Chapter 1

Origins of Modern Algebra

Modern algebra was developed to solve equations.

In this chapter we discuss some of the questions that gave rise to modern algebra. I assume that you are already familiar with some of the language of modern algebra: groups, rings, ideals, homomorphisms, fields, vector spaces, and so on. I have minimized the number of definitions in this chapter. You will find most of the basic definitions and properties of rings, modules, and homomorphisms in chapter 2.

The phrase “modern algebra” is a little vague, but it is commonly used to describe the material that appeared in van der Waerden’s book Modern Algebra that first appeared in 1930. Van der Waerden first encountered this material when he arrived at Göttingen in 1924. Among the primary developers of this material were Dedekind, Weber, Hilbert, Lasker, Macaulay, Steinitz, Noether, Artin, Krull, and Wedderburn, (on rings, ideals, and modules), Schur, Frobenius, Burnside, Schreier, and Galois (on groups and their representations). Van der Waerden had the advantage of attending lectures courses on algebra by Noether at Göttingen and Artin at Hamburg.

Van der Waerden’s book is a marvel. It is as fresh today as when it was written. Although dozens (hundreds?) of books covering similar ground have been written since, none cast the original into shadow.

1.1 From \( \mathbb{N} \) to \( \mathbb{Z} \) to \( \mathbb{Q} \) to \( \mathbb{Q} \bar{\mathbb{Q}} \), \( \mathbb{R} \) and \( \mathbb{C} \)

I disagree with the following quotation:

Die ganze Zahl schuf der liebe Gott, alles Übrige ist Menschenwerk.

God created the integers, all else is the work of man.

\textit{Kronecker}

Even the integers are the work of man. No doubt the first mathematical achievement of man was to recognize when two non-empty sets had the same cardinality. Then came the abstraction, picking a single label, one, two, three,
et cetera, to name/describe sets having the appropriate cardinality. Thus arose the natural numbers 1, 2, 3, \ldots.

There have been a number of primitive cultures which had no numbers beyond one, two, and three. Even cultures with more extended numbering systems have not always had a notion of zero.

The creation of the natural numbers, indeed, of all mathematics, was motivated by man’s desire to understand and manipulate the world. Mathematics is a practical art.

Many equations can be solved within the integers. One can postulate simple arithmetic problems arising from everyday life that can be solved within the integers. A typical example might be find an integer \( x \) such that \( x + 27 = 30 \). At a slightly more sophisticated level, one can imagine simple division problems, such as find \( x \) such that \( 3x = 60 \), that can also be solved within the positive integers. However, a mild modification, such as \( 3x = 67 \), leads to the idea of division with remainder, and suggests how mankind was led to the rational numbers.

One can also imagine the forces that prompted the notion of negative integers.

The construction of the rationals \( \mathbb{Q} \) from the integers \( \mathbb{Z} \) can be formalized in such a way that a similar process applied to any domain produces its field of fractions (see section 1.9). The next result summarizes the utility of the rational numbers in terms of solving certain kinds of equations. Notice that the result holds true if any field is substituted for the rationals.

**Theorem 1.1** If \( a, b, c \) are rational numbers with \( a \neq 0 \), then there is a unique rational number \( x \) such that \( ax + b = c \).

After linear equations come quadratics.

One of the great historical events concerning quadratics is Euclid’s famous proof that \( \sqrt{2} \) is not rational.

**Theorem 1.2** There is no rational number whose square is two.

**Proof.** Suppose to the contrary that \( x \) is a rational number such that \( x^2 = 2 \). Write \( x = a/b \) where \( a \) and \( b \) are integers. By cancelling common factors, we may assume that \( a \) and \( b \) have no common factor. Now, \( 2b^2 = a^2 \), so 2 divides \( a^2 \). Hence 2 divides \( a \), and we may write \( a = 2c \). Hence \( 2b^2 = 4c^2 \), and \( b^2 = 2c^2 \). It follows that \( b^2 \), and hence \( b \), is even. Thus \( a \) and \( b \) are both divisible by 2. This contradicts the fact that \( a \) and \( b \) are relatively prime, so we conclude that 2 cannot be a square in \( \mathbb{Q} \).

This result was no doubt motivated by the problem of computing the length of the hypotenuse of the isosceles right triangle with sides of length one.

But, for now, we focus on the proof itself, the key point of which is the fact that every non-zero element of \( \mathbb{Q} \) can be written as \( a/b \) with \( a \) and \( b \) relatively prime. This fact is a consequence of a still more elementary fact, which we summarize in the next theorem.
1.1. FROM ℕ TO ℤ TO ℚ TO ℂ, ℝ AND ℂ

**Theorem 1.3** Every non-zero integer can be written in an essentially unique way as a product of primes,

\[ p_1^{i_1} \cdots p_n^{i_n} \]

where \( p_1, \ldots, p_n \) are primes.

By a prime we mean an integer \( p \) such that its only divisors are \( \pm 1 \) and \( \pm p \). Thus, the primes are \{\pm 2, \pm 3, \pm 5, \cdots \}. When we say “essentially unique” we mean that factorizations \( 6 = 2 \cdot 3 = 3 \cdot 2 = (-3) \cdot (-1), 2 = 1 \cdot (-2), 3 \cdot (-1) \) are to be viewed as the same; they differ only by order and the inclusion of the terms \( \pm 1 \).

Two integers are relatively prime if the only numbers that divide both of them are \( \pm 1 \).

This theme, the unique factorization of integers and their relatives, reappeared often in the early development of modern algebra, and it remains a staple of introductory algebra courses.

That the Greek’s view of numbers and algebra was intimately connected to geometry is well documented. They had no problem accepting the existence of numbers of the form \( \sqrt{d} \) with \( d \) rational because Pythagoras’s theorem showed that right-angle triangles in which the lengths of two sides were rational numbers led to the conclusion that the length of the third side was of the form \( \sqrt{d} \). Accepting such numbers on an (almost) equal footing with the rationals allowed the solution of a range of quadratic equations with rational coefficients.

Thus, in modern parlance, the Greeks were quite happy computing in fields such as \( \mathbb{Q}(\sqrt{d}) \) when \( d \) is a positive rational number.

The reason why the equation \( x^2 = -1 \) has no solution in \( \mathbb{Q} \) is quite different than the reason why \( x^2 = 2 \) has no solution. One can imagine that the ancients were unconcerned by the fact that \( x^2 = -1 \) has no rational solution. It probably seemed a foolish waste of time to even consider that a problem. However, it is less apparent that an equation such as \( x^2 + 2x + 2 = 0 \) has no rational solution, and the discovery of this fact must surely have been intimately related to the discovery of the general solution to a quadratic equation. Several ancient cultures independently discovered the result that

\[ x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \]

gives the two solutions to the quadratic equation \( ax^2 + bx + c = 0 \). This formula gives a criterion that the quadratic has no solution (within the reals) if \( b^2 - 4ac < 0 \).

This, after many centuries, led to the invention/discovery of \( \sqrt{-1} \) and eventually to the notion of complex numbers. This in turn leads to the following question: if \( f(x) \) a polynomial with coefficients in a field \( k \), is there a field \( K \) containing \( k \) in which \( f \) has a zero? We take up this question in section 1.4.

Having discovered the above formula for the roots of a quadratic polynomial attention turned to the question of whether there are analogous formulas for the
CHAPTER 1. ORIGINS OF MODERN ALGEBRA

solutions to higher degree polynomials. Eventually, Galois gave a comprehensive solution to this problem, and we will encounter Galois’s theory later in this course.

Once the ancients had realized that one could pass beyond the rationals \( \mathbb{Q} \) to include roots of rational numbers and more complicated expressions built from such roots, it was natural to ask if this gave “all” numbers. This question is crystallized by asking whether \( \pi \) is the zero of a polynomial with rational coefficients. More generally, this leads the distinction between algebraic and transcendental elements over an arbitrary field.

1.2 Divisibility and Factorization

After first learning to count and add, children learn how to multiply and divide. Questions about division and factorization are of primary importance in all rings. A great part of the initial impetus for the development of abstract algebra arose from problems of division and factorization, especially in the rings most closely related to the integers, the rings of integers in number fields.

Before turning to these matters let’s first supply some rings that go a little beyond the integers. This will give us a supply of examples, and these particular examples will be important when we discuss Fermat’s theorem in section 1.8.

Let \( d \) be an integer that is not a square. We define

\[
\mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d} \mid a, b \in \mathbb{Z}\}.
\]

This is a subset of \( \mathbb{C} \) and is closed under multiplication, addition, and subtraction, meaning that the product, sum, and difference, of two elements in \( \mathbb{Z}[\sqrt{d}] \) belongs to \( \mathbb{Z}[\sqrt{d}] \). Hence \( \mathbb{Z}[\sqrt{d}] \) is a ring.

The notion of division makes sense in any ring. If \( b \) is a multiple of \( a \) we say that \( a \) divides \( b \), and write \( a \mid b \). Strictly speaking we should be explicit about the ring because \( b \) can be a multiple of \( a \) in one ring but not in another. Hence if \( a \) and \( b \) are elements of a ring \( R \), we say that \( a \) divides \( b \) in \( R \) if \( b = ar \) for some \( r \in R \).

Every element divides zero.

Zero divides no elements other than itself. At the other end of the spectrum, 1 divides every element. But 1 is not the only element with this property.

An element \( u \) in a ring \( R \) is a unit in \( R \) if there is an element \( v \in R \) such that \( uv = vu = 1 \). We call \( v \) the inverse of \( u \) and denote it by \( u^{-1} \). The inverse of an element is unique because if \( v \) and \( b \) are inverses, then \( b = b(uv) = (bu)v = v \).

Exercise.

1. Show that if \( uw = vu = 1 \), then \( v = w \).

2. Let \( V \) be an infinite dimensional vector space. Give an example of a linear map \( u : V \to V \) such that there is an element \( v : V \to V \) such that \( uv = 1 \), but \( vu \neq 1 \). Here 1 denotes the identity map.

3. Show that \( u \) divides every element of \( R \) if and only if it is a unit.
1.2. DIVISIBILITY AND FACTORIZATION

Let $d$ be a non-square integer. Let $x = a + b\sqrt{d}$ be an element of $\mathbb{Z}[\sqrt{d}]$. The norm of $x$ is
\[ N(x) = a^2 - b^2d. \]
Since $d$ is not a square, $N(x) = 0 \iff x = 0$. The other important property of the norm is that $N(xy) = N(x)N(y)$.

Because the norm is an integer, a factorization $a = xy$ in $\mathbb{Z}[\sqrt{d}]$ implies the
the factorization $N(a) = N(x)N(y)$ in $\mathbb{Z}$. This gives us a tool for studying
factorization questions in $\mathbb{Z}[\sqrt{d}]$.

If $d$ is a negative integer the norm of an element $x$ in $\mathbb{Z}[\sqrt{-d}]$ is equal to

\[ \bar{x} = |x|^2, \]

where $\bar{x}$ is its complex conjugate.

**Lemma 2.1** Let $d$ be a negative integer.

1. The element $x = a + b\sqrt{d}$ is a unit in $\mathbb{Z}[\sqrt{d}]$ if and only if $N(x) = 1$.

2. The units in $\mathbb{Z}[i]$ are $\{\pm 1, \pm i\}$.

3. If $d \neq -1$, the units in $\mathbb{Z}[\sqrt{d}]$ are $\{\pm 1\}$.

**Proof.** Since $d < 0$, $N(x) \geq 0$. Certainly, if $x$ is a unit, then $1 = N(1) = N(xx^{-1}) = N(x)N(x^{-1})$, so we conclude that $N(x) = 1$. Conversely, suppose that $N(x) = 1$. Then $x \neq 0$, and it has an inverse in $\mathbb{C}$, namely

\[ x^{-1} = \frac{1}{a + b\sqrt{d}} \frac{a - b\sqrt{d}}{a - b\sqrt{d}} = a - b\sqrt{d}. \]

This belongs to $\mathbb{Z}[\sqrt{d}]$ so $x$ is a unit in $\mathbb{Z}[\sqrt{d}]$.

The only way $a^2 - b^2d$ can equal 1 is if $a^2 = 1$ and $b = 0$, leading to the units $\pm 1$, or if $a = 0$, $d = -1$ and $b^2 = 1$, leading to the units $\pm i$ in $\mathbb{Z}[i]$. \( \square \)

**Definition 2.2** Let $R$ be a commutative ring. A non-zero non-unit $a \in R$ is
irreducible if in every factorization $a = bc$ either $b$ or $c$ is a unit. A non-zero
non-unit $p \in R$ is called a prime if in every division $p|bc$ either $p|b$ or $p|c$. \( \diamond \)

A prime is irreducible: if $p = bc$ then, perhaps after relabelling the factors, $p|b$, so $b = pu$ and $p = puc$, so $1 = uc$, whence $c$ is a unit.

The converse is not usually true though; an irreducible need not be prime.

For example, the only divisors of 2 in $\mathbb{Z}[\sqrt{-5}]$ are $\pm 2$ and $\pm 1$, so 2 is irreducible, and although 2 does not divide either $1 + \sqrt{-5}$ or $1 - \sqrt{-5}$ in $\mathbb{Z}[\sqrt{-5}]$, it divides their product:

\[ (1 + \sqrt{-5})(1 - \sqrt{-5}) = 6 = 2 \cdot 3. \]

Hence 2 is not a prime in $\mathbb{Z}[\sqrt{-5}]$.

This leads to the question of identifying those domains in which every irreducible
element is prime. The answer appears in Lemma 10.2.

Notice that 2 is not a prime in $\mathbb{Z}[i]$ because $2 = (1 + i)(1 - i)$. However, $i + i$ and $1 - i$ are both irreducible because, for example, if $1 + i = xy$, then
$N(x)N(y) = N(1 + i) = 2$ so the norm of either $x$ or $y$ is equal to $\pm 1$, and hence either $x$ or $y$ is a unit.

