/F}(x) := \sum_i \tau_i(x).$$

Since K/F is Galois, if $\sigma \in G(K/F)$, then

$$\sigma\tau_1, \ldots, \sigma\tau_n : E \to K$$

are distinct F-homomorphisms, so these are just a permutation of $\tau_1, \ldots, \tau_n$, hence

$$\mathrm{N}_{E/F}(x) \text{ and } \mathrm{Tr}_{E/F}(x) \text{ lie in } K^G = F$$

just as in the Galois case. (Note we only needed $K/E/F$ with K/F finite and Galois to deduce this i.e., is independent of the finite Galois extension of F containing E.) Moreover, Properties 57.17 hold for E/F with $\{\tau_1, \ldots, \tau_m\}$ replacing the σ in Property 57.17(1) and $\{\sigma_1, \ldots, \sigma_n\}$ in Property 57.17(2) This generalization is useful in applications, especially in studying towers of finite separable extensions of fields. (Cf. the exercises at the end of this section). This can be further generalized to include inseparable extensions, but we shall not pursue this here.

In his book on algebraic number theory, Hilbert had the following result as his 90th Satz (theorem):

Lemma 57.19. (Hilbert Theorem 90) *Let K/F be a Galois extension of fields of degree n. Suppose that K/F is cyclic with $G(K/F) = \langle\sigma\rangle$ and $x \in K$. Then*

$$\mathrm{N}_{K/F}(x) = 1 \text{ if and only if there exists a } y \in K^\times \text{ satisfying } x = \frac{y}{\sigma(y)},$$

i.e., we have an exact sequence

$$1 \to K^\times \xrightarrow{\psi} K^\times \xrightarrow{\mathrm{N}_{K/F}} F^\times$$

with $\psi(y) = y/\sigma(y)$.

PROOF. ($\Leftarrow$): If $x = y/\sigma(y)$, then

$$\mathrm{N}_{K/F}(x) = \frac{\mathrm{N}_{K/F}(y)}{\mathrm{N}_{K/F}\big(\sigma(y)\big)} = 1.$$

($\Rightarrow$): Suppose that

$$1 = \mathrm{N}_{K/F}(x) = x\sigma(x)\cdots\sigma^{n-1}(x).$$

In particular, x is nonzero. By Dedekind's Lemma 51.3, we know that the F-automorphisms $1, \sigma, \ldots, \sigma^{n-1}$ are independent, so if $a_0, \ldots, a_{n-1}$ lie in K, we have $\sum_{i=0}^{n-1} a_i\sigma^i(z) = 0$ for all z in K if and only if all the a_i are zero. Set $a_0 = 1$ and $a_i = x\sigma(x)\cdots\sigma^{i-1}(x)$ in $K^\times$ for $i = 1, \ldots, n-1$, then

$$\begin{aligned} g &= 1 + a_1\sigma + \cdots + a_{n-1}\sigma^{n-1} \\ &= 1 + x\sigma + x\sigma(x)\sigma^2 + \cdots + x\sigma(x)\cdots\sigma^{n-2}(x)\sigma^{n-1} \end{aligned}$$

is not trivial, so there exists an element u in K satisfying

$$0 \neq y := g(u) = u + x\sigma(u) + \cdots + x\sigma(x)\cdots\sigma^{n-2}(x)\sigma^{n-1}(u).$$

Since $\sigma^n = 1_K$ and $1 = \mathrm{N}_{K/F}(x) = x\sigma(x)\cdots\sigma^{n-1}(x)$, we see that

$$x\sigma(y) = x\sigma(u) + x\sigma(x)\sigma^2(u) + \cdots + x\sigma(x)\cdots\sigma^{n-1}(x)\sigma^n(u) = y,$$

i.e., $x = y/\sigma(y)$ as needed. □

Although the computational proof reveals little, generalizations of the lemma above have resulted in very deep theorems in algebra. Hilbert established it to prove the following important result.

Theorem 57.20. *Let K/F be a Galois extension of fields of degree n with the characteristic of F either zero or $\operatorname{char} F \nmid n$. Suppose that K/F is cyclic with $G(K/F) = \langle\sigma\rangle$ and $t^n - 1$ splits over F. Then there exists an element u in K satisfying $K = F(u)$ and $m_F(u) = t^n - a$ in $F[t]$ for some a in F. In particular, K/F is radical.*

PROOF. Let ζ be a primitive nth root of unity in F. By our previous work, we know that $U = \langle\zeta\rangle$ is a cyclic group satisfying $|U| = n$, so $|U| = [K : F]$. As ζ^{-1} lies in F and $\mathrm{N}_{K/F}(\zeta^{-1}) = \zeta^{-n} = 1$, an application of Hilbert Theorem 90 shows that there exists a nonzero element u in K satisfying $\zeta = \sigma(u)/u$, i.e., $\sigma(u) = \zeta u$. Since $\sigma(u^n) = \sigma(u)^n = \zeta^n u^n = u^n$, we have $\sigma^i(u^n) = u^n$ for all i. Therefore, $a := u^n$ lies in $K^{G(K/F)} = F$ as K/F is Galois. Let $f := t^n - a$, a polynomial lying in $F[t]$. This polynomial has n distinct roots: $\zeta^i u$ for $i = 1, \ldots, n$. In particular, $F(u)$ is a splitting field of f over F. To finish, we need to show that $K = F(u)$, as then $\deg f = \deg m_F(u)$. We know that

$$\sigma^i|_{F(u)} : F(u) \to K \text{ maps } u \mapsto \zeta^i u$$

comprise n distinct F-homomorphisms, so by Artin's Lemma 51.8, we have

$$[F(u) : F] \geq n = |U| = [K : F],$$

hence $K = F(u)$. □

This theorem allows us to deduce the desired appropriate converse of Theorem 57.8.

Theorem 57.21. *Suppose that F is a field of characteristic zero and K/F is a finite Galois extension with $G(K/F)$ solvable. Then there exists a finite extension L/K such that L/F is Galois and radical.*

PROOF. Let $n = [K : F] = |G(K/F)|$.

Case 1. The polynomial $t^n - 1$ splits over F:

We show in this case that K/F itself is radical and Galois. As the group $G(K/F)$ is finite and solvable, it is polycyclic, i.e., has a cyclic series say

$$1 = N_1 \lhd N_2 \lhd \cdots \lhd N_r = G$$

with each N_i/N_{i-1} cyclic. Let $F_i = K^{N_i}$. By the Fundamental Theorem of Galois Theory 54.3, we conclude that

$$K/F_{i+1} \text{ is Galois with } G(K/F_{i+1}) = N_{i+1},$$

and, as $N_i \lhd N_{i+1}$,

$$F_i/F_{i+1} \text{ is normal with } G(F_i/F_{i+1}) \cong N_{i+1}/N_i \text{ cyclic.}$$

Let $n_i = |N_{i+1}/N_i| = [F_i : F_{i+1}]$. Since $[F_i : F_{i+1}] \mid [K : F] = |G(K/F)|$, we know that $n_i \mid n$, hence $t^{n_i} - 1 \mid t^n - 1$ in $F[t]$ (why?). Thus $t^{n_i} - 1$ in $F[t] \subset F_{i+1}[t]$ splits over F_i for each i. It follows that K/F is radical and Galois by Theorem 57.20.

Case 2. $t^n - 1$ in $F[t]$ does not split over F:

Let E be a splitting field of $t^n - 1$ over K and ζ a primitive nth root of unity in E.

Claim: $K(\zeta)/F$ is a radical and Galois extension (and, therefore, we are done).

We have already seen that $K(\zeta)/F$ is Galois as K/F is the splitting field of a separable polynomial over F. So we need only show that $K(\zeta)/F$ is a radical extension. As $K(\zeta)/F(\zeta)$ is Galois and $F(\zeta)/F$ is radical and Galois, it suffices to show that $K(\zeta)/F(\zeta)$ is radical.

We reduce to Case 1 by showing:

(*i*) $[K(\zeta) : F(\zeta)] = |G\big(K/F(\zeta)\big)| \mid n$.

(*ii*) $G\big(K(\zeta)/F(\zeta)\big)$ is a solvable group.

Indeed, suppose we have shown (i) and (ii). Then $F(\zeta)$ contains a primitive $|G\big(K(\zeta)/F(\zeta)\big)|$th primitive root of unity, so by Case 1, the extension $K(\zeta)/F(\zeta)$ is Galois and radical. As $F(\zeta)/F$ is radical, it then follows that $K(\zeta)/F$ is radical. So we need only show (i) and (ii). Let $\sigma \in G\big(K(\zeta)/F(\zeta)\big)$, then σ fixes F (and ζ), hence $\sigma|_K : K \to K(\zeta)$ is an F-homomorphism. By Artin's Lemma 51.8, there exist at most $[K : F] = |G(K/F)|$ F-homomorphisms $K \to K(\zeta)$. As each element of $G(K/F)$ is such an F-homomorphism, we must have $\sigma|_K$ lies in $G(K/F)$. Therefore,

$$\psi : G\big(K(\zeta)/F(\zeta)\big) \to G(K/F) \text{ given by the restriction } \sigma \mapsto \sigma|_K$$

is a well-defined group homomorphism. But ψ is monic, as $\sigma|_K = 1_K$ means that σ fixes K and ζ hence $K(\zeta)$, i.e., $\sigma = 1_{K(\zeta)}$. Therefore,

$$|G\big(K(\zeta)/f(\zeta)\big)| \mid |G(K/F)| = n$$

and $G(K\big(\zeta)/F(\zeta)\big)$ is isomorphic to a subgroup of the solvable group $G(K/F)$, so also solvable. □

Corollary 57.22. *Let F be a field of characteristic zero and f a non-constant polynomial in $F[t]$. Then f is solvable by radicals if and only if the Galois group of f is solvable.*

Proof. ($\Rightarrow$) has already been done.

($\Leftarrow$): Let K be a splitting field of f over F. By hypothesis, $G(K/F)$ is solvable, so by the theorem there exists an extension L/K with L/F Galois and radical. Since f splits over L, it is solvable by radicals. □

Corollary 57.23. *Let F be a field of characteristic zero and f a non-constant polynomial in $F[t]$ of degree at most four. Then f is solvable by radicals.*

Proof. The Galois group of f is isomorphic to a subgroup of S_4, a solvable group. □

Exercises 57.24. 1. Let L/F be a field extension and $L/K_i/F$ with K_i/F abelian. Show that $K_1K_2 = K_1(K_2)$ is abelian.

2. Show that Lemma 57.10 is false if p is not a prime.

3. Let K/F be a finite separable extension of fields. Show that $\mathrm{N}_{K/F} = \mathrm{N}_{E/F} \circ \mathrm{N}_{K/E}$ and $\mathrm{Tr}_{K/F} = \mathrm{Tr}_{E/F} \circ \mathrm{Tr}_{K/E}$ for any intermediate field $K/E/F$.

4. Let K/F be a finite separable extension of fields, x an element of K, and $\lambda_x : F(x) \to F(x)$ the F-linear transformation given by $y \mapsto xy$. Show that the characteristic polynomial of λ_x is $\big(m_F(x)\big)^{[K:F(x)]}$. In

particular, if $m_F(x) = t^r + a_{r-1}t^{r-1} + \cdots + a_0$ in $F[t]$, then $\mathrm{N}_{K/F}(x) = (-1)^r a_0^{[K:F(x)]}$ and $\mathrm{Tr}_{K/F}(x) = -[K : F(x)]a_{r-1}$

5. Suppose that K/F is a finite separable extension of fields, $K^* := \mathrm{Hom}_F(K, F)$, the F-linear dual space of K. Show that the map $T : K \to K^*$ defined by $x \mapsto \mathrm{tr}_x : y \mapsto \mathrm{Tr}_{K/F}(xy)$ is an F-linear isomorphism. In particular, if $\{x_1, \ldots, x_n\}$ is an F-basis for K, then there exists a basis $\{y_1, \ldots, y_n\}$ for K satisfying $\mathrm{Tr}_{K/F}(x_i y_j) = \delta_{ij}$ (the Kronecker delta).
6. Let K/F be a finite separable extension of fields, x an element of K, and $\lambda_x : F(x) \to F(x)$ the F-linear transformation given by $y \mapsto xy$. If $\mathcal{B}$ is an F-basis for K, show that $\mathrm{Tr}_{K/F}(x) = \mathrm{trace}[\lambda_x]_{\mathcal{B}}$ and $\mathrm{N}_{K/F} = \det[\lambda_x]_{\mathcal{B}}$.
7. Let K/F be a Galois extension of fields of degree n. Suppose that K/F is cyclic with $G(K/F) = \langle\sigma\rangle$ and $x \in K$. Show $\mathrm{Tr}_{K/F}(x) = 0$ if and only if there exists a $y \in K^\times$ satisfying $x = y - \sigma(y)$.
8. Let F be a field of characteristic zero and K/F an abelian extension. Let n be a positive integer such that $\sigma^n = 1$ for every $\sigma \in G(K/F)$ and F contains a primitive nth root of unity. Show that there exist $\alpha_1, \ldots, \alpha_r$ in K such that $K = F(\alpha_1, \ldots, \alpha_r)$ with $\alpha_i^{n_i} \in F$ for some $n_i \mid n$ for $1 = 1, \ldots, r$.
9. Let G be a multiplicative group and A a (multiplicative) abelian group with $G \to \mathrm{Aut}(A)$ a group homomorphism. This defines a G-action on A by evaluation, that we write by $\sigma(a)$ for $\sigma \in G$ and $a \in A$. We call a map $f : G \to A$ a *crossed homomorphism*, if $f(\sigma\tau) = f(\sigma)\,\sigma f(\tau)$. (So if $G \to \mathrm{Aut}(A)$ is the trivial map, this is just a group homomorphism.) Show all of the following:
   (*i*) In the above, the set $Z(G, A)$ of all crossed homomorphisms $f : G \to A$ is a group.
   (*ii*) In the above, if $a \in A$, then $f : G \to A$ given by $f(\sigma) = a^{-1}\sigma(a)$ is a crossed homomorphism, called a *principal crossed homomorphism*, and the set $B(G, A)$ of all such is a subgroup of $Z(G, A)$. Set $H^1(G, A) := Z(G, A)/B(G, A)$.
   (*iii*) Let L/F be a finite Galois extension and suppose that the group μ_n of nth roots of unity lies in F and consists of n elements. Then $B(G(L/F), \mu_n) = 1$ and $Z(G(L/F), \mu_n) = \mathrm{Hom}(G(L/F), \mu_n)$, i.e., $H^1(G(L/F), L^\times) = \mathrm{Hom}(G(L/F), \mu_n)$.
   (*iv*) Let L/F be a finite Galois extension and suppose that the group of nth roots of unity lies in F and consists of n elements. Then $Z(G(L/F), L^\times) = B(G(L/F), L^\times)$, i.e., $H^1(G(L/F), L^\times) = 1$. (This is Noether's generalization of Hilbert Theorem 90.)

58. Addendum: Kummer Theory

In Section §57, we proved Theorem 57.20 classifying finite cyclic field extensions, i.e., finite Galois extensions with cyclic Galois group, of degree n when the base field contained all the nth roots of unity and the characteristic of the field was zero or did not divide n. Exercise 57.24(8) extends this to finite abelian extension of a field. In this section, we generalize this to arbitrary abelian extensions of a field, i.e., not necessarily finite. To do this, we shall need results from Section 55.

We begin with a preliminary discussion of pairing of groups. Let A, B, and C be abelian multiplicative groups. A *pairing* $\langle\ ,\ \rangle : A\times B \to C$ is a map that is a group homomorphism in each variable. Set

$$A^\perp := \{a \in A \mid \langle a, x\rangle = e_C \ \text{ for all } x \in B\}$$
$$B^\perp := \{b \in B \mid \langle y, b\rangle = e_C \ \text{ for all } y \in A\}.$$

This pairing induces a homomorphism $A \to \mathrm{Hom}(B, C)$ by $a \mapsto \langle a, \ \ \rangle$, hence a monomorphism $A/A^\perp \to \mathrm{Hom}(B/B^\perp, C)$. Similarly, this pairing induces a homomorphism $B \to \mathrm{Hom}(A, C)$ by $b \mapsto \langle\ \ , b\rangle$, hence a monomorphism $B/B^\perp \to \mathrm{Hom}(A/A^\perp, C)$.

Suppose m is a positive integer such that both $A/A^\perp$ and $B/B^\perp$ satisfy $x^m = e_A$ and $y^m = e_B$ for all $x \in A$ and for all $y \in B$, and $C = \mu_m$ is the cyclic group of order m, i.e., m divides the exponent of these three groups, where the *exponent* N of a (multiplicative) group G is the least positive integer N such that $g^N = e_G$ for all $g \in G$ (if such exists). Then in the above,

$$A/A^\perp \ \text{ is finite if and only if } \ B/B^\perp \ \text{ is finite,} \tag{58.1}$$

and if this is the case, then we have an isomorphism:

$$A/A^\perp \xrightarrow{\sim} \widehat{B/B^\perp} := \mathrm{Hom}(B/B^\perp, \mu_m). \tag{58.2}$$

We shall apply this when μ_m is the group of mth roots of unity in a field F of characteristic zero or characteristic not dividing m.

We set up notation to be used in the rest of this section. Let F be a field. Denote by $\widetilde{F}$ a fixed algebraic closure of F and assume all algebraic extensions of F lie in it. So the group of mth roots of unity μ_m lie $\widetilde{F}$. We shall assume that $|\mu_m| = m$, so the characteristic of F is zero or does not divide m. We shall also assume that $\mu_m \subset F$. Therefore, if K/F is a splitting field of $t^m - a \in F[t]$, then $K = F(\alpha)$ with $\alpha^m = a$ and the m distinct roots of $t^m - a$ are given by $\zeta\alpha$ for $\zeta \in \mu_m$ by Theorem 57.20. We write this symbolically as $K = F(a^{\frac{1}{m}})$. Let $(F^\times)^m = \{x^m \mid x \in F\}$ and if $(F^\times)^m \subset P \subset F^\times$ is a subgroup, let

$F_P = F(P^{\frac{1}{m}})$ be the field generated by the $F(a^{\frac{1}{m}})$ with $a \in P$. The (possibly infinite) field extension F_P/F is normal and separable, hence is Galois by Theorem 55.3.

Definition 58.3. Let K/F be a (possibly infinite) Galois extension of fields. We say that K/F has *exponent* m if $G(K/F)$ has finite exponent dividing m. Suppose, in addition, that $\mu_m \subset F$ and $|\mu_m| = m$. We say that K/F is a *Kummer extension of exponent* m if K/F is abelian of exponent m, i.e., Galois with abelian Galois group.

Let F be a field containing the mth roots of unity μ_m with $|\mu_m| = m$. Set

$$\mathcal{F}_m := \{K \mid K/F \text{ is a Kummer extension of exponent } m\}$$
$$\mathcal{G}_m^{ab} := \{P \mid (F^\times)^m \subset P \subset F^\times \text{ is a subgroup}\}.$$

Theorem 58.4. *Let F be a field containing the mth roots of unity μ_m with $|\mu_m| = m$. Then the map $i : \mathcal{G}_m \to \mathcal{F}_m$ given by $P \mapsto F_P$ is a bijection. Moreover, F_P/F is a finite extension if and only if $P/(F^\times)^m$ is a finite group, and if finite, then $[F : F_P] = [P : (F^\times)^m]$.*

Proof. Let $P \in \mathcal{G}_m$. We first show that $F_P \in \mathcal{F}_m$. Let $a \in P$, so $a = \alpha^m$, some $\alpha \in \widetilde{F}$. In the proof of Theorem 57.20 using Hilbert Theorem 90 (Lemma 57.19), we saw that if $\sigma \in G(F_P/F)$, then there exists a unique $\zeta_\sigma \in \mu_m$ satisfying $\sigma(\alpha) = \zeta_\sigma \alpha$. Check that ζ_σ is independent of the choice of α. In particular, $\sigma^m(\alpha) = \alpha$ and if $\tau \in G(F_P/F)$, then

$$\sigma\tau(\alpha) = \zeta_\sigma \zeta_\tau = \tau\sigma(\alpha).$$

It follows that F_P/F is an abelian extension. Since $F_P = F(\{\alpha \in \widetilde{F} \mid \alpha^m \in P\}$, we have $F_P \in \mathcal{F}_m$.

Next let $K \in \mathcal{F}_m$. Set

$$N := \{a \in F^\times \mid \text{there exists } \ \alpha \in K \text{ such that } \alpha^m = a\} \in \mathcal{G}_m.$$

Let $G = G(K/F)$. Define

$$f : N \to \widehat{G} := \mathrm{Hom}(G, \mu_m)$$

as follows: Let $a \in N$. Choose $\alpha \in K$ satisfying $\alpha^m = a$. If $\sigma \in G$, then, as above, there exists a unique $\zeta_\sigma \in \mu_m$ satisfying $\sigma(\alpha) = \zeta_\sigma \alpha$. Define $f_a : N \to \widehat{G} = \mathrm{Hom}(G, \mu_m)$ by $\sigma \mapsto \sigma(\alpha)/\alpha = \zeta_\sigma$.

Claim 58.5. *f_a is a well-defined homomorphism.*

If f_a is well-defined, it is clearly a homomorphism. So we must show that f_a is independent of the choice of α. We leave its proof as an exercise. By the claim we obtain a map $f : N \to \widehat{G}$ given by $a \mapsto f_a$.

We apply this construction to $K = F_P$ with $P \in \mathcal{G}_m$. So we have a pairing

$$\langle\ ,\ \rangle : G \times P \to \mu_m$$

defined by $\langle \sigma, a\rangle = f_a(\sigma) = \sigma(\alpha)/\alpha$ where $a = \alpha^m$.

Claim 58.6. *$G^\perp = 1$ and $P^\perp = (F^\times)^m$.*

That $G^\perp = 1$ is easy, so we turn to $P^\perp$. Suppose that $\alpha \in K = F_P$ satisfies $\alpha^m = \alpha$ and $\langle \sigma, a\rangle = \sigma(\alpha)/\alpha = 1$ for all $\sigma \in G$, but $\alpha \notin F$. Then there exists a $\tau \in G$ such that $\tau(\alpha) \neq \alpha$. Hence $\langle \tau, a\rangle = \tau(\alpha)/\alpha \neq 1$, a contradiction. The Claim now follows easily.

We also conclude by (58.1) that G is finite, i.e., F_P/F is finite, if and only if $P/(F^\times)^m$ is finite. If this is the case then by (58.2), we have

$$[F : F_P] = [P : (F^\times)^m] \quad \text{and} \quad P/(F^\times)^m/(F^\times)^m \cong \widehat{G(F_P/F)}. \tag{58.7}$$

This is the crucial observation.

Claim 58.8. *$P_1 \subset P_2$ in $\mathcal{G}_m$ if and only if $F_{P_1} \subset F_{P_2}$. In particular, $i : \mathcal{G}_m \to \mathcal{F}_m$ is injective.*

The only if statement is clear, so suppose that $F_{P_1} \subset F_{P_2}$. Let $\alpha \in P_1$. Then α lies in F_{P_2}, hence there exists a finitely generated extension $K/F(F^m)$ with F_{P_2}/K and $\alpha \in K$. It follows that we may assume that $F_{P_2}/F(F^m)$ is finite generated, hence a finite field extension. Let P_3 be the group generated by α and P_2. Then $F(P_2) = F(P_3)$, and $[F_{P_2} : F^m] = [F_{P_3} : F^m]$ by (58.7). Hence $P_2 = P_3$. This proves the claim.

So we are left to show that $i : \mathcal{G}_m \to \mathcal{F}_m$ is surjective. Let $K \in \mathcal{F}_m$, say $K = F(P)$ with $P \in \mathcal{G}_m$. For any element x in K, there exists a finite subgroup $P_0 \in \mathcal{G}_m$ such that x lies in $F(P_0)$, i.e., K is the compositum of finite abelian extensions. Therefore, we may assume that K/F is finite abelian extension with $G = G(K/F)$ of finite exponent dividing m. Since G is a finite abelian group of finite exponent dviding m, it is a product of finitely many cyclic groups of exponent dividing m by Theorem 41.6 (or Proposition 14.9). Consequently, $G \cong \widehat{G}$ by Exercise 41.24(11). By Exercise 54.13(13), this reduces us to the case that K/F is finite cyclic which has been done. □

Remark 58.9. Let F contain μ_m with $|\mu_m| = m$ as in the theorem. In the notation of the theorem, if we let $P \in \mathcal{G}_m$ and $G = G(F_P/F)$, then the map

$$f : P \to \widehat{G} \ \text{ given by } \ a \mapsto (f_a : \sigma \to \sigma(\alpha)/\alpha) \ \text{ with } \ \alpha^m = a \tag{58.10}$$

induces a monomorphism $\overline{f} : F/(F^\times)^m \to \widehat{G}$. If we take continuous homomorphisms $G \to \mu_m$ in the Krull topology on $G(\widetilde{F}/F)$, then this map is an isomorphism. [As μ_m lies in F, its topology is discrete.] We leave this as an exercise. Write this as

$$1 \to (F^\times)^m \to P \to \mathrm{Hom}_{\mathrm{cont}}(G, \mu_m) \to 1 \tag{58.11}$$

is exact where $\mathrm{Hom}_{\mathrm{cont}}(G, \mu_m)$ denotes the group of continuous maps $G \to \mu_m$.

More generally, let F_{sep} denote the *separable closure of* F in $\widetilde{F}$, i.e., the maximal separable extension of F in $\widetilde{F}$, then $\Gamma_F := G(\mathrm{F}_{sep}/F)$ is $G(\widetilde{F}/F)$ and called the *absolute Galois group of* F. If $\mu_m \subset F$ and has m elements, then we have an exact sequence

$$1 \to \mu_m \to F^\times_{\mathrm{sep}} \xrightarrow{g} F^\times_{\mathrm{sep}} \to 1.$$

where $g(x) = x^m$ By using 'group cohomology' — a subject that we shall not discuss — one obtains an exact sequence

$$1 \to \mu_m \to F^\times \xrightarrow{g} (F^\times) \xrightarrow{\delta} \mathrm{Hom}_{\mathrm{cont}}(\Gamma_F, \mu_m) \to 1.$$

called the *Kummer sequence* with δ a map coming from group cohomology theory called a connecting map. The surjectivity of δ arises from a homological version of Hilbert Theorem 90, the finite case can be found in Exercise 57.24(9).

Exercise 58.12. 1. Prove 58.1
2. Prove that ζ_σ in the proof of the theorem is independent of the choice of α.
3. Prove Claim 58.5.
4. Prove that (58.10) implies (58.11).

59. Addendum: Galois' Theorem

In this section, we establish Galois' characterization of the Galois group of an irreducible polynomial of prime degree over a field of characteristic zero. We begin with a lemma, previously given as Exercise 21.11 (13). If p is a prime, then a finite group is called an *elementary p-group* if it is isomorphic to $(\mathbb{Z}/p\mathbb{Z})^n$ for positive integer n. Note that any elementary p-group can be viewed as a finite dimensional vector space over $\mathbb{Z}/p\mathbb{Z}$. A *minimal normal* subgroup of a nontrivial group G is a normal subgroup of G such that the trivial group is the only normal subgroup of G properly contained in it.

Lemma 59.1. *Let G be a nontrivial finite solvable group. Then any minimal normal subgroup of G is an elementary p-group.*

PROOF. Let H be a minimal normal subgroup of G. As H is a subgroup of a solvable group, it is solvable. We also know that any characteristic subgroup of H is normal in G (cf. Exercise 11.9(18)). Since the commutator subgroup of any group is a characteristic subgroup, the series obtained by successively taking commutators gives a characteristic series for H (cf. Proposition 16.5). As H is solvable, it contains a characteristic nontrivial abelian normal subgroup, so we may assume H is abelian. Since Sylow groups of an abelian group are characteristic, we may assume that H is a p-group. Since $\{x \in H \mid x^p = 1\}$ is easily seen to be a characteristic subgroup of H, the lemma follows. □

More generally, it can be shown that a minimal normal subgroup of a nontrivial finite group is either simple or a product of simple groups all which are isomorphic.

Let p be a prime and V a finite dimensional vector space over $\mathbb{Z}/p\mathbb{Z}$. If v_0 is a vector in V, then the map $\tau_{v_0} : V \to V$ given by $v \mapsto v + v_0$ is called *translation* by v_0. (Note that it is not an F-linear transformation unless $v_0 = 0$.) The set of all translations of V form a subgroup of the the *group of affine transformations* of V, the group defined by

$$\mathrm{Aff}(V) := \{\tau : V \to V \mid \tau : v \mapsto av + v_0 \text{ with } a \text{ in } \mathbb{Z}/p\mathbb{Z} \text{ and } v_0 \text{ in } V\}.$$

So every element in $\mathrm{Aff}(V)$ is a composition $\tau\lambda_a$ with τ a translation and $\lambda_a : V \to V$ the F-linear map given by $x \mapsto ax$ for some a in F.

Theorem 59.2. *Let p be a prime and S a set with p elements. Suppose that G is a transitive subgroup of $\Sigma(S) \cong S_p$. Then the following are equivalent:*

(1) G is solvable.
(2) Each non-identity element in G fixes at most one element in S.
(3) There exists a transitive normal subgroup T of G satisfying $|T| = p$ and T is its own centralizer in G, i.e., $T = Z_G(T)$.
(4) G is isomorphic to a subgroup of the group of affine transformations $\mathrm{Aff}(\mathbb{Z}/p\mathbb{Z})$ and contains the subgroup of translations.

Moreover, if G is solvable and if T is as in (3), then there exists a cyclic subgroup H in G satisfying $G = HT$, $H \cap T = 1$, and $|H| \mid p - 1$. Furthermore, any non-identity normal subgroup of G is of the form $H'T$ for some subgroup H' of H.

PROOF. We may assume that $G \subset S_p$, so $S = \{1, \ldots, p\}$. As G is transitive, the orbit Gs for any $s \in S$, has order p, so $p \mid |G|$ and all

the Sylow p-groups of G are cyclic of order p, i.e., is generated by a p-cycle.

(1) $\Rightarrow$ (3): Let N be a normal subgroup of G.

Claim. The group N acts transitively on S_p. In particular, $p \mid |N|$:

Let x and y be elements in S. As G acts transitively on S, there exists an element σ in G satisfying $y = \sigma(x)$. As the map $Nx \to Ny$ given by $zx \mapsto z\sigma(x)$ is a well-defined bijection, we see that all orbits of N have the same number of elements. If $\mathcal{O}$ is a system of representatives for the action of N on S, then $p = |S| = \sum_{\mathcal{O}} |Nx| = \sum_{\mathcal{O}} [N : N_x]$, by the Orbit Decomposition Theorem 18.9. So either S has one orbit under the action of N, i.e., N is transitive, or every orbit of S under the action of N has one point, i.e., N is the identity subgroup. This establishes the claim.

Let T be a minimal normal subgroup of G. By the lemma and claim, T is a cyclic subgroup of order p, hence is generated by a p-cycle, say $T = \langle\sigma\rangle$. The normality of T implies that it is the unique Sylow p-subgroup of G. Suppose τ centralizes T, i.e., lies in $Z_G(T)$. If $\tau(s) = s$ for some $s \in S$, then $\tau\sigma(s) = \sigma\tau(s) = \sigma(s)$, hence $\tau\sigma^i(s) = \sigma^i\tau(s) = \sigma^i(s)$ for all i implying $\tau = 1$. Therefore τ is either 1 or is fixed point-free, so a p-cycle.

(3) $\Rightarrow$ (4): Suppose that $T = \langle\sigma\rangle$ is a cyclic subgroup of G of order p. We may assume that $S = \mathbb{Z}/p\mathbb{Z}$ and σ is the p-cycle $(\overline{0}\,\overline{1}\cdots\overline{p-1})$ where $^- : \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ is the canonical epimorphism. Therefore, σ generates the group of translations in $\mathrm{Aff}(\mathbb{Z}/p\mathbb{Z})$. Let τ be an element in G. Then there exists an integer n with $p \nmid n$, unique modulo p, such $\tau\sigma\tau^{-1} = \sigma^n$, i.e., we have a well-defined map $\varphi : G \to (\mathbb{Z}/p\mathbb{Z})^\times$ given by $\tau \mapsto n$. Clearly, this map is a group homomorphism with $\ker\varphi = Z_G(T)$. As $\tau\sigma = \sigma^{\varphi(\tau)}\tau$, we have

$$\tau(x + \overline{1}) = \tau\sigma(x) = \sigma^{\varphi(\tau)}\tau(x) = \tau(x) + \varphi(\tau).$$

In particular, $\tau(\overline{1}) = \tau(\overline{0}) + \varphi(\tau) = \varphi(\tau) + b$ with $b = \tau(\overline{0})$. Inductively, $\tau(x) = \varphi(\tau)x + b$. This yields (4).

(4) $\Rightarrow$ (1): The translations form a cyclic normal subgroup T in G with quotient embedding in $(\mathbb{Z}/p\mathbb{Z})^\times$.

(4) $\Rightarrow$ (2): The equation

$$ax + b \equiv x \mod p,$$

with a and b integers, has a unique solution modulo p if either $a \not\equiv 1 \mod p$ or $b \not\equiv 0 \mod p$. It follows that any non-identity element in G can fix at most one point.

(2) $\Rightarrow$ (3): We apply Exercise 20.26(16) — known as the *Burnside Counting Theorem* – to see that

$$|G| = \sum_G |F_\tau(S)| \text{ where } F_\tau(S) := F_{\langle\tau\rangle}(S).$$

As $\tau = 1$ fixes p points, we have $\sum_{G\setminus\{1\}} |F_\tau(S)| = |G| - p$. By hypothesis, if $\tau \neq 1$ then $|F_\tau(S)| \leq 1$, so we must have precisely $p-1$ non-identity elements τ in G satisfying $|F_\tau(S)| = 0$, i.e., $p-1$ elements in G have no fixed points in S. If σ is one such, then so is σ^i for $i = 1, \ldots, p-1$. It follows that σ is a p-cycle and $T = \langle\sigma\rangle$, a group of order p, contains all the p-cycles in G. As the conjugate of a p-cycle is a p-cycle, $T \lhd G$. Suppose that τ lies in $Z_G(T)$. If $F_\tau(S)$ is empty, then τ lies in T, so assume that $F_\tau(S)$ is not empty, say $\tau(x) = x$. It follows that $\tau\sigma(x) = \sigma\tau(x) = \sigma(x)$, i.e., both x and $\sigma(x)$ lie in $F_\tau(S)$. By assumption this means $\tau = 1$, so $T = Z_G(T)$, establishing (3).

For the last two statements, let G satisfy (4). Set

$$H = \{\tau \in G \mid \tau(x) \equiv ax \mod p \text{ with } a \text{ an integer satisfying } p \nmid a\}.$$

Clearly, $H \cap T = 1$ and $HT = G$. Certainly, we have a monomorphism $H \to (\mathbb{Z}/p\mathbb{Z})^\times$, so H is cyclic and $|H| \mid p-1$. If $N \lhd G$ is not the trivial group, then by the claim, N acts transitively on S and contains T, the unique Sylow p-subgroup of G. It follows that $N = H'T$ for some subgroup H' of H by the Correspondence Principle. □

Theorem 59.3. (Galois) *Suppose that F is a field of characteristic zero and f an irreducible polynomial in $F[t]$ of prime degree p. Let K be a splitting field of f over F. Then f is solvable by radicals if and only if $K = F(\alpha, \beta)$ for any two roots of f in K. Moreover, if f is solvable by radicals, $G(K/F)$ is isomorphic to a subgroup of $\mathrm{Aff}(\mathbb{Z}/p\mathbb{Z})$ containing the subgroup of translations, and there exists an intermediate field $K/E/F$ with E/F cyclic, $[K : E] = p$. Furthermore, if $K/L/F$ is an intermediate field with L/F normal with $F < L$, then $L \subset E$.*

Proof. We know that $G(K/F)$ can be viewed as a transitive subgroup of S_p and f is solvable by radicals if and only if $G(K/F)$ is solvable. By the previous theorem, $G(K/F)$ is solvable if and only if no non-identity element in $G(K/F)$ fixes two roots of f. Let α and β be two roots of f in K. By the Fundamental Theorem of Galois Theory 54.3, we know that $K = F(\alpha, \beta)$ if and only if $G\big(K/F(\alpha, \beta)\big) = 1$ if and only if no non-identity element of $G(K/F)$ fixes α and β. The result now easily follows. □

60. The Discriminant of a Polynomial

Let R be a commutative ring and let the symmetric group S_n act on $R[t_1, \ldots, t_n]$ by $\sigma t_i = t_{\sigma(i)}$ for all i and $\sigma \in S_n$. This induces a group monomorphism $S_n \to \mathrm{Aut}(R[t_1, \ldots, t_n])$. Analogous to the field case of $F(t_1, \ldots, t_n)^{S_n}$, one can show that the set of elements $R[t_1, \ldots, t_n]^{S_n}$ in $R[t_1, \ldots, t_n]$ fixed by S_n, is equal to $R[s_1(t_1, \ldots, t_n), \ldots, s_n(t_1, \ldots, t_n)]$, where

$$s_j(t_1, \ldots, t_n) = \sum_{1 \le i_1 < \cdots < i_j \le n} t_{i_1} \cdots t_{i_j}$$

with $1 \le j \le n$ are the elementary symmetric functions in $t_1, \ldots, t_n$. (Cf. the Fundamental Theorem of Symmetric Functions 66.4.) Let $D = \prod_{i<j}(t_i - t_j)$ in $\mathbb{Z}[t_1, \ldots, t_n]$. If σ is the transposition (ij) in S_n, then $\sigma D = -D$. It follows if σ lies in S_n, then

$$\sigma(D) = \begin{cases} -D & \text{if } \sigma \notin A_n \\ D & \text{if } \sigma \in A_n. \end{cases}$$

Let F be a field of characteristic different from two, f a polynomial in $F[t]$, and K/F a splitting field of f. Suppose that $\deg f = n$ and f has no multiple roots in K, say the roots are $\alpha_1, \ldots, \alpha_n$. Hence K/F is Galois and the Galois group $G(K/F)$ permutes $\alpha_1, \ldots, \alpha_n$, so defines a monomorphism $G(K/F) \to S_n$. If σ lies in $G(K/F)$, we have a commutative diagram:

$$\begin{array}{ccc} F[t_1, \ldots, t_n] & \xrightarrow{\sigma} & F[t_1, \ldots, t_n] \\ {\scriptstyle e_{\alpha_1,\ldots,\alpha_n}}\downarrow & & \downarrow{\scriptstyle e_{\alpha_1,\ldots,\alpha_n}} \\ K & \xrightarrow[\sigma]{} & K, \end{array}$$

where $e_{\alpha_1,\ldots,\alpha_n}$ is the obvious evaluation map. It follows that

$$d := e_{\alpha_1,\ldots,\alpha_n}(D) = \prod_{i<j}(\alpha_i - \alpha_j) \neq 0$$

satisfies

$$\sigma(d) = \begin{cases} -d & \text{if } \sigma \notin A_n \\ d & \text{if } \sigma \in A_n. \end{cases}$$

Therefore, $G\big(K/F(d)\big) = G(K/F) \cap A_n$. Let

$$\Delta := d^2 = \prod_{i<j}(\alpha_i - \alpha_j)^2.$$

Then $\sigma(\Delta) = \Delta$ for every σ in $G(K/F)$, hence $d^2 = \Delta$ lies in $K^{G(K/F)} = F$ and is called the *discriminant* of f. As

$$[G(K/F) : G(K/F(d)] \leq [S_n : A_n] = 2,$$

either $F = F(d)$ or $F(d)/F$ is of degree two, as $d \in F$ if and only if $\Delta \in F^2$ if and only if $G(K/F)$ is a subgroup of A_n (where we view $G(K/F) \to S_n$ as an inclusion).

Computation 60.1. In the setup above, let $s_i = s_i(\alpha_1, \ldots, \alpha_n)$. As S_n fixes each $s_j(t_1, \ldots, t_n)$, we have each s_i lies in $K^{G(K/F)} = F$. By linear algebra, the *Vandermonde determinant*

$$\det \begin{pmatrix} 1 & \cdots & 1 \\ \alpha_1 & \cdots & \alpha_n \\ \vdots & & \vdots \\ \alpha_1^{n-1} & \cdots & \alpha_n^{n-1} \end{pmatrix} = \prod_{i>j} (\alpha_i - \alpha_j) = d, \tag{60.2}$$

so if

$$A = \begin{pmatrix} 1 & \cdots & 1 \\ \alpha_1 & \cdots & \alpha_n \\ \vdots & & \vdots \\ \alpha_1^{n-1} & \cdots & \alpha_n^{n-1} \end{pmatrix}, \tag{60.3}$$

is the *Vandermonde matrix* , we have $\det AA^t = d^2 = \Delta$.

Check 60.4. Let $e_i := \sum_{j=1}^n \alpha_j^i$ for $i = 1, \ldots, 2n-2$. Then

$$AA^t = \begin{pmatrix} n & e_1 & e_2 & \cdots & e_{n-1} \\ e_1 & e_2 & e_3 & \cdots & e_n \\ \vdots & & & & \vdots \\ e_{n-1} & e_n & e_{n+1} & \cdots & e_{2n-2} \end{pmatrix}.$$

As $\sum_{j=1}^n t_j^i$ lies in $F[t_1, \ldots, t_n]^{S_n}$, for every positive integer i, each $\sum_{j=1}^n t_j^i$ is a polynomial in the $s_i(t_1, \ldots, t_n)$'s (in fact, with integer coefficients), hence each e_i lies in $F[s_1, \ldots, s_n]$. We can, therefore, compute formulas for Δ for specific n.

Example 60.5. In the computation and the notation there, we let $n = 2$. Then

$$f = (t - \alpha_1)(t - \alpha_2) = t^2 - (\alpha_1 + \alpha_2)t + \alpha_1\alpha_2 = t^2 - s_1 t + s_2.$$

So $s_1 = \alpha_1 + \alpha_2$ and $s_2 = \alpha_1\alpha_2$, hence $e_1 = s_1 = \alpha_1 + \alpha_2$ and $e_2 = \alpha_1^2 + \alpha_2^2 = (\alpha_1 + \alpha_2)^2 - 2\alpha_1\alpha_2 = s_1^2 - 2s_2$. It follows that

$$\Delta = \det \begin{pmatrix} 2 & e_1 \\ e_1 & e_2 \end{pmatrix} = 2e_2 - e_1^2 = 2(s_1^2 - 2s_2) - s_1^2 = s_1^2 - 4s_2.$$

In general, if f is a polynomial of degree n and splits over K with the roots $\alpha_1, \ldots, \alpha_n$ (possibly not distinct), we define the discriminant of f to be $\Delta = \prod_{i<j}(\alpha_i - \alpha_j)^2$.

If one knows a root of an irreducible polynomial, then we can express the discriminant of it in another way.

Proposition 60.6. *Let f be an irreducible and separable polynomial in $F[t]$ of degree n, L/F a splitting field of f, $\alpha = \alpha_1, \alpha_2, \ldots, \alpha_n$ the (distinct) roots of f in L, $K = F(\alpha)$, and $T_\alpha : L \to L$ the F-linear isomorphism given by $x \mapsto \alpha x$. Then the following are true:*

1. *$\alpha_1, \ldots, \alpha_n$ are the eigenvalues of T_α. In particular, T_α is diagonalizable.*
2. *$\alpha_1^k, \ldots, \alpha_n^k$ are the eigenvalues of T_{α^k} for all positive integers k and T_{α^k} is diagonalizable.*
3. *If $f = t^n + a_{n-1}t^{n-1} + \cdots + a_0$ in $F[t]$, then $f_{T_\alpha} = f^{[L:K]}$, where f_{T_α} is the characteristic polynomial of T_α. In particular,*
   (a) $\mathrm{Tr}_{K/F}(\alpha) = -a_{n-1}$.
   (b) $\mathrm{N}_{K/F}(\alpha) = (-1)^n a_0$.
4. *Let A be the Vandermonde matrix 60.3, then the (ij)th term of AA^t satisfies*

$$(AA^t)_{ij} = \mathrm{Tr}_{K/F}(\alpha^{i-1}\alpha^{j-1}).$$

   In particular, $\det\big(\mathrm{Tr}_{K/F}(\alpha^{i-1}\alpha^{j-1})\big) = \Delta(\alpha)$, *hence nonzero.*
5. *We have*

$$\Delta(\alpha) = (-1)^{\frac{n(n-1)}{2}}\,\mathrm{N}_{K/F}\big(f'(\alpha)\big).$$

Proof. (1) and (2) are left as new exercises and (3) is Exercise 57.24(3).

(4): Using (2), we have $\mathrm{Tr}_{K/F}(\alpha^k) = \alpha_1^k + \cdots + \alpha_n^k$ for all $k \geq 1$, hence

$$(AA^t)_{ij} = \sum_{l=1}^{n} \alpha_l^{i-1}\alpha_l^{j-1} = \sum_{l=1}^{n} \alpha^{i+j-2} = \mathrm{Tr}_{K/F}(\alpha^{i-1}\alpha^{j-1}).$$

(5): We have $\det A = \prod_{i>j}(\alpha_i - \alpha_j)$ and there exist $n(n-1)/2$ pairs of integers (i, j) with $1 \leq j < i \leq n$ in the product $\Delta(\alpha) = \prod_{i>j}(\alpha_i - \alpha_j)^2$. As $(\alpha_i - \alpha_j)^2 = -(\alpha_i - \alpha_j)(\alpha_j - \alpha_i)$, it follows that

$$\Delta(\alpha) = (-1)^{\frac{n(n-1)}{2}} \prod_{i=1}^{n} \prod_{\substack{j=1 \\ j\neq i}}^{n} (\alpha_i - \alpha_j).$$

Since $f = \prod_{i=1}^{n}(t-\alpha_i)$ in $L[t]$, we have

$$f' = \sum_{i=1}^{n}\prod_{\substack{j=1\\ j\neq i}}^{n}(t-\alpha_j), \quad \text{so} \quad f'(\alpha_i) = \prod_{\substack{j=1\\ j\neq i}}^{n}(\alpha_i-\alpha_j)$$

and

$$\Delta(\alpha) = (-1)^{\frac{n(n-1)}{2}}\prod_{i=1}^{n} f'(\alpha_i).$$

Let $\sigma_i : K \to L$ be the F-homomorphisms satisfying $\alpha \mapsto \alpha_i$ for $i = 1, \ldots, n$. Then $\sigma_i\big(f'(\alpha)\big) = f'(\alpha_i)$, hence

$$\Delta(\alpha) = (-1)^{\frac{n(n-1)}{2}}\,\mathrm{N}_{K/F}\left(f'(\alpha)\right)$$

as claimed. □

We use the above in the following computation:

Example 60.7. In Proposition 60.6, set $F = \mathbb{Q}$ and $L = K = \mathbb{Q}(\zeta)$, where ζ a primitive pth root of unity in $\mathbb{C}$ with p an odd prime. The pth cyclotomic polynomial $\Phi_p = t^{p-1} + \cdots + 1 = m_{\mathbb{Q}}(\zeta)$. So by Proposition 60.6(5), we have

$$\Delta(\zeta) = (-1)^{\frac{(p-1)(p-2)}{2}}\,\mathrm{N}_{K/F}\left(\Phi_p'(\zeta)\right) = (-1)^{\frac{p-1}{2}}\,\mathrm{N}_{K/F}\left(\Phi_p'(\zeta)\right).$$

We compute the right hand side of this equation. Since $t^p - 1 = (t-1)\Phi_p$, taking the derivative yields $pt^{p-1} = \Phi_p + (t-1)\Phi_p'$. Hence

$$\Phi_p'(\zeta) = \frac{p\zeta^{p-1}}{\zeta - 1}.$$

By Proposition 60.6(5),

$$\mathrm{N}_{\mathbb{Q}(\zeta)/\mathbb{Q}}\left(\Phi_p'(\zeta)\right) = \frac{\mathrm{N}_{\mathbb{Q}(\zeta)/\mathbb{Q}}(p\zeta^{-1})}{\mathrm{N}_{\mathbb{Q}(\zeta)/\mathbb{Q}}(\zeta-1)} = \frac{p^{p-1}\cdot 1}{\mathrm{N}_{\mathbb{Q}(\zeta)/\mathbb{Q}}(\zeta-1)}.$$

As $\prod_{i=1}^{p-1}(t-\zeta^i) = t^{p-1} + \cdots + t + 1$, we have

$$\prod_{i=1}^{p-1}(\zeta^i - 1) = (-1)^{p-1}\Phi_p(1) = (-1)^{p-1}p = p.$$

Consequently, $\mathrm{N}_{\mathbb{Q}(\zeta)/\mathbb{Q}}\left(\Phi_p'(\zeta)\right) = p^{p-2}$, hence

$$\Delta(\zeta) = (-1)^{\frac{p-1}{2}}p^{p-2}.$$

In particular, $\mathbb{Q}\big(\sqrt{\Delta(\zeta)}\big)$ is the unique quadratic extension of $\mathbb{Q}$ in $\mathbb{Q}(\zeta)$ by Proposition 56.14.

Note, since

$$\prod_{p-1\geq i>j\geq 1}(\zeta^i-\zeta^j)^2=\prod_{p-1\geq i>j\geq 1}\big((\zeta^i-1)-(\zeta^j-1)\big)^2,$$

we also have $\Delta(\zeta)=\Delta(1-\zeta)$.

The technique above can be generalized in order to compute $\Delta(\zeta)$ with ζ a primitive p^mth root of unity (and from it ζ a primitive nth root of unity). To show this, one uses:

Lemma 60.8. *Let ζ be a primitive nth root of unity in $\mathbb{C}$ and $K=\mathbb{Q}(\zeta)$. Then $\Delta(\zeta)\mid n^{\varphi(n)}$.*

Proof. Let $\Phi_n=m_{K/F}(\zeta)$, then $t^n-1=\Phi_n g$ for some $g\in\mathbb{Z}[t]$. Taking derivatives, we have $nt^{n-1}=\Phi_n g'+\Phi_n g$. Evaluating at $t=\zeta$ and taking norms give

$$n^{\varphi(n)}=\mathrm{N}_{K/\mathbb{Q}}(n\zeta^{n-1})=\mathrm{N}_{K/\mathbb{Q}}(\Phi_n')\,\mathrm{N}_{K/\mathbb{Q}}(g)=(-1)^{\frac{n(n-1)}{2}}\Delta(\zeta)\,\mathrm{N}_{K/\mathbb{Q}}(g)$$

in $\mathbb{Z}$. The result follows. □

Exercises 60.9. 1. Show that equation 60.2 is valid.

2. Prove Check 60.4.
3. Let $f=t^3-a_1t^2-a_2t-a_3\in\mathbb{R}[t]$. Show
   (a) The discriminant $\Delta=-4a_1^3a_3+a_1^2a_2^2-18a_1a_2a_3+4a_2^3-27a_3^2$.
   (b) f has multiple roots if and only if $\Delta=0$.
   (c) f has three distinct real roots if and only if $\Delta>0$.
   (d) f has one real root and two non-real roots if and only if $\Delta<0$.
4. Let F be a field and f be an irreducible separable cubic in $F[t]$. Show the Galois group of f is either A_3 or S_3 and if the characteristic of FE is not two, it is A_3 if and only if the discriminant of f is a square in F.
5. Let t^3+pt+q be irreducible over a finite field F of characteristic not 2 or 3. Show that $-4p^3-27q^2$ is a square in F.
6. Prove (1) and (2) of Proposition 60.6
7. Let ζ be a primitive p^mth root of unity in $\mathbb{C}$ and $K=\mathbb{Q}(\zeta)$. Show that $\Delta(\zeta)=\pm p^{m-1(mp-m-1)}$ with the plus sign occurring if and only if either $p\equiv 1 \mod 4$ or $a=2^a$ with $a>2$.

61. Purely Transcendental Extensions

Let K/F be a finitely generated extension of fields in which K is *purely transcendental* over F, i.e., $K\cong F(t_1,\ldots,t_n)$. If $K/E/F$ is an intermediate field, the question arises whether E/F is a purely

transcendental extension. In general, this has been shown to be false, but if $n = 1$ this is true and is the subject of this section.

Lemma 61.1. *Let F be a field and x an element transcendental over F. Suppose that u is an element in $F(x)$ not lying in F, say $u = f/g$ with f and g nonzero non-constant relatively prime polynomials in $F[t]$, and $r = \max\{\deg f, \deg g\}$. Then x is algebraic over $F(u)$ with $m_{F(u)}(X) = f - ug$ in $F(u)[t]$.*

Proof. Let $h = f - ug$ in $F[u][t] \subset F(u)[t]$ with f and g as in the lemma. Suppose that $f = \sum a_i t^i$ and $g = \sum b_i t^i$. If b_j is nonzero then so is $a_j - ub_j$, lest u lies in F. Thus $\deg h = \max\{\deg f, \deg g\}$ and x is algebraic over $F(u)$. It follows that u is transcendental over F as x is. To finish, we must show that h is irreducible in $F(u)[t]$. Suppose this is false, then h is also reducible in the UFD $F[u,t]$ by a consequence of Gauss' Lemma (Lemma 32.7). As h is primitive, we have $h = h_1 h_2$ in $F[u,t]$ with h_1 and h_2 lying in $F[u,t] \setminus F$. As u occurs linearly in h, we may assume that it occurs in h_1, say $h_1 = uh_3 + h_4$ with h_3, h_4 in $F[t]$ and h_2 lies in $F[t]$. So we have

$$g = f - ug = uh_2h_3 - h_2h_4.$$

As u is transcendental over $F(t)$, w must have $f = -h_2h_4$ and $g = -h_2h_3$. Therefore, $h_2 \mid f$ and $h_2 \mid g$ in $F[t]$, a contradiction. □

Corollary 61.2. *Let F be a field, x a transcendental element over F, and u an element of $F(x)$. Then $F(u) = F(x)$ if and only if there exist elements a, b, c, d in F with $ad - bc$ nonzero and $u = \dfrac{ax+b}{cx+d}$.*

[Note the condition on a, b, c, d insures that u does not lie in F.]

Proposition 61.3. *Let F be a field and x a transcendental element over F, then*

$$G(F(x)/F) = \{f : F(x) \to F(x) \mid x \mapsto \frac{ax+b}{cx+d} \text{ an } F\text{-homomorphism with } a, b, c, d \text{ in } F \text{ satisfying } ad - bc \neq 0\}$$
$$\cong \mathrm{GL}_2(F)/Z\big(\mathrm{GL}_2(F)\big).$$

The group $\mathrm{GL}_2(F)/Z\big(\mathrm{GL}_2(F)\big)$ is called the (second) *projective linear group* and denoted by $\mathrm{PGL}_2(F)$. If $F = \mathbb{C}$, then $\mathbb{C} = \mathbb{C}^2$, so $\mathrm{PGL}_2(\mathbb{C}) = \mathrm{PSL}_2(\mathbb{C}) := \mathrm{SL}_2(F)/Z\big(\mathrm{SL}_2(F)\big)$. (If $A \in \mathrm{GL}_2(\mathbb{C})$ multiply it by $\begin{pmatrix} d & 0 \\ 0 & d \end{pmatrix}$ where $d = (\sqrt{\det A})^{-1}$.)

Theorem 61.4. (Lüroth's Theorem) *Let F be a field and x a transcendental element over F. If $F(x)/E/F$ is an intermediate field, then there exists an element y in $F(x)$ such that $E = F(y)$.*

PROOF. We may assume that $F < E \subset F(x)$. Let z be an element in E not in F, then $F(x)/F(z)$ is algebraic by the lemma, hence $F(x)/E$ is algebraic. Let $f = am_E(x)$ with $a \in F[x]$ and $f \in F[x][t]$ primitive. Say

$$f = a_n(x)t^n + \cdots + a_0(x) \text{ with } a_i \in F[x] \text{ for all } i$$

where n is the degree $\deg_t$ of f in t, so $a_n(x)$ is nonzero and $n = \deg m_E(x) = [F(x) : E]$. We also have all $a_i(x)/a_n(x)$ lie in E but as x is transcendental over F, there exists an i such that $u = a_i(x)/a_n(x)$ does not lie in F. We can also write $u = g(x)/h(x)$ with g and h nonzero relatively prime polynomials in $F[t]$. If $r = \max\{\deg g, \deg h\}$ then $r = [F(x) : F(u)]$ by the lemma, so

$$n = [F(x) : E] \leq [F(x) : F(u)] = r.$$

Let $d = g(x)h(t) - h(x)g(t)$ in $F[x,t]$. As g and h are relatively prime in the UFD $F[x,t]$, the polynomial d is nonzero. In addition, the polynomial

$$uh(t) - g(t) = h(x)^{-1}d \text{ in } E[t] \text{ has } x \text{ as a root,}$$

so $m_E(x) \mid h(x)^{-1}d$ in $E[t]$, hence $f = a(x)m_E(x) \mid h(x)^{-1}d$ in $F(x)[t]$. Since f is primitive in $F[x,t]$ and d lies in $F[x,t]$, we must have $f \mid d$ in $F[x,t]$, as a consequence of Gauss' Lemma (e.g., by Exercise 3ii(32.12)). Write $d = \alpha f$ with α in $F[x,t]$ and let $\deg_x$ denote the degree in x. We have $\deg_x d \leq r$ and $\deg_x f \geq \max\{\deg a_j(x) \mid a_j(x) \neq 0\}$. By choice, $u = g(x)/h(x) = a_i(x)/a_n(x)$, so $\deg_x f \geq \max\{\deg g, \deg h\} = r$. Consequently, $d = \alpha f$ in $F[x,t]$ and $f \mid d$ in $F[x,t]$ implies that $\deg_x d = r = \deg_x f$ and $\deg_x \alpha = 0$. In particular, α lies in $F[t]$, so is primitive when viewed as a polynomial over $F[x]$, i.e., an element in $F[x][t]$. By Gauss' Lemma, $d = f\alpha$ is also primitive in $F[x][t]$. Since d is skew symmetric in x and t, it must also be primitive in $F[t][x]$. As α lies in $F[t]$ and $\alpha \mid d$ in $F[x,t]$, we must have α lies in $F[t]^\times = F^\times$, hence

$$n = \deg_t f = \deg_t d = \deg_x d = \deg_x f = r,$$

so $E = F(u)$. □

It is known if F is an algebraically closed field of characteristic zero, e.g., $\mathbb{C}$, that any subfield of $F[t_1, t_2]$ is purely transcendental. (A similar result holds if the characteristic is positive with a mild additional assumption.) However, the result is false for $F[t_1, t_2, t_3]$.

Suppose that F is a field and x is transcendental over F. Let $H \subset G(F(x)/F) \cong \mathrm{PGL}_2(F)$ be a finite subgroup. Then $F(x)/F(x)^H$ is a finite Galois extension. By Lüroth's Theorem, there exists an element u in $F(x)$ such that $F(u) = F(x)^H$. Suppose that $F = \mathbb{C}$, Then using a bit of complex analysis (stereographic projections, etc.), one can show that if $H \subset \mathrm{PGL}_2(\mathbb{C})$ is a finite subgroup, then H is isomorphic to a finite rotation group in $\mathbb{R}^3$. In Section §19 we classified such rotation groups in $\mathbb{R}^3$. Recall that they are the following:

(1) cyclic: $\cong \mu_n$, all n (planar).
(2) dihedral: $\cong D_n$, all n (planar).
(3) tetrahedral: $\cong A_4$ (the rotations of a regular tetrahedron).
(4) octahedral: $\cong S_4$ (the rotations of a cube or a regular octahedron).
(5) icosohedral: $\cong A_5$ (the rotations of a regular dodecahedron or a regular icosahedron).

Exercise 61.5. Prove that the rotations of a regular tetrahedron is isomorphic to A_4, the rotations of a cube or octahedron is isomorphic to S_4, and the rotations of a dodecahedron or icosahedron is isomorphic to A_5 using Exercise 22.22(5).

62. Finite Fields

A major property of a finite field F is that the multiplicative group of F is cyclic. If you have done the appropriate exercises, you see that we have all of the following:

Theorem 62.1. *Let F be a finite field with $|F| = q$. Then all the following are true:*

(1) $F^\times$ is a cyclic group.
(2) The characteristic of F is p for some prime p and $q = p^n$.
(3) If K/F is a finite extension of fields of degree m, then $|K| = q^m$.
(4) If p is a prime and m a positive integer, then there exists a field with p^m elements.
(5) If $|F| = p^n$, with p a prime, then F is a splitting field of $t^{p^n} - t$ over $\mathbb{Z}/p\mathbb{Z}$.
(6) Any two finite fields with the same number of elements are isomorphic. In particular, in a fixed algebraic closure of $\mathbb{Z}/p\mathbb{Z}$, with p a prime, there exists a unique field of order p^m for each positive integer m.

(7) *If* $|F| = q = p^n$, *then* $\varphi_F : F \to F$ *by* $x \mapsto x^q$ *is an* F-*automorphism, the Frobenius automorphism, and* $G(F/\mathbb{Z}/p\mathbb{Z}) = \langle \varphi_F \rangle$.
(8) *If* $|F| = p^n$ *and* K *a field with* $|K| = p^m$ *in an algebraic closure of* F, *then* K/F *if and only if* $n \mid m$.
(9) *If* K/F *is a finite extension of degree* m *and* $\varphi_K : K \to K$ *the Frobenius automorphism* $x \mapsto x^{q^m}$, *then* K/F *is a cyclic extension with* $G(K/F) = \langle \varphi_K^q \rangle$.
(10) *Any finite field is perfect.*
(11) *There exists an* x *in* F *such that* $F = (\mathbb{Z}/p\mathbb{Z})(x)$.

Given a finite field F with $q = p^n$ elements, one would like to find irreducible polynomials in $F[t]$. If f is an irreducible polynomial of degree d in $F[t]$, then by the theorem, $f \mid t^{p^n} - t$ if and only if $d \mid n$. In particular, we must have

$$t^{p^n} - t = \prod_{d|n} f_{i,d}$$

where the $f_{i,d}$ in $F[t]$ run over all the monic irreducible polynomials of degree d. So we have $p^n = \sum_{d|n} dN_{p,d}$, where $N_{p,d}$ is the number of monic irreducible polynomials of degree d in $F[t]$, by comparing degrees. By the general form of Möbius Inversion (Exercise 56.21(6)), we conclude that

Proposition 62.2. *The number* $N_{p,n}$ *of monic irreducible polynomials of degree* n *in* $(\mathbb{Z}/p\mathbb{Z})[t]$, *with* p *a prime, satisfies*

$$N_{p,n} = \frac{1}{n} \sum_{d|n} \mu(\frac{n}{d}) p^d.$$

Using the general form of the Möbius Inversion given by Exercise 56.21(6), we can determine irreducible polynomials, just as in the case of cyclotomic extensions of $\mathbb{Q}$. Let ζ_n be a primitive root of unity over $\mathbb{Z}/p\mathbb{Z}$, i.e., a generator for $F^\times$ if F is a field with p^n elements satisfying $p \nmid n$. Let Φ_n denote the product of $t - \zeta$ where ζ is a primitive nth root of unity in an algebraic closure of $\mathbb{Z}/p\mathbb{Z}$, so $\deg \Phi_n = \varphi(n)$. Unlike the case of $\mathbb{Q}$, the polynomial Φ_n is not irreducible. We have $t^n - 1 = \prod_{d|n} \Phi_d$. Arguing as the case over $\mathbb{Q}$, we conclude that

$$\Phi_n = \prod_{d|n} (t^d - 1)^{\mu(\frac{n}{d})} = \prod_{d|n} (t^{\frac{n}{d}} - 1)^{\mu(d)}.$$

in $(\mathbb{Z}/p\mathbb{Z})[t]$ by Möbius Inversion.

Example 62.3. Let p be a prime not dividing 12, then

$$\begin{aligned}\Phi_{12} &= \prod_{d|12}(t^{\frac{12}{d}}-1)^{\mu(d)}\\ &= (t^{12}-1)^{\mu(1)}(t^6-1)^{\mu(2)}(t^4-1)^{\mu(3)}(t^3-1)^{\mu(4)}\\ &\qquad\qquad (t^2-1)^{\mu(6)}(t-1)^{\mu(12)}\\ &= \frac{(t^{12}-1)(t^2-1)}{(t^6-1)(t^4-1)} = t^4-t^2+1.\end{aligned}$$

Lemma 62.4. *Let F be a finite field with q elements and i a positive integer. Set*

$$S(\widehat{t^i}) := \sum_{x\in F} x^i \ \text{ in } F,$$

then

$$S(\widehat{t^i}) = \begin{cases} -1 & \text{if } q-1 \mid i\\ 0 & \text{otherwise.}\end{cases}$$

For convenience, as $q = 0$ in F, we shall set $S(\widehat{t^0}) = 0$ also, i.e., let $0^0 = 1$.

Proof. Suppose that $q-1 \mid i$, then $x^i = 1$ for all x in $F^\times$ and $S(\widehat{t^i}) = \sum_{F^\times} x^i = |F^{q-1}|1_F = -1$ in F. So we may assume that $q-1 \nmid i$. Then there exists an element y in $F^\times$ such that $y^i \neq 1$ as $F^\times$ is cyclic and the gcd of $|F^\times|$ and i cannot be i. Since $yF^\times = F^\times$, we have

$$S(\widehat{t^i}) = \sum_{F^\times} x^i = \sum_{F^\times}(yx)^i = y^i \sum_{F^\times} x^i,$$

so $(1-y^i)\sum_{F^\times} x^i = 0$. It follows that $S(\widehat{t^i}) = 0$ □

If $f = \sum a_{i_1,\dots,i_n} t_1^{i_1}\cdots t_n^{i_n}$ is a polynomial in $R[t_1,\dots,t_n]$ (R a commutative ring), then define the *total degree* deg of f is defined to be $\max\{i_1+\cdots+i_n \mid a_{i_1,\dots,i_n} \text{ nonzero}\}$. We are interested in solutions of $f = 0$ in R^n for such an f when R is a finite field.

Theorem 62.5. (Chevalley-Warning Theorem) *Let F be a finite field of characteristic p and f a non-constant polynomial in $F[t_1,\dots,t_n]$ having total degree* $\deg f < n$. *Set*

$$V = \{\underline{x} = (x_1,\dots,x_n) \in F^n \mid f(\underline{x}) = 0 \ \text{ in } F\}.$$

Then $|V| \equiv 0 \mod p$.

PROOF. Suppose that F has q elements. Set $P = 1 - f^{q-1}$ in $F[t_1, \dots, t_n]$. As $|F^\times| = q - 1$, if $\underline{x}$ lies in V, then $f(\underline{x}) = 0$ in F, so $P(\underline{x}) = 1$ in F; and if $\underline{x}$ does not lie in V, then $P(\underline{x}) = 0$ in F, i.e.,

$$P(\underline{x}) = \begin{cases} 0 & \text{if } \underline{x} \notin V \\ 1 & \text{if } \underline{x} \in V. \end{cases}$$

(So P is the characteristic function for V.) For any g in $F[t_1, \dots, t_n]$, let

$$S(\widehat{g}) := \sum_{F^n} g(\underline{x}) \text{ in } F,$$

so

$$S(\widehat{P}) = \sum_{F^n} P(\underline{x}) = \sum_{V} 1 = |V| 1_F \text{ in } F.$$

In particular, if we show $S(\widehat{P}) = 0$ in F, then $|V| \equiv 0 \mod p$ as needed.

We know $\deg P < n(q-1)$, so if $P = \sum_{i_1,\dots,i_n} a_{i_1,\dots,i_n} t_1^{i_1} \cdots t_n^{i_n}$, we have $\sum_{j=1}^n i_j < n(q-1)$ if $a_{i_1,\dots,i_n} \neq 0$. In particular, if $a_{i_1,\dots,i_n} \neq 0$ is a coefficient of P, then, there exists a j, $1 \leq j \leq n$, with $i_j < q-1$. Therefore, fixing such a j for each such $a_{i_1,\dots,i_n} \neq 0$, we have

$$\begin{aligned} S(\widehat{P}) &= \sum_{F^n} P(\underline{x}) = \sum_{F^n} \sum_{i_1,\dots,i_n} a_{i_1,\dots,i_n} x_1^{i_1} \cdots x_n^{i_n} \\ &= \sum_{i_1,\dots,i_n} a_{i_1,\dots,i_n} \sum_{F^n} x_1^{i_1} \cdots x_n^{i_n} = \sum_{i_1,\dots,i_n} a_{i_1,\dots,i_n} S(\widehat{t_1^{i_1} \cdots t_n^{i_n}}) \\ &= \sum_{i_1,\dots,i_n} a_{i_1,\dots,i_n} \sum_{F^{n-1}} (\sum_{F} y^{i_j}) x_1^{i_1} \cdots x_{j-1}^{i_{j-1}} x_{j+1}^{i_{j+1}} + \cdots + x_n^{i_n} \\ &= 0, \end{aligned}$$

as $\sum_{\substack{y \in F \\ i_j < q-1}} y^{i_j} = S(\widehat{t^{i_j}}) = 0$ by the lemma. □

Of course, one is interested in solutions of such a non-constant polynomial f in $F[t_1, \dots, t_n]$ if every variable t_i occurs in f nontrivially, since the result is always true otherwise. A polynomial f in $F[t_1, \dots, t_n]$ is called *homogeneous of degree d* if every nonzero monomial in f has the same total degree d.

Corollary 62.6. *Let F be a finite field and f a non-constant polynomial in $F[t_1, \dots, t_n]$ having total degree $\deg f < n$. Suppose that $f(\underline{0}) = 0$. Then there exists a point $\underline{x}$ in F^n besides the origin such that $f(\underline{x}) = 0$. In particular, if f is homogeneous of degree d satisfying $n > d$, then f has a nontrivial solution in F^n.*

A homogeneous polynomial of degree two is called a *quadratic form.* The corollary says that any quadratic form in at least three variables over a finite field has a nontrivial solution. A much deeper theorem (due to Weil) says that if f is any homogeneous polynomial in $\mathbb{Z}[t_1, \ldots, t_n]$ that remains irreducible in $\mathbb{C}[t_1, \ldots, t_n]$, then $f \equiv 0 \mod p$ has a nontrivial zero for all primes $p >> 0$.

A theorem of Jacobson says that if R is a ring (even a rng) in which for each element x there exists an integer $n = n(x) > 1$ satisfying $x^n = x$, then R is commutative. For a proof of this more general result, see Appendix §91. The following result is the starting point.

Theorem 62.7. (Wedderburn) *Every finite division ring is a field.*

PROOF. (Witt) Let D be a division ring. Define the *center* $Z(D)$ of D to be $Z(D) := \{x \in D \mid xy = yx \text{ for all } y \text{ in } D\}$. So $Z(D^\times)$ is the center of the group $D^\times$. Clearly, $Z(D)$ is a field. Now suppose that D is finite. Let $F = Z(D)$. Then F is also finite, so $\operatorname{char} F = p$ for some prime p and $|F| = q = p^n$, for some n. We know that D is an F-vector space, so $|D| = q^m$ where $m = \dim_F D$. Assume that D is not commutative, i.e., $m > 1$. If $x \in D$ does not lie in F, then the conjugacy class of x in $D^\times$ is $\{y \in D^\times \mid y = axa^{-1} \text{ for some } a \text{ in } D^\times\}$ and the centralizer of x in $D^\times$ is $Z_{D^\times}(x)$. Set $Z_D(x) = Z_{D^\times}(x) \cup \{0\}$. Then $F \subset \mathbb{Z}_D(x) \subset D$. Clearly, $Z_D(x)$ is also a division ring and a finite dimensional vector space over F. Set $\delta(x) := \dim_F Z_D(x)$. We have $|Z_D(x)| = q^{\delta(x)}$. The notation and basic theorems about vector spaces over fields hold over division rings (i.e., as (left) modules over division rings) with the same proofs. (The one exception is the Representation Theorem of linear operators where matrix multiplication of matrices A and B must be modified so that the $ij$the entry of AB is $\sum B_{kj}A_{ik}$ – unless linear operators and scalars are written on opposite sides.) In particular, D is a (left) $Z_D(x)$-vector space. It is easy to see that $q^m = |D| = |Z_D(x)|^r = q^{\delta(x)r}$, some r. In particular, $m = \delta(x)r$, so $\delta(x) \mid m$ (cf. with the analog of a tower of fields). By assumption $\delta(x) < m$ as x does not lie in F. Let $\mathcal{C}^*$ be a system of representatives of the conjugacy classes of $D^\times$ not in the center of $D^\times = F^\times$. By the Class Equation 20.3, we have

$$q^m - 1 = |F^\times| + \sum_{\mathbb{C}^*} [D^\times : Z_{D^\times}(x)] = (q-1) + \sum_{\mathcal{C}^*} \frac{q^m - 1}{q^{\delta(x)} - 1}.$$

Let Φ_m be the mth cyclotomic polynomial $\prod_{\zeta \in \mu_m \text{ primitive}} \zeta$ in $\mathbb{Z}[t]$. Then we know that $\Phi_m \mid \dfrac{t^m - 1}{t^\delta - l}$ in $\mathbb{Z}[t]$ if $\delta \mid m$ and $\delta \neq m$, as $t^m - 1 =$

$\prod_{d|m} \Phi_d$. Therefore,

$$\Phi_m(q) \mid q^m - 1 \text{ and } \Phi_m(q) \mid \sum_{\mathcal{C}^*} \frac{q^m - 1}{q^{\delta(x)} - 1}$$

in $\mathbb{Z}$. Consequently, $\Phi_m(q) \mid q - 1$ in $\mathbb{Z}$.

Claim. If ζ is a primitive mth roof of unity, then $|q - \zeta| > |q - 1|$:

This is clear, if you draw a picture or if $\zeta = e^{2\pi\sqrt{-1}r/m}$ with r and m relatively prime, then

$$\begin{aligned} |q - \zeta|^2 &= (q - e^{2\pi\sqrt{-1}r/m})(q - e^{-2\pi\sqrt{-1}r/m}) \\ &= q^2 + 1 - 2q\cos\frac{2\pi r}{m} > q^2 + 1 - 2q = (q-1)^2. \end{aligned}$$

As $-1 < \cos\frac{2\pi r}{m} < 1$, the claim is established. The claim implies that

$$|\Phi_m(q)| = \prod_{\zeta \in \mu_n \text{ primitive}} |q - \zeta| > |q - 1|.$$

Hence $\Phi_m(q) \mid q - 1$ in $\mathbb{Z}$ is impossible. This contradiction shows that $D = F$. □

Exercises 62.8. 1. Prove Theorem 62.1.

2. Show that the nth cyclotomic polynomial $\Phi_n \in \mathbb{Z}/p\mathbb{Z}[t]$ with p a prime not dividing n is a product of $\varphi(n)/d$ of irreducible polynomials of the same degree d.
3. Show that over over finite field F, every element of F is a sum of two squares.
4. Let K/F be a finite extension of finite fields. Prove that the norm map $N_{K/F} : K \to F$ is surjective.
5. Let F be a finite field and f a homogeneous form of degree 2 (i.e., a quadratic form) in two variables with both variables occurring nontrivially. Then every nonzero element x in F is a value of f, i.e., there exists a y in F satisfying $f(y) = x$.

63. Addendum: Hilbert Irreducibility Theorem

Throughout this section, $x, y, t, t_1, \ldots, t_n$ will be independent variables. Hilbert's Irreducibility Theorem says if $f(t_1, \ldots, t_n, t)$ is an irreducible polynomial in $\mathbb{Q}[t_1, \ldots, t_n, t]$, then $f(\alpha_1, \ldots, \alpha_n, t)$ is irreducible in $\mathbb{Q}[t]$ for infinitely many rational numbers $\alpha_1, \ldots, \alpha_n$. Hilbert used this theorem to prove that there exists a finite Galois extension of $\mathbb{Q}$ with Galois group S_n and A_n for every n. We shall apply the theorem to show S_n occurs as a Galois group over $\mathbb{Q}$ for all n. This will show

that there exist irreducible polynomials of any degree over $\mathbb{Q}$ having no root solvable by radicals.

We begin by looking at a polynomial $f(x,t)$ in $\mathbb{Q}[x,t]$. Write it as

$$f(x,t) = a_n(x)t^n + \cdots + a_0(x) \text{ with } a_i(x) \in \mathbb{Q}[x] \text{ and } a_n(x) \text{ nonzero,}$$

so the degree of $f(x,t)$ in t, $\deg_t f$, is n. We can view $f(x,t)$ in $\mathbb{C}[x,t]$. Since $\mathbb{Q}$ is infinite, we can and do identify polynomial functions and polynomials over $\mathbb{C}$. We know by Proposition 39.8 if, in addition, that f is irreducible in $\mathbb{C}[x,y]$ that all but finitely many complex numbers x_0 are regular values, i.e., $a_n(x_0)$ is not zero and $f(x_0,t)$ has n distinct roots in $\mathbb{C}$. We say that a real or complex valued function $\rho(x)$ is *analytic* at x_0 if ρ converges to its Taylor series in a neighborhood of x_0, i.e., has a positive radius of convergence R centered at x_0. If x_0 is a regular value of $f(x,t)$, we shall call an analytic function $\rho = \rho(x)$ a *root function* of $f(x,t)$ in a neighborhood of x_0 if $f\big(x,\rho(x)\big) = 0$ in a neighborhood of x_0, i.e., there exists a real number $R > 0$ such that $\rho(x)$ are power series in $x - x_0$ for all x satisfying $|x\_x_O| < R$. In particular, $\rho(x_0)$ is a root of $f(x_0,t)$. We begin by constructing root functions in a neighborhood of regular value x_0 of any given polynomial in $\mathbb{C}[x,t]$.

Proposition 63.1. *Let*

$$f(x,t) = a_n(x)t^n + \cdots + a_0(x) \text{ with } a_i(x) \in \mathbb{Q}[x]$$

be a polynomial in $\mathbb{Q}[x,t]$ of degree n in t (so $(a_n(x)$ is nonzero) and x_0 a regular value of $f(x,t)$. Then there exist precisely n distinct root functions $\rho_1,\ldots,\rho_n$ of $f(x,t)$ at x_0 and they satisfy

$$(*) \qquad f(x,t) = a_n(x)\prod_{i=1}^{n}\big(t-\rho_i(x)\big) \text{ in a neighborhood of } x_0$$

with the $\rho_i(x)$ power series expansions in $x - x_0$ with coefficients in $\widetilde{\mathbb{Q}}$, the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$, in this neighborhood of x_0.

Proof. Let $\alpha_1,\ldots,\alpha_n$ be the (distinct) roots of $f(x_0,t)$ in $\mathbb{C}$, so lie in a finite extension of $\mathbb{Q}$, e.g., any field containing $K = \mathbb{Q}(\alpha_1,\ldots,\alpha_n)$ in $\widetilde{\mathbb{Q}}$. For each $i = 1,\ldots,n$, we shall construct a unique root function ρ_i satisfying $\rho_i(x_0) = \alpha_i$ with coefficients K in its power series expansion about x_0. We make the following reductions:

(*i*) As $a_n(x_0) \neq 0$, we may assume that $a_n(x_0) = 1$.
(*ii*) We may assume that $x_0 = 0$ by replacing $f(x,t)$ by the polynomial $g(x,t) = f(x + x_0, t)$.

(*iii*) We may assume that $\alpha_i = 0$ by replacing $f(x,t)$ by the polynomial $g(x,t) = f(x, t - \alpha_i)$.

So we can assume that $x_0 = 0$ is a regular value of $f(x,t)$ having root $\alpha_i = 0$ and we must construct a unique root function in a neighborhood of $x = 0$ with the given conditions. Therefore, we must find a unique power series

$$\rho(x) = \sum_{i=1}^{\infty} b_i x^i \text{ with } b_i \in K \text{ converging for all } |x| < R \tag{63.2}$$

for some positive real number R that satisfies $f\big(x, \rho(x)\big) = 0$ in a neighborhood of $x = 0$. As $f(0,0) = 0$, we can write the polynomial

$$f(x,t) = a_{10}x + a_{01}t + \sum_{i+j\geq 2} a_{ij}x^i t^j \text{ in } \mathbb{Q}[x,t]$$

with almost all a_{ij} zero and the $a_{ij} \in K$. Since $x_0 = 0$ is a regular value of $f(x,t)$, we also have $\dfrac{\partial f}{\partial t}(0,0) = a_{01}$ is nonzero, so we can further assume that $a_{01}(0,0) = -1$, i.e.,

$$f(x,t) = a_{10}x - t + \sum_{i+j\geq 2} a_{ij}x^i t^j \text{ in } \mathbb{Q}[x,t]. \tag{63.3}$$

Assume that $\rho(x)$ has a positive radius of convergence about 0. We show that the b_i in (63.2) are uniquely determined by the equation

$$f(x, \sum_{i=1}^{\infty} b_i x^i) = 0 \tag{63.4}$$

in a neighborhood of $x = 0$, i.e., $\rho(x)$ is the unique root function in a neighborhood of 0 and with the b_i lying in K.

The coefficient of x^i in the expansion of the left hand side of equation (63.4) must be zero for each i. In particular, evaluating the x term shows that $0 = a_{10} - b_1$, so we must have $b_1 = a_{10}$. Inductively, we may suppose that we have constructed unique $b_1, \ldots, b_{k-1}$ in K. It is convenient to view y as a new variable to replace t. Let

$$h(x,y) = a_{10}x - y + \sum_{i+j\geq 2} a_{ij}x^i y^j \text{ and } y_0 = \sum_{i=1}^{k-1} b_i x^i.$$

Then the (formal) Taylor expansion for $h(x,y)$ at y_0 exists by Exercise 63.19(2) and is

$$h(x,y) = f(x,y_0) + \frac{\partial f}{\partial y}(x,y_0)(y-y_0) \\ + \text{ terms in higher powers of } y-y_0.$$

By (63.3), we must have

$$\frac{\partial f}{\partial y}(x,y) = -1 + g(x,y)$$

with every term in $g(x,y)$ having total degree at least one. Evaluating y at $\rho(x)$ yields $\rho(x) - y_0 = \sum_{i=k}^{\infty} b_i x^i$, so

$$0 = f(x, \sum_{i=1}^{k-1} b_i x^i) + \big(-1 + g(x, \sum_{i=1}^{k-1} b_i x^i)\big)(\sum_{i=k}^{\infty} b_i x^i) \\ + \text{ terms with power of } x \text{ at least } 2k.$$

As the coefficient of x^k in this equation is zero, the coefficient $-b_k$ of x^k of $f(x,y_0)$ in this equation is uniquely determined by f and $b_1, \ldots, b_{k-1}$ as desired.

We can say more. If $\rho(x)$ is an analytic function, i.e., has positive radius of convergence, then in some neighborhood of $x_0 = 0$, we would have $b_1 = a_{10}$ and $f(x, \sum_{i=1}^{\infty} b_i x^i) = 0$, so

$$\sum_{i=2}^{\infty} b_i x^i = \sum_{i+j \geq 2} a_{ij} x^i (\sum_{l=1}^{\infty} b_l x^l)^j.$$

Expanding this and looking at the coefficient of x^k for $k > 1$ leads to an equation $b_k = \sum_{i+lj=k} c_{ijl} a_{ij} b_l$ for some nonnegative integers c_{ijl}. In particular, $l < k$. For each of the finite nonzero a_{ij} occurring in $f(x,t)$, let t_{ij} be a new variable. By induction, it follows that there exist nonzero polynomials $p_k \in \mathbb{Z}[t_{ij}]_{i,j}$ having positive integer coefficients (in the nonzero terms) and satisfying $b_k = p_k(a_{ij})$.

To finish we must still show that $\rho(x) = \sum b_i x^i$ has a positive radius of convergence in a neighborhood of $x_0 = 0$. Choose a positive integer A satisfying $A > |a_{ij}|$ for all of the finitely many nonzero a_{ij} occurring in $f(x,t)$ and set

$$f_A(x,t) = Ax - t + A(\sum_{i+j \geq 2} x^i t^j).$$

By the argument above, we know if f_A has a positive radius of convergence, then there exist unique $\widetilde{b}_k \in \mathbb{C}$ and $\widetilde{p}_k \in \mathbb{Z}[t_{ij}]_{i,j}$ satisfying $f_A(x, \sum_{i=1}^{\infty} b_i x^i) = 0$ and $\widetilde{b}_k = \widetilde{p}_k(A, \ldots, A)$ are positive integers for

all k. Since $\widetilde{b}_k \geq p_k(|a_{ij}|) \geq |b_k|$ for all k, it suffices to show that $y = \sum_{i=1}^{\infty} \widetilde{b}_i x^i$ converges for $|x| < R$ for some positive real number R. Summing the geometric series, shows

$$\begin{aligned} 0 &= Ax - y + A \sum_{i+j \geq 2} x^i y^j \\ &= Ax - y + Ax^0 \sum_{j=2}^{\infty} y^j + Ax^1 \sum_{j=1}^{\infty} y^j + A \sum_{i=2}^{\infty} x^i (\sum_{j=0}^{\infty} y^j) \\ &= Ax - y + A\frac{y^2}{1-y} + Ax\frac{y}{1-y} + A\frac{x^2}{1-x}\frac{1}{1-y}. \end{aligned}$$

Multiplying by $1 - y$ yields

$$\begin{aligned} 0 &= Ax - Ayx + -y + y^2 + Ay^2 + Axy + A(\frac{x^2}{1-x}) \\ &= (A+1)y^2 - y + A\frac{x}{1-x}. \end{aligned}$$

One of the two solutions for $y(0) = 0$, i.e., when $x = 0$ is

$$y = \frac{1 - \sqrt{1 - 4(A+1)Ax/(1-x)}}{2(A+1)}.$$

Since $\sqrt{1 - \big(bx/(1-x)\big)} = (\sqrt{1-cx})/(\sqrt{1-x})$ has a positive radius of convergence about $x = 0$ for $b = 4A(A+1)$ and $c = b+1$, it follows that f_A is analytic in some neighborhood of 0, hence $\rho(x)$ is a root function in a neighborhood of 0.

We therefore conclude that for each $i = 1, \ldots, n$, there exist a unique root function $\rho_i(x)$ in a neighborhood of x_0 for the root α_i of $f(x_0, y)$ with coefficients in its power series expansion in K. Since the the constant term of $\rho_i(x)$ is α_i for $i = 1, \ldots, n$, the ρ_i are all distinct in some common neighborhood of x_0 and satisfy $f(x,t) = \prod_{i=1}^{n} \big(t - \rho_i(x)\big)$ in this neighborhood. □

We shall need the following well-known lemma whose proof we leave as an exercise.

Lemma 63.5. (Lagrange Interpolation) *Suppose that F is an infinite field, $a_0, \ldots, a_n$ distinct elements in F and $c_0, \ldots, c_n$ elements in F. Then there exists a unique polynomial f in $F[t]$ of degree at most n satisfying $f(a_i) = c_i$ for each $i = 0, \ldots, n$.*

Let $\alpha_0 < \cdots < \alpha_m$ be real numbers. Recall that the $(m+1)$st Vandermonde determinant is given by

$$V_m = V_m(\alpha_0, \alpha_1, \ldots, \alpha_m) = \det \begin{pmatrix} 1 & \alpha_0 & \cdots & \alpha_0^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_m & \cdots & \alpha_m^m \end{pmatrix}$$
$$= \prod_{0 \le i < j \le m} (\alpha_j - \alpha_i) \neq 0.$$

If z is a real-valued function of x, let

$$W_m = W_m(\alpha_0, \alpha_1, \ldots, \alpha_m, z) = \det \begin{pmatrix} 1 & \alpha_0 & \cdots & \alpha_0^{m-1} & z(\alpha_0) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \alpha_m & \cdots & \alpha_m^{m-1} & z(\alpha_m) \end{pmatrix}.$$

With this notation, we have:

Lemma 63.6. *Let $\alpha_0 < \cdots < \alpha_m$ be real numbers and $z(x)$ an m times differential real-valued function on the closed interval $[\alpha_0, \alpha_m]$. Then there exist a real number α satisfying*

$$\frac{z^{(m)}(\alpha)}{m!} = \frac{W_m}{V_m}$$

satisfying $\alpha_0 < \alpha < \alpha_m$ with W_m and V_m as above. In particular, if $z^{(m)}(x)$ is nonzero on $[\alpha_0, \alpha_m]$, then W_m is also nonzero.

Proof. Let $y(t) = a_0 + a_1 t + \cdots + a_m t^m$ in $\mathbb{R}[t]$ be the unique polynomial satisfying $y(\alpha_i) = z(\alpha_i)$ for $i = 0, \ldots, m$ given by Lagrange Interpolation. By Cramer's Rule, the system of equations

$$a_0 + a_1\alpha_i + \cdots + a_m\alpha_i^m = z(\alpha_i) \text{ for } i = 0, \ldots, m$$

satisfies $a_m = W_m/V_m$.

By Rolle's Theorem there exists m distinct real numbers $\overline{\alpha}_i$ with each $\overline{\alpha}_i$ lying in the open interval (α_i, α_{i+1}) and satisfying $y^{(1)}(\overline{\alpha}_i) = z^{(1)}(\overline{\alpha}_i)$ for $i = 0, \ldots, m-1$. Similarly, we see that there exist $m-1$ distinct points on which $y^{(2)}$ and $z^{(2)}$ agree. Inductively, we find that there exists a point α satisfying $\alpha_0 < \alpha < \alpha_m$ on which y and z agree. Since $y^{(m)}(x) = m!\, a_m$, we have $z^{(m)}(\alpha) = y^{(m)}(\alpha) = m!\, a_m$ as needed. □

We now prove the Hilbert Irreducibility Theorem in the special case of two variables. We shall use the following observation.

Observation 63.7. As $\mathrm{char}(\mathbb{C}) = 0$, the domain of polynomials in $\mathbb{C}[t]$, can be identified with the domain of polynomial functions on an interval I (of nonzero length) of $\mathbb{C}$, hence as a subring of the domain

R of convergent functions on I. It follows that we can identify the quotient field of $\mathbb{C}[t]$ with a subfield of the quotient field of R, the field of finite Laurent series on I. (Note we can also work with polynomials in $1/t$ with the appropriate modifications.) We shall utilize this as an identification below.

Proposition 63.8. *Let $f(x,t)$ be an irreducible polynomial in $\mathbb{Q}[x,t]$. Then there exist infinitely many rational numbers α such that $f(\alpha,t)$ is irreducible in $\mathbb{Q}[t]$.*

Proof. Let

$$f(x,t) = a_n(x)t^n + \cdots + a_0(x)$$

with $a_i(x)$ in $\mathbb{Q}[x]$ for $i = 0,\dots,n$ and $a_n \neq 0$. Let $d = \max\{\deg_x a_i(x) \mid 0 \leq i \leq n\}$. Almost all elements of $\mathbb{C}$ are regular values of $f(x,t)$. Let x_0 be one such. Replacing $f(x,t)$ by the irreducible polynomial $h(x,t) = f(x+x_0,t)$, we may assume $x_0 = 0$. By Proposition 63.1, we can write

$$(i) \qquad f(x,t) = a_n(x)\prod_{i=1}^{n}\big(t-\rho_i(x)\big)$$

for each x in some neighborhood of $x_0 = 0$ with ρ_i the n distinct root functions of f at $x = 0$ (so $\rho_i(x_0)$ are the roots of $f(x_0,t)$) and these root functions have coefficients lying in $\widetilde{\mathbb{Q}}$, the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$, in their power series expansion in this neighborhood. It is convenient to work at 'infinity' rather than at 0, where a real or complex valued function $h(x)$ is said to be *analytic at infinity* if it can be represented by a power series $\sum_{i=1}^{\infty} c_i\dfrac{1}{x^i}$ converging for all $|x| > R$ for some positive real number R. Set

$$g(x,t) = x^d f(\frac{1}{x},t),$$

then $g(x,t)$ also lies in $\mathbb{Q}[x,t]$. Moreover, $g(x,t)$ is irreducible. Indeed, suppose that

$$g(x,t) = r(x,t)s(x,t)$$

with $r(x,t)$ and $s(x,t)$ polynomials in $Q[x,t]$ of degrees m_1 and m_2 in x, respectively, then $d = m_1 + m_2$ and

$$f(x,t) = x^d g(\frac{1}{x},t) = x^{m_1}r(\frac{1}{x},t)x^{m_2}s(\frac{1}{x},t)$$

is a factorization of $f(x,t)$ into polynomials. As $f(x,t)$ is irreducible, this must be a trivial factorization, i.e., either r or s lies in $\mathbb{Q}$ as required. Now note if $g(\alpha,t)$ is irreducible for $\alpha \in \mathbb{Q}$, so is $f(\frac{1}{\alpha},t) = \frac{1}{\alpha^d}g(\alpha,t)$. Therefore, we can replace $f(x,t)$ by $g(x,t)$ and work with power series

analytic at infinity, i.e., we may assume all the root functions ρ_i of $f(x,t)$ have power series that converge for all $|x| > R$ and (i) holds for all $|x| > R$ for some real number R.

Let S be a nonempty proper subset of $\{1,\dots,n\}$ and consider the equation

$$(ii) \qquad \prod_{i=1}^{n}\big(t-\rho_i(x)\big) = \prod_{i\in S}\big(t-\rho_i(x)\big)\prod_{i\notin S}\big(t-\rho_i(x)\big).$$

We know that for every value of x with $|x| > R$, we have $f(x,t) = a_n(x)\prod_{i=1}^{n}\big(t-\rho_i(x)\big)$ is a factorization into linear terms in $\widetilde{\mathbb{Q}}[t]$, Since $f(x,t)$ is irreducible in $\mathbb{Q}[x,t]$ and $a_n(x) \in \mathbb{Q}[x]$, neither factor on the right hand side of (ii) can lie in $\mathbb{Q}(x)[t]$ by Lemma 32.7, a consequence of Gauss' Lemma, and Observation 63.7. Consequently,

$$(iii) \qquad \prod_{i\in S}\big(t-\rho_i(x)\big) = t^{|S|} + b_1 t^{|S|-1} + \cdots + b_{|S|} \text{ in } \widetilde{\mathbb{Q}}[t],$$

with each $b_j = b_j(x)$, $1 \le j \le |S|$, analytic at infinity, but at least one, say $b_i = b_i(x)$ does not lie in $\mathbb{Q}(x)$. Let N be the number of such factorizations (ii), so $N = 2^{n-1}-1$. For each such S, let $y_S(x)$ be a coefficient $b_i(x)$ on the right hand side of (iii) analytic at infinity and not in $\mathbb{Q}(x)$.

Claim. Let $R' \ge R$ be arbitrary. Then there exists a rational number x_0 satisfying $x_0 > R'$ (so a regular value of f) satisfying $y_S(x_0)$ is not rational for all nonempty proper subsets S of $\{1,\dots,n\}$. In particular, $f(x_0,t)$ is irreducible and the proposition is true.

Suppose that x_0 satisfies the property of the claim, but that $f(x_0,t) = r(t)s(t)$ in $\mathbb{Q}[t]$ with r and s nonconstant polynomial in $\mathbb{Q}[t]$. Then there exists a nonempty proper subset S of $\{1,\dots,n\}$ satisfying

$$r(t) = a_n(x_0)\prod_{i\in S}\big(t-\rho_i(x_0)\big) \text{ and } s(t) = \prod_{i\notin S}\big(t-\rho_i(x_0)\big)$$

in $\mathbb{Q}[t]$. This means that $y_S(x_0)$ lies in $\mathbb{Q}$, a contradiction. So we need only construct x_0 as claimed, since $R' > R$ is arbitrary.

We construct an x_0 in $\mathbb{Q}$ satisfying the claim. In fact, we shall produce an integer x_0 satisfying the claim. Fix a nonempty proper S and let $y = y_S(x)$. Then y lies in $\mathbb{Q}[\rho_1(x),\dots,\rho_n(x)] \subset \widetilde{\mathbb{Q}}$, so $y(x)$ satisfies an equation

$$(iv) \qquad d_l y^l + d_{l-1}y^{l-1} + \cdots d_0 = 0$$

with $d_i \in \mathbb{Q}(x)$ for $i = 0, \ldots, l$ and d_l nonzero for some integer l. Clearing denominators, we may assume all the d_i lie in $\mathbb{Q}[x]$ and then clearing denominators again that all d_i lie in $\mathbb{Z}[x]$.

Multiplying equation (iv) by d_l^{l-1} shows that $z = z(x) = d_l y(x)$ satisfies an equation

$$z^l + b_{l-1}z^{l-1} + \cdots + b_0 = 0$$

with $b_0, \ldots, b_{l-1}$ all lying in $\mathbb{Z}[x]$. If $\alpha > R$ satisfies $y(\alpha)$ is rational, then $z(\alpha)$ is an integer as each $b_i(\alpha)$ is an integer and any rational root of a monic equation with integer coefficients must be an integer. Therefore, we conclude that $z(x)$ has an integral value at some integer $\alpha > R$ if and only if $y(x)$ has a rational value at that α. Consequently, it suffices to work with $z(x)$. As $z = d_l y$ with $d_l \in \mathbb{Z}[x]$ and $y(x)$ analytic at infinity, we have

$$(v) \qquad z(x) = c_k x^k + \cdots + c_0 + c_{-1}\frac{1}{x} + \text{ higher powers of } \frac{1}{x}$$

with $c_i \in \mathbb{C}$ and z converges for all $|x| > R$.

We must show that $z(\alpha)$ is not an integer for some integer $\alpha > R'$ for infinitely many integers $\alpha > R$, hence for any $R' \geq R$. There are three cases to consider.

Case 1. $z(x)$ is a polynomial in $\mathbb{C}[x]$:

The polynomial $z(x)$ can have at most k integer values by Lagrange Interpolation, lest $z(x)$ hence $y(x)$ be a polynomial in $\mathbb{Q}[x]$, contrary to choice. So $z(x)$ has no integer values at integer points for sufficiently large x.

Case 2. There exists a coefficient c_j of $z(x)$ in (v) that is not real:

We may assume that j has been chosen maximal with c_j not real. Then we have, where Im is the imaginary part,

$$\lim_{x\to\infty} \operatorname{Im}\Big(\frac{z(x)}{x^j}\Big) = \operatorname{Im}(c_j) \neq 0,$$

hence $z(x)$ is not real for all large real values of x.

Case 3. All the c_j in (v) are real and there exist a $j > 0$ with c_{-j} not zero:

Choose $m = m(S) > 0$ such that the mth derivative of $z(x)$ has no non-negative powers of x, say

$$z^{(m)}(x) = \frac{d}{x^k} + \text{ terms with larger powers of } \frac{1}{x}.$$

So d is a nonzero real number and k a positive integer. In particular, we have

$$\lim_{x\to\infty} x^k z^{(m)}(x) = d \neq 0.$$

Choose $R_1 > R'$ such that if $\alpha > R_1$, then

$$0 < |z^{(m)}(\alpha)| < \frac{2|d|}{\alpha^k}.$$

Suppose that there exist $m+1$ integers α_i satisfying

$$R_1 < \alpha_0 < \cdots < \alpha_m \text{ with } z(\alpha_i) \in \mathbb{Z} \text{ for } i = 0, \ldots, m.$$

Let

$$V_m = \det \begin{pmatrix} 1 & \alpha_0 & \cdots & \alpha_0^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_m & \cdots & \alpha_m^m \end{pmatrix}$$

and

$$W_m = \det \begin{pmatrix} 1 & \alpha_0 & \cdots & \alpha_0^{m-1} & z(\alpha_0) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \alpha_m & \cdots & \alpha_m^{m-1} & z(\alpha_m) \end{pmatrix}.$$

Applying Lemma 63.6, we see that both V_m and W_m are nonzero integers and

$$\frac{2|d|}{m!\,\alpha_0^k} \geq \frac{2|d|}{m!\,\alpha^k} \geq \frac{|z^{(m)}(\alpha)|}{m!} = \frac{|W_m|}{|V_m|} \geq \frac{1}{|V_m|}$$

for some α satisfying $\alpha_0 < \alpha < \alpha_m$. Therefore,

$$\frac{m!}{2|d|}\alpha_0^k \leq |V_m| = \prod_{i<j}(\alpha_j - \alpha_i) < (\alpha_m - \alpha_0)^{\frac{m(m+1)}{2}}.$$

Consequently, there exist positive real numbers γ and δ depending on m satisfying

$$(vi) \qquad \alpha_m - \alpha_0 > \gamma\alpha_0^\delta.$$

Note that the right hand side of (vi) goes to infinity as $\alpha_0 \to \infty$, i.e., the α_m's get further and further from α_0 as it gets larger. Recall that $N = 2^{n-1} - 1$ is the number of the $y(x)$'s. (We can replace N by any larger integer.) Choose $R_2 > R_1 > R' \geq R$ satisfying

$$\gamma R_2^\delta > Nm.$$

In the above, we may also assume that $\alpha_0 > R_2$. Therefore, we have $\alpha_0 < \alpha_1 < \cdots < \alpha_m$ with each α_i and each $z(\alpha_i)$ integers, $i = 1, \ldots, m$, and

$$\alpha_m - \alpha_0 > \gamma\alpha_0^\delta > \gamma R_2^\delta > Nm.$$

We conclude that among any $Nm+1$ sufficiently large consecutive integers $l, l+1, \ldots, l+Nm$, the function $z(x)$ takes on integer values

on at most m of these, equivalently $y(x)$ takes on rational values on at at most m of these.

If $z(x)$ satisfies Case 1 or 2, it has only finitely many integer values at integer points. Choose $R_0 > R_2$ such that all such $z(x)$'s have no integer values, hence $y(x)$ no rational values, at integers greater than R_0.

For each of the finite number of S's, we have constructed an integer $m = m(S)$ and a real number $R_0 = R_0(S) > R'$ such that for any $x > R_0$ and string of $Nm+1$ consecutive integers greater than R_0, the analytic function $y_S(x)$ has at most m rational values at such a string of $Nm + 1$ consecutive integers. Let

$$M \geq \max_S \big(m(S)\big) \text{ and } R_3 > \max_S \big(R_0(S)\big).$$

For any string of $NM + 1$ consecutive integers greater than R_3, each $y_S(x)$ can have at most M rational values. In particular, in each string of $NM+1$ consecutive integers greater than R_3, there exists an integer α such that $y_S(\alpha)$ is not rational for every one of the N functions $y_S(x)$. This establishes the claim and completes the proof. □

Corollary 63.9. *Let $f_1(x,t), \ldots, f_m(x,t)$ be irreducible polynomials in $\mathbb{Q}[x,t]$. Then there exist infinitely many rational numbers α such that $f_1(\alpha,t), \ldots, f_m(\alpha,t)$ are all irreducible in $\mathbb{Q}[t]$.*

Proof. By the proof of the proposition, we can work with the union of all the $y(x)$'s that arise from the $f_i(x,t)$'s with N the number of all of these. □

To prove the full version of the Hilbert Irreducibility Theorem, we use a trick developed by Kronecker.

