Algebra II: Rings and Fields

Course Notes
Math 123, Harvard University
C. McMullen

Contents

1 Introduction

Abstract algebra codifies, and brings to the fore, the formal similarities between diverse mathematical structures.

The prime example of abstract algebra is the notion of a *group*, which captures both the idea of symmetry and the idea of multiplication and inverses.

Similarly, linear algebra frees Cartesian space from any distinguished choice of coordinates, and even allows one to vary the underlying field. It simplifies the formal discussion of linear operators, without recourse to matrices.

In this course we will concentrate on commutative rings and fields. These structures mix addition and multiplication. Their interaction consolidates many stories, around the general topic of *factorization*.

The paradigm for factorization comes from the integers, which can be written as products of *prime numbers* in an essentially unique way. We will aim for similar results in other regimes, such as the factorization of polynomials, and deeper results in the theory of *number fields*.

A key intuition that is sometimes lost here is that the prime example of a ring is the ring of functions (of a particular type, such as polynomial, analytic, or continuous) on a *space* (such as Euclidean space, a manifold, or an algebraic variety). We will keep this intuition in mind and see glimpses of *algebraic geometry*, which provides every ring with an underlying space.

We will also see that the theory of fields, which at first sight seem to have little structure, is illuminated by the study of their symmetries. These symmetries are captured by *Galois theory*, a topic which leads to a profusion of important results and provides a deep challenge to one's geometric intuition for algebra.

2 Prequel I: Finitely presented groups

The largest abelian group on two generators is $\mathbb{Z}^2$. In fact, given any other abelian group A with generators x and y, there is a unique, surjective homomorphism

$$\phi : \mathbb{Z}^2 \to A$$

with $\phi(1,0) = x$ and $\phi(0,1) = y$. It is defined by $\phi(a,b) = ax + by$.

What is the largest group G on two generators, a and b, if we do not require that $[a,b] = e$? The answer is provided by the *free group* on two generators, $F_2 \cong \mathbb{Z} * \mathbb{Z}$.

The free group. Let W be the set of all finite words w in the alphabet $(a, \overline{a}, b, \overline{b})$. Words are be multiplied by concatenation, $w_1 * w_2$; the empty word serves as the identity element.

Since words grow in length when multiplied, there are no inverses in W. To remedy this situation, we introduce the relations $a\overline{a} = \overline{a}a = b\overline{b} = \overline{b}b = e$. Canceling such expressions in w leads to a well–defined *reduced word*, $\mathrm{red}(w)$, in which no further cancellation is possible.

These reduced words are the elements of F_2. Multiplication is defined by $w_1 w_2 = \mathrm{red}(w_1 * w_2)$. The inverse of a word is obtained by writing it down backwards and inverting each letter.

Theorem 2.1 *Given any group G generated by two elements x and y, there exists a unique surjective homomorphism $\phi : F_2 \to G$ such that $\phi(a) = x$ and $\phi(b) = y$.*

In this sense F_2 is the biggest group on two generators. The group F_n is defined similarly, as is the general notation of a free product of groups, $G_1 * G_2$.

Generators and relations. Using generators and *relations* one can define the notion of a *finitely presented group*

$$G = \langle g_1, \ldots, g_n : r_1 = s_1, \ldots, r_m = s_m \rangle.$$

Here r_m and s_m are elements of the free group $F = \langle g_1, \ldots, g_n \rangle$. For example:

$$\mathbb{Z}/n = \langle a \ : \ a^n = e \rangle \quad \text{and} \quad \mathbb{Z}^2 = \langle a, b \ : \ ab = ba \rangle.$$

The dihedral group of symmetries of a regular n–gon, generated by a rotation r and a flip f, has the presentation:

$$D_{2n} = \langle r, f \ : \ r^n = f^2 = e, frf = r^{-1} \rangle.$$

To give a formal definition, first note that we can assume all the relations have the form $r_1 = r_2 = \cdots r_m = e$. The formal definition of G is then

$$G = F/R,$$

where R is the smallest *normal* subgroup of F containing the elements $r_1, \ldots, r_m$.

Universal property. Suppose we have another group H, with elements $h_1, \ldots, h_n$. We then certainly have a unique homomorphism $\phi : F \to H$ satisfying $\phi(g_i) = h_i$. But what about a map starting from G? This will exist exactly when $\phi(r_i) = e$ for every $r_i \in F$. Informally, one says:

Theorem 2.2 *There exists a homomorphism $\phi : G \to H$ sending g_i to h_i if and only if ϕ sends each relation r_i to the identity.*

Consequences. Informally, to create G from F we need to 'kill' the elements (r_i) in the free group F. Once we kill r_i, we must also kill gr_ig^{-1} for every $g \in F$, and also all products of elements of this form. The elements of R are the *consequences* of the relations (r_i).

It is difficult to predict what *all* the consequences will be. In fact we have:

Theorem 2.3 *There is no algorithm to determine if a given finitely–presented group is trivial or not.*

Example. It is instructive to prove in detail that the presentation G given for the dihedral group D_{2n} actually determines this group. Certainly we have a surjective map $G \to D_{2n}$, but why is it injective? For the proof, we use the fact that $rf = fr^{-1}$ to show that every element of G has a representative of the form $r^i f^j$, with $0 \leq i < n$ and $0 \leq j < 2$. Since there are just $2n$ expressions of this form, we have $G \cong D_{2n}$.

3 Prequel II: Quadratic forms

We now turn to a topic in linear algebra.

The main subject of linear algebra is *finite–dimensional* vectors spaces V of a field k. Common choices are $k = \mathbb{R}, \mathbb{C}$ or $\mathbb{Q}$.

A central theorem in the subject is that every vector space has a basis. This is equivalent, in the finite–dimensional case, to the assertion that $V \cong k^n$ for a unique $n = \dim V$. In other words we have:

Theorem 3.1 *A finite–dimensional vector space V over k is determined up to isomorphism by the integer $n = \dim(V)$.*

This illustrates a significant theme in linear algebra: if we are willing to consider isomorphic objects as equivalent, they are easier to classify. For example, up to isomorphism, the span V of $v_1 = (1, 2, \sqrt{3}, \pi)$ and $v_2 = (1, -1, 1, -1)$ in $\mathbb{R}^4$ is simply a copy of $\mathbb{R}^2$.

Bilinear forms. Let us now consider the classification of *inner products*, or more generally *symmetric bilinear forms*, on a finite–dimensional vector space. We will treat the case of vector spaces V over $\mathbb{R}$, where the answer is both simple and very useful.

A bilinear form $\langle x, y \rangle$ is a function $V \times V \to \mathbb{R}$ that is linear in each variable individually. For $V = \mathbb{R}^n$ with the standard basis $e_1, \ldots, e_n$, a bilinear form is uniquely determined by its *Gram matrix*

$$A = A_{ij} = \langle e_i, e_j \rangle.$$

In terms of A, we have

$$\langle x, y \rangle = x^t A y = \sum A_{ij} x_i y_j.$$

Nondegeneracy. We say a bilinear form is *nondegenerate* if for every $x \neq 0$, there exists a y such that $\langle x, y\rangle \neq 0$. This is equivalent to the condition $\det(A) \neq 0$.

If $\langle x, y\rangle = \langle y, x\rangle$, we say the bilinear form is *symmetric*.

Quadratic forms. A symmetric bilinear form determines a *quadratic form* by

$$Q(x) = \langle x, x\rangle = x^t A x.$$

In coordinates, $q(x)$ is a homogeneous quadratic polynomial. Conversely, over $\mathbb{R}$ (or any field of characteristic $\neq 2$), a quadratic form determines a symmetric bilinear form by:

$$2\langle x, y\rangle = Q(x + y) - Q(x) - Q(y). \tag{3.1}$$

Examples. The familiar *Euclidean inner product* on $\mathbb{R}^n$ is given by

$$\langle x, y\rangle = \sum_1^n x_i y_i.$$

Its associated *quadratic form* is given by

$$Q(x) = \langle x, x\rangle = \sum x_i^2 = \|x\|^2.$$

By the Pythagorean rule, it measures the length (squared) of the vector x. The level sets of $Q(x)$ are spheres.

More generally, the inner product of *signature* (p, q) on $\mathbb{R}^n$ is given, for $p + q = n$, by

$$Q_{p,q}(x) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_{p+q}^2.$$

The corresponding matrix is

$$A_{p,q} = \begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix}.$$

Relativity and hyperbolic geometry. On $\mathbb{R}^4$, thought of as space–time, we have the famous Lorentz form of special relativity,

$$Q(x, y, z, t) = x^2 + y^2 + z^2 - t^2.$$

It has signature $(3,1)$. We have written this form in coordinates where the speed of light $c = 1$. In this model for physical reality, space and time are part of the same substrate, and the spacelike length of one second of time is 186,282 miles.

An important different between Euclidean and Lorentzian space is that there are vectors of length *zero* in the latter case; these correspond to the trajectories of light rays.

The form of signature $(2,1)$ on $\mathbb{R}^3$ is easier to visualize; its level sets include the 1–sheeted hyperboloid, defined by $Q(x) = 1$; the 2–sheeted hyperboloid, defined by $Q(x) = -1$; and the light cone $Q(x) = 0$.

Just as the sphere S^2 defined by $x^2 + y^2 + z^2 = 1$ is the basis for spherical geometry, the 'sphere of radius $\sqrt{-1}$', defined by

$$x^2 + y^2 - z^2 = -1,$$

can be taken as the basis for hyperbolic geometry.

Classification. We can now state the classification of quadratic forms Q, or equivalently of symmetric bilinear forms.

We note that if $Q(x)$ is nondegenerate and $\dim(V) > 0$, then there exists an x such that $Q(x) \neq 0$, because of formula (3.1).

Theorem 3.2 *Every nondegenerate quadratic form $Q(x)$ on an n–dimensional vector space over $\mathbb{R}$ is isomorphic to one of the standard forms of signature (p,q) on $\mathbb{R}^n$.*

This means there is an isomorphism $\iota : V \cong \mathbb{R}^n$ such that $Q(x) = Q_{p,q}(\iota(x))$ for all $x \in V$. Here are some other equivalent statements:

1. Let $A \in \mathrm{M}_n(\mathbb{R})$ be a symmetric matrix with $\det(A) \neq 0$. Then there exists a matrix $B \in \mathrm{GL}_n(\mathbb{R})$ such that

$$B^t A B = A_{p,q}.$$

2. Given a nondegenerate symmetric form $\langle x, y\rangle$ on V, there exists a basis e_i such that $\langle e_i, e_j\rangle = \pm\delta_{ij}$ for all i, j.

Proof. The proof will be by induction on $\dim V$, the case of dimension 0 being immediate. So assume $\dim(V) \geq 1$. Since Q is nondegenerate, there

exists an $e_1 \in V$ such that $Q(e_1) = \langle e_1, e_1 \rangle \neq 0$. Since $Q(\lambda e_1) = \lambda^2 e_1$, we can arrange that $Q(e_1) = \pm 1$.

Now let $W = e_1^\perp = \{v \in V \ : \ \langle v, e_1 \rangle = 0\}$. Since W is the kernel of a nontrivial linear map, $\dim(W) = \dim(V) - 1$. Also, since Q is nondegenerate, for any $w \neq 0$ in W, there exists a $v = w' + \lambda e_1 \in V = W \oplus \mathbb{R} e_1$ such that

$$\langle w, v \rangle = \langle w, w' \rangle + \lambda \langle w, e_1 \rangle = \langle w, w' \rangle \neq 0.$$

By induction, $(W, Q|W)$ has a basis $e_2, \ldots, e_n$ with $\langle e_i, e_j \rangle = \pm \delta_{ij}$, which can be combined with e_1 to obtain the desired basis for V. ∎

Example. The quadratic form $Q(x, y) = xy$ has signature $1, 1$; the familiar equation

$$xy = (u + v)(u - v) = u^2 - v^2$$

shows the change of variables $x = u + v$, $y = u - v$ transforms Q to the standard form of signature $(1, 1)$.

Positive–definite forms. It is usual to require that an inner product be *positive definite*; that is, it should satisfy $\langle x, x \rangle > 0$ whenever $x \neq 0$. As a special case, the classification theorem for quadratic forms shows that any inner product has an orthonormal basis.

Degenerate forms. A very similar result holds for degenerate forms. In the degenerate case, we have $p + q < n$, and the matrix I_{pq} is extended by zeros to give an $n \times n$ matrix.

Classification in dimension one. In the special case $n = 1$ we find there are just two (nonzero) quadratic forms over $\mathbb{R}$, namely x^2 and $-x^2$. In general, the set of nonzero 1–dimensional quadratic forms over a field k is isomorphic to $k^\times / (k^\times)^2$. In particular, over $\mathbb{Q}$ there are already infinitely many different quadratic forms of dimension one.

Classification in dimension two. Using the fact that $\det(B^t A B) = \det(B)^2 \det(A)$, it is easy to see:

Theorem 3.3 *A nondegenerate quadratic form $x^t A x$ on $\mathbb{R}^2$ has signature $(2, 0)$ or $(0, 2)$ if $\det(A) > 0$; otherwise it has signature $(1, 1)$.*

Conics. We now recapitulate the theory of conics in $\mathbb{R}^2$ from the perspective of quadratic forms. By definition, a conic $C \subset \mathbb{R}^2$ is the zero locus of a polynomial equation of the form

$$(ax^2 + bxy + cy^2) + (dx + ey) + f = P_2(x, y) + P_1(x, y) + P_0 = 0.$$

It is traditional to rule out certain cases as *degenerate* conics, for example when C is empty, or consists of a single point or one or two lines. Otherwise, we have:

Theorem 3.4 *A nondegenerate conic in $\mathbb{R}^2$ is either an ellipse, a hyperbola or a parabola.*

Before proving this, we note that an ellipse is simply the image of the unit circle, defined by $x^2 + y^2 = 1$, under an affine automorphism of the plane, $f(x, y) = ax + by + c$. Similarly a general hyperbola is the affine image of the 'unit hyperbola', defined by $x^2 - y^2 = 1$.

Thus it suffices to show we can make a change of coordinates to put the definition equation of C into one of these two forms. This is always possible, by the classification of quadratic forms, provided $P_2(x, y)$ is nondegenerate. (The linear term is handled by completing the square.)

There remains the case where $P_2(x, y)$ has signature $(1, 0)$ (or $(0, 1)$; in this case, C is the affine image of the standard parabola, $y = x^2$.

Quadrics. A similar discussion can be carried out to describe hypersurfaces of degree 2 in $\mathbb{R}^n$, for any n. Let us summarize the answer for a quadric $Q \subset \mathbb{R}^3$, defined as the zero locus of $P_2(x, y, z) + P_1(x, y, z) + P_0 = 0$. In this case either:

1. Q is an *ellipsoid*, the affine image of the unit sphere, $x^2 + y^2 + z^2 = 1$; or
2. Q is a 1–sheeted hyperboloid, the affine image of the locus $x^2+y^2-z^2 = 1$; or
3. Q is a 2–sheeted hyperboloid, the affine image of the locus $x^2+y^2-z^2 = -1$; or
4. Q is an elliptic paraboloid, the affine image of the locus $z = x^2 + y^2$; or
5. Q is an hyperbolic paraboloid, the affine image of the locus $z = x^2 - y^2$.

Up to a sign, the quadratic form P_2 has signature $(3, 0)$ in the ellipsoid case; $(2, 1)$ in two hyperboloid cases; and $(2, 0)$ or $(1, 1)$ in the paraboloid cases. (The form is degenerate in the last two cases.)

Lie groups. The linear symmetries of $\mathbb{R}^n$ that preserve the standard quadratic form of signature (p, q) form a subgroup $\mathrm{O}(p, q) \subset \mathrm{GL}_n(\mathbb{R})$. The subgroup with $\det(A) = 1$ is denote $\mathrm{SO}(p, q)$.

The group of rotations $\mathrm{SO}(n)$ is the prime example of a compact Lie group, while $\mathrm{SO}(n,1)$ plays a central role in hyperbolic geometry.

Coda: Symplectic forms. The study of symmetric or quadratic forms is tied to both geometry and number theory, making them of central importance.

In the last few decades there has also been many advances in *symplectic geometry*, which is based on alternating forms. These forms are also important in topology, for example to describe the algebraic intersection pairing for curves on a surface.

Alternating forms. A bilinear form is *alternating* if

$$\langle x, y\rangle = -\langle y, x\rangle$$

for all $x, y \in V$. It is convenient to use a different notation for such forms, so we will write this as $x \cdot y$ instead.

As usual, if we choose a basis e_i for V then an alternating form is given by

$$x \cdot y = x^t A y$$

where $A_{ij} = e_i \cdot e_j$. This matrix is antisymmetric — it satisfies $A^t = -A$.

The area form in $\mathbb{R}^2$. The most basic example of an alternating form is the *area form* on $\mathbb{R}^2$, given by

$$x \cdot y = \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}.$$

This form measures the signed area of the parallelogram with sides x and y. Since $e_1 \cdot e_2 = 1$, the matrix of this form is

$$J_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Higher dimensions. The analogous standard form on $\mathbb{R}^{2g}$ is conveniently described in terms of a basis $(a_1, b_1, a_2, b_2, \cdots, a_g, b_g)$. It is characterized by

$$a_i \cdot b_j = \delta_{ij},$$

and its matrix J_{2g} is given simply by g copies of J_2 along the diagonal. Topologically, this form also records the intersection pairing in a standard basis for the homology $H_1(\Sigma_g, \mathbb{Z})$ on an oriented surface of genus g.

We can now state the classification theorem:

Theorem 3.5 *Every symplectic form on a real vector space V of dimension $2g$ is equivalent to the standard symplectic form on $\mathbb{R}^{2g}$.*

There are no symplectic forms on an odd–dimensional vector space.

Proof. The proof is by induction on $\dim V$. Let $x \cdot y$ be a symplectic form on V. Since this pairing is nondegenerate, there exist a pair of vectors $a_1, b_1 \in V$ such that $a_1 \cdot b_1 \neq 0$. Rescaling one of these vectors, we can assume that $a_1 \cdot b_1 = 2$.

Now let $W = \{v \in V \ : \ v \dot{a}_1 = v \cdot b_2 = 0\}$. It is readily verified that $V = W \oplus \mathbb{R}a_1 \oplus \mathbb{R}b_1$ and that $x \cdot y$ induces a symplectic form on W. A standard symplectic basis for W then provides, together with (a_1, b_1), a standard symplectic basis for V. ∎

General fields. It is worth noting that since *squares* did not appear in the argument above, the same result holds for symplectic forms over any field (even a field of characteristic 2). Thus alternating forms are much easier to classify than symmetric forms.

4 Rings and ideals

Bitte vergiss alles, was Du auf der Schule gelernt hast;
denn Du hast es nicht gelernt.

—E. Landau, quoted by M. Artin.

A (commutative) *ring* is a set A equipped with two binary operators, plus and times, such that

1. $(A, +)$ is a commutative group; and for all $a, b, c \in A$:
2. Multiplication is commutative and associative, meaning $a \cdot b = b \cdot a$ and $a \cdot (b \cdot c) = (a \cdot b) \cdot c$;
3. The distributive law $a \cdot (b + c) = a \cdot b + a \cdot c$ holds; and
4. There exists an element $1 \in A$, the (multiplicative) *identity*, such that $1 \cdot a = a$ for all $a \in A$.

The identity is unique, since $1 = 1 \cdot 1' = 1'$ for any other candidate $1'$. Also $0 \cdot a = 0$ for any $a \in A$, since $a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0$.

We usually abbreviate multiplication by writing $a \cdot b$ as ab.

Commutativity. We will use the word *ring*, without other qualifiers, to mean a commutative ring as above. There are also important examples of rings that are not commutative, for example rings of matrices, such as $\mathrm{M}_n(\mathbb{R})$. When discussing such rings we will explicitly mention their non-commutativity.

Units. We say $a \in A$ is a *unit* if it has a multiplicative inverse; that is, $ab = 1$ for some $b \in A$. This element is unique, and we write $b = a^{-1}$. The set of units forms an abelian group under multiplication, denoted $A^\times$.