**Exercise.** Is $2$ prime in $\mathbb{Z}[i]$. Describe exactly which prime integers remain prime in $\mathbb{Z}[i]$.

**Greatest common divisors.** Let $R$ be a domain. A greatest common divisor of two elements $a, b \in R$ is an element $d \in R$ such that

1. $d | a$ and $d | b$, and
2. if $e|a$ and $e|b$, then $e|d$.

We write $d = \gcd(a, b)$, or just $d = (a, b)$. We say that greatest common divisors exist in $R$ if every pair of elements in $R$ has a greatest common divisor in $R$.

The greatest common divisor is not unique. For example, in the ring of integers, both $2$ and $-2$ are greatest common divisors of $6$ and $10$. Similarly, in $\mathbb{Z}[i]$ both $2$ and $2i$ are greatest common divisors of $4$ and $6$. In general, if $d$ and $d'$ are two greatest common divisors of $a$ and $b$, then each is a unit multiple of the other: because each divides the other, we have $d' = du$ and $d = d'v$, so $d(uv - 1) = 0$, whence $uv = 1$.

To obtain uniqueness of a greatest common divisor we need some additional structure on $R$. For example, in $\mathbb{Z}$ if we also insist that the greatest common divisor be positive, then it becomes unique.

Actually, we haven’t even shown that greatest common divisors exist in $\mathbb{Z}$ or $\mathbb{Z}[\sqrt{d}]$. There is something to do here.

We can define the greatest common divisor of any collection of elements by saying that $d$ is a greatest common divisor of $a_1, \ldots, a_n$ if it divides each $a_i$, and if $e$ is any element of $R$ dividing all of them, then $e$ necessarily divides $d$.

### 1.3 Fields

I assume you are familiar with fields such as the real numbers, $\mathbb{R}$, the rational numbers, $\mathbb{Q}$, and the complex numbers $\mathbb{C}$. A field $k$ is a non-empty set of elements which can be added and multiplied with the usual rules holding. The crucial feature of a field is that every non-zero element of it has an inverse; that is, if $\alpha$ is a non-zero element of $k$, there is an element $\alpha^{-1}$ in $k$ such that $\alpha^{-1}\alpha = \alpha\alpha^{-1} = 1$.

**Exercise.** Look up the definition of a field in a textbook. Ponder this point: the examples came before the definition; $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, have some common properties; abstracting from these properties leads to a definition which captures the important features of those examples.

It would be foolish to develop a theory of fields if these were the only examples. Fields abound. Finite fields, the simplest examples of which appear in the next exercise, play a central role in number theory, and in applications of algebra to communications, coding theory, and several other computer-related areas. Number fields, the simplest examples of which are the quadratic extensions $\mathbb{Q}(\sqrt{d})$ of the rationals, occupy a central place in number theory and all
1.3. 

**Fields**

arithmetic questions. Function fields, being fields consisting of ratios of functions defined on geometric objects, are central to algebraic geometry, complex analysis, and other areas.

**Exercise.** Let \( p \) be a positive prime number and define \( \mathbb{F}_p \) to consist of the \( p \) elements \([0], [1], \ldots, [p-1] \) defined as follows:

\[
[i] := \{a \in \mathbb{Z} | p \text{ divides } a - i\}. \tag{3-1}
\]

Actually, we use (3-1) to define \([i]\) for every \( i \in \mathbb{Z} \), but because \([i] = [i + np]\), we only obtain \( p \) distinct \([i]\)'s. Define

\[
[i] + [j] := [i + j] \quad \text{and} \quad [i] \cdot [j] := [ij]. \tag{3-2}
\]

Show that these definitions of + and \( \cdot \) in \( \mathbb{F}_p \) are unambiguous: show that \([i] = [i']\) and \([j] = [j']\), then \([i] + [j] = [i'] + [j']\) and \([i][j] = [i'][j']\). Show that \( \mathbb{F}_p \) is a field. Look up the definition of a field if necessary. Check that \( \mathbb{F}_p \) has inverses: if \( i \) is an integer that is not divisible by \( p \), show there is an integer \( j \) such that \( ij - 1 \) is divisible by \( p \), and hence that \([i][j] = 1 = [1]\). We write \([j] = [i]^{-1}\).

**Exercise.** If \( d \) is a rational number we define

\[
\mathbb{Q}(\sqrt{d}) := \{\alpha + \beta\sqrt{d} | \alpha, \beta \in \mathbb{Q}\}.
\]

This subset of \( \mathbb{C} \) is closed under multiplication and addition, meaning that the product and sum of two in \( \mathbb{Q}(\sqrt{d}) \) belong to \( \mathbb{Q}(\sqrt{d}) \). For that reason we call \( \mathbb{Q}(\sqrt{d}) \) a subring of \( \mathbb{C} \). Check that the inverse (in \( \mathbb{C} \)) of a non-zero element of \( \mathbb{Q}(\sqrt{d}) \) belongs to \( \mathbb{Q}(\sqrt{d}) \). Accordingly, we call \( \mathbb{Q}(\sqrt{d}) \) a subfield of \( \mathbb{C} \) and \( \mathbb{C} \) an extension of \( \mathbb{Q}(\sqrt{d}) \).

**Exercise.** Let \( n \) be a positive integer and \( \zeta = e^{2\pi i/n} \). Show that \( \mathbb{Q}(\zeta) := \mathbb{Q} \oplus \mathbb{Q} \zeta \oplus \cdots \oplus \mathbb{Q} \zeta^{n-1} \) is a subfield of \( \mathbb{C} \).

**Exercise.** Think of six interesting questions about the fields \( \mathbb{F}_p \), \( \mathbb{Q}(\sqrt{d}) \), and \( \mathbb{Q}(\zeta) \).

**Exercise.** The field of rational functions in one variable, denoted \( k(x) \), consists of all ratios \( p/q \) where \( p \) and \( q \) are polynomials in \( x \) having coefficients in \( k \), and \( q \neq 0 \). We add and multiply these in the obvious way. The inverse of a non-zero element \( p/q \) is \( q/p \). This is the field of rational functions on the affine line over \( k \). Likewise, the field \( k(x, y) \) of rational functions on the affine plane over \( k \) consists of all ratios \( p/q \) where \( p \) and \( q \) are polynomials in the variables \( x \) and \( y \), and \( q \neq 0 \). Are the fields \( k(x) \) and \( k(x, y) \) isomorphic? What does the word “isomorphic” mean in this context?

Later, we will examine fields in some detail, but for now we simply introduce them as a necessary preliminary for our discussion of polynomials. Fields provide the coefficients for polynomials.

The letter \( k \) is often used to denote a field because German mathematicians, who were the first to examine fields in some detail, called a field *ein körper*. 

(körper=body, cf. “corpse”). Despite this nomenclature, the study of fields remains a lively topic.

**Notes on notation.** The same symbol is often used for different things in mathematics. If the author is doing a good job, the context will provide enough information to interpret the symbol unambiguously. For example, in (3-2), the + on the left-hand of the = sign is different from the + on the right-hand side. The + on the right-hand side is the usual addition in \( Z \), but + on the left-hand is the new addition in \( F_p \). We are using the old addition in \( Z \) to define the new addition in \( F_p \).

We will use the symbol 0 to denote the zero element in all the rings we meet, so you need to be alert as to which zero is being meant. Likewise, the symbol 1 is used to denote the unit element in a ring. So, rather than writing \([1]\) or \([0]\) for the unit and zero in \( F_p \), we simply write 1 or 0.

If \( i \) is an integer, we might write \( i \), for the element \([i]\) of \( F_p \). If we could all agree to be careful, we could even write \( i \) for \([i]\) ! Think of the time and effort we would save by doing this; the price is eternal vigilance...with apologies to Thomas Jefferson “The price of liberty is eternal vigilance”.

**Exercise.** Suppose that \( R \) is a ring containing a field \( k \) as a subring. Show that the addition and multiplication on \( R \) give it the structure of a vector space over \( k \).

**Exercise.** Sometimes we write \((a, b)\) for the greatest common divisor of two integers \( a \) and \( b \). This notation is also used to denote the ideal generated by \( a \) and \( b \). Show there is an equality of ideals, \((a, b) = (d)\), if \( d \) is a greatest common divisor of \( a \) and \( b \).

### 1.4 The polynomial ring in one variable

Throughout this section \( k \) denotes a field.

In this section we show that the ring of polynomials in one variable with coefficients in \( k \) behaves in many respects like the ring of integers. Our initial focus is on questions of division and factorization.

Let \( R \) be a commutative ring. To begin with you might think of \( R \) being the integers, or the rationals, or the reals.

Polynomials in one variable, say \( x \), with coefficients in \( R \) can be added and multiplied in the obvious way to produce another polynomial with coefficients in \( R \).

We write \( R[x] \) for the set of all polynomials in \( x \) with coefficients in \( R \). An element of \( R[x] \) is an expression

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

where the coefficients \( a_i \) belong to \( R \). Addition and multiplication are defined in the obvious way. Two polynomials are considered to be the same only if all their coefficients are the same. In this way \( R[x] \) becomes a ring, with zero element the zero polynomial \( 0 \), and identity element the constant polynomial \( 1 \).
14. THE POLYNOMIAL RING IN ONE VARIABLE

Definition 4.1 Let $R$ be a ring. The polynomial ring with coefficients in $R$, which we denote by $R[x]$, consists of all formal expressions

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$$

where $a_0, \ldots, a_n \in R$, and this is made into a ring by defining the sum and product of two polynomials by

$$\sum a_i x^i + \sum \beta_i x^i := \sum (a_i + \beta_i) x^i$$

and

$$\left( \sum a_i x^i \right) \left( \sum \beta_i x^i \right) := \sum_n \left( \sum_{j=0}^{n} (a_j \beta_{n-j}) \right) x^n.$$

We call $a_0, \ldots, a_n$ the coefficients of $\sum_{i=0}^{n} a_i x^i$. We say that two polynomials are equal if and only if they have the same coefficients.

We call $x$ an indeterminate. \hfill \Box

We leave it to the reader to check that $R[x]$ is a ring.

We are particularly interested in the case when $R$ is a field.

Recall that if $a$ and $b$ are integers with $b$ non-zero, then there are integers $q$ and $r$ such that $a = bq + r$ and $0 \leq r < |b|$. We usually call $r$ the remainder. This result plays a key role in arithmetic. To show that there is an analogous result for $k[x]$ we need a notion of “size” to replace absolute value.

The degree of a non-zero element $f = a_n x^n + \ldots + a_1 x + a_0$ in $R[x]$ is $n$ provided that $a_n \neq 0$. In that case we call $a_n$ the leading coefficient of $f$. If $f = 0$ it is convenient to define its degree to be $-\infty$. It is a trivial observation that the units in $k[x]$ are precisely the polynomials of degree zero.

Lemma 4.2 Let $R$ be a domain and let $f, g \in R[x]$. Then

1. $\deg(f + g) \leq \max\{\deg f, \deg g\}$;
2. $\deg(fg) = \deg f + \deg g$;
3. $R[x]$ is a domain.

Proposition 4.3 If $f$ and $g$ are non-zero elements of $k[x]$ such that $f$ is non-zero, then there are unique polynomials $q$ and $r$ such that

$$g = fq + r \quad \text{and} \quad \deg r < \deg f.$$

Proof. Existence. We argue by induction on $\deg g$. If $g = 0$, we can take $q = r = 0$. If $\deg g < \deg f$, we can take $q = 0$ and $r = g$. If $m = \deg g \geq \deg f = n$, we can write

$$g = \alpha x^n + \ldots \text{lower degree terms}$$

$$f = \beta x^n + \ldots \text{lower degree terms}.$$
CHAPTER 1. ORIGINS OF MODERN ALGEBRA

Since
\[ \deg(g - \alpha \beta^{-1} x^m - n f) < \deg g, \]
we may apply the induction hypothesis to \( g - \alpha \beta^{-1} x^m - n f \).

Uniqueness. If \( g = f q + r = f q' + r' \), then \( f(q - q') = r' - r \). But \( \deg(r' - r) < \deg f \), so this implies that \( r' - r = 0 \). Hence \( q' = q \) also. \( \square \)

**Proposition 4.4** Every pair of non-zero elements in \( k[x] \) has a greatest common divisor.

**Proof.** To prove this, we need to introduce the Euclidean algorithm. The Euclidean algorithm is a constructive method that produces the greatest common divisor of two polynomials, as we now show. \( \square \)

**The Euclidean algorithm.** Let \( f \) and \( g \) be elements of \( k[x] \) with \( f \) non-zero. By repeatedly using Proposition 4.3 we may write

\[
\begin{align*}
g &= f q_1 + r_1 & \text{with} & \deg r_1 < \deg f, \\
f &= r_1 q_2 + r_2 & \text{with} & \deg r_2 < \deg r_1, \\
r_1 &= r_2 q_3 + r_3 & \text{with} & \deg r_3 < \deg r_2, \\
& \vdots & & \vdots \\
r_i &= r_{i+1} q_{i+2} + r_{i+2} & \text{with} & \deg r_{i+2} < \deg r_{i+1}, \\
& \vdots & & \vdots \\
r_t &= r_{t+1} q_{t+2}. \\
\end{align*}
\]

Since the degrees of the remainders \( r_i \) are strictly decreasing, this process must stop. Stopping means that the remainder must eventually be zero. If \( r_{t+2} = 0 \), and we set \( r_{-1} = g \) and \( r_0 = f \), then the general equation becomes

\[ r_i = r_{i+1} q_{i+2} + r_{i+2} \quad \text{with} \quad \deg r_{i+2} < \deg r_{i+1}, \quad (4-4) \]

and the last equation becomes

\[ r_t = r_{t+1} q_{t+2}. \]

**Claim:** \( r_{t+1} = \gcd(f, g) \). **Proof:** Since \( r_{t+1} \) divides \( r_t \), it follows from (4-4) that \( r_{i+1} \) also divides \( r_{i-1} \). By descending induction, (4-4) implies that \( r_{t+1} \) divides all \( r_i \), \( i \geq -1 \). In particular, \( r_{t+1} \) divides \( f \) and \( g \). On the other hand, if \( e \) divides both \( f \) and \( g \), then it divides \( r_1 \). If \( e \) divides \( r_i \) and \( r_{i+1} \), then it follows from (4-4) that it also divides \( r_{i+2} \). By induction, \( e \) divides \( r_{t+1} \). Hence \( r_{t+1} \) is a greatest common divisor of \( f \) and \( g \). \( \diamond \)

This procedure for finding the greatest common divisor of \( f \) and \( g \) is called the Euclidean algorithm. It completes the proof of Proposition 4.4.