Definition 63.10. Let F be a field, d and n positive integers. Set

$$P_d(n,F) := \{f \in F[t_1, \ldots, t_n] \mid \deg_{t_i} f < d \text{ for all } i = 1, \ldots, n\}$$

and

$$S_d : P_d(n,F) \to P_{d^n}(1,F)$$

defined by

$$S_d\big(f(t_1, \ldots, t_n)\big) = f(t, t^d, t^{d^2}, \ldots, t^{d^{n-1}}).$$

For example, $S_d(t_1^{i_1} \cdots t_n^{i_n}) = t^{i_1 + i_2 d + i_3 d^2 + \cdots i_n d^{n-1}}$. Let m be a positive integer, then m has a unique d-adic expansion. In particular, applying this to all positive integers $m < d^n$, we deduce that S_d is a bijection. Although $P_d(n,F)$ is not closed under products, we do have:

Remark 63.11. If $f, g, fg \in P_d(n, F)$, then

$$S_d(fg) = S_d(f)S_d(g).$$

In particular, since S_d is a bijection, if $S_d(f)$ is irreducible in $F[t]$, then f is irreducible in $F[t_1, \ldots, t_n]$.

The converse of this remark is false. For example, if $F = \mathbb{Q}$, then $f = t_1^2 + t_2^2$ in $\mathbb{Q}[t_1, t_2]$ is irreducible, but $S_3(f) = t^2 + t^6$ is not. Suppose that we have an irreducible polynomial $f \in P_d(n, F)$ but

$$S_d(f) = G(t)H(t) \text{ in } F[t] \text{ with } 0 < \deg G, \deg H < d^n,$$

i.e., $S_d(f)$ is not irreducible. Then G and H lie in P_{d^n}, so there exist unique polynomials g and h in $P_d(n, F)$ satisfying $G = S_d(g)$ and $H = S_d(h)$, i.e.,

$$S_d(f) = S_d(g)S_d(h) = S_d(gh)$$

Theorem 63.12. (Kronecker's Criterion) *Let F be a field and f an element in $P_d(n, F)$. Then f is irreducible in $F[t_1, \ldots, t_n]$ if and only if for all non-trivial factorizations (i.e., polynomials of positive degree)*

$$S_d(f) = S_d(g)S_d(h)$$

with g and h in $P_d(n, F)$, the polynomial gh contains a nonzero monomial of the form $bt_1^{i_1} \cdots t_n^{i_n}$ for some $i_j \geq d$.

PROOF. ($\Leftarrow$): If $f = f_1 f_2$ with $f_1, f_2 \in F[t_1, \ldots, t_n]$, then we have $f_1, f_2 \in P_d(n, f)$, $S_d(f) = S_d(f_1 f_2) = S_d(f_1)S_d(f_2)$, and $f = f_1 f_2$ has no nonzero monomial $bt_1^{i_1} \cdots t_n^{i_n}$, some $i_j > 0$.

($\Rightarrow$): Suppose that f is irreducible and $S_d(f) = S_d(f_1)S_d(f_2)$ for some $f_1, f_2 \in P_d(n, F)$. If $f_1 f_2 \in P_d(n, F)$, then $S_d(f) = S_d(f_1)S_d(f_2) = S_d(f_1 f_2)$. As $S_d : P_d(n, F) \to P_{d^n}(1, F)$ is one-to-one, $f = f_1 f_2$, and one of f_1, f_2 lies in F. □

Example 63.13. If $f = t_1^2 + t_2^2$ in $\mathbb{Q}[t_1, t_2]$, we have $S_3(f) = t^2(1 + t^4)$, $S_3(g) = t^2$ with $g = t_1^2$, $S_3(h) = 1 + t^4$ with $h = 1 + t_1 t_2$, and $gh = t_1^2 + t_1^3 t_2$.

Corollary 63.14. *Let F be a field and $f \in F[t_0, \ldots, t_n]$ irreducible with $\deg_{t_i} < d$ for $i = 1, \ldots, n$. Let $S_d : P_d\big(n, F(t_0)\big) \to P_{d^n}\big(1, F(t_0)\big)$ be defined by $S_d\big(f(t_0, t_1, \ldots t_n)\big) = f(t_0, x, x^d, \ldots, x^{d^n - 1})$. Suppose that*

$$(*) \qquad S_d(f) = \prod_{i=1}^{m} p_i(t_0, x)$$

with $p_i(t_0, x) \in F[t_0][x]$ irreducible for $i = 1, \ldots, m$. Then there exists an element $\varphi \in F(t_0)$ such that if $\alpha \in F$ satisfies $\varphi(\alpha)$ is defined

and nonzero and $p_i(\alpha, x)$ is irreducible in $F[x]$ for $i = 1, \ldots, n$, then $f(\alpha, t_1, \ldots, t_n)$ is irreducible in $F[t_1, \ldots, t_n]$.

PROOF. As $F[t_0]$ is a UFD, all factorizations of $S_d(f)$ into two polynomials in $F[t_0, x]$ arise from (*) (up to constants in F). View f as an irreducible polynomial in $F(t_0)[t_1, \ldots, t_n]$. By Kronecker's Criterion, any such factorization can also be written $S_d(f) = S_d(g)S_d(h)$ with $g, h \in P_d\big(n, F(t_0)\big)$ and satisfying some nonzero monomial $\varphi_{gh} t_1^{i_1} \cdots t_n^{i_n}$ occurs in gh with $\varphi_{gh} \in F(t_0)$ (and, in fact, in $F[t_0]$), for some $i_j > d$. Let φ be the product of all the φ_{gh} arising from such factorizations. If $\alpha \in F$ satisfies $\varphi(\alpha)$ is defined and nonzero and $p_i(\alpha, x)$ is irreducible in $F[x]$ for $1 \le i \le m$, then $S_d(f(\alpha, t_1, \ldots, t_n) = \prod_{i=1}^m p_i(\alpha, x)$ is a factorization into irreducibles in $F[x]$. It follows that $f(\alpha, t_1, \ldots, t_n)$ is irreducible in $F[t_1, \ldots, t_n]$ by Kronecker's Criterion. □

Theorem 63.15. (Hilbert Irreducibility Theorem) *Let $f(x_1, \ldots, x_n, t)$ be an irreducible polynomial in $\mathbb{Q}[x_1, \ldots, x_n, t]$. Then there exist infinitely many $(\alpha_1, \ldots, \alpha_n)$ in $\mathbb{Q}^n$ such that $f(\alpha_1, \ldots, \alpha_n, t)$ is irreducible in $\mathbb{Q}[t]$.*

PROOF. We induct on n. The case of $n = 1$ is Proposition 63.8. We shall apply the previous corollary and its notation to complete the proof. If $f \in P_d\big(n, \mathbb{Q}(x_1)\big)$ and $S_d : P_d\big(n, \mathbb{Q}(x_1)\big) \to P_1\big(d^n, \mathbb{Q}(x_1)\big)$, we are done if $S_d(f)$ is irreducible by induction by choosing α in $\mathbb{Q}(x_1)$ such that $\varphi(\alpha)$ is defined and nonzero. (For almost all x_0, we have $\varphi(x_0)$ is defined and nonzero.) So we may assume that we can factor $S_d(f) = \prod_{i=1}^m p_i(x_1, t)$ as in (*). By Corollary 63.9, there exist infinitely many α in $\mathbb{Q}$ satisfying $p_i(\alpha, t)$ is irreducible in $\mathbb{Q}[t]$ for $i = 1, \ldots, m$. By the previous corollary and induction, the theorem follows. □

It can be shown that the Hilbert Irreducibility Theorem holds with F replacing $\mathbb{Q}$ for any number field, i.e., any finite field extension of $\mathbb{Q}$ or more generally any non-algebraic finitely generated field extension of any field. However, it does not hold for every field replacing $\mathbb{Q}$, e.g., it does not hold for $\mathbb{C}$ (although it holds for $\mathbb{C}(t)$).

We now apply the Hilbert Irreducibility Theorem to show that S_n is a Galois group over $\mathbb{Q}$.

Theorem 63.16. *Let n be any positive integer. Then there exists a finite Galois extension $L/\mathbb{Q}$ with $G(L/\mathbb{Q}) \cong S_n$.*

PROOF. Let $K = \mathbb{Q}(t_1, \ldots, t_n)$ and S_n act on K by $\sigma(t_i) = t_{\sigma(i)}$ for $i = 1, \ldots, n$. Set $F = K^{S_n}$. By Example 54.6(3), we know that $F = \mathbb{Q}(s_1, \ldots, s_n)$ with each s_i is the ith elementary symmetric function

of $t_1, \ldots, t_n$, the extension K/F is a finite Galois extension with Galois group $G(K/F) \cong S_n$, and K is a splitting field of

$$f = \prod_{i=1}^{n}(t - t_i) = t^n - s_1 t^{n-1} + \cdots + (-1)^n s_n.$$

By Remark 54.10 to the Primitive Element Theorem, we can find distinct integers $m_1, \ldots, m_n$ satisfying $K = F(m_1 t_1 + \cdots + m_n t_n)$.

Let $c = m_1 t_1 + \cdots + m_n t_n$. Then the $\sigma(c)$, $\sigma \in G(K/F)$ ($\cong S_n$), give $n!$ distinct elements in K. Let $g = \prod_{G(K/F)} \big(t - \sigma(c)\big) \in K[t]$. Since $\sigma(g) = g$ for all $\sigma \in G(K/F)$, we know that $g \in F[t]$ and $K = F(c)$ is a splitting field of g over F. In particular, as $\deg g = [K : F]$ and $m_F(c) \mid g$ in $F[t]$, we must have $g = m_F(c)$ is irreducible in $F[t]$. Since $s_1, \ldots, s_n$ are algebraically independent, $g(s_1, \ldots, s_n, t)$ is irreducible in $\mathbb{Q}[s_1, \ldots, s_n, t]$, so we can choose $\alpha_1, \ldots, \alpha_n$ in $\mathbb{Q}$, such that $g(\alpha_1, \ldots, \alpha_n, t)$ is irreducible in $\mathbb{Q}[t]$ by the Hilbert Irreducibility Theorem 63.15, i.e, we apply the evaluation map $e_{\alpha_1,\ldots,\alpha_n} : \mathbb{Q}[s_1, \ldots, s_n] \to \mathbb{Q}$ with $s_i \mapsto \alpha_i$ for each i. Let $\beta_1, \ldots, \beta_n$ be the roots of the polynomial $f(\alpha_1, \ldots, \alpha_n, t)$ in $Q[t]$ in its splitting field L in $\mathbb{C}$. Then

$$f(\alpha_1, \ldots, \alpha_n, t) = \prod_{i=1}^{n}(t - \beta_i) = t^n - \alpha_1 t^{n-1} + \cdots + (-1)^n \alpha_n$$

in $\mathbb{Q}[t]$. Let $e = m_1\beta_1 + \cdots + m_n\beta_n$ in $L = \mathbb{Q}(\beta_1, \ldots, \beta_n)$. As $c = m_1 t_1 + \cdots + m_n t_n$ is a root of g with $t_1, \ldots, t_n$ the roots of f, we must have e is a root of the irreducible polynomial $g(\alpha_1, \ldots, \alpha_n, t)$. Therefore,

$$|G(L/\mathbb{Q})| = [L : \mathbb{Q}] \geq \deg m_{\mathbb{Q}}(e) = \deg g(\alpha_1, \ldots, \alpha_n, t) = n!.$$

It follows that $G(L/\mathbb{Q}) \cong S_n$, since $\deg f = n$ and we can view $G(L/F) \subset S_n$. □

Corollary 63.17. *There exists an irreducible polynomial of every degree $n \geq 5$ not solvable by radicals.*

Corollary 63.18. *Let $m \geq 2$. Then there exists an element algebraic of degree 2^m over $\mathbb{Q}$ that is not constructible.*

Proof. Let f be an irreducible polynomial of degree n in $\mathbb{Q}[t]$ with Galois group S_n with $n = 2^m$ and $K = F(\alpha)$ a splitting field of f over $\mathbb{Q}$. If α was constructible (from $z = 0, z = 1$), then $n! = [K : \mathbb{Q}] = 2^e$ for some e by the refined form of the Constructibility Criterion 57.3, which is impossible □

Exercises 63.19. 1. Let F be a field of characteristic zero and g and h two polynomials in $F[t]$. Show if the (formal) derivative $(g-h)' = 0$, then $h = g + \big(h(0) - g(0)\big)$.

2. Let F be a field of characteristic zero, h be a polynomial of degree n in $F[t]$ and t_0 an element of F. Show that the (formal) Taylor expansion of h exists at t_0, i.e.,

$$h(t) = \sum_{i=0}^{n} \frac{h^{(i)}}{i!}(t - t_0)^i$$

where $h^{(i)}$ is the ith derivative of h.

3. Let $f(x,t)$ be an irreducible polynomial in $\mathbb{Q}[x,t]$, Show that the set $\{\alpha \in \mathbb{Q} \mid f(\alpha,t) \in \mathbb{Q}[t] \text{ is irreducible}\}$ is dense in $\mathbb{R}$.

Part 6

Additional Topics

CHAPTER XIII

Transcendental Numbers

64. Liouville Numbers

Although we have seen that there are uncountably many real numbers transcendental over $\mathbb{Q}$, it is usually quite difficult to prove a particular real number that is not easily seen to be algebraic over $\mathbb{Q}$ is transcendental. In this section, we demonstrate that certain numbers are in fact transcendental over $\mathbb{Q}$. The first numbers shown to be transcendental over $\mathbb{Q}$ were constructed by Liouville. His approach depends on the following theorem:

Theorem 64.1. (Liouville) *Let α be a complex number algebraic over $\mathbb{Q}$ of degree $n > 1$. Then there exists a positive constant $c = c(\alpha)$ satisfying*

$$|\alpha - \frac{p}{q}| > \frac{c}{|q|^n} \quad \textit{for all integers } p \textit{ and } q \textit{ with } q \textit{ nonzero.}$$

Proof. By definition, $\deg m_{\mathbb{Q}}(\alpha) = n$. Clearing denominators, we see that there exists a nonzero integer a such that $am_{\mathbb{Q}}(\alpha)$ lies in $\mathbb{Z}[t]$. Dividing $am_{\mathbb{Q}}(\alpha)$ by its content, we see that there exists an irreducible polynomial f in $\mathbb{Z}[t]$ of degree n with $f(\alpha) = 0$. As $m_{\mathbb{Q}}(\alpha)$ has no rational roots, neither does f. (Of course, the ideals $\big(m_{\mathbb{Q}}(\alpha)\big)$ and (f) are the same in $\mathbb{Q}[t]$.) As the characteristic of $\mathbb{Q}$ is zero,

$$f'(\alpha) = \lim_{z\to\alpha} \Big|\frac{f(z) - f(\alpha)}{z - \alpha}\Big| = \lim_{z\to\alpha} \Big|\frac{f(z)}{z - \alpha}\Big|$$

is nonzero, equivalently

$$\lim_{z\to\alpha} \frac{\big|\frac{f(z)}{z-\alpha}\big|}{|f'(\alpha)|} = 1.$$

Choose $\delta > 0$ such that if $0 < |z - \alpha| \le \delta$, then

$$\Big|\frac{f(z)}{z - \alpha}\Big| < 2|f'(\alpha)|, \quad \text{so} \quad |f(z)| < 2|f'(\alpha)||z - \alpha|.$$

Let p and q be integers with q nonzero. Suppose that $0 < |\frac{p}{q} - \alpha| \leq \delta$ (as α is not rational), then

$$|f(\frac{p}{q})| < 2|f'(\alpha)||\frac{p}{q} - \alpha|.$$

Since $\deg f = n$ and $f(\frac{p}{q})$ is nonzero, $f(\frac{p}{q}) = \frac{b}{q^n}$ for some nonzero integer b. It follows, if $|\frac{p}{q} - \alpha| \leq \delta$, then

$$\frac{1}{|q^n|} \leq |f(\frac{p}{q})| < 2|f'(\alpha)||\frac{p}{q} - \alpha|,$$

i.e.,

$$|\frac{p}{q} - \alpha| > \frac{1}{2|f'(\alpha)|}\frac{1}{|q^n|}.$$

Now suppose that $|\frac{p}{q} - \alpha| > \delta$, then $\delta \geq \frac{\delta}{|q^n|}$, so $|\frac{p}{q} - \alpha| > \frac{\delta}{|q^n|}$. Set $c := \min\{\delta, \frac{1}{2|f'(\alpha)|}\}$, then $|\frac{p}{q} - \alpha| > \frac{c}{|q^n|}$ for all nonzero integers q. □

Let $\alpha \in \mathbb{C}$ be a nonrational number. We say that α is a *Liouville number* if there exist no pair of real numbers $c > 0$ and $n \geq 2$ satisfying

$$|\frac{p}{q} - \alpha| > \frac{c}{|q^n|} \text{ for all integers } p \text{ and } q \text{ with } q \text{ nonzero.}$$

Note that if α is a Liouville number, it must be real, since a nonreal complex number has positive distance from any rational number.

Corollary 64.2. *Liouville numbers are transcendental over* $\mathbb{Q}$.

Remark 64.3. Let α be a Liouville number, N a positive integer. Then there exist infinitely many rational numbers $\frac{p}{q}$, with p, q with $q \neq 0$ satisfying

$$0 < |\frac{p}{q} - \alpha| < \frac{1}{|q^N|} \text{ for all integers } p \text{ and } q \text{ with } q \text{ nonzero.}$$

Indeed suppose this is false, then we could choose a real number $c > 0$ smaller than the distance of α to each of these finitely many rational numbers such that

$$0 < |\frac{p}{q} - \alpha| < \frac{c}{|q^N|} \text{ holds for no integers } p \text{ and } q \text{ with } q \text{ nonzero.}$$

We use this to give an alternative characterization of Liouville numbers.

Proposition 64.4. *Let α be a real number. Then α is a Liouville number if and only if for every positive integer N there exist integers p and q, with $q \geq 2$ satisfying*

$$(*) \qquad 0 < |\frac{p}{q} - \alpha| < \frac{1}{q^N}.$$

PROOF. ($\Rightarrow$) is the remark above.

($\Leftarrow$): Let $c > 0$ and $n \geq 2$ be given. Choose an integer m so that $\frac{1}{2^m} < c$ and set $N = n + m$. By assumption, there exist integers p and q, with $q \geq 2$ satisfying

$$|\frac{p}{q} - \alpha| < \frac{1}{q^N} \leq \frac{1}{2^m q^n} < \frac{c}{q^n}.$$

By the theorem, it follows that α is a Liouville number if we can show that α is not a rational number. Suppose to the contrary that $\alpha = r/s$, with r and s integers, s positive. We may assume that we have chosen N so that it satisfies $2^{N-1} > s$. So by assumption, there exist integers p and q, with $q \geq 2$ with satisfying (*). However, we also have

$$|\frac{p}{q} - \alpha| = |\frac{p}{q} - \frac{r}{s}| = |\frac{ps - qr}{qs}| > \frac{1}{sq} > \frac{1}{2^{N-1}q} \geq \frac{1}{q^N},$$

a contradiction. □

Example 64.5. Let $\alpha = \sum_{k=1}^{\infty} \frac{1}{10^{k!}}$, then α is a Liouville number:

Let $\sum_{k=1}^{N} \frac{1}{10^{k!}} = \frac{P_N}{10^{N!}}$, so P_N and $10^{N!}$ are positive integers satisfying

$$\begin{aligned} 0 < |\alpha - \frac{P_N}{10^{N!}}| &= \sum_{k=N+1}^{\infty} \frac{1}{10^{k!}} \leq \frac{1}{10^{(N+1)!}} \sum_{k=0}^{\infty} \frac{1}{10^{k!}} \\ &\leq \frac{10}{9}\left(\frac{1}{(10^{N!})^{N+1}}\right) < \frac{1}{(10^{N!})^N}. \end{aligned}$$

The basic reason that the example is transcendental over $\mathbb{Q}$ is that the series converges very rapidly. Liouville numbers have this property. Of course, most transcendental numbers do not, e.g., π. Mahler showed that $|\pi - \frac{p}{q}| > q^{-42}$ if $q \geq 2$.

Much work was done to strengthen Liouville's theorem. The best theorem was established by Roth who proved the following

Theorem 64.6. (Roth's Theorem) *Let α be a complex number algebraic over the rational numbers but not rational. Suppose that there exists a non-negative real number μ satisfying*

$$|\alpha - \frac{p}{q}| < \frac{1}{q^\mu}$$

for infinitely many p/q with p, q relatively prime integers $q > 0$. Then $\mu \leq 2$. Moreover, this is best possible, i.e., if α is algebraic over $\mathbb{Q}$ and

$\epsilon > 0$, then

$$|\alpha - \frac{p}{q}| > \frac{1}{q^{2+\epsilon}}$$

holds for all but finitely many p/q with p, q relatively prime integers $q > 0$, and there exists a positive constant $c = c(\alpha, \epsilon)$ such that

$$|\alpha - \frac{p}{q}| > \frac{c}{|q|^{2+\epsilon}} \text{ for all integers } p \text{ and } q \text{ with } q \text{ nonzero},$$

i.e., is independent of $n > 0$.

Roth won the Fields Medal for this work. This subject is part of what is called diophantine approximation. The proof is delicate, but does not use a lot of theory. A proof can be found in Leveque's book *Topics in Number Theory, Volume II*, Chapter 4.

Exercises 64.7. 1. Let $b_0 < b_1 < b_2 < \cdots$ an increasing sequence of positive integers satisfying $\limsup_{k\to\infty} b_{k+1}/b_k = \infty$ and let $\{e_k\}$ be a bounded sequence of positive integers. If $a \geq 2$ is an integer, show

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{a^{b_k}} \quad \text{and} \quad \sum_{k=0}^{\infty} \frac{e_k}{a^{b_k}}$$

are Liouville numbers.

2. For any choice of signs, show that

$$1 \pm \frac{1}{2^{1!}} \pm \frac{1}{2^{2!}} \pm \frac{1}{2^{3!}g} \pm \cdots$$

is a Liouville number.

65. Transcendence of e

In this section we prove that the real number e is transcendental over $\mathbb{Q}$. The ideas give the basis for showing that π is transcendental, which we do in the following section. We give Hermite's proof.

Theorem 65.1. *The real number e is transcendental over $\mathbb{Q}$.*

Proof. Suppose that there exist integers $a_0, \ldots, a_m$ not all zero satisfying

$$a_m e^m + a_{m-1} e^{m-1} + \cdots + a_1 e + a_0 = 0,$$

i.e., e is a root of the nonzero polynomial $\sum_{i=0}^{m} a_i t^i$ in $\mathbb{Z}[t]$, equivalently, e is algebraic over $\mathbb{Q}$, by clearing denominators. We may assume that a_m is nonzero and without loss of generality that a_0 is nonzero as well. Define a *polynomial function*

$$f(x) := \frac{x^{p-1}(x-1)^p(x-2)^p \cdots (x-m)^p}{(p-1)!}$$

and set

$$F(x) := f(x) + f'(x) + f^{(2)}(x) + \cdots + f^{(mp+p-1)}(x)$$

where $p > 2$ is a prime number to be specified later and $f^{(j)}$ is the jth derivative of f.
Note that $\deg f = mp + p - 1$, so the $(mp + p)$th derivative of f, $f^{(mp+p)}(x) = 0$. If $0 < x < m$, then we have

$$|f(x)| < \frac{m^{p-1}m^p \cdots m^p}{(p-1)!} = \frac{m^{mp+p-1}}{(p-1)!}. \tag{65.2}$$

Check 65.3. We have the following:

$$\frac{d}{dx}\big(e^{-x}F(x)\big) = e^{-x}\big(F'(x) - F(x)\big) = -e^{-x}f(x) \tag{1}$$

$$a_j \int_0^j e^{-x}f(x)dx = a_j\big[-e^{-x}F(x)\big]_{x=0}^{x=j} = a_jF(0) - a_je^{-j}F(j). \tag{2}$$

Multiplying equation (2) by e^j and adding yields

$$\begin{aligned} A :&= \sum_{j=0}^{m} a_je^j \int_0^j e^{-x}f(x)dx = \sum_{j=0}^{m} a_je^jF(0) - \sum_{j=0}^{m} a_jF(j) \\ &= -\sum_{j=0}^{m} a_jF(j) = -\sum_{j=0}^{m}\sum_{i=0}^{mp+p-1} a_jf^{(i)}(j). \end{aligned} \tag{65.4}$$

Claim: $f^{(i)}(j)$ is an integer for all i and j and $p \mid f^{(i)}(j)$ in $\mathbb{Z}$ if $(i, j) \neq (p-1, 0)$.
The claim will follow from the following:

Lemma 65.5. *Let $h(x) = \dfrac{x^n g(x)}{n!}$ with g a polynomial in $\mathbb{Z}[t]$. Then*

(1) The number $h^{(j)}(0)$ is an integer for all j.
(2) The integer $n + 1$ satisfies $n + 1 \mid h^{(j)}(0)$ in $\mathbb{Z}$ if $j \neq n$.
(3) If $g(0) = 0$, then $n + 1 \mid h^{(n)}(0)$ in $\mathbb{Z}$.

PROOF. Let $t^n g(t) = \sum_{j=0}^{\infty} c_jt^j$ in $\mathbb{Z}[t]$ (almost all $c_j = 0$). So we have c_j are integers and zero if $j < n$. Since $h(x) = \dfrac{x^n g(x)}{n!}$, we see that $h^{(j)}(0) = c_jj!/n!$. If $j < n$, then $c_j = 0$. Hence $h^{(j)}(0) = 0$ and if $j > n$, then $h^{(j)}(0) = c_j(j!/n!)$ in $\mathbb{Z}$. If $j = n$, then $h^{(n)}(0) = c_n$ in $\mathbb{Z}$. Finally, if g has a zero constant term, i.e., $g(0) = 0$, then $c_n = 0$ and the lemma follows. □

We return to the proof of the theorem. We apply the lemma to $f(x), f(x+1), \ldots, f(x+m)$, respectively to see that $f^{(i)}(j)$ is an integer for all i and j and $p \mid f^{(i)}(j)$ with the possible exception when $i = p-1$ and $j = 0$. Therefore, as a_i are integers, we have

$$A = -\sum_{j=0}^{m} \sum_{i=0}^{mp+p-1} a_j f^{(i)}(j) \equiv -a_0 f^{(p-1)}(0) \mod p.$$

We also have $f^{(p-1)}(0) = (-1)^p(-2)^p \cdots (-m)^p$. (This is just $g(0)$, if in the lemma, $g(x) = (x-1)^p \cdots (x-m)^p$.) In particular, if $p > m$, then $p \nmid f^{(p-1)}(0)$. Consequently, as $a_0 \neq 0$, we must have: if p is a prime satisfying $p > |a_0|$ and $p > m$, then

$$A \equiv -a_0 f^{(p-1)}(0) \not\equiv 0 \mod p.$$

In particular, $|A| \geq 1$. By (65.2) and (65.4), we have

$$\begin{aligned} |A| &= \Big| \sum_{j=0}^{m} a_j e^j \int_0^j e^{-x} f(x) dx \Big| \leq \sum_{j=0}^{m} \Big| a_j e^j \int_0^j f(x) dx \Big| \\ &\leq \sum_{j=0}^{m} |a_j| e^j j \frac{m^{mp+p-1}}{(p-1)!} \leq \sum_{j=0}^{m} |a_j| e^m m \frac{m^{mp+p-1}}{(p-1)!} \\ &= \Big(\sum_{j=0}^{m} |a_j| \Big) e^m \frac{m^{mp+p}}{(p-1)!} = \Big(\sum_{j=0}^{m} |a_j| \Big) e^m \frac{(m^{m+1})^p}{(p-1)!}. \end{aligned}$$

Choosing p sufficiently large leads to $|A| < 1$, a contradiction. □

Exercise 65.6. Verify the two statements in Check 65.3.

66. Symmetric Functions

We discussed symmetric functions in the study of Galois Theory. In order to prove that π is transcendental, we shall need a finer discussion. This material is very important in its own right and is fundamental in the study of polynomials as well as its various generalizations.

We first establish some nomenclature for polynomials in finitely many variables over a commutative ring. Let R be a commutative ring and f a nonzero polynomial in $R[t_1, \ldots, t_n]$. Write

$$f = \sum_{i_1, \ldots, i_n} a_{i_1, \ldots, i_n} t_1^{i_1} \cdots t_n^{i_n}.$$

Each nonzero term $a_{i_1, \ldots, i_n} t_1^{i_1} \cdots t_n^{i_n}$ in f is called a *monomial* of *total degree* $i_1 + \cdots + i_n$ of f in $t_1, \ldots, t_n$ and the *total degree* of f in $t_1, \ldots, t_n$

is defined to be the maximum of the total degree of the nonzero monomials of f in $t_1, \ldots, t_n$. For each positive integer k, order the subset

$$\{at_1^{i_1} \cdots t_n^{i_n} \mid a \text{ nonzero in } R \text{ with } k = i_1 + \cdots + i_n\}$$

of $R[t_1, \ldots, t_n]$ *lexicographically*, i.e., if $a, b \in R$ are nonzero and $k = i_1 + \cdots + i_n = j_1 + \cdots + j_n$, then

$$at_1^{i_1} \cdots t_n^{i_n} \leq bt_1^{j_1} \cdots t_n^{j_n}$$

if

$i_l = j_l$ for all l or

there exists an l satisfying $i_s = j_s$ for all $s \leq l$ and $i_{l+1} < j_{l+1}$.

For example,

$$t_1 t_2 t_3^2 < t_1^3 t_2 t_3 \text{ and } t_2^{27} < t_1 t_2^{27}$$

in $R[t_1, t_2, t_3]$ where $<$ of course means $\leq$ but not $=$. We define the *leading term* of a nonzero f in $R[t_1, \ldots, t_n]$ to be the nonzero monomial of f that has maximal total degree relative to the lexicographic ordering on $R[t_1, \ldots, t_n]$. In particular, the total degree of the leading term is the same as the total degree of f, although there can be many nonzero monomials of f of the same total degree. The coefficient of the leading term is called the *leading coefficient*.

Definition 66.1. Let R be a commutative ring. A polynomial f in $R[t_1, \ldots, t_n]$ is called *symmetric* in $t_1, \ldots, t_n$ if for all permutations $\sigma \in S_n$, we have $f(t_1, \ldots, t_n) = f(t_{\sigma(1)}, \ldots, t_{\sigma(n)})$.

Example 66.2. The following are symmetric in $R[t_1, \ldots, t_n]$:

1. Any element of R.
2. $t_1^r + \cdots + t_n^r$ for every positive integer r.
3. $t_1 t_2 + t_1 t_3 + \cdots + t_1 t_n + t_2 t_3 + \cdots + t_{n-1} t_n = \sum_{i<j} t_i t_j$
4. $\displaystyle\sum_{1 \leq i_1 < i_2 < \cdots i_r \leq n} t_{i_1}^s \cdots t_{i_r}^s$ for each $1 \leq r \leq n$ and each positive integer s.

The rth *elementary symmetric function* in $t_1, \ldots, t_n$ is defined to be

$$s_r(t_1, \ldots, t_n) = \sum_{1 \leq i_1 < i_2 < \cdots i_r \leq n} t_{i_1} \cdots t_{i_r}.$$

For example,

$$s_3(t_1, t_2, t_3, t_4) = t_1 t_2 t_3 + t_1 t_2 t_4 + t_1 t_3 t_4 + t_2 t_3 t_4 \text{ and}$$
$$s_4(t_1, t_2, t_3, t_4) = t_1 t_2 t_3 t_4.$$

We shall write s_r for $s_r(t_1, \ldots, t_n)$ if n is clear and we set $s_0 = 1$.

Observation 66.3. Let R be a commutative ring, $S = R[t_1, \ldots, t_n]$, and $s_i(t_1, \ldots, t_n)$ in S, then in $S[t]$, we have

$$f := \prod(t + t_i) := t^n + s_1 t^{n-1} + \cdots + s_n \text{ and}$$
$$g := \prod(t - t_i) := t^n - s_1 t^{n-1} + \cdots + (-1)^n s_n$$

with $-t_1, \ldots, -t_n$ (respectively, $t_1, \ldots, t_n$) the roots of f (respectively, of g) in S. Set $S_0 = R[s_1, \ldots, s_n] \subset S = R[t_1, \ldots, t_n]$. Then the subset $S^{S_n} := \{f \in S \mid f \text{ symmetric in } t_1, \ldots, t_n\}$ satisfies $S_0 \subset S^{S_n} \subset S$. We shall show that $S_0 = S^{S_n}$.

Theorem 66.4. (Fundamental Theorem of Symmetric Functions) *Let R be a commutative ring, $s_i = s_i(t_1, \ldots, t_n)$ in $R[t_1, \ldots, t_n]$. Then*

$$R[s_1, \ldots, s_n] = R[t_1, \ldots, t_n]^{S_n} := \{f \in S \mid F \text{ symmetric in } t_1, \ldots, t_n\}.$$

More specifically, let $f \in R[t_1, \ldots, t_n]$ be symmetric in $t_1, \ldots, t_n$ of total degree k in $t_1, \ldots, t_n$. Then there exists a unique polynomial g in $R[s_1, \ldots, s_n]$ satisfying $f = g$ and the total degree of g in $s_1, \ldots s_n$ is at most k.

PROOF. **Existence**: Let $f \in R[t_1, \ldots, t_n]$ be symmetric in $t_1, \ldots, t_n$ of total degree k and leading term $at_1^{i_1} \ldots t_n^{t_n}$ in the lexicographical ordering. In particular, $k = i_1 + \cdots + i_n$. As f is symmetric, $at_{\sigma(1)}^{i_1} \ldots t_{\sigma(n)}^{i_n}$ is also a nonzero monomial in f for every σ in S_n. It follows that $i_1 \geq i_2 \geq \cdots \geq i_n$. Set

$$j_l := i_l - i_{l-1} \leq i_l \text{ for } l = 1, \ldots, n$$

where $i_{n+1} = 0$. Therefore, $j_l \geq 0$ and $i_l = j_l + j_{l+1} + \cdots j_n$ for every l. The monomial $as_1^{j_1} \cdots s_n^{j_n}$ has total degree in $t_1, \ldots, t_n$ equal to

$$j_1 + 2j_2 + \cdots + nj_n = i_1 + i_2 + \cdots + i_n = k$$

and total degree in $s_1, \ldots, s_n$ equal to

$$j_1 + j_2 + \cdots + j_n \leq i_1 + i_2 + \cdots + i_n = k.$$

The leading term of $as_1^{j_1} \cdots s_n^{j_n}$ as a polynomial in $t_1, \ldots, t_n$ is

$$at_1^{j_1}(t_1t_2)^{j_2}(t_1t_2t_3)^{j_3} \cdots (t_1 \cdots t_n)^{j_n} = at_1^{i_1} \ldots t_n^{i_n},$$

so $f_1 := f - as_1^{j_1} \cdots s_n^{j_n}$ has leading term in $t_1, \ldots, t_n$ less than $as_1^{j_1} \cdots s_n^{j_n}$ in the lexicographic ordering. Iterating the process rids us of all monomials of total degree k in $t_1, \ldots, t_n$ except for monomials in $s_1, \ldots, s_n$ of total degree at most k in $s_1, \ldots, s_n$ that when viewed as polynomials in $t_1, \ldots, t_n$ have total degree k. Induction on the total degree of f produces the existence of the required g.

Uniqueness: Suppose that f has two such expressions. Then we would have an equation

$$\sum_{i_1,\dots,i_n} a_{i_1,\dots,i_n} s_1^{i_1} \cdots s_n^{i_n} = 0$$

not all $a_{i_1,\dots,i_n}$ zero. Let $as_1^{j_1} \cdots s_n^{j_n}$ be the leading term of this expression in $s_1, \dots, s_n$. Writing this term in $t_1, \dots, t_n$ would produce a leading term in $t_1, \dots, t_n$

$$at_1^{i_1} \cdots t_n^{i_n} \text{ with } j_l = i_l - i_{l+1},\ i.e.,\ i_l = j_l + \cdots + j_n$$

just as before. As the t_i are indeterminants, we would have $a = 0$, a contradiction. □

We deduce the immediate specialization of this result.

Corollary 66.5. *Let R be a commutative ring and $f \in R[t_1, \dots, t_n]$ a symmetric polynomial. If $\alpha_1, \dots, \alpha_n$ are elements in R and s_i the elementary symmetric functions in $t_1, \dots, t_n$, then the evaluation $e_{\alpha_1,\dots,\alpha_n}$ of f, $f(\alpha_1, \dots, \alpha_n)$, lies in $R[s_1(\alpha_1, \dots, \alpha_n), \dots, s_n(\alpha_1, \dots, \alpha_n)]$.*

We shall need the following consequence of this.

Corollary 66.6. *Let F be a field and f in $F[t]$ a polynomial of degree n with roots $\alpha_1, \dots, \alpha_n$ in F. Suppose that $P \in F[t_1, \dots, t_n]$ is symmetric in $t_1, \dots, t_n$, then $P(\alpha_1, \dots, \alpha_n)$ lies in F.*

Proof. Using Observation 66.3, we know if $f = a_n t^n + \cdots + a_0$ in $F[t]$, then $\pm s_i(\alpha_1, \dots, \alpha_n) = a_i/a_n$ lies in F. The result follows. □

67. Transcendence of π

The proof of the transcendence of π over the rationals takes the approach of the analogous statement for e, but it is significantly harder. In particular, we shall need the Fundamental Theorem of Symmetric Functions.

We begin with the observation that a function $f : \mathbb{Z} \to \mathbb{Z}$ satisfying $f(p) \to 0$ as $p \to \infty$ must satisfy $f(p) = 0$ for large enough primes, since $|f(p)| < 1$ for all primes large enough.

We give Hilbert's proof of the following famous result.

Theorem 67.1. (Lindemann) *The real number π is transcendental over $\mathbb{Q}$.*

Proof. Suppose that π is algebraic over $\mathbb{Q}$. Then $\alpha_1 := \sqrt{-1}\pi$ is also algebraic over $\mathbb{Q}$. We have

(i) (Euler) $$e^{\alpha_1} + 1 = 0.$$

Let $g_1 = m_{\mathbb{Q}}(\alpha_1)$ in $\mathbb{Q}[t]$. Supposes that $\deg g_1 = n$ and

$$\alpha_1, \ldots, \alpha_n \text{ are all the (distinct) roots of } g_1 \text{ in } \mathbb{C}.$$

Then by (i), we have

(ii) $$\prod_{i=1}^{n}(e^{\alpha_i} + 1) = 0.$$

Expanding (ii) yields

(iii) $$\Big(\sum_{1\le r\le n}\ \sum_{1\le i_1<\cdots<i_r\le n} e^{\alpha_{i_1}+\cdots+\alpha_{i_r}}\Big) + 1 = 0.$$

Let $s_j(t_1, \ldots, t_n) := \sum_{1\le i_1<\cdots<i_j\le n} t_{i_1}\cdots t_{i_j}$ be the jth elementary polynomial in $\mathbb{C}[t_1, \ldots, t_n]$. Then the monic polynomial

$$\prod_{i=1}^{n}(t - \alpha_i) = t^n - s_1(\alpha_1, \ldots, \alpha_n)t^{n-1} + \cdots + (-1)^n s_n(\alpha_1, \ldots, \alpha_n)$$

in $\mathbb{C}[t]$ has precisely $\alpha_1, \ldots, \alpha_n$ as roots. In particular, this polynomial is just g_1. It follows that each $s_j(\alpha_1, \ldots, \alpha_n)$ lies in $\mathbb{Q}$.

Fix r with $1 \le r \le n$ and let

$$T_{i_1,\ldots,i_r}, \quad 1 \le i_1 < \cdots < i_r, \quad 1 \le i \le k := \binom{n}{r}$$

be independent variables and s'_j be the elementary symmetric functions in the $T_{i_1,\ldots,i_r}$ (ordered lexicographically) evaluated at the corresponding $\alpha_{i_1} + \cdots + \alpha_{i_r}$. Set

$$g_r := t^k - s'_1 t^{k-1} + \cdots + (-1)^k s'_k, \text{ a polynomial in } \mathbb{C}[t].$$

Therefore, g_r has roots

$$\alpha_{i_1} + \cdots + \alpha_{i_r} \text{ for } 1 \le i_1 < \cdots < i_r \le n.$$

We can view the evaluation of the elementary symmetric functions in the $T_{i_1,\ldots,i_r}$ in two steps:

- (a) Evaluate these elementary symmetric functions in the $T_{i_1,\ldots,i_r}$ at $t_{i_1} + \cdots + t_{i_r}$ for each $1 \le i_1 < \cdots < i_r \le n$.
- (b) Evaluate the resulting polynomials in (a) at the corresponding $\alpha_{i_1} + \cdots + \alpha_{i_r}$.

Check 67.2. The polynomials in (a) above are symmetric in $t_1, \ldots, t_n$

It follows by Corollary 66.6, the corollary to the Fundamental Theorem of Symmetric Functions 66.4, that the polynomials in (a) lie in $\mathbb{Z}[s_1(t_1,\dots,t_n),\dots,s_n(t_1,\dots,t_n)]$ and

$$s'_j \in \mathbb{Z}[s_1(\alpha_1,\dots,\alpha_n),\dots,s_n(\alpha_1,\dots,\alpha_n)] \subset \mathbb{Q}.$$

Consequently, $g_r \in \mathbb{Q}[t]$ for $r = 1,\dots,n$.

Claim: There exists a polynomial g in $\mathbb{Z}[t]$ having as its roots in $\mathbb{C}$ precisely all the nonzero $\alpha_{i_1}+\cdots+\alpha_{i_r}$, $1 \le i_1 < \cdots < i_r \le n$, $1 \le r \le n$:

Let $\widetilde{g} = g_1 \dots g_n$, a polynomial in $\mathbb{Q}[t]$. We know that $\widetilde{g}$ has as its roots all $\alpha_{i_1}+\cdots+\alpha_{i_r}$, $1 \le i_1 < \cdots < i_r \le n$, $1 \le r \le n$. Let

$$M := \text{precisely the number of } (i_1,\dots,i_r),\ 1 \le i_1 < \cdots < i_r \le n$$
$$1 \le r \le n, \text{ satisfying } \alpha_{i_1}+\cdots+\alpha_{i_r} = 0.$$

Therefore, the non-negative integer M satisfies $t^M \mid \widetilde{g}$, but $t^{M+1} \nmid \widetilde{g}$ in $\mathbb{Q}[t]$. Let $\widehat{g} := \widetilde{g}/t^M$ in $\mathbb{Q}[t]$. Consequently, $\widehat{g}$ has as roots, precisely those $\alpha_{i_1}+\cdots+\alpha_{i_r}$, $1 \le i_1 < \cdots < i_r \le n$, $1 \le r \le n$, that are nonzero. Let m be the least common multiple of all the denominators of the (reduced) coefficients of $\widehat{g}$ (i.e., if $\widehat{g} = \sum \frac{a_i}{b_i} t^i$ with a_i, b_i relatively prime integers ($b_i \neq 0$) if $a_i \neq 0$ (and $b_i = 1$ otherwise), then m is the least common multiple of the b_i's). Set $g = m\widehat{g}$, a polynomial in $\mathbb{Z}[t]$ that satisfies the claim.

Let $r = \deg g$ and rename the roots of g by

$$\beta_1,\dots,\beta_r \text{ not necessarily distinct.}$$

By choice, all the β_i are nonzero, so (iii) becomes

(iv) $\quad e^{\beta_1}+e^{\beta_2}+\cdots+e^{\beta_r}+k = 0$, for some positive integer k.

Write

$$g = ct^r + c_1 t^{r-1} + \cdots + c_r, \text{ a polynomial in } \mathbb{Z}[t].$$

As $\deg g = r$ and $g(0)$ is not zero, we have

$$c \text{ and } c_r \text{ are nonzero integers.}$$

Now fix a prime p and set

$$s = rp - 1.$$

Define

$$f = \frac{c^s t^{p-1} g^p(t)}{(p-1)!}$$
$$F = f + f' + f^{(2)} + \cdots + f^{(s+p)}$$

where $f^{(j)}$ is the jth derivative of f.

Note that $\deg f = p - 1 + pr = p + s$, so $f^{(s+p+1)} = 0$.
It follows that

$$\frac{d}{dt}\big(e^{-t}F(t)\big) = -e^{-t}f(t) \;\text{ so }\; e^{-t}F(t) - F(0) = -\int_0^t e^{-x}f(x)dx.$$

Multiply this equation by e^t and let $x = \lambda t$ (λ is a new variable), to get

$$F(t) - e^tF(0) = t\int_0^1 e^{(1-\lambda)t}f(\lambda t)d\lambda. \tag{v}$$

Setting $t = \beta_1, \ldots, \beta_r$ (one at a time) into (v) and summing, we get, using (iv),

$$\sum_{i=1}^r F(\beta_i) + kF(0) = \sum_{i=1}^r \beta_i \int_0^1 e^{(1-\lambda)\beta_i}f(\beta_i\lambda)d\lambda. \tag{vi}$$

We shall show that this leads to a contradiction by analyzing

$$A = \sum_{i=1}^r F(\beta_i) + kF(0), \text{ the left hand side of (vi), and}$$

$$B = \sum_{i=1}^r \beta_i \int_0^1 e^{(1-\lambda)\beta_i}f(\beta_i\lambda)d\lambda, \text{ the right hand side of (vi).}$$

Claim: If $p >> 0$, then A is a non-zero integer.
To prove this, we have to first evaluate $\sum_{i=1}^r f^{(j)}(\beta_i)$. We show

$$\sum_{i=1}^r f^{(j)}(\beta_i) \text{ is an integer.} \tag{a}$$

$$p \mid \sum_{i=1}^r f^{(j)}(\beta_i) \text{ in } \mathbb{Z}: \tag{b}$$

As $g(\beta_i) = 0$ for all i, all terms in

$$\frac{d^j}{dt^j}\Big|_{t=\beta_i}\Big(\frac{c^st^{p-1}g^p(t)}{(p-1)!}\Big) \tag{$\star$}$$

having a g^l term with $l > 0$ are zero. In particular, (a) and (b) hold if $j < p$.

So we may assume that $j \geq p$.

Check 67.3. $\dfrac{1}{pc^s}f^{(j)}(t)$ lies in $\mathbb{Z}[t]$ for $j \geq p$. (Cf. Example 2.11.)

Therefore, $\dfrac{1}{pc^s}\displaystyle\sum_{i=1}^{r} f^{(j)}(t_i) \in \mathbb{Z}[t_1,\dots,t_r]$ is symmetric in $t_1,\dots,t_r$ of total degree $(p-1)+rp-j = s+p-j \le s$ in $t_1,\dots,t_r$ as $j \ge p$. By the Fundamental Theorem of Symmetric Functions 66.4,

$$(\dagger) \qquad \frac{1}{pc^s}\sum_{i=1}^{r} f^{(j)}(t_i) \text{ lies in } \mathbb{Z}[s_1(t_1,\dots,t_r),\dots,s_r(t_1,\dots,t_r)]$$

and is of total degree in $s_1(t_1,\dots,t_r),\dots,s_r(t_1,\dots,t_r)$ at most $(p-1+pr)-j = s+p-j \le s$.
We conclude by (†) that the sum

$$\frac{1}{pc^s}\sum_{i} f^{(j)}(\beta_i) \text{ lies in } \mathbb{Z}[s_1(\beta_1,\dots,\beta_r),\dots,s_r(\beta_1,\dots,\beta_r)]$$

by evaluating the t_i at the β_i. Since

$$\frac{g}{c} = t^r + \frac{c_1}{c}t^{r-1} + \cdots + \frac{c_r}{c}$$

has roots $\beta_1,\dots,\beta_r$, we see that

$$s_j(\beta_1,\dots,\beta_r) = (-1)^j\frac{c_j}{c} \text{ for } j = 1,\dots,r,$$

so

$$\frac{1}{pc^s}\sum_{i=1}^{r} f^{(j)}(\beta_i) \text{ lies in } \mathbb{Z}[\frac{c_1}{c},\dots,\frac{c_r}{c}].$$

Since (†) has total degree in $s_1(t_1,\dots,t_r),\dots,s_r(t_1,\dots,t_r)$ at most s, it follows that

$$\frac{1}{p}\sum_{i=1}^{r} f^{(j)}(\beta_i) \text{ lies in } \mathbb{Z}[c_1,\dots,c_r] \subset \mathbb{Z},$$

i.e., $\sum_{i=1}^{r} f^{(j)}(\beta_i)$ is an element of $p\mathbb{Z}$. This shows (a) and (b).

Check 67.4. We have

$$f^{(j)}(0) = \begin{cases} 0 & \text{if } j \le p-2 \\ c^s c_r^p & \text{if } j = p-1 \\ l_j p & \text{if } j \ge p \text{ some integer } l_j. \end{cases}$$

We conclude that there exists an integer N satisfying

$$\big(\sum_{i=1}^{r} F(\beta_i)\big) + kF(0) = Np + kc^s c_r^p \text{ in } \mathbb{Z}.$$

As k, c, and c_r are not zero, if p is chosen such that $p > \max\{k, |c|, |c_r|\}$, then $p \nmid \big(\sum_{i=1}^{r} F(\beta_i)\big) + kF(0)$. This shows that A is a nonzero integer for all $p >> 0$.

We turn to computing B.

Claim: $B = 0$ for all $p >> 0$.

Of course, if we show this, then we have finished the proof. For each j with $1 \leq j \leq r$, let

$$m(j) := \sup_{0 \leq \lambda \leq 1} |g(\beta_j \lambda)|.$$

Hence if $0 \leq \lambda \leq 1$, we have

$$|f(\beta_j \lambda)| \leq \frac{|c|^s |\beta_j|^{p-1} m(j)^p}{(p-1)!}.$$

Let

$$N = \max_j \left| e^{(1-\lambda)\beta_j} \right|,$$

then

$$\Big| \sum_{j=1}^{r} \beta_j \int_0^1 e^{(1-\lambda)\beta_j} f(\beta_j \lambda) d\lambda \Big| \leq \sum_{j=1}^{r} \frac{|c|^s |\beta_j|^p m(j)^p}{(p-1)!} N,$$

so the integer

$$\sum_{j=1}^{r} \beta_j \int_0^1 e^{(1-\lambda)\beta_j} f(\beta_j \lambda) d\lambda \to 0 \quad \text{as} \quad p \to \infty.$$

It follows (by the remark at the beginning of the section) that this sum must be zero for all $p >> 0$ □

We state some other results about transcendental numbers over $\mathbb{Q}$ without proof. A complex number α is called an *algebraic integer* if there exists a monic polynomial f in $\mathbb{Z}[t]$ such that α is a root of f. Algebraic Number Theory studies the set

$$\mathcal{A} := \{\alpha \in \mathbb{C} \mid \alpha \text{ is an algebraic integer}\}.$$

This set is, in fact, a domain. If $K/\mathbb{Q}$ is a finite extension of fields, one also studies $\mathcal{A}_K := \mathcal{A} \cap K$, the *ring of algebraic integers* in K. The transcendence of e and π are special cases of a more general result that Lindemann sketched and Weirstraß rigorously proved, viz.,

Theorem 67.5. *Suppose that $\alpha_1, \ldots, \alpha_n$ are distinct algebraic integers and $\beta_1, \ldots, \beta_n$ nonzero algebraic integers. Then*

$$\beta_1 e^{\alpha_1} + \cdots + \beta_n e^{\alpha_n} \quad \textit{is not zero.}$$

It follows from this theorem that if the $\alpha_1, \ldots, \alpha_n$ are linearly independent over $\mathbb{Q}$, then the $e^{\alpha_1}, \ldots, e^{\alpha_n}$ are algebraically independent over $\mathbb{Q}$. In particular e and π are transcendental over $\mathbb{Q}$. In addition, it implies, for example, that $\cos \alpha$, $\sin \alpha$, and $\tan \alpha$ are transcendental

over $\mathbb{Q}$ if α is a nonzero algebraic integer and $\log\alpha$ is transcendental over $\mathbb{Q}$ if α is an algebraic integer different from zero or one.

One of the famous Hilbert Problems is:

Problem 67.6. (Hilbert Seventh Problem) Let α and β be algebraic integers with α not zero or one and β not rational, is it true that α^β is transcendental over $\mathbb{Q}$, e.g., $\sqrt{2}^{\sqrt{2}}$ is transcendental over $\mathbb{Q}$?

This was solved in the affirmative independently by Gelfand and Schneider. They showed

Theorem 67.7. *Suppose that $\alpha_1, \alpha_2, \beta_1, \beta_2$ are nonzero algebraic integers. If $\log\alpha_1$ and $\log\alpha_2$ are linearly independent over $\mathbb{Q}$, then $\beta_1\log\alpha_1 + \beta_2\log\alpha_2$ is not zero.*

This theorem implies the solution of the Hilbert Problem. This was generalized by Baker in the 1960's, who showed

Theorem 67.8. *Suppose that $\alpha_1, \ldots, \alpha_n$ are nonzero algebraic integers such that $\log\alpha_1, \ldots, \log\alpha_n$ are linearly independent over $\mathbb{Q}$. Then $1, \log\alpha_1, \ldots, \log\alpha_n$ are linearly independent over the quotient field of $\mathcal{A}$.*

This theorem implies, for example, that $e^{\beta_0}\alpha_1^{\beta_1}\cdots\alpha_n^{\beta_n}$ is transcendental over $\mathbb{Q}$ for any set of nonzero algebraic integers $\alpha_1, \ldots, \alpha_n, \beta_0, \ldots \beta_n$. Baker won the Fields Medal for this work.

Exercises 67.9. 1. Prove Check 67.2.

2. Prove Check 67.4.

3. Prove Check 67.3.

CHAPTER XIV

The Theory of Formally Real Fields

In this chapter, we develop the theory of formally real fields including the standard results from Artin-Schreier Theory.

68. Orderings

Definition 68.1. A field F is called *formally real* if -1 is not a sum of squares in F, i.e., the polynomial $t_1^2 + \cdots + t_n^2$ has no nontrivial zero over F for any (positive) integer n. A formally real field is called *real closed* if it has no proper algebraic extension that is formally real.

If F is formally real, then the characteristic of F must be zero for if $\operatorname{char} F = p > 0$, then -1 is a sum of $p - 1$ squares.

Notation 68.2. For a commutative ring R, let

$$\sum R^2 := \{x \in R \mid x \text{ is a sum of squares in } R\}.$$

A field F of characteristic different from two is not formally real if and only if $F = \sum F^2$. In general, $\sum F^2$ is closed under addition and multiplication while $\sum (F^2)^\times := \sum (F^2) \setminus \{0\}$ is a multiplicative group.

Definition 68.3. Let R be a commutative ring, $P \subset R$ a subset. We say that P is a *preordering* of R if P satisfies all of the following:

(1) $P + P \subset P$.
(2) $P \cdot P \subset P$.
(3) $-1 \notin P$.
(4) $\sum R^2 \subset P$.

Let

$$\mathcal{Y}(R) := \{P \mid P \subset R \text{ is a preordering }\}.$$

[Of course, $\mathcal{Y}(R)$ may very well be empty.] A preordering $P \in \mathcal{Y}(R)$ is called an *ordering* if, in addition,

$$R = P \cup -P \text{ and } P \cap -P = \{0\}$$

where $-P := \{x \mid -x \in P\}$.

Let P be a preordering on R. We shall show below that if P is a maximal preordering (rel $\subset$), then $R = P \cup -P$ and $P \cap -P$ is a prime ideal in R. In particular, if R is a field, this means that P is a

maximal preordering if and only if P is an ordering. Note: We have not excluded 0 from lying in P.

Let

$$\mathcal{X}(R) := \{P \in \mathcal{Y}(R) \mid P \text{ is an ordering of } R\}.$$

By definition, a field F is formally real if and only if $\sum F^2$ is a preordering. In general, the preordering $\sum F^2$ in a formally real field F is not an ordering. We shall see that $\sum F^2$ is an ordering if and only if $|\mathcal{X}(F)| = 1$. For example, this is the case when a formally real field F is *euclidean*, i.e., if every element in F is a square or the negative of a square. For such a field, we even have $\sum F^2 = F^2$. For example, the real numbers or the real constructible numbers are euclidean fields. Of course, in general $F^2 \neq \sum F^2$. When a field of characteristic different from two satisfies $F^2 = \sum F^2$, it is called *pythagorean*. If F is not formally real (and $\operatorname{char} F \neq 2$), then F is pythagorean if and only if it is *quadratically closed,* i.e., $F = F^2$.

Lemma 68.4. *Let R be a commutative ring, $P \in \mathcal{Y}(R)$.*

(1) If $a, b \in R$ satisfy $ab \in P$ then either $P + aP \in \mathcal{Y}(R)$ or $P - bP \in \mathcal{Y}(R)$.

(2) If $P \in \mathcal{Y}(R)$ is maximal (relative to $\subset$), then $R = P \cup -P$ and $P \cap -P$ is a prime ideal in R. In particular, if R is a field then $P \in \mathcal{X}(R)$.

(3) If R is a field, then P is a maximal preordering if and only if $P \in \mathcal{X}(R)$.

(4) There exists a $\widetilde{P} \in \mathcal{Y}(R)$ containing P and satisfying $R = P \cup -P$ and $P \cap -P$ is a prime ideal in R. In particular, if R is a field, there exists an ordering $\widetilde{P} \in \mathcal{X}(R)$ containing P.

Proof. (1): We need only show that -1 cannot lie in both $P+aP$ and $P - bP$. If it does then there exist $x, y, z, w \in P$ satisfying $-1 = x + ay = z - bw$, so

$$(ay)(-bw) = (-1 - x)(-1 - z) = 1 + x + z + xz,$$

hence $-1 = x + z + xz + abyw \in P$, a contradiction.

(2): If $a \in R$ then $a^2 \in P$, so by (1), either $P + aP \in \mathcal{Y}(R)$ or $P - aP \in \mathcal{Y}(R)$. By maximality, either $a \in P$ or $-a \in P$, so $R = P \cup -P$. Certainly, $P \cap -P$ is closed under addition. If $x \in P \cap -P$ and $y \in R$ then $xy = (-x)(-y) \in P \cap -P$, so $P \cap -P$ is an ideal. Suppose that $ab = (-a)(-b) \in P \cap -P$. If $a \notin P \cap -P$, we may assume that $a \in -P \setminus P$. Maximality and (1) imply that $-b \in P$. As $a(-b) \in P \cap -P$ also, the same argument shows that $b \in -P$.

(3): Fields have no nonzero idempotents.

(4) follows from (2) and Zorn's Lemma. □

Proposition 68.5. *Let F be a formally real field and $P \in \mathcal{Y}(F)$. Then*

$$P = \bigcap_{P \subset \widetilde{P} \in \mathcal{X}(F)} \widetilde{P}.$$

PROOF. The inclusion $P \subset \bigcap_{P \subset \widetilde{P} \in \mathcal{X}(F)} \widetilde{P}$ is clear. Conversely, suppose that $x \notin P$. By definition, $-1 \notin P$. We show that $P - xP \in \mathcal{Y}(F)$. To do this, it suffices to show that $-1 \notin P - xP$. If this is false, write $-1 = y - xz$ with $y, z \in P$. Then $z \neq 0$ so

$$x = 1 \cdot z^{-1} + yz^{-1} = \frac{z}{z^2} + \frac{yz}{z^2} \in P,$$

a contradiction. By the lemma, there exists $\widehat{P} \in \mathcal{X}(F)$ such that $P - xP \subset \widehat{P}$. As $x \notin P - xP$, lest $-x^2 \in P - xP$, it follows that $-1 \in P - xP$, so $x \notin \bigcap_{P \subset \widetilde{P} \in \mathcal{X}(F)} \widetilde{P}$. □

Corollary 68.6. (Artin-Schreier) *Suppose that F is formally real. Then $\sum F^2 = \bigcap_{\mathcal{X}(F)} P$. In particular, $\mathcal{X}(F) \neq \emptyset$.*

PROOF. As F is formally real, $\sum F^2 \in \mathcal{Y}(F)$. □

Remark 68.7. Let F be a formally real field and $P \in \mathcal{X}(F)$. An element $0 \neq x \in F$ is called *positive* (respectively, *negative*) *rel* P and also written $x >_P 0$ (respectively, $x <_P 0$) if $x \in P$ (respectively, $-x \in P$). If $x \in \bigcap_{\mathcal{X}(F)} Q$, i.e., x is positive at every ordering of F then x is called *totally positive* (and $-x$ is called *totally negative*). In particular, we have

$$\bigcap_{\mathcal{X}(F)} Q \setminus \{0\} = (\sum F^2)^\times,$$

i.e., $0 \neq x$ is totally positive if and only if it is a sum of squares.

If $P \in \mathcal{X}(F)$, let $\leq_P$ on F denote the total ordering given by $x \leq_P y$ if $y - x \in P$. Of course, if $x \leq_P y$ and $z \in F$ then $x + z \leq_p y + z$ and if, in addition, $0 \leq_P z$ then $xz \leq_P yz$. We let $>_P$ and $\geq_P$ have the obvious meaning.

Exercises 68.8. 1. Prove that the field of real constructible numbers C is euclidean and $C(\sqrt{-1})$ is quadratically closed.

2. Show that a formally real field is euclidean if and only if it is pythagorean and has a unique ordering.

3. Show the intersection of pythagorean fields is pythagorean.

4. Let F be a formally real field and $\widetilde{F}$ an algebraic closure of F. A pythagorean field L is called a *pythagorean closure* of F if either $L = F$ or whenever $F \subset K < L$, then K is not pythagorean. Show that if there is a pythagorean field in $\widetilde{F}$ containing F, then it is the intersection of all pythagorean fields containing F in $\widetilde{F}$. In particular, if a pythagorean closure of F in $\widetilde{F}$ exists then it is unique and denoted by F_{pyth}.
5. Let F be a formally real field and $\widetilde{F}$ an algebraic closure. Call a square root tower $F \subset F_0 \subset F_1 \subset \cdots \subset F_n$ admissible if for each i, $1 \leq i \leq n-1$, there exist x_i, y_i in F_{i-1} satisfying $F_i = F_{i-1}(\sqrt{x_i^2 + y_i^2})$. Show the union of all admissible square towers over F is the pythagorean closure F_{pyth} of F. In particular, if F is a formally real field, then the pythagorean closure of F in $\widetilde{F}$ exists.

69. Extensions of Ordered Fields

Definition 69.1. Let K/F be a field extension with K formally real. Let $P \in \mathcal{X}(K)$. The pair (K, P) is called an *ordered field*. If $P_0 \in \mathcal{X}(F)$ satisfies $P_0 = P \cap F$, then $(K, P)/(F, P_0)$ is called an *extension* of ordered fields and P is called an *extension* of P_0. If there exists no proper algebraic extension L of F with $Q \in \mathcal{X}(L)$ satisfying $Q \cap F = P_0$, we say that (F, P_0) is *real closed (rel P_0)*. Let $(K, P)/(F, P_0)$ be an extension of ordered fields. If (K, P) real closed (rel P) and K/F is algebraic, (K, P) is called a *real closure of* (F, P_0) *(rel P_0)*.

We shall show that given $P \in \mathcal{X}(F)$, there exists a real closure of (F, P) and it is unique up to a (unique) F-isomorphism. We begin with the following useful result characterizing extensions of orderings under certain conditions.

Theorem 69.2. *Let (F, P) be an ordered field.*

(1) Let $d \in F$ and $K = F(\sqrt{d})$. Then there exists an extension of P to K if and only if $d \in P$.
(2) If K/F is finite of odd degree then there exists an extension of P to K.

Proof. (1): Suppose $d \in P$. We may assume that $d \in P \setminus F^2$. Let

$$S = \{\sum x_i y_i^2 \mid x_i \in P,\ y_i \in K \text{ for all } i\}.$$

We show that $S \in \mathcal{Y}(K)$. Certainly, S is closed under addition and multiplication and contains $\sum K^2$. Thus it suffices to show that $-1 \notin S$. If $-1 \in S$, we can write $-1 = \sum x_i(a_i + b_i\sqrt{d})^2$ for some $a_i, b_i \in F$ and $x_i \in P$. As $\{1, \sqrt{d}\}$ is an F-basis for K, we arrive at the

contradiction that $-1 = \sum x_i(a_i^2 + b_i^2 d) \in P$. This proves that $S \in \mathcal{Y}(K)$. By Lemma 68.4, there exists $\widetilde{P} \in \mathcal{X}(K)$ satisfying $P \subset S \subset \widetilde{P}$. As P is an ordering, $P = \widetilde{P} \cap F$. Suppose that $d \notin P$. If $\widetilde{P} \in \mathcal{X}(K)$ contains P then $d = (\sqrt{d})^2 \in \widetilde{P} \cap F = P$, a contradiction.

(2): As the $\operatorname{char} F$ is zero, by the Primitive Element Theorem, $K = F(x)$, for some $x \in K$. Let f be the minimal polynomial of x and suppose that

$$S = \{\sum a_i y_i^2 \mid a_i \in P,\ y_i \in K \text{ for all } i\} \notin \mathcal{Y}(K).$$

As $K \cong F[t]/(f)$, we have an equation,

$$-1 \equiv \sum a_i g_i(t)^2 \mod (f)$$

for some $a_i \in P$ and $0 \neq g_i \in F[t]$ with $\deg g_i < \deg f$ for all i. Lift this to an equation

$$-1 = \sum a_i g_i(t)^2 + qf$$

in $F[t]$. Since $-1 \notin P$, evaluating at x shows that there exists an i with $\deg g_i > 0$. Let $m = \max_i(\deg g_i) > 0$. Then the coefficient of t^{2m} in $\sum a_i g_i(t)^2$ cannot be zero as it is of the form $\sum a_m b^{2m} >_P 0$. Thus $\deg qf = 2m$. As $\deg f = n$ is odd, so is $\deg q$. By construction, $\deg q < \deg f$. Let $h \mid q$ in $F[t]$ with $h \in F[t]$ irreducible of odd degree. Then $-1 \equiv \sum a_i g_i^2 \mod (h)$, so we may repeat the argument. By induction, this reduces to the case when f is linear where the result is false. □

The first part of the theorem has the immediate corollary.

Corollary 69.3. *Let (F, P_0) be an ordered field. Suppose that there exists an element d in $P_0 \setminus \sum F^2$. Then there exists an proper extension of ordered fields $(F(\sqrt{d}), P)/(F, P_0)$ with $P_0 + dP_0 \subset P$. In particular $F(\sqrt{d})$ is formally real.*

We also conclude that

Corollary 69.4. *(F, P) is real closed (rel P) if and only if F is real closed. Moreover, if this is the case, then $P = F^2$.*

Proof. Suppose that (F, P) is real closed. Let $d \in P$. By the theorem, P extends to $F(\sqrt{d})$ so $d \in F^2$ by hypothesis. As (F, P) is real closed, we must have $P = F^2 \in \mathcal{X}(F)$. Since $F^2 \subset Q$ for all $Q \in \mathcal{X}(F)$, we have $\mathcal{X}(F) = \{F^2\}$. Suppose that L/F is algebraic and L formally real. Let $Q \in \mathcal{X}(L)$. Then $(L, Q)/(F, F^2)$ is an extension of ordered fields as $F^2 \subset Q$, hence $L = F$ and F must be real closed.

If F is real closed, then $F(\sqrt{d})/F$ cannot be a nontrivial extension of formally real fields. It follows by the Corollary 69.3 and Lemma 68.4 that $F^2 \in \mathcal{Y}(F)$ is an ordering hence $\mathcal{X}(F) = \{F^2\}$. If L/F is a proper algebraic extension, then $-1 \in L^2$ so $F \cap L^2$ cannot be a preordering on F. It follows that (F, F^2) is real closed. □

There exist fields such that $\mathcal{X}(F) = \{F^2\}$ but F not real closed, i.e., euclidean fields that are not real closed. The real constructible numbers is such an example. The theorem only implies that a euclidean field has no formally real quadratic extensions. It also means that $F(\sqrt{-1})$ is quadratically closed. Indeed if $\alpha = a+b\sqrt{-1}$ in $F(\sqrt{-1})$ with $a, b \in F$, then there exist $c, d \in F$ satisfying

$$c^2 = \frac{a + \Delta}{2} \quad \text{and} \quad d^2 = \frac{-a + \Delta}{2}$$

where $\Delta = \sqrt{a^2 + b^2}$ is the positive square root in F^2 in euclidean F.

Proposition 69.5. *Every ordered field (F, P) has a real closure $(\widetilde{F}, \widetilde{F}^2)$.*

Proof. Let $\widehat{F}$ be an algebraic closure of F. Let $P \in \mathcal{X}(F)$. The statement follows from a Zorn's Lemma argument on

$$\{\widehat{F}/K/F \mid (K, Q)/(F, P) \text{ is an extension}\}. \qquad \square$$

Because of the last results, if $(K, P)/(F, P_0)$ is a real closure, then $P = K^2$ and K is real closed so we just say K is a real closure of F rel P.

Corollary 69.6. *Let F be a formally real field and $P \in \mathcal{X}(F)$. Then there exists a real closure of (F, P).*

Theorem 69.7. (Fundamental Theorem of Algebra) *The following are equivalent:*

(1) F is real closed.
(2) F is euclidean and every polynomial $f \in F[t]$ of odd degree has a root in F.
(3) F is not algebraically closed but $F(\sqrt{-1})$ is.

Proof. (1) ⇒ (2): We have seen the hypothesis implies that F is euclidean. Let $p \in F[t]$ be irreducible of odd degree. As $K = F[t]/(p)$ is an extension of odd degree any ordering of F would extend, hence p is linear. (2) easily follows.

(2) ⇒ (3): This is the essentially the same proof given for the usual Fundamental Theorem of Algebra 54.12. Let $K = F(\sqrt{-1})$. Then K is quadratically closed. Let E/K be an algebraic extension and L/F the normal closure of K/F. If H is a 2-Sylow subgroup of the Galois

group G of L/F, then $L^H = F(\theta)$ for some θ by the Primitive Element Theorem, so we must have $L^H = F$ by hypothesis. Consequently, $G = H$ and hence must be trivial as K is quadratically closed.

(3) $\Rightarrow$ (1): It suffices to show that F is formally real, i.e., $-1 \notin \sum F^2$. As $\sqrt{-1} \notin F$, it suffices to show F is pythagorean. Let $a, b \in F^\times$. We need only show that $a^2 + b^2 \in F^2$. Let $f = t^4 - 2at^2 + (a^2 + b^2) \in F[t]$. Then $f = \big(t^2 - (a + b\sqrt{-1})\big)\big(t^2 - \big(a - b\sqrt{-1}\big)\big)$ in $F(\sqrt{-1})$. As $F(\sqrt{-1})$ is quadratically closed, there exist $\alpha, \beta \in F(\sqrt{-1})$ satisfying

$$\alpha^2 = a + b\sqrt{-1} \quad \text{and} \quad \beta^2 = a - b\sqrt{-1},$$

hence

$$f = (t - \alpha)(t + \alpha)(t - \beta)(t + \beta) \quad \text{in} \quad F(\sqrt{-1})[t].$$

As $f \in F[t]$ cannot be irreducible and $ab \neq 0$, either $(t - \alpha)(t - \beta)$ or $(t - \alpha)(t + \beta)$ is an irreducible factor of f over F. Thus $\alpha\beta \in F$, hence

$$a^2 + b^2 = (a + b\sqrt{-1})(a - b\sqrt{-1}) = \alpha^2\beta^2$$

lies in F^2. □

It follows that over a real closed field, the only irreducible polynomials are linear polynomials and irreducible quadratic polynomials of the form at^2+bt+c with negative discriminant (in the unique ordering).

Artin and Schreier proved that the conditions of the Fundamental Theorem of Algebra was also equivalent to: F is not algebraically closed and its algebraic closure is a finite extension of it. We shall come back to this below. Before showing this, we turn to the question of uniqueness.

Let A be a finite dimensional commutative F-algebra. If $a \in A$ let $\lambda_a : A \to A$ be the F-linear map given by $x \mapsto ax$. The *trace* $\mathrm{trace} : A \to F$ is the trace of a matrix representation of λ_a relative to a fixed basis. It is independent of the choice of basis. If A is a finite separable field extension of F of characteristic different from two, then this is just $\mathrm{Tr}_{A/F}$ (cf. Exercise 57.24(6). Define the *trace form* of A

$$\varphi = \varphi_A : A \times A \to F$$

to be the symmetric bilinear form given by $\varphi(x, y) = \mathrm{trace}(\lambda_{xy})$.

Let B be a symmetric bilinear space over a formally real field F and $P \in \mathcal{X}(F)$. Then Sylvester's Law of Inertia holds:

If V is the underlying vector space for B, then we have an orthogonal decomposition $V = V_+ \perp V_- \perp V_0$ where if $0 \neq v \in V$ then

$$B(v,v) \;=\; \begin{cases} >_P 0 & \text{if } v \in V_+ \\ <_P 0 & \text{if } v \in V_- \\ = 0 & \text{if } v \in V_0 \end{cases}$$

with the dimensions of V_0, V_+, and V_- independent of such an orthogonal decomposition. The integer

$$\operatorname{sgn}_P B \;:=\; \dim V_+ \,-\, \dim V_-$$

is an invariant of B, called the *signature* of B at P. If K is an extension field of F and $\mathcal{B}$ a basis for V, then we let V_K be the vector space over K on basis $\mathcal{B}$. By restricting scalars to F we see it is also a vector space over F and there exists a (natural) linear embedding of V into V_K by sending $\mathcal{B}$ to $\mathcal{B}$. We call this *extension of scalars*. (Cf. making a real vector space into a complex one by extending scalars.) If B is the bilinear form above, define $B_K : V_K \times V_K \to K$ by $B_K(\alpha v_1, \beta v_2) = \alpha\beta B(v_1, v_2)$ where α and β lie in K and $v_1.v_2$ in V. Then, if $(K,Q)/(F,P)$ is an extension, $\operatorname{sgn}_P B = \operatorname{sgn}_Q(B_K)$.

Computation 69.8. Let (F,P) be an ordered field, $f \in F[t] \setminus F$ and $A = F[t]/(f)$. Let $\bar{}\,: F[t] \to A$ be the canonical epimorphism and $\varphi : A \times A \to F$ the trace form.

Case 1. $f = (t-a)^n$ in $F[t]$:

Let $\mathcal{B} = \{1, \bar{t}-a, \dots, (\bar{t}-a)^{n-1}\}$, an F-basis for A. As $(\bar{t}-a)$ is nilpotent, $\operatorname{trace}(\lambda_{(\bar{t}-a)^i}) = 0$ for all $i \geq 1$. It follows that the matrix representation of φ in the $\mathcal{B}$ basis is given by the diagonal matrix

$$[\varphi]_{\mathcal{B}} \quad = \quad \left(\operatorname{trace}(\lambda_{(\bar{t}-a)^i(\bar{t}-a)^j})\right) \quad = \quad \begin{pmatrix} n & 0 & \cdots & 0 \\ 0 & 0 & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{pmatrix}.$$

Consequently, $\operatorname{sgn}_P \varphi = \operatorname{sgn}[\varphi]_{\mathcal{B}}) = 1$.

Case 2. $f = (t^2 + at + b)^n$ in $F[t]$ with $a^2 - 4b <_P 0$:

Let $\mathcal{B} = \{1, \bar{t}, \bar{t}+1, \bar{t}(\bar{t}+1), (\bar{t}+1)^2, \bar{t}(\bar{t}+1)^2, \dots\}$, an F-basis for A. Computation shows the matrix representation of φ in the $\mathcal{B}$ basis is given by the diagonal matrix

$$[\varphi]_{\mathcal{B}} = \begin{pmatrix} 2n & 0 & \cdots & & 0 \\ 0 & -2n & & & \\ \vdots & & 0 & & \\ & & & \ddots & \\ 0 & \cdots & & & 0 \end{pmatrix}.$$

Consequently, $\operatorname{sgn}_P \varphi = 0$

The next result, due to Sylvester, is the main ingredient in proving the uniqueness of real closures relative to a fixed ordering. It replaces the more common use of Sturm's Theorem.

Lemma 69.9. (Sylvester) *Let (F, P) be an ordered field and K a real closure of (F, P). Suppose that f is a non-constant polynomial in $F[t]$ and $A = F[t]/(f)$. Let $\varphi : A \times A \to F$ be the trace form. Then*

$$\operatorname{sgn}_P \varphi = \text{ the number of roots of } f \text{ in } K.$$

In particular, the number of roots of f in K depends only on P and is independent of the real closure K.

Proof. Let φ be the trace form on $A = F[t]/(f)$ and $f = up_1^{e_1} \cdots p_r^{e_r}$ a factorization of f into monic irreducibles in $K[t]$ with $u \in K^\times$. By the Chinese Remainder Theorem, we have a natural ring (and K-algebra) isomorphism.

$$A_K := K[t]/(f) = \prod K[t]/(p_i^{e_i}) = \prod A_i,$$

where $A_i = K[t]/(p_i^{e_i})$. We view this as an identification of K-algebras. It is easy to check that $\varphi_{A_K}(x_i, x_j) = 0$ if $x_i \in A_i$ and $x_j \in A_j$ with $i \neq j$ (by writing down an appropriate K-basis), hence we can restrict φ_A to each A_i to obtain $\operatorname{sgn}_P \varphi = \operatorname{sgn}_{K^2}(\varphi_{A_K}) = \sum_{i=1}^r \operatorname{sgn}_{K^2} \varphi_{A_i}$. Since K is real closed, it follows from the Fundamental Theorem of Algebra that p_i is either linear or quadratic. Therefore, by the computation, we see that

$$\operatorname{sgn}_P \varphi = \sum_{i=1}^r \operatorname{sgn}_{K^2} \varphi_{A_i} = \sum_{p_{i_j} \text{ linear}} e_{i_j}.$$

The result follows. □

Lemma 69.10. *Let (F, P) be an ordered field and K a real closure of (F, P). Suppose that $E = F(\alpha)$ is an algebraic extension of F with $\varphi = \varphi_E : E \times E \to F$ the trace form. Suppose that $r = \operatorname{sgn}_P \varphi > 0$. Then the minimal polynomial f of α has r roots in K. Let $\alpha_1, \alpha_2, \ldots, \alpha_r$ be these roots. Then there exist at most r distinct and at*

least one extension $\widetilde{P}_i$ of P to E. These are induced by the r distinct F-homomorphisms $\sigma_i : E \to K$ given by $\alpha \mapsto \alpha_i$ with $\widetilde{P}_i = \sigma_i^{-1}(K^2)$.

PROOF. Every F-embedding of E into an algebraic closure of K must take α to a root of f. By Lemma 69.9, the polynomial f has r distinct roots in K. Clearly, $\widetilde{P}_i = \sigma_i^{-1}(K^2) \supset P$. Suppose that $Q \in \mathcal{X}(E)$ extends P with $Q \neq \widetilde{P}_i$ for $1 \leq i \leq r$. Then there exists an $a_i \in E$ satisfying $a_i \in Q$ but $\sigma_i(a_i) \notin K^2$ for all i. By repeated application of Theorem 69.2, there exists $\widetilde{Q}_i \in \mathcal{X}\big(E(\sqrt{a_1}, \ldots, \sqrt{a_r})\big)$ extending Q. By the Primitive Element Theorem, there exists β such that $F(\beta) = E(\sqrt{a_1}, \ldots, \sqrt{a_r})$. Let L be a real closure of $(F(\beta), \widetilde{Q})$ hence of (F, P). The minimal polynomial $g \in F[t]$ of β has root $\beta \in L$, hence a root in K by Lemma 69.9. Therefore, there exists an F-homomorphism $\tau : F(\beta) \to K$. As the F-homomorphism $\tau|_E : E \to K$ must satisfy $\tau|_E = \sigma_i$ for some i, we have $\sigma_i(\alpha_i) = \tau(\alpha_i) = \big(\tau(\sqrt{\alpha_i})\big)^2$ lies in K^2, a contradiction. □

Theorem 69.11. *Let (F, P) be an ordered field.*

1. *Let $(E, Q)/(F, P)$ be an algebraic extension of ordered fields and K a real closure of (F, P). Then there exists an order preserving F-homomorphism $E \to F$.*
2. *Let K_1 and K_2 be two real closures of (F, P). Then there exists a unique F-homomorphism $K_1 \to K_2$ and it is an order preserving isomorphism.*
3. *In Lemma 69.10 and its notation, all the $\widetilde{P}_i$ are distinct. In particular, there exist precisely r extensions of P to E.*

PROOF. (1): Apply Zorn's Lemma to

$$\{M \mid E/M/F \text{ and there exists an order preserving } F\text{-homomorphism } \psi_M : (M, Q \cap M) \to (K, K^2)\}$$

to obtain a maximal such M_0. Suppose $M_0 \neq E$ and $x \in E \setminus M_0$. Let L be a real closure of (E, Q). Then f has a root in L so $\operatorname{sgn}_{Q \cap M_0(x)} \varphi_{M_0(x)} > 0$. By Lemma 69.10, we can extend $\psi_{M_0} : (M_0, Q \cap M_0) \to (K, K^2)$ to $\psi_{M_0(x)} : \big(M_0(x), Q \cap M_0(x)\big) \to (K, K^2)$, a contradiction. So $M_0 = E$.

(2): By (1) there exists an F-homomorphism $\sigma : K_1 \to K_2$. Let $\tau : K_1 \to K_2$ be another F-homomorphism. Let $\alpha \in K_1$ and $\alpha_1, \ldots, \alpha_r$ be the (distinct) roots of the minimal polynomial f of α in K_1. We may assume that

$$\alpha_1 <_{K_1^2} \cdots <_{K_1^2} \alpha_r$$

in K_1. As $\sigma(K_1^2) \subset K_2^2$ and $\tau(K_1^2) \subset K_2^2$ and $\mathcal{X}(K_i) = \{K_i^2\}$ for $i = 1, 2$ both σ and τ are order preserving. Thus

$$\sigma(\alpha_1) <_{K_2^2} \cdots <_{K_2^2} \sigma(\alpha_r) \text{ and } \tau(\alpha_1) <_{K_2^2} \cdots <_{K_2^2} \tau(\alpha_r) \text{ in } K_2.$$

As these are the roots of f in K_2, we have $\sigma(\alpha_i) = \tau(\alpha_i)$ for all i. In particular, $\sigma(\alpha) = \tau(\alpha)$. Hence $\sigma = \tau$.

(3): Suppose in Lemma 69.10 that we have $\widetilde{P}_i = \widetilde{P}_j$ with $i \neq j$. Let $\widetilde{P} = \widetilde{P}_i$ and M a real closure of $(E, \widetilde{P})$. By (1), the embeddings $\sigma_i \neq \sigma_j$ can be extended to F-homomorphisms $M \to K$. This contradicts (2). □

Exercise 69.12. Show that the field of real constructible numbers C has algebraic extensions that are not euclidean. In particular, $C(\sqrt{-1})$ is not algebraically closed.

70. Characterization of Real Closed Fields

To prove the last Artin-Schreier characterization of real closed fields, we need a few facts from field theory. These are

Fact 70.1. Let F be a field.

(1) If $\operatorname{char} F = p > 0$ and $a \in F \setminus F^p$, then $t^{p^e} - a \in F[t]$ is irreducible for all $e \geq 1$. (Cf. Example 47.15(8.)
(2) If $\operatorname{char} F = p > 0$ and $a \in F$ satisfies $a \neq u^p - u$ for any $u \in F$, then $t^p - t - a \in F[t]$ is irreducible. (Cf. Exercise 53.22(9).)
(3) If $\operatorname{char} F = p > 0$ and $a \in F$ satisfies $a \neq u^p - u$ for any $u \in F$ and E is a splitting field of $t^p - t - a$ over F, then there exists and extension K/E of degree p.
(4) Let p be a prime different from $\operatorname{char} F$. If $t^p - 1$ splits in $F[t]$ and K/F is a cyclic extension of degree p, then $K = F(r)$ for some r satisfying $r^p \in F$. (Cf. Theorem 57.20.)
(5) Suppose p is a prime different from $\operatorname{char} F$ and F contains a primitive pth root of unity. If E/F is cyclic of degree p, then there exists an $x \in E$ such that $E = F(x)$ and the minimal polynomial of x over F is $t^p - a \in F[t]$. (Same as the previous.)
(6) If $\operatorname{char} F = p > 0$ and E/F is cyclic of degree p, then there exists an $x \in E$ such that $E = F(x)$ and $x^p - x \in F$. (Same as previous.)

Theorem 70.2. *Let C be algebraically closed and R a proper subfield with C/R finite. Then R is real closed and $C = R(\sqrt{-1})$.*

PROOF. Let $C' = R(\sqrt{-1})$. As C is the algebraic closure of C', it suffices to show $C = C'$. We have:

$$\text{If } C/E/C', \text{ then } [E : C'] \leq [C : C'] < \infty. \tag{70.3}$$

Claim 70.4. C' is perfect:

Suppose not, then $\operatorname{char} F = p > 0$ and there exists an element $a \in C' \setminus (C')^p$. By Fact (1), for each $n \geq 1$, there exists E_n/C' of degree p^n contradicting (70.3). Therefore C' is perfect.

By Claim 70.4, we conclude that C/C' is finite Galois, since C is algebraically closed. Let $G = G(C/C')$. We may assume that $C' < C$. In particular, there exists a prime number p dividing $|G|$ and a cyclic subgroup $H \subset G$ of order p. Let E be the fixed field C^H, so

$$C/E \text{ is cyclic of degree } p.$$

Claim 70.5. E contains a primitive pth root of unity and $p \neq \operatorname{char} E$:

Suppose to the contrary that $p = \operatorname{char} E$. By Fact (4), there exists an $x \in C$ such that $C = E(x)$ is the splitting field of the irreducible polynomial $t^p - t - a \in E[t]$ over E. Hence $a = x^p - x$. By Fact (3), there exists a field extension K/C of degree p which is impossible.

Let $\operatorname{char} E = l \geq 0$. We have shown that $p \neq l$. As $t^p - 1 = (t-1)(t^{p-1} + \cdots + 1)$ splits in $C[t]$ and each irreducible factor has degree at most $p-1$, the polynomial $t^p - 1$ must split in $E[t]$, hence E contains a primitive pth root of unity. As $\operatorname{char} E \neq p$, we have $C = E(r)$ where the irreducible polynomial of r is $t^p - a \in E[t]$ by Fact (4). Let $\zeta \in C$ be a primitive p^2-root of unity. As $t^{p^2} - a$ splits in C, we can write

$$t^{p^2} - a = \prod_{i=0}^{p^2-1} (t - \zeta^i s) \text{ with } a = s^{p^2}$$

in $C[t]$.

Claim 70.6. There exists a primitive p^2-root of unity ε in C such that $C = E(\varepsilon)$:

If $\zeta^i s \in E$ for some i, then $(\zeta^i s)^p \in E$. This would mean that $\big((\zeta^i s)^p\big)^p = s^{p^2} = a$ and $t^p - a$ has a root in E contradicting the irreducibility of $t^p - a \in E[t]$. Consequently, $\zeta^i s \notin E$ for all i. This means that every irreducible factor of $t^{p^2} - a \in E[t]$ must have degree p. Let f be such an irreducible factor, say the minimal polynomial of $\zeta^i s$. The constant term of f must be $b = \zeta^n s^p \in E$ for some n. As $s^p \notin E$, we have

$$C = E(s^p) = E(b\zeta^{-n}) = E(\zeta^n).$$

Since E contains a primitive pth root of unity, $\varepsilon = \zeta^n$ must be a primitive p^2-root of unity as needed.

Let Δ be the prime subfield in C.

Claim 70.7. There exists an r such that $\Delta(\varepsilon)$ contains a primitive p^rth root of unity but not a primitive p^{r+1}th primitive root of unity:

Let ζ_{p^r} be a primitive p^rth root of unity in C. Suppose that $\Delta \cong \mathbf{Q}$. Then $[\mathbf{Q}(\zeta_{p^r}) : \mathbf{Q}] = p^{r-1}(p-1) \to \infty$ as $r \to \infty$. As $[\mathbf{Q}(\varepsilon) : \mathbf{Q}] < \infty$, Claim 70.7 holds in this case. Suppose that $\Delta \cong \mathbf{Z}/l\mathbf{Z}$. By Claim 70.5, we know that $l \neq p$. As $t^{p^r} - 1$ has no repeated roots in C, we have $|\Delta(\zeta_{p^r})| \geq p^r \to \infty$. As $|\Delta(\varepsilon)| < \infty$, Claim 70.7 also holds if Δ has positive characteristic.

Let r be as in Claim 70.7. As $\varepsilon \in F(\varepsilon)$ and ε is a primitive p^2-root of unity, we have $r \geq 2$. Let ω be a primitive p^{r+1}th root of unity in the algebraically closed field C. Let N be the Galois group of $\Delta(\omega)/\Delta$. If Δ is a finite field, then N is cyclic. If $\operatorname{char}\Delta = 0$, then $N \cong (\mathbf{Z}/p^{r+1}\mathbf{Z})^\times$, which is cyclic unless $p = 2$ and $r \geq 2$. (Cf. Theorem 105.6.) Moreover, if N is cyclic, it contains a unique subgroup of order p, hence $\Delta(\omega)$ contains a unique subfield over which it has degree p.

Claim 70.8. $\Delta(\omega)$ contains two subfields over which it has degree p. In particular, $p = 2$ and $\operatorname{char} E = 0$:

Let $f \in E[t]$ be the minimal polynomial of ω. Since $\varepsilon \notin E$ also $\omega \notin E$. Hence $C = E(\omega)$ and $\deg f = p$. We also have $f \mid t^{p^{r+1}} - 1$ in $E[t]$ and $t^{p^{r+1}} - 1 = \prod_{i=0}^{p^{r+1}-1}(t - \omega^i)$ in $C[t]$. Let $K = \Delta(\omega) \cap E$. Then by the above, $f \in K[t]$ so f is the minimal polynomial of ω over K. Consequently, $[\Delta(\omega) : K] = p$ and K is one of the desired subfields.

Let $z = \omega^p$, a primitive p^rth root of unity and $K' = \Delta(z)$. As $\Delta(\omega)$ contains a primitive p^rth root of unity and hence all such, we have $K' = \Delta(z) \subset \Delta(\omega)$. We show that $[\Delta(\omega) : \Delta(z)] = [\Delta(\omega) : K'] = p$. As ω is a root of $t^p - z \in K'[t]$, it suffices to show $t^p - z \in K'[t]$ is irreducible. By Fact (4), it suffices to show that $t^p - z$ has no roots in K'. Since $z \in K'$, certainly K' contains all the primitive pth roots of unity. In particular, if $t^p - z$ has a root in K', it splits in $K'[t]$ and $\omega \in K'$. But this means that $\Delta(\omega) = K' \subset \Delta(\varepsilon)$ contradicting Claim 70.7. Thus $[\Delta(\omega) : K'] = p$.

To prove Claim 70.8, we need only show that $K \neq K'$. Suppose that $K = K'$. Then $z \in K' = K$ means that K contains a primitive p^rth root of unity. As $K = \Delta(\omega) \cap E$, we have $z \in E$, i.e., E contains

a primitive p^rth root of unity for some $r \geq 2$, hence it contains a p^2-root of unity. Thus $\varepsilon \in E$ which is a contradiction. We conclude that $K \neq K'$ and Claim 70.8 is established.

Thus we have shown that $\operatorname{char}\Delta = 0$ and $C = E(\varepsilon)$ with ε a primitive 2^2 root of unity, i.e, $\varepsilon = \pm\sqrt{-1}$. But $\sqrt{-1} \in E$ by Claim 70.5, a contradiction and the theorem is proved. □

Let F be a field and C an algebraic closure. Then

$$F_{sep} := \{x \mid x \in C \text{ separable over } x\},$$

is called the *separable closure* of F. It is the maximal separable field extension of F in C. The Galois group $G(F_{sep}/F)$ is called the *absolute Galois group* of F.

Corollary 70.9. *Let F be a field. If the absolute Galois group of F contains a nontrivial element of finite order, then F is formally real and the element is an involution.*

Proof. By Theorem 70.2 the result follows if F is perfect, so we may assume that $\operatorname{char} F = p > 0$. Let C be an algebraic closure of F and F_{sep} the separable closure of F in C. Suppose that $C/K/F_{sep}$ is an intermediate field and $\sigma : K \to C$ an embedding. Then σ induces a unique isomorphism $K[t]/(t^{p^n} - a) \to \sigma(K)[t]/\big(t^{p^n} - \sigma(a)\big)$ taking the unique root of $t^{p^n} - a$ in $K[t]/(t^{p^n} - a)$ in C to the corresponding root of $t^{p^n} - \sigma(a)$ in $\sigma(K)[t]/\big(t^{p^n} - \sigma(a)\big)$ for all n and all $a \in K$. It follows that if $\sigma : F_{sep} \to F_{sep}$ is an F-automorphism of order r, then σ lifts to a unique F-automorphism of C of order r. The result follows by Theorem 70.2. □

Corollary 70.10. *If the absolute Galois group of a field is finite, then it is isomorphic to* 1 *or* $\mathbb{Z}/2\mathbb{Z}$.

Exercise 70.11. Prove Facts 70.1.

71. Hilbert's 17th Problem

In this section, we present a solution to Hilbert's 17th Problem on positive functions over the reals. We establish the Lang Homomorphism Theorem to solve this problem. We present the proof of the Lang Homomorphism Theorem based on Sylvester's Lemma 69.9. To use it, we shall need the fact (which we do not prove) that the trace form of an algebra over a field of characteristic zero can be diagonalized, i.e., has a diagonal matrix representation. This is in fact true for all symmetric bilinear forms over fields of characteristic not two. We prove this for the special case that we need.

Lemma 71.1. *Let F be a field of characteristic zero and K/E be a finite field extension of degree n. Then there exists an E-basis $\mathcal{B}$ for K such that the matrix representation of the trace form $\varphi : K \times K \to E$ in the basis $\mathcal{B}$ is diagonal.*

PROOF. We must produce an E-basis $\{x_1, \ldots, x_n\}$ for K satisfying $\varphi(x_i x_j) = \delta_{ij}$ for all $i, j = 1, \ldots, n$ As K is a field and E perfect, the trace form is just the non-degenerate field trace $\mathrm{Tr}_{K/E} : K \times K \to E$ by Lemma 73.1. In particular, if x is a nonzero element in E, then there exists a y in E with $\mathrm{Tr}_{K/E}(xy)$ nonzero. As

$$\mathrm{Tr}_{K/E}(xy) = \tfrac{1}{2}\,\mathrm{Tr}_{K/E}\left((x+y)^2\right) - \mathrm{Tr}_{K/E}(x^2) - \mathrm{Tr}_{K/E}(y^2)$$

one of $\mathrm{Tr}_{K/E}\left((x+y)^2\right)$, $\mathrm{Tr}_{K/E}(x^2)$, $\mathrm{Tr}_{K/E}(y)^2)$ is non-zero, i.e., the non-degeneracy of $\mathrm{Tr}_{K/E}$ insures the existence of an element z in K with $\mathrm{Tr}_{K/E}(z^2)$ nonzero. (Of course, we know this is true for the element $z = 1$.) As $\mathrm{Tr}_{K/E}(zx) = 0$ for x in K implies that $x = 0$, the restriction of $\mathrm{Tr}_{K/E}$ to $(Fz)^\perp = \{x \mid \mathrm{Tr}_{K/E}(zy) = 0\}$, gives an 'orthogonal' decomposition $K = Ez \perp (Ez)^\perp$. Repeating the argument yields the result. □

Theorem 71.2. (Lang Homomorphism Theorem) *Let F be a real closed field and $K = F(x_1, \ldots, x_n)$ a proper finitely generated formally real field extension of F (hence not algebraic). Then there exists infinitely many F-algebra homomorphisms $F[x_1, \ldots, x_n] \to F$, i.e., if R is an affine F-algebra that is a domain, then there exist infinitely many F-algebra homomorphisms $R \to F$.*

PROOF. Let $m = \mathrm{tr\ deg}\ K/F$. By hypothesis, $m > 0$. We induct on m.

Suppose that $m = 1$. We may assume that $x = x_1$ is transcendental over F. As $\mathrm{char}\, F = 0$, we can write $K = F(x)[y]$ for some $y \in K$ and furthermore we may assume that y is integral over $F[x]$ (cf. Remark 72.14(1)).

Claim 71.3. There exist (infinitely many) F-algebra homomorphisms $\sigma : F[x, y] \to F$:

Let $f = f(x,t) = t^r + a_{r-1}(x)t^{r-1} + \cdots + a_0(x)$ with the $a_i(x) \in F[x]$ be the minimal polynomial of integral y over $F[x]$. It suffices to show that there exist (infinitely many) $(a, b) \in F^2$ such that $f(a, b) = 0$, since for each such (a, b) the map $\sigma : F[x, y] \to F$ via $x \mapsto a$ and $y \mapsto b$ is well-defined. Let $P \in \mathcal{X}(K)$ and let $\widetilde{K}$ be a real closure of (K, P). So we have $P \cap K = K^2$. Let φ be the trace form of the $F(x)$-algebra K. Since f has a root y in $\widetilde{K}$, by Sylvester's Lemma 69.9, we have

$\operatorname{sgn}_P \varphi > 0$. As $r = [K : F(x)]$ and φ can be diagonalized over $F(x)$, we may assume that there is an $F(x)$-basis $\mathcal{B}$ for K such that

$$[\varphi]_{\mathcal{B}} \quad = \quad \begin{pmatrix} h_1(x) & 0 & \\ 0 & \ddots & \\ & & h_r(x) \end{pmatrix}$$

is diagonal for some $h_i(x) \in F(x)$. Moreover, modifying this basis, we may also assume that all the $h_i(x) \in F[x]$ and are square-free. As F is real closed, each of the $h_i(x)$ can be factored as a product of a constant in F, monic linear polynomials, and monic irreducible quadratic polynomials in x. Each irreducible monic polynomial $x^2 + ax + b = (x + \frac{a}{2})^2 + (b - \frac{a^2}{4})$, a sum of two squares in $F(x)$ hence totally positive as is $c^2 + ac + b$ for all $c \in F$ by substitution. Thus each h_i is a product of a totally positive or totally negative element and finitely many positive factors $x - a$ relative to P and finitely many negative factors $x - b$ relative to P for some $a, b \in F$.

Among all the finitely many a's occurring in all the h_i, let a_0 be the maximum relative to the ordering $F^2 = P \cap F$ and among all the finitely many b's occurring in all the h_i, let b_0 be the minimum relative to the ordering $P \cap F$. (If there are no a's (respectively, b's) choose a_0 (respectively, b_0) such that $a_0 <_P b_0$ in F.) Thus we have $x - a_0 \leq_P x - a$ for all such a and $x - b \leq_P x - b_0$ for all such b. As $x - b_0 <_P x - a_0$, we must also have $a_0 <_P b_0$. Thus there exist infinitely many $c \in F$ with $a_0 <_P c <_P b_0$ (e.g., $a_0 + \dfrac{b_0 - a_0}{2^N}$ for $N > 0$). For each such c, we have $c - a >_P 0$ and $x - a >_P 0$ for all the a's and $c - b <_P 0$ and $x - b <_P 0$ for all the b's. It follows that

$$\operatorname{sgn}_P \varphi = \operatorname{sgn}_{P \cap F} [\varphi]_{\mathcal{B}} \quad = \quad \operatorname{sgn}_{P \cap F} \begin{pmatrix} h_1(c) & 0 & \\ 0 & \ddots & \\ & & h_r(c) \end{pmatrix}$$

for any c satisfying $a_0 <_P c <_P b_0$.

Assertion. Let $c \in F$ satisfy $a_0 <_P c <_P b_0$ and φ_c be the trace form of the F-algebra $A = F[t]/\big(f(c,t)\big)$. Then, except for finitely many exceptions, there exists an F-basis $\mathcal{C}$ for A satisfying

$$[\varphi_c]_{\mathcal{C}} \quad = \quad \begin{pmatrix} h_1(c) & 0 & \\ 0 & \ddots & \\ & & h_r(c) \end{pmatrix}.$$

In particular, for each such c, we have $\operatorname{sgn}_{P \cap F} \varphi_c > 0$.

Let $B = \big(b_{ij}(x)\big)$ be the matrix of the trace form φ relative to the basis $\{1, y, \dots, y^{r-1}\}$ of K over $F(x)$. As y is integral over $F[x]$, each $b_{ij}(x) \in F[x]$. There exists a change of basis matrix $C = \big(c_{ij}(x)\big) \in \mathrm{GL}_r(K)$ such that

$$C^t B C = \begin{pmatrix} h_1(x) & 0 & \\ 0 & \ddots & \\ & & h_r(x) \end{pmatrix}.$$

If c is not a zero of the denominator of any of the $c_{ij}(x)$'s or a zero of $\det C$, then

$$C|^t_{x=c} B|_{x=c} C|_{x=c} = \begin{pmatrix} h_1(c) & 0 & \\ 0 & \ddots & \\ & & h_r(c) \end{pmatrix}$$

as desired. As there exist only finitely many c not satisfying these conditions, the assertion follows.

As F is real closed, for each c in the assertion satisfying $\mathrm{sgn}_{P\cap F}\, \varphi_c > 0$, there exists a $b \in F$ such that $f(c, b) = 0$ by Sylvester's Lemma 69.9. Thus there exist infinitely many (c, b) such that $f(c, b) = 0$ and the Claim is established.

Next we show that almost all of the F-algebra homomorphisms $\sigma : F[x, y] \to F$ extend to $F[x, y, x_2, \dots, x_n] \to F$. As the transcdence degree $m = \mathrm{tr\ deg}_F K = 1$ and $F(x)[y] = F(x_1, \dots x_n)$, we can write

$$x_i = \frac{g_i(x, y)}{q_i(x)} \text{ for some } g_i \in F[x, t], \quad 0 \neq q_i \in F[x].$$

Let $q = \prod_i q_i$. As $q(c) \neq 0$ for almost all of the c constructed above, $\sigma\big(q(x)\big) \neq 0$ for almost all the σ constructed above. Let σ be any such one. By properties of localization with $S = \{t^i \mid i \geq 0\}$m, any such σ extends to an F-algebra homomorphism $\sigma : F[x, y][\frac{1}{q}] \to F$ by Exercise 26.3(6). Since $F[x_1, \dots, x_n] \subset F[x_1, y, x_2, \dots, x_n] \subset F[x, y][\frac{1}{q}]$, the $m = 1$ case is completed.