Fields and integral domains. A field is a special case of a ring: it is a ring in which $A^\times = A - \{0\}$. Note that a field has no *zero divisors*: if $ab = 0$ then $a = 0$ or $b = 0$. A ring A with no zero divisors is called an *integral domain*, provided $0 \neq 1$ in A.

Every subring of a field is an integral domain. Conversely, every integral domain A embeds naturally into its associated *field of fractions* K.

The field of fractions K consists of all expressions of the form a/b, with $a, b \in A$ and $b \neq 0$, modulo the equivalence relation

$$a/b \sim c/d \iff ad = bc. \tag{4.1}$$

The field of fractions of $\mathbb{Z}$ is $\mathbb{Q}$. The field fractions of the polynomial ring $\mathbb{R}[x]$ is $\mathbb{R}(x)$, the field of *rational functions* $f(x) = P(x)/Q(x)$ that can be written as ratios of polynomials.

(Note: $f(x)$ is *rational* because it is a *ratio* of polynomials.)

Basic examples. Rings exist in great variety; here is a sampler.

1. The most basic example of a ring is $A = \mathbb{Z}$. The only units in $\mathbb{Z}$ are ± 1.

2. The next simplest example is the ring $A = \mathbb{Z}/n$. In this ring, $(\mathbb{Z}/n)^\times$ consists of those $x \in \mathbb{Z}/n$ with $\gcd(x, n) = 1$. The inverse of an element can be computed using the Euclidean algorithm to find r, s such that $rx + sn = 1$; then $1/x = r$.

3. Note that if n is composite, $\mathbb{Z}/n$ has zero divisors; for example, $2 \cdot 5 = 0$ in $\mathbb{Z}/10$. On the other hand, if $n = p$ is prime, then

$$\mathbb{Z}/p = \mathbb{F}_p$$

is a *field*, since every nonzero element is relative prime to p.

In $\mathbb{F}_p$ we have the relations $x^p = x$ and $(x+y)^p = x^p + y^p$, for all $x, y \in \mathbb{F}_p$.

4. The familiar fields $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ are also rings.

5. The vector space $\mathbb{R}^n$ becomes a ring if we define $(x_i) \cdot (y_i) = (x_i y_i)$. The identity element is the vector $(1, 1, 1, \ldots 1)$, and the units are the vectors x with $x_i \neq 0$ for all i.

   If $n > 1$, this ring also has zero divisors; for example, in $\mathbb{R}^2$, we have $(1, 0) \cdot (0, 1) = 0$.

6. More generally, the product $A \times B$ of any two rings is again a ring, with $(a, b) \cdot (a', b') = (aa', bb')$. This ring always has zero divisors.

7. The space $C[0, 1]$ of continuous maps $f : [0, 1] \to \mathbb{R}$ is a ring under addition and multiplication of functions. Its units are the nowhere vanishing functions. This ring also has zero divisors.

8. However, the ring $A[0, 1]$ of *real analytic functions* on $[0, 1]$ is an integral domain. This is because, if $f(x) \neq 0$, then the zeros of $f(x)$ form a finite set. The same is true for the ring of polynomial functions on $[0, 1]$.

9. One can form the ring $C(X)$ for any topological space X.

   For example, we can think of $\mathbb{R}^n$ as the ring of $C(\{1, 2, \ldots, n\})$ of (continuous) functions on a discrete set with n elements.

10. The ring of *Gaussian integers* is defined by

$$\mathbb{Z}[i] = \mathbb{Z} \oplus \mathbb{Z}i \subset \mathbb{C};$$

    it forms a square lattice in the complex numbers. This set is closed under multiplication because $i^2 = 1$. Since $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$, it is an integral domain.

    Note: in general, for $t \in \mathbb{C}$ we let $\mathbb{Z}[t]$ denote the smallest subring of $\mathbb{C}$ containing $\mathbb{Z}$ and t. Note that this ring must contain t^d for all $d > 0$.

11. The ring of *Eisenstein integers* $\mathbb{Z}[\omega] = \mathbb{Z} \oplus \mathbb{Z}\omega$ forms a hexagonal or triangular lattice in $\mathbb{C}$. Here

$$\omega = \exp(2\pi i/3) = (-1 + i\sqrt{3})/2$$

is a primitive cube root of unity. This set is closed under multiplication because $\omega^2 + \omega + 1 = 0$.

12. The ring $\mathbb{Z}[1/2] \subset \mathbb{R}$ consists of all rationals of the form $p/2^n$, $n \geq 0$. This is a dense subset of the real line. As an additive group,

$$\mathbb{Z}[1/2] = \bigcup_1^\infty 2^{-n}\mathbb{Z}$$

is a union of cyclic groups, but the union itself is not even finitely generated.

13. The ring $\mathbb{Z}[\sqrt{2}] = \mathbb{Z} \oplus \mathbb{Z}\sqrt{2} \subset \mathbb{R}$ is also dense, but in this case it is isomorphic to $\mathbb{Z}^2$ as an additive group. It is closed under multiplication because $(\sqrt{2})^2 \in \mathbb{Z}$.

14. A complex number t is *transcendental* if it is not the root of any polynomial $p(x) \in \mathbb{Z}[x]$. If t is transcendental, then the map $p(x) \mapsto p(t)$ gives an isomorphism

$$\mathbb{Z}[x] \cong \mathbb{Z}[t] \subset \mathbb{C}.$$

For example, $\mathbb{Z}[\pi]$ is isomorphic to a ring of polynomials.

15. An extreme example of a ring is the zero ring, whose only element is $1 = 0$.

By convention, the zero ring is *not* a field or an integral domain. It is the only ring in which $0 = 1$, and we require these two elements are distinct in integral domains.

16. Any matrix $L \in \mathrm{M}_n(\mathbb{R})$ generates a ring $A = \mathbb{R}[L]$, whose elements have the form

$$a = \sum_0^m a_i L^i, \quad a_i \in \mathbb{R}.$$

Note that $L^0 = I$ is the identity element in this ring. Since $\mathrm{M}_n(\mathbb{R})$ has dimension n^2 over $\mathbb{R}$, we can always take $m < n^2$. (In act one can take

$m < n$.) The same construction can be carried out with other fields or rings in place of $\mathbb{R}$. Note also that, although the full ring of matrices is non–commutative, this ring is commutative.

The ring A — unlike a subring of $\mathbb{C}$ — can have *nilpotent elements*, i.e. nonzero elements a such that $a^k = 0$ for some $k > 0$. For example, when $L = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, we have $L^2 = 0$. In this case A consists of the matrices of the form $\left(\begin{smallmatrix} a & b \\ 0 & a \end{smallmatrix}\right)$, with $a, b \in \mathbb{R}$.

Maps between rings. We now turn to a generalities. It is frequently the case that *morphisms* between objects are as important as the objects themselves, so we begin with this definition.

Given a pair of rings A and B, a map $f : A \to B$ is a *homomorphism* if it respects addition, multiplication, and the identity element; that is, if

$$f(a+b) = f(a) + f(b), \quad f(ab) = f(a)f(b), \quad \text{and} \quad f(1) = 1.$$

For example, the natural map $\mathbb{Z} \to \mathbb{Z}/n$ is a homomorphism. The inclusion of $\mathbb{R} \times 0$ into $\mathbb{R}^2$, however, is not a homomorphism, since it does not send 1 to 1.

Properties of homomorphisms. Let $f : A \to B$ be a homomorphism of rings.

1. Since f is a homomorphism of additive groups, f is injective if and only if $\mathrm{Ker}(f) = (0)$.
2. We have $f(A^\times) \subset B^\times$. Indeed, $f(a^{-1}) = f(a)^{-1}$, since $f(1) = 1$.
3. Thus if A is a field, and B is not the zero ring, the map f is injective.

Ideals. Let A be a ring. An additive subgroup $I \subset A$ is an *ideal* if it satisfies

$$AI = \{ai \,:\, a \in A, i \in I\} \subset I.$$

Thus an ideal is closed under addition and multiplication; but it is usually not a ring, since it usually does not contain the identity. (If it does, then $I = A$). Moreover, an ideal is closed under multiplication by all elements of A, not just by element of I.

Proposition 4.1 *The kernel of a ring homomorphism $f : A \to B$ is an ideal.*

Proof. Since f is a homomorphism of additive groups, its kernel I is an additive subgroup of A. And if $x \in I$, and $a \in A$, then $f(ax) = f(a)f(x) = f(a) \cdot 0 = 0$, so $ax \in I$. ∎

Why ideals? The name 'ideal' was introduced by Dedekind, who also introduced the definition of the real numbers as 'Dedekind cuts' in the rational numbers $\mathbb{Q}$.

In the integers, the ideal $n\mathbb{Z}$ consists of every number divisible by n. Dedekind's idea is that an ideal $I \subset A$ consists of every number divisible by some 'ideal number' ξ. This ideal number need not belong to the ring A at all; it is manifested through the ideal I, just as a real number is manifest by the way it cuts $\mathbb{Q}$ into two sets.

Put differently, the definition of an *ideal* captures the properties enjoyed by the multiples of ξ: for example, if a is a multiple of ξ, so is ab for every $b \in A$, and thus we require that $AI = I$.

Quotient rings. Thus ideal are for rings what normal subgroups are for groups. And indeed, any ideal is a normal subgroup of $(A, +)$, since this group is abelian. The special property of an ideal is that the quotient group A/I can be made into a ring in a natural way, by defining

$$(a + I) \cdot (b + I) = (ab) + I.$$

To see this is well–defined, we note that $(ab)+I$ is the *unique* coset containing the product of the cosets on the left as subsets of A. In fact, we have

$$(a + I)(b + I) = ab + aI + bI + II \subset ab + I + I + I = ab + I.$$

This shows:

Theorem 4.2 *For any ideal I, the natural map $f : A \to A/I$ is a ring homomorphism.*

Generating ideals. Any $x \in A$ generates a *principal ideal*, defined by $I = Ax$ and usually denoted $I = (x)$. More generally, elements $x_1, \ldots, x_n \in A$ generate an ideal

$$I = Ax_1 + Ax_2 + \cdots + Ax_n = (x_1, \ldots, x_n).$$

Every ideal in $\mathbb{Z}$ in principal: it has the form $I = n\mathbb{Z}$, for some $n \geq 0$. Indeed, *every subgroup* of $\mathbb{Z}$ is a *ring*. This is because multiplication by an integer can be carried out by iterated addition.

Ideals, fields and units. At the opposite extreme, it is easy to see:

A is a field iff it has exactly two ideals, namely A and (0).

Proper ideals. One often says an ideal $I \subset A$ is *proper* if

$$0 \neq I \neq A.$$

Thus a field has no proper ideals.

A principal ideal $I = (a) \subset A$ is proper if and only if $a \neq 0$ and a is not a unit. However, $I = (a_1, \ldots, a_n)$ can be equal to A even when none of its generators are units. For example, $(3, 5) = \mathbb{Z}$, and more generally $(k_1, \ldots, k_n) = (\gcd(k_i))$.

Ideals and homomorphisms. Ideals in rings behave a like normal subgroups in groups. For example, we have:

Proposition 4.3 *Let $f : A \to B$ be a homomorphism, and let $I \subset B$ be an ideal. Then $f^{-1}(I)$ is an ideal in A.*

Proof. In fact $f^{-1}(I)$ is the kernel of the map $A \to B \to B/I$. ∎

Images of ideals. It is generally *not true* that $f(I)$ is an ideal in B when I is an ideal in A. For example, $n\mathbb{Z}$ is an ideal in $\mathbb{Z}$, but its image under the natural map $\mathbb{Z} \to \mathbb{Q}$ is never an ideal (unless $n = 0$).

The situation is better for surjective maps.

Theorem 4.4 *If $f : A \to B$ is a surjective homomorphism, and $I \subset A$ is an ideal, then so is $f(I)$.*

Proof. We need only show that $Bf(I) \subset f(I)$. Any $b \in B$ can be written as $b = f(a)$, with $a \in A$, and therefore

$$b \cdot f(I) = f(a) \cdot f(I) = f(a \cdot I) \subset f(I).$$

∎

The following result is sometimes called the 'correspondence theorem'.

Corollary 4.5 *For any ideal $I \subset A$, the ideals in A/I correspond bijectively with the ideals $J \subset A$ that contain I, via the map*

$$J \mapsto J/I.$$

Proof. Use the fact that the quotient map $f : A \to A/I$ is a surjection, so ideals are preserved under both f and f^{-1}. ∎

Example. The ideals in $\mathbb{Z}/n$ are in bijection with the ideals in $\mathbb{Z}$ that contain $n\mathbb{Z}$. We have $n\mathbb{Z} \subset m\mathbb{Z}$ if and only if $m|n$. So the ideals of $\mathbb{Z}/n$ are in bijection with the divisors of n.

Rings of polynomials. Given a ring A, the associated *ring of polynomials* $A[x]$ consists of all formal expressions

$$p(x) = a_0 + a_1 x + \cdots + a_d x^d,$$

with $d \geq 0$ and $a_i \in A$. We adopt the convention that $a_i = 0$ for $i > d$. Sums and products of polynomials are defined in the natural way. If $p(x)$ and $q(x)$ have coefficients (a_i) and (b_i), then the coefficients of $(p+q)(x)$ are $(a_i + b_i)$. Similarly,

$$p(x)q(x) = \sum_{i=0}^{\infty} \left(\sum_{j+k=i} a_j b_k \right) x^i.$$

Although we have written an infinite sum, only finitely many terms are nonzero.

Freedom of polynomials. Like the free group $\langle x \rangle$, the ring $A[x]$ is the *largest ring* extending A and generated by just one additional element, x. This statement can be made precise as follows:

Theorem 4.6 *Suppose A is a subring of B, and $t \in B$. Then there is a unique homomorphism*

$$f : A[x] \to B$$

such that $f(a) = a$ for $a \in A$, and $f(x) = t$. Its image is the smallest subring of B containing both A and t.

The image is usually denoted by $A[t] \subset B$; it should not be confused with the ring of polynomials $A[x]$.

Polynomials as functions. For any polynomial $p(x) \in A[x]$, and $a \in A$, we can plug in a to obtain the value $p(a) \in A$. (This is a special case of the Theorem above, with $B = A$ and $t = a$.)

Thus each polynomial determines a *function* $p : A \to A$, giving a map

$$A[x] \to A^A.$$

It is sometimes said that $A[x]$ *consists of* the functions $p : A \to A$ that can be expressed as polynomials. This is not correct!

For example, $\mathbb{F}_p[x]$ is an infinite ring, since the degree of a polynomial can be arbitrarily large, but there are only finitely many maps $p : \mathbb{F}_p \to \mathbb{F}_p$. In this case many polynomials determine the same function, since $a^p = a$ for all $a \in \mathbb{F}_p$.

However, if K is an infinite field, a polynomial $p(x) \in K[x]$ is in fact determined by the corresponding function $p : K \to K$, since a polynomial with infinitely many zeros must be the zero polynomial.

Degree in $A[x]$. If A is an integral domain, then the product of the leading coefficients in a product $p(x)q(x)$ cannot vanish, so we have

$$\deg(pq) = \deg(p) + \deg(q)$$

for all $p, q \in A[x]$. Because of this natural expectation, we rarely consider polynomials over a ring with zero divisors; and we have:

Theorem 4.7 *If A is an integral domain, then $A[x]$ is an integral domain.*

Many variables. One can also form the ring $A[x_1, \ldots, x_n]$ of polynomials in n variables with coefficients in A. This is the largest extension of A generated by n elements.

There is a natural isomorphism between $A[x, y]$ and $(A[x])[y]$; new variables can be added one at a time, or all at once.

Adding relations. Just as the free group is only the first step in defining group presentations, polynomial rings can be seen as the first step in defining rings using generators and relations.

As a simple example, we have

$$\mathbb{Q}[x]/(x^2 - 2) \cong \mathbb{Q}(\sqrt{2}) \subset \mathbb{R}.$$

This does require some proof. We certainly have a surjective map from $\mathbb{Q}[x]$ to $\mathbb{Q}[\sqrt{2}]$, and its kernel contains the principal ideal $(x^2 - 2)$. But could the kernel be larger? The answer is *no*, since both sides are vector spaces of dimension two over $\mathbb{Q}$.

Nilpotence. For another type of example, consider the ring $A = \mathbb{Q}[x]/x^2$. Heuristically, this is the ring $\mathbb{Q}(\sqrt{0})$. It contains a *nilpotent element* $x \neq 0$, which satisfies $x^2 = 0$.

Prime and maximal ideals. The two special kinds rings — fields and integral domains — correspond to two special kinds of ideals.

Suppose $I \subset A$ is an ideal and $I \neq A$. We say I is *prime* if for all $a, b \in A$,

$$ab \in I \iff (a \in I) \quad \text{or} \quad (b \in I).$$

We say I is *maximal* if it is not contained in any larger, proper ideal.

Theorem 4.8 *The ring A/I is an integral domain iff I is a prime ideal. It is a field iff I is a maximal ideal.*

Proof. For the first statement, just note that the definition of a prime ideal is exactly what is needed to insure that $ab = 0$ in A/I iff $a = 0$ or $b = 0$.

The second statement follows from the correspondence principle, Corollary 4.5: A/I is a field $\iff$ A/I has exactly two ideals $\iff$ there are exactly two ideals in A containing I $\iff$ I is a maximal ideal. ∎

Corollary 4.9 *Every maximal ideal is prime.*

Note that the ideal A itself is neither maximal nor prime. This reflects the fact that we require $0 \neq 1$ in any field or integral domain.

Using the Axiom of Choice, it is routine to show:

Theorem 4.10 *Any ideal $I \subset A$, other than A itself, is contained in a maximal ideal.*

Corollary 4.11 *An element $a \in A$ belongs to a maximal ideal iff a is not a unit.*

Corollary 4.12 *Every ring, other than the zero ring, has at least one maximal ideal.*

Picture of units and maximal ideals. In $\mathbb{R}^n$, the maximal ideals are the coordinate planes, and the rest of the ring consists of units. A similar picture holds in any ring: we can think of $A^\times$ as a 'neighborhood' of the unit 1, and the rest of A as a union of maximal ideals.

Ascending sequences of ideals. In many algebraic situations, the ring of interest A is *Noetherian*, meaning there is no strictly increasing sequences of ideals $I_1 \subset I_2 \subset \cdots$, and hence the existence of maximal ideals is immediate.

On the other hand, in the ring $C(\mathbb{R})$, such sequences do exist, e.g.

$$I_0 \subset I_1 \subset I_2 \subset \cdots$$

where I_n consists of those functions $f(x)$ that vanish for $|x| > n$. Then $J = \bigcup I_n$ consists of the functions with compact support. This ideal is still not maximal. In fact, no one has ever given an explicit maximal ideal containing J; its construction seems to require the Axiom of Choice.

Principal ideal domains. A third important type of ring is a *principal ideal domain* (PID) Such a ring A is an *integral domain* in which every ideal is *principal*: it is generated by *one element*.

For a typical example of a ring that is *not* a PID, consider the ideal:

$$I = (x, y) \subset \mathbb{C}[x, y].$$

This ideal is not principal.

In addition, we have $0 \subset (x) \subset (x, y)$, where all 3 ideals are prime and the last is maximal. This is evidence that $\mathbb{C}[x, y]$ corresponds to a 2–dimensional space. In some sense, PIDs correspond to 1–dimensional spaces; we will later see that every prime ideal $I \neq (0)$ in a PID is maximal.

The ring $\mathbb{Z}[x]$ also fails to be a PID; for example, $J = (2, x)$ is not principal.

The integers and division. Let us now turn to the prototypical example of a PID, and see how it can be generalized.

Theorem 4.13 *The integers $\mathbb{Z}$ are a principal ideal domain.*

Corollary 4.14 *Every ideal in $\mathbb{Z}$ has the form $I = n\mathbb{Z}$ for some $n \geq 0$.*

The fact that $\mathbb{Z}$ is a PID can be explained in terms of the behavior of *division of integers* (with remainder). Let us explain this argument in detail, since we will soon generalize it.

We claim that any nonzero ideal $I \subset \mathbb{Z}$ is generated by its least positive element a; that is, $I = a\mathbb{Z}$. To see this, observe that any other element $b \in I$ can be written as

$$b = na + r$$

with a reminder satisfying $0 \leq r < a$. Since $a, b \in I$, so is r, and thus $r = 0$ by the definition of a, which proves $I = a\mathbb{Z}$.