If \( K \) is a field containing \( k \), then \( K[x] \) contains \( k[x] \). Hence, if \( f \) and \( g \) belong to \( k[x] \), we can ask for their greatest common divisor in \( k[x] \), and for their greatest common divisor in \( K[x] \). These are the same. This is because the uniqueness of \( q \) and \( r \) in Proposition 4.3 ensures that carrying out the Euclidean algorithm in \( k[x] \) for a pair \( f, g \in k[x] \) produces exactly the same result as carrying out the Euclidean algorithm in \( K[x] \) for that pair.
1.4. THE POLYNOMIAL RING IN ONE VARIABLE

Proposition 4.5 Let \( d \) be a greatest common divisor in \( k[x] \) of non-zero elements \( f \) and \( g \). Then \( d = af + bg \) for some \( a \) and \( b \).

Proof. Since a greatest common divisor is unique up to a scalar multiple, we can assume that \( d = r_{t+1} \), the last remainder produced by Euclidean algorithm. Working backwards, we have

\[
r_{t+1} = r_{t-1} - r_t q_{t+1} = r_{t-1} - (r_{t-2} - r_{t-1} q_t) q_{t+1} = \cdots ,
\]

and so on. Eventually we obtain an expression in which every term is a multiple of either \( r_0 = f \) or \( r_{t-1} = g \). Hence the result. \( \square \)

Let \( f \in k[x] \). We write \((f)\) for the set of all multiples of \( f \). That is,

\[
(f) = \{ fg \mid g \in k[x] \}.
\]

It is clear that \((f)\) contains zero. The sum and difference of two multiples of \( f \) are multiples of \( f \). Any multiple of a multiple of \( f \) is a multiple of \( f \). Hence \((f)\) is an ideal of \( k[x] \). We call it the principal ideal generated by \( f \).

Theorem 4.6 Every ideal in \( k[x] \) is principal.

Proof. The zero ideal consists of all multiples of zero, so is principal. If \( I \) is a non-zero ideal, choose a non-zero element \( f \) in it of minimal degree. Clearly \((f) \subseteq I \). If \( g \) is an element of \( I \), we may write \( g =fq+r \) with \( \deg r < \deg f \).

However, if \( r = g - fq \) is a multiple of \( f \), then \( g \in (f) \). Hence \( r = 0 \). Hence we conclude that \( r = 0 \). Hence \( g \in (f) \). Thus \( I = (f) \). \( \square \)

Notice that \((f)\) is generated by \( \lambda f \) if \( \lambda \) is a non-zero element of \( k \). Conversely, if \((f) = (g)\), then \( g \) and \( f \) must be multiples of each other, so \( g = \lambda f \) for some non-zero \( \lambda \) in \( k \). Hence, if \( I \) is a non-zero ideal in \( k[x] \), there is a unique monic polynomial \( f \) such that \( I = (f) \).

The next result is one way to recognize some irreducible polynomials.

Proposition 4.7 (Eisenstein’s criterion) Let \( f = a_n x^n + \cdots + a_1 x + a_0 \in \mathbb{Z}[x] \). Suppose there is a prime \( p \) such that

1. \( p \) does not divide \( a_n \),
2. \( p \) divides all the other coefficients,
3. \( p^2 \) does not divide \( a_0 \).

Then \( f \) is irreducible in \( \mathbb{Q}[x] \).

Proof. Suppose to the contrary that \( f \) is not irreducible in \( \mathbb{Q}[x] \). Part (1) of Lemma 10.5 implies that there are polynomials \( g, h \in \mathbb{Z}[x] \) of positive degree such that \( f = gh \).

Passing to \( \mathbb{Z}_p = \mathbb{Z}/(p) \) and \( \mathbb{Z}_p[x] \), this implies that \( \bar{f} = \bar{g}\bar{h} \) in \( \mathbb{Z}_p[x] \), where \( \bar{f} \) denotes the image of \( f \) in \( \mathbb{Z}_p[x] \) (i.e., the polynomial obtained by reducing all
the coefficients of \( f \) modulo \( p \). Thus \( \bar{g} \bar{h} = \alpha_n x^n \neq 0 \). Hence all the coefficients of \( g \) and \( h \), except their leading ones, are divisible by \( p \). In particular, their constant terms, say \( b_0 \) and \( c_0 \) are divisible by \( p \). Hence \( p^2 \) divides \( b_0 c_0 = a_0 \), a contradiction.

One of the most important applications of Eisenstein’s criterion is to prove the irreducibility of the cyclotomic polynomials

\[
x^{p-1} + \cdots + x + 1,
\]

where \( p \) is a prime. Notice that the zeroes of this are the \( p^{th} \) roots of unity \( e^{2\pi i/p}, 1 \leq n \leq p-1 \).

**Corollary 4.8** Let \( p \) be a prime. The polynomial \( x^{p-1} + \cdots + x + 1 \) is irreducible.

**Proof.** Write \( f(x) = x^{p-1} + \cdots + x + 1 \). If we substitute \( y = x - 1 \) into the equality \((x - 1)f(x) = x^p - 1\), we get

\[
yf(y + 1) = (y + 1)^p - 1 = y^p + \binom{p}{1} y^{p-1} + \cdots + \binom{p}{p-1} y + 1.
\]

If \( 1 \leq i \leq p-1 \), then \( p \) divides \( \binom{p}{i} \), so factoring out \( y \) shows that \( f(y + 1) \) satisfies Eisenstein’s criterion, and is therefore irreducible. Hence \( f(x) \) is irreducible.

We now consider the quotient rings of \( k[x] \).

**Lemma 4.9** If \( f \) is a polynomial of degree \( n \geq 0 \), then \( \dim_k k[x]/(f) = n \), and the images of \( 1, x, \ldots, x^{n-1} \) are a basis for \( k[x]/(f) \).

**Proof.** The natural homomorphism \( \pi : k[x] \to k[x]/(f) \) sends \( k \) to an isomorphic copy of itself in \( k[x]/(f) \), so we think of \( k \) as a subring of \( k[x]/(f) \). Multiplication in \( k[x]/(f) \) therefore gives \( k[x]/(f) \) the structure of a \( k \)-vector space. Since the powers of \( x \) are a basis for \( k[x] \), their images span \( k[x]/(f) \).

If \( g \) is any element of \( k[x] \), then \( x = af + r \) for some \( a \in k[x] \) and some \( r \) of degree \( < n \). Since \( \pi(g) = \pi(r) \) and since \( r \) is a linear combination of \( 1, x, \ldots, x^{n-1} \), \( \{ \pi(x^i) \mid 0 \leq i \leq n-1 \} \) spans \( k[x]/(f) \). These elements are linearly independent too because the only linear combination of \( 1, x, \ldots, x^{n-1} \) that belongs to \( (f) \) is \( 0.1 + 0.2 + \cdots + 0.3^{n-1} \).

An ideal in a ring \( R \) that is not equal to \( R \) and is contained in no ideals other than itself and \( R \) is called a maximal ideal.

One sees easily that \((x - \lambda)\) is a maximal ideal of \( k[x] \), but if \( k \) is not algebraically closed, there will be other maximal ideals.

**Lemma 4.10** An ideal \( I \) in a ring \( R \) is maximal if and only if \( R/I \) is a field.

**Proof.** Suppose that \( I \) is maximal. A non-zero element of \( R/I \) can be written as \([a + I]\) for some \( a \notin I \). Since \( I \) is maximal \( aR + I = R \). Hence there are elements \( b \in R \) and \( c \in I \) such that \( 1 = ab + c \). In \( R/I \),

\[
[a + I][b + I] = [ab + I] = [1 - c + I] = [1 + I] = 1_{R/I}.
\]
Hence \([b + I]\) is the inverse in \(R/I\) of \([a + I]\). This shows that \(R/I\) is a field.

Conversely, suppose that \(R/I\) is a field. Let \(J\) be an ideal of \(R\) that is strictly larger than \(I\). There is an element \(a \in J \setminus I\). Since \([a + I]\) is a non-zero element of \(R/I\), it has an inverse, say \([b + I]\). Since

\[
1_{R/I} = [1 + I] = [a + I][b + I] = [ab + I],
\]

\(1 - ab \in I\), and \(1 \in aR + I \subset J\). Hence \(J = R\), showing that \(I\) is maximal. \(\Box\)

**Algebraic and transcendental elements.** Let \(K\) be a field and \(k\) a subfield of \(K\). An element \(a \in K\) is said to be algebraic over \(k\) if it is a zero of a non-zero polynomial with coefficients in \(k\). That is, if

\[
\lambda_n a^n + \lambda_{n-1} a^{n-1} + \cdots + \lambda_1 a + \lambda_0 = 0
\]

for some \(\lambda_0, \ldots, \lambda_n \in k\), not all zero. An equivalent way of saying this is that the homomorphism \(\varepsilon : k[x] \to K\) given by \(\varepsilon(f) = f(a)\) is not injective.

If \(a\) is not algebraic over \(k\) we say it is transcendental over \(k\).

We say that \(k\) is algebraically closed if the only elements algebraic over \(k\) (whatever \(K\) may be) are the elements of \(k\) itself.

**Proposition 4.11** Let \(k\) be a field. The following are equivalent:

1. \(k\) is algebraically closed;
2. the only irreducible polynomials in \(k[x]\) are the degree one polynomials;
3. every polynomial in \(k[x]\) of positive degree has a zero in \(k\).

### 1.5 Fields of fractions

The formal construction of the ring \(\mathbb{Q}\) as “the field of fractions” of \(\mathbb{Z}\) may be copied for any commutative domain \(R\).

So, suppose that \(R\) is a commutative ring in which every product of non-zero elements is non-zero. let \(R^* = R - \{0\}\). Define a relation on \(R \times R^*\) by

\[
(a, b) \sim (c, d) \quad \text{if} \quad ad = bc. \tag{5.5}
\]

This is an equivalence relation. We will write \([a, b]\) for the equivalence class containing \((a, b)\), and write \(R \times R^*/\sim\) for the set of equivalence classes.

**Lemma 5.1** \(R \times R^*/\sim\) becomes a commutative ring under the definitions

\[
[a/b] + [c/d] := [ad + bc/bd],
\]

\[
[a/b], [c/d] := [ac/bd].
\]

The zero element is \([0/1]\) and the identity is \([1/1]\).
CHAPTER 1. ORIGINS OF MODERN ALGEBRA

Proof. First one must check that these binary operations are defined unambiguously. Then one must check that all the ring axioms hold. I have done that, and you should do this at least once in your life too.

Proposition 5.2 The map \( r \mapsto [r/1] \) is an injective ring homomorphism \( \iota : R \rightarrow R \times R^* / \sim \). Identifying \( R \) with its image, every non-zero element of \( R \) is a unit in \( R \times R^* / \sim \), namely \( r^{-1} = [1/r] \) if \( r \neq 0 \). Every element in \( R \times R^* / \sim \) is of the form \( rs^{-1} \) for some \( r \in R \) and \( s \in R^* \). If \( \varphi : R \rightarrow F \) is any homomorphism from \( R \) to a field \( F \) there is a unique homomorphism \( \alpha : R \times R^* / \sim \rightarrow F \) such that \( \varphi = \alpha \circ \iota \).

We call the ring \( R \times R^* / \sim \) the field of fractions of \( R \) and denote it by \( \text{Fract} \ R \).

1.6 Zeroes of polynomials

One of the great motivating problems for the development of algebra was the question of finding the zeroes, or roots, of a polynomial in one variable.

The question of whether an element \( \alpha \in k \) is a zero of a polynomial \( f \in k[x] \) can be expressed formally as follows: is \( f \) in the kernel of the ring homomorphism \( \varepsilon_\alpha : k[x] \rightarrow k \) defined by

\[
\varepsilon_\alpha(f) = f(\alpha)
\]

You should check that \( \varepsilon_\alpha \) is a ring homomorphism; indeed, the ring structure on \( k[x] \) is defined just so this is a homomorphism. The kernel of \( \varepsilon_\alpha \) is an ideal that contains \( x - \alpha \) and therefore the ideal \( (x - \alpha) \). However, \( (x - \alpha) \) is a maximal ideal. We therefore have the following result.

Lemma 6.1 If \( f \in k[x] \), then \( x - \alpha \) divides \( f \) if and only if \( f(\alpha) = 0 \).

Definition 6.2 Let \( \alpha \in k \) and \( 0 \neq f \in k[x] \). We say that \( \alpha \) is a zero of \( f \) of multiplicity \( n \) if \( (x - \alpha)^n \) divides \( f \) but \( (x - \alpha)^{n+1} \) does not.

Proposition 6.3 Let \( f \) be a monic polynomial in \( k[x] \). If \( \alpha_1, \ldots, \alpha_r \) are the distinct zeroes of \( f \), and \( \alpha_i \) is a zero of multiplicity \( n_i \), then

\[
f = (x - \alpha_1)^{n_1} \cdots (x - \alpha_r)^{n_r} g
\]

where \( g \) is a polynomial having no zeroes in \( k \).

Proof. We argue by induction on the number of zeroes and multiplicity, cancelling a factor of the form \( x - \alpha \) at each step.