Suppose that $m > 1$. Choose $K/E/F$ satisfying $\mathrm{tr\ deg}_E K = 1$. Let $\widetilde{E}$ be the algebraic closure of E in a real closure $\widetilde{K}$ of K. By the Fundamental Theorem of Algebra, $\widetilde{E}$ is also real closed. As $\mathrm{tr\ deg}_{\widetilde{E}} \widetilde{E}(x_1, \dots, x_n) = 1$, there exist infinitely many $\widetilde{E}$-algebra homomorphisms $\sigma : \widetilde{E}(x_1, \dots, x_n) \to \widetilde{E}$ by the $m = 1$ case. Let σ be any one of these. If $\sigma(x_i) \in F$ for all i, then $\sigma|_{F[x_1,\dots,x_n]} : F[x_1, \dots, x_n] \to F$ is an F-algebra homomorphism. So suppose that there exists an i with

$\sigma(x_i) \notin F$. As $\sigma(x_i) \in \widetilde{E}$ for all i, we have tr $\deg_F F\big(\sigma(x_1), \ldots, \sigma(x_n)\big) \leq$ tr $\deg_F \widetilde{E} =$ tr $\deg_F E < m$. By induction there exist infinitely many F-algebra homomorphisms $\tau : F[\sigma(x_1), \ldots, \sigma(x_n)] \to F$. Then the compositions $\tau \circ \sigma|_{F[x_1,\ldots,x_n]} : F[x_1, \ldots, x_n] \to F$ give infinitely many F-algebra homomorphisms. □

A commutative ring is called *semi-real* if -1 is not a sum of squares in R and is called *formally real* if $\sum x_i^2 = 0$ in R implies that $x_i = 0$ for all i. More generally, an ideal $\mathfrak{A}$ in a commutative ring is called *semi-real* (respectively, *formally real*) if $R/\mathfrak{A}$ is semi-real (respectively, formally real). We note that if R is a domain, then it is formally real if and only if its quotient field is and a field is semi-real if and only if formally real.

Lemma 71.4. *Let R be a commutative ring. Then the following are equivalent*

(1) R is semi-real.
(2) There exists $P \in \mathcal{Y}(R)$ such that $P \cap -P$ is a prime ideal.
(3) There exists a formally real prime ideal $\mathfrak{p}$ in R.
(4) There exists a ring homomorphism $R \to F$ for some formally real field F.

Proof. (1) ⇒ (2): As R is semi-real, we have $\sum R^2 \in \mathcal{Y}(R)$. Thus (2) follows by Lemma 68.4.

(2) ⇒ (3): The prime ideal $P \cap -P$ in (2) is clearly semi-real and excludes the multiplicative set $S = 1 + \sum R^2$. By choice $0 \notin S$ and $S + \sum R^2 \subset S$. By Zorn's lemma there exists an ideal $\mathfrak{p}$ containing $P \cap -P$ and excluding S. It is prime (as is well-known or easy to check). The quotient field of $R/\mathfrak{p}$ is checked to be semi-real hence formally real. It follows that $R/\mathfrak{p}$ is formally real.

(3) ⇒ (4): The quotient field of $R/\mathfrak{p}$ works.

(4) ⇒ (1): As a quotient of R is semi-real so is R. □

Corollary 71.5. *Let F be a real closed field and R a domain that is an affine F-algebra. If R is semi-real the there exists an F-algebra homomorphism $R \to F$.*

Proof. There exists a formally real prime ideal $\mathfrak{p}$ in R. Then $R/\mathfrak{p}$ is a real F-affine ring that is a domain so there exists an F-algebra homomorphism $R/\mathfrak{p} \to R$ by the Lang Homomorphism Theorem 71.2. The composition $R \to R/\mathfrak{p} \to R$ now works. □

Corollary 71.6. *Let F be a real closed field and R a real affine F-algebra that is a domain. Let $f_1, \ldots, f_n \in R \setminus \{0\}$. Then there exists an F-algebra homomorphism $\varphi : R \to F$ so that $\varphi(f_i) \neq 0$ for all i.*

PROOF. $R[f_1^{-1}, \dots f_n^{-1}]$ is a real affine F-algebra that is a domain. □

Note that $R = \mathbf{R}[t_1, \dots, t_n]/(t_1^2 + \cdots + t_n^2)$ is a semi-real affine $\mathbf{R}$-algebra that is a domain but it is not formally real since any $\mathbf{R}$-algebra homomorphism $R \to \mathbf{R}$ must take each $t_i \mapsto 0$.

Corollary 71.7. *Let F be a real closed field and R an affine F-algebra. Let $f_1, \dots, f_r$, $g_1, \dots, g_s$ lie in R. Suppose that there exists a maximal preordering $P \in \mathcal{Y}(R)$ such that $0 \neq f_i \in P$ for all i and $g_j \in P$ for all j. Then there exists an F-algebra homomorphism $\varphi : R \to F$ so that $0 \neq \varphi(f_i) \in F^2$ and $\varphi(g_j) \in F^2$ for all i and j.*

PROOF. Let $\mathfrak{p} = P \cap -P$, a prime ideal in R. Replacing R by $R/\mathfrak{p}$ we may assume that $P \cap -P = \mathfrak{p} = 0$. In particular, R is a domain and P is an ordering on R. Clearly, P extends to an ordering of the quotient field K of R. Also call this extension P. We have $f_i >_P 0$ and $g_j \geq_P 0$ for all i and j. Let $E = K(\sqrt{f_1}, \dots, \sqrt{f_r}, \sqrt{g_1}, \dots, \sqrt{g_s})$. By induction on $r + s$, the ordering P extends to an ordering Q on E. Since the domain $A = R[f_1^{-1}, \dots, f_r^{-1}, \sqrt{f_1}, \dots, \sqrt{f_r}, \sqrt{g_1}, \dots, \sqrt{g_s}]$ is a real affine F-algebra, there exists an F-algebra homomorphism $A \to F$. This homomorphism has the desired properties. □

If $\mathfrak{A}$ be an ideal in $F[t_1, \dots, t_n]$ with F infinite, let

(1) $Z_F(\mathfrak{A}) = \{x \in F^n \mid f(x) = 0 \text{ for all } f \in \mathfrak{A}\}$.
(2) $A_F(\mathfrak{A}) = F[t_1, \dots, t_n]/\mathfrak{A}$.

Suppose that $\mathfrak{p} = \mathfrak{A}$ is a prime ideal in $F[t_1, \dots, t_n]$. So $A_F(\mathfrak{p})$ is a domain. As usual we say a rational function f in the quotient field of $A_F(\mathfrak{p})$ is *defined* at $x \in Z_F(\mathfrak{p})$ if $f = g/h$ for some $g, h \in A_F(\mathfrak{A})$ with $h(x) \neq 0$. If F is real closed, a rational function f in the quotient field of $A_F(\mathfrak{p})$ is called *positive semi-definite* if $f(x) \geq 0$ for all points $x \in A_F(\mathfrak{p})$ where f is defined.

Corollary 71.8. (Weak Real Nullstellensatz) *If F is real closed and $\mathfrak{A}$ is an ideal in $F[t_1, \dots, t_n]$, then $\mathfrak{A}$ is semi-real if and only if $Z_F(\mathfrak{A}) \neq \emptyset$.*

PROOF. Let $A_F(\mathfrak{A}) = F[x_1, \dots, x_n]$. If $(a_1, \dots, a_n) \in Z_F(\mathfrak{A})$, then $f : A_F(\mathfrak{A}) \to F$ by $x_i \to a_i$ defines an F-algebra homomorphism and $\mathfrak{A}$ is semi-real by Lemma 71.4. Suppose that $\mathfrak{A}$ is semi-real. By Lemma 71.4 there exists a prime ideal $\mathfrak{p}$ in $F[t_1, \dots, t_n]$ such that $\mathfrak{p}/\mathfrak{A}$ is a formally real prime ideal in $A_F(\mathfrak{A})$ It follows that $F[t_1, \dots, t_n]/\mathfrak{p}$ is a formally real domain so by the Lang Homomorphism Theorem 71.2 there exists an F-algebra homomorphism $\sigma : F[t_1, \dots, t_n]/\mathfrak{p} \to F$. Then $(a_1, \dots, a_n) \in Z_F(\mathfrak{A})$ with $a_i = \sigma(t_i + \mathfrak{p})$. □

Theorem 71.9. (Artin) *Let F be a real closed field and $\mathfrak{p}$ a formally real prime ideal in $F[t_1, \ldots, t_n]$. Let K be the quotient field of $A_F(\mathfrak{p})$. Let $f \in K$. If f is positive, then f is a sum of squares in K.*

PROOF. Suppose that f is not a sum of squares in K. Then there exists $P \in \mathcal{X}(K)$ such that $f <_P 0$ by Corollary 68.6. Thus P extends to $E = K(\sqrt{-f})$ by Theorem 69.2. Let $f = g/h$, with $g, h \in A_F(\mathfrak{p})$ and let $^- : F[t_1, \ldots, t_n] \to A_F(\mathfrak{p})$ be the canonical map. By the Lang Homomorphism Theorem, there exists an F-algebra homomorphism $\sigma : F[\bar{t}_1, \ldots \bar{t}_n, \sqrt{-g}, \dfrac{1}{gh}] \to F$. For each i, let $a_i = \sigma(\bar{t}_i)$. If $q \in \mathfrak{p}$ then

$$q(a_1, \ldots, a_n) = q\big(\sigma(\bar{t}_1), \ldots, \sigma(\bar{t}_n)\big) = \sigma\big(q(\bar{t}_1, \ldots, \bar{t}_n)\big) = 0,$$

so $(a_1, \ldots, a_n) \in Z_F(\mathfrak{p})$. As $\sigma(g) \neq 0$ and $\sigma(h) \neq 0$, the rational function f is defined at $(a_1, \ldots, a_n)$ and is not zero on it. Since $\sigma(-f) \in F^2$, we must have $\sigma(f) <_P 0$. But $\sigma(f) = f(a_1, \ldots, a_n) >_P 0$, a contradiction. □

Artin's Theorem immediately answers Hilbert's 17th problem whether every positive semi-definite function f in $\mathbf{R}(t_1, \ldots, t_n)$ is a sum of squares in K in the affirmative.

Corollary 71.10. (Hilbert's 17th Problem) *Any positive semi-definite function in $\mathbf{R}(t_1, \ldots t_n)$ is a sum of squares in $\mathbf{R}(t_1, \ldots, t_n)$.*

The Motzkin polynomial $t_1^4 t_2^2 + t_1^2 t_2^4 - 3t_1^2 t_2^2 + 1$ in $\mathbf{R}[t_1, t_2]$ is positive semi-definite but not a sum of squares in $\mathbf{R}[t_1, t_2]$. (It is a sum of four squares, but not three squares, in $\mathbf{R}(t_1, t_2)$. Cf. Lam's book *Introduction to quadratic forms over fields* for more detail.)

CHAPTER XV

Dedekind Domains

The study of fields arose from the desire to understand the roots of polynomials with rational coefficients. In a similar way, algebraic number theory arose from the desire to understand the roots of monic polynomials with integer coefficients. In this chapter, we give an introduction to algebraic number theory. Since the generalization to studying Dedekind domains – an appropriate generalization of the integers – is not too difficult, we take this more general approach.

72. Integral Elements

We begin with the study of monic polynomials over (nonzero) commutative rings.

Definition 72.1. Let B be a nonzero commutative ring and A a subring of B. We shall write this as B/A and call it a *ring extension* mimicking how we wrote field extensions. An element x in B is called *integral over* A if x is a root of a monic polynomial f in $A[t]$.

Examples 72.2. 1. Let K/F be an extension of fields. Then an element x in K is integral over F if and only if it is algebraic over F.

2. If A is a nonzero commutative ring, then every element x in A is integral over A as it is a root of $t - x$ in $A[t]$.

3. Let d be a square-free integer and $x = a + b\sqrt{d}$ with a and b in $\mathbb{Q}$. Then x is a root of $t^2 - 2at + (a^2 - b^2 d)$. In particular, one checks that x is integral over $\mathbb{Z}$ if and only if $2a$ and $a^2 - b^2 d$ are integers. For example, the complex numbers $\sqrt{2}$ and $(-1 + \sqrt{-3})/2$ are integral over $\mathbb{Z}$.

4. Every nth root of unity in $\mathbb{C}$, $n > 0$, is integral over $\mathbb{Z}$.

5. Let $C/B/A$ be ring extensions. If x in C is integral over A, then it is integral over B.

Because of the first example, we expect that some of our field theoretic ideas and results should generalize. This is true. First, we must

find a ring theoretic test for an element to be integral. It is given by the following basic result.

Proposition 72.3. *Let B/A be an extension of nonzero commutative rings and x an element in B. Then x is integral over A if and only if $A[x]$ is a finitely generated A-module.*

PROOF. ($\Rightarrow$): Let x be a root of the monic polynomial f in $A[t]$. By the General Division Algorithm 31.4, we can write

$$g = fq + r \text{ with } q, r \in A[t] \text{ and } r = 0 \text{ or } \deg r < \deg f.$$

Therefore, if $n = \deg f$, we have

$$g(x) = r(x) \text{ lies in } \sum_{i=0}^{n-1} Ax^i, \text{ so } A[x] = \sum_{i=0}^{n-1} Ax^i.$$

($\Leftarrow$): We prove a more general statement.

Claim 72.4. Let M be an $A[x]$-module that is finitely generated as an A-module. If $\operatorname{ann}_{A[x]} M = 0$ (i.e., if r in $A[x]$ satisfies $rm = 0$ for all m in M, then $r = 0$), then x is integral over A:

Suppose that $M = Ax_1 + \cdots + Ax_n$. By assumption, M is an $A[x]$-module, i.e.,

$$\lambda_x : M \to M \text{ given by } m \mapsto xm$$

is an A-endomorphism. For each i, $i = 1, \ldots, n$, we can write

$$xm_i = \sum_{j=1}^{n} a_{ij} m_j \text{ some } a_{ij} \in A,\ 1 \le i, j \le n.$$

Therefore,

$$\sum_{j=1}^{n} (\delta_{ij} x - a_{ij}) m_j = 0 \text{ for } i = 1, \ldots, n.$$

Let Δ be the determinant of the $n \times n$ matrix $(\delta_{ij}x - a_{ij})$. By Cramer's Rule, $\Delta m_i = 0$ for $i = 1, \ldots, n$, so $\Delta M = 0$, i.e., Δ lies in $\operatorname{ann}_{A[x]} M = 0$. It follows that x is a root of the monic polynomial $\det((\delta_{ij}t - a_{ij}))$ in $A[t]$. This proves the claim.

Let $M = A[x]$, a finitely generated A-module by assumption. As $A[x]$ is a subring of B it has a one (not zero). Consequently, $\operatorname{ann}_{A[x]} A[x] = 0$, so the claim yields the desired result. □

Corollary 72.5. *Let B/A be an extension of nonzero commutative rings, $x_1, \ldots, x_n$ elements of B. Then $x_1, \ldots, x_n$ are all integral over A if and only if $A[x_1, \ldots, x_n]$ is a finitely generated A-module.*

PROOF. ($\Rightarrow$): The case for $n = 1$ follows from the proposition. By induction, $A[x_1, \ldots, x_{n-1}]$ is a finitely generated A-module, say with generating set S. As x_n is integral over A, it is a root of a monic polynomial in $A[t]$, say of degree d. Then $S \cup \{1, x_n, \ldots, x_n^{d-1}\}$ is a generating set for $A[x_1, \ldots, x_n]$.
($\Leftarrow$): Since $\operatorname{ann}_{A[x_i]} A[x_1, \ldots, x_n] = 0$ for $i = 1, \ldots, n$, this follows from Claim 72.4. □

In the above, we could also have used the following lemma that we leave as an exercise. We shall need it below.

Lemma 72.6. *Let B/A be an extension of nonzero commutative rings. If M is a finitely generated B-module and B a finitely generated A-module, then M is a finitely generated A-module.*

Note the corollary above is analogous with the result that when K/F is a field extension, $x_1, \ldots, x_n$ elements in K, then $F(x_1, \ldots, x_n)/F$ is finite if and only if $x_1, \ldots, x_n$ are all algebraic over F. In a similar way we have the following analogue of another field theoretical result.

Corollary 72.7. *Let B/A be an extension of nonzero commutative rings, x and y elements of B both integral over A. Then $x \pm y$, and xy are integral over A. In particular,*

$$\{x \in B \mid x \text{ integral over } A\}$$

is a subring of B.

PROOF. If z is any element in $A[x, y]$, then $A[x, y, z] = A[x, y]$ is a finitely generated A-module, so z is integral over A. □

Of course, if B/A is an extension of commutative rings with x in B a unit in B, it may be integral over A but its inverse not. For example, $1/2$ is not integral over $\mathbb{Z}$.

Definition 72.8. Let B/A be an extension of nonzero commutative rings and

$$C = \{x \in B \mid x \text{ integral over } A\}.$$

The ring C is called the *integral closure* of A in B. If $C = B$, i.e., every element of B is integral over A, we say B/A is *integral* and if $C = A$, we say that A is *integrally closed* in B.

Therefore, B/A being integral is the analogue of an algebraic extension in field theory. Analogous to the field case, we have:

Corollary 72.9. *Let C/B and B/A be extensions of nonzero commutative rings. Then C/B and B/A are integral extensions if and only if C/A is an integral extension.*

PROOF. We need only show if C/B and B/A are integral extensions so is C/A. Let c be an element in C. As c is integral over B, it satisfies an equation $c^n + b_{n-1}c^{n-1} + \cdots + b_0 = 0$ in B for some $b_0, \ldots, b_{n-1}$ in B. Let $B_0 = A[b_0, \ldots, b_{n-1}]$. As B/A is integral, it follows that $A[c, b_0, \ldots, b_{n-1}]$ is a finitely generated B_0-module by Proposition 72.3 and B_0 is a finitely generated A-module by Corollary 72.5. The result now follows using Lemma 72.6. □

Corollary 72.10. *Let $C/B/A$ be extensions of nonzero commutative rings with B/A integral. Then the integral closure of A in C is the same as the integral closure of B in C.*

We shall mostly be interested in the case that our rings are domains. For this case, we shall use the following notation:

Notation 72.11. If A is a domain and F a field containing A, we shall denote the integral closure of A in F by A_F, i.e.,

$$A_F := \{x \in F \mid x \text{ is integral over } A\}.$$

Definition 72.12. Let A be a domain. We say that A is *integrally closed* or a *normal domain* if A is integrally closed in its quotient field, i.e., if F is the quotient field of A, then $A = A_F$.

An important example of integrally closed domains is given by:

Proposition 72.13. *A UFD is integrally closed. In particular, any PID is integrally closed.*

PROOF. Let A be a UFD and a and b be elements in A with b nonzero. Suppose that a/b is integral over A. As A is a UFD, we may assume that a and b are relatively prime. By definition, a/b satisfies an equation

$$(\frac{a}{b})^n + c_1(\frac{a}{b})^{n-1} + \cdots + c_n = 0$$

for some $c_1, \ldots, c_n$ in A. Therefore, we have the equation

$$a^n = -b(c_1a^{n-1} + \cdots + b^{n-1}c_n) \text{ in } A.$$

It follows that $b \mid a^n$ in the UFD A, hence b is a unit in A and a/b lies in A. □

Remarks 72.14. Let A be a domain, F its quotient field, and $L/K/F$ field extensions.

1. If x in L is algebraic over F, then there exists a nonzero element c in A satisfying cx is integral over A, i.e., lies in A_L. Indeed if x satisfies
$$a_nx^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0$$

in L with $a_0, \dots, a_n$ elements of A and a_n nonzero, then multiplying this equation by a_n^{n-1} shows that $a_n x$ is integral over A.

2. We have $A_L \cap K = A_K$.
3. The quotient field of A_K is K by the first remark.

The case in which we are most interested is when A is the ring of integers. If $K/\mathbb{Q}$ is a finite field extension, then K is called an *algebraic number field* and $\mathbb{Z}_K$ is called the *ring of algebraic integers in* K. Let $\widetilde{\mathbb{Q}}$ be the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$. Then algebraic number theory is the study of $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$. As $\widetilde{\mathbb{Q}}$ is the union of number fields, we have $\mathbb{Z}_K = \mathbb{Z}_{\widetilde{\mathbb{Q}}} \cap K$ for any number field K. Since $\mathbb{Q}$ is perfect, we shall not have to worry about separability that in the general case can cause serious problems.

Exercises 72.15. 1. Let d be a square-free integer and $x = a + b\sqrt{d}$ with a and b in $\mathbb{Q}$. Show that x is integral over $\mathbb{Z}$ if and only if $2a$ and $a^2 - b^2 d$ are integers.

2. Let $\varphi : A \to B$ be a ring homomorphism and M a finitely generated B-module. If B is a finitely generated A-module via the pullback, then M is a finitely generated A-module. In particular, Lemma 72.6 is valid.
3. Let $\varphi : A \to B$ be a ring homomorphism of commutative rings. We say that φ is *integral* if $B/\varphi(A)$ is integral. Show that a composition of integral homomorphisms is integral.
4. Let $\varphi : A \to B$ be a ring homomorphism of commutative rings. We say that φ is *finite* if B is a finitely generated A-module (via φ) and of *finite type* if there exists a ring epimorphism $A[t_1, \dots, t_n] \to B$ for some n lifting φ. Show that φ is finite if and only if φ is integral and of finite type.
5. Let A be a domain with quotient field F and K/F a field extension. If S is a multiplicative set (not containing zero) in A, show that $(S^{-1}A)_K = S^{-1}(A_K)$, i.e., taking localization and integral closures commute. In particular, this applies when $S = A \setminus \mathfrak{p}$ with $\mathfrak{p}$ a prime ideal in A.

73. Integral Extensions of Domains

In this section, we shall investigate the integral closure of a domain in a separable extension of its quotient field. Separability makes life much nicer, and since we are mostly interested in applications to algebraic number theory, this will not be an impairment.

When we investigated cyclic extensions of fields, we needed to use the norm of such an extension. Although we introduced the notion of

trace at the same time, except for exercises, we did not use it. Although the norm and trace of an arbitrary, i.e., not necessarily separable, finite extension can be defined, this is not useful for the trace. Indeed, the trace from a finite inseparable extension turns out to be the zero map. However, for the separable case, the trace is very useful. This was because of Exercise 57.24(5). Because this is important to our current investigation, we now prove this exercise.

Lemma 73.1. *Let K/F be a finite separable extension of fields of degree n, $K^* := \mathrm{Hom}_F(K, F)$, the F-linear dual space of K. Then the map $T : K \to K^*$ defined by $x \mapsto \mathrm{tr}_x : y \mapsto \mathrm{Tr}_{K/F}(xy)$ is an F-linear isomorphism. In particular, if $\mathcal{B} = \{w_1, \ldots, w_n\}$ is an F-basis for K, then there exists an F-basis $\{w_1', \ldots, w_n'\}$ for K satisfying $\mathrm{Tr}_{K/F}(w_i w_j') = \delta_{ij}$ for all $i, j = 1, \ldots, n$.*

Remark 73.2. The basis $\{w_1', \ldots, w_n'\}$ in the lemma is called the *complementary basis* to $\mathcal{B}$

Proof. By Dedekind's Lemma 51.3, the map $\mathrm{Tr}_{K/F} : K \to F$ is nontrivial, so there exists an element z in K with $\mathrm{Tr}_{K/F}(z)$ nonzero. Therefore, $\mathrm{Tr}_{K/F}(xx^{-1}z)$ is not zero for every nonzero x in K. It follows that tr_x is nonzero for all nonzero x in K, i.e., $T : K \to K^*$ is injective, hence an isomorphism by dimension count. Let $\{f_1, \ldots, f_n\}$ be the dual basis to $\mathcal{B}$ and choose w_i' in K to satisfy $T(w_i') = f_i$ for $i = 1, \ldots, n$. Then $\{w_1', \ldots, w_n'\}$ works. □

Suppose that K_i/F is a field extension for $i = 1, 2$ and $\sigma : K_1 \to K_2$ is an F-isomorphism. Let A be a domain lying in F. If x in K_1 is integral over A then $\sigma(x)$ is also integral over A, i.e., $\sigma(A_{K_1}) \subset A_{K_2}$. In particular, if $K = K_1 = K_2$ then $\sigma(A_K) = A_K$. This means that we can use Galois theory. We shall need the results about the trace (and norm) for finite separable extensions as stated in Remark 57.18 as well as Exercise 57.24(3) which we now prove.

Lemma 73.3. *Let K/F be a finite separable extension of fields. Then $\mathrm{N}_{K/F} = \mathrm{N}_{E/F} \circ \mathrm{N}_{K/E}$ and $\mathrm{Tr}_{K/F} = \mathrm{Tr}_{E/F} \circ \mathrm{Tr}_{K/E}$ for any intermediate field $K/E/F$.*

Proof. Let L/F be a finite Galois extension with L/K. Let $\sigma_1, \ldots, \sigma_n : E \to L$ denote all the F-homomorphisms and $\tau_1, \ldots, \tau_m : K \to L$ all the E-homomorphisms. Extend each σ_i to $\widehat{\sigma}_i$ in $G(L/F)$. If $\mu : K \to L$ is an F-homomorphism, then $\mu|_E = \sigma_i$ for some i and $\widehat{\sigma}_i^{-1}\mu = \tau_j$ for some j, hence $\mu = \widehat{\sigma}_i\tau_j$. As in the proof of Proposition 53.11, we see that these are all the distinct F-homomorphisms $K \to L$.

It follows that if x lies in K, then

$$\mathrm{Tr}_{K/F}(x) = \sum \widehat{\sigma}_i \tau_j(x) = \sum \widehat{\sigma}_i(\mathrm{Tr}_{K/E}(x)) = \mathrm{Tr}_{E/F}(\mathrm{Tr}_{K/E}(x))$$

and similarly for the norm. □

Let K/F be finite separable closure with L/K a normal closure of K/F and $\sigma_1, \ldots, \sigma_n : K \to L$ all the F-homomorphisms. If x is an element of K, then any elementary symmetric function in $\sigma_1(x), \ldots, \sigma_n(x)$ is fixed by $G(L/F)$ hence lies in F. It also follows that $\mathrm{N}_{K/F(x)}(x) = x^{[K:F(x)]}$.

Proposition 73.4. *Let A be a domain with quotient field F and K/F a finite separable extension. If x in K is integral over A, then the minimal polynomial $m_F(x)$ of x lies in $A_F[t]$ and both $\mathrm{Tr}_{K/F}(x)$ and $\mathrm{N}_{K/F}(x)$ lie in A_F. In particular, if A is integrally closed, then $m_F(x)$ lies in $A[t]$ and $\mathrm{Tr}_{K/F}(x)$ and $\mathrm{N}_{K/F}(x)$ lie in A.*

Proof. Let L/F be a finite Galois extension satisfying L/K and $\sigma_1, \ldots, \sigma_n : K \to L$ all the distinct F-homomorphisms. If x is an element of K, then any elementary symmetric function in $\sigma_1(x), \ldots, \sigma_n(x)$ is fixed by $G(L/F)$ hence lies in F. Therefore, $m_F(x) = \prod_{i=1}^{n}(t - \sigma_i(x))$ lies in $A_L[t] \cap F[t]$. It follows that $m_F(x)$ lies in $A_F[t]$. By the lemma and Property 57.17(3) (cf. Remark 57.18), we have

$$\mathrm{N}_{F(x)/F}(\mathrm{N}_{K/F(x)}(x)) = (\mathrm{N}_{F(x)/F}(x))^{[K:F(x)]},$$

hence lies in $A_L \cap F = A_F$, and similarly for the trace. □

The result for which we are aiming can now be established.

Theorem 73.5. *Let A be an integrally closed Noetherian domain with quotient field F. Let K/F be a finite separable extension. Then the integral closure of A in K is a finitely generated A-module. In particular, A_K is also an integrally closed Noetherian domain.*

Proof. By Theorem 37.6, we know that any finitely generated A-module is a Noetherian A-module, since A is a Noetherian ring. In particular, it suffices to show there exists a finitely generated A-module M with A_K a submodule of M. Let $n = [K : F]$ and $\mathcal{B} = \{w_1, \ldots, w_n\}$ be an F-basis for K. As K is the quotient field of A_K, by clearing denominators, we may assume that $\mathcal{B}$ lies in A_K. As K/F is a finite separable extension, there exists a complementary basis $\mathcal{B}' = \{w_1', \ldots, w_n'\}$ to $\mathcal{B}$ for K by Lemma 73.1, i.e., $\mathrm{Tr}_{K/F}(w_i w_j') = \delta_{ij}$ for all i and j. For each j, $j = 1, \ldots, n$, there exists a nonzero c_j in A such that $c_j w_j'$ lies in A_K by Remark 72.14(1). Let $c = c_1 \cdots c_n \neq 0$. To finish, it suffices to show the following:

Claim 73.6. A_K is a submodule of $\sum Ac^{-1}w_i$:

Let z be an element of A_K. We can write $z = \sum_{i=1}^{n} b_i w_i$ for some $b_1, \ldots, b_n$ in F. Then for each j, $1 \leq j \leq n$, we have

$$\mathrm{Tr}_{K/F}(czw_j') = c\,\mathrm{Tr}_{K/F}(zw_j') = cb_j.$$

As z and each cw_j' lies in A_K, we conclude that cb_j lies in $A_F = A$ for every j, i.e., $b_j \in c^{-1}A$ for each j. The result follows. □

Remark 73.7. The result is false if we drop the separability assumption. Indeed there exist counterexamples with A a PID with precisely one nonzero prime ideal.

Corollary 73.8. *Let A be a PID with quotient field F and K/F a finite separable field extension. Then A_K is a finitely generated free A-module of rank $[K : F]$.*

Proof. By the theorem, A_K is a finitely generated A-module. Since A_K is a domain, it is a torsion-free A-module, hence A-free by Corollary 41.16 to the Fundamental Theorem of Finite Generated Modules over a PID. Therefore, we need only compute the rank of A_K. We showed in the proof above that there exists an F-basis for K lying in A_K, so the rank of A_K is at most $[K : F]$ using Proposition 41.1. But any A-linearly independent set in A_K is F-linearly independent, so we must have A_K be of rank $[K : F]$. □

Let A be a domain with quotient field F and K/F a finite field extension. If A_K is a free A-module of rank $[K : F]$, then an A-basis for is called an *integral basis*. The corollary applies to any ring of algebraic numbers $\mathbb{Z}_K$. It says that it is a free abelian group of rank $[K : \mathbb{Q}]$, hence has an integral basis. The theorem also applies to the case that $A = F[t]$ with F a field, and $K/F(t)$ a finite separable extension. Then $F[t]_K$ is a free $F[t]$-module of rank $[K : F(t)]$. These are the primary examples of our next study, when we further restrict our domain.

Exercise 73.9. Let K be a quadratic extension of $\mathbb{Q}$. Determine $\mathbb{Z}_K$ and find a $\mathbb{Z}$-basis for $\mathbb{Z}_K$.

74. Dedekind Domains

We know that PIDs are UFDs, but rings of algebraic integers are not always UFDs, e.g., $\mathbb{Z}_{\mathbb{Q}[\sqrt{-5}]}$. We wish to study the appropriate generalization of PIDs that allows us to replace the property of UFDs, which is a property about elements, with an analogous property about ideals. We shall begin using the theory that we have already developed. We need the following observation:

Lemma 74.1. *Let $\varphi : A \to B$ be a ring homomorphism of commutative rings. If $\mathfrak{P}$ is a prime ideal in B, then $\varphi^{-1}(\mathfrak{P})$ is a prime ideal in A. In particular, if φ is the inclusion of rings, then $\mathfrak{P} \cap A$ is a prime ideal in A.*

PROOF. The homomorphism φ induces a monomorphism of rings $\overline{\varphi} : A/\varphi^{-1}(\mathfrak{P}) \to B/\mathfrak{P}$. As $B/\mathfrak{P}$ is a domain so is $A/\varphi^{-1}(\mathfrak{P})$ and the result follows. □

We come to the main subject of this chapter.

Definition 74.2. Let A be a domain, not a field. Then A is called a *Dedekind domain* if A is a Noetherian integrally closed domain in which every nonzero prime ideal is a maximal ideal.

Examples 74.3. The verifications of the following are left as exercises:

1. Let A be a PID that is not a field, then A is a Dedekind domain. In particular, $\mathbb{Z}$ and $F[t]$, with F a field, are Dedekind domains.
2. Let A be a Dedekind domain and S a multiplicative set not containing zero. Then the localization $S^{-1}A$ of A is a Dedekind domain or a field. In particular, if $\mathfrak{p}$ is a maximal ideal in A, then the localization $A_{\mathfrak{p}} = S^{-1}A$ with $S = A \setminus \mathfrak{p}$ is a Dedekind domain.
3. Recall that a *local ring* is a commutative ring with a unique maximal ideal. A Dedekind domain with a unique maximal ideal, so a local integrally closed, Noetherian domain that is not a field, is called a *discrete valuation ring*. If A is a Dedekind domain and $\mathfrak{p}$ a maximal ideal in A, then the localization $A_{\mathfrak{p}}$ of A is a discrete valuation ring, e.g., $\mathbb{Z}_{(p)}$ is a discrete valuation ring where p is a prime in $\mathbb{Z}$ as is $F[t]_{(f)}$ with F a field and f an irreducible polynomial in $F[t]$.

Remark 74.4. If A is a commutative ring, then we define the *(Krull) dimension* of A by

$$\dim A = \max\{n \mid \text{there exists a chain of prime ideals } \mathfrak{p}_0 < \cdots < \mathfrak{p}_n \text{ in } A\},$$

if this a finite number (and to be infinite if not). With this definition, a Dedekind domain is a Noetherian integrally closed domain of dimension one. The geometric analogue of a Dedekind domain turns out to be a smooth affine curve. The points of the curve correspond to the maximal ideals in a Dedekind domain.

Our extension theory allows us to conclude:

Theorem 74.5. *Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. Then A_K is a Dedekind domain.*

PROOF. A_K is integrally closed by definition and Noetherian by Theorem 73.5. So it suffices to show if $\mathfrak{P}$ is a nonzero prime ideal in A_K, then it is a maximal ideal, equivalently, $A_K/\mathfrak{P}$ is a field. By the lemma, $\mathfrak{P} \cap A$ is a prime ideal. If x is a nonzero element in $\mathfrak{P}$, it satisfies an equation

$$(*) \qquad x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0 \text{ for some } a_0, \ldots, a_{n-1} \text{ in } A.$$

As A_K is a domain, we may assume that a_0 is nonzero. Since a_0 lies in $\mathfrak{P} \cap A$, the prime ideal $\mathfrak{p} = \mathfrak{P} \cap A$ is nonzero hence maximal, so $A/\mathfrak{p}$ is a field. But A_K is a finitely generated A-module by Theorem 73.5, so the domain $A_K/\mathfrak{P}$ is a finite dimensional $(A/\mathfrak{p})$-vector space. It follows that $A_K/\mathfrak{P}$ is a field by Corollary 45.18. □

Using equation (*) above, we easily see that the following is true:

Corollary 74.6. *Let A be a domain with quotient field F and K/F a finite field extension. If $\mathfrak{A}$ is a nonzero ideal in A_K, then $\mathfrak{A} \cap A_F$ is nonzero.*

Corollary 74.7. *Let A be a PID with quotient field F and K/F a finite separable extension. Then A_K is a Dedekind domain. Moreover, A_K has an integral basis.*

Remarks 74.8. 1. The corollary shows that every ring of algebraic integers is a Dedekind domain and is a finitely generated free abelian group.

2. The corollary shows that $F[t]_K$ for any field F and finite separable extension K of $F(t)$ is also a Dedekind domain and is a finitely generated free $F[t]$-module. It turns out that the theory about rings of algebraic integers and $F[t]_K$ is very similar when F is a finite field. This case can be viewed as the geometric analogue of the arithmetic case of the ring of algebraic integers. Although one must worry about separability, it is easier in general. A field that is either a finite extension of the rationals or a finite separable extension of $F(t)$ with F a finite field is called a *global field*. Since the expectation that study of the rings $\mathbb{Z}_K$ and $F[t]_K$ in the appropriate global field K are similar, this often leads to the search and study for the appropriate analogue. For example, the solution of the analogue of the Riemann Hypothesis (the Weil Conjecture) for the geometric case by Deligne is one of the major theorems in mathematics proven in the twentieth century.

3. In general, if A is a Dedekind domain with quotient field F and K/F a finite separable field extension, then A_K is not a free A-module.

4. Let A be a domain with quotient field F and K/F a finite field extension. In general, A_K will not be a finitely generated A-module. However, the Krull-Akizuki Theorem 84.17 below says if A is a Noetherian domain of dimension at most one, then B has the same properties for any domain B satisfying $A \subset B \subset K$. In particular, if A is a Dedekind domain, then A_K is also a Dedekind domain. In particular, if $K/F(t)$ is a finite extension, then $F[t]_K$ is a Dedekind domain. Therefore, any finite extension of $F(t)$ is also called a global field.
5. Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. If $\mathfrak{p}$ is a maximal ideal in A, then $(A_\mathfrak{p})_K = (A_K)_\mathfrak{p}$, the localization at $A \setminus \mathfrak{p}$ by Exercise 72.15(5). In particular, $(A_\mathfrak{p})_K$ is a Dedekind domain. Although $A_\mathfrak{p}$ is a discrete valuation ring, $(A_\mathfrak{p})_K$ may not be. However, by Exercise 74.14 (6), it does have only finitely many maximal ideals and both it and $A_\mathfrak{p}$ are PIDs.

We next prove the defining property of Dedekind domains that also achieves our desired generalization of a UFD.

Theorem 74.9. *Let A be a Dedekind domain. Then every nonzero nonunit ideal in A is a product of nonzero prime ideals, unique up to order.*

PROOF. Let $\mathfrak{A} < A$ be a nonzero ideal and F the quotient field of A. We prove the result in a number of steps.

The first step uses the fact that A is Noetherian.

Step 1. There exist nonzero prime ideals (hence maximal ideals) $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ in A, not necessarily distinct, satisfying $\mathfrak{p}_1 \cdots \mathfrak{p}_r \subset \mathfrak{A}$ (Cf. Exercise 27.22 (11)):

If this is not true, by the Maximal Principle, we may assume that $\mathfrak{A}$ is a maximal counterexample, i.e., $\mathfrak{A}$ does not satisfy Step 1, but any nonunit ideal properly containing $\mathfrak{A}$ does. Of course, $\mathfrak{A}$ cannot be a prime ideal, so there exist ideals $\mathfrak{B}_i$, $i = 1, 2$ satisfying $\mathfrak{A} < \mathfrak{B}_i < A$ with $\mathfrak{B}_1\mathfrak{B}_2 \subset \mathfrak{A}$ by Lemma 23.20. As each $\mathfrak{B}_i$ contains a product of prime ideals by the maximality condition, so does $\mathfrak{A}$, a contradiction.

The second step uses the fact that nonzero prime ideals in the Noetherian domain A are maximal.

Step 2. Let $\mathfrak{p}$ be a nonzero prime ideal in A. Set

$$\mathfrak{p}^{-1} := \{x \in F \mid x\mathfrak{p} \subset A\},$$

an A-submodule of F. Then $A < \mathfrak{p}^{-1}$:

Clearly, $A \subset \mathfrak{p}^{-1} \subset F$ and $\mathfrak{p}^{-1}$ is an A-module. Let a be a nonzero element of $\mathfrak{p}$. If $\mathfrak{p} = (a)$, then a^{-1} lies in $\mathfrak{p}^{-1} \setminus A$, so we may assume not. By Step 1, there exist nonzero prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ satisfying $\mathfrak{p}_1 \cdots \mathfrak{p}_r \subset (a) \subset \mathfrak{p} < A$. Moreover, we may assume that we have chosen these prime ideals so that $r > 1$ is minimal. Since $\mathfrak{p}$ is a prime ideal, $\mathfrak{p}_i \subset \mathfrak{p}$ for some i, and as $\mathfrak{p}_i$ is a maximal ideal, we, in fact, must have $\mathfrak{p}_i = \mathfrak{p}$. Changing notation if necessary, we may assume that $\mathfrak{p} = \mathfrak{p}_1$. By the minimality of r, there exists an element b in $\mathfrak{p}_2 \cdots \mathfrak{p}_r \setminus (a)$. We then have

$$b\mathfrak{p} = b\mathfrak{p}_1 \subset \mathfrak{p}_1 \cdots \mathfrak{p}_r \subset (a).$$

In particular, $a^{-1}b\mathfrak{p}$ lies in A, i.e., $a^{-1}b$ lies in $\mathfrak{p}^{-1}$. By the choice of b, we have $a^{-1}b \notin A$.

The third step uses the fact that the dimension one Noetherian domain A is integrally closed.

Step 3. Let $\mathfrak{p}$ be a nonzero prime ideal in A. Then for every nonzero ideal $\mathfrak{B}$ in A, we have $\mathfrak{B} < \mathfrak{B}\mathfrak{p}^{-1}$. In particular, $\mathfrak{p}\mathfrak{p}^{-1} = A$:

Suppose that $\mathfrak{B} = \mathfrak{B}\mathfrak{p}^{-1}$. By Step 2, there exists an element a in $\mathfrak{p}^{-1} \setminus A$ satisfying $a\mathfrak{B} \subset \mathfrak{B}$. Hence $\lambda_a : \mathfrak{B} \to \mathfrak{B}$ is an A-homomorphism. Therefore, $\mathfrak{B}$ is an $A[a]$-module. As $A[a]$ is a nonzero ring, $\operatorname{ann}_{A[a]} \mathfrak{B} = 0$ (as $\mathfrak{B} \subset F$ is $A[a]$-torsion free). Moreover, $\mathfrak{B}$ is finitely generated, since A is a Noetherian ring. Therefore, a is integral over A by Claim 72.4. As A is integrally closed, we conclude that a lies in A, a contradiction. Finally, if $\mathfrak{B} = \mathfrak{p}$, then $\mathfrak{p} < \mathfrak{p}\mathfrak{p}^{-1}$ forces $\mathfrak{p}\mathfrak{p}^{-1}$ to be A, as $\mathfrak{p}$ is a maximal ideal.

Step 4. Finish:

We first show that $\mathfrak{A}$ is a product of nonzero prime ideals. If not, as in the proof of Step 1, we may assume that $\mathfrak{A}$ is a maximal counterexample. Since $\mathfrak{A}$ cannot be a prime ideal, there exists a prime ideal $\mathfrak{p}$ in A with $\mathfrak{A} < \mathfrak{p}$. In particular, $\mathfrak{A}\mathfrak{p}^{-1} \subset \mathfrak{p}\mathfrak{p}^{-1} = A$. By Step 3 and the maximality condition, there exist nonzero prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ such that $\mathfrak{A}\mathfrak{p}^{-1} = \mathfrak{p}_1 \cdots \mathfrak{p}_r$. Applying Step 4 again shows that $\mathfrak{A} = \mathfrak{p}\mathfrak{p}_1 \cdots \mathfrak{p}_r$, a contradiction.

Next we show uniqueness. If we have

$$\mathfrak{p}_1 \cdots \mathfrak{p}_r = \mathfrak{A} = \mathfrak{P}_1 \cdots \mathfrak{P}_s$$

with all the $\mathfrak{p}_i$ and $\mathfrak{P}_j$ maximal ideals, then as in Step 1, we see that $\mathfrak{p}_i = \mathfrak{P}_j$ for some i and j. Changing notation, we may assume that $1 = i = j$. Multiplying the above equation by $\mathfrak{p}^{-1}$ leads to our conclusion as $\mathfrak{p}_1\mathfrak{p}_1^{-1} = A$ by Step 3. $\square$

The converse of the theorem is true, i.e., a domain in which every nonzero nonunit ideal is a product of prime ideals, unique up to order, is a Dedekind domain. This is left as an exercise.

Let A be a Dedekind domain and $\mathfrak{A}$ a nonzero nonunit ideal in A. By the theorem, there exist unique nonzero prime ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_r$ in A and unique positive integers $e_1, \dots, e_r$ satisfying

$$(*) \qquad \mathfrak{A} = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$$

(unique up to order). We call (*) a *factorization* of $\mathfrak{A}$. For each nonzero prime ideal $\mathfrak{p}$ and we define

$$v_{\mathfrak{p}}(\mathfrak{A}) = \begin{cases} e_i & \text{if } \mathfrak{p} = \mathfrak{p}_i, \text{ some } i = 1, \dots, r \\ 0 & \text{otherwise,} \end{cases}$$

then (*) can be written

$$\mathfrak{A} = \prod_{\mathfrak{p}} \mathfrak{p}^{v_{\mathfrak{p}}(\mathfrak{A})}.$$

Let $\mathfrak{A}$ and $\mathfrak{B}$ be two nonzero ideals in a Dedekind domain A. Then, by unique factorization of ideals, $\mathfrak{A} \subset \mathfrak{B}$ if and only if $v_{\mathfrak{p}}(\mathfrak{A}) \geq v_{\mathfrak{p}}(\mathfrak{B})$ for all maximal ideals $\mathfrak{p}$ in A. Mimicking the case for elements, we say $\mathfrak{B}$ *divides* $\mathfrak{A}$ and write $\mathfrak{B} \mid \mathfrak{A}$. In particular, $\mathfrak{p}^n \mid \mathfrak{A}$ if and only if $n \leq v_{\mathfrak{p}}(\mathfrak{A})$. We define the *greatest common divisor* of $\mathfrak{A}$ and $\mathfrak{B}$ in A to be the largest ideal $\mathfrak{D}$ of A containing both $\mathfrak{A}$ and $\mathfrak{B}$. As $\mathfrak{A}$ and $\mathfrak{B}$ lie in $\mathfrak{A} + \mathfrak{B}$, it follows that $\mathfrak{A} + \mathfrak{B} = \mathfrak{D} = \prod_{\mathfrak{p}} \mathfrak{p}^{\min\left(v_{\mathfrak{p}}(\mathfrak{A}), v_{\mathfrak{p}}(\mathfrak{B})\right)}$. We say $\mathfrak{A}$ and $\mathfrak{B}$ are *relatively prime* if $\mathfrak{D} = A$, i.e., $A = \mathfrak{A} + \mathfrak{B}$.

We have shown that if A is a Dedekind domain, then any nonzero prime ideal $\mathfrak{p}$ has an 'inverse', viz., $\mathfrak{p}^{-1}$ as $\mathfrak{p}\mathfrak{p}^{-1} = A$. We next generalize this to a notion of inverses of ideals in Dedekind domains.

Definition 74.10. Let A be a domain with quotient field F. A nonzero A-submodule $\mathfrak{A}$ of F is called a *fractional ideal* of A if there exists a nonzero element x in A satisfying $x\mathfrak{A} \subset A$. If $\mathfrak{A}$ is a fractional ideal, we define

$$\mathfrak{A}^{-1} := \{x \in F \mid x\mathfrak{A} \subset A\}.$$

As this is clearly a submodule of F and $a\mathfrak{A}^{-1} \subset A$ for any a in $\mathfrak{A}$, the module $\mathfrak{A}^{-1}$ is also a fractional ideal. Set

$$I_A := \{\mathfrak{A} \mid \mathfrak{A} \text{ a fractional ideal of } A\}.$$

We say that $\mathfrak{A}$ in I_A is *invertible* if $\mathfrak{A}\mathfrak{A}^{-1} = A$.

As every maximal ideal in a Dedekind domain A is invertible and there exist unique factorization of nonzero nonunit ideals in A, we

expect that every fractional ideal in A is invertible. This is in fact true, as we shall now show.

Corollary 74.11. *Let A be a Dedekind domain. Then every fractional ideal of a Dedekind domain A is invertible. Moreover, I_A is a free abelian group on basis the set of maximal ideals in A.*

PROOF. Let $\mathfrak{A}$ be a fractional ideal, then there exists a nonzero element a in A satisfying $a\mathfrak{A} \subset A$. By the theorem, we have factorizations say $a\mathfrak{A} = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$ and $(a) = \mathfrak{p}_1^{f_1} \cdots \mathfrak{p}_s^{f_s}$. As nonzero prime ideals are invertible, we have $\mathfrak{A} = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r} \mathfrak{p}_1^{-f_1} \cdots \mathfrak{p}_s^{-f_s}$. Cancelling appropriate $\mathfrak{P}_i$'s and $\mathfrak{p}_j$'s gives a factorization that is clearly unique. It follows that I_A is a free abelian group on the set of maximal ideals with unity A. □

Of course, if A is a Dedekind domain and $\mathfrak{A}$ a fractional ideal, we have $\mathfrak{A} = \prod_{\mathfrak{p}} \mathfrak{p}^{v_{\mathfrak{p}}(\mathfrak{A})}$, where $v_{\mathfrak{p}}(\mathfrak{p}^n) = n$ for any integer n.

Definition 74.12. Let A be a Dedekind domain. Set

$$P_A := \{xA \mid x \in F^\times\}$$

a subgroup of I_A called the *group of principal fractional ideals* of A. The quotient group

$$Cl_A = I_A/P_A \text{ is called the } \textit{class group of } A.$$

It measures the obstruction of A from being a PID. It also measures when A is a UFD as we shall now show.

Corollary 74.13. *Let A be a Dedekind domain. Then A is a UFD if and only if A is a PID if and only if Cl_A is trivial.*

PROOF. We know that a domain is a UFD if and only if every nonzero prime ideal contains a prime element by Kaplansky's Theorem 28.1. As A is a Dedekind domain, this is equivalent to every maximal ideal being principal which, by Theorem 74.9, is equivalent to every ideal being principal. □

In general, a Dedekind domain may not be a PID, e.g., $\mathbb{Z}[\sqrt{-5}]$. Also the class group Cl_A may be infinite. In fact L. Claburn has shown that any abelian group can be realized as the class group of some Dedekind domain. However, an important theorem in number theory is that $Cl_{\mathbb{Z}_K}$ is a finite group for any ring of algebraic integers $\mathbb{Z}_K$. This is usually proven by using what is called Minkowski Theory.

Exercises 74.14. 1. Define the least common multiple of two nonzero nonunit ideals in a Dedekind domain and show that it is equal to the intersection of these ideals.

2. Show that $\mathbb{Z}[\sqrt{-5}]$ is a Dedekind domain but not a PID.
3. Let A be a discrete valuation ring, i.e., a Dedekind domain with a unique maximal ideal. Show that A is a PID.
4. Let A be a Dedekind domain and S a multiple set not containing zero. Show that the localization $S^{-1}A$ is either a field or a Dedekind domain. In particular, if $\mathfrak{p}$ is a maximal ideal in A then the localization of A at $\mathfrak{p}$, i.e., $A_{\mathfrak{p}} = S^{-1}A$ where $S = A \setminus \mathfrak{p}$, is a discrete valuation ring.
5. Show that every Dedekind domain with finitely many prime ideals is a PID.
6. Let A be a Dedekind domain with quotient field F and having finitely many prime ideals. Show if K/F is a finite separable field extension, then A_K is a Dedekind domain with finitely many prime ideals.
7. Let R be an arbitrary domain: Show the following:
   (a) Any invertible fractional ideal in R is finitely generated.
   (b) A product of fractional ideals in R is invertible if and only if each of the fractional ideals is invertible.
   (c) If an ideal in R is a product of invertible prime ideals, then the prime ideals are unique up to order.
8. Let R be a domain in which every nonzero ideal is a product of prime ideals. Show that every nonzero prime ideal is invertible and maximal.
9. Show that a domain, not a field, is a Dedekind domain, if and only if every nonzero ideal is a product of prime ideals.
10. Let A be a Dedekind domain and $\mathfrak{p}$ a prime ideal in A. Show the following:
    (i) $\mathfrak{p}^m = \mathfrak{p}^n$ if and only if $m = n$.
    (ii) If $\pi \in \mathfrak{p} \setminus \mathfrak{p}^2$, then $\mathfrak{p}^m = (\pi)^m + \mathfrak{p}^n$ for any positive integers $n \geq m$.
    (iii) The only proper ideals in $A/\mathfrak{p}^n$ are $\mathfrak{p}/\mathfrak{p}^n, \ldots \mathfrak{p}^{n-1}/\mathfrak{p}^n$
11. Let A be a Dedekind domain and $\mathfrak{A}$ a nonzero ideal in A. Using the previous exercise show the following:
    (i) Every ideal in $A/\mathfrak{A}$ is principal.
    (ii) Every ideal in A can be generated by two elements.
12. Let A be a Dedekind domain. Prove the following:
    (i) Let $\mathfrak{A}$ be a fractional ideal of A and $\mathfrak{B} \subset A$ an ideal. Then there exists an element $a \in \mathfrak{A}$ such that $\mathfrak{A}^{-1}a + \mathfrak{B} = A$.

(*ii*) Let $\mathfrak{A}_1$ and $\mathfrak{A}_2$ be two fractional ideals of A. Then the A-module $\mathfrak{A}_1 \coprod \mathfrak{A}_2$ is isomorphic to $A \coprod \mathfrak{A}_1\mathfrak{A}_2$.

(*iii*) Let $\mathfrak{A}_1$ be a fractional ideal of A. Then there exists an A-isomorphism $\mathfrak{A} \coprod \mathfrak{A}^{-1} \cong A^2$. In particular, every fractional ideal over A is A-projective (cf. Exercises 36.12(9), (10), (11)).

(*iv*) Every finitely generated torsion-free A-module is isomorphic to a finite direct sum of ideals in A. (Cf. Exercise 41.24(2).)

(*v*) Let M be a finitely generated torsion-free A-module. Then $M \cong A^n \coprod \mathfrak{A}$ for some n and some ideal $\mathfrak{A}$ in A. (There is a uniqueness statement. Can you guess what it is?)

13. Let A be a local noetherian domain of dimension one with $\mathfrak{m}$ its maximal ideal. Let $k := A/\mathfrak{m}$, called the *residue class field* of A. Show the following are equivalent:
    (*i*) A is a discrete valuation ring.
    (*ii*) A is integrally closed.
    (*iii*) $\mathfrak{m}$ is a principal ideal.
    (*iv*) $\dim_k(\mathfrak{m}/\mathfrak{m}^2) = 1$.
    (*v*) Every non-zero ideal is a power of $\mathfrak{m}$.
    (*vi*) There exists $x \in A$ such that every non-zero ideal is of the form (x^n), $n \geq 0$.
    (*vii*) A is a *valuation ring*, i.e., if a nonzero x is an element in the quotient field of A, then either x is an element of A or its multiplicative inverse is.
    (*viii*) A is a Dedekind domain.

14. Let A be a noetherian domain. Then the following are equivalent:
    (*i*) A is a Dedekind domain.
    (*ii*) $A_{\mathfrak{m}}$ is a discrete valuation ring for all maximal ideals $\mathfrak{m}$ in A.
    (*iii*) Every non-zero fractional ideal in A is invertible.
    (*iv*) Every integral ideal in A factors uniquely into prime ideals.

15. Let A be a Dedekind domain and $\mathfrak{A}$ an nonzero ideal in A with factorization $\mathfrak{A} = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$ into prime ideals in A. Prove all of the following:
    (*i*) There exists a ring isomorphism $A/\mathfrak{A} \cong A/\mathfrak{p}_1^{e_1} \times \cdots A/\mathfrak{p}_r^{e_r}$.
    (*ii*) If M is an A-module, then there exist A-modules M_i for $i = 1, \ldots, r$ and an A-module isomorphism $M \cong \coprod_{i=1}^r M_i$ with $A/\mathfrak{p}_i^{k_j} M_i = 0$ for every $j \neq i$ and each $i = 1, \ldots, r$. In particular, each M_i is an $A/\mathfrak{p}_i^{e_i}$-module.

16. Let A be a Dedekind domain, $\mathfrak{p}$ a nonzero prime ideal in A, and M a finitely generated torsion A-module with *annihilator*, $\operatorname{ann}_A M := \{x \in A \mid xA = 0\} = \bigcap_{x \in A} \operatorname{ann}_A x$. Show all of the following:

(*i*) Let $A_{\mathfrak{p}}$ be the localization of A at $\mathfrak{p}$, a discrete valuation ring hence a PID, then $A/\mathfrak{p}^e \cong A_{\mathfrak{p}}/(\mathfrak{p}^e A_{\mathfrak{p}})$ as rings.
(*ii*) If M satisfies $\mathfrak{p}^e = \mathrm{ann}_A(M)$, then M is isomorphic to a coproduct of A-modules $A/\mathfrak{p}^k$ for various $1 \leq k \leq e$.
(*iii*) If $\mathrm{ann}_A(M) = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$, then M is isomorphic to a coproduct of A-modules $A/\mathfrak{p}_i^{k_{ij}}$ for various $1 \leq k_{ij} \leq e_i$ for $i = 1, \ldots, r$. (There is a uniqueness statement. What is it?)

17. Let A be a Dedekind domain and M a finitely generated A-module. Then $M = M_{tf} \oplus M_t$ with the submodules M_{tf} torsion-free and M_t torsion, respectively. In particular, by (12) and (16) of these Exercises (74.14), $M \cong R^n \coprod \mathfrak{A} \coprod \coprod_{i,j} A/\mathfrak{p}_i^{k_j}$, for some ideal $\mathfrak{A}$ in A and for various $A/\mathfrak{p}_i^{k_{ij}}$, $1 \leq k_{ij} \leq e_i$, where $\mathrm{ann}_A(M) = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$. (There is also a uniqueness statement. What is it?)

75. Extension of Dedekind Domains

Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. We have shown that A_K is also a Dedekind domain. We wish to investigate how nonzero prime ideals in A factor into a product of primes ideals in A_K. We begin by showing any proper ideal in A remains a proper ideal in A_K. This follows from the following (which is a special case of the Cohen-Seidenberg Theorem 82.14(2) below).

Claim 75.1. Let $\mathfrak{p}$ be a nonzero prime ideal in the Dedekind domain A above. Then $\mathfrak{p}A_L < A_L$ for any field extension L/F.

We know that $\mathfrak{p}^2 < \mathfrak{p}$. Let $\pi \in \mathfrak{p} \setminus \mathfrak{p}^2$. It follows that $\pi A = \mathfrak{p}\mathfrak{A}$ for some ideal $\mathfrak{A}$ of A with $\mathfrak{p} \not| \mathfrak{A}$, hence $\mathfrak{p} + \mathfrak{A} = A$. Write $1 = p + a$ with $p \in \mathfrak{p}$ and $a \in \mathfrak{A}$. It follows that $a \notin \mathfrak{p}$ and $a\mathfrak{p} \subset \mathfrak{p}\mathfrak{A} = \pi A$. Suppose that $\mathfrak{p}A_L = A_L$. Then $aA_L = a\mathfrak{p}A_L \subset \pi A_L$, hence $a = \pi x$, for some $x \in A_L \cap F = A$. This implies that a lies in $\mathfrak{p}$, a contradiction. This proves the claim.

In particular, if $\mathfrak{p}$ is a nonzero prime ideal in A, then we have a factorization

$$(*) \qquad \mathfrak{p}A_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r} \text{ in } A_K$$

called the *splitting behavior of* $\mathfrak{p}$ in A_K. We say a prime ideal $\mathfrak{P}$ in A_K *lies over* $\mathfrak{p}$, if $\mathfrak{P} \cap A = \mathfrak{p}$. If $\mathfrak{P}$ is such a prime ideal then $\mathfrak{p}A_K \subset (\mathfrak{P} \cap A)A_K \subset \mathfrak{P} < A_K$, so $\mathfrak{P} = \mathfrak{P}_i$ for some i, as $\mathfrak{P}$ is a maximal ideal in A_K. Conversely if $\mathfrak{P} = \mathfrak{P}_i$, then $\mathfrak{p}A_K \cap A \subset \mathfrak{P} \cap A < A$, so $\mathfrak{P} \cap A = \mathfrak{p}$, as $\mathfrak{p}$ is a maximal ideal. Thus

$$\{\mathfrak{P}_1, \ldots, \mathfrak{P}_r\} = \{\mathfrak{P} \mid \text{with } \mathfrak{P} \text{ a prime ideal in } A_K \text{ satisfying } \mathfrak{P} \mid \mathfrak{p}A_K\},$$

the set of prime ideals in A_K lying over $\mathfrak{p}$ in A_K. We shall write $\mathfrak{P} \mid \mathfrak{p}$ for $\mathfrak{P} \mid \mathfrak{p}A_K$. Let $\mathfrak{P}$ be a nonzero prime ideal in A_K. Then the composition of the inclusion $A \subset A_K$ and the canonical epimorphism $^- : A_K \to A_K/\mathfrak{P}$ shows that the field $A_K/\mathfrak{P}$ is a field extension of $A/\mathfrak{P} \cap A$. We define *ramification index* $e(\mathfrak{P}/\mathfrak{p})$ *of* $\mathfrak{P}$ *over* $\mathfrak{p}$ and if $\mathfrak{P}$ lies over $\mathfrak{p}$, the *inertia index* $f(\mathfrak{P}/\mathfrak{p})$ *of* $\mathfrak{P}$ *over* $\mathfrak{p}$ by

$$e(\mathfrak{P}/\mathfrak{p}) = v_{\mathfrak{P}}(\mathfrak{p}A_K)$$
$$f(\mathfrak{P}/\mathfrak{p}) = [A_K/\mathfrak{P} : A/\mathfrak{p}].$$

respectively. If $e(\mathfrak{P}/\mathfrak{p}) > 1$ or if $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is not separable, we say that $\mathfrak{p}$ *ramifies* in A_K. That the inertia index is a finite number follows from the following important result:

Theorem 75.2. *Let A be a Dedekind domain with quotient field F, K/F a finite separable field extension. If $\mathfrak{p}$ is a nonzero prime ideal in A, then*

$$[K : F] = \sum_{\mathfrak{P}|\mathfrak{p}} e(\mathfrak{P}/\mathfrak{p}) f(\mathfrak{P}/\mathfrak{p}).$$

PROOF. Let $\mathfrak{p}A_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$ be a factorization of $\mathfrak{p}$ in A_K. So $e(\mathfrak{P}_i/\mathfrak{p}) = e_i$ and $f(\mathfrak{P}_i/\mathfrak{p}) = f_i$ for each i. Let $n = [K : F]$ and $^- : A_K \to A_K/\mathfrak{p}A_K$ be the canonical epimorphism, so $\overline{A} = A/\mathfrak{p}$. By the Chinese Remainder Theorem

$$(*) \qquad A_K/\mathfrak{p}A_K = A_K/\mathfrak{P}_1^{e_1} \times \cdots \times A_K/\mathfrak{P}_r^{e_r}.$$

The ring $\overline{A_K} = A_K/\mathfrak{p}A_K$ is a finite dimensional $\overline{A}$-vector space as A_K is a finitely generated A-module, hence so are the $A_K/\mathfrak{P}_i^{e_i}$'s. (Of course, it follows from this that the $f(\mathfrak{P}_i/\mathfrak{p})$ are all finite.)

Step 1. $n = [K : F] = \dim_{\overline{A}} \overline{A_K}$:

Choose $\mathcal{B} = \{x_1, \ldots, x_m\}$ in A_K, so that $\overline{\mathcal{B}} = \{\overline{x}_1, \ldots, \overline{x}_m\}$ is an $\overline{A}$-basis for $\overline{A_K}$.

We shall show that $\mathcal{B}$ is an F-basis for K establishing the claim. We first show $\mathcal{B}$ is linearly independent. Suppose not, then clearing denominators if necessary, we see that we have an equation

$$a_1x_1 + \cdots + a_mx_m = 0 \text{ in } A_K$$

for some $a_1, \ldots, a_m$ in A, not all zero. Let $\mathfrak{A}$ be the nonzero ideal $(a_1, \ldots, a_m)$. As $\mathfrak{A}\mathfrak{A}^{-1} = A$, we can choose a nonzero element a in $\mathfrak{A}^{-1}$ such that not every aa_i lies in $\mathfrak{p}$. As $aa_i \in A$ for every i, we have

$$\overline{aa}_1\overline{x}_1 + \cdots + \overline{aa}_m\overline{x}_m = 0 \text{ in } \overline{A_K}$$

contradicting $\overline{\mathcal{B}}$ is $\overline{A}$-linearly independent. Thus $\mathcal{B}$ is linearly independent.

We next show $\mathcal{B}$ spans. Let $B = Ax_1 + \cdots + Ax_m$, a submodule of A_K. As $\overline{\mathcal{B}}$ is a basis for $\overline{A_K}$, if $b \in B$, there exists an $x \in A_K$ satisfying $x - b$ lies in $\mathfrak{p}A_K$, i.e., $A_K = B + \mathfrak{p}A_K$. Let $M = A_K/B$, a finitely generated A-module. Then $M/\mathfrak{p}M = 0$, since each x_i maps to zero in it so a zero dimensional $\overline{A}$-vector space. Consequently, $M = \mathfrak{p}M$. Set $M = Ay_1 + \cdots + Ay_s$. Then we have equations

$$y_i = \sum_{j=1}^{s} p_{ij}y_j \text{ for some } p_{ij} \in \mathfrak{p},\ 1 \le i, j \le s,$$

hence

$$\sum_{j=1}^{s} (\delta_{ij} - p_{ij})y_j = 0 \text{ for } i = 1, \dots s.$$

Let Δ be the determinant of the $s \times s$ matrix $(\delta_{ij} - p_{ij})$. Then $\Delta y_i = 0$ for every i by Cramer's Rule, so $\Delta M = 0$. It follows that $\Delta A_K \subset B$, hence also that $\Delta K \subset Fx_1 + \cdots Fx_m$. But $\Delta \equiv 1 \mod \mathfrak{p}$, so Δ is nonzero. Therefore, we conclude that

$$K = \Delta K = Fx_1 + \cdots + Fx_m.$$

Therefore, $\mathfrak{B}$ is an F-basis for K and $m = n$. This completes Step 1.

Step 2. Finish:

In view of (*), it suffices to show that $\dim_{\overline{A}} A_K/{\mathfrak{P}_i}^{e_i} = e_i f_i$ for each i, $1 \le i \le r$. We have a descending chain of $\overline{A}$-vector spaces

$$A_K/\mathfrak{P}_i^{e_i} \supset \mathfrak{P}_i/\mathfrak{P}_i^{e_i} \supset \cdots \supset \mathfrak{P}_i^{e_i-1}/\mathfrak{P}_i^{e_i} \supset 0,$$

so

$$\dim_{\overline{A}} A_K/\mathfrak{P}_i^{e_i} = \sum_{m=0}^{e_i-1} \dim_{\overline{A}} \mathfrak{P}_i^m/\mathfrak{P}_i^{m+1}.$$

(Why?) As $\mathfrak{P}_i^{m+1} < \mathfrak{P}_i^m$ by unique factorization of ideals, we can choose an element a in $\mathfrak{P}_i^m \backslash \mathfrak{P}_i^{m+1}$ for each m. Then $\mathfrak{P}_i^m = aA_K + \mathfrak{P}_i^{m+1}$, the greatest common divisor of $\mathfrak{P}_i^{m+1}$ and aA_K and $\mathfrak{P}_i^{m+1} = aA_K \cap \mathfrak{P}_i^{m+1}$ (the least common multiple of $\mathfrak{P}_i^{m+1}$ and aA_K, cf. Exercise 74.14(1)). Let $\lambda_a : A_K \to \mathfrak{P}_i^m/\mathfrak{P}_i^{m+1}$ be the homomorphism defined by $x \mapsto ax + \mathfrak{P}_i^{m+1}$. The kernel of λ_a is $\mathfrak{P}_i$ and λ_a is onto, as

$$\operatorname{im} \lambda_a = (aA_K + \mathfrak{P}_i^{m+1})/\mathfrak{P}_i^{m+1} = \mathfrak{P}_i^m/(aA_K \cap \mathfrak{P}_i^{m+1}) = \mathfrak{P}_i^m/\mathfrak{P}_i^{m+1}.$$

Therefore $f_i = \dim_{\overline{A}} A_k/\mathfrak{P}_i = \dim_{\overline{A}} \mathfrak{P}_i^m/\mathfrak{P}_i^{m+1}$ for any $m > 0$ and $\dim_{\overline{A}} A_K/\mathfrak{P}_i^{e_i} = e_i f_i$ as desired. □

Remark 75.3. If $A = \mathbb{Z}$ in the above, Step 1 can be simplified. Indeed let $\{w_1, \dots, w_n\}$ be an integral basis for $\mathbb{Z}_K$ and $\mathfrak{A}$ any nonzero ideal in $\mathbb{Z}_K$. We know that $\mathfrak{A} \cap \mathbb{Z} = (a)$ for some nonzero integer a. Set

$S = \{\sum r_i w_i \mid 0 \leq r_i < a\}$. Then using the division algorithm, one shows that S is a system of representatives for the cosets $\mathbb{Z}_K/\mathfrak{A}$, so has a^n elements.

Let A be a Dedekind domain with quotient field F and K/F a finite separable extension. Let $\mathfrak{p}$ be a nonzero prime ideal in A and $\mathfrak{P}$ a prime ideal in A_K lying over $\mathfrak{p}$. We say that $\mathfrak{P}$ is *unramified* over A if $e(\mathfrak{P}/\mathfrak{p}) = 1$ and $A_k/\mathfrak{P}$ is a separable extension of $A/\mathfrak{p}$ and *ramified* otherwise (and *totally ramified* if in addition $f(\mathfrak{P}/\mathfrak{p}) = 1$). We say $\mathfrak{p}$ is *unramified* in A_K if every prime ideal in A_K lying over $\mathfrak{p}$ is unramified and the extension K/F (or A_K/A) is called *unramified* if every prime ideal in A is unramified in A_K. We say a nonzero prime ideal $\mathfrak{p}$ in A *splits completely* in A_K if $e(\mathfrak{P}/\mathfrak{p}) = 1 = f(\mathfrak{P}/\mathfrak{p})$ for every prime ideal $\mathfrak{P}$ lying over $\mathfrak{p}$. If K is a global field, with $A = \mathbb{Z}$ or $F[t]$, we do not have to worry about the separability condition of $A_K/\mathfrak{P}$ over $A/\mathfrak{p}$ as either $A_K/\mathfrak{P}$ is of characteristic zero or a finite field. A deep result in number theory states that there exist infinitely many nonzero prime ideals in A that split completely in A_K when K is a global field.

The Kummer-Dedekind Theorem gives a method for computing the splitting behavior of prime ideals for all but finitely many prime ideals in a given separable extension. We first define the set of prime ideals whose splitting behavior this theorem omits. Let A be a Dedekind domain with quotient field F and K/F a finite separable extension. By the Primitive Element Theorem 54.9, $K = F(\alpha)$ for some α in K. We may assume that α lies in A_K. So $A[\alpha]$ is a subring of A_K. Define the *conductor* of $A[\alpha]$ in A_K to be the largest ideal $\mathfrak{f}$ of A_K that lies in $A[\alpha]$, i.e.,

$$\mathfrak{f} = \{x \in A_K \mid xA_K \subset A[\alpha]\}.$$

We know that A_K contains the F-basis $\{1, \alpha_1, \ldots, \alpha^{n-1}\}$ for K, so it follows by Claim 73.6 that $\mathfrak{f} \neq 0$. Note that $\mathfrak{f} \subset A[\alpha]$ as 1 lies in A_K, so the conductor $\mathfrak{f}$ is an ideal of both A_K and $A[\alpha]$.

Kummer-Dedekind's result is the following:

Theorem 75.4. (Kummer-Dedekind) *Let A be a Dedekind domain with quotient field F and $\mathfrak{p}$ a nonzero prime ideal in A. Suppose that K/F is a finite separable field extension with $K = F(\alpha)$, $\alpha \in A_K$, and $\mathfrak{p}$ relatively prime to the conductor $\mathfrak{f}$ of $A[\alpha]$ in A_K. Let $^- : A[t] \to (A/\mathfrak{p})[t]$ be the canonical epimorphism and let $p_1, \ldots, p_r$ be distinct monic polynomials in $A[t]$ such that*

$$\overline{m_F(\alpha)} = \overline{p_1}^{e_1} \cdots \overline{p_r}^{e_r}$$

is a factorization of the image of the minimal polynomial of α *in* $(A/\mathfrak{p})[t]$ *into irreducibles. Set*

$$\mathfrak{P}_i = \mathfrak{p}A_K + p_i(\alpha)A_K \text{ for each } i = 1, \ldots, r.$$

Then $\mathfrak{P}_1, \ldots, \mathfrak{P}_r$ *are all of the distinct primes in* A_K *lying over* $\mathfrak{p}$, $f(\mathfrak{P}_i/\mathfrak{p}) = f_i$, *and* $e_i = e(\mathfrak{P}_i/\mathfrak{p})$ *for* $i = 1, \ldots, r$. *In particular,*

$$\mathfrak{p}A_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$$

is a factorization of $\mathfrak{p}$ *in* A_K.

PROOF. **Step 1**. $A_K/\mathfrak{p}A_K \cong A[\alpha]/\mathfrak{p}A[\alpha]$:

Since $\mathfrak{p}A_K$ and $\mathfrak{f}$ are relatively prime, $A_K = \mathfrak{p}A_K + \mathfrak{f}$. But $\mathfrak{f} \subset A[\alpha]$, so we have $A_K = \mathfrak{p}A_K + A[\alpha]$ and the natural map $A[\alpha] \to A_K/\mathfrak{p}A_K$ is an epimorphism with kernel $\mathfrak{p}A_K \cap A[\alpha]$. As $\mathfrak{f} \cap A$ and $\mathfrak{p}$ must also be relatively prime, we have $\mathfrak{p}A[\alpha] + \mathfrak{f} = A[\alpha]$, so

$$\mathfrak{p}A_K \cap A[\alpha] = (\mathfrak{p}A[\alpha] + \mathfrak{f})(\mathfrak{p}A_K \cap A[\alpha]) \subset \mathfrak{p}A[\alpha].$$

It follows that $\mathfrak{p}A_K \cap A[\alpha] = \mathfrak{p}A[\alpha]$ establishing Step 1.

Step 2. $\overline{A}[t]/(\overline{m_F(\alpha)}) \cong A[\alpha]/\mathfrak{p}A[\alpha]$:

The kernel of the natural epimorphism $A[t] \to \overline{A}[t]/(\overline{m_F(\alpha)})$ is generated by $\mathfrak{p}$ and $(m_F(\alpha))$, Therefore, the evaluation map $A[t] \to A[\alpha]$ induces the desired isomorphism as $A[t]/\mathfrak{p}A[t] \cong (A/\mathfrak{p})[t]$.

Step 3. Finish:

Let $B = \overline{A}[t]/(\overline{m_F(\alpha)})$ and $\widetilde{\ } : \overline{A}[t] \to B$ be the canonical epimorphism. By the Chinese Remainder Theorem,

$$B \cong \overline{A}[t]/(\overline{p_1}^{e_1}) \bigtimes \cdots \bigtimes \overline{A}[t]/(\overline{p_r}^{e_r}).$$

It follows that the nonzero prime ideals in B are the distinct principal ideals $(\widetilde{p_i})$ generated by the $p_i \mod \overline{m_F(\alpha)}$, i.e., the (p_i). We have $\deg p_i = \deg \overline{p_i} = [B/(\widetilde{p_i}) : \overline{A}]$ for $i = 1, \ldots, r$. Further,

$$0 = \widetilde{\left(\overline{m_F(\alpha)}\right)} = \bigcap_{i=1}^{r} (\widetilde{p_i}^{e_i}) \text{ in } B.$$

By Steps 1 and 2, we know that $B \cong A_K/\mathfrak{p}A_K$, with the isomorphism induced by evaluation at α, so by the Correspondence Principle, the prime ideals in B correspond to the prime ideals in $A_K/\mathfrak{p}A_K$. Let $\mathfrak{Q}_i$ in $A_K/\mathfrak{p}A_K$ be the prime ideal corresponding to the prime ideal $(\widetilde{p_i})$ in B for $i = 1, \ldots, r$. It follows that each $\mathfrak{Q}_i$ is a principal ideal generated by $\overline{p_i(\alpha)}$ with $\deg \overline{p_i} = [A_K/\mathfrak{Q}_i : A/\mathfrak{p}]$ for each i, and furthermore, $0 = \bigcap_{i=1}^{r} \mathfrak{Q}_i{}^{e_i}$. Let $\mathfrak{P}_i$ be the ideal in A_K that is the preimage of $\mathfrak{Q}_i$ under the natural epimorphism $A_K \to A_K/\mathfrak{p}A_K$. Then $\mathfrak{P}_i$ is a nonzero prime ideal by Lemma 74.1 and satisfies $\mathfrak{P}_i = \mathfrak{p}A_K + p_i(\alpha)A_K$ for

$i = 1, \ldots, r$. It follows that $\mathfrak{P}_1, \ldots, \mathfrak{P}_r$ are all the prime ideals in A_K lying over $\mathfrak{p}$, and $f_i = [A_K/\mathfrak{P}_i : A/\mathfrak{p}] = \deg \overline{p_i} = \deg p_i$ for each i. Moreover, as

$$e_i = |\{\mathfrak{Q}_i^m \mid m \geq 0\}|, \text{ we have } \overline{\mathfrak{P}_i}^{e_i} = \mathfrak{Q}_i^{e_i},$$

so $\mathfrak{P}_i^{e_i}$ is the preimage of $\mathfrak{Q}_i^{e_i}$. Since $\prod_{i=1}^r \mathfrak{P}_i^{e_i} \subset \bigcap_{i=1}^r \mathfrak{P}_i^{e_i} \subset \mathfrak{p}A_K$, we have $v_\mathfrak{p}(\mathfrak{P}_i) \geq e_i$. As $[K : F] = \sum e_i f_i$, we see that $\mathfrak{p}A_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$ is a factorization of $\mathfrak{p}A_K$. □

Exercises 75.5. 1. Complete the proof of Remark 75.3.

2. Let A be a Dedekind domain with quotient field F and $L/K/F$ finite separable extensions. If $\mathfrak{P}$ is a nonzero prime ideal in A_L, show

$$e(\mathfrak{P}/\mathfrak{P} \cap A) = e(\mathfrak{P}/\mathfrak{P} \cap A_K)e(\mathfrak{P} \cap A_K/\mathfrak{P} \cap A)$$
$$f(\mathfrak{P}/\mathfrak{P} \cap A) = f(\mathfrak{P}/\mathfrak{P} \cap A_K)f(\mathfrak{P} \cap A_K/\mathfrak{P} \cap A).$$

3. (Nakayama's Lemma) Let A be a *local ring*, i.e., a nonzero commutative ring with a unique maximal ideal, say $\mathfrak{m}$. Show if M is a finitely generated A-module satisfying $\mathfrak{m}M = M$, then $M = 0$.

4. Let A be a local ring with maximal ideal $\mathfrak{m}$ and $^- : M \to M/\mathfrak{m}M$ the canonical epimorphism. Let $\mathcal{C} = \{x_1, \ldots, x_n\} \subset M$ and $\overline{\mathcal{C}} = \{\overline{x_1}, \ldots, \overline{x_n}\}$. Show that $\mathcal{C}$ generates M if and only if $\overline{\mathcal{C}}$ spans $\overline{M}$ as an $A/\mathfrak{m}$-vector space and is a minimal generating set for M (obvious definition) if and only if $\overline{\mathcal{C}}$ is an $A/\mathfrak{m}$-basis for $\overline{M}$.

5. Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. If α is an element of A_K, then for each maximal ideal not dividing the conductor $\mathfrak{f}$ of $A[\alpha]$ (hence all but finitely many maximal ideals) $(A_\mathfrak{p})_K = A_\mathfrak{p}[\alpha]$, where $A_\mathfrak{p}$ is the localization of A at $A \setminus \mathfrak{p}$.

6. Let A be a Dedekind domain with quotient field F and K/F a finite separable extension. Then a prime ideal $\mathfrak{p}$ in A ramifies in A_K if and only if the localization $A_\mathfrak{p}$ of A ramifies in $(A_\mathfrak{p})_K$.

76. Hilbert Ramification Theory

Theorem 75.2 becomes even nicer when K/F is Galois. We need the following to prove this.

Proposition 76.1. *Let A be a Dedekind domain with quotient field F and $\mathfrak{p}$ a nonzero prime ideal in A. If K/F is a finite Galois extension, then the Galois group $G(K/F)$ acts transitively on the set of prime ideals in A_K lying over $\mathfrak{p}$.*

PROOF. Let $\mathfrak{P}_i \mid \mathfrak{p}$ be prime ideals in A_K for $i = 1, 2$. Suppose that $\mathfrak{P}_2 \neq \sigma(\mathfrak{P}_1)$ for any $\sigma \in G(K/F)$. By the Chinese Remainder Theorem, there exists an element x in A_K satisfying $x \equiv 0 \mod \mathfrak{P}_2$ and $x \equiv 1 \mod \sigma(\mathfrak{P}_1)$ for all $\sigma \in G(K/F)$, i.e., $x \in \mathfrak{P}_2$ but $\sigma^{-1}(x) \notin \mathfrak{P}_1$ for any σ in the group $G(K/F)$. Since

$$N_{K/F}(x) = \prod_{G(K/F)} \sigma(x) \text{ lies in } \mathfrak{P}_2 \cap A_F = \mathfrak{p} = \mathfrak{P}_1 \cap A \subset \mathfrak{P}_1,$$

we have $\sigma(x) \in \mathfrak{P}_1$ for some $\sigma \in G(K/F)$, a contradiction. □

Proposition 76.2. *Let A be a Dedekind domain with F its quotient field and $\mathfrak{p}$ a nonzero prime ideal in A. If K/F is a finite Galois extension, then $e(\mathfrak{P}/\mathfrak{p}) = e(\mathfrak{P}'/\mathfrak{p})$ and $f(\mathfrak{P}/\mathfrak{p}) = f(\mathfrak{P}'/\mathfrak{p})$ for all prime ideals $\mathfrak{P}$ and $\mathfrak{P}'$ lying over $\mathfrak{p}$ in A_K. In particular, if $\mathfrak{P} \mid \mathfrak{p}$ in A_K is a prime ideal and $e = e(\mathfrak{P}/\mathfrak{p})$, $f = f(\mathfrak{P}/\mathfrak{p})$, then $[K : F] = efr$ where r is the number of primes in A_K lying over $\mathfrak{p}$.*

PROOF. By the last proposition, there exists an element σ in $G(K/F)$ satisfying $\mathfrak{P}' = \sigma(\mathfrak{P})$, hence we have an isomorphism $A_K/\mathfrak{P} \to A_K/\sigma(\mathfrak{P})$ given by $x + \mathfrak{P} \mapsto \sigma(x) + \sigma(\mathfrak{P})$. It follows that $f(\mathfrak{P}/\mathfrak{p}) = f(\mathfrak{P}'/\mathfrak{p})$ Further, if $\mathfrak{P}^m \mid \mathfrak{p}A_K$, then $\sigma(\mathfrak{P})^m \mid \mathfrak{p}A_K$, so $e(\mathfrak{P}/\mathfrak{p}) = e(\mathfrak{P}'/\mathfrak{p})$. The result now follows using Theorem 75.2. □

Proposition 76.2 allows us to use Galois Theory in the subject. This is called Hilbert Ramification Theory. We introduce the subject without going into it deeply (leaving some results as exercises).

For the rest of this section, let A be a Dedekind domain with quotient field F and K/F a finite Galois extension. Let $\mathfrak{p}$ be a nonzero prime ideal in A and $\mathfrak{P}$ a prime ideal in A_K lying over $\mathfrak{p}$. As $G(K/F)$ acts on the set of prime ideals in A_K lying over $\mathfrak{p}$, we can look at the isotropy subgroup

$$G_{\mathfrak{P}} = \{\sigma \in G(K/F) \mid \sigma(\mathfrak{P}) = \mathfrak{P}\}$$

of $\mathfrak{P}$. It is called the *decomposition group* of $\mathfrak{P}$ and its fixed field

$$Z_{\mathfrak{P}} = \{x \in K \mid \sigma(x) = x \text{ for all } \sigma \in G_{\mathfrak{P}}\}$$

is called the *decomposition field* of $\mathfrak{P}$. As usual, we have $\sigma G_{\mathfrak{P}} \sigma^{-1} = G_{\sigma(\mathfrak{P})}$. We also have the index $[G : G_{\mathfrak{P}}]$ = number of primes in A_K lying over $\mathfrak{p}$ as $G(K/F)$ acts transitively on the primes in A_K lying over $\mathfrak{p}$. This means that $\mathfrak{p}$ splits completely in A_K if and only if $G_{\mathfrak{P}} = 1$ and $G(K/F) = G_{\mathfrak{P}}$ if and only if $\mathfrak{P}$ is the only prime ideal in A_K lying over $\mathfrak{p}$.

Proposition 76.3. *Let A be a Dedekind domain with quotient field F and K/F a finite Galois extension, $\mathfrak{P}$ a prime ideal in A_K with $\mathfrak{P}$ lying over $\mathfrak{p}$, then*

(1) $\mathfrak{P}$ is the only prime ideal over $\mathfrak{P} \cap Z_{\mathfrak{P}}$.
(2) $e(\mathfrak{P}/\mathfrak{P} \cap Z_{\mathfrak{P}}) = e(\mathfrak{P}/\mathfrak{p})$ and $f(\mathfrak{P}/\mathfrak{P} \cap Z_{\mathfrak{P}}) = f(\mathfrak{P}/\mathfrak{p})$.
(3) $e(\mathfrak{P} \cap Z_{\mathfrak{P}}/\mathfrak{p}) = 1$ and $f(\mathfrak{P} \cap Z_{\mathfrak{P}}/\mathfrak{p}) = 1$.

PROOF. By Proposition 76.2, we have $[K : Z_{\mathfrak{P}}] = e'f'r'$ where e' is the ramification index of each prime ideal in A_K lying over $\mathfrak{P} \cap Z_{\mathfrak{P}}$, f' its inertia index, and r' is the number of prime ideals in A_K lying over $\mathfrak{P} \cap Z_{\mathfrak{P}}$. Similarly, we have $[Z_{\mathfrak{P}} : F] = e''f''r''$ where e'' is the ramification index of each prime ideal in $Z_{\mathfrak{P}}$ over $\mathfrak{p}$, f'' its inertia index, and r'' is the number of prime ideals in $A_{Z_{\mathfrak{P}}}$ lying over $\mathfrak{p}$, and $[K : F] = efr$ where e is the ramification index of each prime ideal in A_L lying over $\mathfrak{p}$, f is its inertia index, and $r = [G(K/F) : G_{\mathfrak{P}}]$ is the number of prime ideals in A_L lying over $\mathfrak{p}$. By the transitivity of ramification and inertia indices (cf. Exercise 75.5(2), we have $e = e'e''$, $f'f'' = f$, hence $r = r'r''$.

(1): By definition, $G_{\mathfrak{P}} = G(K/Z_{\mathfrak{P}})$, so the primes lying over $\mathfrak{P}_{Z_{\mathfrak{P}}}$ are $\sigma(\mathfrak{P})_{Z_{\mathfrak{P}}}$ by Proposition 76.1, each is $\mathfrak{P}$ and $r' = 1$.

(2), (3): As $r = [G(K/F) : G_{\mathfrak{P}}] = [Z_{\mathfrak{P}} : F] = e''f''r''$, we have $[K : Z_{\mathfrak{P}}] = |G_{\mathfrak{P}}| = ef$. By (1), we have, $e'e''f'f'' = ef = [K : Z_{\mathfrak{P}}] = e'f'$. Thus $e'' = 1 = f''$ and $e' = e$ and $f' = f$ □

We next relate the decomposition group $G_{\mathfrak{P}}$ of $\mathfrak{P}$ in A_K lying over $\mathfrak{p}$ and the Galois group of the field extension $(A_K/\mathfrak{P})/(A/\mathfrak{p})$. Let σ be an element of the $G_{\mathfrak{P}}$, then σ induces an $(A/\mathfrak{p})$-isomorphism

$$\overline{\sigma} : A_K/\mathfrak{P} \to A_K/\mathfrak{P} \text{ given by } x + \mathfrak{P} \mapsto \sigma(x) + \sigma(\mathfrak{P}) = \sigma(x) + \mathfrak{P}.$$

Therefore, we get a group homomorphism

$$^{-} : G_{\mathfrak{P}} \to G\big((A_K/\mathfrak{P})/(A/\mathfrak{p})\big) \text{ given by } \sigma \mapsto \overline{\sigma}. \tag{76.4}$$

Proposition 76.5. *Let A be a Dedekind domain with F its quotient field and $\mathfrak{p}$ a nonzero prime ideal in A. If K/F is a finite Galois extension and $\mathfrak{P}$ a prime in A_K lying over $\mathfrak{p}$, then $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is a finite normal extension and the map $^{-} : G_{\mathfrak{P}} \to G\big((A_K/\mathfrak{P})/(A/\mathfrak{p})\big)$ in (76.4) is surjective.*

PROOF. Let $^{-} : A_K[t] \to (A_K/\mathfrak{P})[t]$ be the natural epimorphism. Since $f(\mathfrak{P} \cap A_{Z_{\mathfrak{P}}}/\mathfrak{p}) = 1$, we have $A/\mathfrak{p} = A_{Z_{\mathfrak{P}}}/(\mathfrak{P} \cap A_{Z_{\mathfrak{P}}})$, so we may assume that $F = Z_{\mathfrak{P}}$, hence $G(K/F) = G_{\mathfrak{P}}$. Let $x \in A_K/\mathfrak{P}$ and choose α in A_K satisfying $\overline{\alpha} = x$. Then we have

$$m_{A/\mathfrak{p}}(x) \mid \overline{m_F(\alpha)} \text{ in } (A/\mathfrak{p})[t]. \tag{*}$$

Since K/F is normal, $m_K(\alpha)$ splits over K, hence $\overline{m_K(\alpha)}$ splits over $A_K/\mathfrak{P}$. It follows that $m_{A/\mathfrak{p}}(x)$ splits over $A_K/\mathfrak{P}$, hence the finite extension $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is normal.

Next, using the Primitive Element Theorem 54.9, we can choose α in K such that $(A/\mathfrak{p})(\overline{\alpha})$ is the maximal separable extension of $A/\mathfrak{p}$ in $A_K/\mathfrak{P}$. We may assume that α lies in A_K. Suppose that $\overline{\sigma}$ lies in $G\big((A_K/\mathfrak{P})/(A/\mathfrak{p})\big)$. Then $\overline{\sigma}$ is completely determined by $\overline{\sigma}(\overline{\alpha})$. As $\overline{\sigma}(\overline{\alpha})$ is a root of $m_{A/\mathfrak{p}}(\overline{\alpha})$, it is a root of $\overline{m_F(\alpha)}$ by (*). Choose a root α' of $m_F(\alpha)$ satisfying $\overline{\alpha'} = \overline{\sigma}(\overline{\alpha})$. There exists an element σ in $G(K/F)$ satisfying $\sigma(\alpha) = \alpha'$ as $G(K/F)$ acts transitively on the roots of $m_F(\alpha)$. Then $\sigma \mapsto \overline{\sigma}$. □

The kernel of (76.4) is denoted by

$$I_{\mathfrak{P}} = \{\sigma \in G_{\mathfrak{P}} \mid \overline{\sigma} = 1_{A_K/\mathfrak{P}}\}$$

and called the *inertia group* of $\mathfrak{P}$, its fixed field

$$T_{\mathfrak{P}} = K^{I_{\mathfrak{P}}}$$

is called the *inertia field* of $\mathfrak{P}$. By Proposition 76.5, we have an isomorphism $G_{\mathfrak{P}}/I_{\mathfrak{P}} \cong G\big((A_K/\mathfrak{P})/(A/\mathfrak{p})\big)$. We know that $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is normal, but it is not Galois in general. However if K is a global field, it is as $A_K/\mathfrak{P}$ is then a finite field.

If the extension $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is separable, then $I_{\mathfrak{P}} = 1$ if and only if $\mathfrak{p}$ is unramified and $G_{\mathfrak{P}} \cong G\big((A_K/\mathfrak{P})/(A/\mathfrak{p})\big)$. More generally, in this case of separability, we have $e(\mathfrak{P}/\mathfrak{p}) = |I_{\mathfrak{P}}| = [K : T_{\mathfrak{P}}]$ and $f(\mathfrak{P}/\mathfrak{p}) = [G_{\mathfrak{P}} : I_{\mathfrak{P}}] = [T_{\mathfrak{P}} : Z_{\mathfrak{P}}]$.

Suppose that K is a global field, so $F = \mathbb{Q}$ and $A = \mathbb{Z}_K$ or F a finite field and $A = F[t]_K$. Of course, if $F = \mathbb{Q}$, then $\mathfrak{p} = (p)$ in $\mathbb{Z}$ for some prime element p. Suppose that $\mathfrak{P}$ is unramified over $\mathfrak{p}$, then there exists a unique element $\sigma_{\mathfrak{P}}$ in $G(K/F)$ satisfying

$$\sigma_{\mathfrak{P}}(x) \equiv x^{f(\mathfrak{P}/\mathfrak{p})} \mod \mathfrak{P}.$$

The automorphism $\sigma_{\mathfrak{P}}$ is called the *Frobenius automorphism*. If K/F is abelian, then $G_{\tau(\mathfrak{P})} = \tau G_{\mathfrak{P}} \tau^{-1} = G_{\mathfrak{P}}$ for all τ in $G(K/F)$. As $G(K/F)$ acts transitively on the primes lying over $\mathfrak{p}$ all decomposition groups over $\mathfrak{p}$ are equal, so the Frobenius homomorphism $\sigma_{\mathfrak{P}}$ depends only on $\mathfrak{p}$. This was the first case, called 'class field theory', thoroughly studied in the first third of the twentieth century.

Exercises 76.6. 1. Let A be a Dedekind domain with quotient field F, $\mathfrak{p}$ a nonzero prime ideal in A. Suppose that K/F is a finite Galois extension and $\mathfrak{P}$ a prime ideal in A_K lying over $\mathfrak{p}$. Show that $T_{\mathfrak{P}}/Z_{\mathfrak{P}}$

is a normal extension satisfying $G(T_{\mathfrak{P}}/Z_{\mathfrak{P}}) \cong G\big(A_K/\mathfrak{P})/(A/\mathfrak{p})\big)$ and $G(K/T_{\mathfrak{P}}) \cong I_{\mathfrak{P}}$.

2. Let A be a Dedekind domain with quotient field F, $\mathfrak{p}$ a nonzero prime ideal in A. Suppose that K/F is a finite Galois extension and $\mathfrak{P}$ a prime ideal in A_K lying over $\mathfrak{p}$ and satisfying $(A_K/\mathfrak{P})/(A/\mathfrak{p})$ is separable. Assuming the previous exercise and Exercise 75.5(2) show
   (*i*) $e(\mathfrak{P}/\mathfrak{p}) = |I_{\mathfrak{P}}| = [K : T_{\mathfrak{P}}]$ and $f(\mathfrak{P}/\mathfrak{p}) = [G_{\mathfrak{P}} : I_{\mathfrak{P}}] = [T_{\mathfrak{P}} : Z_{\mathfrak{P}}]$.
   (*ii*) $e(\mathfrak{P}/\mathfrak{P} \cap T_{\mathfrak{P}}) = e(\mathfrak{P}/\mathfrak{p})$ and $f(\mathfrak{P}/\mathfrak{P} \cap T_{\mathfrak{P}}) = 1$.
   (*iii*) $e(\mathfrak{P} \cap T_{\mathfrak{P}}/\mathfrak{P} \cap Z_{\mathfrak{P}}) = 1$ and $f(\mathfrak{P} \cap T_{\mathfrak{P}}/\mathfrak{P} \cap Z_{\mathfrak{P}}) = f(\mathfrak{P}/\mathfrak{p})$.

77. The Discriminant of a Number Field

In Section 60, we studied the discriminant of a polynomial f in $F[t]$. If K/F is a finite separable extension, then by the Primitive Element Theorem, $K = F(\alpha)$, for some α. Of course there are many such α's, so there is no unique such discriminant. Suppose that F is the quotient field of a Dedekind domain A. Then we know that we can choose α to lie in A. We now have two problems, α is still not unique, and more seriously, A_K is not necessarily $A[\alpha]$ for any α. We investigate what we must do to solve this problem, as the correct definition is crucial to finding the primes in A that ramify in A_K. Although we shall not prove the full result, this addendum should provide a suitable introduction.

Let K/F be a finite, separable field extension of degree n, L/F be a finite Galois extension with L/K, and $\sigma_1, \ldots, \sigma_n : K \to L$ the distinct F-homomorphisms. Let $w_1, \ldots, w_n$ be elements in K and define

$$\Delta(w_1, \ldots, w_n) = \Delta_{K/F}(w_1, \ldots, w_n) := \det\big(\operatorname{Tr}_{K/F}(w_i w_j)\big).$$

Then $\Delta(w_1, \ldots, w_n)$ lies in F as each $\operatorname{Tr}(w_i w_j)$ does. Moreover, if every w_i lies in A_K, then $\Delta(w_1, \ldots, w_n)$ lies in A. We have (cf. also Section 60):

Properties 77.1. Let K/F be a finite, separable field extension of degree n, L/F a Galois extension with L/K, and $\sigma_1, \ldots, \sigma_n : K \to L$ the distinct F-homomorphisms. Let $\mathcal{B} = \{w_1, \ldots, w_n\}$

(1) If $K = F(\alpha)$, then $\{1, \alpha, \ldots, \alpha^{n-1}\}$ is an F-basis for K. Set $\alpha_i = \sigma_i(\alpha)$. Then

$$\Delta(1, \alpha, \ldots, \alpha^{n-1}) = \prod_{i<j} (\alpha_i - \alpha_j)^2$$

is the discriminant of the minimal polynomial $m_F(\alpha)$. We denote it by $\Delta(\alpha)$.

(2) $\Delta(w_1, \ldots, w_n)$ is not zero if and only if $\mathcal{B}$ is an F-basis for K.
(3) If $\mathcal{C}$ is an F-basis for K and $\{w'_1, \ldots, w'_n\}$ its complementary basis, i.e., $\mathrm{Tr}_{K/F}(w_i w'_j) = \delta_{ij}$ for all $1 \leq i, j \leq n$, then
$$\Delta(w_1, \ldots, w_n)\Delta(w'_1, \ldots, w'_n) = 1.$$
(4) If $\mathcal{C}$ is an F-basis for K, then
$$\Delta(w_1, \ldots, w_n) = \det\big((\sigma_i(w_j)\big)^2.$$
(5) If $\mathcal{B}$ and $\{v_1, \ldots, v_n\}$ are two F-bases for K, then
$$\Delta(v_1, \ldots, v_n) = (\det C)^2 \Delta(w_1, \ldots, w_n)$$
with $C \in \mathrm{GL}_n(F)$, the change of basis matrix.
(6) If $F = \mathbb{Q}$ and $\mathcal{B}$ and $\{v_1, \ldots, v_n\}$ are two integral bases for $\mathbb{Z}_K$, then
$$\Delta(v_1, \ldots, v_n) = \Delta(w_1, \ldots, w_n),$$
as the change of basis matrix for these two integral bases must lie in $\mathrm{GL}_n(\mathbb{Z})$. Denote the integer $\Delta(v_1, \ldots, v_n)$, independent of integral basis, by d_K. It is called the *discriminant of K*.
(7) Suppose that $F = \mathbb{Q}$ and $\mathcal{B}$ is an integral basis for $\mathbb{Z}_K$ and $\{v_1, \ldots, v_n\}$ a $\mathbb{Q}$-basis for K with each v_i lying in $\mathbb{Z}_K$. Then there exists an integer a such that $a^2 d_K = \Delta(v_1, \ldots, v_n)$ in $\mathbb{Z}$; and, in fact,
$$|\Delta(v_1, \ldots, v_n)| \geq |d_K|$$
with equality if and only if
$$\{v_1, \ldots, v_n\} \text{ is an integral basis.}$$
[For this reason, an integral basis for $\mathbb{Z}_K$ is often called a *minimal basis*.]

In the case of a ring of algebraic integers, we have the following:

Proposition 77.2. *Let $K/\mathbb{Q}$ be a finite field extension of degree n and $\{\alpha_1, \ldots, \alpha_n\}$ a subset of $\mathbb{Z}_K$ forming a $\mathbb{Q}$-basis for K. Then we have*
$$\mathbb{Z}[\alpha_1, \ldots, \alpha_n] \subset \mathbb{Z}_K \subset \frac{1}{\Delta(\alpha_1, \ldots, \alpha_n)}\mathbb{Z}[\alpha_1, \ldots, \alpha_n].$$

Proof. The first inclusion is trivial. Let $\Delta = \Delta(\alpha_1, \ldots, \alpha_n)$ and $\delta = \det\big((\sigma_i(\alpha_j)\big) \in \mathbb{Z}_K$, where $\sigma_1, \ldots, \sigma_n : K \to \mathbb{C}$ are the $\mathbb{Q}$-embeddings of K. So $\Delta = \delta^2$ by Property 77.1(4). Suppose that $z \in \mathbb{Z}_K$. Then $z = x_1\alpha_1 + \cdots + x_n\alpha_n$ for some $x_1, \ldots, x_n$ in $\mathbb{Q}$. We have
$$\sigma_i(z) = x_1\sigma_i(\alpha_1) + \cdots + x_n\sigma_i(\alpha_n) \text{ for } i = 1, \ldots, n.$$

By Cramer's Rule, $x_i = y_i/\delta$ with y_i the determinant of the matrix $\big(\sigma_i(\alpha_j)\big)$ with the transpose $\big(\sigma_1(z), \dots, \sigma_n(z)\big)^t$, replacing its ith column for each i. Since $y_1, \dots, y_n, \delta$ are algebraic integers, $\Delta x_i = \delta^2 x_i = \delta y_i$ lies in $\mathbb{Z}_K \cap \mathbb{Q} = \mathbb{Z}$, i.e., x_i lies in $\frac{1}{\Delta}\mathbb{Z}$ for $i = 1, \dots, n$. The result follows. □

Definition 77.3. Let A be a Dedekind domain with quotient field F, and K/F a finite separable field extension. The *discriminant ideal* of A_K is the ideal $\mathfrak{D}_{A_K/A}$ of A generated by the elements $\Delta(w_1, \dots, w_n)$ in A for all F-bases $\{w_1, \dots, w_n\}$ of K lying in A_K. Of course, if A is a ring of integers, then $\mathfrak{D}_{A_K/A} = (d_K)$, but in this case it is important to work with the integer d_K as it has a unique sign attached to it.

Proposition 77.4. *Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. Suppose that $K = F(\alpha)$ with α an element of A_K. If $\mathfrak{p}$ is a nonzero prime ideal of A relatively prime to both $\mathfrak{D}_{A_K/A}$ and the conductor $\mathfrak{f}$ of $A[\alpha]$ in A_K, then $\mathfrak{p}$ is unramified in A_K. In particular, only finitely many primes in A ramify in A_K.*

Proof. Let $\bar{}\,: A_K[t] \to (A_K/\mathfrak{p}A_K)[t]$ be the canonical epimorphism and

$$\overline{m_F(\alpha)} = \overline{p_1}^{e_1} \cdots \overline{p_r}^{e_r}$$

be a factorization in $\overline{A}[t]$ into irreducible polynomials with $p_1, \dots, p_r$ monic polynomials in $A[t]$. By Theorem 75.4,

$$\mathfrak{P}_i = \mathfrak{p}A_K + p_i(\alpha)A_K,\ i = 1, \dots r$$

are the prime ideals lying over $\mathfrak{p}$ in A_K and $\mathfrak{p}A_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$. Let $\alpha_1, \dots, \alpha_s$ be the roots of $m_F(\alpha)$ in a finite Galois extension of F containing K. Then $\Delta(\alpha) = \prod_{i<j}(\alpha_i - \alpha_j)^2$, so $\overline{m_F(\alpha)}$ has multiple roots if and only if $\overline{\alpha_i} = \overline{\alpha_j}$ for some $i \neq j$ if and only if $\overline{\Delta(\alpha)} = 0$. Hence $e_i = 1$ for $i = 1, \dots, r$ and each $\overline{p_i}$ has only simple roots. In particular, $A_K/\mathfrak{P}_i^{e_i} = A_K/\mathfrak{P}_i = \overline{A_K}(\overline{\alpha_i})$ is a separable extension of $\overline{A}$. □

Of course in the theorem, if $A_K = A[\alpha]$ for some α, then $\mathfrak{f}$ is the unit ideal, so the condition in this case is that $\mathfrak{p}$ is relatively prime to $\mathfrak{D}_{A_K/A}$. Unfortunately, in general A_K is not $A[\alpha]$ for any α. However, in the number field case, we can show by elementary means that if a prime ideal $p\mathbb{Z}$ in $\mathbb{Z}$ ramifies in $\mathbb{Z}_K$ for $K/\mathbb{Q}$ finite, then p divides d_K.