The smallest nonzero element of an ideal is also the greatest common divisor of its generators; we have:

$$I = (a_1, \ldots, a_n) = (\gcd(a_1, \ldots, a_n)).$$

Recall that $\gcd(a, b)$ can be easily computed by the *Euclidean algorithm*, which is based on the fact that for $a > b$, we have

$$\gcd(a, b) = \gcd(b, a \bmod b).$$

For example,

$$\gcd(187, 55) = \gcd(55, 22) = \gcd(22, 11) = \gcd(11, 0) = 11.$$

We also have:

Theorem 4.15 *The nonzero prime ideals of $\mathbb{Z}$ are given by $p\mathbb{Z}$, where p is prime; and every such ideal is maximal.*

Proof. The ring $\mathbb{Z}/n$ has zero divisors unless n is prime, in which case it is a field. ∎

Polynomials over a field. The ring of polynomials in one variable, over a field, has some close similarities to the ring $\mathbb{Z}$; in particular we have:

Theorem 4.16 *For any field K, the polynomial ring $K[x]$ is a principal ideal domain.*

Division of polynomials. This result stems from the fact that $K[x]$, like $\mathbb{Z}$, admits a *Euclidean division algorithm*: given $p, q \in K[x]$, $q \neq 0$, we can write $p(x)$ uniquely in the form

$$p(x) = q(x)s(x) + r(x),$$

where the *remainder*, if not zero, satisfies $\deg r(x) < \deg q(x)$. Here is a reminder of the procedure to find s and r:

$$\begin{array}{r|l} & x+2 \quad \mathsf{R} \quad -x-1 \\ \cline{2-2} x^2+1 & x^3+2x^2 \;+\; 1 \\ & x^3 \;+\; x \\ \cline{2-2} & 2x^2 - x + 1 \\ & 2x^2 \;+\; 2 \\ \cline{2-2} & -x - 1 \end{array}$$

In this example, we find $p(x) = x^3 + 2x^2 + 1 = (x^2 + 1)(x + 2) - (x + 1)$.

Observe that if $\deg(p) \geq \deg(q)$ and the leading terms of p and q are ax^i and bx^j, then the leading term of $s(x)$ must be $(a/b)x^{i-j}$. It is here we use the fact that K is a *field.*

Proof of Theorem 4.16. The proof follows the same pattern as the proof for $\mathbb{Z}$: given a nonzero ideal $I \subset K[x]$, consider any one of its elements p of *least degree.* We claim $I = (p)$. Indeed, if $q \in I$ were not a multiple of p, division by p would yield a remainder $r \in I$ with $\deg(r) < \deg(p)$. ∎

One can also compute $\gcd(p, q)$ for two polynomials in $K[x]$, just as one would for integers; the result is a single generator for the ideal (p, q).

Monic polynomials. We say $p(x) \in A[x]$ is a *monic* polynomial if its leading coefficient is 1; that is, if

$$p(x) = x^d + a_1 x^{d-1} + \cdots + a_d.$$

When K is a field, any nonzero polynomial in $K[x]$ can be rescaled so it is monic, giving a canonical generator for each ideal.

Corollary 4.17 *The nonzero ideals in $K[x]$ are in bijection with the monic polynomials, via $p \mapsto I = p(x)K[x]$.*

Beyond fields. When A is not a field, the monic polynomials in $A[x]$ play a special role. For example, the division algorithm for $p(x)/q(x)$ works over any ring, provided $q(x)$ is *monic.* In §6, we will see that the complex zeros of monic polynomials in $\mathbb{Z}[x]$ form an important subring of $\mathbb{C}$, the ring of *algebraic integers.*

Maximal ideals and irreducible polynomials. A polynomial $p(x) \in K[x]$ is *irreducible* if $\deg(p) > 0$ and p cannot be written as a product $p = qr$ of polynomials of lower degree.

Theorem 4.18 *The nonzero prime ideals in $K[x]$ are exactly those generated by irreducible polynomials. All such ideals are maximal.*

Proof. If (p) is prime and p factors nontrivially as $p = qr$, then q or r are multiples of p, a contradiction. Thus p is irreducible. Conversely, if (p) is irreducible, then it cannot be properly contained in a large ideal (q), else we could write $p = qr$ for some r. Thus (p) is maximal, and hence prime. ∎

As we will see in §5, irreducible polynomials in $K[x]$ play the same role as primes in $\mathbb{Z}$: every polynomial is a product of irreducibles, in an essentially unique way.

The evaluation map. Every point $a \in K$ gives rise to a natural *evaluation map*

$$E : K[x] \to K$$

defined by $E(p) = p(a)$. The kernel of this map is a maximal ideal, since K is a field. In fact the kernel is exactly $I = (x - a)$. To see this, note that we can write

$$p(x) = (x - a)q(x) + c,$$

and $p(a) = c = 0$ if and only if $p \in (x - a)$.

Examples over $\mathbb{C}$ and $\mathbb{R}$. Thus the points of K always contribute to the maximal ideals of $K[x]$. Since $\mathbb{C}$ is algebraically closed, the only irreducible polynomials are the linear ones. This shows:

Theorem 4.19 *The maximal ideals in $\mathbb{C}[x]$ all have the form $I = (x - a)$, $a \in \mathbb{C}$.*

This simple fact reveals that the *ring* $\mathbb{C}[x]$ of polynomial *functions* on $\mathbb{C}$ implicitly knows about the underlying space:

Corollary 4.20 *The maximal ideals in $\mathbb{C}[x]$ correspond to the points of $\mathbb{C}$.*

Since the roots of real polynomials occur in conjugate pairs, we have:

Corollary 4.21 *The maximal ideals in $\mathbb{R}[x]$ have the form $(x - a)$ and $((x - a)^2 + b)$, where $a, b \in \mathbb{R}$ and $b > 0$.*

Note that the roots of the quadratic polynomial above are $x = a \pm \sqrt{-b}$. This Corollary reflects the well–known fact that in $\mathbb{R}[x]$ any real polynomial can be factored into linear and quadratic terms, the latter with complex roots.

The quotient of $\mathbb{R}[x]$ by a linear irreducible polynomial is $\mathbb{R}$, while the quotient by a quadratic irreducible polynomial is (isomorphic to) $\mathbb{C}$. Its two roots give a pair of complex conjugate points in $\mathbb{C} - \mathbb{R}$, and thus we have:

Corollary 4.22 *The maximal ideals in $\mathbb{R}[x]$ correspond to the points in $\mathbb{C}/(z \sim \overline{z})$.*

Continuous functions. The association between points and maximal ideals is also evident in the ring of real–valued continuous functions, $C[0,1]$.

Theorem 4.23 *Every maximal ideal in $C[0,1]$ has the form*

$$M_a = \{f \,:\, f(a) = 0\}$$

for some $a \in [0,1]$.

Proof. Suppose M is a maximal ideal that is not of this form. Then for every $a \in [0,1]$, the difference $M - M_a$ is nonempty. Hence for every $a \in [0,1]$, there exists a neighborhood U of a and a function $f \in M$ with $f(x) > 0$ on U. By compactness, $[0,1] = \bigcup U_i$ is a finite union of such neighborhoods, and the corresponding functions $f_i \in M$ give rise to a function $g = \sum f_i^2 \in M$ that is nowhere zero. But then g is a unit in $C[0,1]$, a contradiction. ∎

Corollary 4.24 *Let $E : C[0,1] \to K$ be a ring homomorphism, where K is a field. Then $K \cong \mathbb{R}$ and $E(f) = f(a)$ for some $a \in [0,1]$.*

The same theorem holds with $[0,1]$ replaced by any compact Hausdorff space X, and with $C[0,1]$ replaced by the ring of real or complex valued continuous functions $C(X)$ on X.

Algebraic geometry. The idea that quite generally, rings and spaces are linked by maximal ideals, provides a rich and important bridge between algebra and geometry.

In algebraic geometry, the most fundamental objects are the zero sets V of one or more polynomials in $\mathbb{C}^n$. One can associate to V the *ideal*

$$I(V) \subset \mathbb{C}[x_1, \ldots, x_n]$$

of all polynomials vanishing on V. Conversely, given an ideal, we can form the *variety*

$$V(I) = \{a \in \mathbb{C}^n \,:\, p(a) = 0 \;\forall p \in I\}.$$

Point evaluations and maximal ideals. When V consists of a single point $a \in \mathbb{C}^n$, the ideal $I(V)$ is maximal, since it is the kernel of the point evaluation map $E(p) = p(a)$. Indeed, we have

$$I(a) = (x_1 - a_1, \ldots, x_n - a_n).$$

It is a result of central importance that over $\mathbb{C}$, all maximal ideals are obtained in this way. This result is called the *Nullstellensatz* for $\mathbb{C}^n$.

Theorem 4.25 *Every maximal ideal of the polynomial ring $A = \mathbb{C}[x_1, \ldots, x_n]$ has the form*

$$I = (x_1 - a_1, \ldots, x_n - a_n)$$

for some point $a = (a_1, \ldots, a_n) \in \mathbb{C}^n$.

Proof. Let $K = A/I$ be the field associated to a maximal ideal I for A. Then K is a vector space over $\mathbb{C}$ of *countable dimension*, since A is generated over $\mathbb{C}$ by the monomials $x_1^{e_1} x_2^{e_2} \cdots x_n^{e_n}$.

On the other hand, the field $\mathbb{C}(x)$ of rational functions on $\mathbb{C}$ has uncountable dimension over $\mathbb{C}$, because the elements $(x - a)^{-1}$, $a \in \mathbb{C}$ are linearly independent.

The proof pivots on this difference of dimensions. Fix i, and consider the natural map

$$\mathbb{C}[x_i] \to A \to A/I = K.$$

If this map is injective, then — since K is a field, and $\mathbb{C}[x_i]$ is an integral domain — it extends to an injective map from $\mathbb{C}(x_i)$ to K. This contradicts the dimension count above. Thus the map has a kernel J. Since $\mathbb{C}[x_i]/J$ embeds into K, its image is an integral domain, and thus J is a prime ideal. Consequently $J = (x_i - a_i)$ for some $a_i \in \mathbb{C}$, by the classification of ideals in $\mathbb{C}[x]$. But the kernel is just $\mathbb{C}[x_i] \cap I$, so $(x_i - a_i) \subset I$.

Applying this reasoning for each i, we find that I contains the maximal ideal $M = (x_1 - a_1, \ldots, x_n - a_n)$. Since M is maximal and $I \neq A$, we have $I = M$. ∎

Corollary 4.26 *Let K be a field and let $f : \mathbb{C}[x_1, \ldots, x_n] \to K$ be a ring homomorphism. Then $K \cong \mathbb{C}$, and $f(p) = p(a)$ for some $a \in \mathbb{C}^n$.*

Points of a variety. It is now easy to generalize the Nullstellensatz to varieties in $\mathbb{C}^n$.

Theorem 4.27 *For any ideal $I \subset A = \mathbb{C}[x_1, \ldots, x_n]$, the maximal ideals in A/I are in natural bijection with the points of $V(I) \subset \mathbb{C}^n$.*

Proof. Every map from A/I to a field gives rise to a map from A to a field, whose kernel is a maximal ideal of the form $M_a = (x_1 - a_1, \ldots, x_n - a_n)$ by the Nullstellensatz for $\mathbb{C}^n$. Moreover M_a contains I, so every element of I must vanish at a, and therefore $a \in V(I)$. Conversely, every $a \in V(I)$ gives a maximal ideal M_a containing I, since every $p \in I$ vanishes at a. ∎

Corollary 4.28 *If $V(I)$ is empty, then there exists $f_i \in I$ and $g_i \in \mathbb{C}[x_1, \ldots, x_n]$ such that*

$$1 = \sum f_i(x) g_i(x).$$

Proof. Otherwise I contains no units, and therefore it is contained in a maximal ideal, which gives a point of $V(I)$. ∎

This means one can always give a *finite certificate* proving $V(I)$ is empty, namely the polynomials $g_i(x)$.

The meaning of $A/I(V)$. Continuing with A the ring of polynomials on $\mathbb{C}^n$, and $V \subset \mathbb{C}^n$ an algebraic variety, we can now explain the meaning of the quotient ring $A/I(V)$.

To do this, we will consider polynomials $p \in A$ as *functions* $p : \mathbb{C}^n \to \mathbb{C}$. We say $q : V \to \mathbb{C}$ is a polynomial if it can be extended to a polynomial on $\mathbb{C}^n$. We then have:

Theorem 4.29 *For any variety $V \subset \mathbb{C}^n$, the ring $A/I(V)$ coincides with the ring of polynomial functions on V, and*

$$E : A \mapsto A/I(V)$$

is just the restriction map, sending p to $p|V$.

Proof. By definition, every polynomial on V comes from one on $\mathbb{C}^n$, and by definition, $p|V = 0$ if and only if $p \in I(V)$. ∎

Examples. Note that when $V = \{a\}$ is a single point, $E(p) = p(a)$ is exactly the point evaluation map and $I(V) = M_a$. Thus the map $A \mapsto A/I(V)$ can be thought of as a generalized 'evaluation map'. now evaluating p not just at a point, but along the whole variety V.

Note also that when V is *empty*, we obtain the zero ring. Thus the zero ring can be interpreted as the ring of functions on the empty set; this ring contains one element, namely the empty function.

Algebraically closed fields. The Nullstellensatz for $\mathbb{C}^n$ also holds with $\mathbb{C}$ replaced by an arbitrary algebraically closed field, like $\overline{\mathbb{Q}} \subset \mathbb{C}$, the field of algebraic numbers (see §6). However the proof we have given does not work in that case; since $\overline{\mathbb{Q}}$ is countable, the field $\overline{\mathbb{Q}}(x)$ is also countable, so it has countable dimension over $\overline{\mathbb{Q}}$.

Curves in $\mathbb{C}^2$. In the plane, the most interesting objects are the *curves* $C = V(p)$ defined by the vanishing of a single polynomial, $p(x, y) = 0$. There is a slight ambiguity here, because $V(p^n) = V(p)$. It is thus useful to assume that $p(x)$ is not a power of a polynomial of lower degree. Then the following are equivalent:

1. The polynomial $p(x, y)$ is irreducible.
2. The ideal (p) is prime (but not maximal).
3. The curve C is irreducible; that is, we cannot write $C = C_1 \cup C_2$ in a nontrivial way as the union of two other curves.

For example, the curve defined as the zero locus of $p(x, y) = x^2 + y^2 = (x - iy)(x + iy)$ is the union of two lines over $\mathbb{C}$, while the circle defined by $p(x, y) = x^2 + y^2 - 1 = 0$ is irreducible.

Irreducible curves provide good examples of prime ideals that are not maximal. In general, *prime ideals* in rings correspond to *irreducible varieties* in algebraic geometry. Larger ideals correspond to smaller varieties, and the *maximal ideals* correspond to single points.

5 Factorization

In this section we discuss generalizations of prime factorization in the ring of integers $\mathbb{Z}$ to more general rings A.

The best case scenario is where every element of A, with obvious exceptions such as 0 and 1, can be expressed as a product of primes. Then this factorization is (essentially) unique, and we say A is a *unique factorization domain* (or a UFD).

The main results are the following.

1. Any PID is a UFD.
2. Thus, the rings $\mathbb{Z}$ and $K[x]$, with K a field, are UFDs.
3. More generally, any ring with a Euclidean algorithm is a PID, and hence a UFD. This applies to, for example, the rings $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{2}]$.
4. In fact, $\mathbb{Z}[x]$ is a UFD, even though it is not a PID. This is one of the main results in the subject.

5. More generally, whenever A is a UFD, so is the polynomial ring $A[x]$. Thus $\mathbb{Z}[x_1, \ldots, x_n]$ is a UFD, as is $K[x_1, \ldots, x_n]$ for any field K.

We now turn to a detailed discussion.

Primes. Factorization is best discussed in an integral domain A, since in such a domain, the *cancellation law holds*: provided $a \neq 0$,

$$ab = ac \implies b = c.$$

Thus throughout this section we assume that all rings A under consideration are *integral domains*.

Ideals and elements: terminology. It is frequently useful to express properties of elements of A in terms of the principal ideals they generate. Here is some terminology.

We say a is a *proper element* of A if $a \neq 0$ and a is not a unit. Equivalently, (a) is a proper ideal, meaning

$$0 \neq (a) \neq A.$$

We say $a|b$ if there exists a c such that $ac = b$. Equivalently, $a|b$ if $(b) \subset (a)$.

We say proper elements a and b are *associates* if $a = bu$ for some unit u. Equivalently, $(a) = (b)$.

We say a proper element $p \in A$ is a *prime* if, whenever $p|ab$, either $p|a$ or $p|b$. Equivalently, (p) is a prime ideal.

Let us show that the only divisor of a prime p is the prime itself.

Proposition 5.1 *If $a|b$ and $b|a$, then a and b are associates.*

Proof. From $(b) \subset (a) \subset (b)$ we conclude $(a) = (b)$. ∎

Proposition 5.2 *If a proper element $a \in A$ divides a prime p, then a and p are associates.*

Proof. From $a|p$ we get $p = ab$ for some b, and hence $b|p$. Also $p|a$ or $p|b$. In the first case, we find that a and p are associates. In the second case, we find that b and p are associates, and hence a is a unit, contradicting our assumption that a was proper. ∎

Unique factorization domains. We say A is a *unique factorization domain* (UFD) if every proper element of A can be expressed as a product of primes,

$$a = p_1 \cdots p_n.$$

Of course one can always reorder the factors (p_i), and replace (p_i) by $(u_i p_i)$ where u_i are units, $\prod u_i = 1$. Any two factorization that differ in this way are said to be essentially the same.

Theorem 5.3 *Suppose $a \in A$ can be expressed as a product of primes,*

$$a = p_1 \cdots p_n.$$

Then this factorization is essentially unique.

Proof. We proceed as we would for the integers. Suppose

$$a = p_1 \cdots p_n = q_1 \cdots q_m,$$

where p_i and q_j are primes. We may assume that $n \leq m$. Since p_1 divides the product on the right, $p_1 | q_j$ for some j. Then up to units and reordering, we can assume $p_1 = q_1$. Canceling p_1 from both sides and proceeding by induction, we end up with an equation with 1 on one side and $n - m$ primes on the other. Thus $n = m$ and we are done. ∎

Primes versus irreducibles. We say a proper element $a \in A$ is *irreducible* if, whenever $a = bc$, either b or c is a unit. Equivalently, (a) is maximal among proper, principal ideals.

Failure of factorization. In many rings, we can factor elements into products of irreducibles. The main issue is that irreducibles need not be primes.

For a simple example, consider the ring $A = \mathbb{Z}[-5]$. In this ring we have

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

The elements appearing in both factorizations are *irreducible*, but the factorization into irreducibles is not unique.

The problem with $A = \mathbb{Z}[\sqrt{-5}]$ can be traced to the fact that 2 and 3 are irreducibles, but not primes (in A). For example,

$$A/(2) = \mathbb{Z}[x]/(2, x^2 + 5) = \mathbb{F}_2[x]/(x^2 + 1) = \mathbb{F}_2[x]/(x + 1)^2.$$

Thus (2) is a prime ideal, since $(x+1)$ is a nilpotent element. Hence 2 is not a prime in A. On the other hand, $I = (2, 1+\sqrt{5})$ *is* prime, but it is not principal. (In fact $A/I \cong \mathbb{F}_2$.)

To clarify the relationship between these two notions, we show:

Theorem 5.4 *Every prime is irreducible, and in a UFD, every irreducible element is prime.*

Proof. If p is prime and $p = ab$ then we can assume $p|a$; but also $a|p$, so a and p are associates and hence p is irreducible.

In a UFD, an irreducible element can have only one prime factor, so irreducibles are primes. ∎

Theorem 5.5 *If every proper element of A can be factored into a product of irreducible elements, in an (essentially) unique way, then A is a UFD.*

Proof. Let p be irreducible and suppose $p|ab$, so $pc = ab$ for some c. Factor a, b and c into irreducibles. By uniqueness, an associate of p must appear in the factorization of ab, so $p|a$ or $p|b$. Hence irreducibles are prime and we have a UFD. ∎

(Some authors take unique factorization into irreducibles as the *definition* of a UFD.)