The next result is very devious. It shows that if \( f \) is a non-constant polynomial with coefficients in a field \( k \), then there is a larger field \( K \) in which \( f \) has a zero. Of course, the first example that comes to mind is the polynomial \( x^2 + 1 \) in which case \( \mathbb{C} \), the field of complex numbers, contains a zero of the polynomial. However, notice that the proof is essentially a tautology.
1.6. ZEROS OF POLYNOMIALS

Proposition 6.4 Let $f$ be a non-constant polynomial in $k[x]$. Then $f$ has a zero in some extension field $K \supset k$.

Proof. Since every polynomial is a product of irreducible polynomials it suffices to prove this when $f$ is irreducible. When $f$ is irreducible ($f$) is a maximal ideal, so $K := k[x]/(f)$ is a field.

Let $\pi : k[x] \to K$ denote the natural map, and write $\bar{x} = \pi(x)$. If $f = \sum_{i=0}^{n} \lambda_i x^i$, then

$$f(\bar{x}) = \sum_{i=0}^{n} \lambda_i \bar{x}^i = \pi(\sum_{i=0}^{n} \lambda_i x^i) = \pi(f) = 0.$$ 

Hence $\bar{x}$ is a zero of $f$. \qed

This result suggests that we undertake a systematic examination of fields that contain a given field $k$.

A field $K$ is called an extension of a field $k$ if $k$ is a subfield of $K$. We give a more formal definition of an extension field on page ??.

If $K$ is an extension of $k$, the action of $k$ on $K$ by multiplication makes $K$ into a $k$-vector space. We may therefore define the degree of $K$ over $k$ to be

$$[K : k] = \dim_k K.$$ 

We say that $K$ is a finite extension if $[K : k] < \infty$.

The trivial observation that $K$ is a vector space over $k$ already has important consequences. For example, if $p$ is a prime and $K$ is a finite extension of $\mathbb{F}_p$, the field of $p$ elements, then $[K : \mathbb{F}_p] = n$ implies that $[K] = |\mathbb{F}_p^n| = |\mathbb{F}_p|^n = p^n$. On the other hand, if $K$ is a finite field, then the map $\mathbb{Z} \to K$ sending $1$ to $1$ has a non-zero kernel which must be of the form $(p)$ for some prime $p$, so $K$ is an extension of $\mathbb{F}_p$. Hence a finite field must have cardinality $p^n$ for some prime $p$ and integer $n \geq 1$.

It is natural to ask if there is a field of cardinality $p^n$. As we shall see later, there is a unique field of cardinality $p^n$ up to isomorphism. We write $\mathbb{F}_{p^n}$ for the field with $p^n$ elements.

If $f$ is an irreducible polynomial in $\mathbb{F}_p[x]$ of degree $n$, then $\mathbb{F}_p[x]/(f) \cong \mathbb{F}_{p^n}$. You should try some examples with small $n$ and $p$ and see what you can find.

Using the fact that an extension field is a vector space over any of its subfields, if $\mathbb{F}_{p^m}$ is a subfield of $\mathbb{F}_{p^n}$, then $m|n$ because if $[\mathbb{F}_{p^n} : \mathbb{F}_{p^m}] = r$, then $\mathbb{F}_{p^n}$ is isomorphic to the $r$-dimensional vector space over $\mathbb{F}_{p^m}$, and therefore

$$p^n = |\mathbb{F}_{p^n}| = |\mathbb{F}_{p^m}|^r = p^{mr}.$$ 

If $K$ is an extension of $k$, and $\alpha_1, \ldots, \alpha_n \in K$ we write

$$k(\alpha_1, \ldots, \alpha_n)$$

for the smallest subfield of $K$ that contains $k$ and $\alpha_1, \ldots, \alpha_n$. 
Example 6.5 Write \( \omega = \sqrt{2} + \sqrt{3} \) and \( K = \mathbb{Q}(\omega) \). What is \([K : \mathbb{Q}]\)? Since \( \omega^2 = 5 + 2\sqrt{6} \) and \((\omega^2 - 5)^2 = 24\), the minimal polynomial of \( \omega \) divides \((x^2 - 5)^2 - 24\). Hence \([\mathbb{Q}(\omega) : \mathbb{Q}] \leq 4\). The computation of \( \omega^2 \) shows that \( \sqrt{6} \in \mathbb{Q}(\omega) \); taking a \( \mathbb{Q} \)-linear combination of \( \sqrt{6} \omega \) and \( \omega \) shows that \( \sqrt{2} \) and \( \sqrt{3} \) are in \( \mathbb{Q}(\omega) \). Hence \( \mathbb{Q}(\sqrt{2}) \subset \mathbb{Q}(\omega) \). But \([\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2\), so \([\mathbb{Q}(\omega) : \mathbb{Q}] \) is even. An elementary computation shows that \( \sqrt{3} \notin \mathbb{Q}(\sqrt{2}) \), whence \( \mathbb{Q}(\omega) \) is strictly larger than \( \mathbb{Q}(\sqrt{2}) \). It follows that \([\mathbb{Q}(\omega) : \mathbb{Q}] \geq 4\). Hence \([\mathbb{Q}(\omega) : \mathbb{Q}] = 4 \) and the minimal polynomial of \( \omega \) is \((x^2 - 5)^2 - 24\).

Example 6.6 Let’s construct \( \mathbb{F}_{25} \). Since 2 is not a square in \( \mathbb{F}_5 \), \( x^2 - 2 \) is an irreducible polynomial in \( \mathbb{F}_5[x] \), whence \( \mathbb{F}_{25} \cong \mathbb{F}_5[x]/(x^2 - 2) \). Write \( \lambda \) for the image of \( x \) in \( \mathbb{F}_{25} \). Viewing \( \mathbb{F}_{25} \) as a two-dimensional vector space over \( \mathbb{F}_5 \), we have \( \mathbb{F}_{25} = \mathbb{F}_5 \oplus \mathbb{F}_5 \lambda \), so every element of \( \mathbb{F}_{25} \) can be written uniquely as

\[ a + b\lambda, \quad a, b \in \mathbb{F}_5 \]

and the multiplication is given by

\[ (a + b\lambda)(c + d\lambda) = ac + bd\lambda^2 + (ad + bc)\lambda = ac + 2bd + (ad + bc)\lambda. \]

Notice that 3 is not a square in \( \mathbb{F}_5 \). We can ask whether it has a square root in \( \mathbb{F}_{25} \). Now \( (a + b\lambda)^2 = 3 \) if and only if

\[ a^2 + 2ab^2 = 3 \quad \text{and} \quad 2ab = 0. \]

We see that \( a = 0 \) and \( b = 2 \) is a solution. Thus \((2\lambda)^2 = 3\).

Proposition 6.7 Let \( k \subset K \subset L \) be fields. Then \([L : k] = [L : K][K : k]\) if any two of these degrees are finite (and then the third is also finite).

Proof. If \([L : k]\) is finite, then the other two degrees are finite, so suppose that \([L : K]\) and \([K : k]\) are finite. If \( \{\alpha_1, \ldots, \alpha_n\} \) is a \( k \)-basis for \( K \) and \( \{\beta_1, \ldots, \beta_m\} \) is a \( K \)-basis for \( L \), then

\[ L = K\beta_1 \oplus \cdots \oplus K\beta_m = (k\alpha_1 \oplus \cdots \oplus k\alpha_n)\beta_1 \oplus \cdots \oplus (k\alpha_1 \oplus \cdots \oplus k\alpha_n)\beta_m = \bigoplus_{i=1}^{m} \bigoplus_{j=1}^{n} k\alpha_i \beta_j, \]

so \( \{\alpha_i\beta_j\} \) is a \( k \)-basis for \( L \). Hence \([L : k] = mn = [L : K][K : k] \).

If \( R \) is any ring containing \( k \) as a subring and \( \alpha \in R \), there is a unique ring homomorphism \( \psi : k[x] \rightarrow R \) which is the identity on \( k \) and sends \( x \) to \( \alpha \). If \( \psi \) is injective we say that \( \alpha \) is transcendental over \( k \), otherwise we say that \( \alpha \) is algebraic over \( k \), and we call the unique monic generator of the ideal \( \ker \psi \) the minimal polynomial of \( \alpha \).
1.6. ZEROS OF POLYNOMIALS

The element $\alpha$ is transcendental if and only if $\{1, \alpha, \alpha^2, \ldots\}$ is linearly independent over $k$. If $\alpha$ is algebraic, then there is a linear dependence relation between $1, \alpha, \ldots, \alpha^n$ where $n$ is the degree of the minimal polynomial of $\alpha$.

If $K$ is an extension field of $k$, and $\alpha \in K$ is algebraic over $k$, then the image of $k[x]$ is a domain and has finite dimension over $k$ (equal to the degree of the minimal polynomial of $\alpha$). Hence that image is a field, and we deduce that the minimal polynomial of $\alpha$ is irreducible. Furthermore, the image of $k[x]$ is equal to $k(\alpha)$, the subfield of $K$ generated by $k$ and $\alpha$.

**Proposition 6.8** Let $K$ be an extension of $k$ and $\alpha \in K$ an element that is algebraic over $k$. Then

$$[k(\alpha) : k] = \deg p$$

where $p$ is the minimal polynomial of $\alpha$ over $k$.

**Example 6.9** For each positive integer write $\zeta_n = e^{2\pi i/n}$. Since $x^n = 1$, the minimal polynomial of $\zeta_n$ divides $x^n - 1$. Since $x^n - 1 = (x - 1)(x^{n-1} + \cdots + x + 1)$, it follows that the minimal polynomial of $\zeta_n$ divides $x^{n-1} + \cdots + x + 1$. If $n = p$ is prime, the polynomial $x^{p-1} + \cdots + x + 1$ is irreducible by 4.8, so is the minimal polynomial of $\zeta_p$. \hfill $\Diamond$

We say that $K$ is an algebraic extension of $k$ if every element of $K$ is algebraic over $k$.

**Lemma 6.10** If $K$ is a finite extension of $k$, then it is an algebraic extension.

**Proof.** If $\alpha \in k$, then the map $k[x] \to K$, $x \mapsto \alpha$, cannot be injective for dimension reasons, so $\alpha$ is algebraic. \hfill $\square$

Remember that $\overline{Q}$, the algebraic closure of $Q$, is an algebraic extension of $Q$ that is not a finite extension.

**Lemma 6.11** Let $K$ be an extension of $k$. Then $[K : k] < \infty$ if and only if $K$ is an algebraic extension of $k$ and $K = k(\alpha_1, \ldots, \alpha_n)$ for some $\alpha_1, \ldots, \alpha_n$.

**Proof.** ($\Rightarrow$) By Lemma 6.10, $K$ is algebraic over $k$. Also, if $\alpha_1, \ldots, \alpha_n$ is a $k$-basis for $K$, then $K = k(\alpha_1, \ldots, \alpha_n)$.

($\Leftarrow$) We argue by induction on $n$, the case $n = 0$ being trivial. Set $L = k(\alpha_1, \ldots, \alpha_{n-1})$. The induction hypothesis implies that $[L : k] < \infty$. Now $K = L(\alpha_n) \cong L[x]/(p)$ where $p$ is the minimal polynomial of $\alpha_n$ over $L$. Hence $[K : L] < \infty$. Thus $[K : k] = [K : L][L : k] < \infty$. \hfill $\square$

**Three impossible constructions.** The ancients asked whether it was possible, using only a straightedge and compass, to double a cube, trisect an angle, and square the circle.

Using a straightedge and compass one can

1. draw a straight line through two given points, and
2. draw a circle with given center and radius.

Starting with a line segment of length one, we can construct the lattice \( \mathbb{Z}^2 \) of points \((a, b)\) with integer coordinates in the plane \( \mathbb{R}^2 \). For example, it is easy to construct the integer points on the \( x \)-axis, and the construct a perpendicular, and continue doing the obvious thing to obtain \( \mathbb{Z}^2 \).

A point in the plane is constructible if it can be obtained by repeating the constructions (1) and (2) in some order a suitable number of times. More precisely, suppose that at some stage we have constructed the points \( \mathcal{P} \) (initially \( \mathcal{P} \) consists of the points in the lattice \( \mathbb{Z}^2 \)); using (1) we can draw straight lines between distinct points of \( \mathcal{P} \) and the intersection points of such lines are constructible; using (2) we can draw a circle with center \( p \in \mathcal{P} \) and radius equal to a segment joining two points of \( \mathcal{P} \); the points of intersection of such circles are constructible; the points of intersection of such circles with the lines between points of \( \mathcal{P} \) are also constructible.

We say that \( a \in \mathbb{R} \) is constructible if there is \( b \in \mathbb{R} \) such that either \((a, b)\) or \((b, a)\) is constructible. Thus \( a \) is constructible if and only if we can construct a line segment of length \( a \).

The constructible numbers form a field. Obviously the sum and difference of two constructible numbers is constructible. Products can be constructed by constructing similar triangles. Inverses can be constructed similarly. To see this, suppose that \( a > 1 \) has been constructed, and consider the problem of constructing \( a^{-1} \). First construct line segments as below:

![Diagram](image)

Now construct two similar triangles by drawing the line through \((0, a)\) and \((1, 0)\) to give the big hypotenuse, and then construct a line parallel to that through \((0, 1)\) giving the smaller triangle; that line will meet the base at \((a^{-1}, 0)\), showing that \( a^{-1} \) is constructible.

Since the integers are constructible, \( \mathbb{Q} \) is constructible.

Now suppose that we have constructed all elements of a field \( k \) lying between \( \mathbb{Q} \) and \( \mathbb{R} \). \( \mathbb{Q} \subset k \subset \mathbb{R} \). That is the points constructed so far contain all \((a, b)\) in \( k^2 \subset \mathbb{R}^2 \). If we make a single new construction to obtain a point \((a, b)\) then \( a \) belongs to \( k(\sqrt{d}) \) for some \( d \in k \); similarly for \( b \). For example, Pythagoras's theorem shows that the length of the line segment joining two points of \( k^2 \) is of length \( \sqrt{d} \) for some \( d \in k \).
1.7. PYTHAGORAS AND INTEGERS

Corollary 6.12 If \( a \in \mathbb{R} \) is constructible, then there is a sequence of fields \( \mathbb{Q} = k_0 \subset k_1 \subset \cdots \subset k_n \) such that \( [k_i : k_{i-1}] = 2 \) for all \( i \). In particular, the degree of the minimal polynomial of \( a \) is of the form \( 2^m \).