Proposition 77.5. *Let K be a number field and p a prime integer. If (p) ramifies in $\mathbb{Z}_K$, then $p \mid d_K$.*

PROOF. Let L/K be a finite extension with $L/\mathbb{Q}$ a finite Galois extension. Suppose that $p \nmid d_K$ and (p) ramifies in $\mathbb{Z}_K$, say $p\mathbb{Z}_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$ is a factorization of $p\mathbb{Z}_K$ with $e_1 > 1$. Let $\mathfrak{A} = \mathfrak{P}_1^{e_1-1} \cdots \mathfrak{P}_r^{e_r} > p\mathbb{Z}_K$. We have $\mathfrak{P}_i \mid \mathfrak{A}$ for $i = 1, \ldots, r$. Let $\{x_1, \ldots, x_n\}$ be an integral basis for $\mathbb{Z}_K$ where $n = [K : \mathbb{Q}]$. Choose x in $\mathfrak{A} \setminus p\mathbb{Z}_K$ and write

$$x = m_1 x_1 + \cdots + m_n x_n$$

with $m_1, \ldots, m_n$ in $\mathbb{Z}$. By assumption, there exists an i such that $p \nmid m_i$. We may assume that $i = 1$. We have $\{x, x_2, \ldots, x_n\} \subset \mathbb{Z}_K$ is a $\mathbb{Q}$-basis for K and, using properties of determinants and Property 77.1(7), one checks that

$$\begin{aligned} \Delta(x, x_2, \ldots, x_n) &= \Delta(m_1 x_1 + \cdots + m_n x_n, x_2, \ldots, x_n) \\ &= m_1^2 \Delta(x_1, \ldots, x_n). \end{aligned}$$

As x lies in every prime ideal in $\mathbb{Z}_K$ lying over (p), it lies in every prime ideal in $\mathbb{Z}_L$ lying over (p). Let $\mathfrak{Q}$ be a prime ideal in $\mathbb{Z}_L$ lying over (p). As $G(L/\mathbb{Q})$ acts transitively on the set of prime ideals in $\mathbb{Z}_L$ lying over (p), we have $\sigma(x)$ lies in $\mathfrak{Q}$ for every automorphism σ in $G(L/\mathbb{Q})$. By Property 77.1(4), the element $\Delta(x, x_2, \ldots, x_n)$ lies in $\mathfrak{Q} \cap \mathbb{Z} = (p)$. Consequently, $p \mid \Delta(x, x_2, \ldots, x_n) = m_1^2 \Delta(x_1, \ldots, x_n)$. Since $p \nmid m_1$, we must have $p \mid \Delta(x_1, \ldots, x_n) = d_K$, a contradiction. □

It can be shown that the converse to the last proposition is in fact true, i.e., if p is a prime integer, then (p) ramifies in a ring of algebraic integers $\mathbb{Z}_K$ if and only if $p \mid d_K$. Indeed, in Section §78, we shall prove Dedekind's Ramification Theorem that states if A is a Dedekind domain with quotient field F and K/F is a finite separable field extension, then a prime ideal $\mathfrak{p}$ ramifies in A_K if and only if $\mathfrak{p} \mid \mathfrak{D}_{A_K/A}$. In the exercises for this section, we shall use localization techniques to reduce this to the case of a discrete valuation ring. In this section we shall, however, prove the following special case of it.

Proposition 77.6. *Let p be an odd prime, ζ a primitive pth root of unity in $\mathbb{C}$, and $K = \mathbb{Q}(\zeta)$. Then $\mathbb{Z}_K = \mathbb{Z}[\zeta]$. In particular, $\Delta_K = (-1)^{\frac{p-1}{2}} p^{p-2}$. Moreover, $p\mathbb{Z}$ ramifies in $\mathbb{Z}_K$ and is the only prime ideal that ramifies in $\mathbb{Z}_K$.*

PROOF. By Example 60.7, we know that $\Delta(\zeta) = (-1)^{\frac{p-1}{2}} p^{p-2}$ and that $\mathrm{N}(\zeta) = \mathrm{N}(1 - \zeta)$. Certainly, $\mathbb{Z}[\zeta] = \mathbb{Z}[1 - \zeta]$. Suppose that $\mathbb{Z}[1 - \zeta] = \mathbb{Z}[\zeta] < \mathbb{Z}_K$. Using Proposition 77.2, we can find an element

$z \in \mathbb{Z}_K \setminus \mathbb{Z}[1-\zeta]$ satisfying

$$z = \sum_{j=i}^{p-1} \frac{m_j}{p}(1-\zeta)^j \quad \text{for some } i \geq 0 \text{ and } m_j \in \mathbb{Z} \text{ with } p \nmid m_i.$$

(Why?) By Example 60.7, we know that $\prod_{j=0}^{p-1}(\zeta^j - 1) = p$, hence $p/(1-\zeta)^{p-1}$ lies in $\mathbb{Z}[\zeta]$, as $(1-\zeta) \mid (1-\zeta^j)$ for all $j > 0$. Consequently, $p/(1-\zeta)^j$ lies in $\mathbb{Z}[\zeta]$ for $j = 0, \ldots, p-1$ and $zp/(1-\zeta)^j$ lies in $\mathbb{Z}_K$ for all j. It follows easily that this implies that $m_i/(1-\zeta)$ lies in $\mathbb{Z}_K$. But

$$p = \mathrm{N}_{K/\mathbb{Q}}(1-\zeta) \mid \mathrm{N}_{K/\mathbb{Q}}(m_i) = m_i^{p-1},$$

which is impossible. Therefore, $\mathbb{Z}_K = \mathbb{Z}[\zeta]$.

By Property 77.1(6), we have $d_K = \Delta(\zeta)$. Therefore, $\mathbb{Q}\big(\sqrt{\Delta}\big) = \mathbb{Q}\Big(\sqrt{(-1)^{\frac{p-1}{2}}p}\Big)$, the unique quadratic extension of $\mathbb{Q}$ in K by Proposition 56.14. Clearly, $p\mathbb{Z}$ ramifies in $\mathbb{Z}_{\mathbb{Q}(\sqrt{\Delta})}$, so the proposition follows. □

Remark 77.7. If ζ is a primitive p^rth root of unity and ζ_p a primitive pth root of unity in $\mathbb{C}$ with p an odd prime or $p = 2$ and $r > 1$, then a modification of the proof shows that $\mathbb{Z}_{\mathbb{Q}(\zeta)} = \mathbb{Z}[\zeta] = \mathbb{Z}[1-\zeta]$ and the discriminant $d_{\mathbb{Z}_{\mathbb{Q}(\zeta)}}$ is a power of p by Lemma 60.8. Using Proposition 77.5 and the fact that $\mathbb{Q}\Big(\sqrt{(-1)^{\frac{p-1}{2}}p}\Big)$ is a subfield of $\mathbb{Q}(\zeta)$ if p is odd with $p\mathbb{Z}$ ramifying in $\mathbb{Z}_{\mathbb{Q}\left(\sqrt{(-1)^{\frac{p-1}{2}}p}\right)}$ and with 2 ramifying $\mathbb{Z}[\sqrt{-1}]$ (cf. the next section), we see that such $p\mathbb{Z}$ are precisely the prime ideals ramifying in $\mathbb{Z}_K$. In the general case, one shows that $\mathbb{Z}_L = \mathbb{Z}[\omega]$ for a primitive nth root of unity ω for arbitrary n. This is left as an exercise. The primes p that ramify in $\mathbb{Z}_L$ are precisely those odd primes dividing n and 2 if $4 \mid n$ using a similar analysis.

It can be shown that if $\mathbb{Q} < K$ is a finite extension, then $d_K > 1$ and some prime ideal in $\mathbb{Z}$ ramifies in $\mathbb{Z}_K$.

Exercises 77.8. 1. Prove Properties 77.1.

2. (Stickelberger's Criterion) Let K be a number field. Show that d_K is congruent to 0 or 1 modulo 4.
3. (Brill) Let K be a number field and r the number of embedding of K into $\mathbb{C}$ whose image does not lie in $\mathbb{R}$. Show that the sign of d_K is $(-1)^r$.
4. Let ζ be a primitive p^mth root of unity in $\mathbb{C}$ with p an odd prime or $p = 2$ and $m > 1$. Let $K = \mathbb{Q}(\zeta)$. Show that $\Delta_K \mid p^{p^{m-1}(p-1)}$ and $\mathbb{Z}_K = \mathbb{Z}[\zeta] = \mathbb{Z}[1-\zeta]$. In particular, show if p is an odd prime or

$p = 2$ and $m \geq 2$, then $p\mathbb{Z}$ ramifies in $\mathbb{Z}_K$ and is the only prime ideal ramifying in $\mathbb{Z}_K$.

5. Let $K/\mathbb{Q}$ and $L/\mathbb{Q}$ be finite and d the gcd of d_k and d_L. Show that $\mathbb{Z}_K\mathbb{Z}_L \subset \frac{1}{d}\mathbb{Z}_{KL}$ where KL is the compositum $K(L) - L(K)$ of K and L in $\mathbb{C}$. In particular, if $d = 1$, then $\mathbb{Z}_K\mathbb{Z}_L = \mathbb{Z}_{KL}$.
6. Using the last two exercises, show that if $K = \mathbb{Q}(\zeta)$ with ζ a primitive nth root of unity in $\mathbb{C}$, then $\mathbb{Z}_K = \mathbb{Z}[\zeta]$ and a prime p has $p\mathbb{Z}$ ramifies in K if and only if $p \mid n$ if p is odd or 2 if $4 \mid n$.
7. Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. Let $\mathfrak{p}$ be a maximal ideal in A and S a multiplicative set not containing zero. Show that $S^{-1}\mathfrak{D}_{A_K/A} = \mathfrak{D}_{S^{-1}A_K/S^{-1}A}$. In particular, if $S = A \setminus \mathfrak{p}$, then $\mathfrak{p} \mid \mathfrak{D}_{A_k/A}$ if and only if $\mathfrak{p}S^{-1}A \mid \mathfrak{D}_{S^{-1}A_K/S^{-1}A}$. (Cf. Exercise 75.5(6.) [This reduces the theorem that $\mathfrak{p}$ ramifies in A if and only if $\mathfrak{p} \mid \mathfrak{D}_{A_K/A}$ if and only if it does so locally, i.e., A is a discrete valuation ring.]
8. Let A be a discrete valuation ring with quotient field F and K/F a finite separable field extension with A-basis $\{x_1, \ldots, x_n\}$. Then $\mathfrak{D}_{A_K/A} = \Delta_{K/F}(x_1, \ldots, x_n)A$.

78. Dedekind's Theorem on Ramification

In this short section, we prove that a prime in a Dedekind domain ramifies in a finite separable extension of its quotient field if and only if it divides the discriminant ideal. We shall need to extend the definition of the trace of a finite separable field extension to an arbitrary finite field extension. We do this as follows: Let K/F be a finite field extension and K_s be the elements in K separable over F. The set K_s is a field (why?) called the *separable closure* of F in K. Define $\mathrm{Tr}_{K/F} : K \to F$ by $\mathrm{Tr}_{K/F} := [K : K_s]\,\mathrm{Tr}_{K_s/F}$. In particular, K/F is not separable if and only if $\mathrm{Tr}_{K/F} = 0$. (The norm $\mathrm{N}_{K/F}$ is defined by $\mathrm{N}_{K/F} := (\mathrm{N}_{K_s/F})^{[K:K_s]}$.) Exercises in previous sections that entail the use of localization reduce the proof of our desired result to the case of a discrete valuation domain, i.e., a local Dedekind domain. In particular, assuming these exercises and this extension of the trace, we are reduced to proving the following result:

Proposition 78.1. *Let A be a discrete valuation ring with quotient field F and K/F a finite separable extension. Then the maximal maximal ideal $\mathfrak{p}$ in A ramifies in K if and only if $\mathfrak{p} \mid \mathfrak{D}_{A_K/F}$.*

Proof. Let $n = [K : F]$. Since a discrete valuation ring is a PID by Remark 74.8(5), we know that A_K is a free A-module of rank n, say with integral basis $\{x_1, \ldots, x_n\}$. It follows by Exercise 77.8(8)

that $\mathfrak{D}_{A_K/A} = \Delta_{K/F}(x_1, \dots, x_n)A$. Hence $\mathfrak{p} \mid \mathfrak{D}_{A_K/A}$ if and only if $\Delta_{K/F}(x_1, \dots, x_n) \in \mathfrak{p}$.

Let

(*a*) $^- : A_K \to A_K/\mathfrak{p}A_K$ be the canonical epimorphism.

(*b*) $\overline{A} = A/\mathfrak{p}$.

(*c*) $\mathfrak{p}A_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_r^{e_r}$ be the factorization of $\mathfrak{p}$ in A_K into primes.

(*d*) $B_i = A_K/\mathfrak{P}_i^{e_i}$ for $i = 1, \dots, r$.

Using the Chinese Remainder Theorem, we identify $A_K/\mathfrak{p}A_k$ with $\bigtimes_{i=1}^r B_i$. We know that

$$n = [K : F] = \operatorname{rank}_A(A_K) = \dim_{\overline{A}} \overline{A_K}$$

by Theorem 75.2 (and its proof). Let $\mathcal{B}_i = \{u_{j_{i-1}+1}, \dots, u_{j_i}\}$ be an $\overline{A}$-basis for B_i, $i = 1, \dots r$. So $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_r$ is an $_{\overline{A}}$-basis for $\overline{A_K}$ with $|\mathcal{B}| = n$.

For each y in $\overline{A_K}$ define $\lambda_y : \overline{A_K} \to \overline{A_K}$ by $z \mapsto yz$ and $\operatorname{tr} : \overline{A_K} \to \overline{A_K}$ by $y \mapsto \operatorname{tr}(y) := \operatorname{trace}(\lambda_y)$. Let $\operatorname{tr}_i = \operatorname{tr}|_{B_i}$ for $i = 1, \dots, r$.

Note that we have

(*i*) $\{\overline{x_1}, \dots, \overline{x_n}\}$ is an $\overline{A}$-basis for $\overline{A_K}$.

(*ii*) $\operatorname{tr}(u_i u_j) = 0$ if u_i and u_j lie in different B_k's.

Claim. Let $x \in A_K$, then $\overline{\operatorname{Tr}_{K/F}} = \operatorname{tr}(\overline{x})\, [= \operatorname{trace}(\lambda_{\overline{x}})]$. In particular, there exists a change of basis matrix $C \in \mathrm{GL}_n(\overline{A})$ satisfying

$$\overline{\Delta_{K/F}(x_1, \dots, x_n)} = \big(\det(C)\big)^2 \det\big(\operatorname{tr}(u_i u_j)\big),$$

hence

$$\overline{\Delta_{K/F}(x_1, \dots, x_n)} = 0, \text{ i.e., } \Delta_{K/F}(x_1, \dots, x_n) \in \mathfrak{p},$$
$$\text{if and only if } \det\big(\operatorname{tr}(u_i u_j))\big) = 0 :$$

Write $xx_i = \sum_{j=1}^n a_{ij} x_j$ with $a_{ij} \in A$ for all i and j and let α be the matrix $\big(a_{ij}\big)$. Then α is the matrix representation of λ_x relative to the basis $\{x_1, \dots, x_n\}$ and $\overline{\alpha}$ is the matrix representation of $\lambda_{\overline{x}}$ relative to the basis $\{\overline{x_1}, \dots, \overline{x_n}\}$. By Exercise 57.24(5), we have

$$\overline{\operatorname{Tr}_{K/F}(x)} = \overline{\operatorname{trace}(\lambda_x)} = \operatorname{trace}(\lambda_{\overline{x}}) = \operatorname{tr}(\overline{x}).$$

This proves the Claim. Now let $D_i := \big(\operatorname{tr}_i(u_l u_k)\big)$ for $j_{i-1} < l, k \le j_i$, for each $i = 1, \dots, r$, then

$$D := \Big(\operatorname{tr}_i(u_l u_k)\Big) = \begin{pmatrix} D_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & D_r \end{pmatrix} \qquad (\text{with } j_0 = 0).$$

Suppose that $B_i = A_K/\mathfrak{P}_i^{e_i}$ is a field, i.e., $e_i = 1$. Then

$$\det D_i = \Delta_{B_i/\overline{A}}(u_{j_{i-1}+1}, \ldots, u_{j_i}) = \det\big(\operatorname{Tr}_{B_i/\overline{A}}(u_l u_k)\big).$$

Since $B_i/\overline{A}$ is not separable if and only if $\operatorname{Tr}_{B_i/\overline{A}} = 0$, if follows that if B_i is a field for every i, then $\mathfrak{p} \mid \mathfrak{D}_{A_K/A}$ if and only $\det D_i = 0$ for some i if and only if there exists an i such that $B_i/\overline{A}$ is not separable.

So suppose that there exists an i with B_i not a field. We may assume that $i = 1$ and further that u_1 lies in $\mathfrak{P}_1/\mathfrak{P}_1^{e_1}$. Therefore, $(u_1 u_i)^{e_1} = 0$ for all i, hence $\lambda_{u_1 u_i}$ is nilpotent. It follows that the first row of D consists of 0's, hence $\det D = 0$ and $\mathfrak{p} \mid \mathfrak{D}_{A_K/A}$. □

The generalization of the proposition to arbitrary Dedekind domains now follows by the reduction given by Exercise 77.8(7). That is, we the following theorem:

Theorem 78.2. (Dedekind Ramification Theorem) *Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension. The a maximal ideal $\mathfrak{p}$ of A ramifies in A_K if and only if $\mathfrak{p} \mid \mathfrak{D}_{A_K/A}$.*

Immediate consequences of the theorem are:

Corollary 78.3. *Let A be a Dedekind domain with quotient field F and K/F a finite separable field extension, then A_K/A is unramified if and only if $\mathfrak{D}_{A_K/A} = A$.*

Remarks 78.4. Let F be a number field, so a finite extension of $\mathbb{Q}$.

1. Using Minkowski Theory, one can show that $|d_F| > 1$, so $\mathbb{Z}_F/\mathbb{Z}$ cannot be unramified. As $\mathbb{Q}$ and finite fields are perfect, this means that there exists a prime p and a prime ideal $\mathfrak{P}$ in $\mathbb{Z}_F$ lying over $p\mathbb{Z}$ such that $e(\mathfrak{P}/p\mathbb{Z}) > 1$.
2. It can also be shown that there exist finitely many $F/\mathbb{Q}$ satisfying $d = d_F$, for some fixed integer d.
3. There can, however, be finite extensions K of the number field F that are unramified. Indeed a deep theorem of Classfield Theory says that the maximal unramified abelian extension K of F (in $\mathbb{C}$) is finite and of degree $|Cl_{\mathbb{Z}_F}|$ over F, where $CL_{\mathbb{Z}_F}$ $(= I_{Z_F}/P_{Z_K})$ is the class group of Z_K (finite for a ring of algebraic integers by Minkowski Theory). Furthermore, its Galois group $G(K/F)$ is (canonically) isomorphic to $CL_{\mathbb{Z}_F}$. (K is called the *Hilbert class field* of F.) In particular, $F < K$ if Z_F is not a PID.

Corollary 78.5. *Let $F = \mathbb{Q}(\alpha)$ with α integral over $\mathbb{Z}$ satisfying $g(\alpha) = 0$ with $g \in \mathbb{Z}[t]$ monic and p a prime in $\mathbb{Z}$. If $p \nmid \operatorname{N}_{F/\mathbb{Q}}\big((g'(\alpha)\big)$, then $p\mathbb{Z}$ is unramified in $\mathbb{Z}_K$.*

PROOF. Let $f = m_{\mathbb{Q}}(\alpha)$. By Proposition 60.6, we know that $\Delta_{F/\mathbb{Q}}(\alpha) = \pm \mathrm{N}_{F/\mathbb{Q}}\big(f'(\alpha)\big)$. Let $g = fh$ in $\mathbb{Z}[t]$, then $g' = f'h + fh'$, consequently, $d_F \mid \Delta_{F/\mathbb{Q}}(\alpha)\,\mathrm{N}_{F/\mathbb{Q}}\big(h(\alpha)\big) = \pm \mathrm{N}_{F/\mathbb{Q}}\big(g'(\alpha)\big)$ □

79. The Quadratic Case

Let d be a square-free integer. In this section, we compute the ring theory of $\mathbb{Z}_K$ with $K = \mathbb{Z}[\sqrt{d}]$ to illustrate the theory that we have developed.

Throughout this section let $\mathbb{Z}_K$ with $K = \mathbb{Z}[\sqrt{d}]$ and $G(K/F) = \{1_K, \sigma\}$, the Galois group of the Galois extension $K/\mathbb{Q}$. We first determine an integral basis and discriminant of K. We freely use Properties 77.1. We need a special case of Stickelberger's Criterion (Exercise 77.8(2)), viz.,

$$d_K \equiv 0 \mod 4 \text{ or } d_K \equiv 1 \mod 4.$$

Indeed, if $\{w_1, w_2\}$ is an integral basis for $\mathbb{Z}_K$, let $\alpha = w_1\sigma(w_2)$ and $\beta = w_2\sigma(w_1)$. Then

$$d_K = (\alpha - \beta)^2 = (\alpha + \beta)^2 - 4\alpha\beta.$$

As $(\alpha - \beta)^2$, $(\alpha - \beta)^2$ and $\alpha\beta$ are all fixed by $G(K/\mathbb{Q})$, they are all integers, and the result follows. (The proof in the general case is analogous.) The above notation shall be used throughout this section. We also have:

$$\Delta(1, \sqrt{d}) = \det \begin{pmatrix} 1 & \sqrt{d} \\ 1 & \sigma(\sqrt{d}) \end{pmatrix} = \det \begin{pmatrix} 1 & \sqrt{d} \\ 1 & -\sqrt{d} \end{pmatrix} = 4d.$$

As $\mathbb{Z}[\sqrt{d}] \subset \mathbb{Z}_K$, there exists an integer $a \in \mathbb{Z}$, satisfying

$$a^2 d_K = 4d, \text{ so } d_K = 4d \text{ or } d_K = d,$$

since d is square-free. (Note if $a = 2$, we must have d is odd, lest $d_K \equiv 2 \mod 4$ which contradicts Stickelberger's Criterion. So if d is even, $d_K = 4d$ and $8 \mid d_K$.) Using this and Stickelberger's Criterion, we solve our first goal.

Case 1. $d \equiv 2 \text{ or } 3 \mod 4$:

We have $d_K = 4d$ and $\mathbb{Z}_K = \mathbb{Z}[\sqrt{d}]$. In particular, $\{1, \sqrt{d}\}$ is an integral basis for $\mathbb{Z}_K$ by Property 77.1(7), and the conductor $\mathfrak{f}$ of $\mathbb{Z}[\sqrt{d}]$ in $\mathbb{Z}_K$ is $\mathbb{Z}_K$.

Case 2. $d \equiv 1 \mod 4$:

We have $(1-d)/4$ is an integer, hence $(1 \pm \sqrt{d})/2$ lies in $\mathbb{Z}_K$, as they are the roots of $t^2 - t + (1-d)/4$ in $\mathbb{Z}[t]$. It follows that

$$\mathbb{Z}[\sqrt{d}] \subset \mathbb{Z}[\frac{1+\sqrt{d}}{2}] \subset \mathbb{Z}_K.$$

We know that $\Delta\big(1, (1+\sqrt{d})/2\big) = b^2 d_K$, some integer b and computation yields $\Delta\big(1, (1+\sqrt{d})/2\big) = d$. Therefore, by Property 77.1(7), $d_K = d$, $\mathbb{Z}_K = \mathbb{Z}[(1+\sqrt{d})/2]$, and $\{1, (1+\sqrt{d})/2\}$ is an integral basis for $\mathbb{Z}_K$. In particular, the conductor $\mathfrak{f}$ of $\mathbb{Z}[\sqrt{d}]$ in $\mathbb{Z}_K$ contains the ideal $(2, 1-\sqrt{d}) = (2, 1+\sqrt{d})$.
We show $\mathfrak{f} = (2, 1-\sqrt{d})$. The element $(1+\sqrt{d})/2$ does not lie $\mathbb{Z}[\sqrt{d}]$ hence does not lie in $\mathfrak{f}$, since $\mathfrak{f} \subset \mathbb{Z}[\sqrt{d}]$. It follows if $a+b\big((1+\sqrt{d})/2\big)$ lies in $\mathfrak{f}$ with $a, b \in \mathbb{Z}$, we may assume that $a = b = 1$. As $1+(1+\sqrt{d})/2$ does not lie in $\mathbb{Z}[\sqrt{d}]$, we have $\mathfrak{f} = (2, 1-\sqrt{d})$. If a prime ideal $\mathfrak{P} \mid (1-\sqrt{d})$, then $\mathfrak{P}\overline{\mathfrak{P}} \mid (1-d)\mathbb{Z}$. It follows that no factor of $p\mathbb{Z}_K$ with p an odd prime can be a factor of $\mathfrak{f}$.

Because of this computation, if p is a prime integer we can compute the splitting behavior of $p\mathbb{Z}_K$ in $\mathbb{Z}_K$ for all primes p by Theorem 75.4 except if $p = 2$ and $d \equiv 1 \mod 4$. In particular, it is applicable to all odd primes.

Case of an odd prime. Let p be an odd prime. We know that $2 = [K : \mathbb{Q}] = efr$ with e the ramification index of primes over (p), f the inertia index of (p), and r the number of primes in $\mathbb{Z}_K$ lying over (p). We also know that the Legendre symbol $\big(\frac{d}{p}\big)$ determines whether $t^2 - d$ splits over $\mathbb{Z}/p\mathbb{Z}[t]$ or not when d is not divisible by p, i.e., determines f for such p. Using the Kummer-Dedekind Theorem 75.4 and Proposition 77.4, we conclude, if p is an odd prime, that

1. If $p \mid d_K$, then $t^2 - d = t^2$ in $(\mathbb{Z}/p\mathbb{Z})[t]$, so
$$p\mathbb{Z}_K = (p, \sqrt{d})^2.$$
Therefore (p) ramifies in $\mathbb{Z}_K$ if and only if $p \mid d_K$.
2. The ideal $p\mathbb{Z}_K$ splits completely if and only if $\big(\frac{d}{p}\big) = 1$. If this is the case, then
$$p\mathbb{Z}_K = (p, x+\sqrt{d})(p, x-\sqrt{d})$$
a product of two distinct prime ideals with the integer x satisfying $x^2 \equiv d \mod p$.
3. The ideal $p\mathbb{Z}_K$ remains prime in $\mathbb{Z}_K$ if and only if $\big(\frac{d}{p}\big) = -1$.

Case of the even prime 2:

If $d \equiv 2 \mod 4$, then Theorem 75.4 applies, and we see that in either case (2) ramifies (note $2 \mid d_K$), because

1. If $d \equiv 2 \mod 4$, then $2\mathbb{Z}_K = (2, \sqrt{d})^2$.
2. If $d \equiv 3 \mod 4$, then $2\mathbb{Z}_K = (2, d + \sqrt{d})^2 = (2, 1 + \sqrt{d})^2$.

So we are reduced to the case that $d \equiv 1 \mod 4$, and this reduces to the two cases of $d \equiv 1 \mod 8$ and $d \equiv 5 \mod 8$. Since Theorem 75.4 does not apply, we do this by brute force. Notice the computation shows that 2 does not ramify as expected and $(1 + \sqrt{d})/2$ lies in $\mathbb{Z}_K$.

We have

3. If $d \equiv 1 \mod 8$, then (2) splits completely in $\mathbb{Z}_K$:

   We have

   $$(2, \frac{1+\sqrt{d}}{2})(2, \frac{1-\sqrt{d}}{2}) = (4, 1+\sqrt{d}, 1-\sqrt{d}, \frac{1-d}{4}) \subset 2\mathbb{Z}_K$$

   As $2\mathbb{Z}_K \subset (1 + \sqrt{d}, 1 - \sqrt{d})$, we conclude that

   $$2\mathbb{Z}_K = (2, \frac{1+\sqrt{d}}{2})(2, \frac{1-\sqrt{d}}{2})$$

   splits completely.
4. If $d \equiv 5 \mod 8$, then (2) remains a prime ideal in $\mathbb{Z}_K$:

   We must show that $\mathbb{Z}/2\mathbb{Z} < \mathbb{Z}_K/\mathfrak{P}$ if $\mathfrak{P}|2$. If this was false, there are only two cosets of $\mathbb{Z}_K/\mathfrak{P}$, so there exists an integer a satisfying $a \equiv (1+\sqrt{d})/2 \mod \mathfrak{P}$. Since $(1+\sqrt{d})/2$ is a root of $t^2-t+(1-d)/4$, we have $a^2 - a + (1-d)/4$ is even. As $a^2 - a$ is even, we would have $d \equiv 1 \mod 8$, a contradiction. Therefore, (2) remains a prime ideal.

As previously mentioned, a theorem of Kronecker says that any abelian extension of $\mathbb{Q}$ lies in a cyclotomic extension of $\mathbb{Q}$. We have shown that every quadratic extension of $\mathbb{Q}$ is a subfield of a cyclotomic extension of $\mathbb{Q}$ in Theorem 56.17. We use the proposition that implied this together with Hilbert Ramification Theory to interpret when an ideal (p) with p an odd prime in $\mathbb{Z}$ splits completely in $\mathbb{Z}_K$.

Proposition 79.1. *Let p and l be odd rational primes, $L = \mathbb{Q}(\zeta)$ with ζ a primitive l th root of unity, and $K = \mathbb{Q}(\sqrt{(\frac{-1}{l})l}\,)$. Then (p) splits completely in $\mathbb{Z}_K$ if and only if (p) splits into an even number of prime ideals in $\mathbb{Z}_L$.*

PROOF. By Proposition 56.14, we have K is a subfield of L. Suppose that p splits completely in $\mathbb{Z}_K$, say $p\mathbb{Z}_K = \mathfrak{P}_1\mathfrak{P}_2$ is its factorization. There exists an automorphism $\sigma \in G(L/F)$ satisfying $\sigma(\mathfrak{P}_1) = \mathfrak{P}_2$ by Proposition 76.1. Then σ takes the set of prime ideals in $\mathbb{Z}_L$ over $\mathfrak{P}_1$ bijectively onto the set of prime ideals over $\mathfrak{P}_2$. Therefore the set of primes over (p) in $\mathbb{Z}_L$ is even. Conversely, suppose that the number of prime ideals in $\mathbb{Z}_L$ over (p) is even. Let $\mathfrak{P}$ be one such. Then the decomposition group $G_{\mathfrak{P}}$ of $\mathfrak{P}$ in $G(L/F)$ has even index in $G(L/F)$. By Galois Theory, this means the degree of the decomposition field $Z_{\mathfrak{P}}$ over $\mathbb{Q}$ is even. But $L/\mathbb{Q}$ is cyclic, so $K \subset Z_{\mathfrak{P}}$ using Proposition 76.3. □

Using the fact that if ζ is a primitive nth root of unity that $\mathbb{Z}_{\mathbb{Q}(\zeta)} = \mathbb{Z}[\zeta]$, a fact that we did not prove but left as exercises (cf. Exercise 77.8(6)), one can give another proof of the Law of Quadratic Reciprocity essentially due to Kronecker. Although we do not do this, it is key to the generalization of the Law of Quadratic Reciprocity in Number Theory called Artin Reciprocity. (Cf. Ireland and Rosen, *A Classical Introduction to Modern Number Theory*, Chapter 13, §3 for this proof.)

We have determined the ideal structure of the quadratic number field $\mathbb{Z}_K$ and from our work how nonzero ideals in $\mathbb{Z}$ decompose in $\mathbb{Z}_K$ as $\mathbb{Z}_K$ is a Dedekind domain and ideals are products of prime ideals, unique up to order. The remaining problem is to determine the units in $\mathbb{Z}_K$. This depends on whether d is positive or negative. The case of negative d is easy, the case of positive d not so.

Case of $d < 0$. We show that

1. If $d = -1$, then $\mathbb{Z}_K^\times = \{\pm 1, \pm\sqrt{-1}\}$.
2. If $d = -3$, then $\mathbb{Z}_K^\times = \{1, \zeta, \zeta^2\}$, with ζ a primitive cube root of unity.
3. If $d < -3$ or $d = -2$, then $\mathbb{Z}_K^\times = \{\pm 1\}$.

We have seen this for the Gaussian integers, and the proof generalizes, i.e., we take the norm for K to $\mathbb{Q}$. Let ε be a unit in $\mathbb{Z}_K$. Then $\mathrm{N}_{K/F}(\varepsilon) = \pm 1$.

If $d \equiv 2$ or 3 modulo 4, we can write $\varepsilon = x + y\sqrt{d}$ for some integers x and y, so $x^2 + |d|y^2 = 1$. If $d = -1$, we get (1) and if $|d| > 3$ we have $\mathbb{Z}_K^\times = \{\pm 1\}$.

If $d \equiv 1 \mod 4$, then $\{1, (1 + \sqrt{d}/2\}$ is an integral basis, and we see that we can write ε as $(x + y\sqrt{d})/2$ with x and y integers satisfying

$x \equiv y \mod 2$. We then have $x^2 + |d|y^2 = 4$. If $d = -3$ we get (2) and if $|d| > 3$, we see that $\mathbb{Z}_K^\times = \{\pm 1\}$.

To do the case for positive d, we need two results. The first is the following:

Observation 79.2. If M is a positive real number, there exist finitely many α in $\mathbb{Z}_K$ such that $\max(|\alpha|, |\sigma(\alpha)|) < M$.

and the second a famous theorem from elementary number theory, viz.,

Proposition 79.3. *Let d be a square-free positive integer. Then Pell's equation $x^2 - dy^2 = 1$ has infinitely many solutions in integers. Further, there exists an integral solution (x_1, y_1) such that every solution is of the form (x_n, y_n) with $x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})^n$ for some integer n. In particular, the solution set in integers to Pell's equation forms an infinite cyclic group.*

that we prove in Appendix §106.

Using these two results, we prove the case of positive discriminant.

Case of $d > 1$. We show that there exists a unit $u > 1$ in $\mathbb{Z}_K$ such that $(\mathbb{Z}_K)^\times = \{\pm u^m \mid m \in \mathbb{Z}\}$.

By Proposition 79.3, there exist positive integers x and y satisfying $x^2 - dy^2 = 1$. Let $\varepsilon = x + y\sqrt{d}$. Then ε is a unit in $\mathbb{Z}_K$ with $\varepsilon > 1$. Let $M > \varepsilon$ and u be a unit in $\mathbb{Z}_K$ satisfying $1 < u < M$. So $\mathrm{N}_{K/\mathbb{Q}} = u\sigma(u) = \pm 1$. If $\sigma(u) = -1/u$, then $-M < -1/u < M$ and if $\sigma(u) = 1/u$, then we also have $-M < 1/u < M$. By the observation, there exist only finitely many such u and ϵ is at least one of them. Among all of these choose the smallest unit $u > 1$. If $v > 0$ is any other unit, then there exist a unique integer n (not necessarily positive) satisfying $u^n \leq v < u^{n+1}$, so $1 \leq vu^{-n} < u$. By the choice of u, we must have $v = u^n$. If $v < 0$ is a unit in $\mathbb{Z}_K$, then $-v > 0$ is a unit in $\mathbb{Z}_K$ and $-v = u^n$ some integer n. The result follows.

(Note the similarity of the proof with that in Appendix §106.)

The generator of the unit group in this case is called the *fundamental unit*. It can be found using continued fractions, but may not be so obvious. For example, the fundamental unit of $\mathbb{Z}[\sqrt{94}]$ is $2143295 + 221064\sqrt{94}$.

If K is a number field, then the structure of $\mathbb{Z}_K$ can be shown to be $\mu_K \times \mathbb{Z}^n$ where $n = r + s - 1$, with r the number of embeddings of K into $\mathbb{R}$ and $2s$ the number of embeddings of K into $\mathbb{C}$ not contained in $\mathbb{R}$, so the number of embeddings into $\mathbb{C}$ is $r + 2s$ (as the non-real embeddings come in complex conjugate pairs). This is proved using Minkowski Theory.

Exercise 79.4. Prove Observation 79.2.

80. Addendum: Valuation Rings and Prüfer Domains

In this addendum, we shall generalize the concept of Dedekind domain. By Exercise 74.14(14), a Dedekind domain is a Noetherian domain A in which the localization at every nonzero prime ideal $\mathfrak{p}$ in A is a discrete valuation ring. Another characterization of a Dedekind domain was that it was a Noetherian domain in which every fractional ideal in R was invertible. We look at these when we do not assume the domain is Noetherian.

In Exercise 74.14(13) we defined a valuation ring. We begin by recalling the definition.

Definition 80.1. A domain A with quotient field F is called a *valuation ring* of F if for every nonzero element x in F, either x lies in A or x^{-1} lies in A. In particular, if a and b are nonzero elements in A, then $a \mid b$ or $b \mid a$ in A. We call a domain a *valuation ring* if it is a valuation ring in its quotient field.

Every discrete valuation ring is a Noetherian valuation ring by Exercise 74.14(13), hence the localization of any Dedekind domain at a nonzero prime ideal is a valuation ring. Valuation rings are ubiquitous, unfortunately a valuation ring is Noetherian only when it is a discrete valuation ring. The general concept, however, was very important in the algebraization of algebraic geometry.

Properties 80.2. Let A be a valuation ring of F. Then

(1) A is a local ring.
(2) if $a_1, \ldots, a_n$ are nonzero elements in A, then there exists an i such that $a_i \mid a_j$ for $j = 1, \ldots, n$.
(3) If $A \subset B \subset \mathrm{F}$ are subrings, then B is a valuation ring.
(4) A is integrally closed.
(5) Every finitely generated ideal in A is principal. A domain in which every finitely generated ideal is principal is called a *Bézout domain*.

PROOF. (1): Let $\mathfrak{A} = A \setminus A^\times$, the set of nonunits in A. As A is a valuation ring, a lies in $\mathfrak{A}$ if and only if a^{-1} does not lie in $\mathfrak{A}$.

Claim. $\mathfrak{A}$ is an ideal.

Let a, b be nonzero elements in $\mathfrak{A}$ and r a nonzero element in A. Certainly, $ra \in A$. Suppose that $ra \notin \mathfrak{A}$. Then $(ra)^{-1} \in A$, hence $a^{-1} = r(ra)^{-1}$ lies in A, a contradiction. So $\mathfrak{A}$ is closed under multiplication by elements in A. As either ab^{-1} or $a^{-1}b$ lies in A, say ab^{-1},

it follows that $a + b = b(ab^{-1} + 1)$ lies in $\mathfrak{A}$. Hence $\mathfrak{A}$ is an ideal and clearly the unique maximal ideal.

Statements (2) and (3) are immediate, (4) follows by Exercise 80.27(1), and (5) follows form (2). □

We conclude the following:

Corollary 80.3. *Let A be a noetherian domain, not a field. Then A is a valuation ring if and only if A is a discrete valuation ring.*

PROOF. ($\Rightarrow$): By the Properties 80.2, A is a local PID so every nonzero prime ideal is maximal and integrally closed. Since it is Noetherian, it is a local Dedekind domain, hence a discrete valuation ring.

($\Leftarrow$): Is Exercise 74.14(13). A key to the proof is if $\mathfrak{m}$ is the maximal ideal of A, then $x^{\times} \setminus \mathfrak{m}^2$ must satisfy $(x) = \mathfrak{m}$ by uniqueness of factorization of ideals into prime ideals in s Dedekind domain. It follows that A must be a PID and the result follows. □

We next look at fractional ideals in a domain a bit more closely.

Let A be a domain with quotient field F and M a fractional ideal, i.e., a nonzero A-submodule of F such that there exists a nonzero element of A satisfying $xM \subset A$. As before, we shall let I_A denote the set of fractional ideals in A. We say a fractional ideal M is an *invertible fractional ideal* if there exists an A-submodule N of F such that $MN = A$. Let $Inv(A)$ denote the set of invertible fractional ideals. As before, we let $P_A = \{Ax \mid 0 \neq x \in F\}$, the set of *principal fractional ideals*.

Remarks 80.4. Let A be a domain with quotient field F and M an A-submodule of F.

1. Every nonzero ideal of A is a fractional ideal.
2. A nonzero finitely generated A-submodule of F is a fractional ideal — clear denominators.
3. Let x be a nonzero element of F, then $M \in Inv(A)$ if and only if $xM \in Inv(A)$. In particular, $P_A \subset Inv(A)$.
4. If A is Noetherian and $M \in I_A$, then there exists a nonzero x in F such that xM is an ideal. In particular, M is a finitely generated A-module. It follows that if A is Noetherian, I_A is the set of nonzero finitely generated A-submodules of F.

As before if $M \in I_A$, we let $M^{-1} := \{x \in F^{\times} \mid xM \subset A\}$. Of course, $MM^{-1} \subset A$.

Lemma 80.5. *Let A be a domain with quotient field F and M, N be A-submodules of F satisfying $MN = A$. Then $N = M^{-1}$ and M is a finitely generated A-module. In particular, $M \in I_A$.*

PROOF. By definition $N \subset M^{-1}$, so $N \subset M^{-1}MN \subset AN = N$. It follows that $N = M^{-1}$. If M is finitely generated then $M \in I_A$ by Remark 2 above. In any case, since $A = MM^{-1}$, we have an equation $1 = \sum_{i=1}^{n} a_i b_i$ for some a_i in M, some $b_i \in M^{-1}$, and some n. Then $M = \sum_{i=1}^{n} Aa_i$. Indeed, if $x \in M$, then $x = \sum_i a_i(b_i x)$ lies in $\sum_{i=1}^{n} Aa_i$. □

We call a commutative ring a *semi-local ring* if it has only finitely many maximal ideals.

Lemma 80.6. *If A is a semi-local domain, then $Inv(A) = P_A$.*

PROOF. Let $\mathfrak{m}_1, \ldots, \mathfrak{m}_n$ be the maximal ideals in A and suppose that $M \in Inv(A)$, so $MM^{-1} = A$. For each $i = 1, \ldots, n$, choose $a_i \in M$ and $b_i \in M^{-1}$ such that $a_i b_i \notin \mathfrak{m}_i$. Since $\bigcap_{j \neq i} \mathfrak{m}_j \not\subset \mathfrak{m}_i$ for $i = 1, \ldots, n$, there exist $c_i \in \left(\bigcap_{j \neq i} \mathfrak{m}_j\right) \setminus \mathfrak{m}_i$ for $i = 1, \ldots n$. Let $x = \sum c_i b_i$, an element of M^{-1}. Therefore, xM is an ideal of A. If $xM = A$, then x is not zero and $M = x^{-1}A$ is a principal fractional ideal. So suppose that $xM < A$. Then $xM \subset \mathfrak{m}_i$ for some i. In particular, $xa_i \in \mathfrak{m}_i$. As $c_j b_j a_i \in \mathfrak{m}_i$ for every $j \neq i$, we have $c_i b_i a_i$ lies in $\mathfrak{m}_i$. This is impossible. □

Let A be a commutative ring and $S \subset A$ a multiplicative set. If M is an A-module, then $S^{-1}M := \{\frac{m}{s} \mid s \in S, m \in M\}$ is an A- and an $S^{-1}A$-module in the obvious way (which is?). If $\mathfrak{p}$ is a prime ideal in A, we denote by $M_\mathfrak{p}$ the $A_\mathfrak{p}$-module $S^{-1}M$ with $S = A \setminus \mathfrak{p}$. The following is straight-forward:

Lemma 80.7. *If A is a domain, S a multiplicative set in A not containing 0, and $M \in Inv(A)$, then $S^{-1}M \in Inv(S^{-1}A)$.*

We have the following 'local-global' principle:

Proposition 80.8. *Let A be a domain with quotient field F and M an A-submodule of F. Then the following are equivalent:*

1. *$M \in Inv(A)$.*
2. *M is a finitely generated A-module and $M_\mathfrak{p} \in Inv(A)$, for all prime ideals $\mathfrak{p}$ in A.*
3. *M is a finitely generated A-module and $M_\mathfrak{m} \in Inv(A)$, for all maximal ideals $\mathfrak{m}$ in A.*
4. *M is a finitely generated A-module and $M_\mathfrak{m} \in P_A$, for all maximal ideals $\mathfrak{m}$ in A.*

PROOF. (1) $\Rightarrow$ (2) follows from lemmas 80.7 and 80.5, (2) $\Rightarrow$ (3) is immediate, and (3) $\Rightarrow$ (4) follows from Lemma 80.6.

(4) $\Rightarrow$ (1): Suppose that $MM^{-1} < A$. Then there exist a maximal ideal $\mathfrak{m}$ in A with $MM^{-1} \subset \mathfrak{m}$. By assumption, M is finitely generated, say $M = Aa_1 + \cdots + Aa_n$, and there exists a nonzero x in M satisfying $xA_{\mathfrak{m}} = M_{\mathfrak{m}}$. By the definition of localization, there exist nonzero s_i in $A \setminus \mathfrak{m}$ satisfying $s_i a_i \in Ax$, $i = 1, \ldots, n$. Set nonzero $s = s_1 \cdots s_n$. Then $sM \subset Ax$, hence $sx^{-1}M \subset A$, i.e., $sx^{-1} \in M^{-1}$. It follows that $s = sx^{-1}x$ lies in $M^{-1}M \subset \mathfrak{m}$, a contradiction. Therefore, $MM^{-1} = A$ as desired. □

A domain is called a *Bézout domain* if every finitely generated ideal in A is principal and a *Prüfer domain* if every finitely generated ideal is invertible.

Remarks 80.9. 1. Every Bézout domain is a Prüfer domain.

2. A is a Bézout domain if and only if $Inv(A) = P_A$.

3. A domain is a Prüfer domain if and only if every finitely generated fractional ideal is invertible

We leave the following generalization of the fact that a DVR is a PID as an easy exercise:

Proposition 80.10. *Let A be a local domain. Then A is a valuation ring if and only if A is a Bézout domain.*

The following shows that Prúfer domains generalize Dedekind domains when the Noetherian assumption is omitted.

Theorem 80.11. *Let A be a domain. Then the following are equivalent:*

1. *A is a Püfer domain.*

2. *$A_{\mathfrak{p}}$ is a valuation ring for all prime ideals $\mathfrak{p}$ in A.*

3. *$A_{\mathfrak{m}}$ is a valuation ring for all maximal ideals $\mathfrak{m}$ in A.*

PROOF. (1) $\Rightarrow$ (2): Let $\mathfrak{A}$ be a finitely generated ideal in $A_{\mathfrak{p}}$ with $\mathfrak{p}$ a prime ideal in A. Then there exist a_i in $R \setminus \mathfrak{p}$ and s_i in S, $i = 1, \ldots, n$, some n, such that $\mathfrak{A} = \sum_{r=1}^{n} A_{\mathfrak{p}} \frac{a_i}{s_i} = (\sum_{i=1}^{n} Aa_i)_{\mathfrak{p}}$. (Why?) Since $\sum Aa_i$ is invertible so is $\mathfrak{A}$. But then $\mathfrak{A}$ is principal by Proposition 80.8 as $A_{\mathfrak{p}}$ is local. In particular, $A_{\mathfrak{p}}$ is Bézout, hence a valuation ring by the previous proposition.

(2) $\Rightarrow$ (3) is immediate.

(3) $\Rightarrow$ (1): Let $\mathfrak{A}$ be a finitely generated ideal in A. If $\mathfrak{m}$ is a maximal ideal in A, then $\mathfrak{A}_{\mathfrak{m}}$ is principal as $A_{\mathfrak{m}}$ is a valuation ring. Consequently, it is invertible by Proposition 80.8. □

We know that the integral closure of a Dedekind domain in a finite separable closure of its quotient field is also a Dedekind domain (although separable is not needed by the Krull-Akizuki Theorem 84.17 to be proven below). We shall generalize this to a Prüfer domain. Indeed we shall obtain the stronger result that the field extension need only be algebraic, i.e., we do not have to assume that it is finite nor separable. To do so we introduce a concept that leads to the ubiquity of valuation rings.

Definition 80.12. Let A be a subring of B and $\mathfrak{A} < A$ an ideal. We say that $\mathfrak{A}$ *survives* in B if $\mathfrak{A}B < A$.

For example, by Claim 75.1, if A is Dedekind domain then any ideal $\mathfrak{A} < A$ survives in A_L for any field extension L of $qf(A)$. Indeed this is true if B/A is an integral extension for any commutative ring A (as we shall prove in Theorem 82.14 below).

Lemma 80.13. (Chevalley) *Let A be a subring of the commutative ring B and $\mathfrak{A} < A$ an ideal. If u is a unit in B, then $\mathfrak{A}$ survives in either $A[u]$ or in $A[u^{-1}]$.*

PROOF. Suppose this is false. Then there exist elements $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$ in $\mathfrak{A}$ satisfying:

$$1 = a_0 + a_1u + \cdots + a_nu^n \tag{i}$$

$$1 = b_0 + b_1u + \cdots + b_mu^{-m} \tag{ii}$$

in B, where we may assume these have been chosen with $m, n \geq 0$ minimal. Since $\mathfrak{A} < A$, we may also assume that m and n are positive. We also assume that $m \leq n$. Note that $a_0 \neq 1$ and $b_0 \neq 1$ as $\mathfrak{A} < A$. Multiplying equation (ii) by a_nu^n yields

$$a_nu^n(1 - b_0) = b_1a_nu^{n-1} + \cdots + b_ma_nu^{n-m} \tag{iii}$$

and multiplying (i) by $1 - b_0$ yields

$$1 - b_0 = (1 - b_0)a_0 + \cdots + (1 - b_0)a_nu^n. \tag{iv}$$

Plug (iii) into (iv) for $a_n(1 - b_0)u^n$ gives an equation of the form (i) of lesser degree in u than n, a contradiction. □

Theorem 80.14. *Let A be a domain with quotient field F with K/F a field extension, and $\mathfrak{A} < A$ an ideal. Then there exists a valuation ring B of K containing A with $\mathfrak{A}$ surviving in B.*

PROOF. Let

$$\mathcal{S} = \{(A_\alpha, \mathfrak{A}_\alpha) \mid A \subset A_\alpha \subset K \text{ subrings}, \\ \mathfrak{A} \subset \mathfrak{A}_\alpha < A_\alpha \text{ with } \mathfrak{A}_\alpha \text{ an ideal in } A_\alpha\}$$

Partially order $\mathcal{S}$ by $\subseteq$ of pairs. Since $(A, \mathfrak{A}) \in \mathcal{S}$, a Zorn Lemma argument produces a maximal element $(B, \mathfrak{B}) \in \mathcal{S}$. Suppose that B is not a valuation ring of K. Then there exists a nonzero element x in K with neither x nor x^{-1} in B. As $\mathfrak{B}$ survives in $B[x]$ or $B[x^{-1}]$, say $B[x]$, we would have $(B, \mathfrak{B}) < (B[x], B[x]\mathfrak{B})$. Since $B[x] \subset B[x, x^{-1}] \subset K$, the pair $(B[x], B[x]\mathfrak{B})$ lies in $\mathcal{S}$, a contradiction. □

Definition 80.15. Let A and B be local rings with maximal ideals $\mathfrak{m}, \mathfrak{n}$, respectively. We say that B *dominates* A if A is a subring of B and $\mathfrak{m} = A \subset \mathfrak{n}$.

Corollary 80.16. *Let A be a local domain with its quotient field F lying in a field K. Set*

$$\mathcal{D}_K = \{B \text{ a local ring} \mid B \subset K\}$$

partially ordered by domination. Then

(1) There exists a valuation ring V in $\mathcal{D}_K$ dominating A.
(2) A is a valuation ring in $\mathcal{D}_F$ if and only if A is maximal in $\mathcal{D}_F$.

PROOF. (1) follows from Theorem 80.14 with $\mathfrak{A}$ the maximal ideal of A.

(2): The proof of Theorem 80.14 implies the if statement. Conversely, suppose that $A \subset B \subset F$ are valuation rings of F with maximal ideals $\mathfrak{m}_A$ and $\mathfrak{m}_B$, respectively, with B dominating A, we must show that $A = B$. If $A < B$, then there exists a nonzero element x in $B \setminus A \subset F$. Since A is a valuation ring, x^{-1} lies in $A \subset B$, so $x \in B^\times \setminus A$. It follows that x^{-1} lies in $\mathfrak{m}_A = A \setminus A^\times$. Thus $\mathfrak{m}_B \cap A = \mathfrak{m}_A = A \setminus A^\times$, a contradiction. □

This corollary has an application to Prüfer domains.

Corollary 80.17. *Let R be a Prüfer domain with quotient field F. Suppose that A is a valuation ring in F containing R. Then there exists a prime ideal $\mathfrak{p}$ in R such that $A = R_\mathfrak{p}$.*

PROOF. Let $\mathfrak{m}$ be the maximal ideal of A and $\mathfrak{p} = \mathfrak{m} \cap R$, a prime ideal in R. Let $s \in A \setminus \mathfrak{p}$. Then s^{-1} lies in A, lest $s \in \mathfrak{m} \cap R$. Therefore, $R_\mathfrak{p} \subset A$. By Theorem 80.11 $R_\mathfrak{p}$ is a valuation ring, hence by Corollary 80.16, we have $R_\mathfrak{p} = A$. □

This corollary gives a starting point for the theory of algebraic functions in one variable (geometrically curve theory). The algebraic formulation is the following:

Corollary 80.18. *Let F be a field and A a valuation ring satisfying $F \subset A \subset F(t)$. Then there exists an irreducible polynomial f in $F[t]$ satisfying $A = F[t]_{(f)}$ or $A = F[t^{-1}]_{(t^{-1})}$*

PROOF. If t lies in A, then $F[t] \subset A$. As $F[t]$ is a PID, it is Prüfer, so the result follows from Corollary 80.17. So we may assume that $t \notin A$. As A is a valuation ring t^{-1} lies in A. By Corollary 80.17, $A = F[t^{-1}]_{\mathfrak{p}}$ for some prime ideal in $F[t^{-1}]$. Since $t \notin A$, we have $t^{-1} \in \mathfrak{p}$, so $(t^{-1}) = \mathfrak{p}$. □

We now wish to generalize Theorem 74.5 to the more general theorem about Prüfer domains previously mentioned. We need the following lemma.

Lemma 80.19. *Let A be an integrally closed local domain with quotient field F and $a \in F$. Suppose that there exists a polynomial f in $A[t]$ with $f(a) = 0$ and at least one coefficient of f is a unit in A. Then either a or a^{-1} lies in A.*

PROOF. Suppose that $ba^n + ca^{n-1} + g(a) = f(a) = 0$ with $b, c \in A$ and $g \in A[t]$ satisfying $\deg g < n - 1$. If b is a unit in A, then we are done as A is integrally closed, so we may assume that $b \notin A^\times$. Then

$$(*) \qquad 0 = (ba)^n + c(ba)^{n-1} + b^{n-1}g(a) = (ba)^n + c(ba)^{n-1} + g_1(ba)$$

for some $g_1 \in A[t]$ satisfying $\deg g_1 < n-1$. Since A is integrally closed, ba lies in A, as ba is integral over A by (*). If ba is a unit in A, then so is $b^{-1}a^{-1}$. This implies that $a^{-1} \in A$. Therefore, we ay assume that ba lies in the maximal ideal $\mathfrak{m}$ of A. We know that $(ba+c)a^{n-1}+g(a) = 0$. If $c \in A^\times$, then $ba + c \in A^\times$, as A is local and $ba \in \mathfrak{m}$. But this means that a is integral over A, so $a \in A$ as A is integrally closed. Therefore, we may assume that $ba + c$ lies in $\mathfrak{m}$ and one of the coefficients of g is a unit. As $\deg\big((ba + c)t^{n-1} + g\big) < n$, we are done by induction on n. □

Theorem 80.20. *Let A be a Prüfer domain with quotient field F. Let L/F be an algebraic (possibly infinite) algebraic field extension. Then A_L is Prüfer.*

PROOF. Let $\mathfrak{m}$ be a maximal ideal in A_L and $\mathfrak{p} = A \cap \mathfrak{m}$. We need to show that $B_{\mathfrak{m}}$ is a valuation ring. Since A is a Prüfer domain, $A_{\mathfrak{p}}$ is a valuation ring by Theorem 80.11. Let a be an element in L and f a nonzero polynomial in $F[t]$ with $f(a) = 0$. Multiplying by a suitable

element in A, we may assume that $f \in A[t]$. Since $A_{\mathfrak{p}}$ is a valuation ring there exists a coefficient r of f dividing all the other coefficients of f in A. Thus $\frac{1}{r}f \in A_{\mathfrak{p}}[t]$ and has a coefficient that is a unit in $A_{\mathfrak{p}}$, hence in $(A_L)_{\mathfrak{m}}$. By the lemma, either a or a^{-1} lies in $(A_L)_{\mathfrak{m}}$ as needed. □

Corollary 80.21. *Let A be a Dedekind domain with quotient field F. Let L/F be an algebraic (possibly infinite) algebraic field extension, Then A_L is a Prüfer domain. In particular, if $\widetilde{\mathbb{Q}}$ is an algebraic closure of $\mathbb{Q}$, then $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$ is a Prüfer domain.*

Remarks 80.22. 1. If $\widetilde{\mathbb{Q}}$ is an algebraic closure of $\mathbb{Z}$, then $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$ is not Noetherian so not a Dedekind domain.

2. The ring of entire functions on an open Riemann surface is a Prúfer domain. In fact, it is a Bézout domain.
3. If in the proposition, Bézout is substituted for Prúfer, then it can be shown that A_L is also Bézout.
4. Bergman showed that if $A = \mathbb{Z}\big[\sqrt{-5}\,\big] + t\mathbb{Q}\big[\sqrt{-5}\,\big][t]$, i.e., A is the ring of polynomials in $\mathbb{Q}\big[\sqrt{-5}\,\big][t]$ having constant term in $\mathbb{Z}\big[\sqrt{-5}\,\big]$, then A is not Noetherian, not Bézout, but is Prüfer.
5. Every ideal in a Dedekind domain can be generated by two elements by Exercise 74.14(11). The same is false for finitely generated ideals in Prüfer domains. Schulting gave the first counterexample. It was subsequently generalized by Swan.

We give two further applications of valuation rings that are useful in commutative algebra. The first has as a consequence Zariski's Lemma 38.10 that we saw was the key to the Hilbert Nullstellensatz in Section §38.

Theorem 80.23. *Let F be an algebraically closed field and A a domain with $R \subset A$ a subring. Suppose that A is a finitely generated R-algebra and $a \in A$ nonzero. Then there exists an element $0 \neq r \in R$ satisfying the following condition: whenever $\varphi : A \to F$ is a ring homomorphism with $0 \neq \varphi(r)$, then there exists a ring homomorphism $\psi : A \to F$ satisfying $\psi|_R = \varphi$ and $0 \neq \psi(a)$.*

Proof. By induction it suffices to assume that $A = F[u]$.

Case 1. The element u is transcendental over R.

Let $a = \sum_{i=0}^{n} a_i u^i$ in $F[u]$ with a_0 nonzero. We show $r = a_0$ works. So suppose that $\varphi : R \to F$ is a ring homomorphism with $\varphi(a_0) \neq 0$. Since F is infinite, there exists an element $x \in F$ such that $\sum_{i=0}^{n} \varphi(a_i)x^i \neq 0$. Let $\psi : A \to F$ extend φ by setting $\psi(u) = x$. This works.

Case 2. The element u is algebraic over R.

As R is a domain and a nonzero, a^{-1} is also algebraic over the quotient field of R, so there exist equations:

$$a_0u^n + a_1u^{n-1} + \cdots + a_n = 0$$
$$b_0a^{-m} + b_1a^{1-m} + \cdots + b_m = 0$$

for appropriate a_i and b_j in A with a_0 and b_0 nonzero. It follows that u and a^{-1} are integral over $R[(a_0b_0)^{-1}]$. Set $r = a_0b_o$ and suppose that $\varphi : R \to F$ is a ring homomorphism satisfying $\varphi(r) \neq 0$. Set $\mathfrak{p} = \ker \varphi$. Then $\mathfrak{p}$ survives in $R[r^{-1}]$ by choice, so there exists a valuation ring B in the quotient field of $R[r^{-1}]$ such that $\mathfrak{p}R[r^{-1}]$ survives in B. If $b \in B$, we can write $b = r_1/r_2$ with $r_1, r_2 \in A$ and $r_2 \notin \mathfrak{p}$ as $qf(A) = qf\big(R[r^{-1}]\big)$. Hence we have a ring homomorphism $\widetilde{\varphi} : B \to F$ given by $r_1/r_2 \mapsto \varphi(r_1)/\varphi(r_2)$. We show that $A \subset B$ and $\psi : A \to F$ with $\psi = \widetilde{\varphi}|_A$ work. Since the integral closure of $R[r^{-1}]$ lies in B, we have a^{-1} and u lie in B. Thus $A \subset B$ and $a \in B^\times$. It follows that a does not lie in the maximal ideal of B, so $\widetilde{\varphi}(a) \neq 0$. □

Corollary 80.24. Zariski's Lemma *Let K/F be an extension of fields. Suppose that K is a finitely generated F-algebra. Then K/F is a finite field extension.*

Proof. Let $\widetilde{F}$ be an algebraic closure of F. Then apply the theorem with $R = F$, $A = \widetilde{F}$, and $a = 1$. By the Theorem, if x in $\widetilde{F}$ is transcendental over F, the homomorphism $\widetilde{F} \to \widetilde{F}$ sending $x \to 0$ cannot occur, which is false. □

As a second application, we wish to show if A is an integrally closed domain, then so is $A[t]$. To do this we first need the following:

Theorem 80.25. (Krull) *Let A be a domain with quotient field F and*

$$\mathcal{S} = \{B \mid A \text{ is a subring of } B \text{ and } B \text{ is a valuation ring}\}.$$

The A is integrally closed if and only if $A = \bigcap_{\mathcal{S}} B$.

Proof. ($\Leftarrow$): As valuation rings are integrally closed,this follows by Exercise 80.27(5).

($\Rightarrow$): Certainly, $A = \bigcap_{\mathcal{S}} B$, so assume that there exists an element u in $(\bigcap_{\mathcal{S}} B) \setminus A$. Then u does not lie in $A_F = A$, hence $u \notin A[u^{-1}]$ by Exercise 80.27(1). Let $v = u^{-1}$, then $vA[v] < A$, as $v \notin A[v]^\times$. Hence there exists a valuation ring $B_0 \subset F$ such that $vA[v]$ survives in B_0. Since u lies in B for every $u \in \mathcal{S}$, we must have $v \in B_0^\times$. But this means that $vA[v]$ cannot survive in B_0, a contradiction □

Theorem 80.26. *Let A be an integrally closed domain. Then $A[t]$ is integrally closed.*

PROOF. We have

$$(*) \qquad \Big(\bigcap_{\substack{A \subset B \subset F \\ B \text{ a valuation ring}}} B \Big)[t] = \Big(\bigcap_{\substack{A \subset B \subset F \\ B \text{ a valuation ring}}} B[t] \Big).$$

(Why?) and by the previous theorem,

$$A = \Big(\bigcap_{\substack{A \subset B \subset F \\ B \text{ a valuation ring}}} B \Big).$$

As valuation rings are integrally closed by Exercise 80.27(5), we may assume that A is a valuation ring.

Let f be a nonzero element of $F(t)$ (the quotient field of $A[t]$) be integral over $A[t]$. We must show that f lies in $A[t]$. Since $F[t]$ is a UFD hence integrally closed, f lies in $F[t]$. Write $f = \frac{1}{a}g$ with $g \in A[t]$ and $a \in A$ nonzero. If $a \in A^\times$, we are done, so assume not. As A is a valuation ring, there exists a nonzero coefficient b of g dividing all the coefficients of g, so we may write $f = \frac{b}{a}h$ with $h \in A[t]$ and h has some coefficient 1. If $a \mid b$ we are done, so assume not. As A is a valuation ring, we must have $b \mid a$ in A, so $f = \frac{1}{c}h$ for some nonzero element c in A. As c is not a unit in A and A is local it lies in the maximal ideal $\mathfrak{m}$ of A. As f is integral over $A[t]$, there exists polynomials, $\alpha_1(t), \ldots, \alpha_n(t)$ in $A[t]$ for some n satisfying $f^n + \alpha_1(t)f^{n-1} + \cdots + \alpha_n(t) = 0$ in $F[t]$. Hence $h^n + c\alpha_1(t)h^{n-1} + \cdots + c^n\alpha_n(t) = 0$ in $A[t]$. Consequently, $h^n \equiv 0 \mod A[t]\mathfrak{m}$. But $A[t]/\mathfrak{m}A[t] = (A/\mathfrak{m})[t]$ (why?) is a domain, i.e., $A[t]\mathfrak{m}$ is a prime ideal in $A[t]$, so $h \in A[t]\mathfrak{m}$ $(= \mathfrak{m}[t])$. This implies that every coefficient of h lies in $\mathfrak{m}$, a contradiction.

□

Exercises 80.27. 1. Let A be a commutative ring and u a unit in A. If R is a subring of A, show that u^{-1} is integral over R if and only if u^{-1} lies in $R[u]$.

2. Let A be a commutative ring. Then A is a domain if and only if $A_\mathfrak{m}$ is a domain for all maximal ideals $\mathfrak{m}$ of A

3. Show 80.7.

4. Establish Proposition 80.10

5. Prove that the intersection of integrally closed domains each of which lies in a common field is integrally closed.

CHAPTER XVI

Introduction to Commutative Algebra

In this chapter, all rings will be commutative. We shall be particularly interested in developing an algebraic dimension theory that coincides with the intuitive notion of the number of variables needed to describe a geometric object. In addition, we shall establish a generalization of unique factorization into prime ideals as happens in Dedekind domains to arbitrary Noetherian rings.

81. Zariski Topology

Throughout this section, R will denote commutative ring.

For a nonzero commutative ring R, we introduced in §38, the Spectrum of R,

$$\operatorname{Spec}(R) := \{\mathfrak{p} \mid \mathfrak{p} < R \text{ a prime ideal}\},$$

and defined the Zariski topology on it as follows:
If T is a subset of R, define

$$V_R(T) := \{\mathfrak{p} \mid \mathfrak{p} \in \operatorname{Spec}(R) \text{ with } T \subset \mathfrak{p}\}$$

called the *(abstract) variety* of T. We have

Lemma 81.1. *Let R be a commutative ring. Then*

(1) If T is a subset of R, then $V_R(T) = V_R(\langle T\rangle)$.
(2) If $T_1 \subset T_2$ are subsets of R, then $V_R(T_1) \supset V_R(T_2)$.
(3) $V_R(\emptyset) = \operatorname{Spec}(R)$.
(4) $V_R(R) = \emptyset$.
(5) If $T_i, i \in I$, are subsets of R, then $V_R(\bigcup_I T_i) = \bigcap_I V_R(T_i)$.
(6) If $\mathfrak{A}$ and $\mathfrak{B}$ are ideals in R, then

$$V_R(\mathfrak{A}\mathfrak{B}) = V_R(\mathfrak{A} \cap \mathfrak{B}) = V_R(\mathfrak{A}) \cup V_R(\mathfrak{B}).$$

In particular, the collection

$$\mathcal{C} := \{V_R(T) \mid T \subset R\}$$

forms a system of closed sets for the Zariski topology on $\operatorname{Spec}(R)$.

PROOF. This essentially was Exercise 38.15(5).

As an example, we show (6).

Since $\mathfrak{AB} \subset \mathfrak{A} \cap \mathfrak{B}$, we have $V(\mathfrak{AB}) \supset V(\mathfrak{A} \cap \mathfrak{B})$ by (2).

If $\mathfrak{p} \in V(\mathfrak{AB})$, then $\mathfrak{AB} \subset \mathfrak{p}$, hence $\mathfrak{A} \subset \mathfrak{p}$ or $\mathfrak{B} \subset \mathfrak{p}$, equivalently, $\mathfrak{p} \in V(\mathfrak{A})$ or $\mathfrak{p} \in V(\mathfrak{B})$, i.e., $V(\mathfrak{AB}) \subset V_R(\mathfrak{A}) \cup V_R(\mathfrak{B})$.

Finally, if $\mathfrak{p} \in V(\mathfrak{A}) \cup V(\mathfrak{B})$, then $\mathfrak{A} \cap \mathfrak{B} \subset \mathfrak{p}$, so $V(\mathfrak{A}) \cup V(\mathfrak{B}) \subset V(\mathfrak{A} \cap \mathfrak{B})$. □

We shall always assume that $\operatorname{Spec}(R)$ is given the Zariski topology. If R is clear, we shall abbreviate $V_R(T)$ by $V(R)$ and if $\mathfrak{A} = (a_1, \ldots, a_n)$ is a finitely generated ideal in R, we shall write $V(a_1, \ldots, a_n)$ for $V(\mathfrak{A})$.

The Zariski topology is very coarse. In general, as we shall see, it is not a Hausdorff space, i.e., we cannot necessarily find disjoint open neighborhoods of distinct points. Moreover, we shall see that the set of *closed points* in $\operatorname{Spec}(R)$, i.e., those points $\mathfrak{p}$ in $\operatorname{Spec}(R)$ such that $\{\mathfrak{p}\}$ is a closed set, is precisely the set of maximal ideals, so points are usually not closed.

Lemma 81.2. *Let V be a subset of $\operatorname{Spec}(R)$ and $\overline{V}$ its closure in $\operatorname{Spec}(R)$. If $\mathfrak{A} = \bigcap_{\mathfrak{p} \in V} \mathfrak{p}$, then $\overline{V} = V(\mathfrak{A})$.*

PROOF. ($\subset$:) As $\mathfrak{A} = \bigcap_V \mathfrak{p} \subset \mathfrak{P}$ for all $\mathfrak{P} \in V$, we have $V \subset V(\mathfrak{A})$, hence $\overline{V} \subset V(\mathfrak{A})$.
($\supset$:) Let $\overline{V} = V(\mathfrak{B})$, with $\mathfrak{B}$ an ideal in R. If $\mathfrak{p} \in V \subset \overline{V} = V(\mathfrak{B})$, then $\mathfrak{B} \subset \mathfrak{p}$, hence $\mathfrak{B} \subset \bigcap_V \mathfrak{p} = \mathfrak{A}$. Therefore, $\overline{V} = V(\mathfrak{B}) \supset V(\mathfrak{A})$ by Lemma 81.1(2). □

Corollary 81.3. *Let R be a commutative ring. Then*

(1) $\operatorname{Max}(R) \subset \operatorname{Spec}(R)$ is the set of closed points in $\operatorname{Spec}(R)$.
(2) Let $\mathfrak{p}$ and $\mathfrak{P}$ be prime ideals. Then $\mathfrak{P} \in \overline{\{\mathfrak{p}\}}$, the closure of $\{\mathfrak{p}\}$ in $\operatorname{Spec}(R)$, if and only if $\mathfrak{p} \subset \mathfrak{P}$.
(3) $\overline{\{\mathfrak{p}\}} = V(\mathfrak{p})$ for all prime ideals in R.
(4) $\operatorname{Spec}(R)$ is a T_0-topological space, i.e., if $\mathfrak{p}_1$ and $\mathfrak{p}_2$ are two distinct prime ideals in R, then there exists an open set in $\operatorname{Spec}(R)$ containing one of $\mathfrak{p}_1$, $\mathfrak{p}_2$, but not the other.

PROOF. (3) follows from the lemma and (2) follows from (3).

(1): Let $\mathfrak{p} \in \operatorname{Spec}(R)$. Then we have $\mathfrak{p}$ is a closed point if and only if $\overline{\mathfrak{p}} = \{\mathfrak{q} \in \operatorname{Spec}(R) \mid \mathfrak{p} \subset \mathfrak{q}\} = \{\mathfrak{p}\}$ if and only if $\mathfrak{p} \in \operatorname{Max}(R)$.

(4): By (2), if $\mathfrak{p}_1 \not\subset \mathfrak{p}_2$ are prime ideals, then $\mathfrak{p}_2$ lies in the open set $\operatorname{Spec}(R) \setminus V(\mathfrak{p}_1)$. □

Definition 81.4. An ideal $\mathfrak{A} \subset R$ is called a *radical ideal* if $\mathfrak{A} = R$ or $\mathfrak{A} < R$ and

$$\mathfrak{A} = \sqrt{\mathfrak{A}} := \{x \in R \mid x^n \in \mathfrak{A} \text{ some } n \in \mathbb{Z}^+\} = \bigcap_{V(\mathfrak{A})} \mathfrak{p} = \bigcap_{\mathfrak{A} \subset \mathfrak{p} \in \operatorname{Spec}(R)} \mathfrak{p}.$$

Equivalently, if a in R satisfies $a^n \in \mathfrak{A}$, then $a \in \mathfrak{A}$.

Examples 81.5. The following are radical ideals:

(1) Prime ideals.

(2) The nilradical, $\operatorname{nil}(R)$, of R as $\operatorname{nil}(R) = \bigcap_{V(0)} \mathfrak{p} = \sqrt{(0)}$.

Corollary 81.6. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be ideals in R, then $V(\mathfrak{A}) = V(\sqrt{\mathfrak{A}})$ and*

$$V(\mathfrak{A}) \subset V(\mathfrak{B}) \text{ if and only if } \sqrt{\mathfrak{A}} \supset \sqrt{\mathfrak{B}}.$$

Let

$$\mathcal{R} := \{\mathfrak{A} \subset R \mid \mathfrak{A} \text{ is a radical ideal}\}$$
$$\mathcal{V} := \{V(\mathfrak{A}) \mid \mathfrak{A} \text{ is an ideal of } R\}.$$

Then $V : \mathcal{R} \to \mathcal{V}$ given by $\mathfrak{A} \mapsto V(\mathfrak{A})$ is an order-reversing bijection with inverse $\mathcal{I} : \mathcal{V} \to \mathcal{R}$ given by $V \mapsto \bigcap_V \mathfrak{p}$.

Proof. Clearly, $V(\mathfrak{A}) = V(\sqrt{\mathfrak{A}})$ and $V(\mathfrak{A}) \subset V(\mathfrak{B})$ implies $\sqrt{\mathfrak{B}} \subset \bigcap_{V(\mathfrak{A})} \mathfrak{p} = \sqrt{\mathfrak{A}}$. □

We leave the following as an exercise:

Proposition 81.7. *Let r be an element of R and*

$$D(r) := \operatorname{Spec}(R) \setminus V(r) = \{\mathfrak{p} \in \operatorname{Spec}(R) \mid r \notin \mathfrak{p}\}.$$

Then

(1) $\{D(r) \mid r \in R\}$ is a base for the topology of $\operatorname{Spec}(R)$, i.e., every open set in $\operatorname{Spec}(R)$ is a union of such $D(r)$.

(2) $\operatorname{Spec}(R)$ is quasi-compact, i.e., every open cover of $\operatorname{Spec}(R)$ has a finite subcover.

(3) $\operatorname{Spec}(R)$ is a Hausdorff space if and only if $\operatorname{Spec}(R) = \operatorname{Max}(R)$.

The set $D(r)$ above is called a *basic open set* in $\operatorname{Spec}(R)$.

Definition 81.8. A nonempty topological space is called *irreducible* if whenever U, V are nonempty open subsets of X, then $U \cap V \neq \emptyset$. A subspace Y of a topological space X is called *irreducible* if it is irreducible in the induced topology.

Examples 81.9. 1. Points are irreducible.

2. A nonempty Hausdorff space is irreducible if and only if it consists of a single point.

Proposition 81.10. *Let X be a nonempty topological space. Then the following are equivalent:*

(1) X is irreducible.
(2) If $W_1, W_2 < X$ are closed, then $W_1 \cup W_2 < X$.
(3) If $\{W_1, \ldots, W_n\}$ is a finite closed cover of X, then $X = W_i$ for some i.
(4) If U is a nonempty open subset of X, then U is dense in X.
(5) Every open set in X is connected.

We leave the proof as an exercise. For varieties, the concept of irreducibility replaces connectivity.

Definition 81.11. If R is a nonzero commutative ring. Define

$$\operatorname{Min}(R) := \{\mathfrak{p} \in \operatorname{Spec}(R) \mid \text{there exists no prime ideal } \mathfrak{p}_0 < \mathfrak{p}\}$$

called the set of *minimal primes ideals* in R.

As the intersection of a chain of prime ideals is clearly a prime ideal, the set of minimal primes in a nonzero commutative ring is not empty by Zorn's Lemma.

Lemma 81.12. $\operatorname{Spec}(R)$ *is irreducible if and only if* $\operatorname{nil}(R)$ *is a prime ideal if and only if* $|\operatorname{Min}(R)| = 1$. *In particular, if R is a domain, then* $\operatorname{Spec}(R)$ *is irreducible.*

Proof. We know that $\operatorname{nil}(R) = \bigcap_{\operatorname{Spec}(R)} \mathfrak{p} = \bigcap_{\operatorname{Min}(R)} \mathfrak{p} \subset \mathfrak{P}$ for every prime ideal $\mathfrak{P}$ and $V(0) = V(\sqrt{0}) = V\big(\operatorname{nil}(R)\big) = \operatorname{Spec}(R)$.

Check 81.13. Let $\mathfrak{A}$ and $\mathfrak{B}$ be ideals in R, then

(i) $\mathfrak{A} \subset \operatorname{nil}(R)$ if and only if $\sqrt{\mathfrak{A}} \subset \operatorname{nil}(R)$.
(ii) $\sqrt{\mathfrak{AB}} \supset \sqrt{\mathfrak{A}}\sqrt{\mathfrak{B}}$.

Let $\mathfrak{A}$ and $\mathfrak{B}$ be ideals in R, then using the check, we have $|\operatorname{Min}(R)| = 1$ if and only if $\operatorname{nil}(R)$ is a prime ideal if and only if $\mathfrak{AB} \subset \operatorname{nil}(R)$ implies $\mathfrak{A} \subset \operatorname{nil}(R)$ or $\mathfrak{B} \subset \operatorname{nil}(R)$ if and only if $\sqrt{\mathfrak{A}}\sqrt{\mathfrak{B}} \subset \operatorname{nil}(R)$ implies $\sqrt{\mathfrak{A}} \subset \operatorname{nil}(R)$ or $\sqrt{\mathfrak{B}} \subset \operatorname{nil}(R)$ if and only if $V(\mathfrak{AB}) = V(\mathfrak{A}) \cup V(\mathfrak{B}) = \operatorname{Spec}(R)$ implies $V(\mathfrak{A}) = \operatorname{Spec}(R)$ or $V(\mathfrak{B}) = \operatorname{Spec}(R)$ if and only if $\operatorname{Spec}(R)$ is irreducible. □

Since $\operatorname{Spec}(R)$ is a topological space based on an algebraic structure, via the ring R, topological maps of spectra should arise from homomorphisms of the underlying algebraic structures. Indeed, we have:

Definition 81.14. Let $\varphi : A \to B$ be a ring homomorphism of commutative rings. Define the *associated map* to be the map arising from φ by

$$^a\varphi : \operatorname{Spec}(B) \to \operatorname{Spec}(A) \text{ via } {}^a\varphi(\mathfrak{P}) = \varphi^{-1}(\mathfrak{P}).$$

That this map makes sense follows by:

Lemma 81.15. *Let $\varphi : A \to B$ be a ring homomorphism, then ${}^a\varphi(\mathfrak{B})$ is a prime ideal of A for every prime ideal $\mathfrak{P}$ in B, i.e.,*

$$^a\varphi : \operatorname{Spec}(B) \to \operatorname{Spec}(A)$$

is defined. Moreover, if $T \subset A$ is a subset, then

$$({}^a\varphi)^{-1}\big(V(T)\big) = V\big(\varphi(T)\big).$$

In particular, ${}^a\varphi : \operatorname{Spec}(B) \to \operatorname{Spec}(A)$ is a continuous map.

Proof. If $\mathfrak{P} < B$ is a prime ideal, then the ideal $\varphi^{-1}(\mathfrak{P}) < A$ is an ideal and φ induces a monomorphism $A/\varphi^{-1}(\mathfrak{P}) \to B/\mathfrak{P}$ with $B/\mathfrak{B}$ a domain. Consequently, $A/\varphi^{-1}(\mathfrak{P})$ is also a domain, i.e., $\varphi^{-1}(\mathfrak{P})$ is a prime ideal.

We also have

$$\begin{aligned}({}^a\varphi)^{-1}\big(V(T)\big) &= \{\mathfrak{P} \in \operatorname{Spec}(B \mid \varphi^{-1}(\mathfrak{P}) = {}^a\varphi(\mathfrak{P}) \supset T\} \\ &= \{\mathfrak{P} \in \operatorname{Spec}(B) \mid \mathfrak{P} \supset \varphi(T)\} = V\big(\varphi(T)\big). \qquad \square\end{aligned}$$

It is easy to check:

Remark 81.16. If $\varphi : A \to B$ and $\psi : B \to C$ are ring homomorphisms of commutative rings, then ${}^a(\psi \circ \varphi) = {}^a\varphi \circ {}^a\psi$.

Definition 81.17. A map $f : X \to Y$ or topological spaces is called *dominant* if $\operatorname{im} f$ is dense in Y.

The relationship between φ and ${}^a\varphi$ is illustrated by:

Lemma 81.18. *Let $\varphi : A \to B$ be a ring homomorphism.*

1. *If φ is an epimorphism, then ${}^a\varphi : \operatorname{Spec}(B) \to \operatorname{Spec}(A)$ is an injective closed map with image $V(\ker \varphi)$. In particular, ${}^a\varphi$ induces a homeomorphism $V(\ker \varphi) \cong \operatorname{Spec}(B) \cong \operatorname{Spec}(A/\ker \varphi)$.*
2. *If $\mathfrak{A}$ is an ideal, then $V(\mathfrak{A})$ is homeomorphic to $\operatorname{Spec}(R/\mathfrak{A})$.*
3. *If φ is a monomorphism, then ${}^a\varphi : \operatorname{Spec}(B) \to \operatorname{Spec}(A)$ is a dominant map. [It may not be onto.]*

Proof. (1) follows from the Correspondence Principle and (1) implies (2).

(3): Suppose that $a \in A$ satisfies $D(a) := \operatorname{Spec}(A) \setminus V(a)$ is non-empty. We must show that $D(a) \cap \operatorname{im} {}^a\varphi$ is also non-empty. Suppose that this is false, then $\operatorname{im} {}^a\varphi \subset V(a)$, so

$$a \in \bigcap_{\operatorname{Spec}(B)} {}^a\varphi(\mathfrak{P}) = \bigcap_{\operatorname{Spec}(B)} \varphi^{-1}(\mathfrak{P}),$$

hence $\varphi(a) \subset \bigcap_{\operatorname{Spec}(B)} \mathfrak{P} = \operatorname{nil}(B)$, i.e., $\varphi(a)$ is nilpotent in B. In particular, there exists a positive integer n with $0 = \varphi(a)^n = \varphi(a^n)$. As φ is monic, $a^n = 0$ in A, so $a \in \operatorname{nil}(A) \subset \mathfrak{p}$ for every prime ideal $\mathfrak{p}$. It follows that $D(a) = \emptyset$, a contradiction. □

Note: The inclusion $i : \mathbb{Z} \to \mathbb{Z}_{(p)}$, with $\mathbb{Z}_{(p)} = \{\frac{a}{b} \mid a, b \in \mathbb{Z},\, p \not| \, b\}$, the localization of $\mathbb{Z}$ at the prime ideal (p), is injective, but ${}^a i$ is not surjective as $\operatorname{Spec}(\mathbb{Z}_{(p)}) = \{0, p\mathbb{Z}_{(p)}\}$ (but does have dense image as shown).

Localization will play a decisive role in this chapter as well as being a primary tool in commutative algebra. It is, therefore, useful to coalesce some of the properties of localization, many of which appeared in §26 (some as exercises) and subsequent sections, together with some new observations, leaving all the proofs as useful exercises.

Remark 81.19. Let R be a nonzero commutative ring and S a multiplicative set in R. We have seen that we have a canonical ring homomorphism $\varphi_R : R \to S^{-1}R$ given by $r \mapsto r/1$. This makes $S^{-1}R$ into a commutative R-algebra. If $\mathfrak{A}$ in R is an ideal, then

$$S^{-1}\mathfrak{A} := \{\frac{a}{s} \mid a \in A,\, s \in S\} \subset S^{-1}R$$

is an ideal. Let $\mathfrak{B}$ be another ideal in R. The following are true:

(*i*) $S^{-1}\mathfrak{A} = S^{-1}R$ if and only if $\mathfrak{A} \cap S \neq \emptyset$.

(*ii*) $\mathfrak{A} \subset \varphi_R^{-1}(S^{-1}\mathfrak{A}) \subset R$ is an ideal.

(*iii*) If $\mathfrak{D}$ is an ideal in $S^{-1}R$, then there exists an ideal $\mathfrak{C}$ in R satisfying $S^{-1}\mathfrak{C} = \mathfrak{D}$. Moreover, $\mathfrak{D} = S^{-1}\big((\varphi_R)^{-1}(\mathfrak{D})\big)$.

(*iv*) If $\mathfrak{A} < \mathfrak{B}$ and $\mathfrak{A} \cap S = \emptyset$, then $S^{-1}\mathfrak{A} < S^{-1}\mathfrak{B}$.

(*v*) $S^{-1}(\mathfrak{A} + \mathfrak{B}) = S^{-1}\mathfrak{A} + S^{-1}\mathfrak{B}$.

(*vi*) $S^{-1}(\mathfrak{A} \cap \mathfrak{B}) = S^{-1}\mathfrak{A} \cap S^{-1}\mathfrak{B}$.

(*vii*) $S^{-1}(\mathfrak{A}\mathfrak{B}) = S^{-1}\mathfrak{A}\, S^{-1}\mathfrak{B}$.

(*viii*) The map

$$\{\mathfrak{p} \in \operatorname{Spec} R \mid \mathfrak{p} \cap S = \emptyset\} \to \operatorname{Spec}(S^{-1}R) \text{ given by } \mathfrak{p} \mapsto S^{-1}\mathfrak{p}$$

is a bijection and, in fact, a homeomorphism.

As mentioned before, the most important examples of localization of a commutative ring R are:

(*i*) $R_{\mathfrak{p}}$ the localization of R at the multiplicative set $R \setminus \mathfrak{p}$ where $\mathfrak{p}$ is a prime ideal of R, called the localization at $\mathfrak{p}$.

(*ii*) R_a, the localization of R at the multiplicative set $\{a^n \mid n \geq 0\}$ called the localization at a, i.e., at "the open neighborhood $D(a)$ of a".

We return to the topology that interests us. We have seen that irreducible topological spaces are applicable to our theory. However, connected subspaces are not useful, so we need a substitute for the usual decomposition of a space into connected components into a decomposition of another type. This will be a decomposition into irreducible components.

Lemma 81.20. *Let X be a topological space and Y a subset of X.*

(1) Y is irreducible if and only if $\overline{Y}$ in X is irreducible.

(2) Every irreducible subspace of X is contained in some maximal irreducible subspace. Any such is closed.

PROOF. (1): Let $U \subset X$ be an open subset, then $Y \subset \overline{U}$ if and only if $\overline{Y} \subset \overline{U}$. By definition, the closure of $U \cap \overline{Y}$ in $\overline{Y}$ is $\overline{U} \cap \overline{Y}$, so $U \cap Y$ is dense in Y if and only if $U \cap \overline{Y}$ is dense in $\overline{Y}$.

(2) follows from Zorn's Lemma. □

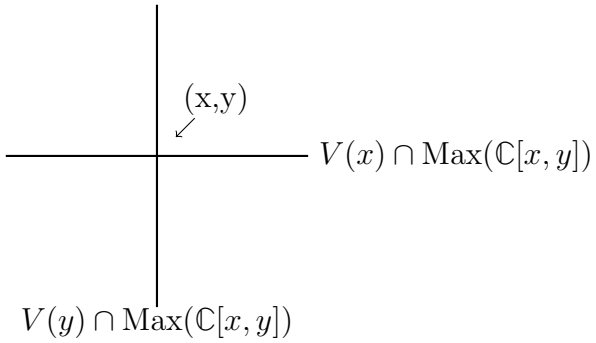
A maximal irreducible subspace of a nonempty topological space X is called an *irreducible component* of X. If X is a Hausdorff space, then $\{x\}$, $x \in X$, are the irreducible components of X, so this is not a useful concept in that case.

Definition 81.21. Let X be a topological space. Then a point $x \in X$ is called a *generic point* of X if $X = \overline{\{x\}}$.

Corollary 81.22. *$\{V(\mathfrak{p}) \mid \mathfrak{p} \in \operatorname{Min}(R)\}$ is the set of irreducible components of $\operatorname{Spec}(R)$ and $\mathfrak{p} \in \operatorname{Spec}(R)$ is a generic point of $V(\mathfrak{p})$.*

Remark 81.23. Irreducible components may intersect non-trivially. For example, let x, y be indeterminants, then $V(xy)$ in $\operatorname{Spec}(\mathbb{C}[x,y])$ has irreducible components $V(x)$ and $V(y)$ and they intersect in (x,y), a maximal ideal in $\mathbb{C}[x,y]$. Topologically, $V(xy) \cong \operatorname{Spec}\big(\mathbb{C}[x,y]/(xy)\big)$. Identifying $\operatorname{Max}(\mathbb{C}[x,y])$ with $\mathbb{C}^2$ by $(x-a, y-b) \mapsto (x,y)$ using the Hilbert Nullstellensatz, we can view the y-axis as the closed points in $V(x)$ and the x-axis as the closed points in $V(y)$. The origin corresponds to the intersection of $V(x)$ and $V(y)$. The following picture illustrates, this where we have identified $V(x) \cap \operatorname{Max}(\mathbb{C}[x,y])$ and the

y-axis, $V(y) \cap \operatorname{Max}(\mathbb{C}[x,y])$ and the x-axis) and (x,y) with the origin, using the Hilbert Nullstellensatz.



Definition 81.24. A topological space X is called *Noetherian* if the collection of open sets in X satisfies the ascending chain condition. Equivalently, the collection of closed sets in X satisfies the descending chain condition.

Example 81.25. If R is a commutative Noetherian ring, then $\operatorname{Spec}(R)$ is a Noetherian topological space, but the converse is false. For example, let F be a field, $A = F[t_1, \dots, t_n, \dots]$, and $\mathfrak{A} = (t_1, t_2^2, \dots, t_n^n, \dots)$. Then the ring $B = A/\mathfrak{A}$ is not Noetherian, as $\operatorname{nil}(B)$ is not nilpotent, but $\operatorname{Spec}(B)$ is a Noetherian space. In fact, $|\operatorname{Spec}(B)| = 1$.

Proposition 81.26. *Let X be a Noetherian space. Then there exist finitely many irreducible components of X and X is their union.*

PROOF. This proof is a typical proof when we have a Noetherian condition. Let

$$\mathcal{F}(X) := \{V \mid V \subset X \text{ closed}\}$$
$$\mathcal{C}(X) := \{C \mid C \subset \mathcal{F}(X), \quad C \text{ is a finite union of closed irreducible sets in } X\}.$$

Claim. $\mathcal{C}(X) = \mathcal{F}(X)$:

If this is false, then by the Minimal Principle (equivalent to the descending chain condition), there exists an element $Y \in \mathcal{F}(X) \setminus \mathcal{C}(X)$ that is minimal. As, $Y \notin \mathcal{C}(X)$, it is not irreducible, $Y = Y_1 \cup Y_2$ with $Y_i < Y$ and $Y_i \in \mathcal{F}(X)$, for $i = 1, 2$. By minimality, $Y_i \in \mathcal{C}(X)$, hence $Y = Y_1 \cup Y_2$ in $\mathcal{C}(X)$. This contradiction establishes the claim.

By the claim, we have $X = \bigcup_{i=1}^n C_i$ for some $C_i \subset X$ closed irreducible. Clearly, we may assume that C_j is not a subset of C_i for $i \neq j$. If C is an irreducible component of X, then we have $C \subset \bigcup_{i=1}^n (C \cap C_i)$ is irreducible. It follows that $C \subset C_i$ for some i, hence $C = C_i$ for

some i as C is a maximal closed irreducible subset of X. The result follows. □

Corollary 81.27. *Let R be a commutative Noetherian ring. Then* $\mathrm{Min}(R)$ *is a finite set and* $\mathrm{Spec}(R) = \bigcup_{\mathrm{Min}(R)} V(\mathfrak{p})$.

Our primary interest in this chapter will be the Krull dimension of a ring that we shall just call the *dimension of a ring*. For the convenience of the reader, we define it again, in an equivalent formulation, that will be of considerable interest for us.

Definition 81.28. Let R be a nonzero commutative ring and $\mathfrak{P}$ a prime ideal in R. We define the *height* of $\mathfrak{P}$ by

$$\mathrm{ht}\,\mathfrak{P} := \sup\{n \mid \mathfrak{p}_0 < \cdots < \mathfrak{p}_n = \mathfrak{P}, \text{ with } \mathfrak{p}_i \in \mathrm{Spec}(R),\ i = 1, \ldots, n\}$$

if finite and $= \infty$ if not, i.e., the maximal number of (proper) links of chains of prime ideal in $\mathfrak{P}$. If $\mathfrak{A} < R$ is an ideal, we define the *height* of $\mathfrak{A}$ to be

$$\mathrm{ht}\,\mathfrak{A} := \inf\{\mathrm{ht}\,\mathfrak{p} \mid \mathfrak{p} \in V(\mathfrak{A})\} = \inf\{\mathrm{ht}\,\mathfrak{p} \mid \mathfrak{p} \in V(\mathfrak{A}) \text{ minimal}\}.$$

Then the *(Krull) dimension* of R is defined to be

$$\dim R := \sup\{\mathrm{ht}\,\mathfrak{p} \mid \mathfrak{p} \in \mathrm{Spec}(R)\} = \sup\{\mathrm{ht}\,\mathfrak{m} \mid \mathfrak{m} \in \mathrm{Max}(R)\}$$

if finite and $= \infty$ otherwise.

If $\mathfrak{A}$ is an ideal, we define the *dimension* of $V(\mathfrak{A})$ to be

$$\dim V(\mathfrak{A}) := \dim R/\mathfrak{A}.$$

More generally, if X is a topological space, the *combinatorial dimension* of X is defined to be

$$\dim X := \sup\{n \mid Y_0 < \cdots < Y_n \text{ each } Y_i \subset X \text{ closed and irreducible}\}$$

if finite and $= \infty$ if not. In particular, if $\mathfrak{A} \subset R$ is an ideal, then

$$\dim R = \dim \mathrm{Spec}(R) \text{ and } \dim V(\mathfrak{A}) = \dim \mathrm{Spec}(R/\mathfrak{A}).$$

Note that this is consistent, since $V(\mathfrak{A})$ is homeomorphic to $\mathrm{Spec}(R/\mathfrak{A})$.

Remarks 81.29. Let R be a nonzero commutative ring and F a field.

1. If R is a domain, then $\dim R = 0$ if and only if R is a field.
2. $\dim R = 0$ if and only if $\mathrm{Spec}(R) = \mathrm{Max}(R)$.
3. Let R be a PID but not a field, then $\mathrm{Spec}(R) = \{(0)\} \cup \mathrm{Max}(R)$ with (0) not maximal, so $\dim R = 1$, e.g., $R = \mathbb{Z}$ or $F[t]$. More generally, any Dedekind domain is of dimension one.

4. $\dim F[t_1, \dots, t_n] \geq n$, since
$$0 < (t_1) < (t_1, t_2) < \cdots < (t_1, \dots, t_n)$$
is a (proper) chain of prime ideals. In fact, $\dim F[t_1, \dots, t_n] = n$, a result that we shall show later.
5. $\dim F[t_1, \dots, t_n, \dots] = \infty$, but there do exist Noetherian rings of infinite dimension.
6. Recall a *local ring* is a commutative ring with a unique maximal ideal. [Such a ring can still have $\operatorname{Spec}(R)$ infinite, and in fact, this is the case if $\dim R > 1$ and R is Noetherian.] If R is a local ring with maximal ideal $\mathfrak{m}$, we say $(R, \mathfrak{m})$ is a local ring. We shall see that a local Noetherian ring has finite dimension.
7. Maximal ideals need not have the same height. For example, let $\mathfrak{p} = (3)$ in $\operatorname{Spec}(\mathbb{Z})$ and $\mathbb{Z}_{\mathfrak{p}}$ the localization of $\mathbb{Z}$ at $\mathfrak{p}$. Then $\mathbb{Z}_{\mathfrak{p}}[t]$ is a Noetherian UFD of dimension two containing maximal ideals $(3, t)$ and $(3t - 1)$, maximal ideals of height one and two, respectively.
8. Suppose that $\mathfrak{A} < R$ is an ideal, then we have:
$$\operatorname{ht} \mathfrak{A} + \dim R/\mathfrak{A} \leq \dim R,$$
i.e.,
$$\operatorname{ht} \mathfrak{A} + \dim V(\mathfrak{A}) \leq \dim \operatorname{Spec}(R).$$
In general, inequality is possible. If we have equality, then we should view $\operatorname{ht} \mathfrak{A}$ as $\operatorname{codim} V(\mathfrak{A}) = \operatorname{codim}_{\operatorname{Spec}(R)} V(\mathfrak{A})$. One of our primary goals of this chapter is to establish this in the special case when R is an affine F-algebra, i.e., a finitely generated commutative F-algebra, and $\mathfrak{A}$ is a prime ideal.
9. If $\mathfrak{p}$ is a prime ideal in R, let $\mathfrak{p}R_{\mathfrak{p}} = \mathfrak{p}_{\mathfrak{p}}$, the localization of $\mathfrak{p}$ at $R \setminus \mathfrak{p}$. Then $(R_{\mathfrak{p}}, \mathfrak{p}_{\mathfrak{p}})$ is a local ring and satisfies $\dim R_{\mathfrak{p}} = \operatorname{ht} \mathfrak{p}R_{\mathfrak{p}}$.

Remark 81.30. We can also generalize the notions above to modules. Let M and N be R-modules. Then
$$(N : M) := \{x \in R \mid xM \subset N\}$$
is a ideal in R. E.g., $(0 : M) = \operatorname{ann}_R M$. We define the *dimension* of M by
$$\dim M := \dim V\big((0 : M)\big) = \dim V\big(\operatorname{ann}_R(M)\big)$$
and the *support* of M by
$$\operatorname{Supp}(M) := \{\mathfrak{p} \in \operatorname{Spec}(R) \mid M_{\mathfrak{p}} \neq 0\},$$
where $M_{\mathfrak{p}} = \{\frac{m}{s} \mid m \in M, s \in R \setminus \mathfrak{p}\}$, the localization of M at $\mathfrak{p}$, an $R_{\mathfrak{p}}$-module. (We leave the proof that we can localize an R-module to an $S^{-1}R$-module with S a multiplicative set in R as an exercise.)

Exercises 81.31. 1. Prove Proposition 81.7.

2. Prove Proposition 81.10.
3. Show Check 81.13 is true.
4. Check Remark 81.16.
5. Prove the assertions in Remark 81.19.
6. Let R be a nonzero commutative ring, S a multiplicative set in R, and M an R-module. Show that $S^{-1}M := \{\frac{r}{s} \mid r \in R,\, s \in S\}$ with the obvious definition is an $S^{-1}R$-module called the *localization* of M at S. If
$$0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$$
is an exact sequence of R-modules, prove that the sequence
$$0 \to S^{-1}M' \xrightarrow{S^{-1}f} S^{-1}M \xrightarrow{S^{-1}g} S^{-1}M'' \to 0$$
of $S^{-1}R$-modules is exact where, e.g., $S^{-1}f : S^{-1}M' \to S^{-1}M$, is defined by $S^{-1}f(\frac{r}{s}m) = \frac{r}{s}f(m)$, i.e., the $S^{-1}R$-module homomorphism induced by f.
7. In the previous exercise, if $S = R \setminus \mathfrak{p}$ with $\mathfrak{p}$ a prime ideal in R, and $f : M \to N$ is an R-homomorphism, let $f_{\mathfrak{p}}$ denote $S^{-1}f$. Then show $\ker(f_{\mathfrak{p}}) = (\ker f)_{\mathfrak{p}}$ and $\operatorname{coker}(f_{\mathfrak{p}}) = (\operatorname{coker} f)_{\mathfrak{p}}$ for all prime ideals $\mathfrak{p}$ in R. Moreover, show that the following are equivalent:
   (*i*) $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is an exact sequence of R-modules.
   (*ii*) $0 \to M'_{\mathfrak{p}} \xrightarrow{f_{\mathfrak{p}}} M_{\mathfrak{p}} \xrightarrow{g_{\mathfrak{p}}} M''_{\mathfrak{p}} \to 0$ is an exact sequence of $R_{\mathfrak{p}}$-modules for all prime ideals $\mathfrak{p}$.
   (*iii*) $0 \to M'_{\mathfrak{m}} \xrightarrow{f_{\mathfrak{m}}} M_{\mathfrak{m}} \xrightarrow{g_{\mathfrak{m}}} M''_{\mathfrak{m}} \to 0$ is an exact sequence of $R_{\mathfrak{m}}$-modules for all maximal ideals $\mathfrak{m}$.

   In particular, f is an R-monomorphism (respectively, R-epimorphism) if and only if $f_{\mathfrak{p}}$ is an $R_{\mathfrak{p}}$-monomorphism (respectively $R_{\mathfrak{p}}$-epimorphism) for all prime ideals $\mathfrak{p}$ in R if and only if $f_{\mathfrak{p}}$ is an $R_{\mathfrak{m}}$-monomorphism (respectively $R_{\mathfrak{m}}$-epimorphism) for all maximal ideals $\mathfrak{m}$ in R
8. Let R be a nonzero commutative ring and $a_1, \ldots, a_n \in R$. Show $D(a_1 \cdots a_n) = \cap_{i=1}^{n} D(a_i)$ and every open set in $\operatorname{Spec}(R)$ is a union of basic open sets.
9. Show the assertions in Example 81.25 are true.
10. Let X be a topological space. Show that X is Noetherian if and only if every open subset of X is quasi-compact.
11. For $f \in R$, let $X = \operatorname{Spec}(R)$ and $X_f = \operatorname{Spec}(R_f)$, where R_f the localization of R at the basic open set $D(f)$. Write $R(X_f)$ for the localization R_f.