PIDs are UFDs. Next we study factorization in principal ideal domains.

Theorem 5.6 *If A is a PID, then it is also a UFD.*

Proof. First, we show that every element in A is a product of irreducibles. If not, we would have an infinite sequence $\ldots a_n|a_{n-1}|\cdots|a_1$, and hence an increasing chain of ideals

$$(a_1) \subset (a_2) \subset (a_3) \subset \cdots,$$

all distinct. But then the ideal $\bigcup(a_i) = (b)$ for some b, and $b \in (a_i)$ for some i, a contradiction. So factorization always terminates.

Second, we show every irreducible element $a \in A$ is prime. We give two proofs.

(i) Every ideal is contained in a maximal ideal, which is prime. Thus $(a) \subset (p)$ with p prime, and thus $p|a$; but then p and a are associates, since a is irreducible, and thus a is prime

(ii) Suppose a is irreducible and $a|bc$, but a does not divide b. Then $(a, b) \neq (a)$, and hence (by irreducibility) $(a, b) = 1$. That is, we can write $ax + by = 1$ for some $x, y \in A$. But then

$$c = c \cdot 1 = c(ax + by) = a(cx) + (bc)y.$$

Since $a|bc$, this shows $a|c$. Thus a is prime. ∎

Primes are maximal. The following simple but important result indicates that PIDs are like one–dimensional spaces.

Theorem 5.7 *In a PID, every nonzero prime ideal is maximal.*

Proof. Let (p) be a nonzero prime ideal. Then (p) is contained in some maximal ideal (m), and hence $m|p$. Since p is prime, this implies $(p) = (m)$. ∎

Greatest common divisors. In a PID, any pair of elements a, b have a greatest common divisor $c = \gcd(a, b)$, characterized (up to a unit) by $(c) = (a, b)$. The element c is also the product of the prime (powers) dividing both a and b.

The element $c = \gcd(a, b)$ can always be expressed as $c = ax + by$ with $x, y \in A$. When a Euclidean algorithm is available, it is practical to compute $\gcd(a, b)$ as well.

Example: Inversion in the field $\mathbb{Q}[x]/(p(x))$. Let $p(x)$ be an irreducible polynomial in $\mathbb{Q}[x]$. Then (p) is maximal, so $K = \mathbb{Q}[x]/(p(x))$ is a field.

Now suppose we are given a nonzero element $[q(x)] \in K$. How can we compute its inverse? That is, how can we compute $r(x)$ such that $[r(x)q(x)] = 1$ in K?

The solution is to use the fact that $\gcd(q, p) = 1$. Thus we can find $r, s \in \mathbb{Q}[x]$ such that $ps + qr = 1$, and then $qr = 1 \bmod p$. The elements r, s in turn can be found by the Euclidean algorithm for polynomials.

As a concrete instance, suppose $A = \mathbb{Q}[x]/(p)$ with $p(x) = x^2 + 1$, and we wish to find $1/(x + 1)$. We observe that

$$(x + 1)(x - 1) = x^2 - 1 = (x^2 + 1) - 2,$$

and hence $1/(x+1)$ is given by $(1-x)/2$. This reflects the familiar fact that $(1+i)(1-i) = 2$.

Euclidean domains. An integral domain A is a *Euclidean domain* if, like $\mathbb{Z}$ and $K[x]$, it has a good division algorithm. This means there is a function $\sigma : A - \{0\} \to \mathbb{N} = \{0, 1, 2, 3, \ldots\}$ such that for any $a, b \in A$, with $b \neq 0$, we can write

$$a = bc + r$$

for some $c \in A$, with $r = 0$ or $\sigma(r) < \sigma(b)$. By a now familiar argument, we have:

Theorem 5.8 *Let A be a Euclidean domain. Then A is also a PID. In fact, every proper ideal $I \subset A$ is generated by (any one of) its smallest nonzero element(s), where $\sigma(a)$ is used to measure size.*

Proof. Let $a \in I$ minimize $\sigma(a)$. Then for any $b \neq 0$ in I we can write $b = ac + r$ with $\sigma(r) < \sigma(a)$. Since $b - ac \in I$ we have $r = 0$ and thus $b \in (a)$. Consequently $I = (a)$. ∎

Corollary 5.9 *A Euclidean domain is a UFD.*

Examples of Euclidean domains. For $\mathbb{Z}$ and $K[x]$, we can take $\sigma(n) = |n|$ and $\sigma(p) = \deg(p)$, respectively.

Corollary 5.10 *The integers $\mathbb{Z}$ form a PID.*

Corollary 5.11 *For any field K, the ring $K[x]$ is a PID.*

Complex quadratic rings. For a new type of example, we now turn to the Gaussian integers.

Theorem 5.12 *The ring $\mathbb{Z}[i]$ is a Euclidean domain with $\sigma(a + ib) = |a + ib|^2 = a^2 + b^2$.*

Proof. First observe that the center of the unit square is distance $\sqrt{2}/2 < 1$ from the vertices. Every other point in the square is closer to the vertices. It follows that for any point $z \in \mathbb{C}$, there is a point $a \in \mathbb{Z}[i]$ with $|z - a|^2 \leq 1/2$.

Now given $a, b \in \mathbb{Z}[i]$, $b \neq 0$, choose a point $c \in \mathbb{Z}[i]$ such that

$$\left|\frac{a}{b} - c\right|^2 < 1/2.$$

Writing $a = bc + r$, we have

$$\sigma(r) = |r|^2 = |a - bc|^2 = |b|^2 \cdot \left|\frac{a}{b} - c\right|^2 \leq |b|^2/2 < |b|^2 = \sigma(b).$$

Here we have used the fact that $\sigma(b) > 0$ since $b \neq 0$. ∎

Corollary 5.13 *The ring $\mathbb{Z}[i]$ is a PID, and hence a UFD.*

Gaussian primes. In the next section we will discuss factorization in $\mathbb{Z}[i]$ in detail. For the moment, we record the fact that *primes* in $\mathbb{Z}[i]$ are of two types:

- *Rational primes* $p \in \mathbb{Z}$, with $p = 3 \bmod 4$; and
- *Complex primes* $\pi = a \pm ib \in \mathbb{Z}[i]$, with $a^2 + b^2 = p \in \mathbb{Z}$ a prime, $p = 1 \bmod 4$.

Put different, every prime in $\mathbb{Z}[i]$ arises as a factor of an ordinary prime $p \in \mathbb{Z}$; if $p = 1 \bmod 4$, then $p = \pi\overline{\pi}$ is a product of two conjugate, complex primes; others p itself is $\mathbb{Z}[i]$.

Eisenstein integers. A similar geometric argument can be used to prove that the *Eisenstein integers* $\mathbb{Z}[\omega]$ are also a PID. We will later investigate in detail the *failure* of $\mathbb{Z}[\sqrt{-d}]$ to be a PID for general d.

The real quadratic ring $\mathbb{Z}[\sqrt{2}]$. For the moment, we note that at key point in the argument above we used that fact that $|zw| = |z| \cdot |w|$ for $z, w \in \mathbb{C}$.

We can carry out a similar argument to show that $\mathbb{Z}[\sqrt{2}]$ is a Euclidean domain. For this we will use the size function

$$\sigma(x + y\sqrt{2}) = |x^2 - 2y^2|.$$

This size function is related to the (Galois) automorphism $a \mapsto a'$ of this ring that sends $\sqrt{2}$ to $-\sqrt{2}$; it can be written as

$$\sigma(a) = |aa'|,$$

from which it follows readily that $\sigma(ab) = \sigma(a)\sigma(b)$.

Theorem 5.14 *The ring $\mathbb{Z}[\sqrt{2}]$ is a Euclidean domain with $\sigma(a) = |aa'|$.*

Proof. For any $x \in \mathbb{R}$ we can choose $n \in \mathbb{Z}$ such that $|x - n| \leq 1/2$. Thus, given $a, b \in \mathbb{Z}[\sqrt{2}]$ with $b \neq 0$, we can choose $c \in \mathbb{Z}[\sqrt{2}]$ such that

$$\frac{a}{b} - c = x + y\sqrt{2}$$

with $|x|$ and $|y|$ both $\leq 1/2$. Then

$$\sigma(a/b - c) = |x^2 - 2y^2| \leq 3/4,$$

and so $a = bc + r$ with

$$\sigma(r) = \sigma(a - bc) = \sigma(b)\sigma(a/b - c) \leq (3/4)\sigma(b) < \sigma(b).$$

■

Corollary 5.15 *The ring $\mathbb{Z}[\sqrt{2}]$ is a PID and a UFD.*

Unique factorization in $\mathbb{Z}[x]$. We now turn the important ring $\mathbb{Z}[x]$.

This ring is *not* a PID. To see it has interesting ideals, choose an integer $n > 1$ and a monic polynomial $q(x)$ with relatively prime coefficients, an let

$$I = (n, q(x)) \subset \mathbb{Z}[x].$$

It is then easy to see that $I \cap \mathbb{Z} = (n)$. Thus if I is principal, it must be generated by $\pm n$, which is impossible since $q(x) \notin (n)$. In fact, if we take $n = p$ to be prime, then

$$\mathbb{Z}[x]/I \cong \mathbb{F}_p[x]/(q(x)),$$

and if $q(x)$ is irreducible mod p then this quotient is a field, so I is a maximal ideal that is not principal.

Nevertheless, we have:

Theorem 5.16 *The ring $\mathbb{Z}[x]$ is a UFD.*

The proof will use the fact that $\mathbb{Z}$ is a UFD, and the same pattern can be used to prove:

Theorem 5.17 *If A is a UFD, so is $A[x]$.*

We will not prove this generalization, but we note an important consequence:

Corollary 5.18 *For any field K, the ring $K[x_1, \ldots, x_n]$ is a UFD.*

This ring is very far from being a PID, for example the ideal

$$I = (x^n, x^{n-1}y, \ldots, y^n) \subset K[x, y]$$

cannot be generated by fewer than $n + 1$ elements.

A remark on terminology: one usually speaks of *irreducible* polynomials, even though such a polynomial could also be called *prime*.

Primitive polynomials. The ring $A = \mathbb{Z}[x]$ is contained in the UFD $B = \mathbb{Q}[x]$, and one might therefore guess that if $p \in A$ is prime in B then it is prime in A. But this is false; for example, $2 \cdot (x + 1)$ is prime in B but it factors in A, since 2 is not a unit in $\mathbb{Z}$.

To take care of this kind of elementary factorization, we introduce the important idea of a *primitive* polynomial $p(x) \in \mathbb{Z}[x]$. We say

$$p(x) = a_0x^d + a_1x^{d-1} + \cdots + a_d$$

is *primitive* if $a_0 > 0$ and $\gcd(a_0, \ldots, a_d) = 1$. The condition on the sign of the leading term is just a convention to obtain uniqueness in the following assertion:

> *Any nonzero polynomial $q(x) \in \mathbb{Q}[x]$ can be written uniquely as*
>
> $$q(x) = Qp(x),$$
>
> *where $Q \in \mathbb{Q}$ and $p(x) \in \mathbb{Z}[x]$ is primitive.*
>
> *If $q(x) \in \mathbb{Z}[x]$, then $Q \in \mathbb{Z}$.*

The factorization above cleanly separates the two different issues that arise for factorization in $\mathbb{Z}[x]$: factorization of integers, and factorization of polynomials. The reason they do not interact further is capture by:

Theorem 5.19 (Gauss's Lemma) *The product of two primitive polynomials in $\mathbb{Z}[x]$ is again primitive.*

Proof. The proof centers on the basic fact that:

> *$p(x) \in \mathbb{Z}[x]$ is primitive if and only its image in $\mathbb{F}_p[x]$ is nonzero for all primes p.*

The theorem then follows from the fact that $\mathbb{F}_p[x]$ is an integral domain: if $p_1(x)p_2(x)$ is not primitive, then it vanishes in $\mathbb{F}_p[x]$ for some p, and then the same must be true for $p_1(x)$ or $p_2(x)$. ∎

(It is an interesting to try to imagine proving Gauss's Lemma directly from the formula for the coefficients of a product of two polynomials.)

Lemma 5.20 *Suppose $p \in \mathbb{Z}[x]$, $q \in \mathbb{Q}[x]$ and $pq \in \mathbb{Z}[x]$. If p is primitive, then $q \in \mathbb{Z}[x]$.*

Proof. Write everything in terms of primitive polynomials and use Gauss's Lemma. In detail, we can write $q = Qp_1$ and $pq = Np_2$ where p_1, p_2 are primitive; then

$$pq = Np_2 = Qpp_1.$$

By Gauss's Lemma, pp_1 is primitive, so by uniqueness, $pp_1 = p_2$ and $Q = N$. This shows that $q = Np_1$ lies in $\mathbb{Z}[x]$. ∎

Note that this Lemma is *false* without the assumption that $p(x)$ is primitive; for example, $(2x)(x + 1/2) = (2x^2 + 1)$.

Corollary 5.21 *If a primitive polynomial $p(x)$ is prime in $\mathbb{Q}[x]$, then it is prime in $\mathbb{Z}[x]$.*

Proof. Suppose $p(x)$ divides the product $a(x)b(x)$ of two polynomials in $\mathbb{Z}[x]$. Then, since $p(x)$ is a prime in $\mathbb{Q}[x]$, it divides one of them — say $a(x)$ — in $\mathbb{Q}[x]$. This means we can write $a(x) = p(x)q(x)$ with $q(x) \in \mathbb{Q}[x]$. But then $q(x) \in \mathbb{Z}[x]$ by the Lemma above, so $p(x)$ divides $a(x)$ in $\mathbb{Z}[x]$. This shows that $p(x)$ is prime. ∎

Proof of Theorem 5.16. Consider a nonzero polynomial $p(x) \in \mathbb{Z}[x]$. It suffices to treat the case where $p(x)$ is primitive. Since $\mathbb{Q}[x]$ is a UFD, we can factor $p(x)$ into a product of rational polynomials:

$$p(x) = q_1(x) \cdots q_n(x).$$

Now by clearing denominators and common factors, we can write each $q_i(x)$ as a rational multiple $Q_i p_i(x)$ of a primitive polynomial in $\mathbb{Z}[x]$. We then find

$$p(x) = Q p_1(x) \cdots p_n(x),$$

where $Q \in \mathbb{Q}$. By Gauss's lemma, the product $p_1(x) \cdots p_n(x)$ is primitive, and $p(x)$ is primitive by assumption. Thus $Q = 1$ and we have obtained a factorization

$$p(x) = p_1(x) \cdots p_n(x)$$

with $p_i(x) \in \mathbb{Z}[x]$ and each $p_i(x)$ a prime in $\mathbb{Q}[x]$. By the preceding Corollary, each $p_i(x)$ is also prime in $\mathbb{Z}[x]$. ∎

The proof shows that to factor a polynomial in $\mathbb{Z}[x]$, it suffices to first remove any common factor from its coefficients, and then factor the resulting primitive polynomial in the Euclidean domain $\mathbb{Q}[x]$.

Eisenstein's criterion. The method of reduction mod p, used to prove Gauss's Lemma, has many other applications. One of them is:

Theorem 5.22 (Eisenstein's criterion) *Let $q(x) = x^d + a_{d-1}x^{d-1} + \cdots + a_d \in \mathbb{Z}[x]$ have the property that, for some prime p, $p|a_i$ for all i but p^2 does not divide a_d.*

Then $q(x)$ is irreducible.

Proof. Suppose to the contrary the $q(x)$ is reducible — say $q(x) = r(x)s(x)$. Reducing mod p, we obtain $x^d = \overline{r}(x)\overline{r}(x)$ in $\mathbb{F}_p[x]$. Since this ring is a UFD, $\overline{r}(x) = x^e$ and $\overline{s}(x) = x^f$ (up to units), where $e + f = d$. Thus the constant terms in $r(x)$ and in $s(x)$ are both equal to zero mod p. But their product gives the constant term a_d in $q(x)$, so $p^2|a_d$. ∎

Cyclotomic example. This criterion can be used to show that, for each prime p, the *cyclotomic polynomial*

$$\Phi_p(x) = \frac{x^p - 1}{x - 1} = x^{p-1} + x^{p-2} + \cdots + 1$$

is irreducible. We use a clever trick: if $\Phi_p(x)$ factors, then so does

$$\Phi_p(x+1) = \frac{(x+1)^p - 1}{x} = x^{p-1} + \binom{p}{1} x^{p-2} + \binom{p}{2} x^{p-3} + \cdots + p,$$

but this polynomial is irreducible by Eisenstein's criterion.

Appendix: Discriminants and resultants. In this section we briefly describe how one can determine if two polynomials over a field have a common root, or if one polynomial has a multiple root.

The resultant. How can one test if $f, g \in K[x]$ have a common factor?

One method is to simply use the Euclidean algorithm to determine $\gcd(f, g)$. Another method, which clarifies the linear algebra underlying the problem, is to compute the *resultant* $R(f, g) \in K$.

To define it, suppose f and g have degrees d and e respectively. The key observation is that these polynomials have a common factor iff there exists a nontrivial solution to the equation

$$rf = sg \tag{5.1}$$

with $r, s \in K[x]$. The *trivial solution* is given by $gf - fg = 0$; for a nontrivial solution, we require that the degrees of rf and sg are both *less* than $d + e$. For example, if h is a common factor, then we obtain the nontrivial solution $(g/h)f = (f/h)g$.

Now notice that equation (5.1) is *linear* in the coefficients of r and s. Thus it has a nonzero solution (r, s) iff there is a linear relation among the polynomials

$$\{f, xf, x^2 f, \ldots, x^{e-1} f, g, xg, x^2 g, \ldots x^{d-1} g\},$$

or equivalently if the corresponding *determinant* $R(f, g)$ vanishes. This determinant is a square matrix of size $d + e$, that can be expressed in terms of the coefficients (a_i) and (b_i) of f and g respectively; when they have degrees

2 and 3, it is given by

$$R(f,g) = \det \begin{pmatrix} a_0 & a_1 & a_2 & 0 & 0 \\ 0 & a_0 & a_1 & a_2 & 0 \\ 0 & 0 & a_0 & a_1 & a_2 \\ b_0 & b_1 & b_2 & b_3 & 0 \\ 0 & b_0 & b_1 & b_2 & b_3 \end{pmatrix}.$$

We then have:

Theorem 5.23 *The polynomials $f, g \in K[x]$ have a common factor iff $R(f,g) = 0$.*

The discriminant. Now suppose $K = \mathbb{C}$. Then any $f(x) \in \mathbb{C}[x]$ is a product of linear polynomials. When does $f(x)$ have a multiple root, i.e. a factor of the form $(x-a)^2$? This happens iff $f(x)$ and $f'(x)$ have a common zero.

To detect this phenomenon, we define the *discriminant* of $f(x)$ by

$$D(f) = (-1)^{d(d-1)/2}(\mathrm{Res}(f(x), f'(x)).$$

(The sign is irrelevant at the moment, since we are only considering if $D(f) = 0$.) We then have:

Theorem 5.24 *The discriminant $D(f) = 0$ iff $f(x)$ has a multiple root.*

Note: when this multiple root is unique, it can be found explicitly by computing $\gcd(f, f')$.

Example: For $f(x) = x^3 + ax + b$, we have $f'(x) = 3x^2 + a$, and hence

$$D(f) = -\det \begin{pmatrix} b & a & 0 & 1 & 0 \\ 0 & b & a & 0 & 1 \\ a & 0 & 3 & 0 & 0 \\ 0 & a & 0 & 3 & 0 \\ 0 & 0 & a & 0 & 3 \end{pmatrix} = -4a^3 - 27b^2.$$

Bad reduction. Suppose $f(x) \in \mathbb{Z}[x]$ has distinct roots over $\mathbb{C}$. Still, it may acquire multiple roots when reduced modulo a prime p. Which primes give rise to this kind of *bad reduction*?