**Proof.** Since \( \mathbb{Q} \subset \mathbb{Q}(a) \subset k_n, 2^m = [k_n : \mathbb{Q}] = [k : \mathbb{Q}(a)](\mathbb{Q}(a) : \mathbb{Q}). \)

Cannot double a cube. Suppose our original cube has sides of length one. The cube of twice the volume has sides of length \( 2^{1/3} \). The minimal polynomial of \( 2^{1/3} \) is \( x^3 - 2 \), so \( [\mathbb{Q}(2^{1/3}) : \mathbb{Q}] = 3 \neq 2^m \).

\( \pi/3 \) cannot be trisected. If it were possible to construct \( \pi/9 \), then \( a = 2 \cos(\pi/9) \) would be constructible. But substituting \( \theta = \pi/3 \) into the identity \( \cos 3\theta = 4 \cos^3 \theta - 3 \cos \theta \) gives \( 8x^3 - 6x - 1 = 0 \). Since \( 8x^3 - 6x - 1 \) is irreducible over \( \mathbb{Q} \) (you can check it has no zero in \( \mathbb{Z}_5 \)) it is the minimal polynomial of \( \alpha \). Hence \( [\mathbb{Q}(\alpha) : \mathbb{Q}] = 3 \neq 2^m \).

Cannot square the circle. The square with area equal to the circle of radius one has sides of length \( \sqrt{\pi} \). If this were constructible, then \( \pi \) would be constructible, and hence algebraic over \( \mathbb{Q} \). But F. Lindemann proved that \( \pi \) is transcendental in 1882.

1.7 Pythagoras and integers

This section is preparation for the discussion of Fermat’s last theorem that appears in the next section.

As already suggested in section 1.1, the fact that integers may be uniquely factored as a product of primes is a powerful tool for the analysis and solution of integer equations. For example, \( 2x = 3 \) has no solutions in the integers because 3 is not divisible by 2. Slightly more thought shows that there are no integers \( x \) and \( y \) such that \( 3x^2 + y^2 = 54 \) — if \( (x, y) \) were a solution to this equation, then \( y \) would be divisible by 3 so, writing \( y = 3a \), we would have \( x^2 + 3a^2 = 18 \), whence 3 divides \( x \), and writing \( x = 3b \), we get \( 3b^2 + a^2 = 6 \), so \( a \) is divisible by 3 and, writing \( a = 3c \), we obtain \( b^2 + 3c^2 = 2 \), an equation that obviously has no integer solutions.

The proof of the following result is elementary, albeit tedious, but notice how it uses the unique factorization property of the integers in an essential way.

**Proposition 7.1** Let \( v \) and \( w \) be relatively prime integers. If \( vw \) is a square, then both \( v \) and \( w \) are squares.

**Proof.** It is helpful in this proof to assume that all the numbers appearing in it are positive. Let’s do that. It is not essential, but it makes things a little cleaner.

Before proving the result, we show that the uniqueness of factorization implies that if a prime \( p \) divides a product \( ab \), then it must divide either \( a \) or \( b \). This is because if \( ab = px \), and we write \( a = p_1^{i_1} \cdots p_m^{i_m}, b = q_1^{j_1} \cdots q_n^{j_n} \), and \( x = r_1^{k_1} \cdots r_t^{k_t} \), as products of positive powers of distinct primes, then

\[
p r_1^{k_1} \cdots r_t^{k_t} = p_1^{i_1} \cdots p_m^{i_m} q_1^{j_1} \cdots q_n^{j_n}
\]
so, by the uniqueness of factorization into primes, \( p \in \{p_1, \ldots, p_m, q_1, \ldots, q_n\}. \) Thus \( p \) divides either \( a \) or \( b. \)

If the result fails we can pick a smallest pair \( v \) and \( w \) for which the result fails. Write \( vw = z^2. \)

Write \( v = p_1^{i_1} \cdots p_m^{i_m} \) and \( w = q_1^{j_1} \cdots q_n^{j_n} \) as products of positive powers of primes. By hypothesis, \( p_1 \) does not divide \( w \) so is not equal to any of the \( q_i \)'s. But \( p_1 \) divides \( z^2 \), so by the previous observation, \( p_1 \) divides \( z \). Hence \( p_1^2 \) divides \( z^2 = vw \). So we can write \( vw = p_1^{2i_1} \cdots r_i^{k_i} \) for some primes \( r_i \).

But uniqueness of factorization says this is the same as the factorization \( vw = p_1^{i_1} \cdots p_m^{i_m} q_1^{j_1} \cdots q_n^{j_n} \), so we conclude that \( i_1 \geq 2 \), whence \( p_1^2 \) divides \( v \). This yields an equation \( (v/p_1^2)w = (z/p_1)^2 \) in integers in which \( v/p_1^2 \) is smaller than \( v \), contradicting our original choice of \( v \) and \( w \). We conclude that no such \( v \) and \( w \) can exist, so this proves the result. \( \square \)

A high point of the application of unique factorization is the classification of the integer solutions to the equation

\[
x^2 + y^2 = z^2. \tag{7-6}
\]

This equation, motivated by Pythagoras’s Theorem, was studied in antiquity, and complete solutions to it were independently found by several ancient cultures.

We will restrict our attention to positive integer solutions because all others can be obtained from these in an obvious way. First, observe that if each of \( x, y, \) and \( z \) is divisible by a number \( d \), then \( xd^{-1}, yd^{-1}, zd^{-1} \) is also a solution to the equation. Thus every solution is obtained from a primitive solution, that is one in which \( x, y, \) and \( z \) have no common factor. It therefore suffices to classify the primitive solutions. However, if \( d \) divides two of \( x, y, \) and \( z \), it must divide the third, so if \( x, y, z \) is a primitive solution, the greatest common divisor of any two of \( x, y, \) and \( z \) is one. Hence, at most one of \( x, y, \) and \( z \) is even, and at least two are odd. But if two of them are odd, the other must be even because a sum or difference of two odd numbers is even. However, if \( x \) and \( y \) are odd, then both \( x^2 \) and \( y^2 \) leave a remainder of one when divided by four, and \( z^2 \) must therefore leave a remainder of two when divided by four. But this is impossible, so we conclude that either \( x \) or \( y \) is even, and \( z \) is odd. We can assume without loss of generality that \( x \) is even and \( y \) is odd.

Now rewrite the equation as \( x^2 = (z+y)(z-y) \). Because \( x, y + z, \) and \( y - z \) are all even, there are integers \( u, v, \) and \( w, \) such that \( x = 2u, y + z = 2v, \) and \( y - z = 2w \). Hence \( u^2 = vw \).

I claim that \( v \) and \( w \) are relatively prime, because if \( p \) divided them both it would divide \( v + w = y \) and \( v - w = z \), which are relatively prime. Now, unique factorization implies that if a product \( vw \) of relatively prime numbers \( v \) and \( w \) is a square, then \( v \) and \( w \) must be squares. Hence there are integers \( a \) and \( b \) such that \( v = a^2 \) and \( w = b^2 \), and \( a \) and \( b \) must be relatively prime because \( v \) and \( w \) are. It follows that \( y = a^2 + b^2 \) and \( z = a^2 - b^2 \). Now,

\[
x^2 = (y + z)(y - z) = 4vw = 4a^2b^2,
\]
and \( x = 2ab. \)

Since we required \( x, y, \) and \( z, \) to be positive, \( a \) and \( b \) are positive and \( a > b. \)

We have therefore proved the following result.

**Theorem 7.2** A complete list of the positive primitive solutions to the equation
\[ x^2 + y^2 = z^2 \]
is given by
\[ x = 2ab, \quad y = a^2 + b^2, \quad z = a^2 - b^2, \]
where \( a \) and \( b \) are arbitrary positive integers with \( a > b. \)

**Exercises.**

1. Suppose that \( a_1, \ldots, a_r \) are pairwise relatively prime integers, i.e., \( \gcd(a_i, a_j) = 1 \) for all \( i \neq j. \) If \( b \) is an integer such that \( a_1 a_2 \cdots a_r = b^n, \) show that each \( a_i \) is an \( n^{th} \) power of an integer. This is an easy exercise, but notice that your proof depends on the fact that every integer can be written as a product of primes in a unique way.

2. Show that \( \mathbb{Z}[\sqrt{-3}] := \{ a + b\sqrt{-3} \mid a, b \in \mathbb{Z} \} \) is a ring; i.e., that sums and products of elements of this form are again of this form.

3. We call a non-zero element \( u \in \mathbb{Z}[\sqrt{-3}] \) a unit if \( u^{-1} \) belongs to \( \mathbb{Z}[\sqrt{-3}] \).

Find all the units in \( \mathbb{Z}[\sqrt{-3}] \).

4. Show that if \( a + b\sqrt{-3} \) divides both \( 1 + \sqrt{-3} \) and \( 1 - \sqrt{-3} \) in \( \mathbb{Z}[\sqrt{-3}], \)

then \( a + b\sqrt{-3} \) is a unit.

5. By the previous exercise, \( v = 1 + \sqrt{-3} \) and \( w = 1 - \sqrt{-3} \) are relatively prime in \( \mathbb{Z}[\sqrt{-3}], \) Show that \( vw \) is a square in \( \mathbb{Z}[\sqrt{-3}] \) despite the fact that neither \( v \) nor \( w \) is a square in \( \mathbb{Z}[\sqrt{-3}] \).

6. Show that the smallest subring of \( \mathbb{C} \) containing \( \mathbb{Z} \) and \( \frac{1}{2}(1 + \sqrt{-3}) \) is
\[
R = \left\{ a + \frac{b}{2}(1 + \sqrt{-3}) \mid a, b \in \mathbb{Z} \right\}.
\]

Is this the same as
\[ \left\{ \frac{1}{2}(a + b\sqrt{-3}) \mid a, b \in \mathbb{Z} \right\}? \]

7. Let \( \zeta_3 = e^{2\pi i/3}. \) Show that the smallest subring of \( \mathbb{C} \) containing \( \mathbb{Z} \) and \( \zeta_3 \) is the ring \( R \) in the previous exercise. We write \( \mathbb{Z}[\zeta_3] \) for this ring.

8. Show that every element in \( \mathbb{Z}[\zeta_3] \) satisfies a monic polynomial with coefficients in \( \mathbb{Z}. \)

9. Working in the ring
\[ \mathbb{Z}[\sqrt{-5}] = \{ a + b\sqrt{-5} \mid a, b \in \mathbb{Z} \} \]

show that
(a) neither $4 + 4\sqrt{-5}$ nor $9 - 9\sqrt{-5}$ is a cube;
(b) their product is a cube;
(c) if $u = a + b\sqrt{-5}$ divides both $4 + 4\sqrt{-5}$ and $9 - 9\sqrt{-5}$, then $\mathbb{Z}[\sqrt{-5}]$ contains $u^{-1}$.

1.8 Fermat’s Last Theorem

The initial impetus for the development of abstract algebra came from number theory, especially attempts to prove Fermat’s conjecture that if $p$ is an integer $\geq 3$ then there no non-zero integers $x, y, z$, such that

$$x^p + y^p = z^p. \quad (8.7)$$

The tale has been told many times. I won’t repeat it. The book Fermat’s Last Theorem by H.M. Edwards is an excellent historical account. The account below is taken from Edwards’s book.

In his papers Fermat left a proof that there are indeed no solutions when $p = 4$. It therefore suffices to establish his conjecture in the case when $p$ is a prime number, so we shall assume that $p$ is prime in our discussion from now on.

One observes easily that if there is a solution to (8.7) in which $x, y, z$ have a common factor, one may cancel that factor to obtain another solution with smaller $x, y, z$. Thus, if one wants to argue by contradiction, one can assume that one has a solution in which no two of $x, y,$ and $z$ has a common factor. Furthermore, one can assume that if there are solutions, then there is a “smallest” solution in an appropriate sense.

The proceedings of the meeting of the Paris Academy on March 1, 1847, serve to illustrate the forces driving the early development of abstract algebra. Lamé announced that he had a proof of Fermat’s conjecture. The starting point of his approach was the factorization

$$x^p + y^p = (x + y)(x + \zeta_p y) \cdots (x + \zeta_p^{p-1} y) \quad (8.8)$$

where $\zeta_p = e^{2\pi i/p}$. He planned to split the argument into two cases: if the factors $(x + \zeta_p y)$ are pairwise relatively prime, then the fact that their product is a $p^h$-power implies that each $(x + \zeta_p y)$ is a $p^h$-power; on the other hand, if the factors are not pairwise relatively prime, Lamé planned to show that they shared a common factor, and then divide through by that common factor obtain a smaller solution to (8.7).

Liouville objected to Lamé’s claim that the only way a product of relatively prime “numbers” could be a $p^h$-power was if each number was itself a $p^h$-power. In modern language, Lamé needed to prove that every number in the ring $\mathbb{Z}[\zeta_p]$, that is every number of the form

$$a_0 + a_1 \zeta_p + \cdots + a_{p-1} \zeta_p^{p-1} \quad (a_0, \ldots, a_{p-1} \in \mathbb{Z}),$$
1.8. **Fermat’s Last Theorem**

could be written as a product of primes in a unique way.