For $f, g, h \in R$ and $U = X_f,\ U' = X_g,\ U'' = X_h,$ show the following:

(a) $R(U)$ depends only on U and not on f.

(b) If $U' \subset U$, then there exists a positive integer n and an element $x \in R$ such that $g^n = xf$. Using this, defines a homomorphism $\rho_{UU'} : R(U) \to R(U')$ via $a/f^m \mapsto ax^m/g^{mn}$. This map depends only on U and U' and is called the *restriction homomorphism.*

(c) If $U = U'$, then ρ_{UU} is the identity map.

(d) If $U'' \subset U' \subset U$, then the diagram

$$\begin{array}{ccc} R(U) & \xrightarrow{\quad\rho_{UU''}\quad} & R(U'') \\ & {\scriptstyle\rho_{UU'}}\searrow \quad \nearrow{\scriptstyle\rho_{U'U''}} & \\ & R(U') & \end{array}$$

commutes. This implies that $X := (X, R)$ is a *presheaf of rings.*

(e) Let $\{U_i \mid i \in I\}$ be a finite covering of a basic open set U in X by basic open sets. Suppose that $a \in U$, $a_i \in U_i$ for each i. Then

(i) If $a \in R(U)$ satisfies $\rho_{UU_i}(a) = 0$ for all i, then $a = 0$.

(ii) If $\rho_{U_i(U_i \cap U_j)}(a_i) = \rho_{U_j(U_i \cap U_j)}(a_j)$ in $R(U_i \cap U_j)$ for all i, j, then there exists an element $r \in R(U)$ satisfying $\rho_{UU_i}(r) = a_i$ for all i.

This implies that X is a *sheaf of rings.* We call (X, R) an *affine scheme.*

[If $x = \mathfrak{p}$, then the local ring $R_{\mathfrak{p}}$ is the *stalk* of X at $\mathfrak{p}$. It is the set of *germs at* $\mathfrak{p}$ on X, i.e., equivalence classes of elements $\overline{r}$ where r is defined in some $R(U)$, U a basic open set, and if r' is defined in $R(U')$, then $r \sim r'$ if there exists a basic open neighborhood V of $\mathfrak{p}$ in $U \cap U'$ such that $\rho_{UV}(r) = \rho_{U'V}(r')$ in $R(V)$.]

82. Integral Extensions of Commutative Rings

As is true in this whole chapter, R is a commutative ring.

In this section, we look further at the notion of integral extensions. We use the material studied in Sections §72 and §73.

Notation 82.1. We begin with some new notation. If A is a commutative R-algebra via the ring homomorphism $\varphi_A : R \to A$, we shall write ra for $\varphi(r)a$, $r \in R$, $a \in A$. If $a_1, \ldots, a_n.$ are elements of A, then we shall write $R[a_1, \ldots, a_n]$ for $\varphi(R)[a_1, \ldots, a_n]$ (even if φ_R is not a monomorphism). If $S \subset R$ is a multiplicative set, we shall write $S^{-1}A$

for $\big(\varphi(S)\big)^{-1}A$. So if $r \in R$, $s \in S$, and $a \in A$, we write $\frac{r}{s}a$ for $\frac{\varphi_A(r)}{\varphi_A(s)}a$. If $\mathfrak{p}$ is a prime ideal in R, we write $A_{\mathfrak{p}}$ for $S^{-1}A$ where $S = R \setminus \mathfrak{p}$.

Definition 82.2. Let $\varphi : R \to A$ be a ring homomorphism of commutative rings, we shall view A as an R-algebra via φ unless otherwise stated. We say

(1) φ is of *finite type* if A is a finitely generated R-algebra.
(2) φ is *finite* if A is a finitely generated R-module (via φ). [One also calls a finitely generated R-module a *finite R-module*.]
(3) φ is *integral* if $A/\varphi(R)$ is an integral extension.

We leave the various remarks about ring homomorphisms of commutative rings and localization as useful exercises for the reader.

Remarks 82.3. Let $\varphi : R \to A$ be a ring homomorphism of commutative rings.

1. φ is a finite map if and only if it is integral and finite type. (Cf. Exercise 72.15(4).)
2. If φ is an epimorphism, then φ is finite, hence integral.
3. If φ is integral and $\mathfrak{A}$ an ideal in A, then the induced map $\widetilde{\varphi} : R/\varphi^{-1}(\mathfrak{A}) \to A/\mathfrak{A}$ is integral and injective.
4. The composition of ring homormorphisms of finite type (respectively, finite, integral) are of finite type (respectively, finite, integral).
5. Suppose that φ is the inclusion and $\mathfrak{A} < R$ an ideal. As before, we write A/R. It is possible that $\mathfrak{A}A = A$ and $\mathfrak{A} < (\mathfrak{A}A) \cap R$. For example, (2) is a prime ideal in $\mathbb{Z}$ but $(2)\mathbb{Q} = \mathbb{Q}$. We shall see that this does not happen if A/R is integral.
6. Suppose that φ is the inclusion with R a domain and $K = A$ a field containing R. As in §72, we shall let R_K denote the integral closure of R in K. In this chapter, if R is integrally closed in its quotient field F, i.e., $R = R_F$, we shall use geometric language and say that R is a *normal domain* instead of saying R is integrally closed.

Remarks 82.4. Let $\varphi : R \to A$ be a ring homomorphism of commutative rings and S a multiplicative subset of R.

1. If φ is a monomorphism, so is $\varphi_{S^{-1}R} : S^{-1}R \to S^{-1}A$. In particular, this applies if φ is the inclusion map. We then view $\varphi_{S^{-1}R}$ as the inclusion. If, in addition, $A = K$ is a field, $\mathfrak{p}$ a prime ideal in R, then $(R_{\mathfrak{p}})_K = (R_K)_{\mathfrak{p}}$.

2. If φ is is of finite type (respectively, finite, integral), then so is $\varphi_{S^{-1}R}$, in which case the integral closure of $S^{-1}R$ in $S^{-1}A$ is $S^{-1}C$ where C is the integral closure of R in A. If $\mathfrak{p}$ is a prime ideal in R, then $(R_{\mathfrak{p}})_K = (R_K)_{\mathfrak{p}}$.

In the more elementary part of this tome, we proved the Hilbert Nullstellsatz. We wish to give deeper insight into why this theorem (rather than the proof we gave) is true. One of the keys to the proof, essentially an equivalent formulation, was Zariski's Lemma 38.10 whose proof we shall derive again later. To do so we must look carefully at integral extensions. We begin with a summary and the methodology that we shall use.

Summary 82.5. Let $i : A \subset B$ be the inclusion of two commutative rings and $S \subset A$ a multiplicative set.

1. $S^{-1}i : S^{-1}A \to S^{-1}B$ is a ring monomorphism, that we view as an inclusion.
2. Let $S^{-1}\varphi_A : A \to S^{-1}A$ by $a \to \frac{a}{1}$, be the canonical homomorphism. Then ${}^a\varphi_A : \operatorname{Spec} S^{-1}A \to \{\mathfrak{p} \in \operatorname{Spec}(A) \mid \mathfrak{p} \cap S = \emptyset\}$ is a homeomorphism. In particular, if $\mathfrak{p}$ is a prime ideal in A such that $\mathfrak{p} \cap S = \emptyset$, then $\mathfrak{p} = (S^{-1}\mathfrak{p}) \cap A := {}^a\varphi_A(S^{-1}\mathfrak{p})$.
3. If $\mathfrak{p}$ is a prime ideal in A, then $(A_{\mathfrak{p}}, \mathfrak{p}A_{\mathfrak{p}})$ is a local ring.
4. A major methodology for trying to prove commutative algebra results is by:
   (a) **Localize**: If $\mathfrak{p}$ is a prime ideal, then $A_{\mathfrak{p}} \subset B_{\mathfrak{p}}$ with $(A_{\mathfrak{p}}, \mathfrak{p}A_{\mathfrak{p}})$ a local ring. Then try to prove the results here and pullback.
   (b) **Quotient**: If $\mathfrak{P}$ is a prime ideal in B, let $\mathfrak{p} = \mathfrak{P} \cap A$ a prime ideal in A with the induced map $A/\mathfrak{p} \to B/\mathfrak{P}$ a homomorphism of domains. Then try to prove the results here and pullback.

Lemma 82.6. *Let A be a commutative R-algebra and u a unit in A. Then u^{-1} is integral over R if and only if u^{-1} lies in $R[u]$.*

PROOF. The element u^{-1} is integral over R if and only if there exists an equation

$$u^{-n} + r_1u^{-n+1} + \cdots + r_n = 0 \text{ in } A \text{ for some } r_1, \ldots, r_n \in R$$

if and only if

$$u(r_1 + r_2u + \cdots + r_nu^{n-1}) = -1 \text{ in } A \text{ for some } r_1, \ldots, r_n \in R$$

if and only if u^{-1} lies in $R[u]$. □

Proposition 82.7. *Let $A \subset B$ be domains with B/A integral. Then A is a field if and only if B is a field.*

PROOF. ($\Leftarrow$): Let $x \in A$ be nonzero. Then x^{-1} lies in the field B and is integral over A if and only if $x^{-1} \in A[x] = A$ by the lemma.

($\Rightarrow$): Let $y \in B$ be nonzero. Then there exists $a_1, \ldots, a_n$ in A such that $y^n + a_1 y^{n-1} + \cdots + a_n = 0$ in B for some n, as B/A is integral. Since B is a domain, we may assume that a_n is nonzero. Therefore, a_n^{-1} lies in $A \subset qf(B)$, the quotient field of B, as A is a field. Consequently, $y^{-1} = -a_n^{-1}(y^{n-1} + \cdots + a_{n-1})$ in $qf(B)$, hence lies in $A[y] \subset B$. □

Corollary 82.8. *Let $\varphi : A \to B$ be an integral ring homomorphism, $\mathfrak{P}$ a prime ideal in B. Then $\mathfrak{P}$ is a maximal ideal in B if and only if ${}^a\varphi(\mathfrak{P})$ is a maximal ideal in A. In particular, the restriction of ${}^a\varphi$ to* $\mathrm{Max}(B)$ *gives the map*

$$ {}^a\varphi|_{\mathrm{Max}(B)} : \mathrm{Max}(B) \to \mathrm{Max}(A). $$

PROOF. let $\mathfrak{P}$ be a prime ideal in B, then φ induces a monomorphism $\overline{\varphi} : A/\varphi^{-1}(\mathfrak{P}) \to B/\mathfrak{P}$ which is integral by Remark 82.3(3). □

Definition 82.9. Let R be a commutative ring. The *Jacobson radical* of R is defined to be $\mathrm{rad}(R) := \bigcap_{\mathrm{Max}(R)} \mathfrak{m}$.

One of the most useful results is the following lemma, apparently first proved independently by Azumaya and Krull:

Lemma 82.10. (Nakayama's Lemma) *Let $\mathfrak{A}$ be an ideal in* $\mathrm{rad}(R)$ *and M a finitely generated R-module. If $M = \mathfrak{A}M$, then $M = 0$.*

PROOF. Suppose that $M \neq 0$ and $M = \sum_{i=1}^n Rm_i$ with n minimal. Then there exist $a_1, \ldots, a_n$ in $\mathfrak{A}$ satisfying $m_1 = \sum_{i=1}^n a_i m_i$, as $M = \mathfrak{A}M$. Thus $(1 - a_1)m_1 = \sum_{i=2}^n a_i m_i$. Since each a_i lies in $\mathrm{rad}(R) \subset \mathfrak{m}$ for every maximal ideal $\mathfrak{m}$ in R, we have $1 - a_1 \notin \mathfrak{m}$ for every maximal ideal $\mathfrak{m}$, hence is a unit in R. It follows that m_1 lies in $\sum_{i=2}^n Rm_i$, contradicting the minimality of n. □

The following consequence is also called Nakayama's Lemma.

Corollary 82.11. *Let $\mathfrak{A} \subset \mathrm{rad}(R)$ be an ideal in R, M a finitely generated R-module, and $N \subset M$ a submodule. If $M = N + \mathfrak{A}M$, then $M = N$.*

PROOF. We have M/N is a finitely generated R-module satisfying $M/N = \mathfrak{A}(M/N)$, so $M/N = 0$ by Nakayama's Lemma. Hence $M = N$. □

We leave the proof of the next useful corollary as an exercise.

Corollary 82.12. *Let $(R, \mathfrak{m})$ be a local ring, M a finitely generated R-module, $^- : M \to M/\mathfrak{m}$, the canonical epimorphism,*

$$\mathcal{S} = \{x_1, \dots, x_n\}, \text{ and } \overline{\mathcal{S}} = \{\overline{x_1}, \dots, \overline{x_n}\}.$$

Then

(1) $\mathcal{S}$ generates M if and only if $\overline{\mathcal{S}}$ spans the $R/\mathfrak{m}$-vector space $\overline{M} = M/\mathfrak{A}M$.
(2) $\mathcal{S}$ is a minimal generating set (obvious definition) for M if and only if $\overline{\mathcal{S}}$ is an $R/\mathfrak{m}$-basis for $\overline{M}$ (and $\dim_{R/\mathfrak{m}} \overline{M} = n$).

Remark 82.13. We shall see later that if $(R, \mathfrak{m})$ is a Noetherian local ring that

$$\dim R = \operatorname{ht} \mathfrak{m} \leq \dim_{R/\mathfrak{m}} \mathfrak{m}/\mathfrak{m}^2 < \infty.$$

If we have equality in the above, then $(R, \mathfrak{m})$ is called a *regular local ring*. This notion of regular is the algebraic replacement for the geometric concept of non-singularity at a point.

A basic theorem in the study of integral extensions, apparently originally due to Krull, is the next result that we now prove. The proof is an excellent illustration of the methods mentioned above.

Theorem 82.14. (Cohen-Seidenberg Theorem)
Let B/A be integral. Then:

(1) (Incomparability) *If $\mathfrak{P}_1 \subset \mathfrak{P}_2$ are prime ideals in B, and satisfy $\mathfrak{P}_1 \cap A = \mathfrak{P}_2 \cap A$, then $\mathfrak{P}_1 = \mathfrak{P}_2$.*
(2) (Lying Over) *${}^a i : \operatorname{Spec}(B) \to \operatorname{Spec}(A)$ is surjective, where $i : A \to B$ is the inclusion map, i.e., (in the notation of Section §75 that we shall continue to use) if $\mathfrak{p}$ is a prime ideal in A, there exists a prime ideal $\mathfrak{P}$ in B lying over $\mathfrak{p}$.*
(3) (Going Up). *If $\mathfrak{p}_1 \subset \mathfrak{p}_2$ are prime ideals in A and $\mathfrak{P}_1$ is a prime ideal in B lying over $\mathfrak{p}_1$, then there exists a prime ideal in $V(\mathfrak{P}_1)$ lying over $\mathfrak{p}_2$.*

(Picture)

$$\begin{array}{ccc} \mathfrak{P}_1 & \subset & \square \\ | & & | \\ \mathfrak{p}_1 & \subset & \mathfrak{p}_2. \end{array}$$

(4) $\dim A = \dim B$.

Proof. (1): Let $\mathfrak{p} = \mathfrak{P}_1 \cap A = \mathfrak{P}_2 \cap A$ in $\operatorname{Spec}(A)$ and $S = A \setminus \mathfrak{p}$. Then $S \cap \mathfrak{P}_i = \emptyset$ for $i = 1, 2$, so $S^{-1}B/S^{-1}A$ is integral, $(S^{-1}A, S^{-1}\mathfrak{p})$ is a local ring with

$$S^{-1}\mathfrak{P}_i \cap S^{-1}A = S^{-1}(\mathfrak{P}_i \cap A) = S^{-1}\mathfrak{p}.$$

Therefore, $S^{-1}\mathfrak{P}_i \in \mathrm{Spec}(S^{-1}B)$ lies over $S^{-1}\mathfrak{p}$, for $i = 1, 2$. By Corollary 82.8, $S^{-1}\mathfrak{P}_i$ is a maximal ideal in $S^{-1}B$ for $i = 1, 2$. It follows that $\mathfrak{P}_1 \subset \mathfrak{P}_2$ implies $S^{-1}\mathfrak{P}_1 = S^{-1}\mathfrak{P}_2$. Consequently, $\mathfrak{P}_1 = \mathfrak{P}_2$ as $\mathfrak{P}_i \cap S = \emptyset$ for $i = 1, 2$.

(2): Let $S = A \setminus \mathfrak{p}$, then $S^{-1}B/S^{-1}A$ is integral. By the diagram

$$\begin{array}{ccc} B & \xrightarrow{\text{loc}} & S^{-1}B \\ | & & | \\ A & \xrightarrow[\text{loc}]{} & \mathrm{S}^{-1}A \end{array} \qquad \text{commutes and} \qquad \begin{array}{c} S^{-1}\mathfrak{P} \\ | \\ S^{-1}\mathfrak{p} \end{array}$$

if $S^{-1}\mathfrak{P}$ lies over $S^{-1}\mathfrak{p}$, we see that it suffices to replace A by $S^{-1}A$ and assume that $(A, \mathfrak{p})$ is a local ring with B/A integral. In particular, it suffices to show that there exists a prime ideal $\mathfrak{P}$ in B lying over $\mathfrak{p}$. Suppose that $\mathfrak{p}B < B$. Then there exists an ideal $\mathfrak{m} \in V(\mathfrak{p}B) \cap \mathrm{Max}(B)$. Since B/A is integral, $\mathfrak{m} \cap A \in \mathrm{Max}(A) = \{\mathfrak{p}\}$, i.e., $\mathfrak{m} \cap A = \mathfrak{p}$ and we would be done. So we may assume that $\mathfrak{p}B = B$. We then have an equation

$$1 = \sum_{i=1}^{m} p_i b_i \text{ for some } p_i \in \mathfrak{p},\ b_i \in B.$$

Set $B' = A[b_1, \ldots, b_m]$ a finitely generated commutative A-algebra. As B/A is integral, B' is a finitely generated A-module satisfying $\mathfrak{p}B' = B'$. It follows by Nakayama's Lemma, that $B' = 0$, which is impossible. Hence $\mathfrak{p}B < B$ and (2) follows.

(3): Let $\mathfrak{P}_1 \in \mathrm{Spec}(B)$ lie over $\mathfrak{p}_1 \in \mathrm{Spec}(A)$. Since the inclusion $A \subset B$ induces an integral monomorphism $A/\mathfrak{p}_1 \to B/\mathfrak{P}_1$, by Lying Over and the Correspondence Principle, there exists a prime ideal $\mathfrak{P}_2$ in $V(\mathfrak{P}_1)$ with $\mathfrak{P}_2/\mathfrak{P}_1$ lying over $\mathfrak{p}_2/\mathfrak{p}_1$ in $\mathrm{Spec}(A/\mathfrak{p}_1)$. Therefore, $\mathfrak{P}_2$ lies over $\mathfrak{p}_1$ by the Correspondence Principle and (3) follows.

(2): Let

$$\mathfrak{P}_1 < \cdots < \mathfrak{P}_n$$

be a chain of prime ideals in B. By Incomparability,

$$\mathfrak{P}_1 \cap A < \cdots < \mathfrak{P}_n \cap A$$

is a chain of prime ideals in A, so $\dim B \leq \dim A$. If

$$\mathfrak{p}_1 < \cdots < \mathfrak{p}_n$$

is a chain of prime ideals in A, then by Lying Over and Going Up, we can construct a chain of prime ideals

$$\mathfrak{P}_1 < \cdots < \mathfrak{P}_n$$

in B, so $\dim A \leq \dim B$. Therefore, $\dim A = \dim B$. □

In the theorem, we needed the condition that $A \subset B$, i.e., that we have a monomorphism $A \to B$. Indeed a counterexample is provided by the canonical epimorphism $^- : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$, $n > 1$ if not.

Corollary 82.15. *Let B/A be an integral extension with B a domain and $\mathfrak{B} < B$ a nonzero ideal. Then $\mathfrak{B} \cap A < A$ is a nonzero ideal.*

Proof. Let $\mathfrak{P} \in V(\mathfrak{B})$. Certainly, $\mathfrak{P} \cap A < A$ as $1 \notin \mathfrak{P}$. Since A and B are domains, the zero ideal in B lies over the zero ideal in A. By Incomparability, $\mathfrak{P}$ does not lie over (0). Suppose that $\mathfrak{B} \cap A = 0$, then the induced mononomorphism $A \hookrightarrow B/\mathfrak{B}$ is integral. Therefore, by Lying Over and the Correspondence Principle, there exists a prime ideal $\mathfrak{P} \in V(\mathfrak{B})$ satisfying $\mathfrak{P}/\mathfrak{B}$ lies over (0) in $\operatorname{Spec}(A)$. As $\mathfrak{B} \cap A = 0$, this means that $\mathfrak{P}$ lies over (0), a contradiction. □

We need to generalize Proposition 76.1, whose former proof can be adapted to this generalization (essentially by localization). Moreover, we shall give a different proof that uses one of the most useful results, called the Prime Avoidance Lemma that we previously left as an exercise, but now prove.

Lemma 82.16. (Prime Avoidance Lemma) *Let $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$ be ideals in R, at least $n-2$ of which are prime. Let $S \subset R$ be a subrng (it does not have to have a 1) contained in $\mathfrak{A}_1 \cup \cdots \cup \mathfrak{A}_n$. Then there exists a j such that $S \subset \mathfrak{A}_j$. In particular, if $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ are prime ideals in R and $\mathfrak{B}$ is an ideal properly contained in S satisfying $S \setminus \mathfrak{B} \subset \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_n$, then S lies in one of the $\mathfrak{p}_i$'s.*

Proof. The case $n = 1$ is trivial, so assume that $n > 1$ and assume that the result is false. By induction, $S \not\subset \mathfrak{A}_1 \cup \cdots \widehat{\mathfrak{A}_i} \cup \cdots \cup \mathfrak{A}_n$ for each i where $\widehat{\ }$ means omit. So for each $i = 1, \ldots, n$, there exists an element $x_i \in S \setminus \mathfrak{A}_1 \cup \cdots \widehat{\mathfrak{A}_i} \cup \cdots \cup \mathfrak{A}_n$. Therefore, we have $x_i \in \mathfrak{A}_i$ for every $i = 1, \ldots, n$ and $x_i \notin \mathfrak{A}_j$ for all $j \neq i$. If $n > 2$, we may assume that $\mathfrak{A}_1$ is a prime ideal. Let $y = x_1 + x_2 \cdots x_n \in S$. If $n = 2$, then $y = x_1 + x_2$ does not lie in $\mathfrak{A}_1 + \mathfrak{A}_2$, a contradiction. So we may assume that $n > 2$. As $\mathfrak{A}_1$ is a prime ideal and $x_i \notin \mathfrak{A}_1$ for all $i > 1$, we have $x_2 \cdots x_n \notin \mathfrak{A}_1$. Consequently, $y \notin \mathfrak{A}_1 \cup \cdots \cup \mathfrak{A}_n$, a contradiction. □

Our generalization of Proposition 76.1 needs us to replace the condition of a finite Galois extension of fields by an arbitrary normal extension of fields. This requires that we extend the definition of the norm of a finite separable field extension. Let L/F be a finite normal extension of fields, and K the separable closure of F in K. We call

$[L:F]_i := [L:K]$ the *purely inseparable degree* of L/F and $[L:F]_s := [K:F]$ the *separable degree* of F in L. So $[L:F] = [L:F]_i[L:F]_s$. Define the norm $\mathrm{N}_{L/F}: L \to F$ by $\mathrm{N}_{L/F}(x) = \big(\prod_{G(L/F)} \sigma(x)\big)^{[L:F]_i}$, i.e., $\mathrm{N}_{L/K}(x) = \mathrm{N}_{K/F}(x^{[L:F]_i})$. Then the norm still satisfies all the expected properties.

Theorem 82.17. *Let A be a normal domain (i.e., an integrally closed domain) with quotient field F and L/F a normal extension of fields. Let $\mathfrak{P}_1$ and $\mathfrak{P}_2$ be prime ideals in A_L lying over $\mathfrak{p}$ in $\mathrm{Spec}(A)$, i.e., $\mathfrak{P}_1 \cap A = \mathfrak{p} = \mathfrak{P}_2 \cap A$. Then there exists an element σ in $G(L/F)$, the Galois group of L/F, satisfying $\mathfrak{P}_1 = \sigma(\mathfrak{P}_2)$, i.e., where $i: A \to A_L$ is the inclusion map, then $G(L/F)$ acts transitively on the fibers of ${}^a i: \mathrm{Spec}(A_L) \to \mathrm{Spec}(A)$.*

Proof. If $b \in A_L$ and $\sigma \in G(L/F)$, then $\sigma(b) \in L$ is integral over $\sigma(A) = A$, i.e., $\sigma(b) \in \big(\sigma(A)\big)_L = A_L$, so $\sigma(A_L) = A_L$. As σ^{-1} lies in $G(L/F)$, the map $\sigma: A_L \to A_L$ is an A-algebra isomomorphism. If $\mathfrak{P}$ is a prime ideal in A_L and $\sigma \in G(L/F)$, then $\sigma(\mathfrak{P}) \in \mathrm{Spec}(A_L)$ and $\mathfrak{P} \cap A = \sigma(\mathfrak{P} \cap A) = \sigma(\mathfrak{P}) \cap A$. So every prime ideal in the orbit $G(L/F)\mathfrak{P} := \{\sigma(\mathfrak{P}) \mid \sigma \in G(L/F)\}$ lies in the fiber $({}^a i)^{-1}(\mathfrak{p})$ of $\mathfrak{p}$.

Case 1. L/F is finite.

Suppose $\mathfrak{P}_2$ lies over $\mathfrak{P}_1 \cap A$. To show that $\mathfrak{P}_2 \in G(L/F)\mathfrak{P}_1$. Suppose not. By Incomparability, $\mathfrak{P}_2 \notin \sigma(\mathfrak{P}_1)$ for any $\sigma \in G(L/F)$. By the Prime Avoidance Lemma, as $G(L/F)\mathfrak{P}_1$ is finite, $\mathfrak{P}_2 \not\subset \bigcup_{G(L/F)} \sigma(\mathfrak{P}_1)$. Let a be an element in $\mathfrak{P}_2 \setminus \bigcup_{G(L/F)} \sigma(\mathfrak{P}_1)$. Then $\sigma(a) \notin \mathfrak{P}_1$ for any $\sigma \in G(L/F)$. As A is a normal domain, we have

$$\mathrm{N}_{L/F}(a) = \Big(\prod_{G(L/F)} \sigma(a)\Big)^{[L:F]_i} \text{ lies in } A_L \cap F = A_F = A.$$

Therefore,

$$\mathrm{N}_{L/F}(a) \in \mathfrak{P}_2 \cap A = \mathfrak{p} = \mathfrak{P}_1 \cap A \subset \mathfrak{P}_1.$$

As $\mathfrak{P}_1$ is a prime ideal and $G(L/F)$ is a finite group, there exists an σ in $G(L/F)$ satisfying $\sigma(a) \in \mathfrak{P}_1$, a contradiction.

Case 2. L/F is possibly infinite.

This is left an exercise (use Zorn's Lemma, since L is a union of finite normal extensions). □

Corollary 82.18. *Let A be a normal domain with quotient field F and K/F a finite field extension. Then ${}^a i: \mathrm{Spec}(A_K) \to \mathrm{Spec}(A)$ has finite fibers, where i is the inclusion map of A in A_K, i.e., $({}^a i)^{-1}(\mathfrak{p})$ is a finite set for all prime ideals $\mathfrak{p}$ in A.*

A ring homomorphism whose associated map has finite fibers is called *quasi-finite*.

PROOF. Let L/F be a normal closure of K/F. Then L/F is a finite normal extension, hence $G(L/F)$ is finite. Let $\mathfrak{p}$ be a prime ideal in A. By Lying Over, there exists a prime ideal $\mathfrak{P}$ in A_L lying over $\mathfrak{p}$. We have $({}^a i_L)^{-1}(\mathfrak{p}) = \{\sigma(\mathfrak{P}) \mid \sigma \in G(L/F)\}$ where $i_L : A \to A_L$ is the inclusion map. By Lying Over, if $\mathfrak{Q}$ is a prime ideal in A_K, then there exists a prime ideal $\mathfrak{P}'$ in A_L lying over $\mathfrak{Q}$ as A_L/A_K is integral. Hence

$$|({}^a i)^{-1}(\mathfrak{p})| \leq |({}^a i_L)^{-1}(\mathfrak{p})| \leq |G(L/F)| < \infty$$

where $i : A \to A_K$ is the inclusion. □

Definition 82.19. A commutative ring R with finitely many maximal ideals is called *semi-local ring*.

Corollary 82.20. *Let A be a semi-local normal domain with quotient field F and K/F a finite field extension. Then A_K is semi-local.*

PROOF. As A_K/A is integral, ${}^a i|_{\mathrm{Max}(A)_K} : \mathrm{Max}(A_K) \to \mathrm{Max}(A)$, where i is the inclusion of A in A_K. As ${}^a i$ has finite fibers, the result follows. □

Proposition 82.21. *Let A be a normal domain with quotient field F and K/F a finite field extension. Suppose that $\mathfrak{P}$, a prime ideal in A_K, lies over the prime ideal $\mathfrak{p}$ in A. Let $\overline{F}$ denote the quotient field of $A/\mathfrak{p}$. Then for all $\alpha \in A_K/\mathfrak{P}$, we have $[\overline{F}(\alpha) : \overline{F}] \leq [K : F]$.*

PROOF. Let $\bar{} : A[t] \to (A/\mathfrak{p})[t] \subset \overline{F}$ be the canonical epimorphism and $x \in A_K$ an element satisfying $\alpha = x + \mathfrak{P}$ in $A_K/\mathfrak{P}$. Then the minimal polynomial $m_F(x)$ of x in $A_F[t] = A[t]$ is monic and $\overline{m_F(x)}|_{t=\alpha} = 0$, so

$$[\overline{F}(\alpha) : \overline{F}] \leq \deg_{\overline{F}} m_{\overline{F}}(\alpha) \leq \deg_F m_F(x) \leq [K : F]. \qquad \square$$

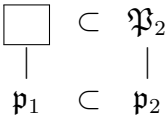
Corollary 82.22. *Let A be a normal domain with quotient field F and K/F a finite field extension. Suppose that there exists an element x in A_K satisfying $A_K = A[x]$. Let $\mathfrak{P}$ be a prime ideal in A_K lying over the prime ideal $\mathfrak{p}$ in A. Set $\overline{K} = qf(A_K/\mathfrak{P})$ and $\overline{F} = qf(A/\mathfrak{p})$. Then $[\overline{K} : \overline{F}] \leq [K : F]$.*

PROOF. $\overline{K} = qf(A/\mathfrak{p})[x + \mathfrak{P}]$. □

The Cohen-Seidenberg Theorem showed that if B/A was integral, then $\dim A = \dim B$. We are also interested in the heights of primes, as we wish to determine the codimension of irreducible subvarieties. This is more delicate and needs stronger hypotheses. We can now determine one such result.

Theorem 82.23. (Going Down Theorem) *Let B/A be integral with A a normal domain and B a domain. Let $\mathfrak{p}_1 \subset \mathfrak{p}_2$ be prime ideals in A and $\mathfrak{P}_2$ a prime ideal in B lying over $\mathfrak{p}_2$. Then there exists a prime ideal $\mathfrak{P}_1$ lying over $\mathfrak{p}_1$ satisfying $\mathfrak{P}_1 \subset \mathfrak{P}_2$. In particular, if $\mathfrak{P}$ is a prime ideal in B, then* $\operatorname{ht}\mathfrak{P} = \operatorname{ht}(\mathfrak{P} \cap A)$.

(Picture)



PROOF. Let F be the quotient field of A and K the quotient field of B. Since B/A is integral, K/F is algebraic. Let L/K be a normal closure of K/F. As A_L/A is integral and $B \subset A_L$, we have A_L/B is also integral. Let $\mathfrak{Q}_1 \in \operatorname{Spec}(A_L)$ lie over $\mathfrak{p}_1$. By Going Up, there exists a prime ideal $\mathfrak{Q}_2 \in V(\mathfrak{Q}_1)$ lying over $\mathfrak{p}_2$. By Lying Over there exists a prime ideal $\mathfrak{Q}_2'$ in A_L lying over $\mathfrak{P}_2$ hence over $\mathfrak{p}_2$. Therefore, there exists an element $\sigma \in G(L/F)$ satisfying $\sigma(\mathfrak{Q}_2) = \mathfrak{Q}_2'$. We also have $\sigma(\mathfrak{Q}_1) \cap A = \mathfrak{p}_1 = \mathfrak{Q}_1 \cap A$. It follows that $\mathfrak{P}_1 = \sigma(\mathfrak{Q}_1) \cap B$ works for the first statement.

Let

$$\mathfrak{P}_1 < \cdots < \mathfrak{P}_n = \mathfrak{P}$$

be a chain of prime ideals in B and $\mathfrak{p} = \mathfrak{P} \cap A$. By Incomparability, we have

$$\mathfrak{P}_1 \cap A < \cdots < \mathfrak{P}_n \cap A = \mathfrak{P} \cap A = \mathfrak{p}$$

a chain of prime ideals in A. Therefore, $\operatorname{ht}(\mathfrak{P} \cap A) \geq \operatorname{ht}\mathfrak{P}$. If

$$\mathfrak{p}_1 < \cdots < \mathfrak{p}_n = \mathfrak{p} = \mathfrak{P} \cap A$$

is a chain of prime ideals in A, then, by Going Down, there exists a chain of prime ideals

$$\mathfrak{P}_1 < \cdots < \mathfrak{P}_n = \mathfrak{P}$$

in B, so $\operatorname{ht}\mathfrak{P} \geq \operatorname{ht}(\mathfrak{P} \cap A)$, and the result follows. □

Corollary 82.24. *Let B/A be integral with A a normal domain and B a domain. Let $\mathfrak{p}$ be a prime ideal in A and $\mathfrak{P}$ a prime ideal in B lying over $\mathfrak{p}$. Then* $\operatorname{ht}\mathfrak{p} = 1$ *if and only if* $\operatorname{ht}\mathfrak{P} = 1$,

Exercises 82.25. 1. Prove the assertions in Remarks 82.3.

2. Prove the assertions in Remarks 82.4.
3. Let $\varphi : A \to B$ be an integral homomorphism. Show that the associated map ${}^a\varphi : \operatorname{Spec}(B) \to \operatorname{Spec}(A)$ is a closed map.
4. Prove Corollary 82.12.

5. Suppose that $\mathfrak{P} = \cap_i \mathfrak{p}_i$ in a commutative ring R with each $\mathfrak{P}, \mathfrak{p}_1, \ldots, \mathfrak{p}_n$ a prime ideal. Show that there exists an i with $\mathfrak{P} = \mathfrak{p}_i$.
6. Prove Case 2 of Theorem 82.17.
7. (Cayley-Hamilton Theorem) Let R be a commutative ring, $\mathfrak{A} < R$ an ideal, and M an R-module that can be generated by n elements. If $\varphi \in \mathrm{End}_R(M)$ satisfies, $\varphi(M) \subset \mathfrak{A}M$, then there exists a monic polynomial $f = t^n + a_{n-1}t^{n-1} + \cdots + a_0$ in $R[t]$ with coefficients $a_0, \ldots, a_{n-1} \in \mathfrak{A}$, such that the R-endomorphism on M, $f(\varphi) = \varphi^n + a_{n-1}\varphi^{n-1} + \cdots + a_0 1_M = 0$.
   [The characterization of integral elements (as well as Nakayama's Lemma 82.10) easy follow from this.]
8. Let $\mathfrak{A}$ be an ideal in $R[t]$ and $^- : R[t] \to R[t]/\mathfrak{A}$, the canonical epimorphism. Set $x = \bar{t}$ and $A = F[t]/\mathfrak{A}$. Prove the following:
   (*i*) As an R-module A is generated by at most n elements if and only if there exists a monic polynomial $f \in \mathfrak{A}$ of degree at most n. If this is the case, then $A = \sum_{i=0}^{n-1} Rx_i$, for some $x_0, \ldots, x_{n-1}$. In particular, A is a finitely generated R-module if and only if $\mathfrak{A}$ contains a monic polynomial.
   (*ii*) A is a free R-module if and only if $\mathfrak{A} = (f)$ for some monic polynomial $f \in R[t]$. If this is the case, say $\mathrm{rank}\, A = n$. Then there exists an $x \in A$ such that $\mathcal{B} = \{1, x, \ldots, x^{n-1}\}$ is a basis for A.
9. Show that the Going Down Theorem 82.23 still holds if the condition on B is weakened to B is an A-torsion-free module.
10. Under the conditions of the previous exercise, show if $\mathfrak{p}$ is a prime ideal in A and $i : A \to B$ is the inclusion map, then
$$({}^a i)^{-1}(\mathfrak{p}) = \{\mathfrak{P} \in \mathrm{Spec}(B)\, \mathfrak{P} \in V(\mathfrak{p}B) \text{ is minimal}\}.$$
11. Let $(R, \mathfrak{m})$ be a local ring. Show that any finitely generated R-projective module is a free R-module.
12. Let R be a subring of a commutative ring A. Show that the Going Down Theorem holds for A/R if ${}^a i : \mathrm{Spec}(A) \to \mathrm{Spec}(R)$ is an open map. (The converse of this is true if R is a Noetherian ring.)

83. Primary Decomposition

As throughout this chapter, R will denote a commutative ring.

In this section, we investigate the generalization of unique factorization of ideals that Dedekind domains satisfy called the Lasker-Noether Theorem. As unique factorization of ideals into a product of prime ideals characterize Dedekind domains, the appropriate generalization

is, as one would expect, weaker. Multiplication of ideals is replaced by intersections, powers of primes by primary ideals, and not all primary ideals are unique. However, this generalization will apply to any Noetherian ring, not just a Noetherian domain.

We shall need the following lemma whose proof we leave as an exercise:

Recall that the radical of an ideal $\mathfrak{A}$ in R is defined to be

$$\sqrt{\mathfrak{A}} := \{x \in R \mid x^n \in \mathfrak{A} \text{ for some } n \in \mathbb{Z}^+\}.$$

Lemma 83.1. *Let $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$ be ideals in R. Then*

$$\sqrt{\bigcap_{i=1}^{n} \mathfrak{A}_i} = \bigcap_{i=1}^{n} \sqrt{\mathfrak{A}_i} = \sqrt{\mathfrak{A}_1 \cdots \mathfrak{A}_n}.$$

Recall that an ideal $\mathfrak{Q} < R$ is is called a *primary ideal* if it satisfies one of the following equivalent conditions:

(*i*) If xy lies in $\mathfrak{Q}$ then either x lies in $\mathfrak{Q}$ or there exists a positive integer n with $y^n \in \mathfrak{Q}$.

(*ii*) If xy lies in $\mathfrak{Q}$ then either x lies in $\mathfrak{Q}$ or $y \in \sqrt{\mathfrak{Q}}$.

(*iii*) The set of zero divisors $\mathrm{zd}(R/\mathfrak{A})$ of $R/\mathfrak{A}$ lies in the nilradical $\mathrm{nil}(R/\mathfrak{Q})$ of $R/\mathfrak{A}$.

Lemma 83.2. *Let $\mathfrak{Q}$ be a primary ideal in R, then $\sqrt{\mathfrak{Q}}$ is a prime ideal in R and is the smallest prime ideal in R containing $\mathfrak{Q}$.*

Proof. Suppose that xy lies in $\mathfrak{Q}$. Then there exists a positive integer n such that $x^n y^n$ lies in $\mathfrak{Q}$. Hence either $x^n \in \mathfrak{Q}$ or $y^{nm} \in \mathfrak{Q}$ i.e., either $x \in \sqrt{\mathfrak{Q}}$ or $y \in \sqrt{\mathfrak{Q}}$. It follows that $\sqrt{\mathfrak{Q}}$ is a prime ideal in R. As $\mathfrak{Q} \subset \sqrt{\mathfrak{Q}}$ and if $\mathfrak{Q}$ lies in a prime ideal $\mathfrak{p}$ so does $\sqrt{\mathfrak{Q}}$, the second statement also follows. □

Definition 83.3. If $\mathfrak{Q}$ be a primary ideal in R and $\mathfrak{p}$ is the prime ideal $\sqrt{\mathfrak{Q}}$, we say that $\mathfrak{Q}$ is a $\mathfrak{p}$-*primary ideal*.

A useful example is given by the following result:

Lemma 83.4. *Let $\mathfrak{A} < R$ be an ideal satisfying $\sqrt{\mathfrak{A}}$ is a maximal ideal in R. Then $\mathfrak{A}$ is a $\sqrt{\mathfrak{A}}$-primary ideal. In particular, if $\mathfrak{m}$ is a maximal ideal in R, then $\mathfrak{m}^n$ is an $\mathfrak{m}$-primary ideal for each positive integer n.*

Proof. Let $\mathfrak{m} = \sqrt{\mathfrak{A}}$. Then $\mathfrak{m} = \bigcap_{V(\mathfrak{A})} \mathfrak{p}$, hence $\mathfrak{m} \subset \mathfrak{P}$ for every prime ideal in $V(\mathfrak{A})$. Since $\mathfrak{m}$ is maximal, $V(\mathfrak{A}) = \{\mathfrak{m}\}$. It follows that $R/\mathfrak{A}$ is a primary ring, i.e., $|\operatorname{Spec}(R/\mathfrak{A})| = 1$. In particular, we have $\mathrm{zd}(R/\mathfrak{A}) = \mathrm{nil}(R/\mathfrak{A})$, so $\mathfrak{A}$ is $\mathfrak{m}$-primary. □

Examples 83.5. 1. Each prime ideal in R is a primary ideal in R.

2. The ideal (n) in $\mathbb{Z}$ is primary if and only if $n = 0$ or n is a power of a prime.
3. Let F be a field and x, y, z indeterminants over F. Then $(xy - z^2)$ is a prime ideal in the the UFD $F[x, y, z]$. Set $R = F[x, y, z]/(xy - z^2)$ and $^- : F[x, y, z] \to R = F[\overline{x}, \overline{y}, \overline{z}]$ the canonical epimorphism. Then $\mathfrak{p} = (\overline{z})$ is a prime ideal in the domain R. We have

$$\overline{xy} = \overline{z}^2 \in \mathfrak{p}^2 := \mathfrak{p}\mathfrak{p}, \text{ but } \overline{x}^2 \notin \mathfrak{p}^2 \text{ and } \overline{y} \notin \sqrt{\mathfrak{p}^2} = \mathfrak{p}.$$

   Therefore, $\mathfrak{p}^2$ is not a primary ideal in R. So, in general, a power of a prime ideal may not be a primary ideal.
4. Let F be a field, x, y indeterminants over F and $R = F[x, y]$. Then $\mathfrak{Q} = (x, y^2)$ is an (x, y)-primary ideal in R as $R/\mathfrak{Q} = F[y]/(y^2)$ satisfies $\mathrm{zd}(R/\mathfrak{Q}) \subset \mathrm{nil}(R/\mathfrak{Q})$. Let $\mathfrak{p} = (x, y)$, then $(x^2, xy, y^2) = \mathfrak{p}^2 < \mathfrak{Q} \subset \mathfrak{p}$. It follows that $\mathfrak{Q} \neq \mathfrak{p}^n$ for any integer n, i.e., a primary ideal may not be a power of a prime ideal.
5. Suppose that $\mathfrak{Q}_i$, $i = 1, \ldots, r$, are all $\mathfrak{p}$-primary ideals. Then so is $\mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_r$:

   We know that $\sqrt{\bigcap \mathfrak{Q}_i} = \bigcap \sqrt{\mathfrak{Q}_i} = \mathfrak{p}$. If $xy \in \bigcap \mathfrak{Q}_i$ with $y \notin \bigcap \mathfrak{Q}_i$, then there exists a j with $y \notin \mathfrak{Q}_j$. As $xy \in \mathfrak{Q}_j$, we have x^n lies in $\mathfrak{Q}_j$ for some n, hence x lies in $\sqrt{\mathfrak{Q}_j} = \mathfrak{p} = \sqrt{\bigcap \mathfrak{Q}_i}$, so $\bigcap \mathfrak{Q}_i$ is $\mathfrak{p}$-primary.

Definition 83.6. Let $\mathfrak{A}$ and $\mathfrak{B}$ be two ideals in R. Set

$$(\mathfrak{A} : \mathfrak{B}) := \{y \in R \mid y\mathfrak{B} \subset \mathfrak{A}\} \subset R,$$

an ideal in R called the *colon ideal*.

If $\mathfrak{B} = (x)$, we write $(\mathfrak{A} : x)$ for $\big(\mathfrak{A} : (x)\big)$. If $\mathfrak{A} = 0$, then $(0 : x) = \{y \in R \mid yx = 0\}$ is the annihilator $\mathrm{ann}_R(x)$ of x in R.
Note the following:

Remarks 83.7. We have:

1. $\mathrm{zd}(R) = \bigcup_{x \neq 0} \sqrt{\mathrm{ann}_R(x)} = \bigcup_{x \neq 0} \sqrt{(0 : x)}$.
2. If $\mathfrak{A}_i \subset R$ are ideals for $i = 1, \ldots, n$, then

$$(\bigcap_{i=1}^{n} \mathfrak{A}_i : x) = \bigcap_{i=1}^{n} (\mathfrak{A}_i : x).$$

Computation 83.8. Let $\mathfrak{Q}$ be a $\mathfrak{p}$-primary ideal in R and $x \in R$. Then we have:

(i) If $x \in \mathfrak{Q}$, then $(\mathfrak{Q} : x) = R$.

(*ii*) If $x \notin \mathfrak{Q}$, then $(\mathfrak{Q} : x)$ is $\mathfrak{p}$-primary. In particular, $\sqrt{(\mathfrak{Q} : x)} = \mathfrak{p}$.
(*iii*) If $x \notin \mathfrak{p}$, then $(\mathfrak{Q} : x) = \mathfrak{Q}$.

PROOF. (*i*) is immediate.
(*iii*): If $x \notin \mathfrak{p}$ and $y \in (\mathfrak{Q} : x)$, then $xy \in \mathfrak{Q}$, hence $y \in \mathfrak{Q}$ by definition.
(*ii*): Let $y \in (\mathfrak{Q} : x)$. As $x \notin \mathfrak{p}$ and $xy \in \mathfrak{Q}$, we have $y \in \sqrt{\mathfrak{Q}} = \mathfrak{p}$. Hence $\mathfrak{Q} \subset (\mathfrak{Q} : x) \subset \mathfrak{p}$, so

$$\mathfrak{p} = \sqrt{\mathfrak{Q}} \subset \sqrt{(\mathfrak{Q} : x)} \subset \sqrt{\mathfrak{p}} = \mathfrak{p}.$$

Next let $y, z \in R$ with $yz \in (\mathfrak{Q} : x)$ and $y \notin \sqrt{\mathfrak{Q}} = \mathfrak{p}$, then $yzx \in \mathfrak{Q}$ implies that $xz \in \mathfrak{Q}$, hence $z \in (\mathfrak{Q} : x)$ showing that $(\mathfrak{Q} : x)$ is $\mathfrak{p}$-primary. □

With the above definitions, we can now define the type of decomposition of ideals in which we shall be interested.

Definition 83.9. Let $\mathfrak{A} < R$ be an ideal. A *primary decomposition* of $\mathfrak{A}$ is an equation

$$(*) \qquad \mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n \text{ with } \mathfrak{Q}_i \text{ primary for } i = 1, \ldots n.$$

We say that (*) is *irredundant* if, in addition,
(*i*) $\sqrt{\mathfrak{Q}_1}, \ldots, \sqrt{\mathfrak{Q}_n}$ are all distinct.
(*ii*) $\bigcap_{j \neq i} \mathfrak{Q}_j \not\subset \mathfrak{Q}_i$ for each $i = 1, \ldots, n$.

Remark 83.10. Suppose that an ideal $\mathfrak{A}$ has a primary decomposition (*). By Example 83.5(5), if $\mathfrak{Q}_{i_1}, \ldots, \mathfrak{Q}_{i_s}$ have the same radical, then $\mathfrak{Q}_{i_1} \cap \cdots \cap \mathfrak{Q}_{i_s}$ is $\sqrt{\mathfrak{Q}_{i_1}}$-primary, so any primary decomposition (*) can be made to satisfy (*i*). Throwing out those $\mathfrak{Q}_i$'s not satisfying (*ii*), then shows that any $\mathfrak{A}$ having a primary decomposition has an irredundant primary decomposition.

We show our first uniqueness statement for irredundant primary decompositions.

Theorem 83.11. *Suppose that $\mathfrak{A} < R$ is an ideal that has an irredundant primary decomposition*

$$\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n \text{ with } \mathfrak{p}_i = \sqrt{\mathfrak{Q}_i} \text{ for } i = 1, \ldots, n.$$

Then

$$\{\mathfrak{p}_1, \ldots, \mathfrak{p}_n\} = \{\sqrt{(\mathfrak{A} : x)} \mid x \in R\} \cap \operatorname{Spec}(R)$$

and is independent of the irredundant primary decomposition, i.e., the radicals of the primary ideals giving any irredundant primary decomposition of $\mathfrak{A}$ are unique.

PROOF. Let $x \in R$. We have $(\mathfrak{A} : x) = (\bigcap \mathfrak{Q}_i : x) = \bigcap (\mathfrak{Q}_i : x)$, so by the Computation 83.8,

$$\sqrt{(\mathfrak{A} : x)} = \bigcap \sqrt{(\mathfrak{Q}_i : x)} = \bigcap_{x \notin \mathfrak{Q}_i} \mathfrak{p}_i.$$

If $\sqrt{(\mathfrak{A} : x)}$ is a prime ideal, it follows that there exists a $\mathfrak{Q}_i$ with $x \notin \mathfrak{Q}_i$ such that $\sqrt{(\mathfrak{A} : x)} = \mathfrak{p}_i$ by Exercise 82.25(5).

Conversely, since our primary decomposition is irredundant, for all i and all x_i satisfying $x_i \in \bigcap_{j \neq i} \mathfrak{Q}_j \setminus \mathfrak{Q}_i$, we have $\sqrt{(\mathfrak{A} : x_i)} = \mathfrak{p}_i$ by Computation 83.8. □

The proof of the above theorem and Computation 83.8 show

Corollary 83.12. *If $\mathfrak{A} < R$ is an ideal having an irredundant primary decomposition $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \ldots, n$, then there exists an element x_i in $\mathfrak{Q}_i$ such that $(\mathfrak{A} : x_i)$ is $\mathfrak{p}_i$-primary.*

Definition 83.13. Let $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ be an irredundant primary decomposition of $\mathfrak{A}$ in R with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \ldots, n$. We set

$$\operatorname{Ass}_R V(\mathfrak{A}) := \{\mathfrak{p}_1, \ldots, \mathfrak{p}_n\},$$

the set of *associated prime ideals of* $\mathfrak{A}$ [actually of $R/\mathfrak{A}$ in modern language]. We partially order $\operatorname{Ass}_R V(\mathfrak{A})$ by set inclusion $\subset$. Elements that are minimal in $\operatorname{Ass}_R V(\mathfrak{A})$ under this partial order are called *isolated prime ideals* of $\mathfrak{A}$. The others (if any) are called *embedded prime ideals* of $\mathfrak{A}$.

Isolated prime ideals in $\operatorname{Ass}_R V(\mathfrak{A})$ are, in fact, precisely the prime ideals minimally containing $\mathfrak{A}$ in R, which we shall show next. In the sequel we shall write $\mathfrak{p} \in V(\mathfrak{A})$ is minimal for such a prime ideal.

Proposition 83.14. *Suppose that the ideal $\mathfrak{A}$ has an irredundant primary decomposition. Then a prime ideal $\mathfrak{p}$ in R is an isolated prime of $\mathfrak{A}$ if and only if $\mathfrak{p} \in V(\mathfrak{A})$ is minimal if and only if $\mathfrak{p}/\mathfrak{A} \in \operatorname{Min}(R/\mathfrak{A})$.*

PROOF. Let $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ be an irredundant primary decomposition of $\mathfrak{A}$ in R with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \ldots, n$. Suppose that $\mathfrak{p}$ lies in $V(\mathfrak{A})$, then

$$\mathfrak{p} = \sqrt{\mathfrak{p}} \supset \sqrt{\mathfrak{A}} = \bigcap \sqrt{\mathfrak{Q}_i} = \bigcap \mathfrak{p}_i.$$

Hence there exists an i such that $\mathfrak{p}_i \subset \mathfrak{p}$, i.e., $\mathfrak{p}$ contains an isolated prime of $\mathfrak{A}$. □

Because of the proposition, isolated primes are also called *minimal primes* of $\mathfrak{A}$.

Proposition 83.15. *Suppose that the ideal $\mathfrak{A}$ in R has an irredundant primary decomposition $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \ldots, n$. Then*

$$\bigcup \mathfrak{p}_i = \{x \in R \mid (\mathfrak{A} : x) > \mathfrak{A}\}.$$

In particular, if $\mathfrak{A} = (0)$, then

$$\operatorname{zd}(R) = \bigcup_{\operatorname{Ass}_R(0)} \mathfrak{p}.$$

PROOF. Let $^- : R \to R/\mathfrak{A}$ be the canonical epimorphism, then **Check**. $(\overline{0}) = \overline{\mathfrak{Q}_1} \cap \cdots \cap \overline{\mathfrak{Q}_n}$ is an irredundant primary decomposition in $\overline{R}$. So it suffices to prove $\operatorname{zd}(R) = \bigcup_{\operatorname{Ass}_R(0)} \mathfrak{p}$.

By the proof of Theorem 83.11, each $\mathfrak{p}_j = \sqrt{(0 : x)}$ for some $x \in R$. The result follows. □

So we now know if (0) has an irredundant primary decomposition, then we have

$$\operatorname{zd}(R) = \bigcup_{\operatorname{Ass}_R(0)} \mathfrak{p}$$

$$\operatorname{nil}(R) = \bigcap_{\operatorname{Spec}(R)} \mathfrak{p} = \bigcap_{\operatorname{Min}(R)} \mathfrak{p}$$

$$\operatorname{Min}(R) \subset \operatorname{Ass}_R V(0).$$

Remark 83.16. If they exist, irredundant primary decompositions are not necessarily unique. An example is given as follows: Let $R = F[x, y]$ with F a field and x, y indeterminants. Then

$$(x) \cap (x, y)^2 = (x^2, xy) = (x) \cap (x^2, y)$$

are two irredundant primary decompositions for (x^2, xy).

We shall obtain a 'weaker' uniqueness statement. We leave the proof of the following as an exercise.

Lemma 83.17. *Primary ideals satisfy:*

1. *Let S be a multiplicative set in R and $\varphi_R : R \to S^{-1}R$ the canonical ring homomorphism given by $r \mapsto r/1$. Then φ_R induces a bijection:*

$$\{\mathfrak{Q} \mid \mathfrak{Q} < R \text{ primary with } \mathfrak{Q} \cap S = \emptyset\} \longrightarrow \{\mathfrak{Q} \mid \mathfrak{Q} < S^{-1}R \text{ primary}\}$$

   via $\mathfrak{Q} \mapsto S^{-1}\mathfrak{Q}$.
2. *If $\psi : A \to B$ is a ring homomorphism of commutative rings, $\mathfrak{Q} < B$ a primary ideal, then $\psi^{-1}(\mathfrak{Q}) \subset A$ is a primary ideal.*

Proposition 83.18. *Suppose that the ideal $\mathfrak{A}$ in R has an irredundant primary decomposition $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \ldots, n$ and S is a multiplicative set in R satisfying $S \cap \mathfrak{Q}_i = \emptyset$ for $1 \leq i \leq m$, and $S \cap \mathfrak{Q}_i \neq \emptyset$ for $i > m$. Then*

$$S^{-1}\mathfrak{A} = S^{-1}\mathfrak{Q}_1 \cap \cdots \cap S^{-1}\mathfrak{Q}_m$$

is an irredundant primary decomposition for $S^{-1}\mathfrak{A}$.

PROOF. We have

$$S^{-1}\mathfrak{A} = S^{-1}(\bigcap_{i=1}^{n} \mathfrak{Q}_i) = \bigcap_{i=1}^{n} S^{-1}\mathfrak{Q}_i = \bigcap_{i=1}^{m} S^{-1}\mathfrak{Q}_i$$

with $S^{-1}\mathfrak{Q}_i$ being $S^{-1}\mathfrak{p}_i$-primary, $i = 1, \ldots, m$ and the $S^{-1}\mathfrak{p}_1, \ldots, S^{-1}\mathfrak{p}_m$ distinct. It follows that $S^{-1}\mathfrak{A} = \bigcap_{i=1}^{m} S^{-1}\mathfrak{Q}_i$ is an irredundant primary decomposition. □

Remark 83.19. In the above, if $\varphi_R : R \to S^{-1}R$ is the canonical ring homomorphism $r \mapsto r/1$, then

$$(\varphi_R)^{-1}(S^{-1}\mathfrak{A}) = (\varphi_R)^{-1}(\bigcap_{i=1}^{m} S^{-1}\mathfrak{Q}_i) = \bigcap_{i=1}^{m} (\varphi_R)^{-1}(S^{-1}\mathfrak{Q}_i) = \bigcap_{i=1}^{m} \mathfrak{Q}_i.$$

Definition 83.20. Let $\mathfrak{A} < R$ be an ideal with an irredundant primary decomposition. A subset $I \subset \operatorname{Ass}_R V(\mathfrak{A})$ is called *isolated* if whenever $\mathfrak{p}, \mathfrak{p}'$ lie in $\operatorname{Ass}_R V(\mathfrak{A})$ and satisfy $\mathfrak{p} \subset \mathfrak{p}'$, then $\mathfrak{p}' \in I$ implies $\mathfrak{p} \in I$, i.e., all descending chains of prime ideals lying in $\operatorname{Ass}_R V(\mathfrak{A})$ beginning with an element of I have all elements in the chain lying in I.

Remark 83.21. Let $I \subset \operatorname{Ass}_R V(\mathfrak{A})$ be an isolated set where $\mathfrak{A}$ has an irredundant primary decomposition. Let $S = R \setminus \bigcup_I \mathfrak{p}$. Then S is a saturated multiplicative set. Moreover, if $\mathfrak{p}' \in \operatorname{Ass}_R V(\mathfrak{A})$, then we have:

(*i*) If $\mathfrak{p}' \in I$, then $\mathfrak{p}' \cap S = \emptyset$.

(*ii*) If $\mathfrak{p}' \notin I$, then $\mathfrak{p}' \not\subset \bigcup_I \mathfrak{p}$ (by the Prime Avoidance Lemma 82.16), hence $\mathfrak{p}' \cap S \neq \emptyset$.

In particular, we obtain our second uniqueness result:

Theorem 83.22. *Suppose that the ideal $\mathfrak{A}$ in R has an irredundant primary decomposition $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \ldots, n$ and $I = \{\mathfrak{p}_{i_1}, \ldots, \mathfrak{p}_{i_m}\} \subset \operatorname{Ass}_R V(\mathfrak{A})$ an isolated subset. Then $\bigcap_{j=1}^{m} \mathfrak{Q}_{i_j}$ is independent of an irredundant primary decomposition for $\mathfrak{A}$. In particular, the $\mathfrak{Q}_i$ with $\mathfrak{p}_i \in V(\mathfrak{A})$ minimal are uniquely determined by $\mathfrak{A}$, i.e., the isolated primes determine unique primary ideals.*

PROOF. Let $S = R \setminus \bigcup_{j=1}^{m} \mathfrak{p}_{i_j}$. Then by the above remark,

$$\mathfrak{Q}_{i_1} \cap \cdots \cap \mathfrak{Q}_{i_m} = \varphi_R^{-1}(S^{-1}\mathfrak{A})$$

where $\varphi_R : R \to S^{-1}R$ is the canonical ring homomorphism $r \mapsto r/1$. □

Example 83.23. Let $\mathfrak{A} < R$ be a radical ideal, i.e., $\mathfrak{A} = \sqrt{\mathfrak{A}}$. Suppose that $|\operatorname{Min}(R/\mathfrak{A})|$ is finite. Then $\mathfrak{A}$ has an irredundant primary decomposition and $\operatorname{Ass}_R V(\mathfrak{A}) = \{\mathfrak{p} \in V(\mathfrak{A}) \mid \mathfrak{p} \text{ minimal}\}$, i.e., there are no embedded primes:

We have

$$\mathfrak{A} = \sqrt{\mathfrak{A}} = \bigcap_{V(\mathfrak{A})} \mathfrak{p} = \bigcap_{\mathfrak{p} \in V(\mathfrak{A}) \text{ minimal}} \mathfrak{p}$$

with the right hand side an irredundant primary decomposition of $\mathfrak{A}$.

If R is Noetherian, we have established in Corollary 81.27 and will establish again below that $|\operatorname{Min}(R/\mathfrak{A})|$ is always finite.

We wish to show that every ideal in a Noetherian ring has an irredundant primary decomposition. To do this we shall use Exercise 27.22(13) which asked to show

Lemma 83.24. *An ideal $\mathfrak{C}$ in a commutative ring R is called irreducible if whenever $\mathfrak{C} = \mathfrak{A} \cap \mathfrak{B}$ for some ideals $\mathfrak{A}$ and $\mathfrak{B}$ in R, then either $\mathfrak{C} = \mathfrak{A}$ or $\mathfrak{C} = \mathfrak{B}$. If R is Noetherian, then every ideal $\mathfrak{A} < R$ is a finite intersection of irreducible ideals of R, i.e., $\mathfrak{A} = \mathfrak{C}_1 \cap \cdots \cap \mathfrak{C}_n$, for some irreducible ideals $\mathfrak{C}_i$ in R. We call such a decomposition an irreducible decomposition of $\mathfrak{A}$.*

This is what is needed as:

Lemma 83.25. *Let R be Noetherian ring and $\mathfrak{A} < R$ an irreducible ideal. Then $\mathfrak{A}$ is a primary ideal.*

PROOF. Since $R/\mathfrak{A}$ is also Noetherian, by the Correspondence Theorem, it suffices to show that if (0) is irreducible, then it is primary. So suppose that (0) is irreducible and $xy = 0$, $x, y \in R$, with $y \neq 0$. By the ascending chain condition,

$$(0 : x) \subset (0 : x^2) \subset \cdots \subset (0 : x^n) \subset \cdots$$

stabilizes, say $(0 : x^N) = (0 : x^{N+i})$ for all $i \geq 1$.

Claim. $(0) = (y) \cap (x^N)$.

If we prove the claim, the result will follow as (0) being irreducible imples $x^N = 0$ by the exercise. To show the claim, let $z \in (y) \cap (x^N)$. As $z \in (y)$, we have $xz = 0$ and as $z \in (x^N)$, we have $z = ax^N$ for some

$a \in R$. Consequently, $0 = xz = ax^{N+1}$, so $a \in (0 : x^{N+1}) = (0 : x^N)$. Therefore, $z = ax^N = 0$ as needed. □

Putting all of what we have done together establishes:

Theorem 83.26. (Lasker-Noether Theorem) *Let R be a commutative Noetherian ring and $\mathfrak{A} < R$ an ideal. Then $\mathfrak{A}$ has an irredundant primary decomposition, say*

$$\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n \text{ with } \mathfrak{p}_i = \sqrt{\mathfrak{Q}_i} \text{ for } i = 1, \ldots, n.$$

Moreover,

(1) $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ are unique.
(2) $\mathfrak{Q}_i$ is unique if $\mathfrak{p}_i$ is isolated.

In particular, $|\operatorname{Min}(R/\mathfrak{A})| \leq |\operatorname{Ass}_R \mathfrak{A}|$ is finite.

We derive some consequences.

Corollary 83.27. *Let R be a Noetherian ring. Then $\operatorname{Min}(R)$ is finite.*

Lemma 83.28. *Let R be a Noetherian ring and $\mathfrak{A} < R$ an ideal. Then there exists a positive integer n satisfying $(\sqrt{\mathfrak{A}})^n \subset \mathfrak{A}$.*

PROOF. Since ideals in a Noetherian ring are finitely generated, $\mathfrak{A} = (a_1, \ldots, a_n)$ for some $a_i \in \sqrt{\mathfrak{A}}$. It follows that there exists a positive integer N such that $a_i^N \in \mathfrak{A}$ for all i, so $(\sqrt{\mathfrak{A}})^N \subset \mathfrak{A}$. □

Corollary 83.29. *Let R be a Noetherian ring, then $\operatorname{nil}(R)$ is nilpotent, i.e., there exists a positive integer N such that $\big(\operatorname{nil}(R)\big)^N = 0$.*

Corollary 83.30. *Let R be a Noetherian ring, $\mathfrak{m} \in \operatorname{Max}(R)$, and $\mathfrak{Q} < R$ an ideal. Then the following are equivalent:*

(1) $\mathfrak{Q}$ is $\mathfrak{m}$-primary.
(2) $\sqrt{\mathfrak{Q}} = \mathfrak{m}$.
(3) There exists a positive integer n such that $\mathfrak{m}^n \subset \mathfrak{Q} \subset \mathfrak{m}$.

PROOF. We have shown all but the implication (3) ⇒ (2) which follows from $\mathfrak{m} = \sqrt{\mathfrak{m}^n} = \sqrt{\mathfrak{Q}} = \sqrt{\mathfrak{m}} = \mathfrak{m}$. □

Proposition 83.31. *Let R be a Noetherian ring and $\mathfrak{A} < R$ an ideal. Then*

$$\operatorname{Ass}_R V(\mathfrak{A}) = \{(\mathfrak{A} : x) \mid x \in R\} \cap \operatorname{Spec}(R)$$

PROOF. Replacing R with $R/\mathfrak{A}$, we may assume that $\mathfrak{A} = 0$. Let $\mathfrak{A} = \mathfrak{Q}_1 \cap \cdots \cap \mathfrak{Q}_n$ be an irredundant primary decomposition with $\mathfrak{p}_i = \sqrt{\mathfrak{Q}_i}$ for $i = 1, \ldots, n$. As the decomposition is irredundant, $\mathfrak{A}_i =$

$\bigcap_{j\neq i}\mathfrak{Q}_j > 0$. By the proof of Theorem 83.11, $\mathfrak{p}_i = \sqrt{(0:x)}$ for all nonzero x in $\mathfrak{A}_i$, since $\mathfrak{A}_i \cap \mathfrak{Q}_i = 0$. Hence

$$(0:x) = \operatorname{ann}_R x \subset \mathfrak{p}_i \text{ for all nonzero } x \in \mathfrak{A}_i.$$

As $\mathfrak{Q}_i$ is $\mathfrak{p}_i$-primary, there exists a positive integer m satisfying $\mathfrak{p}_i^m \subset \mathfrak{Q}_i$ by Lemma 83.28. Consequently, $\mathfrak{A}_i\mathfrak{p}_i^m \subset \mathfrak{A}_i \cap \mathfrak{p}_i^m \subset \mathfrak{A}_i \cap \mathfrak{Q}_i = 0$.

Choose $m > 0$ minimal with $\mathfrak{A}_i \cap \mathfrak{p}_i^m = 0$. If $x \in \mathfrak{A}_i\mathfrak{p}_i^{m-1} \subset \mathfrak{A}_i$, then $x\mathfrak{p}_i = 0$, so we conclude that $\mathfrak{p}_i \subset \operatorname{ann}_R x = (0:x)$ if x is nonzero.

Conversely, if $\mathfrak{p} = (0:x)$ is a prime ideal, then $\sqrt{(0:x)} = \sqrt{\operatorname{ann}_R x} = \mathfrak{p}$, and $\mathfrak{p} \in \operatorname{Ass}_R V(0)$. □

Exercises 83.32. 1. Prove Lemma 83.1

2. Prove Lemma 83.4
3. Show that the colon ideal $(\mathfrak{A}:\mathfrak{B})$ of the ideals $\mathfrak{A}$ and $\mathfrak{B}$ in R is an ideal in R and the largest ideal $\mathfrak{C}$ in R satisfying $\mathfrak{B}\mathfrak{C} \subset \mathfrak{A}$.
4. Prove Remark 83.7
5. Prove Lemma 83.17

84. Akizuki and Krull-Akizuki Theorems

As is true in this chapter, all rings are commutative, although straight-forward modifications of the beginning of this chapter can be shown to be true if R is arbitrary.

Recall that a commutative ring is called *Artinian ring* if it satisfies the *descending chain condition*. Similarly, a left R-module is called *Artinian* if submodules of it satisfies the descending chain condition. We wish first with to prove a theorem of Akizuki that characterizes Artinian commutative rings as those that are Noetherian of dimension zero. We have previously noted that the analogue of the Jordan-Holder Theorem for vector spaces holds and used it to show that dimension was a well defined invariant for finite dimensional vector spaces. We expand on these observations.

Definition 84.1. Let R be a commutative ring. An R-module M is said to have a *composition series* if the exists a finite filtration of submodules of M

$$(*) \qquad M = M_0 > M_1 > \cdots > M_n = 0$$

with each quotient M_i/M_{i+1} *irreducible* (or *simple*), i.e., has no proper submodules. We say that a module having a composition series has *finite length* . The composition series in (*) is called a series of *length* n, i.e., the number of *links* in (*).

Just as in the group theoretical case, we have:

Theorem 84.2. (Jordan-Holder Theorem) *Suppose that an R-module M has a composition of length n. Then every composition series of M is of length n. Moreover, if M has a composition series, any proper chain (i.e, all links proper) of M can be refined to a composition series with unique quotients up to order an isomorphism.*

We omit the proof which is similar, but easier than the group theoretic case.

If M is an R-module having a composition series, the common length of a composition series for M is called the *length* of R and denoted by $l(M)$. We set the length of the zero module as zero. If a module M has no composition series, for convenience, we let $l(M) = \infty$ (with $\infty + n = \infty$ for any integer n). An immediate consequence is:

Corollary 84.3. *Let M be an R-module having a composition series and N a submodule of M, Then $l(N) \leq l(M)$ with equality if and only if $N = M$.*

More generally,

Corollary 84.4. *Let*

$$0 \to M' \to M \to M'' \to 0$$

be an exact sequence of R-modules. Then l is additive, i.e.,

$$l(M) = l(M') + l(M''),$$

whose proof we leave as an exercise. For a module to have finite length is a very strong condition. Indeed:

Lemma 84.5. *Let M be an R-module. Then M has finite length if and only if it is both Noetherian and Artinian.*

Proof. If M has finite length, all proper chains of submodules of M have length bounded by the length of M by the Jordan-Hölder Theorem. Conversely, suppose that M is both Noetherian and Artinian. Since it is Noetherian, there exists a maximal proper submodule $M_1 < M$, i.e., M/M_1 is irreducible by the Correspondence Principle. Since M is Noetherian so is M_1. Continuing gives a descending chain of modules $M = M_0 > M_1 > M_2 > \cdots$ with M_i/M_{i+1} irreducible. This must stop, since M is also Artinian. □

An easy consequence of this (that is left as an exercise) is:

Proposition 84.6. *Let V be a vector space over a field K. Then the following are equivalent:*

(1) V is finite dimensional.
(2) V has finite length.
(3) V is Noetherian.
(4) V is Artinian.

Moreover, the dimension of V is just its length.

Corollary 84.7. *Suppose $\mathfrak{m}_1, ..., \mathfrak{m}_n$ are maximal ideals in R (not necessarily distinct) and further that $\mathfrak{m}_1 \cdots \mathfrak{m}_n = 0$. Then R is Noetherian if and only if R is Artinian.*

PROOF. Each $\mathfrak{m}_1 \cdots \mathfrak{m}_i / \mathfrak{m}_1 \cdots \mathfrak{m}_{i+1}$ is an $(R/\mathfrak{m}_{i+1})$-vector space so is Artinian if and only if it is Noetherian. We know that $\mathfrak{m}_1 \cdots \mathfrak{m}_i$ is Noetherian (resp., Artinian) if and only if both of the ideals $\mathfrak{m}_1 \cdots \mathfrak{m}_{i+1}$ and $\mathfrak{m}_1 \cdots \mathfrak{m}_i / \mathfrak{m}_1 \cdots \mathfrak{m}_{i+1}$ are. (Cf. Proposition 37.4 and Exercise 37.11(4).) The result now follows, since

$$R \supset \mathfrak{m}_1 \supset \mathfrak{m}_1\mathfrak{m}_2 \supset \cdots \supset \mathfrak{m}_1 \cdots \mathfrak{m}_n = 0. \qquad \square$$

Proposition 84.8. *Let R be a nonzero commutative Artinian ring. Then R is semi-local of dimension zero. In particular,* $\operatorname{rad}(R) = \operatorname{nil}(R)$.

PROOF. Let $\mathfrak{p}$ be a prime ideal of R. Then $R/\mathfrak{p}$ is an Artinian domain. Thus to show that $\dim R$ is zero, i.e., $\operatorname{Spec}(R) = \operatorname{Max}(R)$, we need only show that any Artinian domain is a field. But if R is an Artinian domain and x nonzero in R, then the descending chain

$$Rx \supset Rx^2 \supset Rx^3 \supset \cdots$$

must stabilize. Hence $Rx^n = Rx^{n+1}$ for some n. In particular, $yx^{n+1} = x^n$ for some y in the domain R. Thus $yx = 1$ and R is a field.

Now suppose that R is an arbitrary Artinian ring. By the Minimal Principle, there exists a minimal element in the set of finite intersections of maximal ideals,

$$\{\mathfrak{m}_1 \cap \cdots \cap \mathfrak{m}_n \mid \mathfrak{m}_1, \ldots, \mathfrak{m}_m \in \operatorname{Max}(R), \text{ some } m\}.$$

Let $\mathfrak{m}_1 \cap \cdots \cap \mathfrak{m}_n$ be such a minimal element. If $\mathfrak{m}$ is a maximal ideal, then by minimality

$$\mathfrak{m}_1 \cap \cdots \cap \mathfrak{m}_n = \mathfrak{m} \cap \mathfrak{m}_1 \cap \cdots \cap \mathfrak{m}_n \subset \mathfrak{m}.$$

It follows that $\mathfrak{m} = \mathfrak{m}_i$, for some i, hence $\operatorname{Max}(R) = \{\mathfrak{m}_1, \ldots, \mathfrak{m}_n\}$. $\square$

Theorem 84.9. (Akizuki) *Let R be a commutative ring. Then the following are equivalent.*

(1) R is Artinian.
(2) R is Noetherian of dimension zero.
(3) Every finitely generated R-module has finite length.

PROOF. Every module over an Artinian (resp., Noetherian) ring R is Artinian (resp., Noetherian), since it is a quotient of R^n for some n. Thus by the above, we know that R satisfies both (1) and (2) if and only if R satisfies (3). So we need only show that R satisfies (1) if and only if R satisfies (2).

Suppose that (1) holds, i.e., that R is Artinian. Then R is semi-local of dimension zero. Let $\mathrm{Max}(R) = \{\mathfrak{m}_1, ..., \mathfrak{m}_n\}$. By the corollary above, it suffices to show that $0 = \prod_{1=1}^{n} \mathfrak{m}_i^k$ for some k. But $\prod_{i=1}^{n} \mathfrak{m}_i \subset \bigcap_{1=1}^{n} \mathfrak{m}_i = \mathrm{rad}(R) = \mathrm{nil}(R)$ by the proposition. So it suffices to prove the following:

Claim 84.10. *If R is Artinian, then* $\mathrm{nil}(R)$ *is nilpotent:*

By the descending chain condition, $\big(\mathrm{nil}(R)\big)^k = \big(\mathrm{nil}(R)\big)^{k+i}$ for some k and all i. Let $\mathfrak{A} := \big(\mathrm{nil}(R)\big)^k$. We must show that $\mathfrak{A} = 0$. Suppose not. Let

$$\mathcal{S} = \{\mathfrak{B} < R \,|\, \mathfrak{B} \text{ an ideal of } R \text{ with } \mathfrak{A}\mathfrak{B} \neq 0\}.$$

Since $\mathfrak{A} \in \mathcal{S}$ and R is Artinian, there exists an ideal $\mathfrak{B}$ that is minimal in $\mathcal{S}$. In particular, there exists an x in $\mathfrak{B}$ such that $x\mathfrak{A} \neq 0$. By minimality, $\mathfrak{B} = Rx$. Since $(x\mathfrak{A})\mathfrak{A} = x\mathfrak{A}^2 = x\mathfrak{A} \neq 0$, we also have $x\mathfrak{A} = \mathfrak{B}$ by minimality. Choose $y \in \mathfrak{A}$ such that $x = xy$. Then $x = xy^N$ for all positive integers N. But $y \in \mathfrak{A} \subset \mathrm{nil}(R)$, so $y^N = 0$ for some N and hence $x = 0$ also. This is a contradiction. Thus $\mathfrak{A} = 0$ and the Claim is established.

Now suppose that R is Noetherian of dimension zero. By the corollary above, it suffices to show that there exist maximal ideals $\mathfrak{m}_1, ..., \mathfrak{m}_n$ in R, not necessarily distinct, so that $0 = \mathfrak{m}_1 \cdots \mathfrak{m}_n$. Since R is Noetherian the zero ideal contains a finite product of prime ideals. (Cf. Exercise 27.22(12).) Since $\dim R = 0$, these prime ideals are maximal. The result follows. □

Corollary 84.11. *Let R be a domain. Then the following are equivalent:*

(1) R is Noetherian of dimension at most one.
(2) If $0 < \mathfrak{A} < R$ is a ideal then $R/\mathfrak{A}$ has finite length.
(3) If $0 < \mathfrak{A} < R$ is a ideal then $R/\mathfrak{A}$ is Artinian.

PROOF. If (1) holds and $0 < \mathfrak{A} < R$ is a ideal then $R/\mathfrak{A}$ is Noetherian of dimension zero so (2) holds. Clearly, (2) implies (3), so we need only show that (3) implies (1).

If (3) holds then $R/\mathfrak{A}$ is a Noetherian ring of dimension zero for any ideal $0 < \mathfrak{A} < R$. In particular, if $x \in \mathfrak{A}$ is nonzero, then $\mathfrak{A}/Rx$ is a

finitely generated ideal in R/Rx. It follows that $\mathfrak{A}$ is finitely generated as an ideal in R, i.e., R is Noetherian. If $0 < \mathfrak{p}_1 \subset \mathfrak{p}_2$ is a chain of primes in R then $\mathfrak{p}_2/\mathfrak{p}_1 = 0$ in the Artinian domain, hence field, $R/\mathfrak{p}_1$. Thus R has dimension at most one. □

Lemma 84.12. *Let M be a non-trivial R-module and $\mathfrak{p}$ a prime ideal in R containing $\operatorname{ann}_R(M)$ and a minimal such prime ideal, i.e., no prime ideal containing $\operatorname{ann}_R(M)$ properly lies in $\mathfrak{p}$. Then $\mathfrak{p}$ consists of zero divisors of M. In particular,*

$$\bigcup_{\operatorname{Min}(R)} \mathfrak{p} \subset \operatorname{zd}(R).$$

[There is no Noetherian condition or finite generation condition.]

PROOF. Let S be the multiplicative set in R defined by

$$S := \{ab \,|\, a \in R \setminus \mathfrak{p} \text{ and } b \in R \setminus \operatorname{zd}(M)\}.$$

Claim 84.13. $S \cap \operatorname{ann}_R(M) = \emptyset$.

Suppose not. Then there exists an a in $R \setminus \mathfrak{p}$ and b in $R \setminus \operatorname{zd}(M)$ satisfying $abM = 0$. Since b is not a zero divisor on M, we have $aM = 0$, and hence $a \in \operatorname{ann}_R(M) \subset \mathfrak{p}$, a contradiction. This establishes the Claim.

Thus there exists a prime $\mathfrak{P}$ containing $\operatorname{ann}_R(M)$ such that $\mathfrak{P}$ excludes S and is maximal with respect to this property. Since 1 lies in R but not in $\mathfrak{p}$ or $\operatorname{zd}(M)$ and $S = \big(R \setminus \operatorname{zd}(M)\big) \cdot (R \setminus \mathfrak{p})$, we have

$$S \supset \big(R \setminus \operatorname{zd}(M)\big) \text{ and } S \supset (R \setminus \mathfrak{p}).$$

Therefore,

$$\mathfrak{P} \subset \operatorname{zd}(M) \cap \mathfrak{p} \subset \mathfrak{p}.$$

The minimality condition on $\mathfrak{p}$ implies that $\mathfrak{p} = \mathfrak{P}$, so $\mathfrak{p} \subset \operatorname{zd}(M)$ as desired. For the last statement, let $M = R$. Then every prime contains $0 = \operatorname{ann}_R(R)$. It follows from the first part that if $\mathfrak{p}$ is a minimal prime ideal in R, then $\mathfrak{p}$ consists of zero divisors of R as $\operatorname{ann}_R(R) = 0$. □

We need the lemma in the following special case.

Corollary 84.14. *If $\dim R = 0$ then $R^\times = R \setminus \operatorname{zd}(R)$.*

PROOF. We have

$$(*) \qquad \operatorname{zd}(R) \supset \bigcup_{\operatorname{Min}(R)} \mathfrak{p} = \bigcup_{\operatorname{Spec}(R)} \mathfrak{p} = \bigcup_{\operatorname{Max}(R)} \mathfrak{m}.$$

Since $R^\times = R \setminus \bigcup_{\operatorname{Max}(R)} \mathfrak{m}$, in the above (*) is an equality. □

Lemma 84.15. *Let R be a Noetherian domain of dimension one. Let a and c be non-zero elements of R. Let*

$$\mathfrak{A} = \bigcup_{n=0}^{\infty} (Rc : Ra^n) := \{x \in R \mid xa^n \in Rc \text{ for some integer } n\}.$$

Then

$$\mathfrak{A} + Ra = R.$$

PROOF. Let $\mathfrak{A}_k = (Rc : Ra^k) := \{x \in R \,|\, xa^k \in Rc\}$. Since $\mathfrak{A}_k \subset \mathfrak{A}_{k+1}$, for all k, we know that $\mathfrak{A}$ is an ideal. Since R is Noetherian, there exists an integer n such that $\mathfrak{A}_n = \mathfrak{A}_{n+i} = \mathfrak{A}$ for all positive integers i. As c lies in $\mathfrak{A}_k$ for all k, we have $c \in \mathfrak{A}$. In particular, $\mathfrak{A}$ is not trivial. Let $^- : R \to R/\mathfrak{A}$ be the canonical epimorphism. Since $\mathfrak{A} > 0$ and R is a domain, it is clear that $\dim R/\mathfrak{A} = 0$. (We do not need Akizuki's Theorem.) By the corollary above, it suffices to establish the following:

Claim 84.16. *$\overline{a}$ is not a zero divisor in $R/\mathfrak{A}$:*

If this were false, then there would exist a y in $R \setminus \mathfrak{A}$ satisfying $\overline{ay} = 0$, i.e., $ay \in \mathfrak{A} = \mathfrak{A}_n$. We would then have $(ay)a^n \in Rc$ and that would imply that $y \in \mathfrak{A}_{n+1} = \mathfrak{A}_n = \mathfrak{A}$, a contradiction. This establishes the Claim. □

Theorem 84.17. (Krull-Akizuki Theorem) *Let A be a Noetherian domain of dimension one. Let F be the quotient field of A and let K/F be a finite field extension. Let B be a ring such that $A \subset B \subset K$. Then B is a Noetherian domain of dimension at most one and if $\mathfrak{B}$ is a non-trivial ideal of B, then $B/\mathfrak{B}$ is a finitely generated $\big(A/(\mathfrak{B} \cap A)\big)$-module of finite length.*

PROOF. We first make two reductions.

Reduction 1. We may assume that $F = K$:

Certainly, we may assume that K is the quotient field of B and, in fact, $K = F(x_1, \ldots, x_n)$ for some $x_i \in B$. Choose $0 \neq c \in A$ such that each cx_i is integral over A. Let $C = A[cx_1, ..., cx_n]$. Then C is integral over A and a finitely generated A-module. Thus C is a Noetherian domain of dimension one with quotient field K and contained in B. If $\mathfrak{C}$ is an ideal in C then $C/\mathfrak{C}$ is integral over $A/(\mathfrak{C} \cap A)$ and a finitely generated $\big(A/(\mathfrak{C} \cap A)\big)$-module. So $C/\mathfrak{C}$ and $A/(\mathfrak{C} \cap A)$ are Noetherian rings of the same dimension. In particular, one is Artinian if and only if the other is. Consequently, if $\mathfrak{B}$ is a non-trivial ideal in B then by clearing denominators and multiplying by an appropriate power of c,

we see that $B \cap C$ is a non-trivial ideal in C. In particular, $B/\mathfrak{B}$ is a finitely generated $\big(A/(\mathfrak{B} \cap A)\big)$-module if it is a finitely generated $\big(C/(\mathfrak{B} \cap C)\big)$-module and has finite length as an $\big(A/(\mathfrak{B} \cap A)\big)$-module if it has finite length as a $\big(C/(\mathfrak{B} \cap C)\big)$-module. This completes the reduction.

Reduction 2. It suffices to show that the (A/Aa)-module B/Ba is finitely generated for any $0 \neq a \in A$:

To show that B is Noetherian of dimension at most one, it suffices, by Corollary 84.11 to Akizuki's Theorem, to show that $B/\mathfrak{B}$ is Artinian for any nonzero ideal $\mathfrak{B}$ of B. Let $0 < \mathfrak{B}$ be an ideal of B. We must also show that $B/\mathfrak{B}$ has finite length over $A/(\mathfrak{B} \cap A)$.

Claim. $0 < \mathfrak{B} \cap A$:

Let $\mathfrak{B}'$ be any non-trivial finitely generated A-submodule of $\mathfrak{B}$. Then there exists $0 \neq c \in A$ such that $c\mathfrak{B}'$ lies in the domain A by the first reduction. This establishes the Claim.

Let $0 \neq a$ lie in $\mathfrak{B} \cap A$. Then A/Aa is Artinian by Corollary 84.11 to Akizuki's Theorem. By assumption, B/Ba is a finitely generated (A/Aa)-module. Since $B/\mathfrak{B}$ is a cyclic (B/Ba)-module, it is also finitely generated as an (A/Aa)-module, hence has finite length over the Artinian ring A/Aa so also over the Artinian ring $A/(\mathfrak{B} \cap A)$. This also implies that $B/\mathfrak{B}$ is Artinian.

So to finish we are in the following situation. We have $0 \neq a \in A$ is fixed, and we must show that B/Ba is finitely generated as an (A/Aa)-module. We do this in a number of steps. Note, as before, that A/Aa is an Artinian ring.

Step 1. Let $x \in B\,(\subset F)$. Then there exists a positive integer n such that $x \in Aa^{-n} + Ba$:

Write $x = \dfrac{b}{c}$ with $b, c \in A$ and $c \neq 0$. Set

$$\mathfrak{B} = \bigcup_{n=0}^{\infty} (Ac : Aa^n) := \{y \in A \,|\, ya^n \in Ac, \text{ some } n\}.$$

By Lemma 84.15, we have $\mathfrak{B} + Aa = A$ so $1 = y + za$, some $y \in \mathfrak{B}$ and $z \in A$. Consequently, $x = yx + zax$. Since $y \in \mathfrak{B}$, by definition, there exists an integer n such that $ya^n \in Ac$, so

$$x = yx + zax = \frac{ya^n}{a^n}\frac{b}{c} + zax \text{ lies in } Aa^{-n} + Ba$$

as needed.

Step 2. Let $\mathfrak{A}_n := (Ba^n \cap A) + Aa$, an ideal of A. Then there exists a positive integer m such that $\mathfrak{A}_m = \mathfrak{A}_{m+i}$ for all positive integers i:

Each $\mathfrak{A}_n$ contains a so is non-trivial. Moreover, it is clear that the $\mathfrak{A}_n$ form a descending chain of ideals. Since A/Aa is Artinian, the descending chain of ideals

$$\cdots \supset \mathfrak{A}_n/Aa \supset \mathfrak{A}_{n+1}/Aa \supset \cdots$$

stabilizes, hence so does the corresponding chain of $\mathfrak{A}_n$'s.

Step 3. Let m be the integer in Step 2. Then $B \subset Aa^{-m} + Ba$:

Let $x \in B$ be fixed. Then by Step 1, there exists a minimal positive integer n so that $x \in Aa^{-n}+Ba$. If we show that $m \geq n$ then $a^m \in Aa^n$ and $Aa^{-n} \subset Aa^{-m}$ as needed. So we may assume that $n > m$. Write $x = ra^{-n} + ba$, with $r \in A$ and $b \in B$. Then

$$r = a^n(x - ba) \in Ba^n \cap A \subset (Ba^n \cap A) + Aa = \mathfrak{A}_n$$

and $\mathfrak{A}_n = \mathfrak{A}_{n+1} = \mathfrak{A}_m$. Hence $r = b_1a^{n+1} + r_1a$, for some $r_1 \in A$ and $b_1 \in B$ so $x = ra^{-n} + ba = (b_1a^{n+1} + r_1a)a^{-n} + ba$ lies in $Aa^{-n+1} + Ba$. This contradicts the minimality of n, so completes the step.

Step 4. B/Ba is a finitely generated (A/Aa)-module (and hence we are done):

By Step 3, we know that $B/Ba \subset (Aa^{-m} + Ba)/Ba$. Moreover, we know that the A-module $(Aa^{-m} + Ba)/Ba \cong Aa^{-m}/(Aa^{-m} \cap Ba)$ is cyclic, hence Noetherian. Thus B/Ba is a finitely generated A-module as needed. □

Corollary 84.18. *Let A be a Dedekind domain with quotient field F. If K/F is a finite field extension then A_K is also a Dedekind domain.*

Remarks 84.19. 1. If $K = F$ in the theorem then clearly F is the only field between A and F, i.e., the only such B with $\dim B = 0$ is F.

2. If $\mathfrak{B}$ is zero and $B = F$ then F is not a finitely generated A-module when $A < F$.

Exercises 84.20. 1. Prove Corollary 84.4

2. Prove Proposition 84.6

3. Let R be a commutative Artinian ring and M a finitely generated free R-module. Let N be a submodule of M that is also R-free. Show that $\operatorname{rank} N \leq \operatorname{rank} M$.

4. Let R be a commutative ring and M a finitely generated free R-module. Let N be a submodule of M that is also R-free. Show that $\operatorname{rank} N \leq \operatorname{rank} M$.

5. Why does it suffice to prove Claim 84.16.

6. Let A be a domain with quotient field F. Suppose that every ring B with $A \subset B \subset F$ is Noetherian. Show that $\dim A = 1$.

85. Affine Algebras

Throughout this section F will denote a field and R a commutative ring.

Recall a finitely generated commutative F-algebra, with F a field, is called an *affine F-algebra* . By the Hilbert Basis Theorem, such an algebra is a Noetherian ring. We wish to investigate the dimension theory attached to such algebras. We know that an integral extension of rings preserves dimension. As an affine F algebra A is isomorphic to a quotient of a polynomial ring $F[t_1, \ldots, t_n]$ for some n, it seems reasonable to assume that if A is also a domain, then the transcendence degree of $qf(A)/F$ should play a crucial role in such an investigation. A fundamental theorem (whose proof we give is due to Nagata) about affine F-algebras bringing together these two observations is the following:

Theorem 85.1. (Noether Normalization Theorem) *Let F be a field and $A = F[x_1, \ldots, x_n]$ an affine F-algebra that is also a domain. Let $r = \operatorname{tr\,deg}_F qfA$. Then there exist $y_1, \ldots, y_r$ in A algebraically independent over F satisfying $A/F[y_1, \ldots, y_r]$ is integral. In particular, A is a finitely generated $F[y_1, \ldots, y_r]$-module. Let $\varphi : F[t_1, \ldots, t_r] \to A$ be the evaluation map given by $t_i \mapsto y_i$. Then*

(1) φ is an F-algebra monomorphism.
(2) ${}^a\varphi : \operatorname{Spec}(A) \to \operatorname{Spec}(F[t_1, \ldots, t_r])$ is surjective and has finite fibers.
(3) $\dim A = \dim F[y_1, \ldots, y_r] = \dim F[t_1, \ldots, t_r]$.

Proof. Let $A = F[t_1, \ldots, t_n]$. If $qf(A)/F$ is a finite field extension, then it is integral. Consequently, we may assume that $qf(A)/F$ is not algebraic. Relabeling, we may also assume that x_1 is transcendental over F. In addition, we may assume that $x_1, \ldots, x_n$ are algebraically dependent over F.

Let $(\mathbf{j}) := (j_1, \ldots, j_n)$ with j_i non-negative integers for $i = 1, \ldots, n$. As the x_i are algebraically dependent, we have an equation

$$(*) \qquad \sum_{(\mathbf{j})} a_{(\mathbf{j})} x_1^{j_1} \cdots x_n^{j_n} = 0$$

for some $a_{(\mathbf{j})} = a_{j_1,\ldots,j_n} \in F$ not all zero. Let $m_2, \ldots, m_n$ be positive integers (to be chosen) and set

$$y_i = x_i - x_1^{m_i} \text{ for } i = 2, \ldots n$$

and $m_1 = 1$. Define

$$(\mathbf{j}) \cdot (\mathbf{m}) := j_1 + j_2 m_2 + \cdots + j_n m_n.$$

Plugging $x_i = y_i + x_1^{m_i}$, $i = 2, \ldots, n$, into (*) yields an equation

$$(**) \qquad \sum a_{(\mathbf{j})} x_1^{(\mathbf{j})\cdot(\mathbf{m})} + f(x_1, y_2, \ldots, y_n) = 0$$

with $f \in F[x_1, y_2, \ldots y_n]$ and having no monomial solely in x_1. In addition, if there is no cancellation in the term $\sum a_{(\mathbf{j})} x_1^{(\mathbf{j})\cdot(\mathbf{m})}$, then then degree in x_1 of the equation in (**) satisfies.

$$(\dagger) \qquad \deg_{x_1} f < \deg_{x_1} \big(\sum a_{(\mathbf{j})} x_1^{(\mathbf{j})\cdot(\mathbf{m})} \big).$$

Choose $d > \max\{j_i \mid a_{(\mathbf{j})} \neq 0\}$. and set

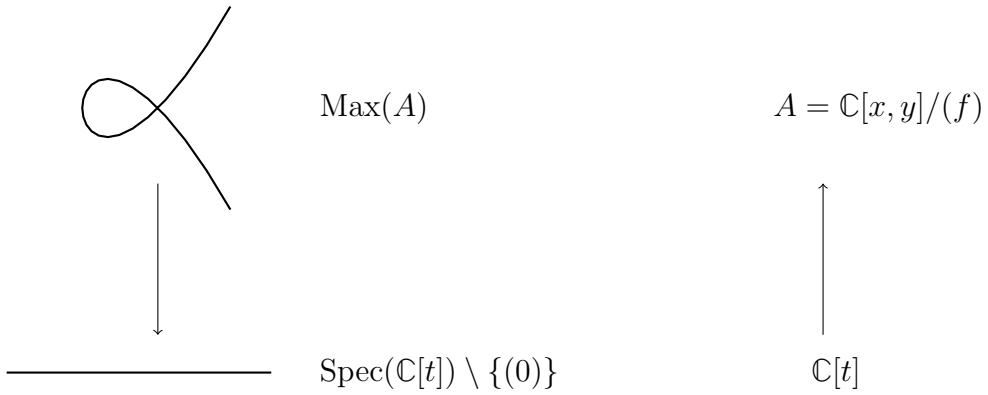
$$(\mathbf{m}) = (1, d, \ldots, d^{n-1}), \ \textit{i.e.}, \ m_i = d^{i-1} \ \text{ for } \ i = 0, \ldots, n-1.$$

As the $(\mathbf{j})$'s are distinct n-tuples, the $(\mathbf{j}) \cdot (\mathbf{m})$'s are all distinct, since they are the d-adic expansions of the $(\mathbf{j}) \cdot (\mathbf{m})$ and x_1 is transcendental over F. Thus (†) holds, so (**) produces an integral equation for x_1 over $F[y_2, \ldots, y_r]$, as the leading coefficient of (**) in x_1 say $a_{(\mathbf{j})}$ is a unit in F. Since $x_i - (x_1^{m_i} + y_i) = 0$ for $i = 2, \ldots, n$ by definition, x_i is integral over $F[x_1, y_2, \ldots, y_n]$ for $i > 1$, hence over $F[y_2, \ldots, y_n]$. Consequently, $F[x_1, x_2, \ldots, x_n]/F[y_2, \ldots, y_n]$ is integral and $F[x_1, x_2, \ldots, x_n]$ is a finitely generated $F[y_2, \ldots, y_n]$-module.

If $y_2, \ldots, y_n$ are algebraically independent over F, we are done. If not we can repeat the process to finish this part of the theorem (as being integral and being a finitely generated module persists in towers).

The other statements in the theorem follow from the above and our previous work, e.g., using Lying Over, as $qf\big(A/F(y_1, \ldots, y_r)\big)$ is a finite algebraic extension and $F[y_1, \ldots, y_r] \cong F[t_1, \ldots, t_r]$ is a normal domain. □

Using the Hilbert Nullstellensatz (cf. the discussion following Lemma 38.14 or another proof of it in 85.9 below), a geometric picture for the Noether Normalization Theorem for a curve when $F = \mathbb{C}$ is given by the following picture, where $f(x, y) = 0$ is the curve in $\mathbb{C}^2$ given by $f = y^2 - x^2(x + 1)$ in $\mathbb{C}[x, y]$ with x, y variables and A is the affine $\mathbb{C}$-algebra $\mathbb{C}[x, y]/(f)$:



Corollary 85.2. *Let F be a field, then*

$$\dim F[t_1, \ldots, t_n] = n = \text{tr deg}_F F(t_1, \ldots, t_n).$$

PROOF. Since $(0) < (t_1) < \cdots < (t_1, \ldots, t_n)$ is a chain of prime ideals in $F[t_1, \ldots, t_n]$, we have $\dim F[t_1, \ldots, t_n] \geq n$. Suppose that $(0) < \mathfrak{p}_1 < \cdots < \mathfrak{p}_m$ is a chain of prime ideals in $F[t_1, \ldots, t_n]$ and $A = F[t_1, \ldots, t_n]/\mathfrak{p}_1$, an affine F-domain, with $^- : F[t_1, \ldots, t_n] \to A$ the canonical epimorphism, so $A = F[\overline{t_1}, \ldots, \overline{t_n}]$. By Noether Normalization, there exist $y_1, \ldots, y_r$ in A, algebraically independent over F satisfying $A/F[y_1, \ldots, y_r]$ is integral. Since $0 < \mathfrak{p}_1$, we must have $\overline{t_1}, \ldots, \overline{t_n}$ are algebraically dependent over F, i.e., $r < n$. By induction, we have

$$\begin{aligned} \dim A &= \dim F[y_1, \ldots, y_r] = \dim F[t_1, \ldots, t_r] \\ &= r = \dim F[t_1, \ldots, t_n]/\mathfrak{p}_1 \geq m - 1, \end{aligned}$$

since $\overline{\mathfrak{p}_1} < \cdots < \overline{\mathfrak{p}_m}$ is a chain of prime ideals in A. Consequently, $n \geq m$, and the result follows. □

Notation 85.3. Let A be an affine F-algebra that is also a domain and $K = qf(A)$. Then there exists a transcendence basis of K over F in A by clearing denominators (or using Noether Normalization). We shall set

$$\text{tr deg}_F A := \text{tr deg}_F K.$$

Corollary 85.4. *Let A be an affine F-algebra that is also a domain. Then* $\dim A = \text{tr deg}_F A$.

PROOF. Let $r' = \text{tr deg}_F A$. By Noether Normalization there exist $y_1, \ldots, y_r$ in A algebraically independent over F with $A/F[y_1, \ldots, y_r]$

integral. Therefore, $qf(A)/F(y_1, \ldots, y_r)$ is finite algebraic, so $r = r'$, and we have

$$\dim A = \dim F[y_1, \ldots, y_r] = \text{tr deg}_F F[y_1, \ldots y_r] = r = \text{tr deg}_F A.$$

□

Corollary 85.5. *Let A be an affine F-algebra and $\mathfrak{p}$ a prime ideal in A. Then* $\dim V(\mathfrak{p}) = \text{tr deg}_F A/\mathfrak{p}$.

PROOF. $\dim V(\mathfrak{p}) = \dim A/\mathfrak{p}$. □

As a consequence of Noether Normalization, we also obtain another proof of Zariski's Lemma.

Corollary 85.6. (Zariski's Lemma) *Let A be an affine F-algebra. If A is a field, then A/F is a finite field extension.*

PROOF. We have $0 = \dim A = \text{tr deg}_F A$, so the finitely generated field extension A/F is algebraic, hence finite. □

Corollary 85.7. *Let $\varphi : A \to B$ be an F-algebra homomorphism of affine F-algebras. Then the restriction of ${}^a\varphi$ to* $\text{Max}(B)$ *satisfies*

$${}^a\varphi|_{\text{Max}(B)} : \text{Max}(B) \to \text{Max}(A).$$

PROOF. Let $\mathfrak{m}$ be a maximal ideal in B, then $\mathfrak{p} = {}^a\varphi(\mathfrak{m})$ is a prime ideal in A and φ induces a ring monomorphism $A/\mathfrak{p} \to B/\mathfrak{m}$. As the quotient of a affine F-algebra is an affine F-algebra, $B/\mathfrak{m}$ is a finite field extension of F by by Zariski's Lemma. Thus $F \subset A/\mathfrak{p} \subset B/\mathfrak{m}$ which implies that $A/\mathfrak{p}$ is a field. Therefore, $\mathfrak{p}$ a maximal ideal in A. □

Recall from §33, if $A = F[t_1, \ldots, t_n]$ and $\mathfrak{A} \subset A$ is an ideal, the *affine variety* of $\mathfrak{A}$ in F^n is defined by

$$Z_F(\mathfrak{A}) = \{\underline{a} = (a_1, \ldots, a_n) \in F^n \mid f(\underline{a}) = 0 \text{ for all } f \in \mathfrak{A}\},$$

the collection of all such forms a system of closed sets for the Zariski topology of F^n. We have seen in Section §38, how Zariski's Lemma implies the (Weak) Hilbert Nullstellensatz):

Theorem 85.8. (Hilbert Nullstellensatz) (Weak Form) *Suppose that F is an algebraically closed field, $A = F[t_1, \ldots, t_n]$, and $\mathfrak{A} = (f_1, \ldots, f_r)$ is an ideal in A. Then $Z_F(\mathfrak{A})$ is the empty set if and only if $\mathfrak{A}$ is the unit ideal. In particular, if $\mathfrak{A} < R$, then there exists a point $\underline{a} \in F^n$ satisfying $f_1(\underline{a}) = 0, \ldots, f_r(\underline{a}) = 0$.*

and using the Rabinowitch Trick, how this implies the (Strong) Hilbert Nullstellensatz):

Theorem 85.9. (Hilbert Nullstellensatz) (Strong Form) *Suppose that F is an algebraically closed field and $A = F[t_1, \ldots, t_n]$. Let $f, f_1, ..., f_r$ be elements in A and $\mathfrak{A} = (f_1, ..., f_r) \subset R$. Suppose that $f(\underline{a}) = 0$ for all $\underline{a} \in Z_F(\mathfrak{A})$. Then there exists an integer m such that $f^m \in \mathfrak{A}$, i.e., $f \in \sqrt{\mathfrak{A}}$. In particular, if $\mathfrak{A}$ is a prime ideal, then $f \in \mathfrak{A}$.*

We give another proof of Strong Form of the Hilbert Nullstellensatz. We begin this proof by reducing it to another form of the theorem.