They are exactly the primes that divide $D(f)$. For example, if $f(x) = x^3 + 2x + 5$, we find $D(f) = -707 = -7 \cdot 101$, and we find that

$$f(x) = (x+2)^2(x+3) \mod 7 \quad \text{and} \quad f(x) = (x+29)^2(x+43) \mod 101.$$

6 Quadratic number fields

A central topic in number theory is the study of finite field extensions $K/\mathbb{Q}$, and their rings of integers $\mathcal{O}_K \subset K$. This setting provides a rich generalization of the familiar field $\mathbb{Q}$ and its ring of integers $\mathbb{Z}$. The step from $\mathbb{Q}$ to K is comparable to the step from the Riemann sphere to general compact Riemann surfaces.

New issues such as the class group and the unit group come to the fore already in the simplest case, where $\dim(K/\mathbb{Q}) = 2$. In this case one can write $K = \mathbb{Q}(\sqrt{d})$ for some square–free integer d. When $d > 0$, we can regard K as a subfield of $\mathbb{R}$; it is a *real* quadratic field; while for $d < 0$, we obtain a *complex quadratic field*, which can be regarded as a subfield of $\mathbb{C}$.

These two cases have different features. In this section we will study the class group of K for complex quadratic fields, and the unit group for real quadratic fields. Here is a brief summary of the main results.

1. Every number field K has a canonical subring $\mathcal{O}_K$, its ring of algebraic integers. This ring is the natural domain for the study of factorization and ideals.

2. For $K = \mathbb{Q}(i)$, the ring of integers $\mathcal{O}_K = \mathbb{Z}[i]$ is a unique factorization domain. We will determine the Gaussian primes and relate them to the Diophantine equation $a^2 + b^2 = n$.

3. For $K = \mathbb{Q}(\sqrt{-5})$, the ring of integers $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$ is *not* a UFD. We introduce the class group, which measures the failure of unique factorization, and show in this case it is isomorphic to $\mathbb{Z}/2$.

4. For general complex quadratic fields, we will show that despite the failure of unique factorization, every *ideal* can be written uniquely as a product of *prime ideals*. This is a central result in number theory.

5. For complex quadratic field, the group of units is always finite. We will see that for real quadratic fields, $\mathcal{O}_K^\times$ is always isomorphic to $\mathbb{Z}/2 \times \mathbb{Z}$. The group of units also plays a central role in number theory.

Degrees. We begin with some generalities that apply to all number fields. Let L be a field, and let $K \subset L$ be a subfield. Then we can regard L as a vector space over K, and we define the *degree* of the *field extension* L/K by

$$[L : K] = \dim_K L.$$

For example, $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$. One sometimes writes $\deg(L/K)$ for $[L : K]$.

Algebraic numbers and algebraic integers. A complex number t is an *algebraic number* if it is a zero of a polynomial $p(x) \in \mathbb{Q}[x]$. Otherwise, t is *transcendental.*

We say t is an algebraic *integer* if it is a zero of a *monic* polynomial

$$p(x) = x^d + a_1 x^{d-1} + \cdots + a_d$$

with *integral* coefficients.

The monic polynomial in $\mathbb{Q}[x]$ of least degree with t as a zero is called the *minimal polynomial* of t. The minimal polynomial is unique and irreducible, since $\mathbb{Q}[t]$ is a UFD.

By Gauss's Lemma, t is an algebraic integer if and only if its minimal polynomial lies in $\mathbb{Z}[x]$. The *degree* of t over $\mathbb{Q}$, denoted

$$\deg_{\mathbb{Q}}(t),$$

is the degree of its minimal polynomial.

Theorem 6.1 *A complex number t is algebraic if and only if the ring $\mathbb{Q}[t]$ is a finite–dimensional vector space over $\mathbb{Q}$. In this case $K = \mathbb{Q}[t]$ is a field, and*

$$\dim(K/\mathbb{Q}) = \deg_{\mathbb{Q}}(t). \tag{6.1}$$

Proof. The ring $\mathbb{Q}[t]$ is finite–dimensional if and only if the evaluation map $E : \mathbb{Q}[x] \to \mathbb{Q}[t]$ has a nontrivial ideal I. Since $\mathbb{Q}[t]$ is a PID, when I is nontrivial it has the form $I = (p)$, for some monic polynomial $p(x)$, and since I consists of all the polynomials that vanish at t, p is the minimal polynomial for t. Thus $K = \mathbb{Q}[t]$ is a field because p is irreducible, and equation (6.1) follows from the fact that $1, t, \ldots, t^{d-1}$ is a basis for K over $\mathbb{Q}$, where $d = \deg(p)$. ∎

By similar reasoning we find:

Theorem 6.2 *A complex number t is an algebraic integer if and only if the ring $\mathbb{Z}[t]$ is isomorphic, as an additive group, to $\mathbb{Z}^d$ for some d. In this case d is the degree of the minimal polynomial for t.*

What is an integer? The definition we have given of an algebraic integer is elegant and subtle, and many of the results that hold for $\mathcal{O}_K$ do not hold for its subrings; it is essential to include *all* integers.

Example: the golden ratio. The *golden ratio* is the element of the field $\mathbb{Q}(\sqrt{5})$ defined by

$$\gamma = \frac{1+\sqrt{5}}{2}.$$

It satisfies the algebraic equation $\gamma^2 = \gamma + 1$. Because of this, the ring it generates is given by:

$$\mathbb{Z}[\gamma] = \mathbb{Z} \oplus \mathbb{Z}\gamma.$$

Thus γ is an algebraic integer, even though the formula for γ has a 2 in the denominator.

Fibonacci series. We remark that the equation $\gamma^2 = \gamma + 1$ explains the close relationship between the golden ratio and the Fibonacci sequence

$$(f_0, f_1, f_2, \ldots) = (1, 1, 2, 3, 5, 8, 13, 21, \ldots).$$

Here $f_{n+2} = f_{n+1} + f_n$. In fact, $f_{n+1}/f_n \to \gamma$, and $f_n \sim C\gamma^n$ as $n \to \infty$.

Golden rectangles and architecture. The number γ occurs geometrically as the aspect ratio of the sides of a *golden rectangle* R, namely a rectangle that is similar to $R - S$, where S is a square. The golden rectangle has been used as a motif in art and architecture, and appears to frame the facade of the Parthenon in Athens.

Sums and products. Clearly the numbers $\sqrt{2}$ and $\sqrt{5}$ are algebraic integers. But what about their sum? Here are two general results that cover this situation.

Theorem 6.3 *The set of all algebraic numbers $\overline{\mathbb{Q}}$ forms a countable subfield of $\mathbb{C}$.*

Proof. The set of algebraic numbers $\overline{\mathbb{Q}}$ is countable because $\mathbb{Q}[x]$ is countable, and any given nonzero polynomial has only finitely many roots.

Next we show that $\overline{\mathbb{Q}}$ is a ring. Given $s, t \in \overline{\mathbb{Q}}$, consider the ring $K = \mathbb{Q}[s,t]$. This ring is spanned, as a vector space over $\mathbb{Q}$, by the monomials $s^i t^j$, $0 \leq i, j$. But using the minimal polynomial for t, we can write t^d, $d = \deg_{\mathbb{Q}}(t)$, as a rational linear combination of smaller powers of t; and similarly for s^e, $e = \deg_{\mathbb{Q}}(s)$. Thus $\dim(K/\mathbb{Q}) \leq de < \infty$. Since $s+t$ and st each generate a subring of K, Theorem 6.1 shows they are both algebraic. Thus $\overline{\mathbb{Q}}$ is a ring.

Finally, it is easy to check that if $t \neq 0$ is algebraic, then so is $1/t$. In fact, the minimal polynomial $p(x)$ for t has degree d, then $1/t$ is zero of $x^d p(1/x) \in \mathbb{Q}[x]$.

Consequently $\overline{\mathbb{Q}}$ is also a field. ∎

By similar reasoning, one can show:

Theorem 6.4 *The set of all algebraic integers forms a subring of* $\overline{\mathbb{Q}}$.

Field notation. Given $t \in \mathbb{C}$, we let $\mathbb{Q}(t) \subset \mathbb{C}$ denote the smallest *field* generated by $t \in \mathbb{C}$. When t is algebraic, we have

$$\mathbb{Q}(t) = \mathbb{Q}[t],$$

and we will generally use the former notation. (When t is transcendental, we have $\mathbb{Q}(t) \cong \mathbb{Q}(x)$, the field of rational functions in one variable with rational coefficients.)

Numbers and integers. We remark that if $t \neq 0$ is an algebraic number, with minimal polynomial

$$p(x) = x^d + a_1 x^{d-1} + \cdots + a_d,$$

then nt is an algebraic integer, for some $n > 0$ in $\mathbb{Z}$. Indeed, the minimal polynomial for nt is given by

$$n^d p(x/n) = x^d + n a_1 x^{d-}1 + \cdots + n^d a_d,$$

and its coefficients lie in $\mathbb{Z}$ provided n is chosen to clear denominators.

Rings and number fields. Every field $K \subset \mathbb{C}$ contains $\mathbb{Z}$, and hence $\mathbb{Q}$. We say K is a *number field* if $d = [K/\mathbb{Q}]$ is finite.

By the results above, every element of K is an algebraic number of degree $\leq d$, and the algebraic integers in K form a subring

$$\mathcal{O}_K \subset K.$$

As we will soon see concretely in the case of quadratic fields, we have:

Theorem 6.5 *If $K \cong \mathbb{Q}^d$ as a vector space, then $\mathcal{O}_K \cong \mathbb{Z}^d$ as an additive group.*

Sketch of the proof. Let e_i be a basis for $K/\mathbb{Q}$. Then for suitable $n_i \in \mathbb{Z}$, the multiples $n_i e_i$ are algebraic integers, and hence $\mathcal{O}_K$ contains a copy of $\mathbb{Z}^d$. To see it cannot be larger, it suffices to show that $\mathcal{O}_K$ is discrete in $K_{\mathbb{R}} = K \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbb{R}^d$. One proof of this is to define the norm function $N : K_{\mathbb{R}} \to \mathbb{R}$; then $U = \{x \in K \ : \ |N(x)| < 1\}$ is a neighborhood of the origin, and the only integral point it contains is $x = 0$, since $N(\mathcal{O}_K) \subset \mathbb{Z}$. ∎

Remark: fields in $\mathbb{C}$ vs. abstract fields. We adopt the perspective that K is a subfield of $\mathbb{C}$ for concreteness. In some ways it is more natural and symmetric to simply regard K as a finite extension of $\mathbb{Q}$, with no *chosen* embedding into $\mathbb{C}$, since there may be many choices.

Prime and maximal ideals. A critical property of the ring of integers in a number field is:

Theorem 6.6 *Any nonzero prime ideal P in $\mathcal{O}_K$ is a maximal ideal.*

Lemma 6.7 *Any finite integral domain A is a field.*

Proof. Since A is finite, given $u \neq 0$ we have $u^i = u^j$ for some $i > j > 0$; canceling, we find $u^n = 1$ some some n, and hence u is a unit, with inverse u^{n-1}. ∎

Proof of Theorem 6.6. Let $a \in P$ be a nonzero element. Then as additive groups, we have

$$(a) = a\,\mathcal{O}_K \cong \mathcal{O}_K \cong \mathbb{Z}^d.$$

It follows that (a) has finite index in $\mathcal{O}_K$. Since $(a) \subset P$, this shows $\mathcal{O}_K/P$ is a finite integral domain. Hence $\mathcal{O}_K/P$ is a field, and therefore P is a maximal ideal. ∎

The dimension of $\mathcal{O}_K$. This result indicates that $\mathcal{O}_K$, like $\mathbb{Z}$ itself, behaves like a 1–dimensional space. (The fact the nonzero prime ideals are maximal is one of the characteristic features of a *Dedekind domain*.)

The quadratic case. For the rest of this section, we specialize to the case where K is a *quadratic* extension of $\mathbb{Q}$. In view of the quadratic formula, we can always write

$$K = \mathbb{Q}(\sqrt{d})$$

for some square–free integer d. It can be shown that this integer is unique.

Galois involution, norm and trace. The field K has a unique nontrivial automorphism, given by

$$a = b + c\sqrt{d} \mapsto a' = b - c\sqrt{d},$$

where $b, c \in \mathbb{Q}$. In particular, the *Galois group* of $K/\mathbb{Q}$ is isomorphic to $\mathbb{Z}/2$.

Note that $a = a'$ if and only if $a \in \mathbb{Q}$. The *norm* and *trace* are maps

$$N : K \to \mathbb{Q} \quad \text{and} \quad \mathrm{Tr} : K \to \mathbb{Q}$$

defined by 'averaging over the Galois group' to obtain rational numbers. They are given by

$$\mathrm{Tr}(a) = a + a' \quad \text{and} \quad N(a) = aa'.$$

These have the useful properties that

$$\mathrm{Tr}(a+b) = \mathrm{Tr}(a) + \mathrm{Tr}(b) \quad \text{and} \quad N(ab) = N(a)N(b).$$

Detecting integers. Note that a satisfies the quadratic equation

$$(x-a)(x-a') = x^2 - (\mathrm{Tr}\, a)x + N(a) = 0.$$

Thus the norm and trace allow us to recognize the algebraic integers in K: we have

$$a \in \mathcal{O}_K \iff N(a) \text{ and } \mathrm{Tr}(a) \text{ lie in } \mathbb{Z}.$$

Units. The norm allows one to easily recognize units and compute their inverses.

Theorem 6.8 *An algebraic integer $a \in \mathcal{O}_K$ is a unit iff $|N(a)| = 1$, in which case $1/a = \pm a'$.*

Proof. If a is a unit, with inverse b, then $N(ab) = N(a)N(b) = N(1) = 1$; since $N(b)$ is an integer, we find $N(a) = \pm 1$. Conversely, if $N(a) = aa' = \pm 1$, then $1/a = \pm a'$ and hence a is a unit. ∎

The ring of integers. By checking the norm and the trace, one can prove:

Theorem 6.9 *Assuming d is square–free, the ring of integers in $\mathbb{Q}(\sqrt{d})$ is given by*

$$\mathcal{O}_K = \mathbb{Z} \oplus \mathbb{Z}\frac{1+\sqrt{d}}{2}$$

if $d = 1 \bmod 4$; otherwise, $\mathcal{O}_K = \mathbb{Z} \oplus \mathbb{Z}\sqrt{d}$.

Proof. The main point is to note that the norm of $(1+\sqrt{d})/2$ is $(1-d)/4$, which is an integer iff $d = 1 \bmod 4$.

For a formal proof, we first note that since d is square-free, whenever c is rational and $dc^2 \in \mathbb{Z}$, we have $c \in \mathbb{Z}$. This is because every prime in the denominator of c^2 occurs to at least second order, but at most to first order in d.

Now suppose $a + b\sqrt{d} \in \mathcal{O}_K$, with $a, b \in \mathbb{Q}$. This is *equivalent* to the norm and trace conditions

$$2a \in \mathbb{Z} \quad \text{and} \quad a^2 - db^2 \in \mathbb{Z}.$$

If $2a$ is even then $a \in \mathbb{Z}$ and hence $db^2 \in \mathbb{Z}$, which implies $b \in \mathbb{Z}$. So in this case we only get elements of $\mathbb{Z}[\sqrt{d}]$, and we already know that all elements of this ring are contained in $\mathcal{O}_K$.

Now suppose $2a$ is odd. Then $a^2 \in \mathbb{Z} + 1/4$, so $db^2 \in \mathbb{Z} + 1/4$ which implies that $4db^2 = d(2b)^2$ is an integer. Hence $(2b)$ is an integer, and

$$4db^2 = d(2b)^2 = 1 \bmod 4.$$

Now the squares mod 4 are 0 and 1, so the equation above implies that $(2b)^2 = 1 \bmod 4$ and hence $d = 1 \bmod 4$ and $2b$ is odd.

Thus the ring of (algebraic) integers is larger than $\mathbb{Z}[\sqrt{d}]$ only when $d = 1 \bmod 4$. Assuming this, we find further than an algebraic integer that is not in $\mathbb{Z}[\sqrt{d}]$ must have the form $a + b\sqrt{d}$ with both a and b in $\mathbb{Z} + 1/2$. These additional elements belong to $\mathbb{Z}[(1+\sqrt{d})/2]$, and we have already seen that conversely the ring generated by $(1+\sqrt{d})/2$ lies in $\mathcal{O}_K$. ∎

Why rings? Why fields? In Hilbert's book on number theory, he refers to $\mathcal{O}_K$ as a *Zahlring*; here *Zahl* means *number* in German. One might speculate that he used the word *ring* because the roots of unity are algebraic integers, or because the powers of an algebraic integer α eventually 'circle back', so that α^d can be expressed in terms of lower powers of α.

Dedekind referred to fields as *Zahlenkörper* (bodies of numbers). The choice to translate this word as field was apparently made by Moore in an 1893 article on Galois theory.

Primes. We can also show that every prime $\pi \in \mathcal{O}_K$ arises by factoring an ordinary prime $p \in \mathbb{Z}$.

Theorem 6.10 *Let $\pi \in \mathcal{O}_K$ be a prime element. Then π divides a unique ordinary $p \in \mathbb{Z}$.*

In fact $|N(\pi)| = p$ or p^2. In the first case p factors as $\pi\pi'$ in $\mathcal{O}_K$, while in the second case p remains prime in $\mathcal{O}_K$ and π is an associate of p.

Proof. Let

$$N(\pi) = \pi\pi' = \pm p_1 \dots p_n$$

be the prime factorization in $\mathbb{Z}$ of the norm of a prime π in $\mathcal{O}_K$. Then π divides some particular prime on the right, say $p = p_i$. Thus $N(\pi)$ divides $N(p) = p^2$, so $|N(\pi)| = p$ or $|N(\pi)| = p^2$.

If $|N(\pi)| = p$, then $p = \pm\pi\pi'$, and we have obtained the prime factorization of p in $\mathcal{O}_K$.

Otherwise, $|N(\pi)| = p^2 = N(p)$. Since $\pi | p$ we can write $p = \pi u$ for some $u \in \mathcal{O}_K$, and we find $N(u) = \pm 1$. Thus u is a unit and hence p is an associate of π. ∎

Complex quadratic fields. Let us now turn to the case where d is a square–free integer with $d < 0$. Then $K = \mathbb{Q}(\sqrt{d})$ is a *complex* quadratic field. In the complex case, the map $z \mapsto z'$ coincides with complex conjugation $\overline{z}$, and hence $N(z) = |z|^2$ for all complex quadratic fields.

As special and important cases, we note that the ring of integers $\mathcal{O}_K$ is given by

$$\begin{cases} \mathbb{Z}[i] & \text{when } d = -1, \\ \mathbb{Z}[\omega] & \text{when } d = -3, \text{ and} \\ \mathbb{Z}[\sqrt{d}] & \text{when } d \neq 1 \bmod 4. \end{cases}$$

Since $N(z) = z\overline{z} = |z|^2$, the units in $\mathcal{O}_K$ are exactly the quadratic integers that lie on the unit circle in $\mathbb{C}$. So we easily find:

Proposition 6.11 *The unit group $\mathcal{O}_K^\times$ in a complex quadratic field is:*

1. *Of order 4, generated by $i = \sqrt{-1}$, for the Gaussian integers $\mathbb{Z}[i]$;*
2. *Of order 6, generated by $(1 + \sqrt{3})/2$, for the Eisenstein integers $\mathbb{Z}[\omega]$; and*
3. *Of order 2, generated by -1, in all other cases.*

The Gaussian integers, *reprise*. We now turn to a detailed study of the Gaussian integers $\mathbb{Z}[i]$, which is the ring of integers in $K = \mathbb{Q}(\sqrt{-1})$. Recall that $N(a) = |a|^2$ and the units in $\mathbb{Z}[i]$ are ± 1 and ± 1.