Liouville’s objection can be appreciated by a consideration of Euler’s “proof” for \( p = 3 \) that appeared in his 1770 book on algebra. By straightforward and solid arguments (see pages 40–41 of Edwards’s book), Euler shows that if there is a solution to \( x^3 + y^3 = z^3 \), then there exist relatively prime integers \( u \) and \( v \), one odd and one even, such that

\[
2u(u^2 + 3v^2) = \text{a cube.} \quad (8.9)
\]

Euler’s proof then breaks into two cases depending on whether or not 3 divides \( u \). Let’s consider the case where 3 does not divide \( u \). Because \( u \) and \( v \) are relatively prime and \( u^2 + 3v^2 \) is odd, it follows easily that \( 2u \) and \( u^2 + 3v^2 \) are relatively prime. Hence \( 2u \) and \( u^2 + 3v^2 \) are cubes. As Edwards explains on page 41 of his book, “one way to find cubes of the form \( u^2 + 3v^2 \) is to choose \( a, b \) at random and to set

\[
\begin{align*}
u &= a^3 - 9ab^2 \\
v &= 3a^2b - 3b^3
\end{align*}
\]

so that \( u^2 + 3v^2 = (a^2 + 3b^2)^3 \). The major gap [...] in Euler’s proof is [his claim] that this is the only way that \( u^2 + 3v^2 \) can be a cube”. The remainder of Euler’s proof is solid.

Euler tried to justify his claim about cubes of the form \( u^2 + 3v^2 \) by arguments involving numbers of the form

\[
a + b\sqrt{-3} \quad (a, b \in \mathbb{Z}).
\]

An exercise in the previous section showed that the set of these numbers is a ring, denoted \( \mathbb{Z}[\sqrt{-3}] \). Euler’s argument was based on the factorization

\[
u^2 + 3v^2 = (u + v\sqrt{-3})(u - v\sqrt{-3})
\]

in \( \mathbb{Z}[\sqrt{-3}] \). Euler noted that if one of these factors is a cube, say \( u + v\sqrt{-3} = (a + b\sqrt{-3})^3 \), then \( u = a^3 - 9ab^2 \) and \( v = 3a^2b - 3b^3 \). He also observed that \( u - v\sqrt{-3} = (a - b\sqrt{-3})^3 \) and so

\[
u^2 + 3v^2 = (u + v\sqrt{-3})(u - v\sqrt{-3})
\]

\[
= (a + b\sqrt{-3})(a - b\sqrt{-3})^3
\]

\[
= (a^2 + 3b^2)^3.
\]

That is, if \( u + v\sqrt{-3} = (a + b\sqrt{-3})^3 \), then \( u^2 + 3v^2 = (a^2 + 3b^2)^3 \) is a cube. Euler’s error was to take this sufficient condition for \( u^2 + 3v^2 \) to be a cube and treat it as if it were a necessary condition. One finds in Euler’s book the statement that “[if \( x \) and \( y \) are relatively prime integers and] \( x^2 + cy^2 \) is [...] a cube, one can certainly conclude that [...] \( x + y\sqrt{-c} \) and \( x - y\sqrt{-c} \) must be cubes, because they are relatively prime in that \( x \) and \( y \) have no common factor.” In this generality, Euler’s statement is false—
Edwards’s speculates that when Euler wrote to Goldbach in 1753 that he had proves the \( p = 3 \) case of Fermat’s last theorem, he had in mind an argument that did not involve the factorization in \( \mathbb{Z}[(\sqrt{-3})] \).

One can imagine that the proof Fermat had in mind when he made his historical marginal note was based on the factorization (8·5). In any case, this factorization was used by Lagrange, by Euler in proving Fermat’s theorem for \( n = 3 \), by Gauss for \( n = 5 \), and by Dirichlet for \( n = 14 \). In his proof for \( n = 3 \), Euler assumed that \( \mathbb{Z}[(\sqrt{-3})] \) is a UFD; since \( \mathbb{Z}[(\sqrt{-3})] \) is a UFD, Euler’s proof is correct. It can be shown that Fermat’s Theorem holds if \( \mathbb{Z}[(\sqrt{-3})] \) is a UFD; unfortunately (or fortunately, depending on your point of view) it is not a UFD for all values of \( n \). Kummer was the first to realize this, and he developed the theory of ideals (= ideal numbers) to recover this lack of unique factorization: every ideal in \( \mathbb{Z}[(\sqrt{-3})] \) can be written as a product of prime ideals in a unique way.

**Exercise.** Is \( \mathbb{Z}[(\sqrt{-3})]/(x^n - 1) \)? If not, what is the relation between these rings?

**Exercise.** The failure of unique factorization in \( \mathbb{Z}[(\sqrt{-3})] \) can be repaired in some sense. The *ideal* generated by 6 can be written in a unique way as a product of prime ideals:

\[
(6) = (2, 1 + \sqrt{-3}), (3, 1 + \sqrt{-3}), (3, 1 - \sqrt{-3}).
\]

To see that \( (2, 1+\sqrt{-3}) \), \( (3, 1+\sqrt{-3}) \), and \( (3, 1-\sqrt{-3}) \) are prime ideals compute the quotient ring and check that it is a domain.

1.9 Domains and fields

A ring \( R \) is a domain, or an integral domain, if every pair of non-zero elements in \( R \) has a non-zero product. Thus, in a domain, a product \( xy \) can only be zero if either \( x \) or \( y \) is zero. We can therefore cancel in a domain: if \( ax = ay \) and \( a \neq 0 \), then \( x = y \) because \( a(x - y) = 0 \).

The ring of integers is a domain: a product of non-zero integers is non-zero. A field is a domain because if \( x \) is a non-zero element and \( xy = 0 \), then \( 0 = x^{-1}0 = x^{-1}xy = 1 \cdot y = y \). Every subring of a field is a domain. We will soon show that every domain is a subring of a field.

It is easy to show that \( \mathbb{Z}/(a) \) is a domain if and only if \( a \) is a prime number or zero. For example, \( \mathbb{Z}/(6) \) is not a domain because \( [2+6][3+6] = [6+6] = 0 \).

A simple geometric example of a commutative ring that is not a domain is provided by the ring of \( k \)-valued functions on a space \( X \) that has more than one element: if \( x \) and \( y \) are different points of \( X \), then the product of the non-zero functions \( f \) and \( g \), defined by \( f(x) = g(y) = 1 \) and \( f(X\{x\}) = g(X\{y\}) = 1 \), is zero. Another geometric example occurs for the ring of continuous \( \mathbb{R} \)-valued functions on the topological space \( X \subset \mathbb{R}^2 \) that is the union of the usual \( x \)- and \( y \)-axes; the functions \( f \) and \( g \) that take, respectively, the \( x \)- and \( y \)-coordinates of a point \( p \in X \) are both non-zero, but their product is zero. This last example is a baby example of the general fact that the coordinate ring of an affine algebraic variety is a domain if and only if the variety is irreducible.
1.9. DOMAINS AND FIELDS

The next two exercises show that under appropriate finiteness conditions a domain must in fact be a field. These exercises can be considered as warm ups for Proposition 12.6 which gives another finiteness condition that ensures a domain is a field.

Exercise. Show that a finite commutative domain is a field.

Exercise. Let \( R \) be a commutative domain containing a field \( k \). Show that \( R \) is a field if \( \dim_k R < \infty \).

As a consequence of the previous exercise, the rings \( \mathbb{Q}[\sqrt{d}] \) are actually fields. More generally, if \( R \) is any subring of \( \mathbb{C} \) containing \( \mathbb{Q} \), \( R \) is a field if \( \dim_{\mathbb{Q}} R < \infty \).

Fields of fractions.
Just as the field of rational numbers can be constructed from the ring of integers, so too does every commutative domain \( R \) have a field of fractions, denoted \( \text{Frac} \ R \), the elements of which can be written as fractions \( a/b \) with \( a, b \in R \) and \( b \neq 0 \). You probably don’t need much persuasion to believe this, but the formal construction of \( \text{Frac} \ R \) is as follows.

Let \( R \) be a domain. Define an equivalence relation on the cartesian product \( R \times R \setminus \{0\} \) as follows:

\[
(a, b) \sim (c, d) \quad \text{if} \quad ad = cb.
\]

Check this is an equivalence relation. We denote the equivalence class of \( (a, b) \) by \( a/b \). We now impose a ring structure on the set of equivalence classes by defining addition by

\[
(a/b) + (c/d) = (ad + bc)/bd,
\]

and multiplication by

\[
(a/b) \cdot (c/d) = ac/bd.
\]

Check that these two binary operations are well-defined. Check that under +, the equivalence classes form an abelian group with zero element 0/1. Check that 1/1 is an identity element for the multiplication. Check that the equivalence classes form a ring with identity. We denote this ring by \( \text{Frac} \ R \) and call it the field of fractions of \( R \).

Check that there is an injective homomorphism \( \rho : R \to \text{Frac} \ R \) defined by \( r \mapsto r/1 \). We usually identify \( R \) with its image in \( \text{Frac} \ R \) under this map, and think of \( R \) as a subring of \( \text{Frac} \ R \). Each non-zero element \( b \in R \) has an inverse in \( \text{Frac} \ R \), namely \( 1/b \). We often write \( b^{-1} \) for \( 1/b \).

Exercise. Let \( R \) be a commutative domain, and \( \varphi : R \to S \) a ring homomorphism such that \( \varphi(b) \) is a unit in \( S \) for every non-zero element \( b \in R \). Show there is a unique ring homomorphism \( \psi : \text{Frac} \ R \to S \) such that \( \varphi = \psi \rho \), where \( \rho : R \to \text{Frac} \ R \) is the map in the previous paragraph.

It is often useful when dealing with a domain \( R \) to consider rings between \( R \) and \( \text{Frac} \ R \) that are obtained by inverting only some of the non-zero elements in \( R \). We now consider this matter.
If $S$ is a subset of $R$ that does not contain zero and $R'$ is a ring lying between $R$ and Fract $R$ in which every element of $S$ is a unit, then every product of elements from $S$ is also a unit in $R'$ because $(st)^{-1} = s^{-1}t^{-1}$. It therefore makes sense to assume that $S$ is closed under multiplication; such a subset of $R$ is said to be multiplicatively closed.

**Proposition 9.1** Let $R$ be a commutative domain and $S$ a multiplicatively closed subset of $R$ that does not contain 0. Then

$$R_S := R[S^{-1}] := \{as^{-1} \mid a \in R, s \in S\}$$

is a subring of Fract $R$ containing $R$. Moreover,

1. $R[S^{-1}]$ is the smallest subring of Fract $R$ containing $R$ in which every element of $S$ is a unit;

2. every ideal of $R_S$ is of the form $IR_S$ for some ideal $I$ in $R$;

3. if $J$ is an ideal of $R_S$, then $J = (J \cap R)R_S$;

4. if every ideal of $R$ is finitely generated, so is every ideal of $R_S$.

**Proof.** To see that $R_S$ is a ring, we need to check that it is closed under products and sums. If $x, y \in R_S$, we can write $x = as^{-1}$ and $y = bt^{-1}$ for some $a, b \in R$ and $s, t \in S$. It follows that $xy = ab(st)^{-1}$ and $x + y = (at + bs)(st)^{-1}$. Since $st \in S$, $xy$ and $x + y$ belong to $R_S$.

(1) If $s \in S$, then $s^{-1} = 1s^{-1}$ belongs to $R_S$, so every element of $S$ is a unit in $R_S$. On the other hand, if $R'$ is a ring lying between $R$ and Fract $R$ and contains $s^{-1}$ for each $s \in S$, then $R'$ contains $as^{-1}$ for every $a \in R$, so contains $R_S$.

(2) and (3). Let $J$ be an ideal of $R_S$. Then $I := R \cap J$ is an ideal of $R$, so (2) follows from (3). Since $J$ contains $I$, it contains $IR_S$. To prove the converse, suppose that $x \in J$. Then $x = as^{-1}$ for some $a \in R$ and $s \in S$. Since $J$ is an ideal it contains $xs = a$. Thus $a \in J \cap R = I$, and $as^{-1} \in IR_S$.

(4) If $J$ is an ideal of $R_S$, then $J \cap R$ is a finitely generated ideal of $R$ by hypothesis, so $J = (J \cap R)R_S$ is generated as an ideal of $R_S$ by that same finite set of generators. \qed

For example, the field of fractions of $\mathbb{Z}[\sqrt{d}]$ is $\mathbb{Q}(\sqrt{d})$.

### 1.10 Unique factorization domains

**Definition 10.1** A commutative domain $R$ is a unique factorization domain, or UFD, if every element of $R$ can be written uniquely as a product of irreducible elements, and the irreducibles that occur in the factorization are unique up to order and multiplication by units. \diamond
To see what “uniqueness” means in this definition, consider the factorizations
\[ 6 = 2 \cdot 3 = (-3) \cdot (-2) = (-3) \cdot (-1) \cdot (2) \cdot (-1) \cdot (-1) \]
in \( \mathbb{Z} \). The uniqueness means this: if we have two factorizations of an element
as a product of irreducibles, and \( x \) is an irreducible appearing in one of those
factorizations, then some unit multiple of \( x \) must appear in the other factorization.

**Lemma 10.2** In a unique factorization domain, primes and irreducibles are
the same.

**Proof.** We observed on page 5 that a prime is irreducible.

Suppose that \( x \) is an irreducible and that \( x \mid bc \). Then \( bc = xy \) for some \( y \).
We can write each of \( b, c \), and \( y \), as a product of irreducibles. Doing so gives
two factorizations of \( bc \) as a product of irreducibles. Buy the uniqueness of such
a factorization, at least one of the irreducibles in the factorizations of \( b \) and \( c \)
must be a unit multiple of \( x \). But that implies that \( x \) divides either \( b \) or \( c \), thus
showing that \( x \) is prime. \( \square \)

This puts into perspective the non-unique factorization
\[ 6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}) \]
in the ring \( \mathbb{Z} \sqrt{-5} \).

The proof that \( k[x_1, \cdots, x_n] \) is a unique factorization domain proceeds by
induction on the number of variables. We will show that \( R[x] \) is a unique
factorization domain if \( R \) is.