PROOF. Let $A = F[t_1, \ldots, , t_n]$ and $\mathrm{Max}(A) \subset \mathrm{Spec}(A)$ have the induced topology. The Weak Form of the Hilbert Nullstellensatz implies that the map

$$F^n \to \mathrm{Max}(A) \text{ given by } \underline{a} = (a_1, \ldots, a_n) \mapsto \mathfrak{m}_{\underline{a}} := (t_1 - a_1, \ldots, t_n - a_n)$$

is a homeomorphism. Let $\mathfrak{A} < A$ be an ideal. Recall that $\mathcal{I}(V) := \bigcap_V \mathfrak{p}$ if $V \subset \mathrm{Spec}(A)$, so $\sqrt{\mathfrak{A}} = \mathcal{I}\big(V(\mathfrak{A})\big) = \bigcap_{V(\mathfrak{A})} \mathfrak{p}$. Therefore, it suffices to prove that

$$\sqrt{\mathfrak{A}} = \bigcap_{V(\mathfrak{A})} \mathfrak{p} = \bigcap_{V(\mathfrak{A}) \cap \mathrm{Max}(A)} \mathfrak{m}$$

in the above.

In fact, this is independent of the fact that F be algebraically closed and is an algebraic version of the Strong Form of the Hilbert Nullstellensatz that we shall now state and prove.

Theorem 85.10. (Hilbert Nullstellensatz) (Algebraic Form) *Let A be an affine F-algebra and $\mathfrak{A} < A$ an ideal. Then*

$$\sqrt{\mathfrak{A}} = \bigcap_{V(\mathfrak{A}) \cap \mathrm{Max}(A)} \mathfrak{m},$$

i.e., the closed points in $\mathrm{Spec}(A)$ *determines the variety $V(\mathfrak{A})$.*

PROOF. Let $\mathfrak{B} = \bigcap_{V(\mathfrak{A}) \cap \mathrm{Max}(A)} \mathfrak{m}$. Therefore, $\sqrt{\mathfrak{A}} \subset \mathfrak{B}$. Suppose that $\sqrt{\mathfrak{A}} < \mathfrak{B}$. Let $b \in \mathfrak{B} \setminus \sqrt{\mathfrak{A}}$ and $S = \{b^n \mid n \geq 0\}$. Then $S \bigcap \sqrt{\mathfrak{A}} = \emptyset$. Let $\varphi : A \to S^{-1}A$ be the canonical homomorphism given by $a \mapsto \frac{a}{1}$. Since $S^{-1}A = A[b^{-1}]$, it is also an affine F-algebra, hence φ is an F-algebra homomorphism of affine F-algebras, so ${}^a\varphi$ takes maximal ideals to maximal ideals. By choice, $S^{-1}\mathfrak{A} < S^{-1}A$, so there exists a maximal ideal $\mathfrak{n}$ in $V(S^{-1}\mathfrak{A}) \cap \mathrm{Max}(S^{-1}A)$. Since $\varphi(b)$ is a unit in $S^{-1}A$, we know that $\varphi(b) \notin \mathfrak{n}$. We also have $\mathfrak{A} = \varphi^{-1}(S^{-1}\mathfrak{A}) \subset \varphi^{-1}(\mathfrak{n}) = {}^a\varphi(\mathfrak{n})$, so ${}^a\varphi(\mathfrak{n})$ lies in $V(\mathfrak{A}) \cap \mathrm{Max}(A)$. Therefore, we have

$$b \in \mathfrak{B} = \bigcap_{V(\mathfrak{A}) \cap \mathrm{Max}(A)} \mathfrak{m} \subset {}^a\varphi(\mathfrak{n}),$$

hence $\varphi(b) \in \mathfrak{n}$, a contradiction. □

A ring satisfying the conclusion of this form of the Nullstellensatz is called a *Jacobson* (or *Hilbert*) ring.

Corollary 85.11. *Let A be an affine F-algebra. Then* $\mathrm{Max}(A)$ *is a dense subset of* $\mathrm{Spec}(A)$.

Proof. Let $\mathfrak{A} < A$ be an ideal and $U = \mathrm{Spec}(A) \setminus V(\mathfrak{A})$. If $\mathrm{Max}(A) \subset V(\mathfrak{A})$, then we have

$$\mathrm{nil}(A) = \bigcap_{\mathrm{Spec}(A)} \mathfrak{p} = \bigcap_{\mathrm{Max}(A)} \mathfrak{m} = \bigcap_{\mathrm{Max}(A)\cap V(\mathfrak{A})} \mathfrak{m} = \sqrt{\mathfrak{A}}$$

by the Algebraic Form of the Hilbert Nullstellensatz, so $\mathrm{Spec}(A) = V(\sqrt{\mathfrak{A}}) = V(\mathfrak{A})$ and U is empty. The result follows. □

Let A be an affine F-algebra. Then $A \cong F[t_1, \ldots, t_N]/\mathfrak{A}$ for some N and ideal $\mathfrak{A} < F[t_1, \ldots, t_N]$. As $V(\mathfrak{A}) \cong \mathrm{Spec}(F[t_1, \ldots, t_N]/\mathfrak{A}) \cong \mathrm{Spec}(A)$ and A is a a Noetherian ring by the Hilbert Basis Theorem, $\mathrm{Spec}(A)$ is a Noetherian space. Therefore, it decomposes into finitely many indecomposable components, i.e., the irreducible varieties $V(\mathfrak{p})$ with $\mathfrak{p}$ in the finite set $\mathrm{Min}(A)$. Of course, each prime ideal (respectively, minimal prime ideal, maximal prime ideal) $\mathfrak{p}$ in $\mathrm{Spec}(A)$ is isomorphic to $\mathfrak{P}/\mathfrak{A}$ for some prime ideal $\mathfrak{P}$ (and respectively, minimal, maximal) in $V(\mathfrak{A})$. If $\mathfrak{p} \in \mathrm{Min}(A)$, then $\dim V(\mathfrak{p}) = \mathrm{tr\ deg}_F A/\mathfrak{p}$. Different minimal primes can result in different dimensions. Of course, one is interested in determining when all irreducible components have the same dimension, a subject that we shall not pursue. Suppose that $\mathfrak{P}$ is a prime ideal in A. We define $\mathrm{codim}_{\mathrm{Spec}(A)} V(\mathfrak{P}) := \dim A - \dim V(\mathfrak{P})$, the *codimension* of $V(\mathfrak{P})$ in $\mathrm{Spec}(A)$ and if $\mathfrak{p} \subset \mathfrak{P}$ is another prime ideal, hence $V(\mathfrak{P}) \subset V(\mathfrak{p})$, we define $\mathrm{codim}_{V(\mathfrak{p})} V(\mathfrak{P}) := \dim V(\mathfrak{p}) - \dim V(\mathfrak{P})$, the *codimension* of $V(\mathfrak{P})$ in $V(\mathfrak{p})$. We wish to show in all cases that

$$\mathrm{ht}\,\mathfrak{p} = \mathrm{codim}_{V(\mathfrak{p})} V(\mathfrak{P}) = \mathrm{tr\ deg}_F A/\mathfrak{p} - \mathrm{tr\ deg}_F A/\mathfrak{P}.$$

In particular,

$$\dim A/\mathfrak{P} + \mathrm{ht}\,\mathfrak{P} = \dim A = \dim V(\mathfrak{P}) + \mathrm{ht}\,\mathfrak{P}.$$

Of course, if height is to be codimension, we would also want if $\mathfrak{p} \subset \mathfrak{P}$ are prime ideals in A that

$$\mathrm{ht}\,\mathfrak{P} = \mathrm{ht}\,\mathfrak{p} + \mathrm{ht}\,\mathfrak{P}/\mathfrak{p}.$$

Intuitively, still more should be expected. If $\mathfrak{B} = (f_1, \ldots, f_n)$ is an ideal in A, geometrically, we would want to view the f_i as hypersurfaces in $V(\mathfrak{B})$, e.g., if $A = F[t_1, \ldots, t_N]$, then the f_i are polynomials in $F[t_1, \ldots, t_N]$, and we can view this as the hypersurface $f_i = 0$ in F^N.

Taking our cue from linear algebra, we would then expect if $V(\mathfrak{p})$ is an irreducible component of $V(\mathfrak{B})$ that

$$\dim V(\mathfrak{p}) \geq \dim A - n,$$

i.e., $\operatorname{codim}_{\operatorname{Spec}(A)} V(\mathfrak{p}) \leq n$ or if height is the same as codimension that

$$\operatorname{ht} \mathfrak{p} \leq n.$$

We shall, in fact, show all of this to be true for an affine F-algebra. Note that if F is algebraically closed, the above still will hold with $Z(\mathfrak{A})$ replacing $V(\mathfrak{A})$, etc. It will therefore follow that the algebraic notions of Krull dimension and height translate into the correct geometric notions.

Most of the above will follow by our next theorem. We first make the following definition:

Definition 85.12. A chain of prime ideals

$$\mathfrak{p}_0 < \mathfrak{p}_1 < \cdots < \mathfrak{p}_m = \mathfrak{p}$$

in a commutative ring is called *saturated* if no further new primes can be added to this chain.

Theorem 85.13. *Let F be a field and A be an affine F-algebra that is a domain, $\mathfrak{p}$ a prime ideal in A, and*

$$0 < \mathfrak{p}_1 < \cdots < \mathfrak{p}_m = \mathfrak{p}$$

a saturated chain of prime ideals in A. Then

(1) $\operatorname{ht} \mathfrak{p} = m$.
(2) $\dim A = \dim A/\mathfrak{p} + \operatorname{ht} \mathfrak{p}$.

Moreover, $\operatorname{ht} \mathfrak{p} = \operatorname{codim}_{\operatorname{Spec}(A)} V(\mathfrak{p})$ *and all maximal saturated chains of prime ideals in $\mathfrak{p}$ (hence starting from 0 and ending at $\mathfrak{p}$) have the same number of links. In particular,* $\dim A = \operatorname{ht} \mathfrak{m}$ *for all maximal ideals in A.*

Proof. Clearly, we need only prove (1) and (2). Let $n = \dim A$, finite by Noether Normalization, and $A = F[x_1, \ldots, x_r]$.

Case 1. $m = 1$, then (1) and (2) hold.

As $0 < \mathfrak{p}$ is a saturate chain of prime ideals, $\operatorname{ht} \mathfrak{p} = 1$, so (1) holds.

Subcase 1. $n = r$.

Since $\dim A = \operatorname{tr} \deg_F A$, the elements $x_1, \ldots, x_n$ are algebraically independent over F. Therefore, $A \cong F[t_1, \ldots, t_n]$ is a UFD. In particular, $\mathfrak{p}$ must contain a prime element f, so we must have $\operatorname{ht} \mathfrak{p} = 1$ and $\mathfrak{p} = (f)$. Since f is nonzero, there exists an i such that x_i occurs non-trivially in f, say $i = n$. Let $\bar{} : A \to A/\mathfrak{p}$ be the canonical epimorphism, so $\overline{f(x_1, \ldots, x_n)} = f(\overline{x_1}, \ldots, \overline{x_n}) = 0$ in the domain

$\overline{A} = F[\overline{x_1}, \ldots, \overline{x_n}]$, i.e., $\overline{x_1}, \ldots, \overline{x_n}$ are algebraically dependent over F. Therefore, $\text{tr deg}_F \overline{A} = \text{tr deg} F[\overline{x_1}, \ldots, \overline{x_n}] < n$. If $g \in A$ satisfies $\overline{g} = g(\overline{x_1}, \ldots, \overline{x_n}) = 0$, then $g \in \ker{}^{-} = (f)$, so $f \mid g$ in A. In particular, $g = 0$ or x_n occurs non-trivially in g. In particular, if $h \in F[x_1, \ldots, x_{n-1}] \cong F[t_1, \ldots, t_{n-1}]$ satisfies $h(\overline{x_1}, \ldots, \overline{x_{n-1}}) = 0$, then $h = 0$. It follows that $\overline{x_1}, \ldots, \overline{x_{n-1}}$ are algebraically independent over F, so $\dim \overline{A} = \text{tr deg}_F \overline{A} = n-1$ and $\dim A = \dim \overline{A} + 1 = \dim \overline{A} + \text{ht}\, \mathfrak{p}$.

Subcase 2. $n \neq r$.

By Noether Normalization there exist $y_1, \ldots, y_n$ in A algebraically independent over F satisfying $A/F[y_1, \ldots, y_n]$ is integral. Let $R = F[y_1, \ldots, y_n]$ and $\mathfrak{p}' = \mathfrak{p} \cap R$, a prime ideal in R. We have

$$\dim A = \dim R \text{ and } \text{ht}\, \mathfrak{p} = \text{ht}\, \mathfrak{p}',$$

the second equality by Going Down as R is a normal domain. We also have the monomorphism $R/\mathfrak{p}' \to A/\mathfrak{p}$ is integral, hence $\dim R/\mathfrak{p}' = \dim A/\mathfrak{p}$. By Subcase 1 applied to R, we have

$$\dim A = \dim R = \dim R/\mathfrak{p}' + \text{ht}\, \mathfrak{p}' = \dim A/\mathfrak{p} + \text{ht}\, \mathfrak{p} = \dim(A/\mathfrak{p}) + 1,$$

which is (2).

Case 2. Suppose that $\mathfrak{p}$ is a maximal ideal in A. Then $\text{ht}\, \mathfrak{p} = m = \dim A$.

We have a saturated chain

$$0 < \mathfrak{p}_1 < \cdots < \mathfrak{p}_m = \mathfrak{p}.$$

We show that $m = \text{ht}\, \mathfrak{p} = \dim A$ by induction on $\dim A$. If $\dim A = 1$, the result is immediate, so we may assume that $\dim A > 1$. By Noether Normalization, there exist $y_1, \ldots, y_n$ in A algebraically independent over F satisfying $A/F[y_1, \ldots, y_n]$ is integral. Let $R = F[y_1, \ldots, y_n]$, a normal domain, so by Going Down, we have $\text{ht}\, \mathfrak{p} = \text{ht}(\mathfrak{p} \cap R)$. Moreover, the inclusion $R \hookrightarrow A$ is an integral map, so $\mathfrak{p} \cap R$ is a maximal ideal in R. Since $\dim A = \dim R$ and

$$0 < \mathfrak{p}_1 \cap R < \cdots < \mathfrak{p}_m \cap R = \mathfrak{p} \cap R$$

by Comparability, it suffices to show that $m = n$ in the case that $A = R$. So assume that $A = R$. We know, by the definition of $\dim A$, that

$$m \leq \dim A = \text{tr deg}_F A = n.$$

By the argument in Case 1, Subcase 1, we have

$$n - 1 = \dim A/\mathfrak{p}_1 = \text{tr deg}_F A - 1.$$

Let ${}^{-} : A \to A/\mathfrak{p}_1$, the canonical epimorphism. Then we have

$$0 = \overline{\mathfrak{p}_1} < \cdots < \overline{\mathfrak{p}_m} = \overline{\mathfrak{p}}$$

with $\overline{\mathfrak{p}}$ a maximal ideal in $\overline{A}$ by the Correspondence Principle. By induction on n, we have $m-1=n-1$, so $m=n$. This establishes Case 2.

Case 3. For any prime ideal $\mathfrak{p}$ in A, (1) and (2) hold.

Let $\mathfrak{m} \in V(\mathfrak{p}) \cap \operatorname{Max}(A)$. By Case 2, every saturated chain of prime ideals in $\mathfrak{m}$ has $n = \operatorname{ht}\mathfrak{m} = \dim A$ links. Extend the saturated chain

$$0 < \mathfrak{p}_1 < \cdots < \mathfrak{p}_m = \mathfrak{p}$$

in A to a saturated chain of prime ideals

$$0 < \mathfrak{p}_1 < \cdots < \mathfrak{p}_m < \cdots < \mathfrak{p}_n = \mathfrak{m}$$

in A using the fact that A is Noetherian. Since $\mathfrak{m}/\mathfrak{p}_m$ is a maximal ideal in $A/\mathfrak{p}_m$, by Case 2, we have

$$\operatorname{ht}\mathfrak{m}/\mathfrak{p}_m = \dim A/\mathfrak{p}_m = n - m.$$

Clearly,

$$\dim A \geq \dim A/\mathfrak{p}_m + \operatorname{ht}\mathfrak{p}_m \text{ and } \operatorname{ht}\mathfrak{p}_m \geq m,$$

so

$$\begin{aligned}\dim A &\geq \dim A/\mathfrak{p}_m + \operatorname{ht}\mathfrak{p}_m \\ &= (n-m) + \operatorname{ht}\mathfrak{p}_m \geq (n-m) + m = n = \dim A,\end{aligned}$$

hence

$$\dim A = \dim A/\mathfrak{p} + \operatorname{ht}\mathfrak{p} \text{ and } \operatorname{ht}\mathfrak{p} = m$$

as needed. □

A Noetherian ring R is called *catenary* if all prime ideals $\mathfrak{p} < \mathfrak{P}$ in R and saturated chains of prime ideals beginning at $\mathfrak{p}$ and ending at $\mathfrak{P}$ have the same finite number of links and is called *universally catenary* if every finitely generated commutative R-algebra is catenary. It is easy to show the following:

Lemma 85.14. *If R is a catenary ring and $\mathfrak{A} < R$ an ideal, then $R/\mathfrak{A}$ is a catenary ring.*

As every affine F-algebra that is also a domain, e.g., $F[t_1, \ldots, t_n]$, is catenary by the theorem, by the lemma, we conclude that:

Corollary 85.15. *Every field is universally catenary.*

We turn to the last fact that we wish to prove, viz., if $\mathfrak{A} = (f_1, \ldots, f_n)$ is an ideal in an affine F-algebra, then $\operatorname{ht}\mathfrak{p} \leq n$. We prove this in greater generality. This result, due to Krull, is one of the foundational theorems in commutative algebra.

Recall if $\mathfrak{A} < R$ is an ideal, the *height* of $\mathfrak{A}$ is

$$\operatorname{ht}\mathfrak{A} := \inf\{\operatorname{ht}\mathfrak{p} \mid \mathfrak{p} \in V(\mathfrak{A})\} = \inf\{\operatorname{ht}\mathfrak{p} \mid \mathfrak{p} \in V(\mathfrak{A}) \text{ minimal}\}.$$

Theorem 85.16. (Principal Ideal Theorem) *Let R be a commutative Noetherian ring, $\mathfrak{A} < R$ an ideal and $\mathfrak{p} \in V(\mathfrak{A})$ minimal. If $\mathfrak{A} = (x_1, \ldots, x_n)$, then $\operatorname{ht}\mathfrak{p} \leq n$. In particular, $\operatorname{ht}\mathfrak{A} \leq n$.*

PROOF. The crucial result is the case that $n = 1$, hence the theorem's name.

Case 1: $\mathfrak{A} = (x)$ is principal.

If $x = 0$ or even nilpotent, then $\mathfrak{p} \in \operatorname{Min}(R)$ and $\operatorname{ht}\mathfrak{p} = 0$. So we may assume that x is nonzero. Suppose that the result is false, i.e., $\operatorname{ht}\mathfrak{p} > 1$. Then we have a chain of prime ideals in R,

$$\mathfrak{p}_0 < \mathfrak{p}_1 < \mathfrak{p} \text{ with } \mathfrak{p}_0, \mathfrak{p}_1 \notin V(\mathfrak{A}),$$

since $\mathfrak{p} \in V(\mathfrak{A})$ is minimal. We make some reductions. By the Correspondence Principle, we have $\mathfrak{p}/\mathfrak{p}_0 \in V\Big(\big((x) + \mathfrak{p}_0\big)/\mathfrak{p}_0\Big)$ is minimal and $\operatorname{ht}\mathfrak{p}/\mathfrak{p}_0 \geq 2$. Replacing R by $R/\mathfrak{p}_0$, we may assume that R is a domain and $\mathfrak{p}_0 = 0$. Next let $S = R \setminus \mathfrak{p}$, then $R_{\mathfrak{p}} = S^{-1}R$ satisfies $S^{-1}\mathfrak{p} \in V(R_{\mathfrak{p}}x)$ is minimal and $0 < S^{-1}\mathfrak{p}_1 < S^{-1}\mathfrak{p}$. So replacing R by $R_{\mathfrak{p}}$ and changing notation (considerably), we are reduced to R satisfying the following conditions:

(1) R is a Noetherian domain
(2) $(R, \mathfrak{m})$ is a local ring.
(3) $\operatorname{ht}\mathfrak{m} \geq 2$, so there exists a chain of prime ideals $0 < \mathfrak{P} < \mathfrak{m}$ in R.
(4) The element $x \in \mathfrak{m}$ satisfies $V(x) = \{\mathfrak{m}\}$. In particular, $x \notin \mathfrak{P}$.

Let $\mathfrak{A}_i = (\mathfrak{P}R_{\mathfrak{P}})^i \cap R < R$ a finitely generated ideal in R for every non-negative integer i.

Note that $(R_{\mathfrak{P}}, \mathfrak{P}R_{\mathfrak{P}})$ is a Noetherian local domain with $R \subset R_{\mathfrak{P}} \subset qf(R)$.

We have a descending chain of ideals in R:

$$\mathfrak{P} = \mathfrak{A}_1 \supset \mathfrak{A}_2 \supset \cdots \supset \mathfrak{A}_i \supset \cdots .$$

As $\mathfrak{m} \in V(x)$ is minimal, $\operatorname{Spec}\big(R/(x)\big) = \{\mathfrak{m}/(x)\}$, hence $R/(x)$ is a Noetherian ring of dimension zero. By Akizuki's Theorem 84.9, $R/(x)$ is an Artinian ring. Therefore, the descending chain of ideals

$$\big(\mathfrak{A}_1 + (x)\big)/(x) \supset \cdots \supset \big(\mathfrak{A}_i + (x)\big)/(x) \supset \cdots$$

in $R/(x)$ stabilizes, say $\big(\mathfrak{A}_j + (x)\big)/(x) = \big(\mathfrak{A}_{j+n} + (x)\big)/(x)$ for all $n \geq 0$. By the Correspondence Theorem, $\big(\mathfrak{A}_j + (x)\big) = \big(\mathfrak{A}_{j+n} + (x)\big)$.

Let $a \in \mathfrak{A}_j$. Then

$$a = b + rx, \text{ for some } b \in \mathfrak{A}_{j+1}, \text{ and } r \in R.$$

In particular, $rx = a - b \subset \mathfrak{A}_j \subset (\mathfrak{P}R_{\mathfrak{P}})^j \subset R_{\mathfrak{P}}$. As $x \notin \mathfrak{P}$, it must be a unit in $R_{\mathfrak{P}}$, consequently, $r \in (\mathfrak{P}R_{\mathfrak{P}})^j$. It follows that $r \in (\mathfrak{P}R_{\mathfrak{P}})^j \cap R = \mathfrak{A}_j$, i.e.,

$$\mathfrak{A}_j = \mathfrak{A}_{j+1} + \mathfrak{A}_j x \subset \mathfrak{A}_{j+1} + \mathfrak{A}_j \mathfrak{m} \subset \mathfrak{A}_j.$$

Therefore, $\mathfrak{A}_j = \mathfrak{A}_{j+1} + \mathfrak{A}_j \mathfrak{m}$ in the local ring $(R, \mathfrak{m})$ with $\mathfrak{A}_j$ finitely generated. By alternate form of Nakayama's Lemma, Corollary 82.11, this implies that we have $\mathfrak{A}_j = \mathfrak{A}_{j+1}$. Thus,

$$(\mathfrak{P}R_{\mathfrak{P}})^j = \mathfrak{A}_j R_{\mathfrak{P}} = \mathfrak{A}_{j+1} R_{\mathfrak{P}} = (\mathfrak{P}R_{\mathfrak{P}})^{j+1}$$

in the local Noetherian ring $(R_{\mathfrak{P}}, \mathfrak{P}R_{\mathfrak{P}})$. Hence $(\mathfrak{P}R_{\mathfrak{P}})^j = 0$ in the domain $R_{\mathfrak{P}}$, so $\mathfrak{P}R_{\mathfrak{P}} = 0$ and $\mathfrak{P} = 0$, a contradiction. This proves Case 1.

Case 2. General case (the proof is due to Akizuki).

Let $\mathfrak{A} = (a_1, \ldots, a_n) < R$ be an ideal with $\mathfrak{p} \in V(\mathfrak{A})$ minimal. Suppose that $\operatorname{ht} \mathfrak{p} > n$. Choose a prime ideal $\mathfrak{p}_1$ in R with $\mathfrak{p}_1 < \mathfrak{p}$, and $\operatorname{ht} \mathfrak{p}_1 \geq n$. Replacing R by $R_{\mathfrak{p}}$ if necessary, we may assume that $(R, \mathfrak{m})$ is a local ring with $\mathfrak{m} = \mathfrak{p}$, so $V(\mathfrak{A}) = \{\mathfrak{m}\}$. As R is Noetherian, we may assume that $\mathfrak{p}_1 < \mathfrak{m}$ is saturated, i.e., there exists no prime ideal $\mathfrak{q}$ in R with $\mathfrak{p}_1 < \mathfrak{q} < \mathfrak{m}$. (Why?) Since $\mathfrak{p}_1 \notin V(\mathfrak{A})$, there exists an i such that $a_i \notin \mathfrak{p}_1$. We may assume that $a_1 \notin \mathfrak{p}_1$. Therefore, we have

$$\mathfrak{p}_1 < \mathfrak{p}_1 + (a_1) \subset \mathfrak{p} = \mathfrak{m}.$$

Hence $V\big(\mathfrak{p}_1 + (a_1)\big) = \{\mathfrak{m}\}$, as $\mathfrak{p}_1 < \mathfrak{m}$ is saturated. Consequently, we have

$$(*) \qquad \sqrt{\mathfrak{p}_1 + (a_1)} = \bigcap_{V\left(\mathfrak{p}_1 + (a_1)\right)} \mathfrak{P} = \mathfrak{m}.$$

Since $\mathfrak{m}$ is finitely generated, there exists a positive integer k satisfying $\mathfrak{m}^k \subset \mathfrak{p}_1 + (a_1)$ by (*) (and the definition of the radical). As we also have $\mathfrak{A} \subset \mathfrak{m}$, we conclude that $\mathfrak{A}^k \subset \mathfrak{m}^k \subset \mathfrak{p}_1 + (a_1)$. Therefore, we have equations

$$a_i^k = b_i + c_i a_1 \text{ for some } b_i \in \mathfrak{p}_i \text{ and } c_i \in R, \ i = 2, \ldots, n.$$

Let $\mathfrak{B} = (b_2, \ldots, b_n) \subset \mathfrak{p}_1$, so $\mathfrak{p}_1 \in V(\mathfrak{B})$. As $\operatorname{ht} \mathfrak{p}_1 \geq n$, by induction, we conclude that there exists a prime ideal $\mathfrak{q} \in V(\mathfrak{B})$ minimal with $\mathfrak{B} \subset \mathfrak{q} < \mathfrak{p}_1$. However, $a_1^k, \ldots, a_n^k$ all lie in $\mathfrak{B} + (a_1) \subset \mathfrak{q} + (a_1)$, so

$$\{\mathfrak{m}\} = V(\mathfrak{A}) = V(a_1, \ldots, a_n) = V(a_1^k, \ldots a_n^k) \supset V\big(\mathfrak{q} + (a_1)\big).$$

Therefore, $V\big(\mathfrak{q}+(a_1)\big)=\{\mathfrak{m}\}$ and $\mathfrak{m}/\mathfrak{q}\in V\big((\mathfrak{q}+(a_1))/\mathfrak{q}\big)$ is minimal (working in the ring $R/\mathfrak{q}$). By Case 1, we know that $\operatorname{ht}\mathfrak{m}/\mathfrak{q}\leq 1$, which contradicts the existence of the chain of primes $0=\mathfrak{q}/\mathfrak{q}<\mathfrak{p}_1/\mathfrak{q}<\mathfrak{m}/\mathfrak{q}$ in $R/\mathfrak{q}$. □

Since ideals in a Noetherian ring are finitely generated, we have:

Corollary 85.17. *Let R be a Noetherian ring and $\mathfrak{p}$ a prime ideal in R. Then $\operatorname{ht}\mathfrak{p}$ is finite. In particular, the set of prime ideals in R satisfies the descending chain condition.*

Corollary 85.18. *Every local Noetherian ring has finite dimension.*

Corollary 85.19. *Let R be a Noetherian domain. Then R is a UFD if and only if every height one prime in R is principal if and only if the set of height one primes is equal to the set of nonzero principal prime ideals.*

Proof. We know by Kaplansky's Theorem 28.1 that R is a UFD if and only if every nonzero prime ideal in R contains a prime element. As every nonzero prime ideal in R contains a prime ideal of height one by Corollary 85.17, the result follows. □

The next result can be viewed as a converse to the Principal Ideal Theorem.

Corollary 85.20. *Let R be a Noetherian ring, $\mathfrak{A}<R$ an ideal, and $\mathfrak{p}<R$ a prime ideal. Then*

(1) If $\operatorname{ht}\mathfrak{A}=n\geq 1$, then there exist $a_1,\dots,a_n\in\mathfrak{A}$ satisfying $\operatorname{ht}(a_1,\dots,a_i)=i$ for $i=1,\dots,n$.

(2) If $\operatorname{ht}\mathfrak{p}=n\geq 1$, then there exist $a_1,\dots,a_n\in\mathfrak{p}$ satisfying $\mathfrak{p}\in V(a_1,\dots,a_n)$ is minimal and $\operatorname{ht}(a_1,\dots,a_i)=i$ for $i=1,\dots,n$.

Proof. (1): As $\operatorname{ht}\mathfrak{A}\geq 1$ and $\operatorname{Min}(R)$ is finite, there exists an element $a_1\in\mathfrak{A}\setminus\bigcup_{\operatorname{Min}(R)}\mathfrak{p}$ by the Prime Avoidance Lemma. Therefore, $\mathfrak{p}\notin V(a_1)$ for all primes $\mathfrak{p}$ in $\operatorname{Min}(R)$, so $\operatorname{ht}(a_1)>0$. By the Principal Ideal Theorem, $\operatorname{ht}(a_1)\leq 1$, hence $\operatorname{ht}(a_1)=1$. By induction, there exists $a_1,\dots,a_i$ in $\mathfrak{A}$ satisfying $\operatorname{ht}(a_1,\dots,a_i)=i<n$. Let $\mathfrak{p}\in V(a_1,\dots,a_i)$ be minimal. By the Principal Ideal Theorem, $\operatorname{ht}\mathfrak{p}\leq i$ and hence $\operatorname{ht}\mathfrak{p}=i$. As $\{\mathfrak{P}\mid\mathfrak{P}\in V(a_1,\dots,a_i)\,\text{minimal}\}$ is a finite set, there exists an element

$$a_{i+1}\in\mathfrak{A}\setminus\bigcup_{\substack{\mathfrak{P}\in V(a_1,\dots a_i)\\ \text{minimal}}}\mathfrak{P}$$

by the Prime Avoidance Lemma. Let $\mathfrak{B} = (a_1, \ldots, a_{i+1})$ and $\mathfrak{Q} \in V(\mathfrak{B}) \subset V(a_1, \ldots, a_i)$. Then there exists a prime ideal $\mathfrak{P}$ in $V(a_1, \ldots, a_i)$ minimal satisfying $\mathfrak{P} \subset \mathfrak{Q}$. Consequently, $\operatorname{ht}\mathfrak{Q} > \operatorname{ht}\mathfrak{P} = i$. By the Principal Ideal Theorem, $\operatorname{ht}\mathfrak{B} \leq i+1$, hence $\operatorname{ht}\mathfrak{B} = i+1$ and (1) follows by induction.

(2): Choose a_i as in (1) applied to $\mathfrak{A} = \mathfrak{p}$. Since $\mathfrak{p} \in V(A_1, \ldots, a_n)$ satisfies $\operatorname{ht}\mathfrak{p} = \operatorname{ht}(a_1, \ldots, a_n)$, we have $\mathfrak{p} \in V(a_1, \ldots, a_n)$ is minimal. □

The second statement of the last corollary implies if F is an algebraically closed field and Z is a irreducible affine variety of codimension r in F^n, then Z is an irreducible component of an affine variety defined as the intersection of r hypersurfaces.

Remark 85.21. Let $(R, \mathfrak{m})$ be a Noetherian local ring and $a_1, \ldots, a_n$ elements in $\mathfrak{m}$ that satisfy $\operatorname{ht}(a_1, \ldots, a_n) = \operatorname{ht}\mathfrak{m}$. We say that the elements $a_1, \ldots, a_n$ form a *system of parameters* for R. As $V(a_1, \ldots, a_n) = \{\mathfrak{m}\}$, the ideal $\mathfrak{m}/(a_1, \ldots, a_n)$ in $R/(a_1, \ldots, a_n)$ is nilpotent. We have $\mathfrak{m}^k \subset (a_1, \ldots, a_n) \subset \mathfrak{m}$ for some integer k, since $\mathfrak{m}$ is finitely generated. Define $\text{V-}\dim R := \dim_{R/\mathfrak{m}} \mathfrak{m}/\mathfrak{m}^2$. By Corollary 82.12 of Nakayama's Lemma, V-dim is the size of a minimal generating set for $\mathfrak{m}$. By the Principal Ideal Theorem,

$$\dim R = \operatorname{ht}\mathfrak{m} \leq \text{V-}\dim R.$$

If there exists a system of parameters generating $\mathfrak{m}$, such a system of parameters is called a *regular system of parameters* for R and R is called a *regular local ring*.

Example 85.22. Let F be a field and $R = F[[t_1, \ldots, t_n]]$, the formal power series in $t_1, \ldots, t_n$. Then R is a local ring with maximal ideal $\mathfrak{m} = (t_1, \ldots, t_n)$ by Exercise 23.21(3) and is Noetherian by Example 27.13(2). Therefore, it is a regular local ring by Exercise 34.11(2).

By Theorem 34.9, the regular local ring $R = F[[t_1, \ldots, t_n]]$ with F a field is a UFD. More generally, a deeper result says any regular local ring is a domain and a UFD. The proof requires homological algebra and was one of that subjects first deep algebraic applications. The geometric case of this result (i.e., the local ring attached to a nonsingular point of an (affine) variety) can be proven using Theorem 34.9 without homological algebra.

Exercises 85.23. 1. Prove the following generalization of the Normalization Theorem 85.1: Let F be a field and A an affine F-algebra (not necessarily a domain). If

$$\mathfrak{A}_1 < \cdots < \mathfrak{A}_m < A$$

is a chain of ideals in A, then there exists a non-negative integer n and elements $y_1, \ldots, y_n$ in A algebraically independent over F and integers

$$0 \leq h_1 \leq \cdots \leq h_m \leq n$$

satisfying the following:

(a) $F[y_1, \ldots, y_n]$ is integral in A.
(b) A is a finitely generated $F[y_1, \ldots, y_n]$-module.
(c) $\mathfrak{A}_i \cap F[y_1, \ldots, y_n] = (y_1, \ldots, y_{h_i})$ for $i = 1, \ldots m$.

2. Prove Lemma 85.14.
3. Let R be a catenary ring and S a multiplicative set in R. Then $S^{-1}R$ is catenary.
   [Note. If A is an affine F-algebra, this says that localizations of it are catenary. Such a localization may not be an affine F-algebra.]
4. Let R be a Noetherian ring of dimension greater than one. Prove that $\operatorname{Spec}(R)$ is infinite.
5. Prove the following generalization of the Principal Ideal Theorem: Let R be a Noetherian ring, $\mathfrak{A} < R$ an ideal generated by n elements, and $\mathfrak{p}$ an element of $V(\mathfrak{A})$. Suppose that $\operatorname{ht} \mathfrak{p}/\mathfrak{A} = m$ in $R/\mathfrak{A}$. Prove that $\operatorname{ht} \mathfrak{p} \leq m + n$.
6. If F is a field, show that $\text{V-dim}(F[[t_1, \ldots, t_n]]) = n$.

86. Addendum: Japanese Rings

Let A be a Noetherian domain with quotient field F. We say that A is a *Japanese ring* if whenever K/F is a finite extension of fields, the integral closure A_K of A is a finitely generated A-module. We saw in Theorem 73.5 that if A is a Noetherian normal domain with quotient field F and K/F a finite separable extension, then the integral closure A_K of A in K is a finitely generated A-module. In particular, if $\operatorname{char}(F) = 0$ a normal domain A is Japanese. The hypothesis that K/F is separable is a nontrivial condition if $\operatorname{char}(F) = p > 0$. Indeed in the positive characteristic case, A may not be Japanese, even if A is normal. In fact, Schmidt-Nagata showed if F is satisfies $[F : F^p]$ is infinite and

$$\mathcal{S} := \{F_\alpha \mid F^p \subset F_\alpha \subset F \text{ are fields with } F_\alpha/F^p \text{ a finite extension}\}$$

partially ordered by $\subset$ with $\Gamma = \{\alpha \mid F_\alpha \in \mathcal{S}\}$ and $A_\alpha = F_\alpha[[t]]$, the formal power series ring with coefficients in A_α. then A_α is a DVR with maximal ideal A_α. Further let $A := \bigcup_\Gamma A_\alpha$, a ring with the obvious structure. Then it can be shown that A is a discrete valuation ring but is not Japanese, i.e., the result, in general, is false for dimension one

noetherian rings. Of course fields are Japanese. We show even more is true. To investigate inseparable extensions, we need to use properties of purely inseparable extensions (cf. Exercise 50.10(17)). We begin with the following.

Lemma 86.1. *Let A be a Noetherian domain with quotient field F. Suppose that for all finite, purely inseparable field extensions E of F that A_E is a finitely generated A-module. Then A is Japanese.*

PROOF. Let K/F be a finite field extension and L/K a normal closure of K/F. As $A_K \subset A_L$ and A is a Noetherian ring, it suffices to show that A_L is a finitely generated A-module by Theorem 37.6. So we may assume that K/F is a normal extension. Let $E = K^{G(K/F)}$, then E/F is Galois, so separable and E/F is purely inseparable by Exercise 50.10(17). As A_E is a finitely generated A-module by hypothesis, it is a Noetherian normal domain, so A_K is a finitely generated A_E-module by Theorem 73.5. It follows that A_K is a finitely generated A-module. □

A Noetherian domain A is called *universally Japanese* if every domain B that is a finitely generated A-algebra B is Japanese.

Theorem 86.2. *Every field is universally Japanese.*

PROOF. Let A be an affine F-algebra that is also a domain with quotient field K and L/K a finite field extension. We must show that A_L is a finitely generated A-module. By the Noether Normalization Theorem 85.1, there exist $x_1, \ldots, x_n$ in A algebraically independent over F with $F[x_1, \ldots, x_n] \subset A$ integral. As $K/F(x_1, \ldots, x_n)$ is a finite field extension. so is $L/F(x_1, \ldots, x_n)$. Therefore, it suffices to show that A_L is a finitely generated $F[x_1, \ldots, x_n]$-module, i.e., we may assume that $A = F[x_1, \ldots, x_n]$. Since $F[x_1, \ldots, x_n] \cong F[t_1, \ldots, t_n]$, we can assume further that $A = F[t_1, \ldots, t_n]$ and $K = F(t_1, \ldots, t_n)$. Moreover, by Lemma 86.1, we may assume that L/K is purely inseparable. As A is a normal Noetherian domain, we may also assume that $\mathrm{char}(F) = p > 0$.

Since L/K is finite and purely inseparable, there exists a positive integer m satisfying $L^{p^m} \subset K = qf(A)$. Let

$$\begin{aligned} \mathcal{C} := &\text{ the set of the } p^m\text{th roots} \\ &\text{of all the coefficients of all the } f_i, g_i \end{aligned}$$

and

$$\begin{aligned} L &= K(a_1, \ldots, a_n) \text{ with} \\ a_i^{p^m} &= \frac{f_i}{g_i} \text{ with } f_i, g_i, \in A,\ g_i \neq 0 \text{ for all } i \\ \widetilde{F} &= F(\mathcal{C}) = F[\mathcal{C}] \\ E &= \widetilde{F}(t_1^{\frac{1}{p^m}}, \ldots, t_n^{\frac{1}{p^m}}). \end{aligned}$$

We have $K \subset L \subset E$. As A is a Noetherian ring, it suffices to show that A_E is a finitely generated A-module. Let

$$B = \widetilde{F}[t_1^{\frac{1}{p^m}}, \ldots, t_n^{\frac{1}{p^m}}] = F[\mathcal{C}, t_1^{\frac{1}{p^m}}, \ldots, t_n^{\frac{1}{p^m}}].$$

Each of the finitely many elements, $c \in \mathcal{C}, t_1^{\frac{1}{p^m}}, \ldots, t_n^{\frac{1}{p^m}}$, is integral over A, so B is a finitely generated A-module. We also have that $B \subset A_E$ with $qf(B) = E$ and $B = \widetilde{F}[t_1^{\frac{1}{p^m}}, \ldots, t_n^{\frac{1}{p^m}}]$ a UFD, hence normal. Thus $A_E = B$ and the result follows □

Corollary 86.3. *Let F be a field and A an affine F-algebra that is also a domain with $K = qf(A)$. If L/F be a finite field extension, then A_L is an affine F-algebra. In particular, the integral closure of A is a affine F-algebra.*

A noetherian ring A is called a *Nagata ring* if $A/\mathfrak{p}$ is Japanese for every prime ideal $\mathfrak{p}$ in $\mathrm{Spec}(A)$. It is a deep theorem that a noetherian domain A is a Nagata ring (a property internal to A) if and only if it is universally Japanese. In particular, if A is a Dedekind domain of characteristic zero, then A is a Nagata ring as it is Japanese and its quotients determined by maximal ideals are finite fields. Hence any Dedekind domain of characteristic zero is univerally Japanese. This is important in arithmetic algebraic geometry.

CHAPTER XVII

Division and Semisimple Rings

In this chapter, we study the simplest non-commutative rings, viz., division rings. More generally, we study matrix rings over division rings. In the first section, we show that simple left Artinian rings are up to isomorphism matrix rings over division rings and uniquely so. We then study products of such rings. This is important as it is the foundation of the theory of finite group representations that we shall study in the next chapter. We then investigate another way of obtaining division rings important in number theory. Finally, we generalize Wedderburn's Theorem that finite division rings are fields to a theorem of Jacobson, showing that a ring is commutative if for every element x in the ring, there exists an integer n satisfying $x^n = x$.

87. Wedderburn Theory

Modules over a division ring are just vector spaces with a line in a vector space a 'simple' module. As vector spaces are direct sums of lines, a module over a division ring is a direct sum of 'simple' submodules. modules over division rings, i.e., vector spaces are direct sums of lines. In this section, we shall generalizes this to a module over a simple left Artinian ring, i.e., any such is a direct sum of 'simple' submodules. We shall also prove Wedderburn's Theorem that any simple Artinian ring is, in fact, isomorphic to a matrix ring over a division ring. We begin with properties of modules.

Definition 87.1. Let R a nonzero ring. We say

1. A nonzero (left) R-module is *irreducible* (or *simple*) if it contains no proper submodules. An irreducible left ideal in R is also called a *minimal left ideal*.
2. A (left) R-module is called *completely reducible* if every submodule of it is a direct summand of it.
3. The ring R is *(left) semisimple* if it is a completely reducible module over itself.

Of course, we have the analogous definitions for right R-modules. When we wish to consider right modules we shall write the word right.

Examples 87.2. 1. Over a division ring, every vector space is completely reducible, since every vector space is free on a basis.

2. Let D be a division ring. Then the kth column space of $\mathbb{M}_n(D)$,

$$M_k := \{A = (a_{ij}) \in \mathbb{M}_n(D) \mid a_{ij} = 0 \text{ if } j \neq k\},$$

is a minimal left ideal of R and is an n-dimension vector space over D. Moreover, as $\mathbb{M}_n(D)$-modules, we have

$$\mathbb{M}_n(D) = M_1 \oplus \cdots \oplus M_n \text{ and } M_i \cong M_j \text{ for all } i, j.$$

Lemma 87.3. *Let $\widetilde{M}$ be a nonzero R-module and M_i irreducible submodules of $\widetilde{M}$ for i in I (not necessarily non-isomorphic). Let M be the R-module $\sum_I M_i$. Then there exists a subset J of I satisfying M is the direct sum $\bigoplus_J M_j$.*

Proof. Using Zorn's Lemma, we see that there exists a subset J of I maximal such that $N = \sum_J M_j = \bigoplus_J M_j$. Let i_o be an element in $I \setminus J$. Then the R-module M_{i_o} is irreducible and contains the submodule $N \cap M_{i_o}$, so either $N \cap M_{i_o} = M_{i_o}$, i.e., $M_{i_o} \subset N$ or $N \cap M_{i_o} = 0$, i.e., $\sum_{J \cup \{i_o\}} M_j = \bigoplus_{J \cup \{i_o\}} M_j$. By maximality, this second case does not occur, so $M = N$. □

Proposition 87.4. *Let M be a nonzero R-module. Then the following are equivalent:*

(1) M is a sum of irreducible R-submodules.
(2) M is a direct sum of irreducible R-modules.
(3) M is a completely reducible R-module.

Proof. (1) ⇒ (2) follows from the lemma.

(2) ⇒ (3): Let N be a submodule of M and $M = \bigoplus_I M_i$ with every M_i, $i \in I$, irreducible. Using Zorn's Lemma, there exists a maximal subset J of I such that $N + \bigoplus_J M_j = N \oplus \bigoplus_J M_j$. By the argument in the proof of the lemma, this must be M.

(3) ⇒ (1): We prove the following:

Claim 87.5. If M is a completely reducible R-module and M_0 a nonzero submodule of M, then there exists an irreducible submodule of M_0:

Let m be a nonzero element of M_0. Then we have an R-epimorphism

$$\rho_m : R \to Rm \text{ given by } r \mapsto rm.$$

As the the annihilator $\operatorname{ann}_R m$ is a left ideal in R, the map ρ_m induces an R-isomorphism

$$\overline{\rho_m} : M/\operatorname{ann}_R m \to Rm.$$

Using Zorn's Lemma, we know that there is a maximal left ideal $\mathfrak{m}$ in R with $\operatorname{ann}_R m \subset \mathfrak{m}$. By the Correspondence Principle, $\mathfrak{m}m < Rm$ is maximal, hence $Rm/\mathfrak{m}m$ is irreducible. By hypothesis, M is completely reducible, so $M = M' \oplus \mathfrak{m}m$ for some submodule M' of M. Let x be an element of Rm. Then there exists an $r \in \mathfrak{m}$ and an $m' \in M'$ satisfying

$$x = m' + rm \text{ so } m' = x - rm \text{ lies in } M' \cap Rm.$$

It follows that

$$Rm = (M' \cap Rm) \oplus \mathfrak{m}m \subset M_0 \text{ and } M' \cap Rm \cong Rm/\mathfrak{m}m \text{ is irreducible.}$$

This establishes the claim. Now let M_0 be the sum of all the irreducible submodules of M. By the claim $M_0 \neq 0$. As M is completely reducible, $M = M_0 \oplus M_1$ for some submodule M_1. If $M_1 \neq 0$, then applying the claim again shows that M_1 contains an irreducible submodule of M. This contradicts the choice of M_0. □

Corollary 87.6. *Let D be a division ring. Then $\mathbb{M}_n(D)$ is semisimple.*

Remark 87.7. No nonzero ideal in $\mathbb{Z}$ is irreducible, so $\mathbb{Z}$ is not semisimple. If $p > 0$ is a prime in $\mathbb{Z}$, then $\mathbb{Z}/p\mathbb{Z}$ is semisimple as it is a field. More generally, $\mathbb{Z}/n\mathbb{Z}$ with $n > 0$ a product of distinct primes is semisimple. Note that $\mathbb{Z}/n\mathbb{Z}$, $n > 0$ a product of distinct primes, satisfies the descending chain condition, i.e., is an Artinian ring, and has no nonzero nilpotent (left) ideals $\mathfrak{A}$, i.e., (left) ideals $\mathfrak{A}$ such that $\mathfrak{A}^n = 0$ for some positive integer n. Note also that $\mathbb{Z}/n\mathbb{Z}$ with $n > 1$ not square free is Artinian but has nonzero nilpotent elements and is not semi-simple. (Cf. Exercise 88.7(4) below.)

Corollary 87.8. *Let M_i, $i \in I$, be completely reducible R-modules. Then $\coprod_I M_i$ is completely reducible.*

Corollary 87.9. *Let M be a completely reducible R-module, and N a submodule of M. Then N and M/N are completely reducible.*

Proof. We may assume that $N \neq 0$. By Claim 87.5, there exists a submodule N_0 of N that is the sum of all the irreducible submodules of N. As M is completely reducible, $M = N_0 \oplus N_1$ for some submodule N_1. As in the proof of the proposition, we see that $N = N_0 \oplus (N_1 \cap N)$. By Claim 87.5, we must have $N_1 \cap N = 0$, so $N = N_0$ and N is completely reducible. As $M = N \oplus N'$ for some submodule N' of M, N' is completely reducible by what we have just shown, hence so is $M/N \cong N'$. □

Definition 87.10. Let R be a ring. An element e in R is called an *idempotent* if $e^2 = e$. If $e_1, \ldots, e_n$ are idempotents, they are called *orthogonal* if $e_i e_j = \delta_{ij} e_i$ for all i and j.

Examples 87.11. Let R be a ring.

1. The elements 0 and 1 of R are idempotents. They are called the *trivial idempotents* and are orthogonal.
2. If e is an idempotent so is $1-e$, and then $e, 1-e$ are orthogonal idempotents.
3. The sum of any orthogonal idempotents is an idempotent.
4. Let $M = M_1 \oplus M_2$ and $p_i : M \to M$ be the projection $m_1 + m_2 \mapsto m_i$, where $m_i \in M_i$, for $i = 1, 2$. Then p_1, p_2 are orthogonal idempotents in the ring $\mathrm{End}_R(M)$.
5. Let $S = \mathbb{M}_n(R)$ and e_{ij} the matrix with 1 in the ijth entry, zero elsewhere. Then $e_{11}, \ldots, e_{nn}$ are orthogonal idempotents in S. We also note that
$$e_{ii} S e_{ii} \cong R \text{ as rings.}$$

Proposition 87.12. *Let R be a nonzero ring. Then the following are equivalent:*

(1) Every R-module is completely reducible.
(2) Every short exact sequence in R-modules splits.
(3) The ring R is semisimple.
(4) $R = \bigoplus_{i=1}^{n} \mathfrak{A}_i$ for some n and some left ideals $\mathfrak{A}_i$ with each $\mathfrak{A}_i$, $i = 1, \ldots, n$, a minimal left ideal; and, furthermore, $\mathfrak{A}_i = Re_i$, $i = 1, \ldots, n$, with $e_1, \ldots, e_n$ orthogonal idempotents. Moreover if $\mathfrak{A}$ is a minimal left ideal, then $\mathfrak{A} = \mathfrak{A}_i$ for some i. [The $\mathfrak{A}_i$ need not be mutually non-isomorphic.]

Proof. The equivalence of (1) and (2) follows from the corollaries above and (4) $\Rightarrow$ (3) from the proposition. So we need only show (3) $\Rightarrow$ (4).

As R is completely reducible, $R = \bigoplus_I \mathfrak{A}_i$ for some minimal left ideals $\mathfrak{A}_i$. There exist nonzero $e_{i_1}, \ldots, e_{i_n}$ in R for some n with $e_{i_j} \in \mathfrak{A}_{i_j}$ and $1 = e_{i_1} + \cdots + e_{i_n}$. It follows that $R \subset \mathfrak{A}_{i_1} \oplus \cdots \oplus \mathfrak{A}_{i_n}$, hence we must have equality. Changing notation, we may assume that $R = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_n$, with $e_i \in \mathfrak{A}_i$ for each i and $1 = e_1 + \cdots + e_n$. It follows that $e_j = \sum e_i e_j$ in $\mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_n$, hence $e_1, \ldots, e_j$ are orthogonal idempotents. As $0 < Re_j \subset \mathfrak{A}_j$, we have $Re_j = \mathfrak{A}_j$, since $\mathfrak{A}_j$ is irreducible for each j. Finally, if $\mathfrak{A}$ is a minimal left ideal in R, then $\mathfrak{A} \cap \mathfrak{A}_j > 0$ for some j, so $\mathfrak{A} = \mathfrak{A}_j$. □

Let D be a division ring and $R = \mathbb{M}_n(D)$. Then R is a simple ring, i.e., the only 2-sided ideals in R are 0 and R and R is left and right semisimple, left and right Noetherian (i.e, ACC on left and right ideals), left and right Artinian (i.e., DCC on left and right ideals), since

R is a finite dimensional D-vector space. More generally, if $D_1, \ldots, D_r$ are division rings and $A = \mathbb{M}_{n_1}(D_1) \times \cdots \times \mathbb{M}_{n_r}(D_r)$ then A is left and right semisimple, left and right Noetherian, and left and right Artinian.

One problem that arises when working with vector spaces over non-commutative division rings is that the composition of linear operators and the multiplication of their matrix representations (relative to some fixed bases) of these operators does not correspond, in fact, is reversed. One way around this is to have linear operators and scalars operate on vectors on different sides. Another way is indicated by the following lemma.

Lemma 87.13. *Let M be an R-module, then the rings $\operatorname{End}_R(\coprod_{i=1}^n M)$ and $\mathbb{M}_n\big(\operatorname{End}_R(M)\big)$ are isomorphic.*

PROOF. Let $N = \coprod_{i=1}^n M_i$ with $M = M_i$ for all i. Then we have the usual maps

$$\iota_i : M_i \to N \text{ given by } m \mapsto (0, \ldots, \underbrace{m}_{i}, \ldots, 0)$$

and

$$\pi_j : N \to M_j \text{ given by } (m_1, \ldots, m_n) \mapsto m_j$$

are an R-monomorphism and R-epimorphism, respectively, satisfying

$$\pi_j \iota_i = \delta_{ij} 1_{M_i} \text{ and } \sum_{j=1}^n \iota_j \pi_j = 1_N.$$

Let $f \in \operatorname{End}_R(N)$ and $f_{ij} := \pi_i f \iota_j$ in $\operatorname{End}_R(M)$. Define

$$\varphi : \operatorname{End}_R(N) \to \mathbb{M}_n\big(\operatorname{End}_R(M)\big) \text{ by } f \mapsto (f_{ij}).$$

This map is clearly additive and it is a ring homomorphism, for if f, g lie in $\operatorname{End}_R(N)$, we have

$$(f_{ij})(g_{lk}) = (\pi_i f \iota_j)(\pi_l g \iota_k) = (\pi_i f g \iota_k) \text{ and } 1 \mapsto (\pi_l \iota_j) = I.$$

Now define

$$\psi : \mathbb{M}_n\big(\operatorname{End}_R(M)\big) \to \operatorname{End}_R(N) \text{ by } (f_{ij}) \mapsto \sum_{i=1}^n \sum_{j=1}^n \iota_i f_{ij} \pi_j,$$

an additive map that is a ring homomorphism, for if $(f_{ij}), (g_{ij})$ lie in $\mathbb{M}_n\big(\operatorname{End}_R(M)\big)$, we have $\psi\big(\operatorname{diag}(1_M, \dots 1_M)\big) = 1_N$ and

$$\psi\big((f_{il})(g_{kj})\big) = \psi\big((\sum_{l=1}^{n} f_{il}g_{lj})\big) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} \iota_i f_{il} g_{lj} \pi_j$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} \iota_i f_{il}\pi_l\iota_l g_{lj}\pi_j = \psi\big((f_{il})\big)\psi\big((g_{kj})\big).$$

Clearly, $\varphi \circ \psi = 1_{\mathbb{M}_n\left(\operatorname{End}_R(M)\right)}$ and $\psi \circ \varphi = 1_{\operatorname{End}_R(N)}$, so φ and ψ are inverse isomorphisms. □

For the next lemma, we do write endomorphisms and scalars on different sides.

Lemma 87.14. *Let e be an idempotent in the ring R and define*

$$\rho : eRe \to \operatorname{End}_R(Re) \text{ by } eae \mapsto \rho_{eae} : xe \mapsto xe \cdot eae.$$

View Re as a right $\big(\operatorname{End}_R(Re)\big)$-module. Then ρ is an isomorphism of rings.

PROOF. Check that ρ_{eae} is an R-homomorphism of Re when we write endomorphisms on the right, i.e., $(re)\rho_{eae} = r(e\rho_{eae})$. [Notice when we write it in this way it looks like the associative law.] It is also easily checked that ρ is a ring homomorphism.

ρ is a $1-1$: If $0 = re \cdot eae = reae$ for all r in R, then, setting $r = e$, shows that $eae = 0$.

ρ is onto: Let f lie in $\operatorname{End}_R(Re)$. Then there exists an element a in R satisfying $(e)f = ae$ in $Re \subset R$. Therefore, for all r in R, we have

$$(re)f = (re \cdot e)f = re\big((e)f\big) = reae = re \cdot eae = (re)\rho_{eae},$$

so $f = \rho_{eae}$. □

The following is essentially an immediate observation, but of great use.

Lemma 87.15. (Schur's Lemma) *Let M be an irreducible R-module. Then $\operatorname{End}_R(M)$ is a division ring.*

PROOF. Let f in $\operatorname{End}_R(M)$ be nonzero. Then $\ker f < M$ and $0 < \operatorname{im} f \subset M$. As M is irreducible, $\ker f = 0$ and $\operatorname{im} f = M$, hence f is an R-isomorphism. □

We can now classify nonzero simple, left Artinian rings.

Theorem 87.16. (Wedderburn's Theorem) *Let R be a nonzero ring. Then the following are equivalent:*

(1) R is simple and left Artinian.
(2) R is simple and semisimple.
(3) There exists a division ring D and a positive integer n such that $R \cong \mathbb{M}_n(D)$.

Suppose that R satisfies (3). Then in (3), D is unique up to a ring isomorphism and n is unique. More precisely, if R satisfies (2) and e is a nonzero idempotent in R, then the following are true:

(i) All minimal left ideals in R are isomorphic.
(ii) If $\mathfrak{A}$ is a minimal left ideal in R and $D = \mathrm{End}_R(\mathfrak{A})$, then $eRe \cong \mathbb{M}_m(D)$ for some positive integer m.
(iii) In (ii), the ring D is a division ring and the center $Z(eRe)$ of eRe is isomorphic to $Z(D)$, the center of D.
(iv) If D and E are division rings and $\mathbb{M}_m(D) \cong \mathbb{M}_n(E)$, with m, n positive integers, then $D \cong E$ and $m = n$.

PROOF. We have already shown that (3) $\Rightarrow$ (1) and (3) $\Rightarrow$ (2).

(1) $\Rightarrow$ (2): As R is left Artinian, there exists minimal left ideal $\mathfrak{A}$ of R by the Minimal Principle. Let

$$0 < \mathfrak{B} := \sum_{r \in R} \mathfrak{A}r \subset R,$$

a (2-sided) ideal. As R is simple, $\sum_{r\in R} \mathfrak{A}r = \mathfrak{B} = R$. Let $\rho_r : \mathfrak{A} \to \mathfrak{A}r$ be the R-epimorphism defined by $a \mapsto ar$. Since $\mathfrak{A}$ is irreducible, either $\rho_r = 0$ or ρ_r is an R-isomorphism, i.e., either $\mathfrak{A}r = 0$ or $\mathfrak{A}r \cong \mathfrak{A}$. It follows that R is completely reducible.

(2) $\Rightarrow$ (3): It suffices to show that (2) $\Rightarrow$ (i) — (iv), since 1 is an idempotent and $1R1 \cong R$. The argument to prove (1) $\Rightarrow$ (2) shows:

$R = \sum_{r\in R} \mathfrak{A}r$ with $\mathfrak{A}$ a minimal left ideal in R and
$\mathfrak{A} \cong \mathfrak{A}r$ for all r in R satisfying $\mathfrak{A}r \neq 0$,

as semisimple rings contain minimal left ideals. We now show conditions (i) — (iv) are satisfied.

(i): As argued before, we see that

There exists a minimal left ideal $\mathfrak{A}$ in R.
$R = \mathfrak{A}r_1 \oplus \cdots \oplus \mathfrak{A}r_n$ for some $r_1, \ldots, r_n$ in R.
Every minimal left ideal in R is $\mathfrak{A}r_i$ for some i, hence is isomorphic to $\mathfrak{A}$.

(ii). As R is completely reducible, there exists a nonzero idempotent e in R. As $0 < Re \subset R$, Re is also a completely reducible R-module.

By Claim 87.5, Re contains an irreducible submodule, hence a minimal left ideal of R. Using the notation in the proof of (i), we have

$$Re = \bigoplus_{j=1}^{m} (\mathfrak{A} r_{i_j} \cap Re) \cong \coprod_{j=1}^{m} \mathfrak{A},$$

where the i_j, $1 \leq j \leq m$, are those integers satisfying $\mathfrak{A} r_{i_j} \cap Re \neq 0$. By Lemmas 87.13 and 87.14 (with $e = 1$ as a special case), we have

$$eRe \cong \mathrm{End}_R(Re) \cong \mathrm{End}_R(\coprod_{j=1}^{m} \mathfrak{A}) \cong \mathbb{M}_m\big(\mathrm{End}_R(\mathfrak{A})\big).$$

This establishes (ii).

(iii) follows by Schur's Lemma and Exercise 87.20(4).

(iv): Let $A = \mathbb{M}_m(D)$, $B = \mathbb{M}_n(E)$, and $e_{11} = (\delta_{1i}\delta_{1j})$ in $\mathbb{M}_m(D)$. We have $Ae_{11} \subset \mathbb{M}_m(D)$ is a minimal left ideal. It is unique up to isomorphism by (i). Similarly, if $e'_{11} = (\delta_{1i}\delta_{1j})$ in $\mathbb{M}_n(E)$, then Be'_{11} is a minimal left ideal in B, unique up to isomorphism. By Lemma 87.14 and Example 87.11(5), we have

$$D \cong e_{11}Ae_{11} \cong \mathrm{End}_A(Ae_{11})$$
$$E \cong e'_{11}Be'_{11} \cong \mathrm{End}_B(Be'_{11}).$$

As $A \cong B$, their unique minimal left ideals (up to isomorphism) must correspond, so $D \cong E$ by the above. It follows that

$$m^2 = \dim_D A = \dim_E B = n^2,$$

so $m = n$ also. □

Corollary 87.17. *Let R be a simple ring. then R is left Artinian if and only if R is right Artinian if and only if R is left semisimple if and only if R is right semisimple.*

PROOF. This is true for $\mathbb{M}_n(D)$ with D a division ring. □

Corollary 87.18. *Let R be a simple, left Artinian ring. Then the following are equivalent:*

(1) R is a division ring.
(2) The only zero divisor of R is zero.
(3) The only idempotents of R are 0 and 1.
(4) The only nilpotent element in R is zero.

Let F be a field. Recall that a nonzero ring A is called an *F-algebra* if there exists a ring homomorphism $F \to Z(A)$ where $Z(A)$ is the center of A. This must be a monomorphism and we usually

identify F with $F1_A$. This ring homomorphism makes A into an F-module, i.e., a vector space over F. We say that the F-algebra A is a *finite dimensional F-algebra* if $\dim_F A$ is finite.

Corollary 87.19. *Let F be a field and A a finite dimensional F-algebra. Suppose that A is also a simple ring. Then $A \cong \mathbb{M}_n(D)$ for some division ring D containing F in its center. The division ring D is a finite dimensional F-algebra, unique up to isomorphism and n is unique.*

PROOF. The ring A is left Artinian since a finite dimensional vector space over F. □

Let F be a field and A a finite dimensional F-algebra. If A is also simple and *central*, i.e., $Z(A) = F$, then A is called a *central simple F-algebra*. By Wedderburn Theorem, $A \cong \mathbb{M}_m(D)$ for some division ring D. More over $Z(D) = F$ and D is a finite dimensional vector space over F. If $D = F$, we say that A is *split*. If A and B are central simple F-algebras, we say that A and B are *similar* if $A \cong \mathbb{M}_m(D)$ and $B \cong \mathbb{M}_n(E)$ with D and E isomorphic division rings. This is an equivalence relation by Wedderburn's Theorem. The classes of central simple F-algebras under this equivalence relation can be given an abelian group structure and form what is called the *Brauer group of F*, an important group in algebra and number theory. The Brauer group of $\mathbb{R}$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$ and the Brauer group of $\mathbb{C}$ is trivial, i.e., every central simple algebra over $\mathbb{C}$ splits. This follows from the results in Section §89. A theorem of Tsen shows that the Brauer group of a field of transcendence degree one over $\mathbb{C}$ is also trivial. This is harder to prove. The determination of the Brauer group of a number field is one of the crowning achievements of twentieth century number theory.

Exercises 87.20. 1. Let V be a nonzero finite dimensional vector space over a field (or division ring) F. Then V is is an $\mathrm{End}_F(V)$-module via evaluation, i.e., $fv := f(cv)$ for all f in $\mathrm{End}_F(V)$ and v in V. Show that V is $\mathrm{End}_F(V)$-irreducible.

2. Let R be a commutative ring with ideals $\mathfrak{A}_i$, $i = 1, 2$. Set
$$V_i := \{\mathfrak{p} \mid \mathfrak{p} \text{ a prime ideal with } \mathfrak{A}_i \subset \mathfrak{p}\}.$$
Suppose that $V_1 \cap V_2 = \emptyset$ and every prime ideal in R lies in V_1 or V_2, i.e., the set of prime ideals is the disjoint union of V_1 and V_2. Show that R contains a non-trivial idempotent.

3. Up to isomorphism, show that exists a unique irreducible R-module if R is simple and left Artinian.

4. Show if R is a ring, the map $R \to \mathbb{M}_n(R)$ given by $r \mapsto rI$ defines an isomorphism between the center $Z(R)$ of R and the center $Z\big(\mathbb{M}_n(R)\big)$ of $\mathbb{M}_n(R)$.

88. The Artin-Wedderburn Theorem

In the last section, we classified simple semisimple rings. In this section, we classify semisimple rings. The prototype for the simple case was a matrix ring over a division ring. The prototype for the general case is is a finite product of matrix rings over division rings. This generalization is quite useful, as it gives a foundation for that part of the Theory of Group Representations devoted to finite groups over fields of characteristic zero.

If we have a finitely generated completely reducible module over a ring R, then we would expect a nice decomposition for it. Indeed, if we look at the case of a finite dimensional vector space then it has a composition series. The analog gives us the following:

Lemma 88.1. *Let $M_1, \dots, M_r$ and $N_1, \dots, N_s$ be two finite collections of non-isomorphic irreducible R-modules. If*

$$M_1^{m_1} \coprod \cdots \coprod M_r^{m_r} \cong N_1^{n_1} \coprod \cdots \coprod N_s^{n_s}$$

for some positive integers $m_1, \dots, m_r$, $n_1, \dots, n_s$, then $r = s$ and there exists a permutation $\sigma \in S_r$ such that $M_i \cong N_{\sigma(i)}$ for all i. In particular, if $M = N$, then $M_i = N_{\sigma(i)}$ for all i.

Proof. Let $M = M_1^{m_1} \coprod \cdots \coprod M_r^{m_r}$, $N = N_1^{n_1} \coprod \cdots \coprod N_s^{n_s}$, and $\varphi : M \to N$ an R-isomorphism. Clearly, φ and φ^{-1} sets up a bijection

$$\{M_1, \dots, M_r\} \longleftrightarrow \{N_1, \dots, N_s\},$$

so $r = s$. Changing notation, we may assume that $M_i \cong N_i$ for all i and we are reduced to showing the following

Claim. If N is an irreducible R-module and $N^m \cong N^n$, then $m = n$:
Since $D = \operatorname{End}_R(N)$ is a division ring by Schur's Lemma, and

$$\mathbb{M}_m\big(\operatorname{End}_R(N)\big) \cong \operatorname{End}_R(N^m) \cong \operatorname{End}_R(N^n) \cong \mathbb{M}_n\big(\operatorname{End}_R(N)\big)$$

by Lemma 87.13, we have $m = n$ by Wedderburn's Theorem 87.16. The last statement is immediate using $\varphi = 1_M$ □

An important generalization of this lemma, the Krull-Schmidt Theorem, occurs when we replace the word irreducible by the word indecomposable, where an R-module is called *indecomposable* if it cannot be written as a non-trivial direct sum of two submodules. We shall not prove this generalization.

Lemma 88.2. *Let R be a semisimple ring. Then R is left Artinian. More precisely, if $\mathfrak{B}$ is a non-trivial left ideal of R, then there exist unique minimal left ideals $\mathfrak{A}_1, \ldots, \mathfrak{A}_m$ in R satisfying $\mathfrak{B} = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_m$.*

PROOF. We have shown that $R = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_n$ for minimal left ideals $\mathfrak{A}_i$ and if $\mathfrak{A}$ is a minimal left ideal, then $\mathfrak{A} = \mathfrak{A}_i$ for some i. It follows, as before, that

$$\mathfrak{B} = \bigoplus_{i=1}^{m} (\mathfrak{A}_{j_i} \cap \mathfrak{B}) \text{ with the } j_i \text{ satisfying } \mathfrak{A}_{j_i} \cap \mathfrak{B} \neq 0.$$

The last lemma says that this is unique. As any submodule of $\mathfrak{B}$ is a direct sum of some of the $\mathfrak{A}_{j_1}, \ldots, \mathfrak{A}_{j_m}$, the result follows. □

Our first formulation of the generalization of Wedderburn's Theorem is the next result. However, in the proof we shall show much more. Since notation will be used in the proof, we shall postpone this more precise statement until after the proof.

Theorem 88.3. (Artin-Wedderburn Theorem) *Let R be a semisimple ring. Then R is an (internal) direct sum of finitely many simple left Artinian rings, all unique (up to order). In particular, any semisimple right is both left and right Artinian (and left and right semisimple).*

PROOF. We prove this in a number of steps.

Step 1. Let $\mathfrak{A}$ be a minimal left ideal in R. Set

$$S_{\mathfrak{A}} := \{\mathfrak{B} \mid \mathfrak{B} \subset R \text{ a minimal left ideal with } \mathfrak{B} \cong \mathfrak{A}\}$$
$$B_{\mathfrak{A}} := \sum_{S_{\mathfrak{A}}} \mathfrak{B}.$$

Then the following are true:

(*i*) $B_{\mathfrak{A}} \subset R$ is a 2-sided ideal.

(*ii*) If $\mathfrak{A}'$ is a minimal left ideal in R, then

$$\mathfrak{A} \cong \mathfrak{A}' \text{ if and only if } B_{\mathfrak{A}} B_{\mathfrak{A}'} \text{ is not zero:}$$

(*i*): Let $\mathfrak{B} \in S_{\mathfrak{A}}$ and $x \in R$. As $\rho_x : \mathfrak{B} \to \mathfrak{B}x$ by $y \mapsto yx$ is an R-epimorphism, $\mathfrak{B}x = 0$ or $\mathfrak{B}x \cong \mathfrak{B} \cong \mathfrak{A}$. In either case, $\mathfrak{B}x \subset B_{\mathfrak{A}}$, so $B_{\mathfrak{A}}$ is a 2-sided ideal.

(*ii*): ($\Rightarrow$): We know that $\mathfrak{A} = Re > 0$ with e an idempotent. Then

$$0 \neq e^2 = e \cdot e \in B_{\mathfrak{A}} B_{\mathfrak{A}} = B_{\mathfrak{A}} B_{\mathfrak{A}'}.$$

($\Leftarrow$): Suppose that $B_{\mathfrak{A}} B_{\mathfrak{A}'}$ is not zero. Then there exist a $\mathfrak{B} \in S_{\mathfrak{A}}$ and a nonzero b' in $\mathfrak{B}' \in S_{\mathfrak{A}'}$ with $\mathfrak{B}b' \neq 0$. As $\mathfrak{B}$ is irreducible, $\mathfrak{B} \to \mathfrak{B}b'$

by $x \mapsto xb'$ is an R-isomorphism. As $0 < \mathfrak{B}b' \subset \mathfrak{B}'$ with $\mathfrak{B}'$ irreducible, we have

$$\mathfrak{A} \cong \mathfrak{B} \cong \mathfrak{B}b' \cong \mathfrak{B}' \cong \mathfrak{A}'.$$

Step 2. Let $R = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_q$ with $\mathfrak{A}_i$, $i = 1, \ldots, q$, the minimal left ideals in R, arranged such that $\mathfrak{A}_1, \ldots, \mathfrak{A}_m$, $m \leq q$, are all the mutually non-isomorphic ones. [So if $i > m$, there exists a $j \leq m$ with $\mathfrak{A}_i \cong \mathfrak{A}_j$ and every minimal left ideal in $\mathfrak{A}$ is isomorphic to some $\mathfrak{A}_i$ with $i \leq m$.] Set $B_i = B_{\mathfrak{A}_i}$ for $1 \leq i \leq m$. Then $R = \sum_{i=1}^m B_i$ and $B_i B_j = 0$ if $i \neq j$ (by Step 1). Moreover,

(*i*) $R = \bigoplus_{i=1}^m B_i$.

(*ii*) If $\mathfrak{B}$ is a nonzero 2-sided ideal in R, then $\mathfrak{B} = B_{i_1} \oplus \cdots \oplus B_{i_n}$, some i_j, $1 \leq i_j \leq m$:

(*i*): Let $C := B_i \cap \sum_{i \neq j} B_j$, then

$$B_i B_j = 0 \text{ so } CB_j = 0 = B_j C \text{ as } C \subset B_i,$$
$$B_j B_i = 0 \text{ so } CB_i = 0 = B_i C \text{ as } C \subset \sum_{j \neq i} B_j,$$

so $C = RC = 0$.

(*ii*): As R is completely reducible, there exists a minimal left ideal $\mathfrak{A}$ in R contained in $\mathfrak{B}$, and, in fact, $\mathfrak{B}$ is completely reducible so a sum of minimal left ideals of R. As $\mathfrak{A}_1, \ldots, \mathfrak{A}_m$ are all the minimal left ideals of R up to isomorphism, we may assume that $\mathfrak{A} \cong \mathfrak{A}_i$ lies in $\mathfrak{B}$ and need only show

Claim. $B_i \subset \mathfrak{B}$:

Write $\mathfrak{A} = Re$ with e a nonzero idempotent. We must show if $\mathfrak{A}' \in S_{\mathfrak{A}}$, then $\mathfrak{A}' \subset \mathfrak{B}$. Let $a \in \mathfrak{A} = Re = Re \cdot e = \mathfrak{A}e$. Therefore, there exists an element b in $\mathfrak{A}$ such that $a = be = be \cdot e = a \cdot e$, i.e., if $a \in \mathfrak{A}$ then $a = ae$. Let $\varphi : \mathfrak{A} \to \mathfrak{A}'$ be an R-isomorphism. Then for every a in $\mathfrak{A}$, we have $\varphi(a) = \varphi(ae) = a\varphi(e)$ in $\mathfrak{A}'$. Since $\mathfrak{B}$ is a 2-sided ideal, $\mathfrak{A}' = \mathfrak{A}\varphi(e) \subset \mathfrak{B}$, proving the claim.

Step 3. Let B_i be as in Step 2, then B_i is a simple, left Artinian ring:

We know $R = B_1 \oplus \cdots \oplus B_m$, so $1 = f_1 + \cdots + f_m$ for some $f_i \in B_i$.

Claim. The elements $f_1, \ldots, f_m$ are *central* orthogonal idempotents, i.e., in addition to the f_i being orthogonal idempotents, they lie in the center of R:

We know that $f_1, \ldots, f_m$ are orthogonal idempotents, so we need only show that they are central. Let x be an element of R. Then we have $x = \sum_i x_i$ for some $x_i \in B_i$ for each i. As $1_R x = x = x 1_R$, we must

have $f_i x_i = x_i = x_i f_i$ for all i and $f_j x_i = 0 = x_i f_j$ for all $j \neq i$. It follows that $f_i x = x f_i$ as needed.

It follows by the claim that B_i is a ring with $1_{B_i} = f_i$. Let $\mathfrak{B}$ be a 2-sided ideal in B_i. Since $B_i B_j = 0 = B_j B_i$ for all $j \neq i$, we have

$$\mathfrak{B} = B_i \mathfrak{B} = (\sum_j B_j) B_i \mathfrak{B} = \sum_j B_j \mathfrak{B} = R\mathfrak{B},$$

and similarly $\mathfrak{B} = \mathfrak{B}R$. Therefore, $\mathfrak{B}$ is a nonzero 2-sided ideal in R. By Step 2, $\mathfrak{B}$ is a sum of some of the B_j's, hence must be B_i, so B_i is simple. As R is completely reducible, there exists a minimal left ideal $\mathfrak{A}$ of R lying in B_i. If $\mathfrak{A}' < \mathfrak{A}$ is a left ideal in B_i, then $B_j \mathfrak{A}' \subset B_j B_i = 0$ for all $j \neq i$, so $\mathfrak{A}'$ is a left ideal in R, hence $\mathfrak{A}' = 0$. Therefore, $\mathfrak{A}$ is a minimal left ideal in B_i. It follows easily that B_i is semisimple, hence left Artinian.

Step 4. Suppose that $R = \bigoplus_{i=1}^m B_i = \bigoplus_{j=1}^n C_j$ with the B_i's as in Step 2 and each C_j a 2-sided ideal in R that is also a simple, left Artinian ring. Then $m = n$ and there exists a permutation $\sigma \in S_m$ such that $B_i = C_{\sigma(i)}$ for every i. The $B_1, \ldots, B_m$ are called the *simple components* of R and the decomposition $R = B_1 \oplus \cdots \oplus B_m$ is called a *Wedderburn decomposition* of R:

Each $B_i \cap C_j$ is a 2-sided ideal in both simple rings B_i and C_j. Therefore, for each i, there exists a k with $B_i \cap C_k$ nonzero, hence $B_i = C_k$. Similarly, for each j, there exists an l such that $B_l \cap C_j$ is nonzero, hence $B_l = C_j$. The result follows.

Step 5. Finish:

Each B_i is a simple, left Artinian ring, hence also right Artinian (and right semisimple). It follows easily that R is also right Artinian. □

The proof of the theorem also shows the following:

Theorem 88.4. (Artin-Wedderburn) *Let R be a nonzero ring. Then the following are equivalent:*

(1) R is semisimple.
(2) R is a finite direct product of simple, left Artinian rings.
(3) There exist (possibly isomorphic) division rings $D_1, \ldots, D_m$, some m, unique up to isomorphism (and order), and unique positive integers $n_1, \ldots, n_m$ and a ring isomorphism

$$R \cong \mathbb{M}_{n_1}(D_1) \times \cdots \times \mathbb{M}_{n_m}(D_m) \text{ up to order.}$$

If R is a semisimple ring and $\mathfrak{A}_1, \ldots, \mathfrak{A}_m$ are all the non-isomorphic minimal left ideals in R, then $\{\mathfrak{A}_1, \ldots, \mathfrak{A}_m\}$ is called a *basic set* for R.

We leave the proof of the following two corollaries as exercises:

Corollary 88.5. *Let R be a semisimple ring and $\{\mathfrak{A}_1, \ldots, \mathfrak{A}_m\}$ a basic set. If M is an irreducible R-module then $M \cong \mathfrak{A}_i$ for some i. In particular, if R is also simple, then, up to isomorphism, there exists a unique irreducible R-module.*

Corollary 88.6. *Let R be a semisimple ring and $\{\mathfrak{A}_1 \ldots, \mathfrak{A}_m\}$ a basic sets for R. Let $B_i = \sum_{S_{\mathfrak{A}_i}} \mathfrak{A}$ for $i = 1, \ldots, r$ and M a nonzero R-module. Then B_iM is a submodule of M and is a sum of irreducible submodules each isomorphic to $\mathfrak{A}_i$. Further, $M = \bigoplus_{i=1}^m B_iM$.*

Exercises 88.7. 1. If R is a semisimple ring and M a non-trivial R-module, then M is a direct sum of irreducible modules, unique up to isomorphism and order.

2. Prove Corollary 88.5.
3. Prove Corollary 88.6.
4. Show that a ring is semisimple if and only if it is left artinian and has no nonzero nilpotent left ideals.

89. Finite Dimensional Real Division Algebras

We have seen that the (Hamiltonian) quaternion algebra $\mathcal{H}$, the four dimensional real vector space with basis $\{1, i, j, k\}$, made into a ring by defining a multiplication on this basis by

$$i^2 = -1,\ j^2 = -1,\ ij = -ji = k$$

and extending this linearly to the whole space defines a division ring. The center $\{z \in \mathcal{H} \mid yz = zy \text{ for all } y \text{ in } \mathcal{H}\}$ of $\mathcal{H}$ is $\mathbb{R}$. Frobenius solved the problem of finding all division algebras D that are finite dimensional real vector spaces with $\mathbb{R}$ in its center, i.e.,

$$\mathbb{R} = \{x \mid xy = yx \text{ for all } y \text{ in } D\}.$$

We call such algebras *finite dimensional real division algebras.* If D is such a division algebra, then $\mathbb{R} \to D$ given by $r \mapsto r1_D$ is a ring monomorphism that we view as an inclusion. The result is the following:

Theorem 89.1. (Frobenius) *Let D be a finite dimensional real division algebra. Then D is isomorphic to $\mathbb{R}$, $\mathbb{C}$, or $\mathcal{H}$.*

Proof. We may assume that $\mathbb{R} < D$. As D is a finite dimensional real vector space, for any element x in D, we must have x is a root of a nonzero polynomial in $\mathbb{R}[t]$ and $\mathbb{R}[x]$ is a commutative ring. As $\mathbb{R}[x] \subset D$ with D a division ring, $\mathbb{R}[x]$ must be a domain. It follows, just as in the proof of Theorem 45.12, that $\mathbb{R}[x] = \mathbb{R}(x)$ is a field and a finite

extension of $\mathbb{R}$. (Cf. Exercise 45.24(1).) By the Fundamental Theorem of Algebra, we must have $\dim_{\mathbb{R}} \mathbb{R}(x) \leq 2$ and if x is not a real number, $\{1, x\}$ is a real basis for $\mathbb{R}(x)$ and $\mathbb{R}(x) \cong \mathbb{C}$. In particular, there exists an i in D such that $i^2 = -1$. Then $\mathbb{R}(i) \cong \mathbb{C}$, so we can identify $\mathbb{R}(i)$ and $\mathbb{C}$, which we do, i.e., we may view $\mathbb{R} \subset \mathbb{C} \subset D$. Therefore, D is a (left) finite dimensional complex vector space. Let $T : D \to D$ be the $\mathbb{C}$-linear transformation given by $x \mapsto xi$. Then T satisfies $T^2 = -1_D$, so the minimal polynomial q_T of T satisfies $q_T \mid t^2 + 1$ in $\mathbb{C}[t]$. It follows that the only possible eigenvalues of T are $\pm i$ and at least one of them most be an eigenvalue. Let $E_T(\alpha) = \{v \in D \mid Tv = \alpha v\}$ for α in $\mathbb{C}$, so either $E_T(i) \neq 0$ or $E_T(-i) \neq 0$. Since $E_T(i) \cap E_T(-i) = 0$, we have $E_T(i) \oplus E_T(-i) \subset D$. Let x be an element in D, then we have

$$T(x - ixi) = xi + ix = i(x - ixi), \text{ so } x - ixi \text{ lies in } E_T(i)$$
$$T(x + ixi) = xi - ix = -i(x - ixi), \text{ so } x + ixi \text{ lies in } E_T(-i).$$

As

$$x = \frac{1}{2}(x - ixi) + \frac{1}{2}(x + ixi) \text{ lies in } E_T(i) + E_T(-i),$$

we have $D = E_T(i) \oplus E_T(-i)$ as a complex vector space.

Check: $E_T(i) = \{x \in D \mid ix = xi\}$ is a commutative division ring, i.e., a field.

Therefore, by the Fundamental Theorem of Algebra, $E_T(i) = \mathbb{C}$, so $\dim_{\mathbb{C}} E_T(i) = 1$. Consequently, if $E_T(-i) = 0$, we must have $D = \mathbb{C}$. So we may assume that $E_T(-i)$ is nonzero. If x and y lie in $E_T(-i)$, then

$$(*) \qquad xyi = x(-i)y = ixy,$$

so xy lies in $E_T(-i)$. Let y be a nonzero element in $E_T(-i)$. Define a $\mathbb{C}$-linear transformation

$$\rho_y : D \to D \text{ by } x \mapsto xy.$$

As D is a division ring, ρ_y must be a monomorphism (why?). By (*), we have

$$\rho_y|_{E_T(-i)} : E_T(-i) \to E_T(i).$$

As the linear transformation $\rho_y|_{E_T(-i)}$ is injective and $\dim_{\mathbb{C}} E_T(i) = 1$, we must also have $\dim_{\mathbb{C}} E_T(-i) = 1$.

Claim: If x is a nonzero vector in $E_T(-i)$, then x^2 is real and $x^2 < 0$: We know that $x^2 \in \mathbb{R}(x) \cap E_T(i) = \mathbb{R}(x) \cap \mathbb{C}$ by (*). As $\mathbb{R}[x]$ is a two dimensional real vector space on basis $\{1, x\}$ with x not lying in $\mathbb{C}$, if $a + bx$, a, b real, lies in $\mathbb{C}$, we must have $b = 0$. It follows that x^2 is real. If $x^2 > 0$, then $x^2 = \theta^2$ for some real number θ. This means that $\pm\theta, x$

would be three distinct roots of $t^2 - x^2$ in $\mathbb{R}[t]$, which is impossible. Therefore, $x^2 < 0$ and the claim is established.

Now let $j = x/\sqrt{|x^2|}$, then $j^2 = -1$. Set $k = ij$, Then $\{j, k\}$ is a real basis for the vector space $E_T(-i)$ — check — and $\{1, i, j, k\}$ is a basis for the real vector space D satisfying $i^2 = -1 = -j^2$ and $k = ij = -ji$. If follows that $D \cong \mathcal{H}$. □

90. Addendum: Cyclic Algebras

Wedderburn's Theorem shows that any central simple algebra is a matrix ring over a division ring. However, it does not give any indication on how to find division rings. Wedderburn was also interested in discovering new division rings. The simplest construction of new central simple algebras became important in Number Theory. In this section, we introduce this construction. Behind its foundation is Hilbert's Theorem 90 (and its cohomological formulation that we shall not discuss).

We begin with the construction of new rings.

Definition 90.1. Let R be any nonzero ring and σ a ring automorphism of R. Define the *twisted polynomial ring* $R[t, \sigma]$ to be the ring with the usual addition of polynomials with multiplication given by

$$(\sum_i a_i t^i)(\sum_j b_j t^j) := \sum_{i,j} a_i \sigma^i(b_j) t^{i+j}$$

where the a_i, b_j lie in R. So, in general, t does not commute with elements in R, rather we have $tb = \sigma(b)t$ for all b in R.

Remarks 90.2. Let R be a nonzero ring and σ a ring automorphism of R.

1. If for any a and b in R, satisfying $ab = 0$, we have $a = 0$ or $b = 0$, i.e., R is a *non-commutative domain*, then so is the twisted polynomial ring $R[t, \sigma]$.
2. Let $R = K$ be a field. Then the left division algorithm holds in $K[t, \sigma]$. In particular, every left ideal is principal (i.e., $K[t, \sigma]$ is a *left PID*). If $F \subset K^{\langle\sigma\rangle}$ is a subfield and f is a polynomial in $F[t]$, then $K[t, \sigma]f$ is a two sided ideal, so $K[t, \sigma]/K[t, \sigma]f$ is a ring.

Definition 90.3. Let K/F be a (finite) cyclic field extension of degree n, i.e., Galois with cyclic Galois group, with $G(K/F) = \langle\sigma\rangle$ and $f = t^n - a$ a polynomial in $F[t]$ with a nonzero. Then $(K/F, \sigma, a) := K[t, \sigma]/K[t, \sigma]f$ is called a *cyclic algebra of degree n* over F. Equivalently, using the canonical epimorphism $^- : K[t, \sigma] \to K[t, \sigma]/K[t, \sigma]f$,

we have $(K/F, \sigma, a)$ is a vector space over K on basis $\{1, x, \ldots, x^{n-1}\}$ with $x = \bar{t}$ and satisfying

$$x^n = a \quad \text{and} \quad x\beta = \sigma(\beta)x \text{ for all } \beta \text{ in } K.$$

It is convenient to use both of these formulations, so we switch between them with this notation without further comment.

Note that a cyclic algebra $(K/F, \sigma, a)$ of degree n over F is a finite dimensional F-algebra of F-dimension n^2.

Examples 90.4. 1. The Hamiltonian quaternions is the cyclic algebra $(\mathbb{C}/\mathbb{R}, {}^{-}, -1)$ over $\mathbb{R}$ of degree 2 with ${}^{-}$ complex conjugation and $x = j$.

2. Let F be a field of characteristic zero or of prime degree not dividing n and containing a primitive nth root of unity. Suppose that K/F is a cyclic extension of degree n with $G(K/F) = \langle\sigma\rangle$. Then $K = F(y)$ with $m_F(y) = t^n - b$ irreducible and $\sigma y = \zeta y$ for some primitive nth root of unity ζ by Theorem 57.20 and its proof. So $xy = \zeta yx$. This cyclic algebra is usually written $\left(\frac{a, b}{F, \zeta}\right)$. For example, for the Hamiltonian quaternions, we have $y = i$ and $a = -1 = b$.

We shall also see that if F has a cyclic extension of degree n, then $\mathbb{M}_n(F)$ is also an example of a cyclic algebra. To do this, we show that cyclic algebras are central simple algebras.

Proposition 90.5. *Let $R = (K/F, \sigma, a)$ be a cyclic algebra over F of degree n. Then R is a central simple algebra over F. Moreover, the centralizer $Z_R(K) := \{x \in R \mid xb = bx \text{ for all } b \in K\}$ is K and the only field L satisfying $K \subset L \subset R$ is K.*

Proof. We first show that R is simple. Write

$$R = K1_R \oplus \cdots \oplus Kx^{n-1}$$
$$x^n = a \text{ with } a \in F \text{ and } x\alpha = \sigma(\alpha)x \text{ for all } \alpha \in K.$$

Let $0 < \mathfrak{A} \subset R$ be a 2-sided ideal. Choose a nonzero element $z \in \mathfrak{A}$ satisfying

$$z = \alpha_{i_1}x^{i_1} + \cdots + \alpha_{i_r}x^{i_r} \text{ with } \alpha_{i_1}, \ldots, \alpha_{i_r} \in K^\times,\ 0 \le i_1 < \cdots < i_r < n$$

and r minimal. If $r = 1$, then z is a unit in R as α_{i_1}, x^{i_1} are units, so we may assume that $r > 1$. As $\sigma^{i_1} \neq \sigma^{i_r}$, there exists and element β

in K satisfying $\sigma^{i_1}(\beta) \neq \sigma^{i_r}(\beta)$. In the ideal $\mathfrak{A}$, we have

$$\begin{aligned} (1) \qquad z\beta &= (\alpha_{i_1}x^{i_1} + \cdots + \alpha_{i_r}x^{i_r})\beta \\ &= \alpha_{i_1}\sigma^{i_1}(\beta)x^{i_1} + \cdots + \alpha_{i_r}\sigma^{i_r}(\beta)x^{i_r} \\ (2) \quad \sigma^{i_1}(\beta)z &= \alpha_{i_1}\sigma^{i_1}(\beta)x^{i_1} + \cdots + \alpha_{i_r}\sigma^{i_1}(\beta)x^{i_r}. \end{aligned}$$

Subtracting (2) from (1) yields the nonzero element

$$z\beta - \sigma(\beta)z = \alpha_{i_2}\big(\sigma^{i_2}(\beta) - \sigma^{i_1}\big(\beta\big)\big)x^{i_2} + \cdots + \alpha_{i_r}\big(\sigma^{i_r}(\beta) - \sigma^{i_1}(\beta)\big)x^{i_r},$$

contradicting the minimality of r. Therefore, R is simple.

Suppose that $z = \sum_{i=0}^{n-1} \alpha_i x^i$ is nonzero in R and commutes with all β in K, then $\alpha_i\sigma(\beta)x^i = \alpha_i\beta x^i$ for $i = 0, \ldots, n-1$. If β does not lie in F, then $\alpha_i = 0$ for $i = 1, \ldots, n-1$ and β must lie in K. If $K \subset L \subset R$ is a subfield, then $L \subset Z_R(K) = K$.

Finally, if $z \in Z(R)$, then $z \in Z_R(K) = K$. As $z\beta = \sigma(\beta)z$ for all $\beta \in K$, we have $z \in K^{\langle\sigma\rangle} = F$. □

A field L lying in the cyclic algebra $R = (K/F, \sigma, a)$ not properly contained in any larger in R is called *a maximal subfield* of R. So K is one such. There can be many non-isomorphic ones.

The proposition allows us to show:

Computation 90.6. Let K/F be a cyclic extension of fields of degree n with Galois group $\langle\sigma\rangle$ and $R = (K/F, \sigma, 1)$. We show that $R \cong \mathbb{M}_n(F)$:

Since $t^n - 1 = (t-1)(t^{n-1} + \cdots + t + 1)$ in $K[t, \sigma]$ and $R = K[t, \sigma]/(t^n - 1)$, we have $t^n - 1$ lies in the principal ideal $R(t-1)$, a maximal left ideal in R by Remark 90.2(2). It follows that $M := K[t, \sigma]/K[t, \sigma](t-1) \cong K$ is a simple R-module of F-dimension n. Since R is a simple ring, the F-algebra homomomorphism $\rho : R \to \mathrm{End}_F(M)$ defined by $r \mapsto \rho_r : m \mapsto mr$ must be a monomorphism. Since $\dim_F R = \dim_F \mathrm{End}_F(M)$ and $\mathbb{M}_n(F) \cong \mathrm{End}_F(M)$, the result is established.

Corollary 90.7. *Let $R = (K/F, \sigma, a)$ be a cyclic algebra over F of prime degree p. Then R is either a division ring or $R \cong \mathbb{M}_p(F)$.*

Recall that a central simple F algebra is called *split* if it is isomorphic to a matrix ring over F.

Proof. By Wedderburn's Theorem 87.16, $R \cong \mathbb{M}_r(D)$ for some division ring D, so $p^2 = r^2 \dim_F(D)$. As p is a prime number, $p = r$ and $\dim_F(D) = 1$ splits, (i.e., R splits) or R is a division algebra (i.e., $r = 1$ and $R = D$). □

Theorem 90.8. *Let $R = (K/F, \sigma, a)$ be a cyclic algebra of degree n over F. Then $R \cong \mathbb{M}_n(F)$ if and only if $a \in \mathrm{N}_{K/F}(K^\times)$.*

PROOF. ($\Leftarrow$): Suppose that $a \in \mathrm{N}_{K/F}(K^\times)$, then there exists an $\alpha \in K^\times$ such that $aN_{K/F}(\alpha) = 1$. Let $y = \alpha x$ where x is the image of t in $K[t, \sigma]$ under the canonical epimorphism. Set $y = \alpha x$ in R. As $x\alpha = \sigma(\alpha)x$ and $x^n = a$, we have

$$y^n = (\alpha x)^n = \sigma^{n-1}(\alpha) \cdots \sigma(\alpha)\alpha x^n = \mathrm{N}_{K/F}(\alpha)a = 1,$$

so, if $\beta \in K$,

$$y\beta = \alpha x\beta = \alpha\sigma(\beta)x = \sigma(\beta)\alpha x = \sigma(\beta)y.$$

The F-algebra map $R \to (K/F, \sigma, 1)$ sending x to y and fixing K is therefore an isomorphism. It follows by the computation that $R \cong \mathbb{M}_n(F)$.

($\Rightarrow$): Suppose that $R \cong \mathbb{M}_n(F)$. Then R has a simple left R-module M of dimension n over F. As $K[t, \sigma]$ is a left PID, we have $M \cong K[t, \sigma]/K[t, \sigma]f$ for some $f \in K[t, \sigma]$ with $K[t, \sigma]f \supset K[t, \sigma](t^n - a)$. Since

$$n = \dim_K M = (\dim_F K)(\deg f),$$

we must have $\deg f = 1$, i.e., $f = t - c$ for some c in K. As $K[t, \sigma]f \supset K[t, \sigma](t^n - a)$,

$$t^n - a = (b_{n-1}t^{n-1} + \cdots b_1 t + b_0)(t - c),$$

for some $b_0, \ldots, b_{n-1}$ in K. Multiplying out and comparing coefficients yields

$$\begin{aligned} b_{n-1} &= 1 \\ b_{n-2} &= \sigma^{n-1}(c) \\ b_{n-3} &= \sigma^{n-1}(c)\sigma^{n-2}(c) \\ &\vdots \\ b_0 &= \sigma^{n-1}(c) \cdots \sigma(c). \end{aligned}$$

So $a = b_0 c = \sigma^{n-1}(c) \cdots \sigma(c)c = \mathrm{N}_{K/F}(c)$ as needed. (Cf. this proof with that of Hilbert Theorem 90.) □

Let K/F be a cyclic Galois field extension of degree n with Galois group $\langle\sigma\rangle$ and $\bar{} : K^\times \to K^\times/\mathrm{N}_{K/F}(K^\times)$ be the canonical map. Wedderburn showed that $(K/F, \sigma, a)$ is a division ring if $\overline{a}$ has order n in $K^\times/\mathrm{N}_{K/F}(K^\times)$. The proof uses a generalization of the norm map $\mathrm{N}_{K/F}$ called the *reduced norm*. This generalizes Corollary 90.7. A deep theorem in Algebraic Number Theory shows that every division algebra over a number field is a cyclic algebra. This is not true in general.

There is a more general construction for an arbitrary finite Galois extension K/F called a *crossed product algebra* that also produces central simple algebras. It is constructed as follows. Let K/F be a finite Galois extension of fields. Let R be a vector space over K of dimension $[K : F]$ on basis $\mathcal{B} := \{u_\sigma \mid \sigma \in G(K/F)\}$. We define multiplication on R as follows: Let σ, τ, η be arbitrary elements in $G(K/F)$ and set

$$u_\sigma \beta = \sigma(\beta) u_\sigma \qquad \text{for all } \beta \in K \tag{90.9}$$

$$u_\sigma u_\tau = \alpha_{\sigma,\tau} u_{\sigma\tau} \qquad \text{for some } \alpha_{\sigma,\tau} \in K \tag{90.10}$$

$$\alpha_{\sigma,\tau}\alpha_{\sigma\tau,\eta} = \sigma(\alpha_{\tau,\eta})\alpha_{\sigma,\tau\eta} \qquad . \tag{90.11}$$

The last relation is to guarantee associativity. It also shows that $\alpha_{1.\sigma} = \alpha_{1,1}$ and $\alpha_{\sigma,1} = \sigma(\alpha_{1,1})$ so the one of R is $\alpha_{1,1}^{-1} u_1$. One also checks that R is an F-algebra, i.e., $a(\alpha\beta) = (a\alpha)\beta = \alpha(a\beta)$ for all $a \in F$ and $\alpha, \beta \in R$. It is a fact that, although every equivalence class of central simple algebras contains a crossed product algebra, not every central simple algebra is one. So the theory is still not finished.

Exercises 90.12. 1. Let R be a ring and σ a ring automorphism of R. Define the *twisted power series ring* $R[[t, \sigma]]$ over R, to be the ring with the usual addition of (formal) power series and multiplication induced by $tx = \sigma(x)t$ for all x in F. Show if R is a non-commutative domain, so is $R[[t, \sigma]]$.

2. Let R be a field and σ a ring automorphism of R. Define the *twisted Laurent series ring*

$$R((t, \sigma)) := \{\sum_{-\infty}^{\infty} a_i t^i \mid a_i \in R \text{ with } a_i = 0 \text{ for almost all negative } i\}$$

over R, to be the ring with the usual addition of (formal) Laurent series (i.e., $\sum_{-\infty}^{\infty} a_i t^i + \sum_{-\infty}^{\infty} b_i t^i = \sum_{-\infty}^{\infty} a_i b_i t^i$) and multiplication induced by $tx = \sigma(x)t$ for all x in R. If R is a division ring, show that $R((t, \sigma))$ is also a division ring and never a field if σ is not the identity.

3. Let R be a field and σ a ring automorphism of R of finite order n. Show that the center of $R((t, \sigma))$ is $R((t^n))[= R((t^n, 1_R))]$. In particular, $R((t, \sigma))$ is not finite dimensional over its center.

4. Let F be a field of characteristic zero of of positive degree not dividing n containing a primitive nth root of unity. Suppose that K/F is a cyclic extension of degree n and $t^n - a$ and $t^n - b$ are irreducible polynomials in $F[t]$. Show that $\left(\frac{b, a}{F, \zeta^{-1}}\right) \cong \left(\frac{a, b}{F, \zeta}\right)$

5. Show equations 90.9 define a ring.

6. Show the ring defined by equations 90.9 is a central simple F-algebra.

91. Addendum: Jacobson's Theorem

Let R be any ring. Then its *center*, $Z(R) := \{x \in R \mid xy = yx \text{ for all } y \text{ in } R\}$, is a commutative subring. If R is a division ring, then $Z(D)$ is a field. In addition, if $x \in D$ and F any subfield of $Z(D)$, then $F[x]$ is a commutative subring of D and its quotient field $F(x)$ is a field and also lies in D. Moreover, if x is algebraic over F, i.e., $F[x]$ is a finite F-vector space, then $F(x) = F[x]$ and $F(x)/F$ is finite. We have looked at two cases for D, the case that D is finite and the case that D is a finite dimensional $\mathbb{R}$-vector space. In the first case, D was commutative, but in the second case this might not be true. When D is a finite division ring, we have $x^{|D|} = 1$ for all x in D. Jacobson's Theorem says that a ring R that satisfies $x^n = x$ for each x in R for some integer $n = n(x) > 1$ is commutative. We prove this in this section. If D is a division ring, then two nonzero elements x and y in D commute if and only if $yxy^{-1} = x$. We want to exploit this in the special case of Jacobson's Theorem when R is a division ring (and reduce to the finite division ring case that we have answered).