Primes and factorization. As we have seen in §5, $\mathbb{Z}[i]$ is a Euclidean domain, and hence a UFD. That is, every nonzero $a \in \mathbb{Z}[i]$ can be written as a product of primes in an essentially unique way. But what are the primes in this ring? We will show:

Theorem 6.12 *Every prime $\pi \in \mathbb{Z}[i]$ arises as a divisor of a prime $p \in \mathbb{Z}$. If $p = 3 \bmod 4$, then p is a prime in $\mathbb{Z}[i]$. Otherwise:*

1. *We can write $p = a^2 + b^2$ as a sum of two squares, with $a, b \in \mathbb{Z}$;*
2. *The integer $\pi = a + bi$ is prime in $\mathbb{Z}[i]$, and*
3. *The prime factorization of p in $\mathbb{Z}[i]$ is given by $p = \pi\overline{\pi}$.*

Ramification and splitting. One way to look at this result is the following: the primes in $\mathbb{Z}[i]$ can all be found by factoring the primes in $\mathbb{Z}$. There are really 3 cases:

1. When $p = 3 \bmod 4$, the prime is *inert*; it remains prime in $\mathbb{Z}[i]$.
2. When $p = 1 \bmod 4$, the prime *splits*; it factors as a product of distinct primes, $p = \pi\overline{\pi}$.
3. Finally when $p = 2$, the prime *ramifies*; we have $2 = \pi\overline{\pi}$, but the primes π and $\overline{\pi}$ are associates. Indeed, we can take $\pi = (1 + i)$; then $\overline{\pi} = 1 - i = -i\pi$, and hence $2 = -i\pi^2$. On the level of ideals, we have $(2) = (\pi^2)$.

If we think of the primes of $\mathbb{Z}[i]$ as the points of a space, then we can regard $\mathbb{Z}[i]$ geometrically as a 2–sheeted covering space of $\mathbb{Z}$, branched over the primes $2, 3, 7, 11$ etc.; see Figure 1. Note that the geometry over the ramified prime $p = 2$ is special; the tangent space to $\mathbb{Z}[i]$ becomes vertical.

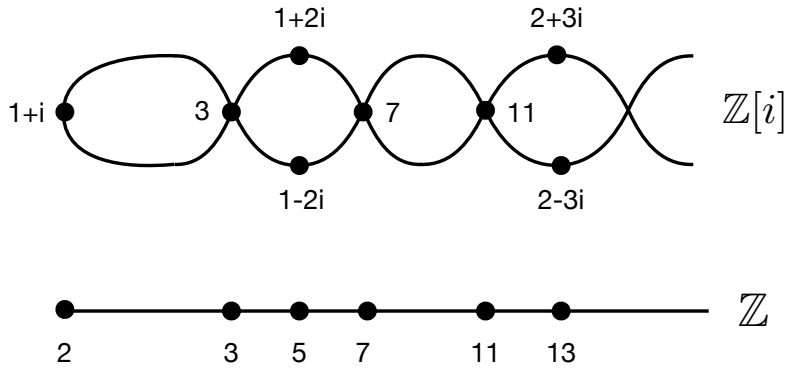


Figure 1. The Gaussian primes over the primes of $\mathbb{Z}$.

Sums of squares. Which whole numbers are sums of two squares? The analysis of prime factorization in $\mathbb{Z}[i]$ allows one, somewhat surprisingly, to solve the Diophantine equation

$$a^2 + b^2 = n$$

over $\mathbb{Z}$.

Corollary 6.13 *An integer $n > 0$ is a sum of two squares iff every prime with $p = 3 \bmod 4$ divides n to even order.*

Proof. Since $N(a + bi) = a^2 + b^2$, n is a sum of two squares iff it is the norm of a Gaussian integer $c = a + bi$. Up to a unit u, we can factor such an integer c into Gaussian primes:

$$c = up_1 \cdots p_n \pi_1 \cdots \pi_m,$$

where the primes p_i lie in *zed* and satisfy $p_i = 3 \bmod 4$, and the primes π_i have norms $N(\pi_i) = q_i$, with q_i prime and $q_i \neq 3 \bmod 4$. Since $N(p_i) = p_i^2$, the prime factorization of $n = N(c)$ is given by

$$n = c\bar{c} = (p_1 \cdots p_n)^2 q_1 \cdots q_m.$$

This shows the condition on the prime divisors of n is necessary for n to be a sum of two squares. It is also sufficient, since by factoring each q_i as $\pi_i\overline{\pi}_i$ we obtain an equation for $c \in \mathbb{Z}[i]$ satisfying $N(c) = n$. ∎

Points on circles. Note that the *number of solutions* to $a^2 + b^2 = n$ is the number of points in $\mathbb{Z}^2$ the circle of radius $\sqrt{n}$ in $\mathbb{R}^2$ centered at the origin; this number can also be determined by factoring n in $\mathbb{Z}[i]$.

The square–root of −1 in finite fields. To prepare for the description of Gaussian primes, we first describe the units in a finite field. We then determine when $x^2 + 1 = 0$ has a solution in $\mathbb{F}_p$.

Theorem 6.14 *Let F be a finite field. Then its unit group $F^\times$ is cyclic.*

Proof. By the theory of finite abelian groups, we have

$$F^\times \cong \prod_1^n \mathbb{Z}/d_i,$$

for some sequence of integers $d_i > 1$ with $d_i | d_{i+1}$. In particular, every element of $F^\times$ is a root of the polynomial $x^{d_n} - 1$. But this polynomial has at most d_n roots. Thus

$$d_1 \cdots d_n = |F^\times| \leq d_n,$$

which implies $n = 1$ and $F^\times$ is cyclic. ∎

Corollary 6.15 *The equation $x^2 = -1$ has a solution in $\mathbb{F}_p$ iff $p = 2$ or $p = 1 \bmod 4$.*

Proof. If $p = 2$ then $-1 = 1 \bmod p$, so $x = 1$ is a solution. Otherwise, solutions to $x^2 = -1$ correspond to elements of order 4 in $\mathbb{F}_p^\times \cong \mathbb{Z}/(p-1)$. And an element of order 4 exists iff 4 divides $p - 1$. ∎

Informally, $\mathbb{F}_p$ contains $\sqrt{-1}$ iff $p = 1 \bmod 4$.

Proof of Theorem 6.12. By Theorem 6.10, any prime $\pi \in \mathbb{Z}[i]$ divides a unique prime $p \in \mathbb{Z}$. Now p itself is prime in $\mathbb{Z}[i]$ iff the ring

$$\mathbb{Z}[i]/(p) = \mathbb{Z}[x]/(p, x^2 + 1) = \mathbb{F}_p[x]/(x^2 + 1)$$

is a field. This happens iff x^2+1 is irreducible, i.e it has no root in $\mathbb{F}_p$, which as we have just seen is equivalent to the condition $p = 3 \bmod 4$. So in this case, π is an associate of p.

Now suppose $p = 2$ or $p = 1 \bmod 4$. Then p is not a prime in $\mathbb{Z}[i]$; rather, by Theorem 6.10, it factors as a product of primes $p = \pi\overline{\pi}$. Writing $\pi = a+bi$, we find $p = a^2 + b^2$ is a sum of squares. ∎

Open problem. It is not easy to find a generator for the cyclic group $\mathbb{F}_p^\times$. In fact E. Artin conjectured that $\mathbb{F}_p^\times = \langle 2 \rangle$ for infinitely many p; this is still an open problem.

7 Ideals and lattices

In this section we define the notion of *multiplication of ideals*,

$$I, J \mapsto I \cdot J,$$

which makes sense in any ring. With this concept in place, we formulate and prove the main result on factorization in complex quadratic field K:

> *Any nonzero ideal $I \subset \mathcal{O}_K$ factors uniquely as a product of prime ideals, $I = P_1 \cdot P_2 \cdots P_n$.*

In particular, by factoring the *principal ideal* (a) generated a proper element $a \in \mathcal{O}_K$, we obtain a factorization of a into 'ideal prime numbers' that works even when $\mathcal{O}_K$ itself is not a UFD.

In preparation for the proof, we first discuss the geometric theory of lattices, and introduce the *ideal class group* $\mathrm{Cl}(K)$ of a general number field K.

This section concludes with a brief study of the unit group of *real* quadratic fields, where the geometry of lattices again plays an important role.

Lattices. A *lattice* $L \subset \mathbb{R}^n$ is a discrete additive subgroup isomorphic to $\mathbb{Z}^n$. Equivalently, L can be written in the form

$$L = \mathbb{Z}e_1 \oplus \mathbb{Z}e_2 \oplus \cdots \oplus \mathbb{Z}e_n,$$

where $(e_1, \ldots, e_n)$ is a basis for $\mathbb{R}^n$. For the standard basis, we just obtain the usual group of integral points $L = \mathbb{Z}^n \subset \mathbb{R}^n$.

We will be interested in lattices $L \subset \mathbb{C} \cong \mathbb{R}^2$. In this case, we say L and L' are *similar* if

$$aL = L'$$

for some $a \in \mathbb{C}^*$.

Shortest vector. Since L is discrete, it has one or more *shortest vectors*, i.e. elements $0 \neq a \in L$ such that $|a| \leq |b|$ for all nonzero $b \in L$. Abusing language, we will refer to a as *the shortest vector* in L, even though it is not unique. The length of the shortest vector is denoted by

$$|L| = \inf\{|a| \ : \ a \in L, a \neq 0\}.$$

Normalized lattices. Let us say L is *normalized* if $|L| = 1$ and $1 \in L$. We can always rescale L (multiply by a complex scalar) to make it normalized.

Let $\mathbb{H} = \{\tau \in \mathbb{C} : \operatorname{Im} \tau > 0\}$ denote the upper halfplane. For a normalized lattice, we can choose $\tau \in \mathbb{H}$ such that

$$L = \mathbb{Z} \oplus \mathbb{Z}\tau.$$

The quotient of the complex plane by L is a torus, with

$$\operatorname{area}(\mathbb{C}/L) = \operatorname{Im} \tau.$$

Equivalently, this is the area of the fundamental parallelogram with vertices $(0, 1, \tau, \tau + 1)$.

The fundamental domain. Now let us refine our choice of basis for a normalized lattice. Let

$$F = \{\tau \in \mathbb{H} \ : \ |\tau| \geq 1 \text{ and } |\operatorname{Re} \tau| \leq 1/2\}.$$

See Figure 2. (It is not an accident that F is also a fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{H}$.) Note that ∂F contains the circular arc joining ω to ω^2, and passing through i.

Theorem 7.1 *If $1 \in L$ is the shortest vector in L, then we can write*

$$L = \mathbb{Z} \oplus \mathbb{Z}\tau$$

with $\tau \in F$.

Proof. Take $\tau \in \mathbb{H}$ to be one of the shortest vectors in $L - \mathbb{Z}$. Then $|\tau| \geq |L| = 1$, and $|\operatorname{Re}\tau| \leq 1/2$ else one of $\tau \pm 1$ would be shorter, so $\tau \in F$.

Now consider any $a + b\tau \in L$, with $a, b \in \mathbb{R}$. Then we can translate by an element of $\mathbb{Z} \oplus \mathbb{Z}\tau$ to arrange that $|a|, |b| \leq 1/2$. Then

$$|a + b\tau| < 1/2 + |\tau|/2 \leq |\tau|,$$

which implies $a + b\tau = 0$ and hence $L = \mathbb{Z} \oplus \mathbb{Z}\tau$. ∎

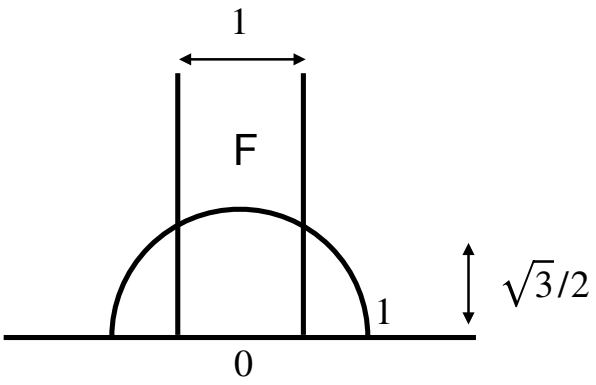


Figure 2. Normalized lattices have the form $L = \mathbb{Z} \oplus \mathbb{Z}\tau$, $\tau \in F$.

Corollary 7.2 *For any lattice L, we have*

$$\operatorname{area}(\mathbb{C}/L) \geq (\sqrt{3}/2)|L|^2.$$

Proof. It suffices to prove this for normalized lattices, with $|L| = 1$; and for these, $\operatorname{Im}\tau$ is minimized when $\tau = (1 \pm \sqrt{-3})/2$, in which case $\operatorname{Im}\tau = \sqrt{3}/2$. ∎

Corollary 7.3 *The densest normalized lattice in $\mathbb{C}$ is $\mathbb{Z}[\omega]$.*

Rectangles and diamonds. Consider a normalized lattice

$$L = \mathbb{Z} \oplus \mathbb{Z}\tau,$$

with $\tau \in F$.

We say L is a *rectangular lattice* if $\tau = iy$, some $y \geq 1$.

We say L is a *diamond lattice* if $\tau = iy \pm 1/2$. In this case, $2\tau \pm 1 = iy$, and L *contains* the rectangular lattice $L' = \mathbb{Z} \oplus \mathbb{Z}iy$ with index 2.

Conversely, every rectangular lattice $\mathbb{Z} \oplus \mathbb{Z}iy$ with $y > \sqrt{3}$ has a natural extension to a diamond lattice $\mathbb{Z} \oplus \mathbb{Z}\tau$, $\tau = (1 + iy)/2$.

Every lattice is similar to a normalized one, so we can use the same terminology for general lattices. Geometrically, L is a rectangular lattice if it has a basis of orthogonal vectors, and L is a diamond lattice if it is obtained from a rectangular lattice L' by adding in the barycenter of each rectangle. See Figure 3.

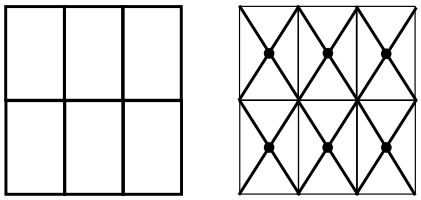


Figure 3. Rectangles and diamonds.

Examples. Given a square-free integer $d < 0$, consider the rectangular lattice

$$L = \mathbb{Z} \oplus \mathbb{Z}\sqrt{d};$$

and let $K = \mathbb{Q}(\sqrt{d})$. Then $\mathcal{O}_K = L$ provided $d = 2$ or $3 \bmod 4$; otherwise, $d = 1 \bmod 4$ and

$$\mathcal{O}_K = \mathbb{Z} \oplus \mathbb{Z}\frac{1 + \sqrt{d}}{2}$$

is the diamond lattice extending L.

Extensions of lattices. In the classification of ideals it will be important to look for lattices L that extend $\mathcal{O}_K$, but have the same shortest vector. The following result helps one to limit the search.

Theorem 7.4 *Suppose we have a pair of normalized lattices with $\sigma, \tau \in F$, such that*

$$\mathbb{Z} \oplus \mathbb{Z}\tau \subset \mathbb{Z} \oplus \mathbb{Z}\sigma.$$

Then $\sigma = (\tau + k)/n$ for some $n, k \in \mathbb{Z}$, $n > 0$.

Proof. This statement just rephrases the fact that τ lies in the lattice $\mathbb{Z} \oplus \mathbb{Z}\sigma$. ∎

Corollary 7.5 *There are only finitely many normalized lattices extending a given one.*

Proof. Since $\mathrm{Im}(\sigma) \geq \sqrt{3}/2$, the imaginary part of τ gives an upper bound on n, and for each n, the bound $|\operatorname{Re}\sigma| \leq 1/2$ limits the possibilities for k to a finite set. ∎

Example. Suppose $L = \mathbb{Z} \oplus \mathbb{Z}iy$, with $y \geq 1$ so this lattice is normalized. Then the normalized extensions of L all have the form $\mathbb{Z} \oplus \mathbb{Z}\sigma$, with

$$\sigma = \frac{iy + k}{n} \in F.$$

Then we have

$$0 < n \leq \frac{2y}{\sqrt{3}} \quad \text{and} \quad |k| \leq n/2,$$

since $\operatorname{Im}\sigma \geq \sqrt{3}/2$ and $|\operatorname{Re}\sigma| \leq 1/2$.

The ideal class group: $\mathbb{Z}[\sqrt{-5}]$. We now turn out attention to one of the simplest rings in which unique factorization fails. Let $K = \mathbb{Q}(\sqrt{-5})$; then $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$, since $-5 = 3 \bmod 4$.

In this ring, unique factorizations fails because

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

To verify that factorization has failed, we note:

Proposition 7.6 *In the ring $A = \mathbb{Z}[\sqrt{-5}]$, the elements 2, 3 and $1 \pm \sqrt{-5}$ are irreducibles.*

Proof. We will use the norm on A, given by $N(a+b\sqrt{-5}) = a^2+5b^2$. Clearly the equations $a^2 + 5b^2 = 2$ and $a^2 + 5b^2 = 3$ have no integral solutions, since we must have $b = 0$. Thus A has no elements of norm 2 or 3. Now suppose 3 factors as xy in A, and neither x nor y is a unit. Then $N(3) = 9 = N(x)N(y)$, so $N(x) = N(y) = 3$, a contradiction. Similar reasoning applies in the other cases, since $N(2) = 4$ and $N(1 \pm \sqrt{5}) = 6$. ∎

It follows that $\mathbb{Z}[\sqrt{-5}]$ is not a PID. We can also see this directly: the ideal

$$I = (2, 1+\sqrt{-5}) = 2\left(\mathbb{Z} \oplus \mathbb{Z}\frac{1+\sqrt{-5}}{2}\right)$$

is a diamond lattice, while $\mathcal{O}_K$ is a square, so we cannot have $I = (a) = a\,\mathcal{O}_K$.

The ideal class group. This example suggest it is useful to classify the different *shapes of ideals*. In fact the different shapes form a *finite group*, called the *ideal class group* $\mathrm{Cl}(K)$. Let us define this group for an arbitrary number field.

Products of ideals. Let I, J be ideals in a ring A. Their *product* $I \cdot J$ is the additive subgroup generated by

$$IJ = \{ab \,:\, a \in I, b \in J\}.$$

Remarkably, $I \cdot J$ is an ideal for A. In fact, $AI = I$ so $AIJ = AIJ$, and as a general fact we have:

Theorem 7.7 *Suppose $S \subset A$ is a subset of a ring, and $AS \subset S$. Then the additive group I generated by S is an ideal.*

Proof. The group I consists of all finite sums of the form $b = \sum n_i s_i$, with $n_i \in \mathbb{Z}$ and $s_i \in S$. Since $AS \subset A$, for any $a \in A$ the product $ab = \sum n_i(a_i s_i) = \sum n_i s_i'$, with $s_i' \in S$, so it also belongs to I. Thus $AI \subset I$ and hence I is an ideal. ∎

Corollary 7.8 *For any pair of ideals I, J, the additive group $I \cdot J$ generated by IJ is also an ideal; in fact it is the smallest ideal containing IJ.*

Similar considerations allow one to compute additive generators for $I \cdot J$ from generators for I and J.

Theorem 7.9 *Let I, J be ideals in A generated, as additive groups, by $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_m)$ respectively. Then $I \cdot J$ is generated, as an additive group, by the set of all possible products $a_i b_j$.*

Proof. This follows from the fact that

$$IJ = \left(\sum \mathbb{Z}a_i\right)\left(\sum \mathbb{Z}b_j\right) \subset \sum \mathbb{Z}a_i b_j.$$

∎

It is a remarkable fact that the product of ideals I and J in A can be computed using just IJ and addition, and without further reference to multiplication.

Fractional ideals. Now let K be a number field. We say $J \subset K$ is a *fractional ideal* if $J \neq (0)$ and

$$I = aJ \subset \mathcal{O}_K$$

is an ordinary, nonzero ideal, for some $a \in K^*$. Two fractional ideals are *equivalent* if $J_1 = aJ_2$ for some $a \in K^*$.

Note that a fractional ideal satisfies

$$\mathcal{O}_K \cdot I \subset I.$$

The definition of $I \cdot J$ extends naturally to fractional ideals and preserves equivalence classes, since

$$(aI) \cdot (bJ) = ab(I \cdot J).$$

The class group. The *ideal class group* of K, denoted $\mathrm{Cl}(K)$, is the set of equivalence classes of fractional ideals in K, with multiplication defined by $[I] \cdot [J] = [I \cdot J]$. One of the cornerstones of the theory of number fields is:

Theorem 7.10 *For any number field K, multiplication of ideals makes $\mathrm{Cl}(K)$ into a finite abelian group.*

It is not at all obvious that inverses exist in $\mathrm{Cl}(K)$. A proof in the case of complex quadratic fields will be given below.