**Definition 10.3** Let \( R \) be a unique factorization domain, and \( R[x] \) the poly-
nomial ring over it. The content of \( f = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n \in R[x] \)
is
\[ c(f) := \gcd(\alpha_0, \ldots, \alpha_n). \]
This is only well-defined up to a unit multiple, so when we write \( c(f) = c(g) \) we
will mean that \( c(f) \) is a unit multiple of \( c(g) \).

We call \( f \) a primitive polynomial if \( c(f) \) is a unit.

**Remark.** If \( f \in R[x] \), then \( c(f)^{-1} f \) is primitive so every polynomial is a
scalar multiple of a primitive polynomial. Thus, every irreducible polynomial
in \( R[x] \) is primitive.

**Lemma 10.4 (Gauss)** Let \( R \) be a unique factorization domain. If \( f, g \in R[x] \),
then
\[ 1. \ c(fg) = uc(f)c(g) \text{ for some unit } u, \text{ and} \]
\[ 2. \ a \text{ product of primitive polynomials is primitive.} \]
CHAPTER 1. ORIGINS OF MODERN ALGEBRA

Proof. By the above remark, we can write \( f = a f_1 \) and \( g = b g_1 \) where \( a, b \in R \) and \( f_1, g_1 \in R[x] \) are primitive. It suffices to prove that \( f_1 g_1 \) is primitive. Hence (1) follows from (2).

To prove (2), let \( f = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n \) and \( g = \beta_0 + \beta_1 x + \cdots + \beta_m x^m \) be primitive. Suppose, contrary to what we wish to show, that \( fg \) is not primitive. Let \( p \in R \) be an irreducible divisor that divides all the coefficients of \( fg \). Since \( f \) and \( g \) are primitive we can choose \( s \) and \( t \) minimal such that \( \alpha_s \) and \( \beta_t \) are not divisible by \( p \). Since \( p \) is prime, it does not divide \( \alpha_s \beta_t \). The coefficient of \( x^{s+t} \) in \( fg \) is

\[
\cdots + \alpha_{s-1} \beta_{t+1} + \alpha_s \beta_t + \alpha_{s+1} \beta_{t-1} + \cdots.
\]

Since \( p \) divides all the terms in this sum except \( \alpha_s \beta_t \), it does not divide the coefficient of \( x^{s+t} \). This is a contradiction. \( \square \)

Lemma 10.5 Let \( R \) be a unique factorization domain, and let \( k = \text{Fract } R \).

1. A polynomial in \( R[x] \) is divisible by a polynomial of degree \( d \) in \( k[x] \) if and only if it is divisible by a polynomial of degree \( d \) in \( R[x] \).

2. A primitive element of \( R[x] \) is irreducible in \( R[x] \) if and only if it is irreducible in \( k[x] \).

Proof. (1) (\( \Rightarrow \)) This is obvious.

(\( \Rightarrow \)) Let \( f \in R[x] \) and suppose that \( f = gh \) is a factorization in \( k[x] \) with \( \deg g = d \). Write \( g = \sum_{i=0}^{n} \alpha_i \beta_i^{-1} x^i \) where all \( \alpha_i \) and \( \beta_i \) are in \( R \). Let \( \beta \) be the product of all the \( \beta_i \)'s, and set \( \gamma_i = \beta \alpha_i \beta_i^{-1} \). Thus \( g = \beta^{-1} \sum_{i=0}^{n} \gamma_i x^i \) where each \( \gamma_i \in R \). Since \( \beta g \in R[x] \), we can write it as \( \beta g = c(\beta g) \) where \( g_1 \in R[x] \) is primitive. Hence \( g = \beta^{-1} c(\beta g) g_1 \). In a similar way, we can write \( h = \delta^{-1} c(\delta h) h_1 \) where \( \delta \in R, \delta h \in R[x] \), and \( h_1 \in R[x] \) is primitive.

Because \( \beta \delta f = \beta \delta gh = c(\beta g) c(\delta h) g_1 h_1 \) belongs to \( R[x] \), we can take the content of both sides and apply Gauss’s Lemma to conclude that \( \beta \delta = c(\beta g) c(\delta h) \), whence \( f = g_1 h_1 \). This is a factorization of \( f \) in \( R[x] \) and \( \deg g_1 = \deg g \).

(2) (\( \Leftarrow \)) If \( f \in R[x] \) factors as \( f = gh \) in \( R[x] \), then that factorization is also a factorization in \( k[x] \), so either \( g \) or \( h \) is a unit in \( k[x] \). If \( g \in k \), then \( g \in k \cap R[x] = R \), so \( g \) divides all the coefficients of \( f = gh \). Since \( f \) is primitive in \( R[x] \) it follows that \( g \) is a unit. Hence \( f \) is irreducible in \( R[x] \).

(\( \Rightarrow \)) Let \( f \in R[x] \) be primitive, and irreducible in \( R[x] \). Suppose that \( f = gh \) where \( g, h \in k[x] \). We must show that the degree of either \( g \) or \( h \) is zero. By (1) we can also write \( f = g_1 h_1 \) where \( g_1, h_1 \in R[x] \) and \( \deg g_1 = \deg g \) and \( \deg h_1 = \deg h \). But \( f \) is irreducible in \( R[x] \) so either \( g_1 \) or \( h_1 \) is a unit in \( R[x] \). Suppose \( g_1 \) is a unit in \( R[x] \). Then \( g_1 \in R \subset k \), and hence \( g \in k \) too, showing that \( f \) is irreducible in \( k[x] \).

\( \square \)

Theorem 10.6 If \( R \) is a unique factorization domain, so is \( R[x] \).

Proof. Let \( k = \text{Fract } R \).
1.11. **Principal ideal domains**

Let \( f \in R[x] \). We first show that \( f \) is a product of irreducibles. Write \( f = \alpha f_1 \) with \( \alpha \in R \) and \( f_1 \in R[x] \) primitive. By hypothesis, \( \alpha \) is a product of irreducibles in \( R \); these irreducibles remain irreducible in \( R[x] \), so it suffices to show that \( f_1 \) is a product of irreducibles. Replacing \( f \) by \( f_1 \), we can therefore assume that \( f \) is primitive.

Since \( f \) belongs to \( k[x] \), \( f = g_1 \cdots g_n \) where each \( g_i \) is an irreducible in \( k[x] \). By the proof of the previous lemma, we can write \( g_i = \beta_i^{-1} \gamma_i h_i \), where \( \beta_i, \gamma_i \in R \) and \( h_i \in R[x] \) is primitive. Since \( g_i \) is irreducible in \( k[x] \), so is \( h_i \). By the previous lemma, \( h_i \) is therefore irreducible in \( R[x] \). Taking the content of both sides of \( \beta_1 \cdots \beta_n f = \gamma_1 \cdots \gamma_n h_1 \cdots h_n \) gives \( \beta_1 \cdots \beta_n = \gamma_1 \cdots \gamma_n \), whence \( f = h_1 \cdots h_n \). This expresses \( f \) as a product of irreducibles in \( R[x] \).

To show that \( R[x] \) is a unique factorization domain, it suffices to show that if an irreducible \( p \in R[x] \) divides a product \( ab \), then \( p \) divides either \( a \) or \( b \). Since \( p \) is irreducible in \( R[x] \) it is primitive, and hence irreducible in \( k[x] \). Since \( k[x] \) is a unique factorization domain, \( p \) divides either \( a \) or \( b \) in \( k[x] \). Without loss of generality, we may assume that \( b \) divides \( b \). Hence \( b = pd \) with \( c \in k[x] \). It now suffices to show that \( d \in R[x] \). Write \( b = c(b)b_1 \) with \( b_1 \in R[x] \) primitive. By the proof of the previous lemma, we may write \( d = \alpha \beta^{-1} d_1 \) with \( \alpha, \beta \in R \) and \( d_1 \in R[x] \) primitive. Taking the content of both sides of \( \beta c(b)b_1 = \alpha p d_1 \) gives \( \beta c(b) = \alpha d_1 \), so \( d = c(b)d_1 \) belongs to \( R[x] \).

**Theorem 10.7** Let \( R \) be a unique factorization domain and \( k = \text{Fract } R \). Let \( f \in R[x] \). If \( f = gh \) with \( g, h \in R[x] \), then there exist \( G, H \in R[x] \) such that \( f = GH \) and \( \deg G = \deg g \) and \( \deg H = \deg h \).

**Proof.** First suppose that \( f \) is primitive. \( \square \)

**Example 10.8** The ring \( R = k[t, t^{1/2}, t^{1/4}, \cdots] \) is a domain in which prime and irreducible elements are the same but it is not a UFD. It fails to be a UFD because some elements, \( t \) for example, cannot be written as a product of irreducibles. To see that every irreducible is prime, suppose that \( x \) is irreducible and that \( x \mid yz \). There is a suitably large \( n \) such that \( x, y, \) and \( z \), all belong to \( k[t, t^{1/2}, \cdots, t^{1/n}] \); this subring is equal to \( k[t^{1/2n}] \) which is a polynomial ring in one variable (so a UFD); since \( x \) is still irreducible as an element of \( k[t^{1/2n}] \) it is prime in \( k[t^{1/2n}] \), so must divide either \( y \) or \( z \); hence \( x \) is prime in \( R \). Notice that \( R \) is not noetherian: the chain \((t^{1/2}) \subset (t^{1/4}) \subset \cdots \) does not stabilize. \( \diamond \)

1.11 **Principal ideal domains**

Recall that every ideal in \( \mathbb{Z} \) is of the form \((d)\) for some \( d \). Similarly, every ideal in \( k[x] \) is of the form \((f)\) (Theorem 4.6).

An ideal of the form \((r)\) in a ring \( R \) is said to be principal.

**Definition 11.1** A principal ideal domain is a domain in which every ideal is principal, i.e., every ideal consists of multiples of a single element. \( \diamond \)
Using the Euclidean algorithm is the standard method to show that a ring is a principal ideal domain. The argument in Theorem 4.6 is typical.

**Principal ideal domains.** Principal ideal domains abound. They are of great importance in both number theory and algebraic geometry. Later we will study them in detail.

Number theorists are interested in finite extension fields of \( \mathbb{Q} \). By definition, such a field \( k \) is a subfield of \( \mathbb{C} \) that is obtained by adjoining to \( \mathbb{Q} \) the zeroes of a polynomial in \( \mathbb{Q}[x] \). One then has the notion of the subring of integers in \( k \), by that one means the elements of \( k \) that are a zero of a *monic* polynomial in \( \mathbb{Z}[x] \). It is remarkable that such elements form a ring. If you don’t believe this try to see why \( \alpha + \beta \) and \( \alpha \beta \) are zeroes of monic polynomials in \( \mathbb{Z}[x] \) given that \( \alpha \) and \( \beta \) are. It is an important question to decide when such a ring of integers is a principal ideal domain. Here is an easy example. Adjoining to \( \mathbb{Q} \) the zeroes of \( x^2 + 1 \) gives the field \( \mathbb{Q}(i) = \{a + bi \mid a, b \in \mathbb{Q}\} \). The ring of integers in this is \( \mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\} \). To show that \( \mathbb{Z}[i] \) is a principal ideal domain, one uses a version of the Euclidean algorithm. They key point is to introduce a notion of “size” that allows us to prove an analogue of Proposition 4.3 saying that when we divide we can obtain a remainder that is smaller than the number we are dividing by.

Determining the ring of integers is more subtle than the example of \( \mathbb{Z}[i] \) suggests. For example, the ring of integers in \( \mathbb{Q}(\sqrt{5}) \) is \( \mathbb{Z}[\frac{1}{2}(1 + \sqrt{5})] \). Is this a principal ideal domain?

The polynomial ring in two variables is not a principal ideal domain. The ideal \( (x^n, x^{n-1}y, \ldots, xy^{n-1}, y^n) \) can not be generated by less than \( n+1 \) elements. Try to prove this. If it proves difficult, start with the case \( (x, y) \). You should also pay attention to the fact that the \( n+1 \) generators I listed are all homogeneous, so \( I = \oplus_{d=n} I \cap k[x,y]d \).

**Proposition 11.2** Let \( R \) be a principal ideal domain. Then

1. greatest common divisors exist in \( R \);
2. if \( d = \gcd(a, b) \), then \( d = ax + by \) for some \( x, y \in R \);
3. every irreducible in \( R \) is prime.

**Proof.** (1) The ideal \( aR + bR \) is principal, so is equal to \( dR \) for some \( d \in R \). Clearly, \( d = ax + by \) for some \( x, y \in R \), so it remains to show that \( d \) is a greatest common divisor of \( a \) and \( b \). First, since \( a \) and \( b \) belong to \( dR \), they are both divisible by \( d \). Second, if \( e \) divides both \( a \) and \( b \), then \( aR + bR \) is contained in \( eR \), so \( d \) is a multiple of \( e \). Hence \( d \) is a greatest common divisor of \( a \) and \( b \).

(2) Let \( a \) be irreducible, and suppose that \( a | bc \). To show that \( a \) is prime, we must show it divides either \( b \) or \( c \). Suppose \( a \) does not divide \( b \). Let \( d = ax + by = \gcd(a, b) \). Since \( d \) divides \( a \), either \( d \) is a unit or \( a = du \) with \( u \) a unit. But \( d | b \), so the second alternative implies that \( b | d \). Hence \( d \) must be a unit. Since \( a \) divides \( bc \), it therefore divides \( acd^{-1} + bcyd^{-1} = c(ax + by)d^{-1} = c \). Hence \( a \) is prime.
1.11. PRINCIPAL IDEAL DOMAINS

Theorem 11.3 Every principal ideal domain is a unique factorization domain.

Proof. Let \( R \) be a PID and \( a \) a non-zero non-unit in \( R \). We must show that \( a \) is a product of irreducibles in a unique way.