Proposition 91.1. *Let D be a division ring that satisfies $x^n = x$ for each x in D and some integer $n > 1$ depending on x. Then D is a field.*

Proof. We know that the center of D contains a prime field, so $n1_D$ lies in the center for all integers n. Let x be a nonzero element of D. Then there exist positive integers n and n' such that $x^n = x$ and $(2x)^{n'} = 2x$. It follows with $N = (n-1)(n'-1)+1$ that $(2^N-2)x = 0$ in D, so $2^N - 2 = 0$. We conclude that the center of D has characteristic p for some prime p. Let F be the prime field in the center of D, so $F \cong \mathbb{Z}/p\mathbb{Z}$. Suppose that D is not commutative. Let x in D not lie in the center. As x is algebraic over F, since it is a root of $t^n - t$ in $F[t]$ for some positive integer n, the field extension $F(x)/F$ is finite by the discussion prior to the proposition. Therefore, $F(x)$ is a finite field. Indeed the same argument shows that $F(y)$ is a finite field for all y in D. Let $|F(x)| = p^m$. We also know that any element z in $F(x)$ satisfies $z^{p^m} = z$.

Claim: There exists an element y in D and an integer k such that $x^k \neq x$ and $yxy^{-1} = x^k$:

Let $T_x : D \to D$ be the F-linear transformation defined by $z \mapsto zx - xz$. As x is not in the center of D, this map is not the trivial map, i.e.,

$T_x(z) \neq 0$ for some z. Since

$$T_x^2(z) = T_x(zx - xz) = zx^2 - 2xzx + x^2z,$$

inductively, we see that

$$T_x^l(z) = \sum_{i=0}^{l} (-1)^i \binom{l}{i} x^i z x^{l-i} \text{ for all positive integers } l \text{ and all } z \text{ in } D.$$

Since $\operatorname{char} F = p$, the Children's Binomial Theorem implies that $T_x^p(z) = zx^p - x^p z$, hence

$$T_x^{p^m}(z) = zx^{p^m} - x^{p^m}x = zx - xz = T_x(z) \text{ for all } z \text{ in } D,$$

i.e., $T_x^{p^m} = T_x$. For each w in D, left multiplication by w, $\lambda_w : D \to D$, i.e., $z \mapsto wz$ is also an F-linear transformation. We have for each w in $F(x)$ (so $wx = xw$)

$$T_x\lambda_w(z) - T_x(wz) = wzx - x(wz) = w(zx - xz) = \lambda_w T_x(z),$$

i.e., $T_x\lambda_w = \lambda_w T_x$ for all w in $F(x)$ and z in D. As $\lambda_w\lambda_{w'} = \lambda_{w'}\lambda_w$ if w, w' lie in $F(x)$, we have

$$(*) \qquad 0 = T_x^{p^m} - T_x = \prod_{F(x)} (T_x - \lambda_w).$$

By assumption, there exists an element y in D satisfying $xy - yx \neq 0$. As D is a division ring, $D^\times$ satisfies the cancellation law and y is invertible in D. It follows that there exists a nonzero element w in $F(x)$ such that $0 = (T_x - \lambda_w)(y) = yx - xy - wy$, i.e., $yxy^{-1} = x + w$ in $F(x)$ by (*). Let $s = p^m - 1$, the order of $F(x)^\times = \langle x \rangle$. As $1, x, \ldots, x^{s-1}$ are all the roots of $t^s - 1$ in $F(x)[t]$, we must have $yxy^{-1} = x^k$ for some $k > 1$. This proves the claim.

Let y in D be as in the claim. We know by the argument above that $F(y)$ is also a finite field with say p^n elements. So we have $y^{p^n} = y$ and $xy \neq yx$. Set

$$W := \{z \in D \mid z = \sum_{i=0}^{p^m} \sum_{j=0}^{p^n} a_{ij} x^i y^j \text{ with } a_{ij} \text{ in } F\}.$$

We have W a finite set closed under addition. By the claim, W is closed under multiplication, so W is a finite subring of D. If z lies in W, $F(z) = F[z] \subset W$, so W is a finite division ring. Consequently, W is a field by Wedderburn's Theorem, contradicting x and y do not commute. $\square$

Let F be a field and D a division ring containing F. If $x \in D$, then $F(x)$ is a field in D. We say that D is *algebraic* over F if every element in D is algebraic over F.

Corollary 91.2. *Let D be a division algebra that is algebraic over a finite field. Then D is commutative. In particular, D/F is an algebraic extension of fields.*

PROOF. If F is the finite field in D and of characteristic p, then $F(x)$ is a finite field for each $x \in D$, hence $x^{p^n} = x$ for some n. It follows that D is commutative by the proposition. □

We next generalize the proposition.

Theorem 91.3. (Jacobson) *Let R be a rng (i.e., a ring possibly without a one). Suppose for each element x in R, there exists an integer $n > 1$ depending on x such that $x^n = x$. Then R is commutative.*

PROOF. We may assume that R is not the trivial rng. We want to reduce to the proposition. Let x be an element in R and $n > 1$ an integer such that $x^n = x$. Since for every element z in R, $z^m = z$ for some integer $m > 1$, R cannot have any nonzero nilpotent elements, i.e. we cannot have $z^m = 0$ unless $z = 0$ – why?

Step 1. The element x^{n-1} is a *central idempotent* in R, i.e., x^{n-1} lies in the center of R and $(x^{n-1})^2 = x^{n-1}$:

Set $e = x^{n-1}$. We have $e^2 = x^{2n-2} = x^n x^{n-2} = xx^{n-2} = x^{n-1} = e$, so e is an idempotent in R. If y lies in R, then $(ye - eye)^2 = 0 = (ey - eye)^2$. As R has no nonzero nilpotent elements, we have $ye = eye = ye$, so e is central.

Step 2. Every right ideal $\mathfrak{A}$ in R is a (two-sided) ideal:

Let r be an element in R and x an element in $\mathfrak{A}$. Then, as $\mathfrak{A}$ is a right ideal, with x^{n-1} central and lying in $\mathfrak{A}$, we have $rx = rx^{n-1}x = \big((x^{n-1})r\big)x$ lies in $\mathfrak{A}$.

Step 3. Suppose that R is a ring (i.e., has a one). Define the *right radical* $\mathfrak{R}$ of R to be the intersection of all maximal right ideals in R. [It can be shown that the *left radical* of R and the right radical of R are the same, but we do not need this.] Then for all y in R, the element $xy - yx$ lies in $\mathfrak{R}$:

Since 1 lies in R, the Zorn's Lemma argument showing that nonzero commutative rings have maximal ideals, shows that maximal right ideals exist in R. Let $\mathfrak{m}$ be any such maximal right ideal. By Step 2, $\mathfrak{m}$ is an ideal in R. By the Correspondence Principle, $R/\mathfrak{m}$ is a ring with no nontrivial right ideals. It follows that every nonzero element

in $R/\mathfrak{m}$ has a right inverse and therefore an inverse, i.e., $R/\mathfrak{m}$ is a division ring – why? If z is an element in R with $z^m = z$ and $m > 1$, we have $(z + \mathfrak{m})^m = z^m + \mathfrak{m} = z + \mathfrak{m}$, so $R/\mathfrak{m}$ is a division ring satisfying the hypothesis of the proposition, hence commutative. Therefore, $(z + \mathfrak{m})(y + \mathfrak{m}) = (y + \mathfrak{m})(z + \mathfrak{m})$ for all y, z in R, i.e., $zy - yz$ lies in $\mathfrak{m}$ for any right maximal ideal in R as needed.

Step 4. If R is a ring, then R is commutative:

It suffices to show the (right) radical $\mathfrak{R}$ in R is zero. Suppose not and x is a nonzero element in $\mathfrak{R}$. Let $x^n = x$ with $n > 1$ and $e = x^{n-1}$. By Step 1, e is a central idempotent and lies in $\mathfrak{R}$ as x does. Suppose that $1 - e$ is not a unit. Then $1 - e$ lies in a maximal right ideal $\mathfrak{m}$ in R. This implies that 1 lies in $\mathfrak{m}$ as $e \in \mathfrak{R} \subset \mathfrak{m}$, which is impossible. Therefore, the right ideal $(1 - e)R = R$, so $(1 - e)r = e$ for some r in R. We conclude that $0 = e(1 - e)r = e^2 = e = x^{n-1}$, contradicting x cannot be nilpotent.

Step 5. Finish:

By Step 4, we may assume that R does not have a one. Let e be a central idempotent in R as in Step 1 and set $T = eR = Re$. Check that T is a ring with $e = 1$ and for each y in T, there exists an integer $n > 1$ such that $y^n = y$. By Step 4, we know that T is a commutative ring. Let x and y be elements of R, then using the commutativity of T, we have $xye = (xe)(ye) = (ye)(xe) = yxe$. It follows that $(xy - yx)e = 0$. Since e was a central idempotent in R $(xy - yx)e = 0$ for every central idempotent e in R. In particular, if $m > 1$ an integer such that $(xy - yx)^m = xy - yx$ in R, then $(xy - yx)^{m-1}$ is a central idempotent by Step 1, so $xy - yx = (xy - yx)(xy - yx)^{m-1} = 0$. This proves that R is commutative. □

Exercises 91.4. 1. Show that if R is a rng in which every element x satisfies $x^n = x$ for some integer n (depending on x), that R has no nontrivial nilpotent elements.

2. If R is a nontrivial ring in which every nonzero element has an inverse, show that R is a division ring.

CHAPTER XVIII

Introduction to Representation Theory

In this chapter, we give applications of the Artin-Wedderburn Theorem to finite group theory. In particular, we study group homomorphisms $\varphi : G \to \mathrm{GL}_n(F)$ with G a finite group and F a field, especially the case that the characteristic of F is zero. We give two important applications, the first a famous theorem of Hurwitz on the products of sums of squares, and the second the theorem of Burnside showing that groups of order $p^a q^b$ are solvable where p and q are primes and a, b non-negative integers. To prove the second we introduce and study characters of representations, i.e., the trace of a representation $\varphi : G \to \mathrm{GL}_n(F)$.

Throughout this chapter F will denote a field and G a group.

92. Representations

Throughout this section R will denote a nonzero commutative ring.

Definition 92.1. Let R be a commutative ring and G a group (respectively monoid) (usually written multiplicatively) . The *group ring* (respectively, *monoid ring*) $R[G]$ of G over R is the free R-module on basis $G = \{g \mid g \in G\}$ made into a ring by the following multiplication:

$$(\sum_G a_g g)(\sum_G b_h h) = \sum_G c_k k \text{ with } a_h, b_h \in G \text{ almost all zero where}$$

$$c_k = \sum_{gh=k} a_g b_h.$$

We have $1_{R[G]} = 1_R e_G$ where e_G is the unity of G.

The map $\varphi_{R[G]} : R \to R[G]$ given by $r \mapsto r1_{R[G]}$ is a ring monomorphism, so we view $R \subset R[G]$. As $R \subset Z(R[G])$, the center of $R[G]$, the ring $R[G]$ is an R-algebra.

Examples 92.2. Let R be a commutative ring and G a group.

1. If $H = \{t^i \mid i \geq 0\}$, then $R[H] = R[t]$.
2. If $H = \mathbb{N}^n$ with $\mathbb{N} = \mathbb{Z}^+ \cup \{0\}$, (an additive monoid), then $R[H] \cong R[t_1, \ldots, t_n]$.

3. $R[\mathbb{Z}] \cong R[t, t^{-1}]$.
4. $R[\mathbb{Z}^n] \cong R[t_1, t_1^{-1}, \ldots, t_n, t_n^{-1}]$.
5. $R[G]$ is commutative if and only if G is abelian.
6. $G \subset R[G]^\times$.
7. If G is cyclic of order n, then $R[G] \cong R[t]/(t^n - 1)$.

Remarks 92.3. Let R be a commutative ring, G a (multiplicative) group, M an $R[G]$-module, and N an R-module.

1. The map $\lambda : G \to \mathrm{Aut}_R(M)$ given by $g \mapsto \lambda_g = \lambda(g) : m \mapsto gm$, is a group homomomorphism. (The inverse of λ_g is $\lambda_{g^{-1}}$.)
2. A group homomorphism $\varphi : G \to \mathrm{Aut}_R N$ is called a *representation* of G. Such a φ makes N into an $R[G]$-module via

$$g \cdot x := \varphi(g)(x) \text{ for all } g \in G,\ x \in N.$$

   Conclusion. A representation $\varphi : G \to \mathrm{Aut}_R N$ is equivalent to an $R[G]$-structure on N. We say that φ is *irreducible* if N is an irreducible $R[G]$-module (via φ).
3. If $\tau : G' \to G$ is a group homomorphism, then any representation of G induces a representation of G' by composition.
4. As G is an R-basis for the R-free module $R[G]$, any representation $\varphi : G \to \mathrm{Aut}_R N$ can be extended to an R-algebra homomorphism which we also write as $\varphi : R[G] \to \mathrm{End}_R N$ and also call a *representation*. Conversely, any R-algebra homomorphism $\varphi : R[G] \to \mathrm{End}_R N$ restricts to a representation $\varphi : G \to \mathrm{Aut}_R N$.

   Conclusion. A representation $\varphi : G \to \mathrm{Aut}_R N$ is equivalent to an R-algebra homomorphism $\varphi : R[G] \to \mathrm{End}_R N$.
5. Let $\lambda : G \to \sum(G)$ be the left regular representation, i.e., $x \mapsto \lambda_x : g \mapsto xg$. Then λ induces a representation $\lambda : R[G] \to \mathrm{End}_R(R[G])$ called the *(left) regular representation* of G (relative to R). It is *faithful* , i.e., $\ker \lambda = 0$.
6. Suppose that N is a finitely generated free R-module of rank n on (an ordered) basis $\mathcal{B}$. Then $N \cong R^n$ and we have a group isomorphism $\psi^{-1} : \mathrm{GL}_n(R) \to \mathrm{Aut}_R(N)$ where $\psi(T) = [T]_{\mathcal{B}}$, the matrix representation of T relative to the basis $\mathcal{B}$. If $\varphi : G \to \mathrm{Aut}_R N$ is a representation, we get a group representation $\widetilde{\varphi} = \psi\varphi : G \to \mathrm{GL}_n(R)$ called an *R-representation* of *degree n*. We say that $\widetilde{\varphi}$ *affords* N or φ *affords* N if $\mathcal{B}$ is clear. [It is also common to write $\mathrm{GL}(N)$ for $\mathrm{Aut}_R(N)$.] Hence fixing a basis $\mathcal{B}$ for a finitely generated free R-module N gives rise to a representation $\widetilde{\varphi} : G \to \mathrm{GL}_n(R)$ where n

is the rank of N, hence an $R[G]$-module structure to R^n. One often writes $\widetilde{\varphi}(n)$, $n \in N$, for $\varphi(g)(n)$ if we know N.

7. Suppose that N and N' are two free R-modules of the same rank n with $\varphi : G \to \mathrm{Aut}_R(N)$ and $\psi : R \to \mathrm{Aut}_R(N')$ affording N and N', respectively. We say φ and ψ are *equivalent* and write $\varphi \sim \psi$, if there exists an R-isomorphism $T : N \to N'$ satisfying

$$\begin{array}{ccc} N & \xrightarrow{T} & N' \\ {\scriptstyle \varphi(g)}\downarrow & & \downarrow{\scriptstyle \psi(g)} \\ N & \xrightarrow[T]{} & N'. \end{array}$$

commutes for all g in G, i.e., if N is an $R[G]$-module via φ and N' is an $R[G]$-module via ψ in the above, then $T : N \to N'$ is an $R[G]$-isomorphism. If $N = R^n = N'$, of course, we write $GL_n(R)$ for $\mathrm{Aut}_R\, N$.

Conclusion. We have $\varphi \sim \psi$ in the above if and only if $N \cong N'$ as $R[G]$-modules.

8. We say the $R[G]$-module M is *G-trivial* if $gm = m$ for all m in M and all g in G.

9. Define

$$M^G := \{m \in M \mid gm = m \text{ for all } g \in G\},$$

a submodule of M called the set of *G-fixed points* of M.

10. Let $\varphi : G \to \mathrm{GL}_n(R)$ afford the finitely generated free R-module N. Then N is G-trivial via φ if and only if φ is the trivial map, i.e., $g \mapsto I$ for all g in G. We call this the *trivial representation* of G. If $n = 1$, it is irreducible.

11. Suppose both M, N, and P are $R[G]$-modules. Then $\mathrm{Hom}_R(M, N)$ is an $R[G]$-module defined by

$$(\sigma f)(m) := \sigma\Big(f\big(\sigma^{-1}(m)\big)\Big)$$

for all $\sigma \in G$, $m \in M$, and $f \in \mathrm{Hom}_R(M, N)$.

Check. If $f \in \mathrm{Hom}_R(M, N)$, $g \in \mathrm{Hom}_R(N, P)$, then $\sigma(g \circ f) = (\sigma g) \circ (\sigma f)$ and

$$\mathrm{Hom}_{R[G]}(M, N) = \big(\mathrm{Hom}_R(M, N)\big)^G.$$

12. We view R as a trivial $R[G]$-module. If M is an $R[G]$-mmodule, then the isomorphism of $R[G]$-modules $\mathrm{Hom}_R(R, M) \to M$ given by $f \mapsto f(1)$ induces an isomorphism $\mathrm{Hom}_{R[G]}(R, M) \cong M^G$.

13. Suppose that G is a finite group. Define the *norm* of G by
$$N_G := \sum_G g \text{ in } R[G].$$
As $\sigma N_G = N_G = N_G \sigma$ for all σ in G, we have N_G lies in $R[G]^G$ and $N_G x$ lies in M^G for all x in M, i.e., $N_G : M \to M^G$ given by $x \mapsto N_G\, x$ is an $R[G]$-homomorphism, so $N_G M \subset M^G$. The quotient $M^G/N_G M$ is an object of study in algebraic number theory. Of course, this also says that N_G lies in the center of $R[G]$. More generally, if H is a normal subgroup of G, then N_H lies in the center of $R[G]$.
14. Suppose that G is a finite group and both M and N are $R[G]$-modules. Let $f : M \to N$ be an R-homomorphism. Define the *trace* of f to be the $R[G]$-homomorphism
$$\operatorname{Tr}_G f := N_G f = \sum_G \sigma f : M \to M.$$
In particular, if f is an $R[G]$-homomorphism, then $\operatorname{Tr}_G f = |G| f$.

Lemma 92.4. *Let R be a commutative ring and G a finite group. Suppose that M, M', N, N' are $R[G]$-modules and $f : M \to N$ an R-homomorphism. If $\varphi : M' \to M$ and $\psi : N \to N'$ are $R[G]$-homomorphisms, then $\operatorname{Tr}_G(\psi f \varphi) = \psi \circ \operatorname{Tr}_G f \circ \varphi$.*

PROOF. By definition, $(\sigma f)(m) = \sigma\big(f(\sigma^{-1} m)\big)$, so
$$\begin{aligned}\operatorname{Tr}_G(\psi f \varphi) &= \sum_G \sigma \circ (\psi f \varphi) = \sum_G (\sigma \circ \psi) \circ (\sigma \circ f) \circ (\sigma \circ \varphi) \\ &= \psi \circ \Big(\sum_G \sigma f\Big) \circ \varphi = \psi \circ \operatorname{Tr}_G f \circ \varphi. \qquad \square\end{aligned}$$

Theorem 92.5. (Maschke's Theorem) *Let F be a field and G a finite group. Suppose that either* $\operatorname{char} F = 0$ *or* $\operatorname{char} F \not| \, |G|$. *Then $F[G]$ is semisimple.*

PROOF. Let V be an $F[G]$-module and W an $F[G]$-submodule. We must show that W is a direct summand of V as an $F[G]$-module. Let $i_W : W \to V$ denote the inclusion map. We must show that this is a split $F[G]$-monomorphism by Exercise 35.18(15). Write $V = W \oplus W'$ as F-vector spaces. (Of course, W' need not be an $F[G]$-module.) Let $\pi_W : V \to W$ be the vector space projection of V onto W. By assumption, $1/|G| \in F^\times$, so we can define the 'average'
$$\varphi := \frac{1}{|G|} \operatorname{Tr}_G \pi_W : V \to W.$$

The map φ is an $F[G]$-homomorphism and, by the lemma, we have

$$\operatorname{Tr}_G(i_W \pi_W i_W) = i_W \circ \operatorname{Tr}_G \pi_W \circ i_W = |G|\, i_W \circ \varphi \circ i_W.$$

Since

$$\operatorname{Tr}_G(i_W \pi_W i_W) = \operatorname{Tr}_G \big(i_W(\pi_W i_W)\big) = \operatorname{Tr}_G(i_W 1_W) = \operatorname{Tr}_G i_W = |G|\, i_W,$$

we conclude that $i_W \circ (\varphi \circ i_W) = i_W = i_W 1_W$ as $|G|$ is a unit in $F[G]$. As i_W is a monomorphism, we have $\varphi \circ i_W = 1_W$, i.e., i_W is a split $F[G]$-monomorphism as desired. □

Maschke's Theorem allows us to apply the Artin-Wedderburn Theorem to $F[G]$ when G is a finite group and $\operatorname{char} F = 0$ or $\operatorname{char} F$ does not divide $|G|$. We also get one further piece of useful information, as we can compute the dimension of the center of $F[G]$, which we now do.

Definition 92.6. Let R be a commutative ring and G a group. If g is an element of G and the conjugacy class $\mathcal{C}(g)$ of g in G finite, we let $C_g := \sum_{\mathcal{C}(g)} h$ called the *class sum* of g in $R[G]$.

Lemma 92.7. *Let R be a commutative ring and G a finite group. If $C_{g_1}, \ldots, C_{g_r}$ are the distinct class sums of G in $R[G]$, then the center $Z(R[G])$ is a free R-module on basis $\mathcal{B} := \{C_{g_1}, \ldots, C_{g_r}\}$. In particular, the rank of $Z(R[G])$ is equal to the number of conjugacy classes of G.*

Proof. Let σ be an element of G. Then for all g in G, we have $\sigma C_g \sigma^{-1} = C_g$, so C_g lies in $Z(R[G])$. Since $G = \bigvee \mathcal{C}(g_i)$ is an R-basis for $R[G]$, the set $\mathcal{B}$ is R-linearly independent. Thus we need only show that $\mathcal{B}$ spans $Z(R[G])$. Let $y = \sum_G a_g g$ lie in $Z(R[G])$. Then for each σ in G, we have $y = \sigma y \sigma^{-1} = \sum_G a_g \sigma g \sigma^{-1}$. As G is an R-basis for $R[G]$, we have $a_g = a_{\sigma g \sigma^{-1}}$ for all σ in G, and the result follows. □

We now apply the Artin-Wedderburn Theorem to the case that F is an algebraically closed field and G is a finite group with $\operatorname{char} F = 0$ or $\operatorname{char} F \nmid |G|$, i.e., Maschke's Theorem holds, e.g., $F = \mathbb{C}$, to conclude the following:

Theorem 92.8. *Let G be a finite group and F an algebraically closed field of characteristic zero or positive characteristic not dividing the order of G. Let $\mathfrak{A}_1, \ldots, \mathfrak{A}_r$ be a basic set for $F[G]$ and $n_i = \dim_F \mathfrak{A}_i$, $i = 1, \ldots, r$. Then*

1. *$F[G] \cong \coprod_{i=1}^r \mathfrak{A}_i^{n_i}$ as $F[G]$ modules.*
2. *$F[G] \cong \bigtimes_{i=1}^r \mathbb{M}_{n_i}(F)$ as rings.*
3. *r is the number of conjugacy classes of G.*
4. *$|G| = \sum_{i=1}^r n_i^2$ and at least one of the n_i is one.*

PROOF. There are no finite dimensional F-division rings over F except for F, as any such would contain an element x not in F, but then $F(x)$ would be a commutative division ring distinct from F. It also follows that if B_i is the Wedderburn component corresponding to $\mathbb{M}_{n_i}(F)$, we have $\dim_F B_i = n_i$ is the dimension of a column space of $\mathbb{M}_{n_i}(F)$. Therefore, we have (1), (2), and (3). Finally, $Z\big(M_{n_i}(F)\big) = F$, so

$$G = \dim_{\mathbb{C}} \mathbb{C}[G] = \dim_{\mathbb{C}} \left(\bigtimes_{i=1}^{r} \mathbb{M}_{n_i}(\mathbb{C})\right) = \sum_{i=1}^{r} n_i^2,$$

and the trivial representation of G into $\mathrm{GL}_1(F) = F^\times$ is irreducible of degree one giving (4). □

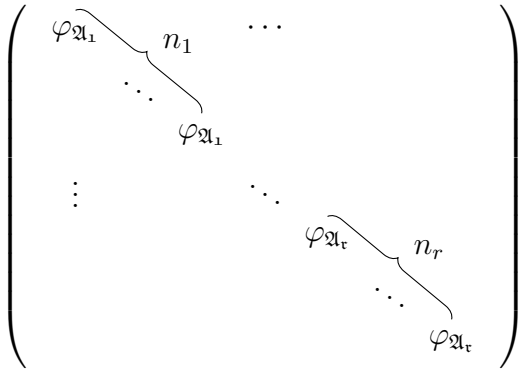
We can weaken the hypothesis of the theorem that F be algebraically close to having all the Wedderburn components matrix rings over F. We shall investigate this in the next section.

Remarks 92.9. Let G be a finite group and F a field of characteristic zero or of positive characteristic not dividing $|G|$.

1. A representation $\varphi : G \to \mathrm{GL}_1(F)$, i.e., of degree one is called a *linear representation*). By Maschke's Theorem, it must be irreducible.
2. As a special case of (1), suppose that G is an *elementary 2-group*, i.e., $G \cong (\mathbb{Z}/2\mathbb{Z})^n$ for some n and F is a field of characteristic different from two. Then Maschke's Theorem holds. There exist at least 2^n linear inequivalent representations of G given by the trivial representation and the $2^n - 1$ representations each defined by taking one nonzero element of G to -1 and all other elements if G to 1. Since $\dim_F F[G] = 2^n$ and $F[G]$ is semisimple, these must be all the irreducible representations of G.
3. Let $G^{ab} = G/[G, G]$. Then any representation of G^{ab} gives rise to a representation of G of the same degree via composition with the canonical epimorphism $^- : G \to G^{ab}$.
4. If F is algebraically closed, then any irreducible representation $\varphi : G^{ab} \to \mathbb{M}_n(F)$ must be of degree one.

Example 92.10. Let G be a finite group of order n. We give some examples of complex representations, i.e., representations $\varphi : G \to \mathrm{GL}_n(\mathbb{C})$.

1. We interpret Theorem 92.8 in the case of complex representations. Let $\mathfrak{A}_1, \ldots, \mathfrak{A}_r$ be a basic set for $\mathbb{C}[G]$ and $n_i = \dim_{\mathbb{C}} \mathfrak{A}_i$, $i = 1, \ldots, r$. Let φ_i be the irreducible representation afforded by $\mathfrak{A}_i$. Then the regular representation $\lambda : G \to \mathrm{GL}_n(\mathbb{C})$ is similar to



2. Let $G = \langle x \rangle$ be a cyclic group of order n and ω a primitive root of unity in $\mathbb{C}$. Then $\rho_i : G \to \mathbb{C}^\times = \mathrm{GL}_1(\mathbb{C})$ determined by $x \to \omega^i$ for $i = 1, \dots n$ are all the inequivalent irreducible representations of G.
3. Let G be the dihedral group D_3 with $G = \{a, b \mid a^3 = 1 = b^2,\ bab^{-1} = a^{-1}\}$. We know the trivial representation is of degree one, and there are three conjugacy classes in D_3. Therefore, there is another a linear character and one of degree 2. The non-trivial linear character $G^{ab} \cong \mathbb{Z}/2\mathbb{Z}$ induces the irreducible character $G \to \mathbb{C}^\times$ determined by $a \mapsto 1$ and $b \mapsto -1$. Let ω be a primitive cube root of unity. The representation $\varphi : G \to \mathrm{GL}_n(\mathbb{C})$ determined by
$$a \mapsto \begin{pmatrix} \omega & 0 \\ 0 & \omega^{-1} \end{pmatrix} \text{ and } b \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
is the third irreducible representation.

Exercise 92.11. 1. Let F be an algebraically closed field of characteristic zero, G be a finite group and $\varphi : G \to \mathrm{GL}_n(F)$ a representation. Show that φ is irreducible if and only if for each x in G, there exists an element λ in F such that $\lambda(x) = \lambda I$.
2. Let F be an algebraically closed field of characteristic zero, G be a finite group, and z an element of the center of $F[G]$. Show if V is an irreducible $F[G]$-module, then there exists an element λ in F satisfying $zv = \lambda v$ for all $v \in V$.
3. Let F be an algebraically closed field of characteristic zero and G be a finite group. Show if there exists a faithful irreducible $F[G]$-module M (i.e., irreducible and $gm = m$ for all $m \in M$, then $g = e_G$), then the center of G is cyclic.
4. Determine all complex representations of a finite abelian group.
5. Let $G = \langle x \rangle$ be a cyclic group of order n and ω a primitive root of unity in $\mathbb{C}$. In the notation of Example 92.10(2) determine the $\mathbb{C}[G]$-submodule of $\mathbb{C}[G]$ determined by ρ_1, by ρ_i.

6. Let G be the dihedral group D_3 and ω a primitive root of unity in $\mathbb{C}$. In the notation of Example 92.10(2) determine the irreducible $\mathbb{C}[G]$-submodules of $\mathbb{C}[G]$.
7. Find all complex irreducible representations of the dihedral group D_4.
8. Find all complex irreducible representations of the quaternion group Q_3.
9. Find all real irreducible representations of the quaternion group Q_3.

93. Split Group Rings

The purpose of this section is to extend theorems about of group rings over fields with the group finite and Maschke's Theorem holds when the underlying field is algebraically closed to the case that the underlying division rings in the Wedderburn decomposition of a group ring are always the base field. This is useful in the general theory.

Definition 93.1. Let D be a division ring and M a (left) D-vector space. If A is a subring of $\mathrm{End}_D(M)$, we say that A *act densely* on M if given any D-linear independent vectors $v_1, \ldots, v_m$ in M, $m > 0$, and vectors $v_1', \ldots, v_n'$ in M, then there exists a $T \in A$ satisfying $Tv_i = v_i'$ for $i = 1, \ldots, n$.

Theorem 93.2. (Jacobson Density Theorem) *Let R be a nonzero commutative ring, M an irreducible R-module, and $D = \mathrm{End}_R(M)$ (a division ring by Schur's Lemma). Then*

$$\lambda : R \to \mathrm{End}_D(M) \ \text{ by } \ a \mapsto \lambda_a : m \mapsto am$$

is a ring homomorphism. Set

$$L_R(M) := \mathrm{im}\,\lambda \subset \mathrm{End}_D(M).$$

Then $L_R(M)$ acts densely on M as a D-vector space. In particular, if $\dim_D M$ is finite, then λ is a ring epimorphism.

Proof. Let a be an element on R and f an element of $\mathrm{End}_D(M)$. Then for all m in M, we have

$$\lambda_a f(m) = af(m) = f(am) = f\lambda_a(m),$$

so λ_a lies in $\mathrm{End}_D(M)$ for all a in R, and clearly, λ is a ring homomorphism. Let $v_1, \ldots, v_m$ in M be D-linearly independent with $m \geq 1$ and $v_1', \ldots, v_n'$ in M. As M is R-irreducible, $M = Rv_i$ for every $i = 1, \ldots, n$. Thus the result is trivial if $m = 1$. So assume that $m > 1$. We are finished if we can establish the following:

Claim: It suffices to produce $s_m \in R$ satisfying

$$s_m v_i = \begin{cases} 0, & \text{if } i < m. \\ \neq 0, & \text{if } i = m. \end{cases}$$

Indeed, suppose that such an s_m exists. As $s_m v_m \neq 0$, we have $M = Rs_m v_m$ by irreducibility. Write $v'_m = rs_m v_m$ with $r \in R$ and let $b_m = rs_m$. In an analogous way, there exist $b_j \in R$ satisfying $b_j v_j = \delta_{ij} v'_j$, with δ_{ij} the Kronecker delta. Then $b = b_1 + \cdots + b_m$ works. This establishes the claim.

Now assume that the result is false. This means, we have

(*) If $a \in R$ satisfies $av_i = 0$, $1 \leq i \leq m-1$, then $av_m = 0$.

By induction, $M^{m-1} = \{(av_1, \ldots, av_{m-1}) \mid a \in R\}$, so (*) implies if $a, b \in R$, then

$$(av_1, \ldots, av_{m-1}) = (bv_1, \ldots, bv_{m-1}) \Rightarrow av_m = bv_m.$$

This means that

$$\mu : M^{m-1} \to M \text{ given by } (av_1, \ldots, av_{m-1}) \mapsto av_m$$

is a well-defined R-homomorphism. Let $\iota_j : M \to M^{m-1}$ be the R-homomorphism given by $m \mapsto (0, \ldots, \underbrace{m}_{j}, 0, \ldots, 0)$ as usual, and set $\varphi_j = \mu\, \iota_j$ in $D = \mathrm{End}_R(M)$. Then

$$v_m = \mu(av_1, \ldots, av_{m-1}) = \sum_{j=1}^{m-1} \mu(0, \ldots, \underbrace{m}_{j}, 0, \ldots, 0) = \sum_{j=1}^{m-1} \varphi_j(v_j),$$

contradicting $v_1, \ldots, v_m$ are D-linearly independent. □

Remark 93.3. Let F be a field and A a finite dimensional F-algebra (i.e., $\dim_F A < \infty$). Let M be a finitely generated A-module, hence a finite dimensional F-vector space. Then $\mathrm{End}_A(M) \subset \mathrm{End}_F(M)$ is a subring. By definition $F \subset Z(A)$, the center of A, so

$$\lambda : A \to \mathrm{End}_F(M) \text{ by } a \mapsto \lambda_a : m \mapsto am$$

is a ring homomorphism. Set

$$L_A(M) := \mathrm{im}\, \lambda \cong A/\ker \lambda = A/\mathrm{ann}_R\, M.$$

We view

$$F \subset \mathrm{End}_A(M) \subset \mathrm{End}_F(M)$$

as subrings via $a \mapsto a1_M$.

Theorem 93.4. (Burnside) *Let F be a field, A an F-algebra (not necessarily finitely generated), and M an irreducible A-module that is also a finite-dimensional vector space over F (so cyclic as finitely generated).*

(1) If F is algebraically closed, then $F = \mathrm{End}_A(M)$.
(2) If $F = \mathrm{End}_A(M)$, then $L_A(M) = \mathrm{End}_F(M)$.

PROOF. $\mathrm{End}_A(M)$ is a finite dimensional vector space over F as M is. By Schur's Lemma 87.15, $D = \mathrm{End}_A(M)$ is a division ring.

(1): Let $x \in D$. Then $F \subset Z(D)$, so $F(x) \subset D$ is a commutative division ring, i.e., $F(x)$ is a field. As $[F(x) : F] \leq \dim_F D < \infty$, the field extension $F(x)/F$ is algebraic, hence $F(x) = F$, as F is algebraically closed.

(2): As $\dim_R(M)$ is finite, this follows from the Jacobson Density Theorem 93.2. □

Definition 93.5. Let A be a semi-simple ring and a finite dimensional F-algebra with F a field. Suppose that $A = B_1 \oplus \cdots \oplus B_m$ is a Wedderburn decomposition. We say that A is *F-split* if B_i is a matrix ring over F for every $i = 1, \ldots, m$.

We leave the following as an exercise:

Proposition 93.6. *Let F be a field, A a semi-simple finite dimensional F-algebra. Then A is F-split if and only if $F \cong \mathrm{End}_A(M)$ for every irreducible A-module M. In particularly, if F is algebraically closed, then A is F-split.*

Remark 93.7. Suppose that F is a field, G a finite group not divisible by the characteristic of F. Then $F[G]$ is semi-simple. Suppose, in addition, that $F[G]$ is F-split and $M_1, \ldots, M_r$ a complete set of representatives for the isomorphism classes of irreducible $F[G]$-modules. Then $Z(F[G]) = \prod_{i=1}^{r} F$ and r is the number of conjugacy classes of G by Lemma 92.7.

Using the Artin-Wedderburn Theorem, Maschke's Theorem, Burnside's Lemma, and the above, we now have the following generalization of Theorem 92.8:

Summary 93.8. Let F be a field, G a finite group such that $\mathrm{char}\, F \nmid |G|$ (so $F[G]$ is semi-simple by Maschke's Theorem). Suppose that $F[G]$ is F-split, e.g., if F is algebraically closed. Let

$$M_1, \ldots, M_r$$

be a complete set of representatives for the isomorphism classes of irreducible $F[G]$-modules and $n_i = \dim_F M_i$ for $1 \leq i \leq r$. Then

1. $F[G] \cong \prod_{1=1}^{r} M_i^{n_i}$ as $F[G]$-modules.
2. $F[G] \cong \times_{i=1}^{r} \mathbb{M}_{n_i}(F)$ as rings.
3. r is the number of conjugacy classes of G.
4. $|G| = \sum_{i=1}^{r} n_i^2$.
5. $L_{F[G]}(M_i) = \mathrm{End}_F(M_i)$ for $i = 1, \ldots, r$.

Exercises 93.9. 1. Let F be a field, A an F-algebra (not necessarily finitely generated), and M a completely reducible A-module. Show if $\mathrm{End}_A(M) \cong F$, then M is an irreducible A-module.

2. Prove Proposition 93.6.

94. Addendum: Hurwitz's Theorem

We shall give a nice application of the theory developed so far. If $n \geq 3$, the *general quaternion group* Q_n is the group on n generators $a_1, \ldots, a_{n-1}, \varepsilon$ subject to the relations

$$\begin{aligned} \varepsilon^2 &= 1 \\ a_i^2 &= \varepsilon \qquad \text{for each } i = 1, \ldots, n-1 \\ a_i a_j &= \varepsilon a_j a_i \qquad \text{for all } i, j = 1, \ldots, n-1 \text{ satisfying } i \neq j. \end{aligned}$$

For example, Q_3 is the usual quaternion group. We shall see below that this group has order 2^n.

Properties 94.1. Let Q_n, $n \geq 3$, be the general quaternion group.

1. The element ε lies in the center of Q_n.
2. The element ε satisfies $\varepsilon = a_i a_j a_i^{-1} a_j^{-1}$ for all $i \neq j$. In particular, ε lies in $[Q_n, Q_n]$.
3. $[Q_n, Q_n] = \langle \varepsilon \rangle = \{1, \varepsilon\}$ (as $Q_n/\langle \varepsilon \rangle$ is abelian and $\varepsilon \in [Q_n, Q_n]$). In particular, $Q_n/\langle \varepsilon \rangle$ is an elementary 2-group on the image of the generators $a_1, \ldots, a_{n-1}$. Hence $|Q_n/\langle \varepsilon \rangle| = 2^{n-1}$ and $|Q_n| = 2^n$.
4. There exist (at least) 2^{n-1}-linear representations of Q_n over any field of characteristic different from two by Remark 92.9(2).
5. If n is odd, then the center $Z(Q_n)$ of Q_n, is $\langle \varepsilon \rangle$:

   It suffices to show that $z = a_1 \cdots a_l$ for $1 \leq l \leq n$ is not central. But if such a z lies in $Z(Q_n)$, then

   $$a_l a_1 \cdots a_l = a_1 \cdots a_l a_l = \varepsilon^{l-1} a_l a_1 \cdots a_l,$$

   so $l - 1$ is even, hence l is odd and $l < n - 1$. But

   $$a_{n-1} a_1 \cdots a_l = a_1 \cdots a_l a_{n-1} = \varepsilon^l a_{n-1} a_1 \cdots a_l,$$

   so l is even, a contradiction.

6. If n is even, then $Z(Q_n) = \langle \varepsilon, a_1 \cdots a_{n-1} \rangle$:
   This is similar to the previous argument, using $a_1 \cdots a_n$ is central as $n-1$ is odd, so
   $$a_1 \cdots a_{n-1} a_j = \varepsilon^{n-2} a_j a_1 \cdot a_{n-1} = a_j a_1 \cdots a_{n-1},$$
   since a_j occurs once in $a_1 \cdots a_{n-1}$.
7. If g in Q_n is not central, then its conjugacy class $\mathcal{C}(g) = \{g, \varepsilon g\}$:
   As $g \notin Z(Q_n)$, we have $|\mathcal{C}(g)| > 1$. Let $^- : Q_n \to Q_n/\langle \varepsilon \rangle$ be the canonical surjection. As $\overline{Q_n}$ is abelian, $g_o \in \{g, \varepsilon g\}$ if $g_o = xgx^{-1}$ with $x \in Q_n$.

We next look at the degrees of the irreducible representations of the general quaternion group over an algebraically closed field of characteristic different from two.

Calculation 94.2. Let Q_n, $n \geq 3$, be the general quaternion group and r the number of conjugacy classes of Q_n. Let F be a algebraically closed field of characteristic different from two. As Q_n is a 2-group, we know that r is the number of irreducible F-representations of Q_n and there exist 2^{n-1} linear F-representations by the properties of established above. Therefore, using the properties above, we have

$$2^n = |G| = \sum_{i=1}^{r} d_i^2 \text{ with } d_i = 1 \text{ for } 1 \leq i \leq 2^{n-1}.$$

Case 1. n is odd:
We have

$$r = |Z(Q_n)| + \frac{|Q_n| - |Z(Q_n)|}{2} = 2 + \frac{2^n - 2}{2} = 2^{n-1} + 1.$$

So there exists one further irreducible F-representation and it is of degree d_r. Since $2^n = 1 \cdot 2^{n-1} + d_r^2$, we have $d_r = 2^{\frac{n-1}{2}}$.

Case 2. n is even:
We have

$$r = |Z(Q_n)| + \frac{|Q_n| - |Z(Q_n)|}{2} = 4 + \frac{2^n - 4}{2} = 2^{n-1} + 2,$$

so there exist two further irreducible F-representations of degrees d_{r-1} and d_r and these satisfy $2^n = 1 \cdot 2^{n-1} + d_{r-1}^2 + d_r^2$, hence $2^{n-1} = d_{r-1}^2 + d_r^2$. Write $d_{r-1} = 2^k f_1$ and $d_r = 2^l f_2$ with f_1, f_2 odd. It follows that $k = l$, hence

$$2^{n-2k-1} = f_1^2 + f_2^2 \equiv 2 \mod 4.$$

If either f_1 or f_2 is greater than one, we would have $4 \mid f_1^2 + f_2^2$, which is impossible. Consequently, $d_{r-1} = d_r = 2^{\frac{n-2}{2}}$.

We shall use the calculation above to solve a very nice problem in field theory. Let $x_1, \dots, x_n$ and $y_1, \dots, y_n$ be variables over a field F. We wish to know when there exists a formula

$$(*) \qquad z_1^2 + \cdots + z_n^2 = (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2)$$

over $F[x_1, \dots, x_n, y_1, \dots y_n]$ with the z_i bilinear in the x_i's and y_j's, i.e., bilinear when plugging in any values in F for the variables. We know that there are such formulae in certain cases, e.g.,

If $n = 1$. (This is trivial.)
If $n = 2$. (Cf. with the norm from $\mathbb{C}$ to $\mathbb{R}$.)
If $n = 4$. (Cf. with the norm form of a quaternion algebra.)
If $n = 8$. (Cf. with the norm form Cayley's octonion algebra — a non-associate algebra, i.e., satisfies all properties of an algebra except associativity, and in addition all nonzero elements have inverses.)

Theorem 94.3. (Hurwitz) *Let F be a field of characteristic not two. Then (*) exists if and only if $n = 1, 2, 4,$ or 8.*

Proof. (Eckmann) If $n = 1, 2, 4, 8$, such a formula exists. We have written it down except for the case of $n = 8$, which we leave to the reader to look up. Let $X = (x_1, \dots, x_n)$ and suppose that (*) holds. Let $z_i = \sum_{j=1}^n a_{ij}(X) y_j$ for $i = 1, \dots, n$. Each of the $a_{ij}(X)$ must be linear in the x_i's. We have

$$\begin{aligned} \sum_{i=1}^n z_i^2 &= \sum_{i=1}^n \big(\sum_{j=1}^n a_{ij}(X) y_j \big)^2 \\ &= \sum_{i,j=1}^n a_{ij}(X)^2 y_j^2 + 2 \sum_{i=1}^n \sum_{1 \le j < k \le n} a_{ij}(X) a_{ik}(X) y_j y_k. \end{aligned}$$

Using (*) and comparing coefficients, we see that

$$\sum_{i=1}^n x_i^2 = \sum_{i=1}^n a_{ij}(X)^2 \qquad \text{for } j = 1, \dots, n.$$

$$\sum_{i=1}^n a_{ij}(X) a_{ik}(X) = 0 \qquad \text{for all } j \neq k.$$

Let $A = \big(a_{ij}(X)\big)$ in $\mathbb{M}_n\big(F[x_1, \dots, x_n, y_1, \dots, y_n]\big)$. We have

$$(\dagger) \qquad A^t A = \big(\sum_{i=1}^n x_i^2 \big) I.$$

Write

$$A = A_1x_1 + \cdots + A_nx_n \text{ with } A_i \in \mathbb{M}_n(F).$$

Substitute this into (†) and multiply out. Comparing coefficients yields equations:

$$\begin{aligned} &(1) \quad A_i^tA_j + A_j^tA_i = 0 \qquad \text{for all } i,j = 1,\dots,n \text{ and } i \neq j.\\ &(2) \quad A_i^tA_i = I \qquad\qquad\quad \text{for all } i = 1,\dots,n. \end{aligned}$$

In particular, the A_i are orthogonal matrices. We next normalize these equations by setting

$$B_i = A_iA_n^t \text{ for } 1 \leq i \leq n, \text{ hence } B_n = A_nA_n^t = I.$$

The B_i's then satisfy:

$$\begin{aligned} &(1') \quad B_i^tB_j + B_j^tB_i = 0 \qquad \text{for all } i,j = 1,\dots,n-1 \text{ and } i \neq j.\\ &(2') \quad B_i^tB_i = I \qquad\qquad\quad \text{for all } i = 1,\dots,n. \end{aligned}$$

As $B_n = I$, we have $B_i^tB_n + B_n^tB_i = 0$ for $1 \leq i < n$, hence

$$B_i^t = -B_i \text{ for } i = 1,\dots,n-1,$$

so $(1')$ and $(2')$ become

$$\begin{aligned} &(1'') \quad B_iB_j + B_jB_i = 0 \qquad \text{for all } i,j = 1,\dots,n-1 \text{ and } i \neq j.\\ &(2'') \quad B_i^2 = -I \qquad\qquad\quad \text{for all } i = 1,\dots,n. \end{aligned}$$

Let $\widetilde{F}$ be an algebraic closure of F and $\rho : Q_n \to \mathrm{GL}_n(\widetilde{F})$ be the representation induced by $a_i \mapsto B_i$, for $i = 1,\dots,n-1$ and $\varepsilon \mapsto -I$. Each linear representation of Q_n must take $\varepsilon \mapsto 1$ as $a_ia_j = \varepsilon a_ja_i$ for $i \neq j$ and F is commutative. Since $\operatorname{char} F \neq 2$, ρ is a direct sum of irreducible representations of degree greater than one using Exercise 88.7(1). In particular,

$$\begin{aligned} &\text{If } n \text{ is odd, then } 2^{\frac{n-1}{2}} \mid n, \text{ hence } n = 1.\\ &\text{If } n \text{ is even, then } 2^{\frac{n-2}{2}} \mid n. \end{aligned}$$

Since $2^{(n-2)/2} > n$ for $n \geq 10$, we have $n \leq 8$ and we check:

$$\begin{aligned} &\text{If } n = 8, \text{ then } 2^{\frac{8-2}{2}} \mid 8.\\ &\text{If } n = 6, \text{ then } 2^{\frac{6-2}{2}} \not\mid 6.\\ &\text{If } n = 2,4, \text{ then } 2^{\frac{n-2}{2}} \mid n. \end{aligned}$$

□

A more general theorem (that we do not prove) is

Theorem 94.4. (Hurwitz-Radon Theorem) *Let $x_1, \ldots, x_l$ and $y_1, \ldots, y_n$ be variables over a field F of characteristic not two. Then there exists a formula*

$$z_1^2 + \cdots + z_n^2 = (x_1^2 + \cdots + x_l^2)(y_1^2 + \cdots + y_n^2)$$

over $F[x_1, \ldots, x_l, y_1, \ldots y_n]$ with the z_i bilinear in the x_i's and y_j's if and only if $n = 2^{4\alpha+\beta}$ with $l \leq 8\alpha + 2^\beta$.

If one does not insist on the z_i being bilinear in the x_i's and y_i's and works in $F(x_1, \ldots, x_n, y_1, \ldots y_n)$, then Pfister showed that the product of two sums of 2^n squares is a sum of 2^n squares.

Exercise 94.5. Find all complex irreducible representations of the general quaternion group Q_n, $n \geq 3$.

95. Characters

Throughout this section F will be a field and G a group.

Definition 95.1. Let $\varphi : G \to \mathrm{GL}_n(F)$ be a representation. The map $\chi_\varphi : G \to F = \mathbb{M}_1(F)$ defined by

$$\chi_\varphi(x) := \operatorname{trace} \varphi(x) \text{ for all } x \in G$$

is called the *character* of the representation φ. The degree of the representation φ is also called the *degree* of χ_φ , so it is the integer $\chi_\varphi(e_G)$ which we shall denoted by $\chi_\varphi(1)$, i.e., write e_G as 1. In particular, χ_φ is not a group homomorphism in general. We say that χ_φ is an *irreducible character* if φ is an irreducible representation. For example, the character associated to the trivial representation called the *trivial character* is an irreducible character of degree one.

Remarks 95.2. Let V be an $F[G]$-module of dimension n as a vector space over F and $\varphi = \varphi_V : G \to \mathrm{GL}_n(F)$ a representation affording $V \cong F^n$ relative to some fixed basis. If $\psi : G \to \mathrm{GL}_n(F)$ affords V relative to another basis, then $\varphi(x)$ and $\psi(x)$ are conjugate for all x in V, hence $\chi_\varphi(x) = \chi_\psi(x)$ for all v in V. Therefore, χ_φ depends only on the equivalence class of φ. We often write χ_V for χ_φ. As $F[G]$ is F-free on basis G, the character χ_φ induces an F-linear functional $F[G] \to F$ that we also denote by χ_φ.

Proposition 95.3. *Let V, V', and V'' be $F[G]$-modules, finite dimensional over F. Then*

(1) $\chi_V(\sigma x \sigma^{-1}) = \chi_V(x)$ for all σ in G.

(2) If

$$0 \to V' \to V \to V'' \to 0$$

is an exact sequence of finite dimensional $F[G]$-modules, then

$$\chi_V = \chi_{V'} + \chi_{V''}.$$

PROOF. We already showed (1). As for (2), we may assume that $V' \subset V$ with $\mathcal{B}'$ a basis for V' extended to a basis to a basis $\mathcal{B}$ for V. If $\varphi_V : G \to \mathrm{GL}_n(F)$, then with the obvious notation,

$$\varphi_V(x) = \begin{pmatrix} \varphi_{V'}(x) & * \\ 0 & \varphi_{V''}(x) \end{pmatrix},$$

and the result follows. □

Example 95.4. Let G be a finite group with $\operatorname{char} F$ not dividing $|G|$, so $F[G]$ is semi-simple by Maschke's Theorem 92.5. Let

$$\lambda : G \to \mathrm{Aut}_F\big(F[G]\big) \text{ be given by } x \mapsto \lambda_x : y \mapsto xy.$$

Extend this representation linearly to an F-algebra homomorphism $\lambda : F[G] \to \mathrm{End}_F\big(F[G]\big)$. This is just the *(left) regular representation.* Let $\{\mathfrak{A}_1, \ldots, \mathfrak{A}_r\}$ be the basic set for $F[G]$.

Suppose that $F[G]$ is F-split, e.g., if F is algebraically closed, and $\varphi_{\mathfrak{A}_1}, \ldots, \varphi_{\mathfrak{A}_r}$ are all the irreducible representations of G afforded by the $\mathfrak{A}_i$ with the $\chi_{\mathfrak{A}_i}$ the corresponding characters. Set $n_i = \deg \chi_{\mathfrak{A}_i} = \chi_{\mathfrak{A}_i}(1)$. Then we have

$$\chi_{F[G]} = \chi_\lambda = \sum_{i=1}^{r} n_i \chi_{\mathfrak{A}_i} = \sum_{i=1}^{r} \chi_{\mathfrak{A}_i}(1)\chi_{\mathfrak{A}_i}.$$

Note: λ_x with x not e_G in G permutes G without fixed points and G is a basis for $F[G]$, so

$$\operatorname{trace} \varphi_\lambda(x) = \chi_\lambda(x) = 0 \text{ for all } x \neq e_G.$$

More generally, if $\varphi : G \to \mathrm{GL}_n(F)$ is a representation affording a finite dimensional F-vector space V, then there exist unique integers $m_i \geq 0$ satisfying

1. $V \cong \prod \mathfrak{A}_i^{m_i}$ as $F[G]$-modules.
2. $\chi_V = \sum m_i \chi_{\mathfrak{A}_i}$.
3. $\chi_V(1) = \dim_F V$.

We now show that if F is a field of characteristic zero, then the equivalence class of a representation of a finite group $G \to \mathrm{GL}_n(F)$ is completely determined by its character. This is false if F has positive characteristic dividing the order of G.

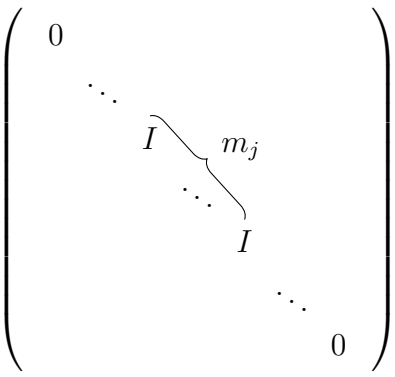
Theorem 95.5. *Let F be a field of characteristic zero, G a finite group, and V, V' two finitely generated $F[G]$-modules. Then*

$$V \cong V' \text{ if and only if } \chi_V = \chi_{V'}.$$

PROOF. ($\Rightarrow$) is trivial.

($\Leftarrow$): We know that the F-algebra homomorphism $\lambda : F[G] \to \mathrm{End}_F(V)$ induced by $x \mapsto \lambda_x : v \to xv$ affords V and that $F[G]$ is semi-simple by Maschke's Theorem. Let $\{\mathfrak{A}_1, \ldots, \mathfrak{A}_r\}$ be a basic set and $B_i = B_{\mathfrak{A}_i}$, $1 \le i \le r$ the simple components, so $F[G] = B_1 \oplus \cdots \oplus B_r$ is the Wedderburn decomposition. Let $e_i = 1_{B_i}$, $1 \le i \le r$, so $e_1, \ldots, e_r$ are central orthogonal idempotents. V is a finitely generated $F[G]$-module, so there exist unique $m_i \ge 0$ satisfying $V \cong \coprod \mathfrak{A}_i^{m_i}$ as $F[G]$-modules, which we view as an equality.

To prove the theorem, we need to show that the F-linear functional $\chi_V : F[G] \to F$ determines all the m_i. Fix a j, $1 \le j \le r$, and let $d_j = \dim_F \mathfrak{A}_j \ne 0$. [Note: We are not assuming $F[G]$ is F-split, so the value d_j is not so clear.] As e_j is central, for each i, the map $\lambda_{e_j} : \mathfrak{A}_i \to \mathfrak{A}_j$ given by $a \mapsto e_j a$ is an $F[G]$-homomorphism. Viewing $\mathfrak{A}_i$ as a B_i-module, we have $\lambda_{e_j} = \delta_{ij} 1_{B_j}$. Thus we can view $\lambda_{e_j} : V \to V$ by $v \mapsto e_j v$ as an $F[G]$-homomorphism and conclude that $\operatorname{im} \lambda_{e_j} = \lambda_{e_j}(V) = \mathfrak{A}_j^{m_j}$. Let $\mathcal{B}_{kj}$ be an (ordered) F-basis for the kth component of $\mathfrak{A}_j^{m_j}$. Then $\mathcal{B}_j = \mathcal{B}_{1j} \cup \cdots \cup \mathcal{B}_{m_j j}$ is an F-basis for $\mathfrak{A}_j^{m_j}$ and $\mathcal{B} = \bigcup \mathcal{B}_j$ is an F-basis for V. We then have $[\lambda_{e_j}]_{\mathfrak{B}}$ is the matrix



with I the $d_j \times d_j$ identity matrix. Consequently,

$$\chi_V(e_j) = \mathrm{trace}[\lambda_{e_j}]_{\mathfrak{B}} = d_j m_j,$$

so

$$m_j = \frac{\chi_V(e_j)}{d_j} = \frac{\chi_V(e_j)}{\dim_F \mathfrak{A}_j} \text{ in } F$$

is determined by χ_V. □

Exercise 95.6. Let G be a finite group and F be a field of positive characteristic p dividing $|G|$. Suppose that $\varphi : G \to \mathrm{GL}_n(F)$ is an irreducible representation, show that φ is the trivial representation, hence $\chi_\varphi = n1_F$. In particular, if V and V' are finitely generated $F[G]$-modules (hence finite dimensional vector spaces over F), then $\chi_V = \chi_{V'}$ if $\dim V \equiv \dim V' \mod p$.

96. Orthogonality Relations

Throughout this section, we shall let $\widetilde{F}$ denote an algebraic closure of the field F.

Definition 96.1. A group G is called a *torsion group* if every element of G has finite order. (G may be infinite.)

Proposition 96.2. *Let G be a torsion group, V an n-dimension vector space over F, V an $F[G]$-module. Let x be an element of G and N the order of the cyclic subgroup $\langle x\rangle$ in G. Then $\chi_V(x)$ is a sum of n terms in which each term is an Nth root of unity in an algebraic closure $\widetilde{F}$ of F.*

Proof. Let $N = \dim_F V$ and $\varphi : G \to \mathrm{GL}_N(F)$ afford V. Set $\alpha = \varphi(x)$ and $n = |\langle x\rangle|$. Taking a Jordan canonical form of α in $\mathrm{GL}_N(\widetilde{F})$ we see that $\operatorname{trace}\alpha$ is a sum of N elements each of which is an eigenvalue of α (counted with multiplicity). So it suffices to show that every eigenvalue of α is an nth root of unity in $\widetilde{F}$. If $v \in V$ is a nonzero eigenvector of α with eigenvalue ε, then $x^n = e_G$ implies that $\alpha^n = 1$, so $v = \alpha^n v = \varepsilon^n v$. Consequently, $1 = \varepsilon^n$ as needed. □

Recall that the integral closure of $\mathbb{Z}$ in K, $K/\mathbb{Q}$ a field extension, is denoted by $\mathbb{Z}_K$. Then the proposition implies:

Corollary 96.3. *Let G be a torsion group and n a positive integer satisfying $x^n = e$ for every element x in G. Suppose that F is an algebraically closed field of characteristic zero and ω a primitive nth root of unity in F. If $\varphi : G \to \mathrm{GL}_N(F)$ is a representation, then $\chi_\varphi : G \to \mathbb{Z}_{\mathbb{Q}(\omega)}$.*

Definition 96.4. Let $\chi : G \to F$ be a character. The *kernel* of χ is defined to be the set

$$\ker\chi := \chi^{-1}\big(\chi(1)\big) = \{x \in G \mid \chi(x) = \chi(1)\}.$$

Proposition 96.5. *Let G be a finite group, F a field of characteristic zero, and $\varphi : G \to \mathrm{GL}_n(F)$ a representation. Then $\ker\chi_\varphi = \ker\varphi$. In particular, $\ker\chi_\varphi$ is a normal subgroup of G.*

PROOF. Certainly, $\ker\varphi \subset \ker\chi_\varphi$, so we need only show the reverse inclusion. So suppose that $\chi_\varphi(x) = \chi_\varphi(1)$, then by the proposition, we have

$$n = \chi_\varphi(1) = \varepsilon_1 + \cdots + \varepsilon_n$$

with each ε an $|G|$th root of unity over F hence over $\mathbb{Q}$ (and viewed in $\mathbb{C}$). So $n \le |\varepsilon_1| + \cdots + |\varepsilon_n| \le n$. Check if $|\varepsilon_1 + \cdots + \varepsilon_n| = |\varepsilon_1| + \cdots + |\varepsilon_n|$, then $\varepsilon_i = \varepsilon_j$ for all i, j. It follows that $n = \varepsilon_1 + \cdots + \varepsilon_n = n\varepsilon_1$, hence $\varepsilon_i = \varepsilon_1$ for all i. Since the ε_i are the eigenvalues of $\varphi(x)$ in an algebraic closure of F by the previous proof, we must have the characteristic polynomial of $\varphi(x)$ is $f_{\varphi(x)} = (t-1)^n$. Since $\varphi(y)$ is a root of $t^{|G|} - 1$ for all $y \in G$, i.e, $\varphi(y)^{|G|} - 1 = 0$, the matrix $\varphi(x)$ is a root of the gcd of $(t-1)^n$ and of $t^{|G|} - 1$. The latter polynomial cannot have multiple roots, as we are working in characteristic zero, so this gcd is $t - 1$. It follows that $\varphi(x) = I$ and x lies in $\ker\varphi$. □

Theorem 96.6. *Let G be a finite group and $\varphi : G \to \mathrm{GL}_n(F)$ and $\varphi' : G \to \mathrm{GL}_m(F)$ two irreducible representations. Let*

$$a_{ij}, a'_{kl} : G \to F \ \text{ satisfy } \ \varphi(x) = \big(a_{ij}(x)\big),\ \varphi'(x) = \big(a'_{kl}(x)\big)$$

for all x in G, i.e., a_{ij}, a'_{kl} are the coordinate functions of φ, φ', respectively. Then:

(1) If φ and φ' are not equivalent, then

$$\sum_G a_{ij}(x) a'_{kl}(x^{-1}) = 0 \ \text{ for all } \ i, j, k, l.$$

(2) Suppose that $\operatorname{char} F$ *is zero or does not divide the order of G and $F[G]$ is F-split. Then n is nonzero in F and*

$$\sum_G a_{ij}(x) a'_{kl}(x^{-1}) = \delta ij\, \delta jk\, \frac{|G|}{n}.$$

PROOF. Let V with (ordered) basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ be afforded by φ and V' with (ordered) basis $\mathcal{B}' = \{v'_1, \ldots, v'_m\}$ be afforded by φ'. Check if $f : V \to V'$ is an F-linear transformation, then $Tf : V \to V$ given by

$$Tf = \sum_G \varphi'(x^{-1}) \circ f \circ \varphi(x),$$

i.e.,

$$Tf\big(\varphi(\sigma)v\big) = Tf\big(\sigma(v)\big) = \sigma Tf(v) = \varphi'(\sigma) Tf(v)$$

for all v in V and for all σ in G, is an $F[G]$-homomorphism. (Cf. Tr_G.) For fixed i and l, define an F-linear transformation f_o by

$$f_o : V \to V' \text{ given by } \begin{cases} v_i \mapsto v_l \\ v_j \mapsto 0 \quad \text{if } j \neq i. \end{cases}$$

(1): By assumption $V \ncong V'$ as $F[G]$-modules. Since V and V' are irreducible $F[G]$-modules, we must have $f_o = 0$. As $[f_o]_{\mathcal{B},\mathcal{B}'} = \big((\delta_{rl}\delta_{si})_{rs}\big) = \big((f_o)_{rs}\big)$ and

$$\begin{aligned} (Tf_o)_{kj} &= \sum_G \sum_{r,s} \big(\varphi'(x^{-1})\big)_{kr} (f_o)_{rs} \varphi(x)_{sj} \\ &= \sum_G \sum_{r,s} a'_{kr}(x^{-1})(f_o)_{rs}\delta_{rl}\delta_{si}a_{sj}(x) \qquad (*) \\ &= \sum_G a'_{kl}(x^{-1})a_{ij}(x), \end{aligned}$$

we have proven (1) as $Tf_o = 0$.

(2): As V is $F[G]$-irreducible and F-split, we have $F = \mathrm{End}_{F[G]}(V)$. Let $\mathcal{B}' = \mathcal{B}$ and $v'_i = v_i$ for all v_i, then $Tf_o = \lambda(f_o)I$ for some $\lambda(f_o)$ in F. Consequently, $\mathrm{trace}(Tf_o) = n\lambda(f_o)$. However, we also have

$$\begin{aligned} \mathrm{trace}(Tf_o) &= \mathrm{trace}\big(\sum_G \varphi(x^{-1})[f_o]_{\mathcal{B}}\varphi(x)\big) \\ &= \sum_G \mathrm{trace}\big(\sum_G \varphi(x^{-1})[f_o]_{\mathcal{B}}\varphi(x)\big) \\ &= |G|\,\mathrm{trace}([f_o]_{\mathcal{B}}) = |G|\,\delta_{il}. \end{aligned}$$

Hence n is nonzero in F and

$$Tf_o = \lambda(f_o)I = \frac{|G|}{n}\,\delta_{il}\,I.$$

Plugging this into (*) yields

$$\frac{|G|}{n}\,\delta_{il}\delta_{kj} = (Tf_o)_{kj} = \sum_G a_{kl}(x^{-1})a_{ij}(x). \qquad \square$$

Of course, we call two characters $\chi, \chi' : G \to F$ *inequivalent* if they are not equivalent.

Theorem 96.7. (Orthogonal Relations) *Let G be a finite group and χ, χ' two inequivalent characters.*

(1) If both χ and χ' are irreducible, then

$$\sum_G \chi(x)\chi'(x^{-1}) = 0.$$

(2) *Suppose that* $\operatorname{char} F$ *is zero or does not divide the order of* G *and* $F[G]$ *is* F*-split. If* χ *is irreducible of degree* n*, then* $\chi(1) = n \neq 0$ *in* F *and*

$$\frac{1}{|G|}\sum_G \chi(x)\chi(x^{-1}) = 1.$$

(3) *If* $\operatorname{char} F = 0$*, then*
   (a) $\sum_G \chi(x)\chi(x^{-1})$ *is an integer.*
   (b) $|G|$ *divides* $\sum_G \chi(x)\chi(x^{-1})$ *in* $\mathbb{Z}$*.*
   (c) *If* $|G| = \sum_G \chi(x)\chi(x^{-1})$*, then* χ *is irreducible.*

PROOF. Let $\chi = \chi_\varphi$ and $\chi' = \chi_{\varphi'}$ with $\varphi = (a_{ij})$ and $\varphi' = (a'_{ij})$ the representations giving χ and χ' respectively.

(1): By the theorem,

$$\begin{aligned}\sum_G \chi(x)\chi'(x^{-1}) &= \sum_G \Big(\big(\sum_i a_{ii}(x)\big)\big(\sum_j a_{jj}(x^{-1})\big)\Big) \\ &= \sum_{i,j}\sum_G a_{ii}(x)a'_{jj}(x^{-1}) = 0.\end{aligned}$$

(2): If $n = \chi(1)$, we know n is nonzero in F by the theorem and

$$\begin{aligned}\sum_G \chi(x)\chi(x^{-1}) &= \sum_G \Big(\big(\sum_i a_{ii}(x)\big)\big(\sum_j a_{jj}(x^{-1})\big)\Big) \\ &= \sum_{i,j}\sum_G a_{ii}(x)a'_{jj}(x^{-1}) = \sum_{ij}\frac{|G|}{n} = n\frac{|G|}{n} = |G|.\end{aligned}$$

(3): If $\widetilde{F}$ is an algebraic closure of F, then we can view χ as a character $\chi : G \to \widetilde{F}$. Let $\chi_1, \ldots, \chi_r$ be all the irreducible characters $G \to \widetilde{F}$. Then $\chi = \sum m_i\chi_i$ for some non-negative integers m_i. We see that

$$\begin{aligned}\sum_G \chi(x)\chi(x^{-1}) &= \sum_G \big(\sum_i m_i\chi_i(x)\big)\big(\sum_j m_j\chi_j(x^{-1})\big) \\ &= |G|\sum_i m_i^2\end{aligned} \tag{*}$$

by (1) and (2). Since $\sum m_i^2$ is an integer and characteristic of F is zero, we have $\sum_G \chi(x)\chi(x^{-1})$ and $|G|$ dividing $\sum_G \chi(x)\chi(x^{-1})$.

Finally suppose that $|G| = \sum_G \chi(x)\chi(x^{-1})$. Then by (*), we have $1 = \sum m_i^2$, so there exists a i such that $m_i = 1$ and $m_j = 0$ for all $j \neq i$, i.e., χ is irreducible as a character when viewed as $\chi : G \to \widetilde{F}$.

Claim. $\chi : G \to F$ is an irreducible:

Let V be the F-vector space with basis $\mathcal{B}$ afforded by φ (recall $\chi = \chi_\varphi$) and W a nonzero $F[G]$-submodule of V with F-basis $\mathcal{C}$. Let $\widetilde{W} \subset \widetilde{V}$ be the $\widetilde{F}$-vector spaces with bases $\mathcal{C}$ and $\mathcal{B}$, respectively. Both are $\widetilde{F}[G]$-modules with $\widetilde{W}$ a submodule of $\widetilde{V}$. As $\chi : G \to \widetilde{F}$ is afforded by $\widetilde{V}$ and is irreducible, we see that $\widetilde{W} = \widetilde{V}$. So

$$\dim_F W = \dim_{\widetilde{F}} \widetilde{W} = \dim_{\widetilde{F}} \widetilde{V} = \dim_F V < \infty.$$

It follows that $W = V$ and $\chi : G \to F$ is irreducible. □

Theorem 96.8. *Let G be a finite group. Suppose that* $\operatorname{char} F$ *is zero or does not divide the order of G and $F[G]$ is F-split. If $\chi_1, \dots, \chi_r$ are all the irreducible characters of G and x, y lie in G, then*

$$\sum_{i=1}^{r} \chi_i(x)\chi_i(y^{-1}) = \begin{cases} 0, & \text{if } y \notin \mathcal{C}(x). \\ |Z_G(x)|, & \text{if } y \in \mathcal{C}(x). \end{cases}$$

Proof. Let $x_1, \dots x_{r'}$ be a system of representatives for the conjugacy classes of G. As $F[G]$ is F-split, $r = r'$. Let $C = (c_{ij})$ with $c_{ij} = \chi_i(x_j)$ $\big($independent of the choice of $x \in \mathcal{C}(x_j)\big)$. This matrix is called the *character table* of G. Let $D = (d_{ij})$ with

$$d_{ij} = \frac{|\mathcal{C}(x_i)|}{|G|}\chi_j(x_i^{-1}) = \frac{1}{|Z_G(x_i)|}\chi_j(x_i^{-1}).$$

Check that

$$\begin{aligned} (CD)_{ij} &= \sum_k \frac{|\mathcal{C}(x_k)|}{|G|}\chi_i(x_k)\chi_j(x_k^{-1}) \\ &= \frac{1}{|G|}\sum_G \chi_i(x)\chi_j(x^{-1}) \\ &= \frac{1}{|G|}\,|G|\delta_{ij} = \delta_{ij}. \end{aligned}$$

Thus $CD = I$, hence $DC = I$ and

$$(DC)_{ij} = \sum_{k=1}^{r} \frac{1}{|Z_G(x_i)|}\chi_k(x_i^{-1})\chi_k(x_j).$$

Let $x \in \mathcal{C}(x_i)$ and $y \in \mathcal{C}(x_j)$. If $i \neq j$, we see that $(DC)_{ij} = 0$ leads to $\sum_{k=1}^{r} \chi_k(x)\chi_k(y^{-1}) = 0$ and if $i = j$, then $(DC)_{ii} = 1$ leads to

$$\frac{1}{|Z_G(x_i)|} = \frac{|G|}{|\mathcal{C}(x)|} = \sum_{k=1}^{r} \chi_k(x)\chi_k(x^{-1}) = \sum_{k=1}^{r} \chi_k(x)\chi_k(y^{-1}). \qquad \square$$

Remarks 96.9. Let G be a finite group. We look at complex characters for G. Let r be the number of conjugacy classes of G and $x_1, \ldots, x_r$ a system of representatives of the conjugacy classes. Let $\chi, \chi' : G \to \mathbb{C}$ be two characters. Then we have

$$\frac{1}{|G|}\sum_G \chi(x)\chi'(x^{-1}) = \sum_{i=1}^{r} \frac{\chi(x)\chi'(x^{-1})}{|Z_G(x_i)|}.$$

Define an inner product on $(\mathbb{C}[G])^* = \mathrm{Hom}_{\mathbb{C}}(\mathbb{C}[G], \mathbb{C})$, the dual space of $\mathbb{C}[G]$, by

$$\langle\ ,\ \rangle : (\mathbb{C}[G])^* \times (\mathbb{C}[G])^* \to \mathbb{C}$$

given by

$$\langle f, h\rangle = \frac{1}{|G|}\sum_G f(x)\overline{h}(x)$$

where $\overline{h}$ is defined by $\overline{h}(x) := \overline{h(x)}$ for x in $\mathbb{C}[G]$. Since $\chi(x)$ is a sum of roots of unity and $\chi(x^{-1}) = \overline{\chi(x)}$, we have

$$\langle \chi, \chi'\rangle = \frac{1}{|G|}\sum_G \chi(x)\overline{\chi(x^{-1})}$$

where we restrict $\langle\ ,\ \rangle$ to the subspace spanned by the characters. We also have

$$\overline{\chi} : G \to \mathbb{C} \text{ given by } \overline{\chi(x)} := \chi(x^{-1})$$

is a character. Moreover, χ is irreducible if and only if $\overline{\chi}$ is. So complex characters come in two flavors, *real*, i.e., those have values in $\mathbb{R}$ and *complex pairs* χ, $\overline{\chi}$. i.e., those complex characters arising with non-real values.

Examples 96.10. 1. Let G be the cyclic group $\langle a\rangle$ of order 3 and ω a cube root of unity. Then the character table over the complex numbers is:

	1	a	a^2
χ_1	1	1	1
χ_2	1	ω	ω^2
χ_3	1	ω^2	ω

2. Let G be the dihedral group $D_3 = \langle a, b \mid a^3 = 1 = b^2,\ bab^{-1} = a^{-1}\rangle$ with 1, a, b a system of representatives of the conjugacy classes. Then the character table over the complex numbers is:

	1	a	b
χ_1	1	1	1
χ_2	-1	1	-1
χ_3	2	-1	0

Exercise 96.11. 1. (Idempotent Theorem) Let G be a finite group. Suppose that $\operatorname{char} F$ is zero or does not divide the order of G and $F[G]$ is F-split. Let $\mathfrak{A}$ be an irreducible $F[G]$-module and $B_{\mathfrak{A}}$ the simple component corresponding to $\mathfrak{A}$. Then show the unit $f = 1_{B_{\mathfrak{A}}}$ of $B_{\mathfrak{A}}$ satisfies

$$f = \frac{\chi_{\mathfrak{A}}(1)}{|G|} \sum_G \chi_{\mathfrak{A}}(x^{-1})x.$$

2. Let G be the dihedral group $D_4 = \langle a, b \mid a^4 = 1 = b^2,\ bab^{-1} = a^{-1}\rangle$ with 1, a^2, a, b, ab a system of representatives of the conjugacy classes. Show the character table over the complex numbers is

	1	a^2	a	b	ab
χ_1	1	1	1	1	1
χ_2	1	1	1	-1	-1
χ_3	1	1	-1	1	-1
χ_4	1	1	-1	-1	1
χ_5	2	-2	0	0	0

97. Burnside's p^aq^b Theorem

Throughout this section, we let $\widetilde{F}$ denote an algebraic closure of the field F. Recall that if K is an algebraically closed field of characteristic zero, we shall view an algebraic closure of $\mathbb{Q}$ in K to lie in $\mathbb{C}$ by taking an appropriate isomorphic image.

Theorem 97.1. (Arithmetic Lemma) (Frobenius) *Let F be a field of characteristic zero with algebraic closure $\widetilde{F}$ and G a finite group such that $F[G]$ is F-split. If $\chi : G \to F$ is an irreducible character and x an element in G, then*

$$\frac{|\mathcal{C}(x)|}{\chi(1)}\chi(x) \text{ is an algebraic integer, i.e., lies in } \mathbb{Z}_{\widetilde{\mathbb{Q}}},$$

where $\widetilde{\mathbb{Q}}$ is the algebraic closure of $\mathbb{Q}$ in $\widetilde{F}$.

PROOF. Let $\chi = \chi_V$ with V the irreducible $F[G]$-module afforded by $\varphi : F[G] \to \mathrm{End}_F(V)$ and $C = C_x = \sum_{\mathcal{C}(x)} y$, the class sum of x in $F[G]$. We know that C lies in the center $Z\big(F[G]\big)$ by Lemma 92.7. By definition, if $z \in Z\big(F[G]\big)$, then $\varphi(y)\varphi(z) = \varphi(z)\varphi(y)$ for all $y \in F[G]$ means that $\varphi(z)$ lies in $\mathrm{End}_{F[G]}(V)$. As $F[G]$ is F-split, $\mathrm{End}_{F[G]}(V) = F1_V$ by Burnside's Theorem 93.4. Therefore, $\varphi(z)$ lies in $F1_V$ for all $z \in Z\big(F[G]\big)$. In particular, $\varphi(C) = \lambda 1_v$ for some λ in F. Let $n = \chi(1) = \dim V$ and $\mathcal{B}$ a (ordered) basis for V. We know that $\chi(x)$ lies in $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$ with $\widetilde{\mathbb{Q}}$ an algebraic closure of $\mathbb{Q}$ by Lemma 96.3, so

$$\chi(1)\lambda = n\lambda = \mathrm{trace}\left([\varphi(C)]_{\mathcal{B}}\right) = \mathrm{trace}\left([\varphi\big(\sum_{\mathcal{C}(x)}(y)\big)]_{\mathcal{B}}\right) = |\mathcal{C}(x)|\chi(x)$$

lies in $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$, hence

$$\lambda = \frac{|\mathcal{C}(x)|}{\chi(1)}\chi(x) \text{ lies in } \widetilde{\mathbb{Q}}.$$

We must show that λ lies in $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$.

We restrict the map φ to a ring homomorphism (hence a $\mathbb{Z}$-algebra homomorphism)

$$\varphi = \varphi|_{\mathbb{Z}[G]} : \mathbb{Z}[G] \to \mathrm{End}_F(V).$$

Since $C = \sum_{\mathcal{C}(x)} 1 \cdot z$ lies in $\mathbb{Z}(G)$, it must lie in $Z\big(\mathbb{Z}[G]\big)$. Further restrict φ to a ring homomorphism

$$\varphi = \varphi|_{Z(\mathbb{Z}[G])} : Z(\mathbb{Z}[G]) \to \mathrm{End}_F(V).$$

As above, we have $\varphi\Big(Z\big(\mathbb{Z}[G]\big)\Big) \subset \mathrm{End}_{F[G]}(V) = F1_V$, i.e., this restriction means $\varphi : Z(\mathbb{Z}[G]) \to F1_V = \mathrm{End}_{F[G]}(V)$ and C lies in $Z\big(\mathbb{Z}[G]\big)$. Since G is a finite group, $\mathbb{Z}[G]$ is a finitely generated abelian group, hence so is the subring $\varphi\Big(Z\big(\mathbb{Z}[G]\big)\Big)$ – using either $Z\big(\mathbb{Z}[G]\big)$ is generated by class sums or $\mathbb{Z}$ is noetherian. In particular, every element of $\varphi\Big(Z\big(\mathbb{Z}[G]\big)\Big)$ is integral over $\mathbb{Z}$ by Claim 72.4. So $\varphi(C) = \lambda 1_V$ implies that λ lies in $\widetilde{\mathbb{Q}} \cap \mathbb{Z}_F \subset \widetilde{\mathbb{Q}} \cap \mathbb{Z}_{\widetilde{F}} = \mathbb{Z}_{\widetilde{\mathbb{Q}}}$, where $\widetilde{F}$ is an algebraic closure of F. □

Theorem 97.2. *Let F be a field of characteristic zero and G a finite group such that $F[G]$ is F-split. If $\chi : G \to F$ is an irreducible character, then $\chi(1) \mid |G|$ in $\mathbb{Z}$.*

PROOF. Let $x_1, \ldots, x_r$ be a system of representatives for the conjugacy classes of G and $\widetilde{\mathbb{Q}}$ an algebraic closure of $\mathbb{Q}$. Then $\dfrac{|\mathcal{C}(x_i)|}{\chi(1)}\chi(x_i)$ and $\chi(x^{-1})$ lie in $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$, hence

$$\frac{|G|}{\chi(1)} = \sum_G \frac{1}{\chi(1)}\chi(x)\chi(x^{-1}) = \sum_{i=1}^{r} \frac{\mathcal{C}(x_i)}{\chi(1)}\chi(x)\chi(x^{-1})$$

lies in $\mathbb{Z}_{\widetilde{\mathbb{Q}}} \cap \mathbb{Q} = \mathbb{Z}$. □

Schur proved a stronger result, viz., under the hypothesis of the theorem, $\chi(1) \mid [G : Z(G)]$ in $\mathbb{Z}$.

Corollary 97.3. *Let F be a field of characteristic zero and G a finite group such that $F[G]$ is F-split, say $F[G] \cong \bigtimes_{i=1}^{r} \mathbb{M}_{n_i}(F)$. Then*

(1) r is the number of conjugacy classes of G.
(2) $|G| = \sum_{i=1}^{r} n_i^2$.
(3) There exists an i, $1 \le i \le r$ satisfying $n_1 = 1$.
(4) $n_i \mid |G|$ for $i = 1, \ldots, r$.

PROOF. We have previously shown (1) — (3) and (4) follows from the theorem. □

Lemma 97.4. (Burnside's Lemma) *Let F be an algebraically closed field of characteristic zero, G a finite group, and $\varphi : G \to \mathrm{GL}_n(F)$ an irreducible representation. If x in G satisfies $|\mathcal{C}(x)|$ is relatively prime to $n = \chi_\varphi(1)$, then either $\chi_\varphi(x) = 0$ or $\varphi(x) = \lambda I$ for some λ in $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$, with $\widetilde{\mathbb{Q}}$ the algebraic closure of $\mathbb{Q}$.*

PROOF. Let $\chi = \chi_\varphi$ and $m = |\mathcal{C}(x)|$, so m and n are relatively prime. We can write $1 = am + bn$ for some integers a, b; hence

$$\lambda := \frac{\chi(x)}{n} = a\frac{m}{n}\chi(b) + b\chi(x)$$

lies in $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$, with $\widetilde{\mathbb{Q}}$ an algebraic closure of $\mathbb{Q}$ by the Arithmetic Lemma. By Proposition 96.2 and the proof of Proposition 96.5, we know that $\chi(x) = \varepsilon_1 + \cdots + \varepsilon_n$ for some $|G|$th roots of unity ε_i in $\widetilde{\mathbb{Q}}$ that are also the eigenvalues of $\varphi(x)$ and satisfy

$$(*) \qquad |\lambda| = \frac{1}{n}|\chi(x)| = \frac{1}{n}|\varepsilon_1 + \cdots + \varepsilon_n| \le \frac{1}{n}\sum_{i=1}^{n} |\varepsilon_i|$$

with equality if and only if $\varepsilon_i = \varepsilon_j$ for all i, j.

Case 1. Not all the ε_i are equal:

By (*), we have $|\lambda| < 1$. Let ω be a primitive $|G|$th root of unity over $\mathbb{Q}$ and $L = \mathbb{Q}(\omega)$. Then λ lies in $\mathbb{Z}_{\widetilde{\mathbb{Q}}} \cap L = \mathbb{Z}_L$ and ε_k lies in $\mathbb{Z}_L$ for all k. Suppose that $\varepsilon_i \neq \varepsilon_j$. Since $L/\mathbb{Q}$ is a finite Galois extension (even abelian), $\sigma(\varepsilon_i) \neq \sigma(\varepsilon_j)$ for all σ in $G(L/\mathbb{Q})$. Hence by (*), we have $|\sigma(\lambda)| < 1$ for all σ in $G(L/\mathbb{Q})$. It follows that $\mathrm{N}_{L/\mathbb{Q}}(\lambda)| = \prod_{G(L/\mathbb{Q})} |\sigma(\lambda)| < 1$ and $\mathrm{N}_{L/\mathbb{Q}}(\lambda)$ lies in $\mathbb{Z}_L \cap \mathbb{Q} = \mathbb{Z}$, so $\mathrm{N}_{L/\mathbb{Q}}(\lambda) = 0$. Therefore, $\lambda = 0$ and $\chi_\varphi(x) = \chi(x) = \lambda n = 0$.

Case 2. All the ε_i are equal, say $\varepsilon = \varepsilon_i$:

We have $\lambda = \frac{1}{n}\sum_{i=1}^n \varepsilon_i = \varepsilon$ and ε is the only eigenvalue of $\varphi(x)$, so the characteristic polynomial of $\varphi(x)$ is $f_{\varphi(x)} = (t-\varepsilon)^n$ in $L[t]$. As $x^{|G|} = 1$, the matrix $\varphi(x)$ is a root of $t^{|G|} - 1$. Hence $\varphi(x)$ also a root of the gcd of $(t-\varepsilon)^n$ and $t^{|G|} - 1$ which is $t - \varepsilon$ in $L[t]$, since $t^{|G|} - 1$ has no multiple roots. Therefore, $\varphi(x) = \varepsilon I = \lambda I$. □

Theorem 97.5. (Burnside) *Let G be a finite group and p a prime. If there exists an element x in G satisfying $\mathcal{C}(x) = p^e > 1$, then G is not a simple group.*

Proof. Let $\chi_1, \ldots, \chi_r : G \to \mathbb{C}$ be all the distinct irreducible complex characters. One of these must be the trivial character, say it is χ_1. Let $n_i = \chi_i(1)$, so $n_1 = 1$ (and $\chi_1(x) = 1$). Since $\mathcal{C}(x) \neq \{x\}$, we know that $x \neq 1$, hence $1 \notin \mathcal{C}(x)$. Therefore, we have

$$0 = \sum_{i=1}^r \chi_i(x)\chi_i(1^{-1}) = \sum_{i=1}^r \chi_i(1)\chi_i(x) = 1 + \sum_{i=2}^r n_i\chi_i(x).$$

Case 1. There exists an integer $1 < j \leq r$ such that $p \nmid n_j$ and $\chi_j(x) \neq 0$:

By assumption, $\chi_j(1)$ and $|\mathcal{C}(x)|$ are relatively prime. Since $\chi_j(x) \neq 0$, by Burnside's Lemma, there exists a complex number λ satisfying $\varphi_j(x) = \lambda I$, where $\chi_j = \chi_{\varphi_j}$. In particular, $\varphi_j(x)$ lies in $Z\big(\mathbb{M}_{n_j}(\mathbb{C})\big)$, so $\varphi_j(x)\varphi_j(y) = \varphi_j(y)\varphi_j(x)$ for all $y \in G$. As $|\mathcal{C}(x)| > 1$, $x \notin Z(G)$, so there exists a $y \in G$ satisfying $1 \neq xyx^{-1}y^{-1}$ lying in $\ker\varphi_j$. As $j > 1$, χ_j is not the trivial character, so $1 < \ker\varphi_j = \ker\chi_j < G$ by Proposition 96.5. Consequently, G is not simple.

Case 2. For all $1 < j \leq r$, we have $\chi_j(x) = 0$ whenever $p \nmid n_j$:

In this case, we have

$$(*) \qquad 0 = 1 + \sum_{i=2}^r n_j\chi_j(x) = 1 + p\sum_{\substack{p>1 \\ p|n_j}} \frac{n_j}{p}\chi_j(x).$$

As n_j/p is an integer when $p|n_j$, we have $1/p$ lies in $\mathbb{Z}_{\mathbb{C}} \cap \mathbb{Q} = \mathbb{Z}$ by (*), a contradiction. Thus Case 2 cannot occur. □

Theorem 97.6. (Burnside's p^aq^b-Theorem) *Let G be a finite group of order p^aq^b with p, q distinct primes and a, b non-negative integers. Then G is a solvable group.*

PROOF. By our previous work and induction, it suffices to show that G is not simple if a and b are both positive. Let Q be a Sylow q-subgroup of G and $e_G \neq x$ an element in $Z(Q)$. Then $Q \subset Z_G(x)$, hence $|\mathcal{C}(x)| = [G : Z_G(x)] = p^n$ for some integer $n \geq 0$. If $n = 0$, then $Z_G(x) = G$, so $1 < Z(G) \triangleleft G$ as $x \neq e_G$, and the result follows. If $n > 0$, then G is not simple by Burnside's Theorem. □

When Burnside wrote the first edition of his historic book on finite group theory, he decided not discuss representation theory as it was not intrinsic to the theory of groups. After he proved the p^aq^b-Theorem in 1904, he realized that representation theory was now an essential tool in studying group theory, so he included it in a revised edition of his book. A proof of his theorem was proven avoiding representation in the 1970's. It is much more difficult than the one using representation theory.

98. Addendum: Torsion Linear Groups

In this section, we introduce further work of Burnside that focuses on the problem of conditions that force a group G to be finite. Of course, one would need that the group is a torsion group, i.e., every element has finite order. Burnside looked at the stronger condition on a group G, viz., one in which there exist a positive integer n such that $x^n = e$ for every element in the group G. Since the group $\bigtimes_{i=1}^{\infty} \mathbb{Z}/n\mathbb{Z}$ satisfies this condition yet is an infinite group, other conditions are necessary. Burnside showed if, in addition, such a group G was also a subgroup of $\mathrm{GL}_n(F)$, with F a field of characteristic zero, then G is a finite group. This is the principal result that we prove in this section. This led to a long research project on attempts to generalize it.

We begin with the following lemma:

Lemma 98.1. (Frobenius-Schur) *Let F be an algebraically closed field and G a (not necessarily finite) group, $\varphi : G \to \mathrm{GL}_n(F)$ an irreducible representation, and $a_{ij} : G \to F$, $1 \leq i, j \leq n$, the coordinate functions of φ, i.e., $\varphi = \big(a_{ij}\big)$. Then*

(1) $\mathbb{M}_n(F) = \langle \varphi(x) \mid x \in G \rangle$.

(2) $\mathcal{B} = \{a_{ij} \mid 1 \leq i, j \leq n\}$ *is a set of* F*-linearly independent functions.*

PROOF. (1): Extending φ to the F-algebra map $\varphi : F[G] \to \mathbb{M}_n(F)$ yields (1) by Burnside's Theorem 93.4.
(2): If $\mathcal{B}$ is F-linearly dependent, then it is F-linearly dependent as a set of F-functionals $F[G] \to F$, say

$$0 = \sum_{i,j} b_{ij} a_{ij}, \quad b_{ij} \in F \text{ not all zero.}$$

Suppose that $b_{i_0 j_0} \neq 0$. Let $e_{i_0 j_0}$ be the $i_0 j_0$th matrix unit. By (1), there exists an x in $F[G]$ such that $\varphi(x) = e_{i_o j_o}$. Therefore, we have

$$0 = \sum_{i,j} b_{ij} a_{ij}(x) = \sum_{i,j} b_{ij} \delta_{i i_0} \delta_{j j_0} = b_{i_0 j_0},$$

a contradiction. Hence $\mathcal{B}$ is F-linearly independent in $\mathrm{Hom}_F(F[G], F)$. □

Corollary 98.2. *Let* F *be an algebraically closed field and* G *a group. Suppose that* $\varphi : G \to \mathrm{GL}_n(F)$ *is an irreducible representation and* $\varphi = \big(a_{ij}\big)$, $1 \leq i, j \leq n$, *the coordinate functions of* G. *Then there exist* $x_1, \ldots, x_{n^2}$ *elements in* G *such that* $\{x_1, \ldots, x_{n^2}\}$ *is an* F*-basis for* $\mathbb{M}_n(F)$. *Moreover, if* $v_k = \varphi(x_l)$, $1 \leq k \leq n^2$, *is viewed in* F^{n^2}, *then* $\{v_1, \ldots, v_{n^2}\}$ *is a basis for* F^{n^2}.

Definition 98.3. Let G be a group. We say that G is of *bounded period* if there exists a positive integer N such that $x^N = e$ for all elements x in G.

Theorem 98.4. (Burnside) *Let* F *be a field of characteristic zero and* G *a subgroup of* $\mathrm{GL}_n(F)$ *(i.e., there exists a faithful representation of* G *of degree* n*). Suppose that* G *has bounded period* N. *Then* G *is a finite group. In fact,* $|G| \leq N^{n^3}$.

PROOF. We may assume that F is algebraically closed, the inclusion $\iota : G \subset \mathrm{GL}_n(F)$ affords the $F[G]$-module $V = F^n$, and $\iota = \big(a_{ij}\big)$ with the a_{ij} the coordinate functions of ι. Let $\chi = \chi_V$.
Case 1. V is an irreducible $F[G]$-module.
By Corollary 98.2, there exists $\{x_1, \ldots, x_{n^2}\} \subset G$, an F-basis for $\mathbb{M}_n(F)$ with $v_k = \big(a_{ij}(x_k)\big)$, $k = 1, \ldots n^2$, (viewed in F^{n^2}) giving a basis $\mathcal{B} = \{v_1, \ldots, v_{n^2}\}$ for F^{n^2}. As ι is a group homomorphism.

$$(*) \qquad \chi(x_k x) = \sum_{i=1}^{n} a_{ii}(x_k x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(x_k) a_{ji}(x), \;\; k = 1, \ldots, n^2$$

are n^2-linear equations in the unknowns $a_{ji}(x)$. As every $\chi(x)$, $x \in G$, is a sum of n roots of unity in μ_N, we have

$$s := \big|\{\chi(x) \mid x \in G\}\big| \leq N^n < \infty.$$

Since $\mathcal{B}$ is linearly independent in F^{n^2}, the matrix of coefficients of the system (*) is invertible, hence by Cramer's Rule each system (*) has a unique solution. It follows that

$$|G| = \big|\{a_{ij}(x) \mid x \in G\}\big| \leq s^{n^2} \leq N^{n^3}.$$

Case 2. V is not an irreducible $F[G]$-module.

As V is reducible, $n > 1$. Since G cannot act transitively on a basis for V, there exist representations $\varphi_i : G \to GL_{n_i}(F)$ with $1 = 1, 2$ each of degree less than n, satisfying

$$\iota(x) = \begin{pmatrix} \varphi_1(x) & * \\ 0 & \varphi_2(x) \end{pmatrix} = \begin{pmatrix} \varphi_1(x) & U(x) \\ 0 & \varphi_2(x) \end{pmatrix}$$

with $U : G \to F^{n_1 n_2}$. Set $G_i = \{\varphi_i(x) \mid x \in G\}$, $i = 1, 2$. Then G_i is a group of bounded period N of degree $n_i < n$, so finite with $|G_i| \leq N^{n_i^3}$ for $i = 1, 2$ by induction. Set $H_i = \ker \varphi_i$, then $G_i \cong G/H_i$ and $[G : H_i] < \infty$ for $i = 1, 2$. By Poincaré's Lemma $\big($Exercise 10.16(4)$\big)$, we have $[G : H_1 \cap H_2] < \infty$. Let $x \in H_1 \cap H_2$, then

$$\iota(x) = \begin{pmatrix} I & U(x) \\ 0 & I \end{pmatrix} \quad \textit{and} \quad 0 = \iota(x)^N = \begin{pmatrix} I & NU(x) \\ 0 & I \end{pmatrix}.$$

Consequently, $NU(x) = 0$ and $U(x) = 0$ as F is a field of characteristic zero. Therefore, $H_1 \cap H_2 = 1$, hence G is a finite group and the representation $\psi : G \to G_1 \times G_2$ given by $x \mapsto \big(\varphi_1(x), \varphi_2(x)\big)$ is a monomorphism. It follows that

$$|G| \leq |G_1||G_2| \leq N^{n_1} N^{n_2} \leq N^{n^3}. \qquad \square$$

Remark 98.5. In the above proof we did not needed that F be a field of characteristic zero in Case 1; and in Case 2, we only needed that $\mathrm{char}(F) \nmid N$. In particular, the lemma holds if V is irreducible or if V is reducible and $\mathrm{char}(F) \nmid N$.

Using the proof above, we obtain another theorem of Burnside.

Theorem 98.6. (Burnside) *Let F be an arbitrary field and G a subgroup of $\mathrm{GL}_n(F)$. Then G is finite if and only if G has finitely many conjugacy classes.*

Proof. Certainly if G is finite, then it has finitely many conjugacy classes, so we need only show the converse. We use the notation as in the proof of the previous theorem. So ι is the inclusion map. If the

character χ associated to ι is irreducible then the proof of Case 1 for Theorem 98 still works as $\{\chi(x) \mid x \in G\}$ is a finite set, which implies that G is a finite group. So we may assume that ι is not irreducible. As in the notation of Case 2 in the previous proof, we may assume that

$$\iota(x) = \begin{pmatrix} \varphi_1(x) & U(x) \\ 0 & \varphi_2(x) \end{pmatrix}$$

where $G_i = \{\varphi_i(x) \mid x \in G\}$, a representation of degree $n_i < n$, $\varphi : G_i \to \mathrm{GL}_{n_i}(F)$ and $U : G \to F^{n_1 n_2}$, $i = 1, 2$. Let $\psi : G \to G_1 \times G_2$ be given by $x \mapsto \big(\varphi_1(x), \varphi_2(x)\big)$ as before, and set $H = \ker \psi \lhd G$. Then $H \subset \ker \varphi_1 \cap \ker \varphi_2$. Suppose that $\left(\begin{smallmatrix} I & U(h) \\ 0 & I \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} I & U(h') \\ 0 & I \end{smallmatrix}\right)$ lie in H, then

$$\begin{aligned} \iota(x) &= \begin{pmatrix} I & U(h) \\ 0 & I \end{pmatrix} \begin{pmatrix} I & U(h') \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & U(h+h')) \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} I & U(h') \\ 0 & I \end{pmatrix} \begin{pmatrix} I & U(h) \\ 0 & I \end{pmatrix}. \end{aligned}$$

In particular, H is abelian and $H \subset Z_G\Big(\left(\begin{smallmatrix} I & U(h) \\ 0 & I \end{smallmatrix}\right)\Big)$. It follows that the conjugacy class $\mathcal{C}\Big(\begin{pmatrix} I & U(h) \\ 0 & I \end{pmatrix}\Big)$ satisfies

$$\Big|\mathcal{C}\Big(\begin{pmatrix} I & U(h) \\ 0 & I \end{pmatrix}\Big)\Big| = \Big[G : Z_G\Big(\begin{pmatrix} I & U(h) \\ 0 & I \end{pmatrix}\Big)\Big] \le [G : H].$$

As G has finitely many conjugacy classes, so do G_1 and G_2, hence both are finite by induction. Therefore, $[G : H] \le |G_1|\,|G_2| < \infty$. As G has finitely many conjugacy classes and H is an abelian subgroup of G, we have H is also finite. It follows that G is a finite group. □

As mentioned in the introduction to this section, $\bigtimes_{i=1}^{\infty} \mathbb{Z}/p\mathbb{Z}$ is an infinite group of bounded period p, so certainly not finitely generated. Burnside conjectured in 1902 that any finitely generated torsion group G was finite. This is called the Burnside Conjecture. (Torsion groups are also called *periodic groups*.) We next improve the Burnside's theorems in the linear group case.

We need two lemmas.

Lemma 98.7. *Let G be a finitely generated torsion group. Suppose that G contains an abelian subgroup H of finite index. Then G is a finite group.*

Proof. Let $G = \bigvee_{1=1}^{m} g_i H$ be a coset decomposition. As G is finitely generated, there exists a finite subset

$$\mathcal{S} = \{g_1, \ldots, g_m, g_{m+1}, \ldots, g_n\} \subset G$$

generating G and closed under taking inverses. For each ordered pair (i,j), $1 \leq i,j \leq n$, there exists an integer $r = r(i,j)$, $1 \leq r \leq m$, satisfying

$$(*) \qquad g_i g_j = g_r h_{ij},\ h_{ij} \in H.$$

Set

$$H_0 = \langle h_{ij} \mid h_{ij} \ \text{ occurring as in } (*)\rangle.$$

Therefore, H_0 is a finitely generated group. As H_0 is a subgroup of the abelian torsion subgroup H, it is finite. Suppose that (i,j,r) is as in (*) and let $1 \leq s \leq n$. Then for each $m = 1, \ldots, m$, there exists a positive integer $v = v(r,s)$ satisfying

$$g_s g_i g_j = g_s g_r h_{ij} = g_v h_{sr} h_{ij} \ \text{ lies in the coset } \ g_v H_0.$$

It follows by induction that the group

$$\langle g_1, \ldots, \ldots, g_n\rangle = \{\text{words in the } g_i\}$$

lies in $\bigcup_{i=1}^m g_i H_0$, hence $G \subset \bigcup_{i=1}^m g_i H_0$, so is finite. □

You should compare this proof with the proof of Theorem 15.3. We can use the lemma in the linear group case to see

Lemma 98.8. *Let F be a field and G a finitely generated torsion subgroup of $\mathrm{GL}_n(F)$. Then G has bounded period. In particular, if $\mathrm{char}(F) = 0$, then G is finite.*

Proof. Let Δ be the prime subfield of F. We may assume that $\Delta = \mathbb{Q}$ or $\Delta = \mathbb{Z}/p\mathbb{Z}$, depending on the characteristic of F. Since G is finitely generated, we may assume that F/Δ is a finitely generated field extension. In particular, there exists an intermediate field $F/F_0/\Delta$ with F_0/Δ a finitely generated purely transcendental field extension and F/F_0 a finite field extension. Let $r = [F : F_0]$. The subgroup G of $\mathrm{GL}_n(F)$ acts faithfully on F^n, so viewing F^n as F_0^{nr}, we have G acts faithfully on the finite dimensional F_0-vector space F_0^{nr}. In particular, we may assume that $F = F_0$ is a finitely generated purely transcendental extension.

If $x \in G \subset \mathrm{GL}_n(F)$, let q_x be the minimal polynomial of x in $F[t]$. In particular, $\deg q_x \leq \deg f_x = n$, where f_x is the characteristic polynomial of x in $\mathbb{M}_n(F)$.

Case 1. $\Delta = \mathbb{Q}$.

As x in G has finite order, the roots of q_x are roots of unity, so the coefficients of q_x lie in $\mathbb{Z}_{\widetilde{\mathbb{Q}}} \cap F$, where $\widetilde{\mathbb{Q}}$ is the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$ and $\mathbb{Z}_{\widetilde{\mathbb{Q}}}$ is the integral closure of $\mathbb{Z}$ in $\widetilde{\mathbb{Q}}$ by Remarks 72.14. Since $F/\mathbb{Q}$ is purely transcendental, $\mathbb{Z}_{\widetilde{\mathbb{Q}}} \cap F = \mathbb{Z}_{\widetilde{\mathbb{Q}}} \cap \mathbb{Q} = \mathbb{Z}$ as $\mathbb{Z}$ is an integrally

closed domain. It follows that q_x lies in $\mathbb{Z}[t]$ (as each irreducible factor of q_x does) by Proposition 73.4. The coefficients of q_x are elementary symmetric functions in the roots of q_x and these roots are roots of unity, hence the coefficients for each q_x must have bounded absolute value. So there can only be finitely many q_x with roots in a finite set of roots of unity.