The *class number* of K is given by

$$h(K) = |\,\mathrm{Cl}(K)|.$$

The fractional ideal $(1) = \mathcal{O}_K$ serves as the *identity element* in $\mathrm{Cl}(K)$. An ideal is equivalent to (1) if and only if it principal, and thus:

The field K has class number $h(K) = 1$ iff $\mathcal{O}_K$ is a PID .

In this case $\mathcal{O}_K$ is also a UFD.

Ideal classes and lattices. Now let us return to the case where K is a complex quadratic field. It is then easy to see:

Ideals I, J are equivalent if and only if they are similar as lattices.

The same is true for fractional ideals. Thus $\mathrm{Cl}(K)$ classifies the different possible shapes of fractional ideals in K.

Ideal classes for $\mathbb{Q}(\sqrt{-5})$. Let us further specialize to the case $K = \mathbb{Q}(\sqrt{-5})$. In this case, we have already found two different fractional ideals:

$$\mathcal{O}_K = \mathbb{Z} \oplus \mathbb{Z}\sqrt{-5} \quad \text{and} \quad J = \mathbb{Z} \oplus \mathbb{Z}\frac{1+\sqrt{-5}}{2}.$$

Note that both these fractional ideals are *normalized lattices*, in fact J is the diamond lattice associated to $\mathcal{O}_K$. Note also that $\mathcal{O}_K$ is an ordinary ideal, but J itself is not; however

$$2J = (2, 1+\sqrt{5}) \subset \mathcal{O}_K$$

is the non–principal ideal discussed before.

Lemma 7.11 *For $\tau = \sqrt{-5}$, the only normalized lattices extending $\mathbb{Z} \oplus \mathbb{Z}\tau$ are given by*

$$L_k = \mathbb{Z} \oplus \mathbb{Z}\frac{\tau + k}{2},$$

where $k = 0$ or 1. The lattice $L_1 = J$ is a fractional ideal, while the lattice L_0 is not.

Proof. Applying Theorem 7.4 with $y = \sqrt{5}$, and noting that $0 < n \leq 2y/\sqrt{3} = \sqrt{20/3} < \sqrt{7} < 3$, we get the two possible extensions L_0 and L_1 above. To see that L_0 is not a fractional ideal, note that $\sqrt{-5} \cdot \sqrt{-5}/2 = -5/2 \notin L_0$. ∎

Theorem 7.12 *Let I be a nonzero ideal in $\mathbb{Z}[\sqrt{-5}]$ with shortest vector a. Then either $I = a\,\mathcal{O}_K$ or $I = aJ$.*

Proof. Replacing I with $(1/a)I$, we can assume that I is a normalized lattice as well as a fractional ideal. In particular, we have $\mathcal{O}_K \cdot 1 \subset I$, and thus I is a normalized lattice extending $\mathcal{O}_K$. By the preceding Lemma, either $I = \mathcal{O}_K$ or $I = J$, completing the proof. ∎

Corollary 7.13 *The ideal class group of* $\mathbb{Q}(\sqrt{-5})$ *is* $\mathbb{Z}/2$.

Computing the square. Let us verify that $[J^2] = [\mathcal{O}_K]$ in $\mathrm{Cl}(K)$. The fractional ideal J is equivalent to the ordinary ideal

$$I = 2J = (2, 1 + \sqrt{-5}).$$

As an additive group, I is generated by $a = 2$ and $b = 1+\sqrt{-5}$. By Theorem 7.9, $I \cdot I = I \cdot \overline{I}$ is generated, as an additive group, by the products

$$\langle a^2, a\bar{b}, ab, b\bar{b}\rangle = \langle 4, 2 + 2\sqrt{-5}, 2 - 2\sqrt{-5}, 6\rangle.$$

Since $6 - 4 = 2$, this group simplifies to:

$$I^2 = 2\mathbb{Z} \oplus 2\sqrt{-5}\mathbb{Z} = 2\,\mathcal{O}_K = (2).$$

Thus $[J^2] = [I^2] = [(2)] = [\mathcal{O}_K]$ in $\mathrm{Cl}(K)$, as expected.

Inversion of ideals. We now come to a critical result in the theory. This result allows us to compute the *inverse* of an ideal in the ideal class group. It is inspired by the fact that the inverse of a complex *unit* u is given by $\pm\overline{u}$.

Theorem 7.14 *For any nonzero ideal* $I \subset \mathcal{O}_K$*, there is an integer* $n > 0$ *such that*

$$I \cdot \overline{I} = (n).$$

In particular, $[\overline{I}]$ *is the inverse of* $[I]$ *in the ideal class group* $\mathrm{Cl}(K)$.

Proof. Choose $a, b \in \mathcal{O}_K$ such that $I = \mathbb{Z}a \oplus \mathbb{Z}b$. Then, by Theorem 7.9, we find that $J = I \cdot \overline{I}$ is generated, as an *additive group*, by

$$|a|^2, |b|^2, a\bar{b}, \quad \text{and} \quad \overline{a}b.$$

Note that $t = a\bar{b} + \overline{a}b = \mathrm{Tr}(a\bar{b})$ is an integer. Let

$$n = \gcd(|a|^2, |b|^2, t).$$

Clearly $n \in J$. We wish to show that $J = (n)$.

For this, it remains only to show that $a\bar{b} \in (n)$. Since J is an ideal for $\mathcal{O}_K$, it suffices to show that $a\bar{b}/n$ is an *algebraic integer*. (This is the key insight!) And for this, we need only verify the statement:

$$\mathrm{Tr}(a\bar{b})/n = t/n \in \mathbb{Z} \quad \text{and} \quad N(a\bar{b})/n^2 = |a|^2|b|^2/n^2 \in \mathbb{Z}.$$

But both these statements are true by the definition of n. ∎

The norm of an ideal. We define the *norm* of an ideal $A \subset \mathcal{O}_K$ by

$$N(A) = n > 0,$$

where $A \cdot \overline{A} = (n)$. Note that for principal ideals, we have

$$N((a)) = |a|^2,$$

so this definition generalizes the norm of elements. Moreover, we have

$$N(A \cdot B) = N(A)N(B),$$

since

$$A \cdot B \cdot \overline{A} \cdot \overline{B} = (A \cdot \overline{A}) \cdot (B \cdot \overline{B}) = (N(A)) \cdot (N(B)) = (N(A)N(B)).$$

Finally, we note that

$$N(A) = 1 \iff A = \mathcal{O}_K,$$

since this condition implies that

$$\mathcal{O}_K = (1) = \overline{A}A \subset A \subset \mathcal{O}_K\,.$$

Theorem 7.15 *For any complex quadratic field K, the ideal classes* $\mathrm{Cl}(K)$ *form a finite group with respect to multiplication of ideals.*

Proof. To check that $\mathrm{Cl}(K)$ is a group, we simply observe that (i) $\mathcal{O}_K \cdot I = I$ for any ideal I, so $[\mathcal{O}_K]$ serves as the identity; (ii) inverses exist by the previous result; indeed, $[I]^{-1} = [\overline{I}]$, since $[(n)] = [\mathcal{O}_K]$; and (iii) ideal multiplication is associative, as is readily verified.

To prove finiteness of the class group, we will use the theory of lattices. First, since 1 is the shortest vector, we can present the ring of integers in K as a normalized lattice:

$$\mathcal{O}_K = \mathbb{Z} \oplus \mathbb{Z}\tau,$$

with $\tau \in F$. (In fact we can take $\tau = \sqrt{-d}$ or $(1 + \sqrt{d})/2$.)

Now let $[I]$ be the class of a fractional ideal in $\mathrm{Cl}(K)$. Dividing by its shortest vector if necessary, we can also assume that I is a normalized lattice. Since I is an ideal containing 1, it gives a lattice extending $\mathcal{O}_K$. By Corollary 7.5, there are only finitely many such lattices, and thus $\mathrm{Cl}(K)$ is finite. ∎

Finale: Factorization into prime ideals. We can now finally achieve our main goal: we will establish a theory of unique factorization of *ideals*.

Theorem 7.16 *Let K be a complex quadratic field. Then every nonzero ideal $I \subset \mathcal{O}_K$ can be factored as a product of prime ideals,*

$$I = P_1 \cdots P_n,$$

in an essentially unique way.

In particular, every proper principal ideal (a) can be factored into prime ideals as above. One can think of the ideals as an enrichment of $\mathcal{O}_K$ to restore unique factorization when it fails in $\mathcal{O}_K$ itself.

Setup. For the remainder of this section, the ambient ring $\mathcal{O}_K$ will be fixed as above. We will use $A, B, C, \ldots$ to denote nonzero ideals, and abbreviate $A \cdot B$ by AB.

The proof of Theorem 7.16 pivots on the fact, already established, that for any proper ideal $A \subset \mathcal{O}_K$ we have

$$A\overline{A} = (n)$$

for a unique $n > 0$ in $\mathbb{Z}$. Recall $N(A) = n$ and $N(AB) = N(A)N(B)$.

Note that $(N(A)) = A\overline{A} \subset A$, and thus

$$N(A) = 1 \iff A = (1) = \mathcal{O}_K\,.$$

In particular $N(P) > 1$ for any prime divisor P. (In fact $N(P) = p$ or p^2 for some prime $p \in \mathbb{Z}$.

Divisibility. Let us write $A|B$ if there exists a C such that $BC = A$. We will establish the following facts about division.

1. If $AB = AC$ then $B = C$.
2. If $P|AB$ and P is a prime ideal, then $P|A$ or $P|B$.
3. Every proper ideal has a prime divisor.
4. If P and Q are prime ideals and $P|Q$, then $P = Q$.

Assuming these facts for the moment, it is easy to complete:

Proof of Theorem 7.16. First we claim any proper ideal A can be factored into prime ideals. Indeed, by (3) above, A has *some* prime divisor P, and if $A = P$ we are done. Otherwise, $A = PB$ with $1 < N(B) < N(A)$, and the proof is completed by induction on $N(A)$.

Uniqueness (up to reordering) then follows as usual, by induction on the length of the factorization. If we have two factorizations of A into prime ideals,

$$P_1 \cdots P_n = Q_1 \cdots Q_m,$$

then $P_1 | Q_i$ for some i, and hence $P_1 = Q_i$; we can then cancel P_1 from both sides, and apply induction to complete the proof. ∎

Division and containment. We now turn to the proof of the basic facts above. We will repeatedly use two facts: $A\overline{A} = (n)$ for some $n > 0$, and prime ideals are maximal (Theorem 6.6).

Theorem 7.17 *If $AB = AC$ then $B = C$*

Proof. Let $n = N(A)$. Then from $AB = AC$ we get

$$\overline{A}AB = nB = \overline{A}AC = nC,$$

and dividing by n gives $B = C$. ∎

Theorem 7.18 *We have $A|B$ if and only if $B \subset A$.*

Proof. If $A|B$ then for some C, we have $B = AC \subset A$, since A is an ideal.

Now assume $B \subset A$. Let us start with the case where $A = (n)$. Then every element of B is divisible by n, so $C = (1/n)B$ is an ideal; and

$$AC = n \cdot (1/n) \cdot B = B,$$

so $A|B$. To handle the general case, we use the fact that if $B \subset A$ then

$$\overline{A}B \subset \overline{A}A = (n),$$

and hence $\overline{A}A | \overline{A}B$; canceling, we get $A|B$. ∎

Theorem 7.19 *If $P|AB$ and P is a prime ideal, then $P|A$ or $P|B$.*

Proof. If $P|AB$ then we have $AB \subset P$. If $A \subset P$ then $P|A$ and we are done, so suppose we have an element $a \in A - P$. Then $aB \subset P$, and hence $B \subset P$, by the definition of a prime ideal; and therefore $P|B$. ∎

Theorem 7.20 *Every proper ideal A is divisible by some prime ideal.*

Proof. We have $A \subset P$ for some maximal ideal P, which is necessarily prime; and $P|A$ by the preceding result. ∎

Theorem 7.21 *If P and Q are prime ideals, and $P|Q$, then $P = Q$.*

Proof. If $P|Q$ then $P \subset Q$ and hence $P = Q$ since P is a maximal ideal. ∎

Factoring primes in $\mathbb{Z}[\sqrt{-d}]$. One way to approach factoring an ideal is to first factor $N(I)$ into ordinary primes, and then factor these primes in $\mathcal{O}_K$. Here is an algorithm for factoring ordinary primes (p) in $\mathbb{Z}[\sqrt{-d}] = \mathcal{O}_K$, $d \neq 1 \bmod 4$.

First, if $x^2 + d = 0$ has no solution $\bmod p$, then (p) is already a prime ideal. In this case

$$\mathcal{O}_K/(p) \cong \mathbb{F}_p[x]/(x^2 + d),$$

which is a field since $x^2 + d$ is irreducible.

Otherwise, $a^2 + d = 0 \bmod p$ for some a, and we can write $p = I\overline{I}$, with $I = (a + \sqrt{-d}, p)$. In this case, $x^2 + d = (x + a)(x - a) \bmod p$, and

$$\mathcal{O}_K/(p) = \mathbb{F}_p[x]/(x^2 + d, x + a) = \mathbb{F}_p[x]/(x - a) \cong \mathbb{F}_p.$$

Factoring in $\mathbb{Z}[i]$. Note that when $\mathcal{O}_K$ is has class number one, we can then proceed further to write $I = (b)$, and then $b \doteq p$.

Here is an example in $\mathbb{Z}[i]$. Let $p = 29$. Since $p = 1 \bmod 4$, we know that $x^2 + 1$ has a root in $\mathbb{F}_p$, but what is it? One guess is to try $a^{(p-1)/4}$ for various a. And in fact $2^7 = 128 = 12 \bmod 29$ satisfies $12^2 = -1 \bmod 29$. This $p = I \cdot \overline{I}$, where

$$I = (29, 12 + i).$$

We now use division to find a generator. We can write $29 = 2(12+i)-2i+5$, so

$$I = (12 + i, 5 - 2i).$$

When the notice that $N(5 - 2i) = 5^2 + 2^2 = 29$, so in fact $I = (5 - 2i)$.

Intermission

Coda: The meaning of $N(A)$ and $N(a)$. It is natural to wonder what the integer $N(A)$ measures about an ideal, and it turns out its meaning is purely 'geometric'. That is, $N(A)$ can be expressed using only the additive group $(\mathcal{O}_K, +)$: we have

$$N(A) = [\mathcal{O}_K : A] = \frac{\text{area}(\mathbb{C}/A)}{\text{area}(\mathbb{C}/\mathcal{O}_K)}.$$

This is not obvious from the line of development we have followed. It can be proved by first studying the case of prime ideals, and then using the factorization theorem.

The case of a *principal ideal* is simpler. Indeed, multiplication by $a \in \mathcal{O}_K$ makes sense not just on K but also on $\mathbb{C} \cong \mathbb{R}^2$. If we write $a = x + iy$ use $(1, i)$ as a basis, and regard a as a linear map T_a, then

$$T_a = \begin{pmatrix} x & -y \\ y & x \end{pmatrix},$$

and $\det(T_a) = x^2 + y^2 = N(a)$.

Appendix: Computing the class group. Let d be a square–free integer, $d < 0$, and let

$$\epsilon(d) = \begin{cases} 1 & \text{if } d = 1 \bmod 4, \text{ and,} \\ 2 & \text{otherwise.} \end{cases}$$

As we have see, for $K = \mathbb{Q}(\sqrt{d})$ we have $\mathcal{O}_K = \mathbb{Z} \oplus \mathbb{Z}\tau$, with $\tau = \sqrt{d}$ or $(1 + \sqrt{d})/2$. Thus $\text{Im}(\tau) = \epsilon(d)|\sqrt{d}|/2$.

on extensions of $\mathcal{O}_K = \mathbb{Z} \oplus \mathbb{Z}\tau$ allows one to generate all elements of the class group effectively and without repetition. Namely, with τ as above, we examine the finite list of candidate fractional ideals $I = \mathbb{Z} \oplus \mathbb{Z}\sigma$, $\sigma \in F$, $\sigma = (\tau + k)/n$. To check if I is actually a fractional ideal, it suffices to check if $\tau I \subset I$.

This argument shows:

Theorem 7.22 *For $d < 0$ a square free integer, every ideal class for $\mathbb{Q}(\sqrt{d})$ is represented by an ideal I with*

$$(n) \subset I \subset (1),$$

where $n \leq \epsilon(d)\sqrt{|d/3|}$.

Proof. Every ideal class is represented by a normalized lattice J, containing 1 and hence containing $\mathcal{O}_K$. By Theorem 7.4, there exists an n and k such that $J = \mathbb{Z} \oplus \mathbb{Z}\sigma$, with $\sigma = (\tau + k)/n \in F$, and hence

$$\frac{\mathrm{Im}(\tau)}{n} = \frac{\epsilon(d)|\sqrt{d}|}{2n} \geq \frac{\sqrt{3}}{2}.$$

We then have

$$\mathcal{O}_K \subset J \subset (1/n)\,\mathcal{O}_K,$$

since $1 \in J$ and $\tau \in \mathcal{O}_K$. Setting $I = nJ$ gives the desired representative. ■

Since $(n) \subset I$ exactly when $I|(n)$, we have:

Corollary 7.23 *The prime ideals in $\mathcal{O}_K$ occurring as the factors of the ordinary primes $p \leq \epsilon(d)\sqrt{|d/3|}$ generate* $\mathrm{Cl}(K)$.

Example: $d = -2n$. For $d < 0$ it is rare for $K = \mathbb{Q}(\sqrt{d})$ to have class number one. (All cases will be listed below.)

Let us give an indication of this fact by showing that $h(K) > 1$ whenever $d > 2$ is *even*, that is when $d = -2n$ and $n > 1$ is a square–free odd number. In this case $\mathcal{O}_K = \mathbb{Z}[\sqrt{-d}]$. We give two proofs that $\mathcal{O}_K$ is not a PID.

(i) We observe that $\mathcal{O}_K/(2) \cong \mathbb{F}_2[x]/(x^2+2n) \cong \mathbb{F}_2[x]/(x^2)$ is not a field. Thus (2) can be factored. If K is a UFD, we can write $2 = ab$ as a product of two primes in $\mathcal{O}_K$, with $N(a) = N(b) = 2$. But we have

$$N(x + y\sqrt{d}) = x^2 + dy^2,$$

and since $d > 2$ the equation $N(a) = 2$ has no solutions!

(ii) We observe that $J = \mathbb{Z} \oplus \mathbb{Z}\sqrt{2n}/2$ is a fractional ideal with a different shape than $\mathcal{O}_K$. It is an ideal since $\sqrt{2n} \cdot J \subset J$, because $\sqrt{2n} \cdot \sqrt{2n}/2 = n \in \mathbb{Z}$. Thus $h(K) \geq 2$.

Example: $d = -163$. Let us show that $h(K) = 1$ when $K = \mathbb{Q}(\sqrt{-163})$. Since $-163 = 1 \bmod 4$,

$$\mathcal{O}_K = \mathbb{Z} \oplus \mathbb{Z}\tau,$$

where $\tau = (1 + \sqrt{-163})/2$ has $N(\tau) = 41$ and $\mathrm{Tr}(\tau) = 1$. Thus

$$\mathcal{O}_K \cong \mathbb{Z}[x]/(x^2 - x + 41).$$

By the Corollary above, $\mathrm{Cl}(K)$ is generated by the factors of (p) where $p = 2, 3, 5$ and 7, since $\sqrt{163/3} \leq \sqrt{55} < 8$. For each of these primes, once can quickly check that $x^2 - x + 41$ has no root mod p.

Indeed, the values of this polynomial for $0 \leq x < 7$ are $(41, 41, 43, 47, 53, 61, 71)$, and all of these are prime (!). So (p) is already a prime ideal, and hence every ideal in $\mathcal{O}_K$ is principal.

A formula for primes? In fact, $p(n) = n^2 - n + 41$ is prime for all n with $0 \leq n < 41$. The first nonprime value $p(41) = 1681 = 41^2$. (Note that $p(1 - n) = p(n)$, so negative values work as well.)