Uniqueness. Suppose that \( a = a_1 \cdots a_m = b_1 \cdots b_n \) and that each \( a_i \) and \( b_j \) is irreducible. Without loss of generality we can assume that \( m \leq n \). If \( m = 1 \), then we would be done. By Proposition 11.2, \( a_1 \) divides some \( b_j \); relabel the \( b_j \)s so that \( a_1 | b_1 \). Since \( a_1 \) and \( b_1 \) are irreducible, \( b_1 = a_1 u \) for some unit \( u \). Thus \( a_2 \cdots a_m = (ub_2) \cdots b_n \). If \( m = 1 \), we would have \( 1 = (ub_2) \cdots b_n \) so \( n \) would have to be one also, and we would be finished. However, if \( m > 1 \) and by an induction argument we can reduce to the case \( m = 1 \).

Existence. Suppose to the contrary that \( a \) is not a product of irreducibles. Then \( a \) is not irreducible, so \( a = a_1 b_1 \) with \( a_1 \) and \( b_1 \) non-units. Since \( a \) is not a product of irreducibles, at least one of \( a_1 \) and \( b_1 \) is not a product of irreducibles. Relabelling if necessary, we can assume that \( a_1 \) is not a product of irreducibles. Thus \( a_1 \) is not irreducible, and we may write \( a_1 = a_2 b_2 \) with \( a_2 \) and \( b_2 \) non-units.

Continuing in this way, we obtain a sequence \( a_1, a_2, \ldots \) of irreducible elements, and factorizations \( a_i = a_{i+1} b_{i+1} \) into a product of non-units. This yields a chain

\[
Ra \subset Ra_1 \subset Ra_2 \subset \cdots
\]

of ideals. The union of an ascending chain of ideals is an ideal of \( R \), and it is a principal ideal, say \( Rz \), by hypothesis. Now \( z \) must belong to some \( Ra_i \), but then \( Rz \subset Ra_i \subset Ra_i + 1 \subset Rz \), so these ideals are equal. In particular, \( a_{i+1} \in Ra_i \), so \( a_{i+1} = a_i u \). It follows that \( a_i = a_{i+1} b_{i+1} = a_i u b_{i+1} \), whence \( b_{i+1} \) is a unit. This is a contradiction.

We conclude that \( a \) must be a product of irreducibles. \( \square \)

Proposition 11.4 Let \( f \) be an element in a principal ideal domain \( R \). The following are equivalent:

1. \( f \) is irreducible;
2. \( (f) \) is a maximal ideal;
3. \( R/(f) \) is a field;
4. \( f \) is a prime.

Proof. Lemma 4.10 shows that conditions (2) and (3) are equivalent. Theorem 11.3 and Lemma 10.2 shows that conditions (2) and (4) are equivalent.

(1) \( \Rightarrow \) (2). Suppose \( J \) is an ideal of \( R \) that contains \( (f) \). By hypothesis, \( J \) is principal, say \( J = (g) \). Thus \( f = gh \) for some \( h \in R \). Since \( f \) is irreducible either \( g \) is a unit, in which case \( J = R \), or \( h \) is a unit, in which case \( g = fh^{-1} \) and \( (g) = (f) \).

(2) \( \Rightarrow \) (1). Suppose that \( f = gh \). Then \( (f) \subset (g) \) so either \( (g) = R \), in which case \( g \) is a unit, or \( (g) = (f) \), in which case \( g = fv \) for some \( v \in R \) and \( hv = 1 \) so \( h \) is a unit. Thus \( f \) is irreducible. \( \square \)
1.12 Integrality

Let $R \subset S$ be commutative rings. We say that $a \in S$ is integral over $R$ if it satisfies a monic polynomial with coefficients in $R$; that is, if there are elements $\lambda_0, \ldots, \lambda_{n-1}$ in $R$ such that

$$a^n + \lambda_{n-1}a^{n-1} + \cdots + \lambda_1 a + \lambda_0 = 0.$$ 

It is clear that every element of $R$ is integral over $R$, because $a \in R$ is a zero of the monic polynomial $x - a \in R[x]$.

If $d \in \mathbb{Z}$, then all $n$th roots of $d$ are integral over $\mathbb{Z}$ because they are zeroes of the monic polynomials $x^n - d$.

If the only elements of $S$ that are integral over $R$ are the elements of $R$, we say that $R$ is integrally closed in $S$. A domain that is integrally closed in its field of fractions is sometimes said to be integrally closed.

The integral closure of $R$ in $S$ is the set of all elements of $S$ that are integral over $R$. We sometimes write $\bar{R}$ for this integral closure even though the notation does not indicate its dependence on $S$. Proposition 12.5 shows that $\bar{R}$ is a ring. To see that this is really not obvious, try to show that $\sqrt{2} + \sqrt{3} + \sqrt{5}$ is integral over $\mathbb{Z}$.

For example, $\mathbb{Z}$ is integrally closed in $\mathbb{Q}$. To see this, suppose that $q \in \mathbb{Q}$ satisfies

$$q^n + \lambda_{n-1}q^{n-1} + \cdots + \lambda_1 q + \lambda_0 = 0$$

with all $\lambda_i$ in $\mathbb{Z}$. Write $q = a/b$ with $a, b \in \mathbb{Z}$. We can assume that $a$ and $b$ have no common factor. Since

$$a^n + \lambda_{n-1}a^{n-1}b + \cdots + \lambda_1 ab^{n-1} + \lambda_0 b^n = 0,$$

every prime dividing $b$ must also divide $a^n$ and hence $a$. Since $a$ and $b$ have no common prime factor, we conclude that $b = \pm 1$, whence $q \in \mathbb{Z}$.

This argument depends only on the fact that $\mathbb{Z}$ is a unique factorization domain. Hence we have the next result.

Proposition 12.1 A unique factorization domain is integrally closed.

In particular, the polynomial ring $k[t]$ is integrally closed. In contrast, its subring $k[t^2, t^3]$ is not: for example, $t$ is a zero of the monic polynomial $x^2 - t^2$ with coefficients in $k[t^2, t^3]$. The next result shows that every element of $k[t]$ is integral over $k[t^2, t^3]$. Since Fract $k[t^2, t^3]$ contains $t^{-1} = t^2(t^3)^{-1}$, it follows that Fract $k[t^2, t^3] = k(t)$. Hence $k[t]$ is the integral closure of $k[t^2, t^3]$.

Proposition 12.2 Let $R \subset S$ be commutative rings, and $a \in S$. The following are equivalent:

1. $a$ is integral over $R$;
2. $R[a]$ is a finitely generated $R$-module;
3. there is a subring $S'$ of $S$ such that $R[a] \subset S' \subset S$ and $S'$ is a finitely generated $R$-module.

Proof. (1) \implies (2) If $a^n + \lambda_{n-1}a^{n-1} + \cdots + \lambda_1 a + \lambda_0 = 0$, then $R[a] = R + Ra + \cdots + Ra^{n-1}$.

(2) \implies (3) Take $S' = R[a]$.

(3) \implies (1) Write $S' = R{s_1} + \cdots + R{s_n}$ where $s_1 = 1$. For each $i$, $as_i \in S'$, so $as_i = \sum_j \lambda_{ij} s_j$ for some $\lambda_{ij}$ in $R$. Rewrite this equation as $\sum_{j=1}^n (a\delta_{ij} - \lambda_{ij}) s_j = 0$. Let $M$ be the $n \times n$ matrix with $ij$th entry $a\delta_{ij} - \lambda_{ij}$; set $\Delta = \det M$ and write $\mathbf{s}$ for the column vector $(s_1, \ldots, s_n)^T$. Thus $M\mathbf{s} = 0$, and

$$0 = (M^{\text{adj}})M\mathbf{s} = \Delta\mathbf{s}$$

where $M^{\text{adj}}$ is the adjoint matrix. Hence $\Delta s_i = 0$ for all $i$; in particular, $0 = \Delta s_1 = \Delta = \det(a\delta_{ij} - \lambda_{ij})$. Writing out this determinant explicitly gives a monic polynomial of degree $n$ in $a$ with coefficients in $R$. Hence $a$ is integral over $R$.

\[\square\]

Corollary 12.3 Let $R \subset S$ be commutative rings and $a_1, \ldots, a_n$ elements of $S$. If all the $a_i$s are integral over $R$, then $R[a_1, \ldots, a_n]$ is a finitely generated $R$-module.

Proof. We argue by induction on $n$, the case $n = 1$ being given by Proposition 12.2. The induction hypothesis is that $R[a_1, \ldots, a_{n-1}]$ is a finitely generated $R$-module, say equal to $Rs_1 + \cdots + Rs_m$. Since $a_n$ is integral over $R$, it is integral over $R[a_1, \ldots, a_{n-1}]$, so Proposition 12.2 shows that $R[a_1, \ldots, a_n]$ is a finitely generated $R[a_1, \ldots, a_{n-1}]$-module, say

$$R[a_1, \ldots, a_n] = R[a_1, \ldots, a_{n-1}]t_1 + \cdots + R[a_1, \ldots, a_{n-1}]t_k.$$ 

It follows that $R[a_1, \ldots, a_n] = \sum_{i=1}^m \sum_{j=1}^k Rs_it_j$. \[\square\]

Let $R \subset S$ be commutative rings. We say that $S$ is integral over $R$ if every element of $S$ is integral over $R$.

Corollary 12.4 Let $R \subset S$ be commutative rings. If $S$ is a finitely generated $R$-algebra, then $S$ is integral over $R$ if and only if it is a finitely generated $R$-module.

Proposition 12.5 Let $R \subset S$ be commutative rings. The integral closure of $R$ in $S$ is a subring of $S$.

Proof. Write $\bar{R}$ for the integral closure of $R$ in $S$. Thus,

$$\bar{R} = \{ a \in S \mid a \text{ is integral over } R \}.$$ 

Obviously, $R \subset \bar{R}$. We must show that if $a, b \in \bar{R}$, then $a \pm b$ and $ab$ are in $\bar{R}$. By Proposition 12.2, $R[a]$ and $R[b]$ are finitely generated $R$-modules, say
$R[a] = Ra_1 + \cdots + Ra_m$ and $R[b] = Rb_1 + \cdots + Rb_n$. We can assume that $a_1 = b_1 = 1$. This ensures that the finitely generated $R$-module

$$S' := \sum_{i=1}^{m} \sum_{j=1}^{n} Ra_i b_j$$

is in fact a ring. Both $R[a]$ and $R[b]$ are subrings of $S'$, so $a \pm b$ and $ab$ belong to $S'$. By Proposition 12.2, $a \pm b$ and $ab$ are integral over $R$. □

The next result should be compared with the two exercises in Section 1.9 which showed that suitable finiteness conditions imply that a domain must be a field.

**Proposition 12.6** Let $T$ be a domain and $R$ a subring of $T$ such that $T$ is integral over $R$. Then $R$ is a field if and only if $T$ is.

**Proof.** ($\Rightarrow$) Let $a$ be a non-zero element of $T$. Let $n$ be minimal such that

$$a^n + \lambda_{n-1} a^{n-1} + \cdots + \lambda_1 a + \lambda_0 = 0$$

for some elements $\lambda_{n-1}, \ldots, \lambda_0 \in R$. We must have $\lambda_0 \neq 0$ because if it were not, we could cancel a common factor of $a$ and so reduce the minimal $n$. But now $\lambda_0$ is a unit in $R$, so

$$a(-\lambda_{n-1} a^{n-1} + \cdots - \lambda_1)\lambda_0^{-1} = 1,$$

thus showing that $a$ is a unit in $T$.

($\Leftarrow$) Let $b$ be a non-zero element of $R$. Then $b^{-1} \in T$, so satisfies a monic polynomial

$$b^{-n} + \lambda_{n-1} b^{-n+1} + \cdots + \lambda_1 b^{-1} + \lambda_0 = 0$$

with coefficients in $R$. Multiplying through by $b^{n-1}$ gives

$$b^{-1} = -\lambda_{n-1} - \lambda_{n-2} b - \cdots - \lambda_1 b^{n-2} - \lambda_0 b^{n-1},$$

thus showing that $b^{-1} \in R$, and hence that $R$ is a field. □

**Example 12.7** The polynomial ring $k[t]$ is a PID, hence a UFD, and is therefore integrally closed in its field of fractions $k(t)$. The field $k(t)$ is also the field of fractions of the subring $R = k[t^2, t^3] = k + kt^2 + kt^3 + kt^4 + \cdots$ of $k[t]$. Now $R$ is not integrally closed in $k(t)$ because $t$ is integral over $R$ but not an element of $R$. It is a zero of the polynomial $x^2 - t^2 \in R[x]$. Because $k[t] = R[t]$ it therefore follows from Proposition 12.2 that every element of $k[t] = R[t]$ is integral over $R$. Hence the integral closure of $R$ in $k(t)$ is $k[t]$. ◇

**Example 12.8** Since $\mathbb{Z}$ is a UFD it is integrally closed in $\mathbb{Q}$. However, it is not integrally closed in the larger field $\mathbb{Q}(\sqrt{2})$ because $\sqrt{2}$ is integral over $\mathbb{Z}$. It is a
zero of $x^2 - 2 \in \mathbb{Z}[x]$. One can show that the integral closure of $\mathbb{Z}$ in $\mathbb{Q}(\sqrt{2})$ is equal to $\mathbb{Z}[\sqrt{2}] = \mathbb{Z} \oplus \mathbb{Z}\sqrt{2}$.

On the basis of the example of $\mathbb{Q}(\sqrt{2})$ one might guess that the integral closure of $\text{In}(\sqrt{-3})$ is $\mathbb{Z}[\sqrt{-3}]$. That would be wrong because although $\mathbb{Z}[\sqrt{-3}]$ is integral over $\mathbb{Z}$, it is not the integral closure. For example, $\alpha = \frac{1}{2}(1 + \sqrt{-3})$ is integral over $\mathbb{Z}$. It is a zero of $x^2 - x + 1 \in \mathbb{Z}$.

1.13 Integers in number fields

Definition 13.1 A subfield $F$ of $\mathbb{C}$ is called a number field if $\dim_{\mathbb{Q}} F < \infty$. The integral closure of $\mathbb{Z}$ in $F$ is called the ring of integers in $F$. 

\(\diamond\)