Let ω be a root of q_x. Then $m_{\mathbb{Q}}(\omega) \mid q_x \mid f_x$, where $m_{\mathbb{Q}}(\omega)$ is the minimal polynomial of ω in $\mathbb{C}$. It follows that $\deg m_{\mathbb{Q}}(\omega) \leq n$ by the Cayley-Hamilton Theorem. As a primitive mth root of unity ε in $\mathbb{Q}$ satisfies $\deg m_{\mathbb{Q}}(\varepsilon) = \phi(m)$ and $\phi(m) \to \infty$ as $m \to \infty$, there can only be finitely many ω that are roots of the q_x, all uniformly bounded, i.e., $\{q_x \mid x \in G\}$ is a finite set. Therefore, there exists a positive integer N such that $x^N = 1$ for all x in G. In particular, the subgroup G of $GL_n(F)$ has bounded period. By Burnside's Theorem 98.4, G is a finite group.

Case 2. $\Delta = \mathbb{Z}/p\mathbb{Z}$ with $p > 0$.

As in Case 1, we have $q_x \in (\mathbb{Z}/p\mathbb{Z})[t]$ satisfies $\deg q_x \leq n$, so there exist finitely many such q_x, $x \in G$. The result follows by an analogous argument as in Case 1. □

Theorem 98.9. (Schur) *Let F be a field and G a finitely generated torsion subgroup of $\mathrm{GL}_n(F)$. Then G is a finite group.*

Proof. By Lemma 98.8, G has bounded period, so by Lemma 98.7, it suffices to show that G contains an abelian subgroup of finite index. By Remark 98.5, the only part of the proof of Burnside's Theorem 98.4 in the characteristic zero case that must be modified is the case that the inclusion $\iota : G \to GL_n(F)$ is not irreducible. But the proof of Burnside's Theorem 98.6 produces an abelian subgroup of finite index. □

In 1964, Golod-Shafarevich produced an infinite group on three generators in which every element has order a power of a prime, so the Burnside Conjecture is false. A modification of this conjecture was made to the conjecture that any finitely generated torsion group of bounded period is finite. Adrian-Novikov proved this too was false in 1968 (although Novikov claimed to have proven it in 1959). It was modified again to the so-called Restricted Burnside Problem which said: There exist only finitely many m-generator groups G of bounded period n. In particular, there exists an integer $N = N(m, n)$ such that $|G| < N$ for all such G. This was proven by Zelmanov in 1994 for which he won a Fields Medal.

Exercise 98.10. Complete the proof of Case 2 of Theorem 98.9.

CHAPTER XIX

Universal Properties and Multilinear Algebra

Given an algebraic object, it is usually very difficult to define a homomorphism from that object to another. Even if we have a well-defined set map, to show that it preserves the structure, i.e., is a homomorphism, we have to show that it preserves all the relations of that object under the map. We saw that free modules were those modules defined by the universal property that a homomorphism from a free module on a basis $\mathcal{B}$ was completely determined by where the basis $\mathcal{B}$ went. This meant that free modules had no non-trivial relations. Many structures in algebra can be defined by such universal properties, we shall call such definitions *categorical definitions*. For example, we also showed that the quotient field of a domain was defined by one (cf. Theorem 24.14). An algebraic object A is defined by a universal property (P) if it satisfies property (P) together with a map (or maps) and if C satisfies property (P), then there is a unique homomorphism(s) $A \to C$ (or $C \to A$) together with a commutative diagram (or commutative diagrams) respecting the given map(s). For example, F is a free R-module on a set $\mathcal{B}$ if there is a fixed set map $i : \mathcal{B} \to F$ and if M is an R-module and $j : \mathcal{B} \to M$ set map, then there exists a unique R-homomorphism $f : F \to M$ satisfying

$$\begin{array}{ccc} \mathfrak{B} & \xrightarrow{i} & F \\ & \searrow^{j} & \downarrow f \\ & & M \end{array}$$

commutes. This also means that if F' is another free R-module on $\mathcal{B}$ with $i' : \mathcal{B} \to F'$ the given map, then there exists a unique R-isomorphism $f : F \to F'$. Thus the module F together with the set map i satisfies: there is a canonical bijection

$$\mathrm{Hom}_{\mathrm{sets}}(\mathcal{B}, N) \to \mathrm{Hom}_R(F, N)$$

relative to $i : \mathcal{B} \to F$, where the left hand side are set maps and canonical means the the map given uniquely by the commutative diagram.

In this chapter, we shall investigate other universal properties. Many of the details will be left to the reader. As an application we shall define the determinant of an R-homomorphism when R is a commutative ring and satisfying those properties of determinants of matrices that you know.

99. Some Universal Properties of Modules

Let R be a ring. Given an R-homomorphism $f : M \to N$ of R-modules, we know that we should determine its kernel and image. We also saw that cokernels existed. The kernel was a submodule of M on which f vanished. Of course, f can vanish on other submodules. Does $\ker f$ satisfy a universal property? The answer is yes. Formally we may define the kernel of an R-homomorphism as follows:

Definition 99.1. Let $f : A \to B$ be an R-homomorphism of R-modules. A *kernel* of f is an R-module K together with a R-homomorphism $i : K \to A$ satisfying

$$\begin{array}{ccc} K & & \\ {\scriptstyle i}\downarrow & \searrow^{0} & \\ A & \xrightarrow[f]{} & B \end{array}$$

commutes (where 0 is the zero map), and satisfies the following universal property: If

$$\begin{array}{ccc} M & & \\ {\scriptstyle g}\downarrow & \searrow^{0} & \\ A & \xrightarrow[f]{} & B \end{array}$$

is a commutative diagram of R-modules and R-homomorphisms, then there exists a unique R-homomorphism $h : M \to K$ satisfying

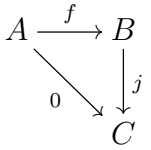
$$\begin{array}{ccccc} K & \xrightarrow{i} & A & \xrightarrow{f} & B \\ & {\scriptstyle h}\nwarrow & \uparrow{\scriptstyle g} & \nearrow{\scriptstyle 0} & \\ & & M & & \end{array}$$

commutes. Moreover, if the kernel exists, it is unique up to a unique isomorphism.

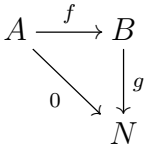
Our original definition of kernel satisfies the above with $K = \ker f$ and i the inclusion map. We have just made a choice of a representation of a kernel object. Since it is unique relative to the inclusion map, there can be no other one. Note also that f is a monomorphism if and only if the kernel is the zero module (of which there is only one).

We can define a cokernel of an R-homomorphism $f : A \to B$ by reversing arrows (called *duality*), i.e.,

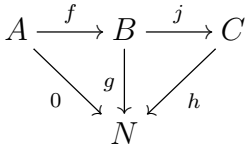
Definition 99.2. Let $f : A \to B$ be an R-homomorphism of R-modules. A *cokernel* of f is an R-module C together with a R-homomorphism $j : B \to C$ satisfying



commutes, and satisfying the following universal property: If



is a commutative diagram of R-modules and R-homomorphisms, then there exists a unique R-homomorphism $h : C \to N$ satisfying



commutes. Moreover, if the cokernel exists, it is unique up to a unique isomorphism.

Then our definition of cokernel is the unique quotient object $B/\operatorname{im} f$ relative to the canonical epimorphism. One problem here is that we want to define the image by a universal property, and it will depend on the definition of cokernel. Note that it is automatic that the map to a cokernel is surjective, so isomorphic to a quotient of B by some submodule. We define the image of an R-homomorphism as follows:

Definition 99.3. Let $f : A \to B$ be an R-homomorphism of R-modules. An *image* of f is kernel of $j : B \to C$, where C is a cokernel.

Remarks 99.4. By choice of nomenclature, a cokernel of a kernel of $f : A \to B$, should be called a *coimage* of $f : A \to B$. By the First Isomorphism Theorem, then one chosen is $A/\ker f$ under the canonical epimorphism defining the cokernel. Indeed in this categorical language, the First Isomorphism Theorem is the fact arising from the definitions that there is a canonical map from the coimage of an R-homomorphism to its image.

When we gave examples of modules, two that arose were the direct product and direct sum of modules. We now define these by categorical definitions.

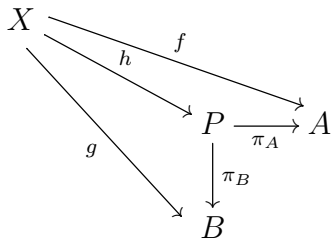
Definition 99.5. Let A and B be R-modules. Then a *product* of A and B is an R-module P together with R-homomorphisms

$$\begin{array}{ccc} P & \xrightarrow{\pi_A} & A \\ {\scriptstyle \pi_B}\downarrow & & \\ B & & \end{array}$$

satisfying the following universal property: If

$$\begin{array}{ccc} X & \xrightarrow{f} & A \\ {\scriptstyle g}\downarrow & & \\ B & & \end{array}$$

are R-homomorphisms, then there exists a unique R-homomorphism $h : X \to P$ such that



commutes. Moreover, if the product exists, it is unique up to a unique isomorphism and denoted by $A \prod B$.

Of course, this product is the external direct product previously defined via the projection maps.

The *coproduct* Y of two R-modules A and B is the dual notion of product defined by reversing arrows, i.e., it is a diagram

$$\begin{array}{ccc} A & \xrightarrow{\iota_A} & Y \\ & & \uparrow{\scriptstyle \iota_B} \\ & & B \end{array}$$

satisfying the obvious universal property. It is unique up to a unique isomorphism and denoted by $A \coprod B$. Of course, this coproduct is the external direct sum previously defined via the inclusion maps of coordinates.

One defines a product and a coproduct of an arbitrary collection $\{M_i\}_I$ of R-modules in the obvious way and denotes it by $\prod_I M_i$ and $\coprod_I M_i$ respectively. For all j, they come equipped with both epimorphisms $\pi_j : \prod_I M_i \to M_j$ and monomorphisms $\iota_j : M_j \to \coprod_I M_i$ respectively which we identify as the ones that previously arose. If I is finite these are isomorphic and are identified as before.

As an example of how to use universal properties, we show:

Lemma 99.6. *An R-module A is a coproduct of the R-submodules $A_1, \ldots, A_n$ if and only if there exist R-homomorphisms $\iota_j : A_j \to A$ and $\pi_j : A \to A_j$ for $j, k = 1, \ldots, n$ satisfying $\pi_k \iota_j = \delta_{kj} 1_{A_j}$ and $\sum_j \iota_j \pi_j = 1_A$.*

Proof. ($\Rightarrow$): The definition of coproduct gives rise to the ι_j's. For a fixed k and and each j, we have a diagram

$$\begin{array}{ccc} A_j & \xrightarrow{\delta_{kj} 1_{A_j}} & A_k \ . \\ {\scriptstyle \iota_j}\downarrow & & \\ A & & \end{array}$$

By the universal property of coproduct, there exists a unique R-homomorphism $\pi_k : A \to A_k$ such that

$$\begin{array}{ccc} A_j & \xrightarrow{\delta_{kj} 1_{A_j}} & A_k \\ {\scriptstyle \iota_j}\downarrow & \nearrow_{\pi_k} & \\ A & & \end{array}$$

commutes. By distributivity, $(\iota_1\pi_1 + \cdots + \iota_n\pi_n)\iota_j = \iota_j$. Then

$$\begin{array}{ccc} A_j & \xrightarrow{\iota_j} & A \\ & {\scriptstyle \iota_j}\searrow & \downarrow{\scriptstyle 1_A} \\ & & A \end{array} \qquad \text{and} \qquad \begin{array}{ccc} A_j & \xrightarrow{\iota_j} & A \\ & {\scriptstyle \iota_j}\searrow & \downarrow{\scriptstyle \iota_1\pi_1 + \cdots + \iota_n\pi_n} \\ & & A \end{array}$$

both commute for each j. Uniqueness gives $1_A = \iota_1\pi_1 + \cdots + \iota_n\pi_n$.

($\Leftarrow$): If $f_j : A_j \to B$ is an R-homomorphisms for each j, then defining $f : A \to B$ by $f = \sum_i f_i \pi_i$ works. □

Suppose that we have R-homomorphisms of R-modules

$$A \xrightarrow{f} B \xrightarrow{h} D \quad \text{and} \quad A \xrightarrow{g} C \xrightarrow{j} D,$$

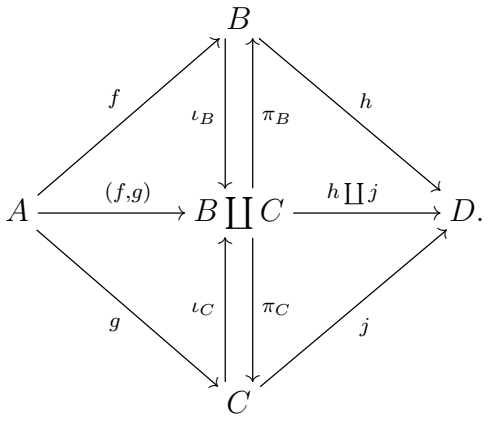
then by the universal properties of product and coproduct, there exist unique R-homomorphisms

$$A \xrightarrow{(f,g)} B \coprod C \xrightarrow{h \coprod j} D$$

given by

$$(f,g) := \iota_B f + \iota_C g$$
$$h \coprod j := h\pi_B + j\pi_C,$$

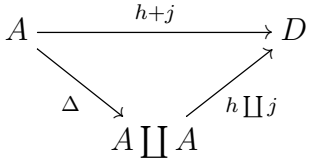
i.e., as we are identifying $A \prod B$ and $A \coprod B$, i.e., we have a diagram



In particular, we have

$$(h \coprod g)(f,g) = hf + jg.$$

An interesting example of this is when $A = B = C$ and $f = g = 1_A$. Set $\Delta = (1_A, 1_A) : A \to A \coprod A$, called the *diagonal map*, then the above shows



commutes. Note that this commutative diagram categorically determines addition in $\mathrm{Hom}_R(A, D)$.

Exercise 99.7. 1. Define a free group, i.e., a group satisfying the obvious definition. Do you know of any examples of such an item.
2. Define the product and coproduct of a collection of R-modules $\{M_i \mid i \in I\}$ and show that these are the external direct product and external direct sum of R-modules.
3. Define the product and coproduct of two groups. Can you identify them?

4. Let $f : M \to N$ be an R-homomorphism of R-modules. Show that there exists a natural R-isomorphism $\operatorname{coim} f \to \operatorname{im} f$ arising from universal properties without using the First Isomorphism Theorem.
5. Let R be a commutative ring, $S \subset R$ a multiplicative set, and M an R-module. Let $S^{-1}M := \{\frac{m}{s} \mid m \in M, s \in S\}$. Define an addition and R-action on $S^{-1}M$ making it an R-module and an $S^{-1}R$-module. This is called the *localization of M at S*. Determine the universal property that such a module satisfies.

100. Tensor Products

Let V be a vector space over F. If v and w are vectors in V, in general, there is no way to define a product of v and w in V. We want to rectify this, by creating a vector space in which a 'product' of v and w makes sense, called the tensor product. We shall do this in generality based upon a universal property. For vector spaces, as they are free modules, what we get is what one desires, but in general, unexpected results can occur.

Let $M_1, \ldots, M_n$, and N be R-modules and $M_1 \times \cdots \times M_n$ the cartesian product of the M_i. A map

$$f : M_1 \times \cdots \times M_n \to N$$

is called *(R-) n-linear* or (R-) *n-multilinear* if f is linear in each variable, i.e.,

$$f(x_1, \ldots, rx_i + x_i', \ldots, x_n) = rf(x_1, \ldots, x_i, \ldots, x_n) + f(x_1, \ldots, x_i', \ldots, x_n)$$

for all $x_j \in M_j$, $x_{i'} \in M_i$, and $r \in R$. If $n = 2$ an n-linear form is called an *R-bilinear form*. If R is a commutative ring we would like to convert bilinear maps into R-homomorphisms. This is done as follows:

Definition 100.1. Let R be a commutative ring and M, N two R-modules. An R-module T is called a *tensor product* of M and N if there exists an R-bilinear map $\iota : M \times N \to T$ satisfying the following universal property: If A is an R-module and $j : M \times N \to A$ an R-bilinear map, then there exists a unique R-homomorphism $f : T \to A$ such that

$$\begin{array}{ccc} M \times N & \xrightarrow{\iota} & T \\ & \searrow^{j} & \downarrow f \\ & & A \end{array}$$

commutes. If a tensor product $\iota : M \times N \to T$ exists, it is unique up to a unique isomorphism and T is denoted by $M \otimes_R N$.

Lemma 100.2. *Let R be a commutative ring and M, N two R-modules. Then a tensor product of M and N exists.*

PROOF. Let P be the free R-module on basis $\mathcal{B} = M \times N$. Let W be the submodule of P generated by the following:

$$\begin{aligned}(m + m', n) - (m, n) - (m', n)\\ (m, n + n') - (m, n) - (m, n')\\ (rm, n) - r(m, n)\\ (m, rn) - r(m, n)\end{aligned}$$

for all m, m' in M, n, n' in N, and r in R. Let $^- : P \to P/W$ by $x \mapsto \overline{x} = x + W$ be the be the canonical R-epimorphism. Then the composition ι of the maps $M \times N \xrightarrow{inc} P \xrightarrow{-} P/W$ is R-bilinear. Set $T = P/W$ and $M \times N \to T$. We denote $\iota(m, n)$ by $m \otimes n$. Then every element in T is a finite sum

$$\sum_{\text{finite}} r_i(m_i \otimes n_i) = \sum_{\text{finite}} (r_i m_i) \otimes n_i = \sum_{\text{finite}} m_i \otimes (r_i n_i),$$

for appropriate elements m_i in M, n_i in N, and r_i in R, as $r(m \otimes n) = (rm) \otimes n = m \otimes (rn)$ for all m in M, n in N, and r in R. Suppose that $j : M \times N \to A$ is R-bilinear. Define $f : T \to A$ as follows: By the universal property of freeness on basis $M \times N$, there exists a unique R-homomorphism $g : P \to A$ satisfying

$$\begin{array}{ccc} M \times N & \xrightarrow{\iota} & P \\ & \searrow^{j} & \downarrow{\scriptstyle g} \\ & & A \end{array}$$

commutes. Now, with $^- : P \to P/W$ the canonical R-epimorphism,

$$\begin{array}{ccccc} M \times N & \longrightarrow & P & \xrightarrow{-} & P/W \\ & \searrow^{j} & \downarrow{\scriptstyle g} & & \\ & & A & & \end{array}$$

and clearly, $g|_{\ker -} = 0$, i.e., $g|_W = 0$ as j is R-bilinear. Hence

$$\begin{array}{ccccc} 0 \longrightarrow W & \xrightarrow{\text{inc}} & P & \xrightarrow{-} & P/W \\ & \searrow^{0} & \downarrow{\scriptstyle g} & & \\ & & A & & \end{array}$$

is exact. But P/W is the cokernel of the inclusion map $\text{inc} : W \to P$, so there exists a unique R-homomorphism $\overline{g} : P/W \to A$ satisfying

$$\begin{array}{ccc} P & \xrightarrow{-} & P/W \\ {\scriptstyle g}\downarrow & \swarrow{\scriptstyle \overline{g}} & \\ A & & \end{array}$$

commutes. As

$$\begin{array}{ccc} M \times N & \xrightarrow{\iota} & P/W \\ & {\scriptstyle j}\searrow & \downarrow{\scriptstyle \overline{g}} \\ & & A \end{array}$$

commutes and $\overline{g}$ is unique, we have established existence.
(If you do not like using the universal property of cokernels, you can prove this directly, i.e., if A is a submodule of an R-module B and $f : B \to C$ is an R-homomorphism satisfying $A \subset \ker \varphi$, then there exists a unique R-homomorphism $\overline{\varphi} : B/A \to C$ satisfying

$$\begin{array}{ccc} B & \longrightarrow & C. \\ {\scriptstyle -}\downarrow & \nearrow & \\ B/A & & \end{array}$$

commutes.) We leave the proof of uniqueness to the reader. □

We adopt the notation above, i.e., elements of $M \otimes_R N$ are finite sums $\sum r_i(m_i \otimes n_i)$, etc.

Remark 100.3. Let R be a commutative ring.

1. If M, N, and A are R-modules, by the universal property of tensor products, the module structure of $M \otimes_R N \to A$ is determined by a R-bilinear map $M \times N \to A$ (hence on generators (m, n) in $M \times N$). In particular, an R-homomorphism $\varphi : M \otimes_R N \to A$, is completely determined by the $\varphi(m \otimes n)$.
2. Let M_i, N_i be R-modules, and $\varphi_i : M_i \to N_i$ an R-homomorphism for $i = 1, 2$. By the universal property of tensor products, there exists a unique homomorphism $\varphi_1 \otimes \varphi_2 : M_1 \otimes_R M_2 \to N_1 \otimes_R N_2$ induced by $m_1 \otimes m_2 \mapsto \varphi(m_1) \otimes \varphi(m_2)$ as $M_1 \times M_2 \to N_1 \otimes_R N_2$ given by $(m_1, m_2) \mapsto \varphi(m_1) \otimes \varphi(m_2)$ is R-bilinear.

If R is a commutative ring, then the lemma shows that there exists a bijection of sets

$$\{f : M \times N \to A \mid f \text{ an } R\text{-bilinear map}\} \\ \to \{f : M \otimes_R N \to A \mid f \text{ an } R\text{-bilinear map}\}$$

i.e., $\otimes_R$ converts R-bilinear maps to R-linear maps, as desired.

Properties 100.4. Let R be a commutative ring and P, M, N R-modules.

1. $0 \otimes 0 = m \otimes 0 = 0 \otimes n$ is the zero of $M \otimes_R N$ for any m in M, any n in N.
2. The map $\iota_M : R \otimes_R M \to M$ induced by $r \otimes m \mapsto rm$ is a *canonical R-homomorphism*, where canonical means if $f : M \to N$ is an R-homomorphism, then we have commutative diagram

$$\begin{array}{ccc} R \otimes_R M & \xrightarrow{\iota_M} & M \\ {\scriptstyle 1_R \otimes \varphi}\downarrow & & \downarrow{\scriptstyle \varphi} \\ R \otimes_R N & \xrightarrow{\iota_N} & N. \end{array}$$

   (Why is it well defined?) Check that $M \to R \otimes_R M$ given by $m \mapsto 1 \otimes m$ is the inverse.
3. Let M_i, $i \in I$, N be R-homomorphisms. Then there exists a canonical R-isomorphism

$$\coprod_I (M_i \otimes_R N) \to (\coprod_I M_i) \otimes_R N,$$

   i.e, $\prod_I$ and $\otimes_R$ commute. (Similarly, on the other side.):

   Since the map

$$(\coprod_I M_i) \times N \to \coprod_I (M_i \otimes_R N_i) \text{ given by } \{\{m_i\}_I, n\} \mapsto \{m_i \otimes n\}_I$$

   is R-bilinear, the universal property of the tensor products induces a unique R-homomorphism $f : (\coprod_I M_i) \otimes_R N \to \coprod_I (M_i \otimes_R N_i)$. For each $j \in I$, we have an R-bilinear map $M_j \times N \to (\coprod_I M_i) \otimes_R N$ given by $(m, n) \mapsto \{\delta_{ij} m_i\} \otimes n$, inducing a unique R-homomorphism $M_j \otimes_R N \to (\coprod_I M_i) \otimes_R N$ by the universal property of tensor products. Hence, by the universal property of coproducts, there exists a unique R-homomorphism $g : \coprod_I (M_i \otimes_R N) \to (\coprod_I M_i) \otimes_R N$. Check that f and g are inverses of each other.
4. $M \otimes_R N$ is canonically isomorphic to $N \otimes_R M$ induced by the R-bilinear map $M \times N \to N \otimes_R M$ given by $(m, n) \mapsto n \otimes m$.

5. $(M \otimes_R N) \otimes_R P$ is canonically isomorphic to $M \otimes_R (N \otimes_R P)$
6. $(\coprod_I R) \otimes_R M \cong \coprod_I (R \otimes_R M) \cong \coprod_I M$ with all R-isomorphisms canonical.
7. Suppose that M is a free R-module on basis $\mathcal{B}$ and N is a free R-module on basis $\mathcal{C}$. Then $M \otimes_R N$ is a free R-module on basis $\{b \otimes c \mid b \in \mathcal{B}, c \in \mathcal{C}\}$. In particular,
$$\operatorname{rank}_R(M \otimes_R N) = \operatorname{rank}_R M \operatorname{rank}_R N :$$
As $R \cong Rx$ for all x in $\mathcal{B}$ or $\mathcal{C}$, we have $M \cong \coprod_{\mathcal{B}} R$ and $N \cong \coprod_{\mathcal{C}} R$. Therefore,
$$M \otimes_R N \cong (\coprod_{\mathcal{B}} R) \otimes_R (\coprod_{\mathcal{C}} R) \cong \coprod_{\mathcal{B}} \left(R \otimes_R (\coprod_{\mathcal{C}} R)\right) \cong \coprod_{\mathcal{B}} \coprod_{\mathcal{C}} R.$$
8. If M, N are not free R-modules, bad things can occur. For example, the abelian group $\mathbb{Z}/2\mathbb{Z} \otimes_{\mathbb{Z}} /3\mathbb{Z} = 0$. For if $[a]_n$ is the congruence class of the integer a modulo n, then $[1]_2 \otimes [1]_3 = [1]_2 \otimes 4[1]_3 = 4[1]_2 \otimes [1]_3 = 0$. Similarly, $[1]_2 \otimes [-1]_3 = 0$.
9. (Base Extension) Let $\varphi : R \to T$ be a ring homomorphism of commutative rings. We know that any T-module B becomes an R-module via the pullback, i.e., $rx := \varphi(r)x$ for all $r \in R$, $x \in B$, i.e., via φ. In particular, T becomes an R-module, hence so does $T \otimes_R M$. We can now make $T \otimes_R M$ into a T-module by defining
$$\alpha(\beta \otimes m) := (\alpha\beta) \otimes m,$$
for all α and β in T and m in M. We call the T-module $T \otimes_R M$ the *base extension* of the R-module M to a T-module (via φ).

Examples 100.5. Let K/F be a field extension. Then K is an F-vector space, say on basis $\mathcal{C}$. Let V be an F-vector space on basis $\mathcal{B}$. The $K \otimes_F V$ is an F-vector space on basis $\{c \otimes b \mid c \in \mathcal{C}, b \in \mathcal{B}\}$, so $\dim_F K \otimes_F V = \dim_F K \dim_F V$ and $K \otimes_F V$ is a K-vector space on basis $\{1 \otimes b \mid b \in \mathcal{B}\}$. For example, $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$ is a four dimensional real vector space and a 2-dimensional complex vector space.

Question 100.6. Can you make $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$ into a ring? If so, is it a field?, a domain?, commutative?

Proposition 100.7. *Let V and W be finite dimensional F-vector spaces, $V^* = \operatorname{Hom}_F(V, F)$, the dual space of V (or finitely generated R-free modules with R commutative). Then the natural map*
$$\varphi : V^* \otimes_F W \to \operatorname{Hom}_F(V, W)$$
induced by $f \otimes w \mapsto (v \mapsto f(v)w)$ is an F-linear isomorphism.

PROOF. Although the map in the proposition is independent of any basis, we use bases to prove it. Let $\mathcal{B} = \{v_1, \dots, v_n\}$ be an F-basis for V and $\mathcal{B}^* = \{f_1, \dots, f_n\}$ the dual F-basis for V^*, i.e., $f_i(v_j) = \delta_{ij}$, so f_i is the coordinate function on v_i. Therefore, if v is an element of V, $v = \sum_i f_i(v)v_i$. Define

$$\psi : \mathrm{Hom}_F(V, W) \to V^* \otimes_F W \text{ by } T \mapsto \sum_i f_i \otimes Tv_i.$$

Check: φ and ψ are both F-linear.
Then we have

$$\varphi\psi T(v) = \varphi(\sum f_i \otimes Tv_i)(v) = \sum f_i(v)Tv_i = \sum (T(f_i(v)v_i) = T(v),$$

so $\varphi\psi = 1_{\mathrm{Hom}_F(V,W)}$. Hence φ is an epimorphism. As

$$\dim_F(V^* \otimes_F W) = \dim \mathrm{Hom}_F(V, W) < \infty,$$

φ is an F-isomorphism (as both are finite dimensional F-vector spaces). [If V and W are only assumed to be finitely generated free R-modules, one must show that $\psi\varphi = 1_{V^* \otimes_R W}$.] □

Remarks 100.8. We look at a interesting special case of the proposition above. Let $V = W$ and $\mathcal{B}, \mathcal{B}^*$ as in the proposition. Then the proposition gives an F-linear transformation

$$\varphi : V^* \otimes_F V \to \mathrm{End}_F(V).$$

Moreover, the inverse map ψ satisfies $\psi(1_{\mathrm{End}_F(V)}) = \sum f_i \otimes v_i$. Let

$$\mathrm{End}_F(V) \xrightarrow{\psi} V^* \otimes_f V \xrightarrow{\rho} F,$$

where ρ is induced by $f \otimes v \mapsto f(v)$. It follows that for $T \in \mathrm{End}_F(V)$, we have

$$\rho\psi(T) = \rho(\sum f_i \otimes Tv_i) = \sum f_i(Tv_i).$$

In particular, if $Tv_i = \sum \alpha_{ji}v_j$, then

$$\begin{aligned}\rho\psi(T) &= \sum_i f_i(\sum_j \alpha_{ji}v_j) = \sum_{i,j} \alpha_{ji}f_i(v_j) \\ &= \sum_i \alpha_{ii} = \mathrm{trace}\,[T]_{\mathcal{B}}.\end{aligned}$$

Since φ is defined independently of $\mathcal{B}$ and ψ is its inverse, ψ is also independent of $\mathcal{B}$. therefore, $\rho\psi(T)$ is independent of $\mathcal{B}$, i.e., the *trace* of T defined by

$$\mathrm{trace}\, T := \rho\psi(T) = \mathrm{trace}\,[T]_{\mathcal{B}}$$

is independent of $\mathcal{B}$. Since $\psi(1_{\mathrm{End}_F(V)}) = \sum f_i \otimes v_i$, we have $\varphi(f_i \otimes v_i) = 1_{\mathrm{End}_F(V)}$. The element $\sum f_i \otimes v_i$ is called the *Casimir element*. It too is independent of $\mathcal{B}$ and $\mathcal{B}^*$

Exercise 100.9. 1. Prove that the tensor product of two modules over a commutative ring is unique up to a unique isomorphism.

2. Show $0 \otimes 0 = m \otimes 0 = 0 \otimes n$ is the zero of $M \otimes_R N$ for any m in M, any n in N.
3. Prove Property 100.4(5).
4. Let m and n be positive integer and d the greatest common divisor of m and n. Then $\mathbb{Z}/m\mathbb{Z} \otimes_{\mathbb{Z}} \mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/d\mathbb{Z}$ (but not canonically).
5. In the proof of Proposition 100.7, show that $\psi\varphi = 1_{V^* \otimes_R W}$ if V and W are finitely generated free R-modules.
6. Let R be a commutative ring, M, N, P R-modules. The there exists a canonical bijection

$$\mathrm{Hom}_R\big(M, \mathrm{Hom}_R(N, P)\big) \cong \mathrm{Hom}_R(M \otimes_R N, P).$$

We say that $- \otimes_R N$ and $\mathrm{Hom}_R(N, -)$ are *adjoints*.

7. Let R be a commutative ring, M an R-module, and

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

a short exact sequence of R-modules. Show that

$$A \otimes_R M \xrightarrow{f \otimes 1_M} B \otimes_R M \xrightarrow{g \otimes 1_M} C \otimes_R M \to 0$$

and

$$M \otimes_R A \xrightarrow{1_M \otimes f} M \otimes_R B \xrightarrow{1_M \otimes g} M \otimes_R C \to 0$$

are exact.

8. Let R be a commutative ring, $S \subset R$ a multiplicative set, and M an R-module. Show that there exists a natural $S^{-1}R$-isomorphism $S^{-1}M \to S^{-1}R \otimes_R M$. (Cf. Exercise 99.7(5).) Show further that if

$$0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$$

is a short exact sequence of R-modules, then so is

$$0 \to S^{-1}R \otimes_R M' \xrightarrow{1 \otimes f} S^{-1}R \otimes_R M \xrightarrow{1 \otimes g} S^{-1}R \otimes_R M'' \to 0.$$

9. Let R be a commutative ring and N an R-module. We say that N is a *flat* R-module if whenever

$$0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$$

is exact, so is

$$0 \to M' \otimes_R N \xrightarrow{f \otimes 1_N} M \otimes_R N \xrightarrow{g \otimes 1_N} M'' \otimes_R N \to 0.$$

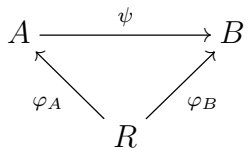
Show that if N is a projective R-module, then it is R-flat.

10. Let R be an arbitrary ring, i.e., not necessarily commutative, M a (left) R-module and N a right R-module. Define $M \otimes_R N$ and show that it exists.

101. Tensor, Symmetric, and Exterior Algebras

In the last section, we defined the tensor product of modules over a commutative ring. In this section, we show how to construct algebras from given algebras and given modules.

First let us recall a ring A over a commutative ring R is called an *R-algebra* if there exists a ring homomorphism $\varphi_A : R \to Z(A)$ (where $Z(A)$ is the center of A) and an *R-algebra homomorphism* of R-algebras is a ring homomorphism $\psi : A \to B$ of R-algebras such that



commutes. The algebra structure on A makes A into a left and right R-module. For example, every ring is a $\mathbb{Z}$-algebra. If R is a field F and A a nonzero F-algebra, we often identify F and $F1_A$, i.e., view φ_A as an inclusion as before. An R-algebra A is called a *graded R-algebra* if $A = \coprod_{i=0}^{\infty} A_i$ as an additive group with $\varphi_A(R) \subset A_0$ and $A_i A_j \subset A_{i+j}$ for all i, j. In particular, A_0 is a ring. Elements in A_i are called *homogeneous elements of degree i*. If $B = \coprod_{i \geq 0} B_i$ is another graded R-algebra, an R-algebra homomorphism $\psi : A \to B$ is called *graded* if $\varphi(A_i) \subset B_i$ for all $i \geq 0$. We want to construct some graded R-algebras.

Let R be a commutative ring and S a set. Then the *free commutative R-algebra* on S is the R-algebra satisfying the appropriate universal property — what is it? It exists and is unique up to a unique isomorphism, i.e., if two R-algebras satisfy the universal property there is a unique R-algebra isomorphism between them. Indeed, $R[t_s]_S$, the polynomial ring on $|S|$-variables, works.

Construction 101.1. Let R be a commutative ring and A, B two R-algebras. If $(a, b) \in A \times B$, then

$$m_{a,b} : A \times B \to A \otimes_R B \text{ given by } (a', b') \mapsto aa' \otimes bb'$$

is R-bilinear, so induces a unique R-homomorphism

$$\mu_{a,b} : A \otimes_R B \to A \otimes_R B \text{ given by } a' \otimes b' \mapsto aa' \otimes bb'$$

and hence

$(A \otimes_R B) \times (A \otimes_R B) \to A \otimes_R B$ given by $(a \otimes b, a' \otimes b') \mapsto aa' \otimes bb'$

is a well defined map that makes $A \otimes_R B$ into a ring and into an R-algebra by $R \to A \otimes_R B$ by $r \mapsto r \otimes 1$, where $r \otimes 1 = r1_A \otimes 1_B = 1_A \otimes r1_B = 1 \otimes r = \varphi_A(r)1_A \otimes 1_B = 1_A \otimes \varphi_B(r)1_B$ for all r in R. We also call this algebra the *tensor product* of A and B.

Definition 101.2. Let R be a commutative ring and M, N be R-modules. For each positive integer n, let $T^n(M)$ be the R-module $\underbrace{M \otimes_R \cdots \otimes_R M}_{n}$. Then $T^n(M)$ is generated by $m_1 \otimes \cdots \otimes m_n$ with $m_i \in M$. Define $T^0(M) := R$. If $f : M \to N$ is an R-homomorphism, then there exists a unique R-homomorphism $T^n f : T^n(M) \to T^n(N)$ induced by the (R-)n-linear map $M_1 \times \cdots \times M_n \to T^n(N)$ given by $(m_1, \cdots, m_n) \mapsto f(m_1) \otimes \cdots \otimes f(m_n)$ (why?).

Example 101.3. Let n be a positive integer, V an m-dimensional F-vector space on basis $\mathcal{B} = \{v_1, \ldots, v_m\}$. Then $T^n(V)$ is an F-vector space basis $\{v_{i_1} \otimes \cdots \otimes v_{i_n} \mid 1 \leq i_j \leq m\}$ (repetitions are allowed), a vector space of dimension m^n. (The same is true if we have M a free R-module of on a basis $\mathcal{B}$ with R a commutative ring.) We also define $T^{-n}(V) := T^n(V^*)$, with V^* the dual space $\mathrm{Hom}_F(V, F)$.

Construction 101.4. Let R be a commutative ring and M an R-module. Associativity of the tensor products implies that

$$T^m(M) \times T^n(M) \to T^{m+n}(M) \text{ given by } (x, y) \mapsto x \otimes y$$

is R-bilinear, so induces a (non-commutative) ring structure on the R-module

$$T(M) := \coprod_{i=0}^{\infty} T^n(M)$$

called the *tensor algebra* of M. If $x, y \in T(M)$, we write xy for the product. Hence if $x \in T^m(M)$ and $y \in T^n(M)$, then $xy = x \otimes y$. Under this ring structure, we have

$$\begin{aligned} 0_{T(M)} &:= \coprod_{n=0}^{\infty} 0_{T^n(M)} \\ 1_{T(M)} &:= 1_R \\ R &:= T^0(M), \text{ a subring of } T(M) \\ T(M) &\text{ is a } R\text{-module} \\ T(M) &\text{ is an } R\text{-algebra.} \end{aligned}$$

The R-algebra $T(M)$ also satisfies the following universal property: If $\varphi : M \to A$ is a R-homomorphism of R-algebras, then there exists a unique R-algebra homomorphism $\Phi : T(M) \to A$ satisfying $\Phi|_M = \varphi$.

Let R be a commutative ring and M an R-module. As the tensor algebra $T(M)$ is a ring, it has ideals, hence quotient rings. We use this to construct other algebras. But $T(M)$ has an additional structure, it is a graded R-algebra by $T(M) = \coprod T^n(M)$ as the ring multiplication on $T(M)$ satisfies $T^m(M) \times T^n(M) \to T^{m+n}(M)$. We want to construct graded R-algebras from $T(M)$. The first is the *symmetric algebra*.

Construction 101.5. Let R be a commutative ring and M an R-module. For each positive integer n, let $\mathfrak{A}_n$ be the submodule of $T^n(M)$ generated by all elements $m_1 \otimes \cdots \otimes m_n - m_{\sigma(1)} \otimes \cdots \otimes m_{\sigma(n)}$ with $m_1, \ldots, m_n$ in M and σ in S_n. Define $\mathfrak{A}_0 = 0$. Let

$$\begin{aligned} \mathfrak{A} &:= \coprod_{i \geq 0} \mathfrak{A}_n \\ S^n(M) &:= T^n(M)/\mathfrak{A}_n \\ S(M) &:= \coprod_{i \geq 0} S^n(M) = \coprod_{i \geq 0} T^n(M)/\mathfrak{A}_n. \end{aligned}$$

Check that $\mathfrak{A}$ is a ideal in $T(M)$ and $S(M) = T(M)/\mathfrak{A}$ is a graded R-algebra, i.e., the ring multiplication on the algebra $S(M)$ satisfies $S^m(M) \times S^n(M) \to S^{m+n}(M)$ for all $m, n \geq 0$. The R-algebra $S(M)$ is a commutative algebra called the *symmetric algebra* of M and we have ring homomorphisms $R \xrightarrow{\varphi_{T(M)}} T(M) \xrightarrow{-} S(M)$ with $^-$ the canonical epimorphism.

Example 101.6. Let V be an n-dimensional F-vector space. Then $S(V) \cong F[t_1, \ldots, t_n]$ (An analogous statement holds if V is an R-free module of rank n with R commutative):

Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be an F-basis for V and $^- : T^m(V) \to S^m(V)$ the canonical R-module epimorphism. Then

$$\{\overline{v_{i_1} \otimes \cdots \otimes v_{i_m}} \mid 1 \leq i_1 \leq \cdots \leq i_m \leq n\}$$

is an F-basis for $S^m(V)$. In particular,

$$\dim_F S^m(V) = \binom{m+n-1}{n-1} = \binom{m+n-1}{m},$$

the number of (*homogeneous*) monic monomials of total degree m in $t_1, \ldots, t_n$. (Why?) In particular, $\dim_F S(V) = \infty$.

In the construction above, the ideal $\mathfrak{A} := \coprod_{n\geq 0} \mathfrak{A}_n$ satisfies

$$\mathfrak{A} \cap T^n(M) = \mathfrak{A}_n \text{ and } \mathfrak{A}_m\mathfrak{A}_n \subset \mathfrak{A}_{mn} \text{ for all } m, n \geq 0.$$

We say $\mathfrak{A} = \coprod_{n\geq 0} \mathfrak{A}_n$ is a *graded* or *homogeneous ideal* in the graded ring $T(M)$.

We want universal properties for $S^n(M)$ and $S(M)$. Call a map $f : \underbrace{X \times \cdots \times X}_{n} \to Y$ of R-modules *symmetric* if $f(x_1, \ldots, x_n) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for all $\sigma \in S_n$.

Let R be a commutative ring and M an R-module. Then we have a canonical n-linear map $f : \underbrace{M \times \cdots \times M}_{n} \to S^n(M)$ given by the composition $\underbrace{M \times \cdots \times M}_{n} \to T^n(M) \to T^n(M)/\mathfrak{A}_n$. Then $S^r(M)$ satisfies the following universal property:

Let $g : M \times \cdots \times M \to N$ be an n-linear and symmetric map of R-modules, then there exists a unique R-homomorphism $\varphi : S^n(M) \to N$ such that

(101.7)
$$\begin{array}{ccc} M \times \cdots \times M & \xrightarrow{\ f\ } & S^n(M) \\ & \searrow^{g} & \downarrow{\scriptstyle \varphi} \\ & & N \end{array}$$

commutes and the commutative R-algebra $S(M)$ satisfies the following universal property:

If A is a commutative R-algebra and $\varphi : M \to A$ is an R-homomorphism, then there exists a unique R-algebra homomorphism

(101.8) $$\Phi : S(M) \to A \text{ satisfying } \Phi|_M = \varphi.$$

We next construct another quotient of the tensor algebra, this time a non-commutative algebra.

Construction 101.9. Let R be a commutative ring and M an R-module. Define a submodule $\mathfrak{B}_n$ of $T^n(M)$ to be the submodule generated by all $m_1 \otimes \cdots \otimes m_n$ with $m_1, \ldots, m_n$ in M and $m_i = m_j$ for some $i \neq j$. Set $\bigwedge^n(M) := T^n(M)/\mathfrak{B}_n$. The map $f : \underbrace{M \times \cdots \times M}_{n} \to \bigwedge^n(M)$ defined by the composition $\underbrace{M \times \cdots \times M}_{n} \to T^n(M) \to T^n(M)/\mathfrak{B}_n$ is n-linear and *alternating*, i.e., satisfies $f(x_1, \ldots, x_n) = 0$ if there exists an $i \neq j$ such that $x_i = x_j$. It satisfies the following universal property:

If N is an R-module and $g : M \times \cdots \times M \to N$ is n-linear and alternating, then there exists a unique R-homomorphism $\varphi : \bigwedge^n(M) \to N$ such that

$$\begin{array}{ccc} M \times \cdots \times M & \xrightarrow{f} & \bigwedge^n(M) \\ & \searrow^{g} & \downarrow{\scriptstyle \varphi} \\ & & N \end{array}$$

commutes. This follows as there exists a unique R-homomorphism $\psi : T^n(M) \to N$ such that

$$\begin{array}{ccc} M \times \cdots \times M & \xrightarrow{f} & T^n(M) \\ & \searrow^{g} & \downarrow{\scriptstyle \psi} \\ & & N \end{array}$$

commutes. Then

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathfrak{B}_n & \xrightarrow{f} & T^n(M) & \xrightarrow{-} & \bigwedge^n(M) & \longrightarrow & 0 \\ & & & \searrow^{0} & \downarrow{\scriptstyle \psi} & & & & \\ & & & & N & & & & \end{array}$$

is exact and commutes, so φ exists and is unique by the universal property of the cokernel.

We denote the image of $(m_1, \ldots, m_n)$ in $\bigwedge^n(M)$ by $m_1 \wedge \cdots \wedge m_n$.

If $\varphi : M \to N$ is an R-homomorphism, then

$$\underbrace{M \times \cdots \times M}_{n} \to \bigwedge^n(N)$$

defined by $m_1, \ldots, m_n \mapsto \varphi(m_1) \wedge \cdots \wedge \varphi(m_n)$ is n-linear and alternating, so induces a unique R-homomorphism $\wedge^n\varphi : \bigwedge^n(M) \to \bigwedge^n(N)$. Set $\bigwedge^0(M) := R$, $\bigwedge^1(M) = M$ and $\bigwedge(M) := \coprod_{n\geq 0} \bigwedge^n(M)$.

Check. $\mathfrak{B} := \coprod_{n\geq 0} \mathfrak{B}_n$ is a graded ideal in $T(M)$ by using

$$T^m(M) \times \mathfrak{B}_n \text{ and } \mathfrak{B}_m \times T^n(M) \text{ lie in } \mathfrak{B}_{m+n},$$

so induces an R-bilinear map $\bigwedge^n(M) \times \bigwedge^m(M) \to \bigwedge^{m+n}(M)$ and (as we have associativity) an R-homomorphism

$$\bigwedge^m(M) \otimes_R \bigwedge^n(M) \to \bigwedge^{m+n}(M).$$

This defines a graded R-algebra structure $\bigwedge(M)$ with multiplication $\bigwedge(M) \times \bigwedge(M) \to \bigwedge(M)$ defined by

$$(x_1 \wedge \cdots \wedge x_m) \cdot (y_1 \wedge \cdots \wedge y_n) := x_1 \wedge \cdots \wedge x_m \wedge y_1 \wedge \cdots \wedge y_n.$$

(Why does it suffice to define multiplication on generators?) This R-algebra is called the *exterior algebra* of M.

Remark 101.10. Let R be a commutative ring and M an R-module.

1. $\bigwedge(M)$ can be defined to be the R-algebra generated by M with defining relations
$$m \wedge m = 0 \text{ for all } m \in M :$$
If m, n are elements of M, then
$$mn + nm = (n+m)^2 - n^2 - m^2 \text{ in } T(M),$$
so
$$m \wedge n = -n \wedge m \text{ in } \bigwedge(M)$$
and it follows that the graded ideal $\mathfrak{B}$ is generated in the algebra $T(M)$ by m^2, $m \in M$.
2. Using (1), one checks $\bigwedge(M)$ satisfies the following universal property: If A is an R-algebra and $\varphi : M \to A$ an R-homomorphism satisfying $\varphi(m)^2 = 0$ for all $m \in M$, then there exists a unique R-algebra homomorphism $\Phi : \bigwedge(M) \to A$ satisfying $\Phi|_M = \varphi$. In particular, any R-homomorphism $\varphi : M \to N$ extends to a unique graded R-algebra homomorphism $\wedge\varphi : \bigwedge(M) \to \bigwedge(N)$, i.e., $\wedge\varphi\big(\bigwedge^n(M)\big) \subset \bigwedge^n(N)$. For example, $\varphi(1_M) = 1_{\bigwedge(M)}$.
3. If $\varphi : M \to N$ and $\psi : N \to P$ are R-homomorphisms, then $\wedge(\psi\varphi) = \wedge\psi \circ \wedge\varphi$.

Exercise 101.11. 1. Let V be an n-dimensional F-vector space. Show
$$\dim_F S^m(V) = \binom{m+n-1}{n-1} = \binom{m+n-1}{m}.$$

2. Let R be a commutative ring. Define the free commutative R-algebra on a set S and show it is a polynomial ring on $|S|$ variables.
3. Let R be a commutative ring. Define the free R-algebra on a set S and show it exists.
4. Let R be a commutative ring and A and B commutative R-algebras. Show that $A \otimes_R B$ is the coproduct of A and B as rings (definition?). [If A and B are not commutative, this is not true.]
5. Let K_i/F be (arbitrary) field extensions for $i = 1, 2$. A *composite* of K_1, K_2 over F is a field L such that there exist (field) homomorphisms $\varphi_i : K_i \to L$ fixing F, i.e., F-homomorphisms, such that L is generated by $\varphi_1(K_1)\varphi_2(K_2)$. Show that there is a bijection between $\operatorname{Spec}(K_1 \otimes_F K_2)$ and F-isomorphism classes of composites of K_1 and K_2 over F.

6. If R is a commutative ring and M an R-module, prove that the tensor algebra $T(M)$ satisfies the universal property in Construction 101.4.
7. Fill in the details needed to justify the construction of the symmetric algebra in Construction 101.5.
8. Prove $S^r(M)$ satisfies the universal property (101.7).
9. Prove $S(M)$ satisfies the universal property (101.8).
10. Prove $\bigwedge(M)$ satisfies the universal property in Remark 101.10(2).

102. The Determinant

In this section, we show that the determinant exists for matrices over a commutative ring using the exterior algebra of a finitely generated free module M. We construct the determinant of an R-endomorphism of M that gives rise to the determinant of its matrix representations, so is more intrinsic than the more elementary matrix proof. Recall over a commutative ring all bases for a finitely generated free have the same rank. Unlike the tensor algebra over a finitely free module that has infinite rank, the exterior algebra has finite rank. More precisely, we have the following:

Theorem 102.1. *Let R be a commutative ring and M a free R-module of rank n on basis $\{x_1, \dots, x_n\}$. Then $\bigwedge(M)$ is a free R-module of rank 2^n. More precisely,*

$$\mathcal{B}_r := \{x_{i_1} \wedge \cdots \wedge x_{i_r} \mid 1 \le i_1 < \cdots < i_r \le n\}$$

is a basis for $\bigwedge^r(M)$ for $r \le n$. In particular,

$$\operatorname{rank} \bigwedge^r(M) = \begin{cases} \binom{n}{r} & \text{if } r \le n \\ 0 & \text{otherwise.} \end{cases}$$

To prove this we need a preliminary idea and its development in the case of exterior algebras.

Construction 102.2. Let R be a commutative ring, M an R-module, and $A = \coprod_{i \ge 0} A_i$ a graded R-algebra.

For each $i \ge 0$, define an R-homomorphism $\sigma : A_i \to A_i$ by $z \mapsto (-1)^i z$. Then σ induces a graded R-algebra homomorphism $\sigma : A \to A$, since $(-1)^i(-1)^j = (-1)^{i+j}$. Note that $\sigma^2 = 1_A$.

An R-homomorphism $\delta : A \to A$ is called an *anti-derivation* if

$$\delta(ab) = \delta(a)b + \sigma(a)\delta(b) \text{ for all } a, b \in A.$$

If A is $T(M)$, $S(M)$, or $\bigwedge(M)$, then M generates A as an R-algebra, hence δ is determined by what it does to $A_1 = M$.

Case. $A = T(M)$:
Let $\widetilde{\varphi} : M \to T(M)$ be an R-homomorphism. Then $M \times M \to T(M)$ defined by

$$(*) \qquad (a, b) \mapsto \widetilde{\varphi}(a)b - a\widetilde{\varphi}(b) = \widetilde{\varphi}(a)b + \sigma(a)\widetilde{\varphi}(b)$$

is R-bilinear, so induces an R-homomorphism $T^2(M) \to T(M)$. Inductively, we obtain R-homomorphisms $T^n(M) \to T(M)$, i.e., $\widetilde{\varphi}$ induces an anti-derivation $T(M) \to T(M)$. We also call this map $\widetilde{\varphi}$.

Case. $A = \bigwedge(M)$:
Let φ be the composition

$$M \xrightarrow{\widetilde{\varphi}} T(M) \xrightarrow{^-} \bigwedge(M),$$

with $\widetilde{\varphi}$ the anti-derivation of the previous case.

Claim. If φ satisfies

$$\varphi(x) \wedge x - x \wedge \varphi(x) = 0 \text{ for all } x \in M, \tag{102.3}$$

then φ induces an anti-derivation $\bigwedge(M) \to \bigwedge(M)$ that we shall also call φ:
As $\widetilde{\varphi} : T(M) \to T(M)$ is an anti-derivation, to show that φ induces an anti-derivation $\bigwedge(M) \to \bigwedge(M)$, it suffices to show that

$$\widetilde{\varphi}(\mathfrak{B}) \subset \mathfrak{B} = \ker\big({}^- : T(M) \to \bigwedge(M)\big).$$

Let $x \in M$, then by (*), we have

$$(\dagger) \qquad \widetilde{\varphi}(x^2) = \widetilde{\varphi}(x \cdot x) = \widetilde{\varphi}(x)x + \sigma(x)\widetilde{\varphi}(x) = \widetilde{\varphi}(x)x - x\widetilde{\varphi}(x)$$

that lies in $\mathfrak{B}$. Let a and b lie in $T(M)$ and x in M, then

$$\begin{aligned}\widetilde{\varphi}(ax^2b) &= \widetilde{\varphi}(a)x^2b + \sigma(a)\widetilde{\varphi}(x^2b)\\ &= \widetilde{\varphi}(a)x^2b + \sigma(a)\widetilde{\varphi}(x^2)b + \sigma(a)\sigma(x^2)\widetilde{\varphi}(b)\\ &= \widetilde{\varphi}(a)x^2b + \sigma(a)\widetilde{\varphi}(x^2)b + \sigma(a)\sigma(x)^2\widetilde{\varphi}(b).\end{aligned}$$

Since $x^2, \sigma(x)^2$ lie in $\mathfrak{B}$ and $\widetilde{\varphi}(x^2)$ lies in $\mathfrak{B}$ by (†) with $\mathfrak{B}$ a 2-sided ideal in $T(M)$, we conclude that $\widetilde{\varphi}(ax^2b)$ lies in $\mathfrak{B}$ for all a, b in $T(M)$ and for all x in M. Since $\mathfrak{B}$ is generated by x^2, $x \in M$, by Remark 101.10(1), we have

$$\mathfrak{B} = \langle ax^2b \mid a, b \in T(M), x \in M\rangle.$$

Therefore, $\widetilde{\varphi}(\mathfrak{B}) \subset \mathfrak{B}$, and the claim is established.

Examples 102.4. Let R be a commutative ring and M an R-module. Suppose that we are in the case $A = \bigwedge(M)$ above together with its notation.

1. If $\varphi(x)$ lies in M for all x in M, then equation (102.3) becomes
$$2\varphi(x) \wedge x = 0 \text{ for all } x \in M.$$
2. If $\varphi(x)$ lies in R for all x in M, then equation (102.3) always holds. In particular, as $R = \bigwedge^0(M) \subset \bigwedge(M)$, if φ lies in $M^* = \mathrm{Hom}_R(M, R)$, then φ induces an anti-derivation $\bigwedge(M) \to \bigwedge(M)$ that we also call φ.
3. Suppose that M is R-free on basis $\mathcal{B} = \{x_1, \dots, x_n\}$ and $\mathcal{B}^* = \{f_1, \dots, f_n\}$, the dual basis for $M^* = \mathrm{Hom}_R(M, R)$. Then each f_i defines an anti-derivation $f_i : \bigwedge(M) \to \bigwedge(M)$.

We now proceed to the proof of Theorem 102.1.

PROOF. We use the notation of Example 102.4(3). As
$$x_{\sigma(i_1)} \wedge \cdots \wedge x_{\sigma(x_{i_r})} = \pm x_{i_1} \wedge \cdots \wedge x_{i_r}, \text{ for all } \sigma \in S_r,$$
we have
$$\mathcal{B}_r = \{x_{i_1} \wedge \cdots \wedge x_{i_r} \mid 1 \le i_1 < \cdots < i_r \le n\}$$
generates $\bigwedge^r(M)$. As $\bigwedge(M) = \coprod_{i \ge 0} \bigwedge^r(M)$ is graded, it suffices to show that $\mathcal{B}_r$ is R-linearly independent. Suppose this is false. Since $\mathcal{B} = \mathcal{B}_1$ is linearly independent, there exists a mimimal r such that $\mathcal{B}_r$ is linearly dependent. Let
$$(*) \qquad \sum_{1 \le i_1 < \cdots < i_r \le n} a_{i_1, \dots, i_r} x_{i_1} \wedge \cdots \wedge x_{i_r} = 0 \text{ in } \bigwedge^r(M),$$
not all $a_{i_1,\dots,i_r}$ zero. Changing notation, we may assume that $a_{1,j_2,\dots j_r}$ is nonzero with $1 < j_2 < \cdots < j_r \le n$. Applying the anti-derivation arising from f_1 in $\mathcal{B}^*$ to (*), we get
$$\sum_{1 \le i_2 < \cdots < i_r \le n} a_{1,i_2,\dots,i_r} x_{i_2} \wedge \cdots \wedge x_{i_r} = 0,$$
since $f_1(x_i) = \delta_{i1}$. This implies that $\mathcal{B}_{r-1}$ is linearly dependent, a contradiction. □

Corollary 102.5. *The sign map* $\mathrm{sgn} : S_n \to \{\pm 1\}$ *given by*
$$\mathrm{sgn}\,\sigma = \begin{cases} 1 & \text{if } \sigma \text{ is a product of an even number of transpositions} \\ -1 & \text{otherwise.} \end{cases}$$
is a well-defined group homomorphism.

PROOF. Let R be a commutative ring and M a free R-module on basis $\{x_1, \dots, x_n\}$, then $\bigwedge^n(M)$ is R-free on basis $\{x_1 \wedge \cdots \wedge x_n\}$. Hence if $\sigma \in S_n$. there exists a unique $\mathrm{sgn}(\sigma)$ in $\{\pm 1\}$ satisfying
$$(102.6) \qquad x_{\sigma(1)} \wedge \cdots \wedge x_{\sigma(n)} = \mathrm{sgn}(\sigma) x_1 \wedge \cdots \wedge x_n.$$

It is easily checked that sgn is a group homomorphism. □

Construction 102.7. Let R be a commutative ring and M an R-free module on basis $\{x_1, \ldots, x_n\}$. Let $f : M \to M$ be an R-homomorphism. Write

$$f(x_i) = \sum_{j=1}^{n} \alpha_{ji} x_j \text{ for } i = 1, \ldots, n$$

and $A = (\alpha_{ij}) = [f]_{\mathcal{B}}$, the matrix representation of f relative to the (ordered) basis $\mathcal{B}$. Then f induces a unique graded R-algebra homomorphism $\wedge f : \bigwedge(M) \to \bigwedge(M)$. Since $\{x_1 \wedge \cdots \wedge x_n\}$ is a basis for $\bigwedge^n(M)$, there exists a unique element λ in R satisfying $\wedge^n f = \lambda 1_{\bigwedge^n(M)}$, where $\wedge f|_{\bigwedge^n(M)} = \wedge^n f$. In particular, λ is an eigenvalue of $\wedge^n f$ on the rank one free R-module $\bigwedge^n(M)$. Using equation (102.6) and the alternating property, we see that

$$\begin{aligned}
\lambda 1_{\bigwedge^n(M)} &= \wedge^n f(x_1 \wedge \cdots \wedge x_n) \\
&= f(x_1) \wedge \cdots \wedge f(x_n) = \sum_{j=1}^{n} \alpha_{j1} x_j \wedge \cdots \wedge \sum_{j=1}^{n} \alpha_{jn} x_j \\
&= \sum_{S_n} \operatorname{sgn}(\sigma) \alpha_{\sigma(1)1} \cdots \alpha_{\sigma(n)n} x_1 \wedge \cdots \wedge x_n.
\end{aligned}$$

The element

$$\lambda = \sum_{S_n} \operatorname{sgn}(\sigma) \alpha_{\sigma(1)1} \cdots \alpha_{\sigma(n)n}$$

is called the *determinant* of f and denoted $\det f$ (or $\det A$). As it is an eigenvalue for $\wedge^n f$ on a rank one free R-module, it is unique and independent of $\mathcal{B}$.

Next let $g : M \to M$ be another R-homomorphism and $B = [g]_{\mathcal{B}}$. Then

$$\wedge^n(fg) = \wedge^n f \circ \wedge^n g \text{ so } \det fg = \deg f \det g, \text{ } i.e., \det AB = \det A \det B.$$

More generally, f, g induce $\wedge^r f, \wedge^r g : \bigwedge^r(M) \to \bigwedge^r(M)$ for each $r = 1, \ldots, n$. Set $A^{(r)} = [\wedge^r f]_{\mathcal{B}_r}$ and $B^{(r)} = [\wedge^r g]_{\mathcal{B}_r}$ in $\mathbb{M}_{\binom{n}{r}}(R)$. Then

$$(AB)^{(r)} = A^{(r)} B^{(r)}$$

which gives identities of degree r in the entries of A and B called the *Binet-Cauchy Equations*. Note that $A^{(r)}$ is an $(\binom{n}{r} \times \binom{n}{r})$-matrix whose entries are the rth order minors of A, etc.

To get additional formulae, let $A = (a_{ij})$ be an $n \times n$-matrix over a commutative ring R. Then we view A as the matrix representation of the appropriate $f : R^n \to R^n$ in the standard basis $\{e_1, \ldots, e_n\}$. Set

$S = \{1, \dots, n\}$. If H is a subset of S consisting of r elements, let $\underline{H}$ be the ordered r-tuple $(i_1, \dots, i_r)$ where $1 \le i_1 < \cdots < i_r \le n$ with all $i_j \in H$. and $e_{\underline{H}} = e_{i_1} \wedge \cdots \wedge e_{i_r}$. If K is another subset of S, then

$$e_{\underline{H}} \wedge e_{\underline{K}} = \begin{cases} \varepsilon_{\underline{HK}} e_{\underline{H \cup K}} & \text{if if } H \cap K = \emptyset \\ 0 & \text{if } H \cap K \neq \emptyset, \end{cases} \tag{102.8}$$

where $\varepsilon_{\underline{HK}} = (-1)^m$, if $H \cap K = \emptyset$ and m is the number of transpositions modulo two to get $\underline{H \cup K}$ from $\underline{H} \cup \underline{K}$. If H is a subset of S, we denote by H' the complement of H in S. Fix r, $1 \le r \le n$, and let $S(r) = \{I \mid I \subset S \text{ with } |I| = r\}$. Let $A_{\underline{IH}}$ denote the submatrix of A with rows of A associated to $\underline{I}$, $I \in S(r)$, and columns of A associated to $\underline{H}$ i.e, $\det A_{\underline{IH}}$ is the r-order minor associated to $\underline{I'}$ and $\underline{H'}$.

Fix $H \subset S(r)$. Then we have

$$\wedge^r(f)(e_{\underline{H}}) = \sum_{I \in S(r)} \det(A_{\underline{IH}})\, e_{\underline{I}}.$$

Therefore, for such a fixed H, we have

$$\begin{aligned} (\det A \cdot \varepsilon_{\underline{HH'}})\, e_{\underline{S}} &= \big(\wedge^r(f)\big)(e_{\underline{H}}) \wedge \big(\wedge^{n-r}(f)\big)(e_{\underline{H'}}) \\ &= \sum_{I \in S(r)} \sum_{J \in S(n-r)} \det(A_{\underline{IH}}) \det(A_{\underline{JH'}})\, e_{\underline{I}} \wedge e_{\underline{J}} \\ &= \sum_{I \in S(r)} \sum_{J \in S(n-r)} \det(A_{\underline{IH}}) \det(A_{\underline{JH'}})\, \varepsilon_{\underline{IJ}} e_{\underline{S}} \\ &= \sum_{I \in S(r)} \varepsilon_{\underline{II'}} \det(A_{\underline{IH}}) \det(A_{\underline{I'H'}})\, e_{\underline{S}} \end{aligned}$$

by (102.8). In particular,

$$\det A = \sum_{I \in S(r)} \varepsilon_{\underline{HH'}} \varepsilon_{\underline{II'}} \det(A_{\underline{IH}}) \det(A_{\underline{I'H'}}).$$

This is called the *Laplace expansion* of $\det A$. For $r = 1$, this is the well-known expansion by minors along rows. By symmetry, we have an analogous result on columns by using right modules. If we let $A(i, j)$ be the $(n-1)$-order minor of A, i.e., the determinant of the submatrix of A by deleting the ith row and jth column, then $(-1)^{i+j} A(i, j)$ is called the (i, j)*th cofactor* of A. Define the *classical adjoint* $\operatorname{adj}(A)$ of A to be the transpose of the matrix of cofactors of A, i.e., the (i, j) entry of $\operatorname{adj}(A)$ is $(-1)^{i+j} A(j, i)$. By Laplace expansion, we check that

$$\sum_{i=1}^{n} a_{i1}(-1)^{i+1} A(j, 1) + \cdots + a_{in}(-1)^{i+1} A(j, n)$$

is the expansion of a determinant whose jth row equals the ith row of A while the other rows are as in A. If $i \neq j$ this means this matrix has two rows the same so is zero, hence is zero off the main diagonal. It follows that

$$A \cdot \mathrm{adj}(A) = \det A \cdot I$$

with I the $n \times n$ identity matrix. Similarly, $\mathrm{adj}(A) \cdot A = \det A \cdot I$.

Let $M = \sum_{i=1}^{n} Rx_i$ be an R-module. If we have a system of equations $\sum_{j=1}^{n} a_{ji}x_j = 0$ for $i = 1, \ldots, n$, then we must have $\det(A)M = 0$. In particular, $\det A$ is in the annihilator of M. It follows that if $\mathrm{ann}_R(M) = 0$ and $M \neq 0$, then $\det A = 0$. Next suppose that A is invertible, i.e., $\det A$ is a unit (respectively, R is a domain and $M = R$) and we want to solve a system of equations $\sum_{j=1}^{n} a_{ji}x_j = b_i$ for $i = 1, \ldots, n$ in M (respectively the quotient field of R), i.e., if $x = (x_1 \cdots x_n)^t$ and $b = (b_1 \cdots b_n)^t$ column matrices, then we want to solve the matrix equation $Ax = b$. Then the solution is $x = (\det A)^{-1}\mathrm{adj}(A)b$. This is called *Cramer's Rule*.

Exercise 102.9. 1. Let V be an n-dimensional F-vector space. If $y_1, \ldots, y_n$ are elements in V, show $\{y_1, \ldots, y_n\}$ is linearly dependent if and only if $y_1 \wedge \cdots \wedge y_n = 0$.

2. Let R be a commutative ring and

$$0 \to M' \to M \to M'' \to 0$$

an exact sequence of free R-modules of ranks r, n, s, respectively. Show that there exists a natural isomorphism

$$\varphi : \bigwedge^{r}(M') \otimes_R \bigwedge^{s}(M'') \to \bigwedge^{n}(M).$$

3. Let $M = M \oplus M''$ be a direct sum of free R-modules of finite rank. Then for all positive integers n, show that

$$\bigwedge^{n}(M) \cong \coprod_{r+s=n} \bigwedge^{r}(M') \otimes_R \bigwedge^{s}(M'').$$

Check that Cramer's Rule in the text is the usual one.

4. Let R be a commutative ring and A an R-algebra. A *derivation* $\delta : A \to A$ is an R-homomorphism satisfying $\delta(ab) = \delta(a)b + a\delta(b)$ for all $a, b \in A$. (E.g., the derivative on rings of real differentiable functions.) Show that if M is an R-module, then any R-homomorphism $\varphi : M \to S(M)$ extends to a derivation $S(M) \to S(M)$.

Appendices

103. Matrix Representations

We review the notion of **matrix representation** of a linear transformation relative to a pair of ordered bases and what happens when we change bases. We do this in somewhat more generality then in linear algebra. All proofs, however, are the same, hence omitted. Throughout R will be a commutative ring. [So if R is a field, and R-module is just a vector space over R and an R-homomorphism is just a linear transformation.]

Setup.

Let R be a commutative ring.
V and W are free R-modules of rank n and m respectively.

$\mathcal{B} := \{v_1, ..., v_n\}$ is an ordered basis for V.
$\mathcal{B}'$ is a second ordered basis of V.

$\mathcal{C} := \{w_1, ..., w_m\}$ is an ordered basis for W.
$\mathcal{C}'$ is a second ordered basis of W.

$T : V \to W$ is an R-homomorphism.

Let v be an element in the free R-module V. Then there exist unique scalars $\alpha_1, ..., \alpha_n$ in R such that

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

The scalars $\alpha_1, ..., \alpha_n$ are called the *coordinates* of v relative to the (ordered) basis $\mathcal{B}$.

Let

$$[v]_{\mathcal{B}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

the *coordinate matrix* of v relative to the basis $\mathcal{B}$ and

$$V_{\mathcal{B}} := \{\, [v]_{\mathcal{B}} \mid v \in V \} = R^{n\times 1}.$$

We have an R-isomorphism

$$V \to V_{\mathcal{B}} \text{ given by } v \mapsto [v]_{\mathcal{B}}.$$

Similarly, if w is a vector in the free R-module W, then there exist unique scalars, $\beta_1, ..., \beta_m$, in R such that

$$w = \beta_1 w_1 + \cdots + \beta_m w_m.$$

Let

$$[w]_{\mathcal{C}} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$$

the *coordinate matrix* of w relative to the basis $\mathcal{C}$ and

$$W_{\mathcal{C}} := \{\, [w]_{\mathcal{C}} \mid w \in W\} = R^{m\times 1}.$$

Examples 103.1. 1.

$$[v_1]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \ldots \ , [v_n]_{\mathcal{B}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

2. Suppose that $n = 3$. Then

$$[2v_1 - 3v_3]_{\mathcal{B}} = \begin{pmatrix} 2 \\ 0 \\ -3 \end{pmatrix} \qquad [7v_2 - v_3]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 7 \\ -1 \end{pmatrix}.$$

We turn to the R-homomorphism $T : V \to W$. By the Universal Property of Free Modules, there exists a unique matrix

$$[T]_{\mathcal{B},\mathcal{C}} \in R^{m\times n}$$

called the *matrix representation* of T relative to the ordered bases $\mathcal{B}, \mathcal{C}$ that satisfies

$$[T]_{\mathcal{B},\mathcal{C}}[v]_{\mathcal{B}} = [T(v)]_{\mathcal{C}} \quad \forall v \in V.$$

If $V = W$ and $\mathcal{B} = \mathcal{C}$, we let $[T]_{\mathcal{B}} = [T]_{\mathcal{B},\mathcal{C}}$.

The matrix $[T]_{\mathcal{B},\mathcal{C}}[v]_{\mathcal{B}} = [T(v)]_{\mathcal{C}}$ is computed as follows. Write Tv_j in the $\mathcal{C}$ basis. Then the coordinate matrix of Tv_j relative to the $\mathcal{C}$ basis is the jth column of the matrix $[T]_{\mathcal{B},\mathcal{C}}$, i.e., if

$$Tv_j = \beta_{1j}w_1 + \cdots + \beta_{mj}w_m \text{ then } [T(v_j)]_{\mathcal{C}} = \begin{pmatrix} \beta_{1j} \\ \vdots \\ \beta_{mj} \end{pmatrix}$$

and this is the jth column of $[T]_{\mathcal{B},\mathcal{C}}$.

Definition 103.2. A matrix of the form $[Id]_{\mathcal{B},\mathcal{B}'}$ is called a *change of basis matrix.*

It arises by writing the elements in the $\mathcal{B}$ basis in terms of the elements in the $\mathcal{B}'$ basis.

The main results are:

Proposition 103.3. *Let V be an R-free module of rank n. Let $\mathcal{B}$ and $\mathcal{B}'$ be ordered bases for V. Then $[Id]_{\mathcal{B},\mathcal{B}'}$ is an invertible matrix and*

$$[Id]^{-1}_{\mathcal{B},\mathcal{B}'} = [Id]_{\mathcal{B}',\mathcal{B}}$$

Remarks 103.4. It can be shown that

$$\mathrm{GL}_n R = \{[T]_{\mathcal{B},\mathcal{C}} \,|\, T \text{ an isomorphism, } \mathcal{B},\mathcal{C} \text{ bases for } V\}.$$

Since R is commutative, determinants exist. It can also be shown that

$$\mathrm{GL}_n R = \{A \in \mathbf{M}_n R \,|\, \det(A) \in R^\times\}$$

Theorem 103.5. *Let V, W, and X be finitely generated free R-modules. Let $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ be ordered bases for V, W, and X, respectively. Let $T : V \to W$ and $S : W \to X$ be R-homomorphisms. Then*

$$[S \circ T]_{\mathcal{B},\mathcal{D}} = [S]_{\mathcal{C},\mathcal{D}}[T]_{\mathcal{B},\mathcal{C}}$$

where the right hand side is matrix multiplication.

Theorem 103.6. *Let V be a finitely generated free R-module and $\mathcal{B}$ an ordered basis for V. Then*

$$\mathrm{End}_R V \to \mathbb{M}_n R \text{ via } T \mapsto [T]_{\mathcal{B}}$$

is a ring isomorphism and induces a group isomorphism

$$\mathrm{Aut}_R V \to \mathrm{GL}_n R.$$

Theorem 103.7. (Change of Basis Theorem.) *Let V and W be finitely generated R-free modules. Let $\mathcal{B}$ and $\mathcal{B}'$ be ordered bases for V and $\mathcal{C}$ and $\mathcal{C}'$ be ordered basis for W. Let $T : V \to W$ be an R-homomorphism. Then*

$$[T]_{\mathcal{B}',\mathcal{C}'} = [Id]_{\mathcal{C},\mathcal{C}'}[T]_{\mathcal{B},\mathcal{C}}[Id]_{\mathcal{B}',\mathcal{B}} = [Id]_{\mathcal{C},\mathcal{C}'}[T]_{\mathcal{B},\mathcal{C}}[Id]^{-1}_{\mathcal{B},\mathcal{B}'}.$$

The Change of Basis Theorem states that the following diagram commutes

$$\begin{array}{ccc} V_{\mathcal{B}} & \xrightarrow{[T]_{\mathcal{B},\mathcal{C}}} & W_{\mathcal{C}} \\ {\scriptstyle [Id]_{\mathcal{B},\mathcal{B}'}}\downarrow & & \downarrow{\scriptstyle [Id]_{\mathcal{C},\mathcal{C}'}} \\ V_{\mathcal{B}'} & \xrightarrow{[T]_{\mathcal{B}',\mathcal{C}'}} & W_{\mathcal{C}'} \end{array}$$

Note that the inverses of the change of bases matrices go in the reverse direction. One can fill in more of the diagram. For example, the maps from the diagonals can also be read off, e.g.,

$$[T]_{\mathcal{B},\mathcal{C}'} = [T]_{\mathcal{B}',\mathcal{C}'}[Id]_{\mathcal{B},\mathcal{B}'}$$

$$[T]_{\mathcal{B}',\mathcal{C}} = [Id]^{-1}_{\mathcal{C},\mathcal{C}'}[T]_{\mathcal{B}',\mathcal{C}'} = [Id]_{\mathcal{C}',\mathcal{C}}[T]_{\mathcal{B}',\mathcal{C}'}$$

Warning 103.8. Usually T is not an isomorphism, so $[T]_{\mathcal{B},\mathcal{C}'}$ is not invertible. So you cannot reverse arrows having T in them. If T is an isomorphism, then the matrix representation of T^{-1} is the inverse of the corresponding matrix representation of T.

Historically, one defined equivalence relationships between matrices. We define these and tell how they are related to R-homomorphisms of finitely generated free R-modules.

Definition 103.9. If $A, B \in \mathbb{M}_n R$, we say that they are *similar* and write $A \sim B$ if there is an invertible matrix $C \in \mathrm{GL}_n R$ such that $A = C^{-1}BC$.

Clearly, $\sim$ is an equivalence relation. An important problem is to find good representatives for the classes under this equivalence relation when R is a nice ring, e.g., a PID. This leads to the study of Rational Canonical Forms and Jordan Canonical Forms in linear algebra.

Theorem 103.10. *Let $A, B \in \mathbb{M}_n R$. Then A is similar to B in $\mathbb{M}_n R$ if and only if there exist a free R-module V of rank n with bases $\mathcal{B}, \mathcal{B}'$ and $T \in \mathrm{End}_R V$ such that $A = [T]_{\mathcal{B}}$ and $B = [T]_{\mathcal{B}'}$.*

Definition 103.11. If $A, B \in R^{m\times n}$, we say that they are *equivalent* write $A \simeq B$ if there is an invertible matrices $P \in \mathrm{GL}_m R$ and $Q \in \mathrm{GL}_n R$ such that $A = PBQ$.

Clearly, $\simeq$ is an equivalence relation. An important problem is to find good representatives for the classes under this equivalence relation when R is a nice ring, e.g., a PID. This leads to the study of Smith Normal Forms in linear algebra.

Theorem 103.12. *Let $A, B \in R^{m\times n}$. Then $A \simeq B$ in $R^{m\times n}$ if and only if there exist free R-modules V, W of rank n, m, respectively, with bases $\mathcal{B}, \mathcal{B}'$ for V and bases $\mathcal{C}, \mathcal{C}'$ for W and an R-homomorphism $T : V \to W$ such that $A = [T]_{\mathcal{B},\mathcal{C}}$ and $B = [T]_{\mathcal{B}',\mathcal{C}'}$.*

104. Smith Normal Form over a Euclidean Ring

Let R be a domain and $A \in R^{m\times n}$. Recall that A is in *Smith Normal Form* if A is a matrix of the form

$$\begin{pmatrix} q_1 & 0 & \dots & & & & \\ 0 & q_2 & & & & & \\ \vdots & & \ddots & & & & \\ & & & q_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & \end{pmatrix}$$

with $q_1 \mid q_2 \mid q_3 \mid \cdots \mid q_r$ in R and $q_r \neq 0$, i.e., the diagonal entries of A are $q_1, \dots, q_r, 0, \dots, 0$ with $q_1 \mid q_2 \mid q_3 \mid \cdots \mid q_r$ in R and $q_r \neq 0$ and all entries off the diagonal are 0.

Let R be a domain and $A \in \mathbb{M}_n(R)$. We say that $A = (a_{ij})$ is an *elementary matrix* if

(i) *Type I:* if there exists $\lambda \in R$ and $l \neq k$ such that

$$a_{ij} = \begin{cases} 1 & \text{if } i = j \\ \lambda & \text{if } (i,j) = (k,l) \\ 0 & \text{otherwise} \end{cases}$$

(ii) *Type II:* if there exists $k \neq l$ such that

$$a_{ij} = \begin{cases} 1 & \text{if } i = j \neq l \text{ and } i = j \neq k \\ 0 & \text{if } i = j = l \text{ or } i = j = k \\ 1 & \text{if } (k,l) = (i,j) \text{ or } (k,l) = (j,i) \\ 0 & \text{otherwise} \end{cases}$$

(iii) *Type III:* if there exists a unit u in R and l such that

$$a_{ij} = \begin{cases} 1 & \text{if } i = j \neq l \\ u & \text{if } i = j = l \\ 0 & \text{otherwise} \end{cases}$$

Remarks 104.1. 1. Let $A \in R^{m\times n}$. Multiplying A on the left (respectively right) by a suitable size

(i) Type I is adding a multiple of a row (respectively column) of A to another row (respectively column) of A.

(*ii*) Type II is interchanging two rows (respectively columns) of A.

(*iii*) Type III is multiplying a row (respectively column) of A by a unit.

2. All elementary matrices are invertible.

Recall if R is a ring, two $m \times n$ matrices A and B in $R^{m\times n}$ are called *equivalent* if there exist invertible matrices $P \in \mathrm{GL}_m(R)$ and $Q \in \mathrm{GL}_n(R)$ such that $A = PAQ$.

Theorem 104.2. *Let R be a euclidean ring, $A = (a_{ij}) \in R^{m\times n}$. Then A is equivalent to a matrix in Smith Normal Form. Moreover, there exist matrices $P \in GL_m(R)$ and $Q \in \mathrm{GL}_n(R)$, each a product of matrices of Type I, Type II, and Type III, such that PAQ is in Smith Normal Form.*

Proof. The proof will, in fact, be an algorithm to find a Smith Normal Form for A. Let δ be the euclidean function on R. If $A = 0$ there is nothing to do, so assume that $A \neq 0$.

Step 1. Choose $a = a_{ij} \neq 0$ such that $\delta(a)$ is minimal among all the $\delta(a_{lk})$, $a_{lk} \neq 0$. Put a in the $(1, 1)$ spot using matrices of Type II. In particular, we may assume that $a = a_{11}$.
[If a is a unit in R, use a Type III matrix to make $a = 1$.]

Step 2. If $a \mid a_{ij}$ for all i and j, use Type I matrices to transform A into a matrix of the form

$$\begin{pmatrix} a & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix}$$

Note that a divides every entry of A_1 [Check]. If a non-zero entry of A_1 has smaller δ value, $\delta(a)$ is not minimal so go back to Step 1.
[As δa is a non-negative integer this cannot happen infinitely often.]
If δa is still minimal, take A_1 and go back to Step 1.
[Note. If this occurs, by induction there exist matrices Q_1, P_1 such that $P_1A_1Q_1$ is in Smith Normal Form. Let

$$P = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & P_1 & \\ 0 & & & \end{pmatrix} \text{ and } Q = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & Q_1 & \\ 0 & & & \end{pmatrix}$$

then

$$P \begin{pmatrix} a & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix} Q$$

is in Smith Normal Form.]

Step 3. Step 2 does not apply and there exists an entry $b = a_{ij}$ in either the first row or first column such that $a \nmid b$:

Write $b = qa + r$ in R with $r \neq 0$ and $\delta(r) < \delta(a)$. Use Type I matrices to change A into a matrix with r in it. Since $\delta(a)$ is not longer minimal, go back to Step 1.

[Since $\delta(a)$ is a non-negative integer and $\delta(r) < \delta(a)$, this must eventually stop.]

Step 4. Neither Step 2 nor Step 3 apply. Thus $a \mid a_{ij}$ whenever $i = 1$ or $j = 1$:

Use Type I matrices to convert A to

$$\begin{pmatrix} a & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix}$$

Now one of the following occurs.

(1) There exists a non-zero entry b in A_1 such that $\delta(b) < \delta(a)$. So $\delta(a)$ is no longer minimal. Go back to Step 1.

(2) $a \mid b$ for all entries b in A_1. This is impossible — You should have been in Step 2.
[No matter, take A_1 and go to Step 1.]

(3) There exists an entry b in A_1 such that $a \nmid b$:
Write $b = qa + r$ in R with $r \neq 0$ and $\delta(r) < \delta(a)$. Use Type I matrices to get b into the first column. (This does not change the $(1, 1)$ entry a.) Now use Type I matrices to change b to r. Since $\delta(a)$ is no longer minimal, go back to Step 1.

Clearly this algorithm yields a Smith Normal Form of A.

□

Remarks 104.3. 1. If R is a ring, two $m \times n$ matrices A and B in $R^{m\times n}$ are called *equivalent* if there exist invertible matrices $P \in \mathrm{GL}_m(R)$ and $Q \in \mathrm{GL}_n(R)$ such that $A = PAQ$. Compare this to change of basis in linear algebra of matrix representations of linear transformations. So the theorem says if R is a euclidean ring than

any $m \times n$ matrix over R is equivalent to a matrix in Smith Normal Form.

2. Note that we do not really need Type III matrices in the above. We only used them to transform those diagonal entries q_i that were units into one.

3. We did not really need Step 2 as it is incorporated into Step 4, but it is useful to isolate it to complete the induction step of the proof.

Exercise 104.4. Let R be a commutative ring. Let $E_n(R)$ be the subgroup of $\mathrm{GL}_n(R)$ generated by all matrices of the form $Id + \lambda$ where λ is a matrix with precisely one non zero entry and this entry does not occur on the diagonal. Suppose that R is a euclidean ring. Show that $\mathrm{SL}_n(R) = E_n(R)$.

105. Primitive Roots

Let G be a finite abelian group. In this section, we compute the automorphism group of G. This is useful in the study of cyclotomic extensions of a field. We know that $G \cong \mathbb{Z}/n\mathbb{Z}$ for some positive integer n and $\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z}) \cong (\mathbb{Z}/n\mathbb{Z})^\times$ (as generators must go to generators), so it is abelian of order $\varphi(n)$. By the Chinese Remainder Theorem, we are reduced to the case of of computing $(\mathbb{Z}/p^n\mathbb{Z})^\times$ where p is a (positive) prime. We also know that $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic, so can assume that $n > 1$. We shall show that $(\mathbb{Z}/p^n\mathbb{Z})^\times$ is also cyclic unless $p = 2$. A generator for this automorphism group is called a *primitive root*. The proofs rely on the binomial theorem and its special case, the *Children's Binomial Theorem*: that $(a+b)^p \equiv a^p + b^p \mod p$ for any integers a and b. In particular, it implies

Lemma 105.1. *Let p be a (positive) prime and $n > 1$. If a, b, and k are integers, then $b^p \equiv b^p + kp^n \mod p^{n+1}$. In particular, if $a \equiv b \mod p^n$, then $a^p \equiv b^p \mod p^{n+1}$.*

Lemma 105.2. *Let p be a (positive) odd prime and $n \geq 2$. Then for every integer a, we have $(1+ap)^{p^{n-2}} \equiv 1 + ap^{n-1} \mod p^n$.*

Proof. We induct on n. The case $n = 2$ is trivial, so we may assume that $n > 2$. In particular, $2(n-1) > n+1$, so $p(n-1) > n+1$. By induction, we may assume the result for $n-2$ and show the result for $n-1$. By the lemma, induction, and the Binomial Theorem, we

have

$$\begin{aligned}(1+ap)^{p^{n-1}} &\equiv \left((1+ap)^{p^{n-2}}\right)^p \equiv (1+ap^{n-1})^p \\ &\equiv 1+\binom{p}{1}ap^{n-1}+\binom{p}{2}a^2p^{2(n-1)}+\cdots+(ap^{n-1})^p \\ &\equiv 1+\binom{p}{1}ap^{n-1} \equiv 1+ap^n \mod p^{n+1}.\end{aligned}$$

□

Lemma 105.3. *Let p be a (positive) odd prime and a an integer not divisible by p. Then for every integer $n \geq 2$, the congruence class of $1+ap$ in $\mathbb{Z}/p^n\mathbb{Z}$ has order p^{n-1}.*

Proof. Let $^- : \mathbb{Z} \to \mathbb{Z}/p^n\mathbb{Z}$ the canonical epimorphism. By the previous lemma, we know that $(1+ap)^{p^{n-1}} \equiv 1+ap^n \mod p^{n+1}$, hence $(1+ap)^{p^{n-1}} \equiv 1 \mod p^n$ and the order of $\overline{(1+ap)}$ divides p^{n-1}. The previous lemma also implies that $(1+ap)^{p^{n-1}} \equiv 1+ap^{n-1} \not\equiv 1 \mod p^n$, so the order of $\overline{(1+ap)}$ must be p^{n-1}. □

Proposition 105.4. *Let p be a (positive) odd prime, then $(\mathbb{Z}/p^n\mathbb{Z})^\times$ is cyclic for all n.*

Proof. Let a be an integer relatively prime to p. We know that $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic of order $p-1$, so generated by the residue class of a.

Claim 1. We may assume that $a^{p-1} \not\equiv 1 \mod p^2$:

Suppose that $a^{p-1} \equiv 1 \mod p^2$. Let $b = a+p$, then

$$\begin{aligned}b^{p-1} = (a+p)^{p-1} &\equiv a^{p-1}+(p-1)a^{p-2}p \\ &\equiv 1+(p-1)a^{p-2}p \not\equiv 1 \mod p^2\end{aligned}$$

by the Binomial Theorem, so replacing a by b works.

Claim 2. Let $^- : \mathbb{Z} \to \mathbb{Z}/p^n\mathbb{Z}$ the canonical epimorphism. If a is as in the first claim, then $\langle\overline{a}\rangle = (\mathbb{Z}/p^n\mathbb{Z})^\times$:

We know that $\varphi(p^n) = p^{n-1}(p-1)$ is the order of $(\mathbb{Z}/p^n\mathbb{Z})^\times$, so it suffices to prove that if $\overline{a}^N = 1$, then $p^{n-1} \mid N$ and $p-1 \mid N$ in $\mathbb{Z}$, as p and $p-1$ are relatively prime. By the first claim, we can write $a = 1+xp$ for some integer x not divisible by p. By the last lemma, $\overline{1+xp}$ has order p^{n-1} in $(\mathbb{Z}/p^n\mathbb{Z})^\times$. As $\overline{a}^N = 1$, we must have $p^{n-1} \mid N$. Write $N = p^{n-1}M$ for some integer M. By Fermat's Little Theorem, $1 \equiv (a^N) = (a^M)^{p^{n-1}} \equiv a^M \mod p$, so $p-1 \mid M$ as a is relatively prime to p. Hence $\overline{a}^M$ has order $p-1$. This establishes Claim 2 and the theorem. □

We know that $(\mathbb{Z}/2\mathbb{Z})^\times \cong 1$ and $(\mathbb{Z}/2^2\mathbb{Z})^\times \cong \mathbb{Z}/2\mathbb{Z}$ are cyclic, however, this is not the case for $(\mathbb{Z}/2^n\mathbb{Z})^\times$ when $n > 2$.

Proposition 105.5. *Let $n \geq 3$ be an integer. Then $(\mathbb{Z}/2^n\mathbb{Z})^\times \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2^{n-2}\mathbb{Z}$,*

PROOF. Let $n \geq 3$. We begin with the following:

Claim. $5^{2^{n-3}} \equiv 1 + 2^{n-1} \mod 2^n$:

This is true for $n = 3$, so we proceed by induction. As $n \geq 3$, we have $2n - 2 \geq n + 1$. Assuming the result for $n - 3$ we show it holds for $n - 2$. By Lemma 105.1

$$5^{2^{n-2}} = (5^{2^{n-3}})^2 = (1 + 2^{n-1})^2 = 1 + 2^n + 2^{2n-2} \equiv 1 + 2^n \mod 2^{n+1},$$

proving the claim. The claim implies that $5^{2^{n-3}} \equiv 1+2^{n-1} \not\equiv 1 \mod 2^n$ and $5^{2^{n-2}} \equiv 1 + 2^n \mod 2^{n+1}$. It follows that $5^{2^{n-2}} \equiv 1 \mod 2^n$, and the order of $\overline{5}$ in $(\mathbb{Z}/2^n\mathbb{Z})^\times$ is 2^{n-2}, where $^- : \mathbb{Z} \to \mathbb{Z}/2^n\mathbb{Z}$ is the canonical epimorphism. To finish, it suffices to show that the subgroup $\langle\overline{-1}, \overline{5}\rangle$ in $\mathbb{Z}/2^n\mathbb{Z}$ has order 2^{n-1}, as this would imply that $(\mathbb{Z}/2^n\mathbb{Z})^\times \cong \langle\overline{-1}\rangle \times \langle\overline{5}\rangle$. So suppose that $(-1)^a 5^b \equiv (-1)^{a'} 5^{b'} \mod 2^n$ for some positive integers a, a', b, b'. As $n \geq 2$ and $5 \equiv 1 \mod 4$, we have $(-1)^a \equiv (-1)^{a'} \mod 4$, hence $a \equiv a' \mod 2$. It follows that $5^b \equiv 5^{b'} \mod 2^n$. We may assume that $b \geq b'$, then $5^{b-b'} \equiv 1 \mod 2^n$. Since $\overline{5}$ has order 2^{n-2}, we have $2^{n-2} \mid b - b'$, i.e., $b \equiv b' \mod 2^n$. The result follows. □

The two propositions and the Chinese Remainder Theorem yield the desired result.

Theorem 105.6. *Let n be a positive integer. Then $(\mathbb{Z}/n\mathbb{Z})^\times$ is cyclic if and only if $n = 2, 4, p^r$, or $2p^r$ where p is an odd prime.*

106. Pell's Equation

In this section, we show that the well-known diophantine equation

$$(*) \qquad x^2 - dy^2 = 1 \text{ with } d \text{ a positive integer},$$

is solvable, i.e., there exist integers x, y satisfying (*). In elementary number theory, it is shown that the solutions may be found using continued fractions, a method based on repeated use of the division algorithm, although it may not so easy to calculate. We need two lemmas.

Lemma 106.1. *Let α be an irrational number. Then there exist infinitely many relatively prime integers x, y satisfying*

$$|\frac{x}{y} - \alpha| < \frac{1}{y^2}.$$

PROOF. Let n be a positive integer. Partition the half-open interval $[0, 1)$ into n half-open subintervals,

$$(\dagger) \qquad [0,1] = \bigvee_{j=0}^{n-1} [\frac{j}{n}, \frac{j+1}{n}).$$

Recall if β is a real number then $[\beta]$ is defined to be the largest integer $\leq \beta$, and $\beta - [\beta]$ is then called the *fractional part* of β. Consider the $n+1$ fractional parts of the elements $0, \alpha, 2\alpha, \ldots, n\alpha$. By the Dirichlet Pigeon Hole Principle, two of these fractional parts must lie in the same subinterval in (†). Hence there exist integers j and k satisfying $0 \leq j < k \leq n$ and

$$|(k\alpha - [k\alpha]) - (j\alpha - [j\alpha])| < \frac{1}{n}.$$

Set $y = k - j$ and $x = [j\alpha] - [k\alpha]$. We have $0 < y < n$ and

$$0 < |\frac{x}{y} - \alpha| < \frac{1}{yn} < \frac{1}{y^2}.$$

Therefore, x/y is one solution. We may assume that x and y are relatively prime, since dividing by the gcd of x and y would also lead to a solution. Now choose an integer $n_1 > 1/|(x/y) - \alpha|$. Repeating the construction would yield a new solution x_1, y_1 in relatively prime integers. If follows that we can construct infinitely many such solutions. □

Lemma 106.2. *Let d be a positive square-free integer. Then there exists a constant c satisfying*

$$|x^2 - dy^2| < c$$

has infinitely many solutions in integers x, y.

PROOF. We have $x^2 - dy^2 = (x - y\sqrt{d})(x + y\sqrt{d})$ in $\mathbb{Z}[d]$. By the lemma, there exist infinitely many relatively prime integers x, y with $y > 0$ and satisfying $|x - y\sqrt{d}| < 1/y$. Therefore,

$$|x + y\sqrt{d}| < |x - y\sqrt{d}| + 2y\sqrt{d} < \frac{1}{y} + 2y\sqrt{d}$$

and

$$|x^2 - dy^2| < \frac{1}{y^2} + 2\sqrt{d} < 1 + 2\sqrt{d},$$

so $c = 1 + 2\sqrt{d}$ works. □

Proposition 106.3. *Let d be a square-free positive integer. Then Pell's equation $x^2 - dy^2 = 1$ has infinitely many solutions in integers. Further, there exists an integral solution (x_1, y_1) such that every solution is of the form (x_n, y_n) with $x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})^n$ for some integer n, In particular, the solution set in integers to Pell's equation forms an infinite cyclic group.*

PROOF. We begin with some notation. If $\gamma = x + y\sqrt{d}$, x, y in $\mathbb{Q}$, let $\overline{\gamma} = x - y\sqrt{d}$ and $\mathrm{N}(\gamma) = \gamma\overline{\gamma} = x^2 - dy^2$. By the last lemma, we know that there exists a nonzero integer m such that $x^2 - dy^2 = m$ has infinitely many solutions x, y in positive integers. We may also assume that we have reduced this set of solutions to those having different x's. In particular, among these infinitely many solutions in positive integers, there exist two such pairs x_1, y_1 and x_2, y_2 to this equation with $x_1 \neq x_2$, $x_1 \equiv x_2 \mod |m|$, and $y_1 \equiv y_2 \mod |m|$. Let $\alpha = x_1 + y_1\sqrt{d}$ and $\beta = x_2 + y_2\sqrt{d}$. By choice, $\mathrm{N}(\alpha) = m = \mathrm{N}(\beta)$. Write $\alpha\overline{\beta} = r + s\sqrt{d}$ with r and s integers satisfying $r \equiv 0 \equiv s \mod |m|$. It follows that

$$\alpha\overline{\beta} = m(a + b\sqrt{d})$$

for some integers a and b. Taking N of this equation yields $a^2 - db^2 = 1$. Suppose that $b = 0$, then $a = \pm 1$ and $\alpha m = \alpha\beta\overline{\beta} = \pm m\beta$, i.e., $\alpha = \pm\beta$. This implies $x_1 = x_2$, a contradiction. It follows that Pell's equation has a solution a, b to $a^2 - db^2 = 1$ with $ab \neq 0$.

Among all the solutions a, b to Pell's equation choose one in positive integers a, b with the real number $\alpha = a + b\sqrt{d} > 0$ minimal. Suppose that x, y is another solution in positive integers. Set $\beta = x + y\sqrt{d}$. We show that $\beta = \alpha^n$ for some positive integer n. Suppose not. Then there exists a positive integer n such that $\alpha^n < \beta < \alpha^{n+1}$. Since $\alpha\overline{\alpha} = 1$, i.e., $\alpha^{-1} = \overline{\alpha}$, we have

$$1 < \beta\alpha^{-n} < \alpha.$$

We can write $1 < \beta\alpha^{-n} = \beta\overline{\alpha}^n = r + s\sqrt{d}$ with r and s integers. As $\mathrm{N}(\beta\overline{\alpha}^n) = \mathrm{N}(\beta)(\overline{\alpha}) = 1$, we have r, s is also a solution of Pell's equation. Since $r + s\sqrt{d} > 0$, we have $0 < r - s\sqrt{d} = (r + s\sqrt{d})^{-1} < 1$. In particular, r is positive and $s\sqrt{d} > r - 1 \geq 0$. Therefore, s is also positive, contradicting the minimality of a, b. If x, y is a solution in integers to Pell's equation with $x > 0$ and $y < 0$, then $\beta = x + y\sqrt{d}$ satisfies $\mathrm{N}(\beta) = 1$, so $\overline{\beta} = \alpha^n$ for some n, hence $\beta = \alpha^{-n}$. The cases $x < 0$, $y > 0$ and $x < 0, y < 0$ lead to solutions $-\alpha^n$ for some n in $\mathbb{Z}$. □

Bibliography

Texts on Abstract Algebra

[1] M. Artin, *Algebra*, Prentice Hall Inc., Englewood Cliffs, NJ, 1991.

[2] Cohn, P. M., *Algebra. Vol. 1*, Second Edition, John Wiley & Sons Ltd., Chichester, 1982.

[3] Cohn, P. M., *Algebra. Vol. 2*, Second Edition, John Wiley & Sons Ltd., Chichester, 1989.

[4] Cohn, P. M., *Algebra. Vol. 3*, Second Edition, John Wiley & Sons Ltd., Chichester, 1991.

[5] Dummit, David S. and Foote, Richard M., *Abstract algebra*, Third Edition, John Wiley & Sons Inc., Hoboken, NJ, 2004.

[6] Herstein, I. N., *Topics in algebra*, Second Edition, Xerox College Publishing, Lexington, Mass., 1975.

[7] Jacobson, Nathan, *Basic algebra. I*, Second Edition, W. H. Freeman and Company, New York, 1985.

[8] Jacobson, Nathan, *Basic algebra. II*, Second Edition, W. H. Freeman and Company, New York, 1989.

[9] Lang, Serge, *Algebra*, Graduate Texts in Mathematics, Vol. 211, Third Edition, Springer-Verlag, New York, 2002.

[10] Rotman, Joseph J., *Advanced modern algebra*, Graduate Studies in Mathematics, Vol. 114, Second Edition, American Mathematical Society, Providence, RI, 2010.

Books on Group Theory

[11] Rotman, Joseph J., *An introduction to the theory of groups*, Graduate Texts in Mathematics, Vol. 148, Fourth Edition, Springer-Verlag, New York, 1995.

[12] Rose, John S., *A course on group theory*, Reprint of the 1978 original [Dover, New York]], Dover Publications Inc., New York, 1994.

Books on Field Theory

[13] Artin, Emil, *Galois theory*, Second Edition, Edited and with a supplemental chapter by Arthur N. Milgram, Dover Publications Inc., 1998.

[14] Cox, David A., *Galois theory*, Pure and Applied Mathematics (Hoboken), Second Edition, John Wiley & Sons Inc., Hoboken, NJ, 2012.

[15] Hadlock, Charles Robert, *Field theory and its classical problems*, Carus Mathematical Monographs, Vol. 19, Mathematical Association of America, Washington, D.C., 1978.

[16] Kaplansky, Irving, *Fields and rings*, Chicago Lectures in Mathematics, Reprint of the second (1972) edition, University of Chicago Press, Chicago, IL, 1995.

[17] McCarthy, Paul J., *Algebraic extensions of fields*, Second Edition, Dover Publications Inc., New York, 1991.

Books on Number Theory

[18] Apostol, Tom M., *Introduction to analytic number theory*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1976.

[19] LeVeque, William Judson, *Topics in number theory. Vol. I, II*, Reprint of the 1956 original [Addison-Wesley Publishing Co., Inc., Reading, Mass.; MR0080682 (18,283d)], with separate errata list for this edition by the author, Dover Publications Inc., Mineola, NY, 2002.

[20] LeVeque, William J., *Fundamentals of number theory*, Reprint of the 1977 original, Dover Publications Inc., Mineola, NY, 1996.

[21] Ireland, Kenneth and Rosen, Michael, *A classical introduction to modern number theory*, Graduate Texts in Mathematics, Vol. 84, Second Edition, Springer-Verlag, New York, 1990.

[22] Marcus, Daniel A., *Number fields*, Universitext, Springer-Verlag, New York, 1977.

[23] Neukirch, Jürgen, *Algebraic number theory*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 322, Translated from the 1992 German original and with a note by Norbert Schappacher, With a foreword by G. Harder, Springer-Verlag, Berlin, 1999.

Book on Commutative Algebra

[24] Atiyah, M. F. and Macdonald, I. G., *Introduction to commutative algebra*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1969.

[25] Eisenbud, David, *Commutative algebra – With a view toward algebraic geometry*, Graduate Texts in Mathematics, Vol. 150, Springer-Verlag, New York, 1995.

[26] Kaplansky, Irving, *Commutative rings*, Revised, The University of Chicago Press, Chicago, Ill.-London, 1974.

Books on Non-Commutative Algebra

[27] Lam, T. Y., *A first course in noncommutative rings*, Graduate Texts in Mathematics, Vol. 131, Second Edition, Springer-Verlag, 2001.

[28] Pierce, Richard S., *Associative algebras*, Graduate Texts in Mathematics, Vol. 88, Studies in the History of Modern Science, 9, Springer-Verlag, New York, 1982.

Book on with Material on Formally Real Fields

[29] Scharlau, Winfried, *Quadratic and Hermitian forms*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 270, Springer-Verlag, Berlin, 1985.

[30] Lam, T. Y., *Introduction to quadratic forms over fields*, Graduate Studies in Mathematics, Vol. 67, American Mathematical Society, Providence, RI, 2005.

Notation

Index