Why is this the case? Since

$$p(n) = (n - \tau)(n - \overline{\tau}),$$

the number–theoretic meaning of this polynomial is:

$$p(n) = N(\tau - n).$$

Theorem 7.24 *The algebraic integer $\tau + n$ is a prime in $\mathcal{O}_K$ for all n, $0 \leq n < 41$.*

Proof. Note that $N(\tau) = 41$ and τ minimizes both $N(a)$ and $|\operatorname{Im} a|$ over all $a \in \mathcal{O}_K - \mathbb{Z}$. Now suppose $\tau + n$ factors nontrivially as ab in $\mathcal{O}_K$. Since τ minimizes $|\operatorname{Im} a|$, we cannot have $a \in \mathbb{Z}$ or $b \in \mathbb{Z}$. Thus

$$p(n) = N(\tau + n) = N(a)N(b) \geq 41^2.$$

But $p(n) < 41^2$ for $0 \leq n < 41$; thus $\tau + n$ is prime. ∎

Corollary 7.25 *The value of $p(n)$ is prime for $0 \leq n < 41$.*

Proof. Since $\tau - n$ is prime, by Theorem 6.10, there exists an ordinary prime p such that $p(n) = N(\tau - n) = p$ or p^2. The second case can only occur when $\tau - n$ is an associate of p, which is impossible since the only units in $\mathcal{O}_K$ are ± 1. ∎

Class number one. The complete list of complex quadratic fields with class number one is known. There are nine:

$$h(K) = 1 \iff d = -1, -2, -3, -7, -11, -19, -43, -67, \text{ or } -163.$$

Aside from -1 and -2, these are all congruent to $1 \bmod 4$, and we have the case of a diamond lattice. The same reasoning we have applied to $d = -163$ also shows, for example, that $p(n) = n^2 - n + 17$ is prime for $0 \leq n < 17$, by consider the case $d = -67$.

For reference, we also note:

$$\begin{aligned} h(K) &= 2 \quad \text{if} \quad d = -5, -10, -13, -15, -22, \quad \text{and} \\ h(K) &= 4 \quad \text{if} \quad d = -17, -21. \end{aligned}$$

Aside from the cases of class number one, $\mathcal{O}_K$ is not a UFD; in fact:

Theorem 7.26 *For a complex quadratic field, $\mathcal{O}_K$ is a UFD iff it is a PID.*

Proof. Suppose $\mathcal{O}_K$ is a UFD. We will show every nonzero prime ideal $P \subset \mathcal{O}_K$ is principal. To this end, let $a = p_1 \cdots p_n$ be the prime factorization of some nonzero element $a \in P$. Then $p_i \in P$ for some i, by the definition of a prime ideal, and thus $(p_i) \subset P$. But (p_i) is also a maximal ideal. Thus $P = (p_i)$ is principal.

Since every proper ideal in $\mathcal{O}_K$ is a product of prime ideals, $\mathcal{O}_K$ is a PID. The converse is a general fact. ∎

The same result holds for an arbitrary number field.

PID but not Euclidean. As mentioned above, for $K = \mathbb{Q}(\sqrt{-19})$ the ring $\mathcal{O}_K$ is a PID. However it is *not* a Euclidean domain.

To see this, recall that $\mathcal{O}_K = \mathbb{Z}[\tau]$ where $\tau = (1 + \sqrt{-19})/2$ is a root of $p(x) = x^2 - x + 5$. Moreover, the units in $\mathcal{O}_K$ are just ± 1.

Suppose this ring is a Euclidean function with size function $\sigma(a)$. One can show there exists an $a \in \mathcal{O}_K$, not a unit, such that $\sigma(b) < \sigma(a)$ implies that $b \in \{0, 1, -1\}$.

Now consider the ring $A = \mathcal{O}_K / (a)$. By the division algorithm, $\{0, 1, -1\}$ maps onto A, so $|A| = 2$ or $|A| = 3$, and in fact $A = \mathbb{F}_2$ or $\mathbb{F}_3$. But $p(x)$ has no root in these fields — a contradiction.

Appendix: Complex multiplication. There is a close connection between complex quadratic number fields K and elliptic curves. Namely, every fractional ideal $I \subset K$ determines a elliptic curve

$$E = \mathbb{C}/I$$

with $\mathrm{End}(E) = \mathcal{O}_K$. The action of $a \in \mathcal{O}_K$ on E is given simply by $z \mapsto az$, which is well–defined because $aI \subset I$. One says that E admits *complex multiplication* (CM) by $\mathcal{O}_K$.

Note that equivalent ideals give isomorphic elliptic curves. Thus the class number $h(K)$ describes the number of distinct isomorphism classes of elliptic curves with (CM) by $\mathcal{O}_K$.

Appendix: What else is special about 163? The observation the $\mathbb{Q}(\sqrt{-163})$ has class number one underlies the miraculous fact that $\exp(\pi\sqrt{163})$ is almost in $\mathbb{Z}$. Indeed,

$$\exp(\pi\sqrt{163}) \approx 2.625 \times 10^{17}$$

differs from an integer by less than 10^{-12}.

An explanation of this fact goes through the connection with elliptic curves, and uses the analytic modular function $j(\tau)$ on $\mathbb{H}$. It is known that $j(\tau)$ is an algebraic integer at the points $\tau \in \mathbb{H}$ such that $\mathbb{Z}[\tau]$ is an ideal in the ring of integers in a quadratic imaginary field K. In fact the degree of $j(\tau)$ over $\mathbb{Q}$ is the class number $h(K)$. In particular, if $h(K) = 1$ (as is the case for $d = 163$), then $j(\tau)$ is an integer. This holds when $\tau = (1 + \sqrt{-163})/2$, since $\mathbb{Z}[\tau] = \mathcal{O}_K$.

Now the j-function has an expansion

$$j(q) = q^{-1} + 744 + O(q)$$

where $q = \exp(2\pi i\tau)$. Setting $\tau = (1+\sqrt{-163})/2$, we obtain $q = -\exp(-\pi\sqrt{163})$. Plugging into the power series of j above, we have that $j(q)$ is an integer, and at the same time $j(q)$ is very close to $1/q + 744$. This explains why $\exp(\pi\sqrt{-163})$ is very nearly a (huge) integer N. In fact, it shows that the deviation from N is on the order of $1/N$ itself.

Sequel: Real quadratic fields. We now briefly turn to a study of fields of the form $K = \mathbb{R}(\sqrt{d}) \subset \mathbb{R}$, where d is a *positive* square-free integer.

To illustrate a significant difference between the real and complex cases, the main result we will prove is:

Theorem 7.27 *For K a real quadratic field, the multiplicative group of units in $\mathcal{O}_K$ is isomorphic to $\mathbb{Z} \times (\mathbb{Z}/2)$.*

The $\mathbb{Z}/2$ factor comes from (± 1). One should contrast with the complex case, where U was a torsion group.

Norm and trace. Recall that we are considering K as a subfield of $\mathbb{R}$, and that $a \mapsto a'$ denote the Galois automorphism of $K/\mathbb{Q}$, sending $\sqrt{d}$ to $-\sqrt{d}$.

For real quadratic fields we have a natural homomorphism of rings,

$$\iota : K \to \mathbb{R}^2,$$

defined by $\iota(a) = (a, a')$. In these coordinates, if we define the norm and trace on $\mathbb{R}^2$ by

$$N(x, y) = xy \quad \text{and} \quad \mathrm{Tr}(x, y) = x + y,$$

then these functions restrict to give the usual norm and trace on K.

Lattices. The embedding ι provides us with a geometric realization of the ring of integers as a lattice in $\mathbb{R}^2$; explicitly, we have

$$\mathcal{O}_K \cong \mathbb{Z}(1, 1) \oplus ((1 + \sqrt{d})/2, (1 - \sqrt{d})/2)$$

when $d = 1 \bmod 4$, and otherwise

$$\mathcal{O}_K \cong \mathbb{Z}(1, 1) \oplus (\sqrt{d}, -\sqrt{d}).$$

This embedding of $\mathcal{O}_K$ into $\mathbb{R}^2$ in the real case takes the place of the embedding of $\mathcal{O}_K$ into $\mathbb{C}$ in the complex case.

Units. The unit group of $\mathcal{O}_K$ consists of those elements with $N(a) = \pm 1$. Clearly $\mathcal{O}_K^\times$ contains ± 1, and any other element of U has infinite order.

Proposition 7.28 *The units in $\mathcal{O}_K$ form a discrete subgroup of $\mathbb{R}^\times$.*

Proof. Let H be the locus in $\mathbb{R}^2$ defined by $xy = \pm 1$. Then $H \cap \mathcal{O}_K$ is a discrete set, consisting of the points (a, a') such that a is a unit. Since the projection $H \to \mathbb{R}^*$ given by $(x, y) \mapsto x$ is proper, the image of $H \cap \mathcal{O}_K$ in $\mathbb{R}^*$ is also discrete. ∎

Fundamental units. The unique generator $\epsilon_d > 1$ for $\mathcal{O}_D^\times/(\pm 1)$ is called the *fundamental unit* for K. For example,

$$1+\sqrt{2}, 2+\sqrt{3}, (1+\sqrt{5})/2 \quad \text{and} \quad 5+2\sqrt{6}$$

are all fundamental units in the fields they generate. Remarkably,

$$2143295 + 221064\sqrt{94}$$

is also a fundamental unit. It is known that

$$\log \epsilon_d = O(\sqrt{d} \log d),$$

and this bound is approximately the right size when $|\operatorname{Cl}(K)| = 1$.

Class numbers in the real quadratic case. It is a major open problem to show to establish:

Conjecture 7.29 *There are infinitely many real quadratic fields with class number one.*

Nevertheless there is strong numerical and theoretical evidence (Cohen–Lenstra) that about 75% of the fields of the form $\mathbb{Q}(\sqrt{p})$, with p prime, have class number one. The field $\mathbb{Q}(\sqrt{d})$ has class number one (and hence is a UFD) for

$$d = 2, 3, 5, 6, 7, 11, 13, 14, 17, 19, 21, 22, 23, 29, 31, 33, 37, 38, 41, 43, \ldots$$

Quadratic forms. We note that the norm forms in the real and complex cases, given by

$$N(x,y) = xy \quad \text{and} \quad N(x+iy) = x^2 + y^2,$$

have signature $(1,1)$ and $(2,0)$ respectively. The difference between the unit groups in the real and complex cases can be traced to the fact that $\mathrm{SO}(2)$ is compact, with $\mathrm{SO}(1,1)$ is not.

Existence of units. We now turn to the proof of Theorem 7.27. We begin with a fact about lattices in $\mathbb{R}^2$. Given $a, b > 0$, we let

$$R(a,b) = [-a,a] \times [-b,b] \subset \mathbb{R}^2.$$

Theorem 7.30 *Suppose $L \subset \mathbb{R}^2$ is a lattice, and*

$$ab \geq (2/\sqrt{3})\,\mathrm{area}(\mathbb{R}^2/L).$$

Then the rectangle $R(a,b)$ contains a point $(x,y) \in L$, $(x,y) \neq (0,0)$.

Proof. By applying a linear change of coordinates, we can reduce to the case where $a = b$ and the shortest vector in L has length 1. Then by Corollary 7.3, we have

$$\mathrm{area}(\mathbb{R}^2/L) \geq \sqrt{3}/2 = \mathrm{area}(\mathbb{C}/\mathbb{Z}[\omega]).$$

Thus $a^2 \geq 1$ and hence the square $R(a,a)$ contains the shortest vector in L. ∎

Theorem 7.31 *For any real quadratic field K, there exists an n such that $N(a) = n$ for infinitely many $a \in \mathcal{O}_K$.*

Proof. Choose r such that $r^2 \geq (2/\sqrt{3})\,\mathrm{area}(\mathbb{R}^2/\mathcal{O}_K)$. For each integer $k > 0$, let $R_k = R(r/k, k/r)$. By the preceding result, every rectangle R_k contains a point $(a_k, a_k') \in \mathcal{O}_K$ with $a_k \neq 0$. Since $|a_k| \leq r/k \to 0$, the sequence (a_k) contains infinitely many distinct elements. Since the integer $|N(a_k)|$ is bounded by r^2 there are also infinitely many distinct $a_k \in \mathcal{O}_K$ with the same norm n. ∎

Theorem 7.32 *If $a, b \in \mathcal{O}_K$, $N(a) = N(b) = n$, and $a = b \bmod (n)$, then a/b is a unit in $\mathcal{O}_K$.*

Proof. Note that $N(a/b) = N(a)/N(b) = 1$. Thus to show a/b is a unit, it suffices to show that a/b is an algebraic integer. Having already tested the norm, it remains only to check that

$$\mathrm{Tr}(a/b) = \frac{ab' + a'b}{bb'} = \frac{\mathrm{Tr}(ab')}{N(b)}$$

is an integer. But $a - b \in (n)$, so

$$N(a-b) = N(a) + N(b) - \mathrm{Tr}(ab') \in (n).$$

Since $N(a) = N(b) = n$, this gives $\mathrm{Tr}(ab') \in (n)$ and hence $\mathrm{Tr}(a/b) \in \mathbb{Z}$. ∎

Proof of Theorem 7.27. Since the ring $\mathcal{O}_K/(n)$ is finite, once one has infinitely many solutions to $N(a) = n$, one can find infinitely many solutions in the same residue class mod (n). Their ratios then provide infinitely many units in $\mathcal{O}_K$.

Since the unit group is a discrete subset of $\mathbb{R}^*$, containing (± 1), it is isomorphic to $\mathbb{Z}/2 \times \mathbb{Z}$. ∎

Proof 2: Right and left cosets. Here is a more sophisticated proof of the existence of units. Let $G = \mathrm{SL}_2(\mathbb{R})$, let $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, and let $A \subset G$ be the diagonal subgroup. Choose $M \in \mathrm{SL}_2(\mathbb{R})$ such that $M(\mathbb{Z}^2) = \lambda\,\mathcal{O}_K$, $\lambda > 0$. Then our task is to show that $A \cdot [M] \subset \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$ is compact, for in this case $[M]$ generates a closed geodesic, and its stabilizer in $A \cong \mathbb{R}$ is the desired group of units.

Since $N(x, y) = xy$ is invariant under A, it is easy to see that $A \cdot [M]$ is bounded in the space of lattices G/Γ. On the other hand, since $N(x, y)$ only assumes integral values on $\mathcal{O}_K$, it is easy to see that $[M] \cdot \Gamma$ is discrete in the space of quadratic forms $A\backslash G$. It follows that $A \cdot [M]$ is closed, and hence compact, completing the proof.

Quadratic equations in the integers. We now return to the norm form. Assume that $d > 0$, $d \neq 1 \bmod 4$, and $K = \mathbb{Q}(\sqrt{d})$ has class number one. Then the equation

$$a^2 - db^2 = n$$

can be profitably studied using factorization in

$$\mathcal{O}_K = \mathbb{Z}[\sqrt{d}]$$

and Theorem 6.10, just as we did for $\mathbb{Z}[i]$. That is, we rely on the fact that

$$N(a + b\sqrt{d}) = a^2 - db^2.$$

One difference in the real case is that the norm can be negative. So, for example, when a prime p *splits* in K, we are only guaranteed that we can solve

$$a^2 - db^2 = \pm p.$$

Indeed, provided d is odd, only one sign will be possible.

Example. Let us try to solve

$$a^2 - 11b^2 = \pm 7.$$

The left hand side is never $3 \bmod 4$, so the only possible sign is -7. Since $11 = 2^2 \bmod 7$, the prime 7 splits, and hence we have a solution; indeed $2^2 - 11 \cdot 1^2 = -7$.

On the other hand, $11 = 2 \bmod 3$ is not a square in $\mathbb{F}_3$, so the equation

$$a^2 - 11b^2 = \pm 3$$

has no solutions.

8 Fields

We now turn to the study of fields. Our motivating interest will again be applications to number fields, that is fields with $\mathbb{Q} \subset K \subset \mathbb{C}$ and $[K : \mathbb{Q}] < \infty$.

As highlights, we will use field theory to show that certain geometric constructions are *impossible*; for example, there is no way to construct a regular 7–sided polygon using a ruler and compass. We will also address the question of solving polynomials by *radicals*, and in particular explain the *unsolvability of the quintic.*

A fundamental insight, going back to Galois, is that the secrets of a field extension L/K are unlocked by studying its symmetry group $G = \mathrm{Gal}(L/K)$, the group of automorphism of L that fix K pointwise. Indeed, for suitable extensions, we will find a bijection between the subfields of L and the subgroups of G.

Examples of fields. To set the stage, we illustrate some of the many types of fields that can arise.

1. The most fundamental field is $\mathbb{Q}$, the field of fractions of $\mathbb{Z}$.

2. The fields $\mathbb{R}$ is obtained from $\mathbb{Q}$ by taking its *completion* with respect to the usual absolute value, $|x| = \max(x, -x)$.

3. The field $\mathbb{C} = \mathbb{R}[i]$ is an algebraic extension of $\mathbb{R}$ of degree two. It has the important property that it is algebraically closed. That is, every polynomial in $\mathbb{C}[x]$ has a root in $\mathbb{C}$.

4. The set of all algebraic numbers $\overline{\mathbb{Q}}$ also forms an algebraically closed subfield of $\mathbb{C}$.

5. Given finitely many algebraic numbers $t_1, \ldots, t_n \in \overline{\mathbb{Q}}$, one can consider the *number field* $K = \mathbb{Q}(t_1, \ldots, t_n)$. All finite extensions of $\mathbb{Q}$ are obtained in this way. (In fact, a single element will do.)

6. For each prime we have the finite field $\mathbb{F}_p$. In fact for every power of p, $q = p^n$, there is a unique field $\mathbb{F}_q$ with q elements.

7. The field of rational functions $\mathbb{C}(x)$ is called a *function field*; its elements can be thought of as functions on $\mathbb{C}$ (with poles allowed).

   Similarly, $\mathbb{C}(x_1, \ldots, x_n)$ is the function field for $\mathbb{C}^n$.

8. Recall that an irreducible algebraic variety $V \subset \mathbb{C}^n$ corresponds to a prime ideal $I(V)$, with associated ring of polynomial functions $\mathcal{O}(V) = A/I(V)$. The field of fractions $K(V)$ of $\mathcal{O}(V)$ is called the *function field* of V. Its element are represented by ratios of polynomials, $f(x) = p(x)/q(x)$, with $q|V$ not identically zero.

9. Let $\mathbb{C}[[t]]$ denote the ring of formal power series, whose elements are infinite sums of the form

$$f(t) = \sum_0^\infty a_n t^n$$

   $a_n \in \mathbb{C}$. This ring can be obtained as the completion of $\mathbb{C}[t]$ with respect to the metric defined by $d(p, q) = 2^{-n}$ if $p(x) - q(x) = t^n r(x)$, where $r(0) \neq 0$.

   The units in $\mathbb{C}[[t]]$ consists of those series with $f(0) = a_0 \neq 0$. For example,

$$(1 - t)^{-1} = \sum_0^\infty t^n.$$

   Because of this, the field of fractions of $\mathbb{C}[[t]]$ consists of the formal *Laurent series*,

$$f(t) = \sum_{-N}^\infty a_n t^n.$$

   This field is usually denoted $\mathbb{C}((t))$.

10. The ring of p–adic integers $\mathbb{Z}_p$, similarly, consists of all formal sums of the form

$$a = \sum_0^\infty a_n p^n,$$

with $0 \le a_n < p$. It arises as the completion of $\mathbb{Z}$ with respect to a suitable absolute value, which satisfies $|k|_p = 1$ if $\gcd(k,p) = 1$, and $|p^i|_p = p^{-i}$. Any integer relatively prime to p is a *unit* in $\mathbb{Z}_p$.

One can think of an element of $\mathbb{Z}_p$ as a number written in base p, but with digits that can continued infinitely far to the left. Thus in $\mathbb{Z}_2$,

$$-1 = \ldots 11111. \quad \text{and} \quad -1/3 = 1/(1-4) = 1+4+4^2+\cdots = \ldots 1010101.$$

(Check that $(-1/3) + 2(-1/3) = 1$.)

The field of fractions of $\mathbb{Z}_p$ is the field of p–adic numbers, $\mathbb{Q}_p$. Its elements are similary 'Laurent series' in p, i.e. they have the form

$$a = \sum_{-N}^\infty a_n p^n.$$

These can be thought of as base p numbers that are allowed to have finitely many digits to the *right* of the decimal point.

We note that $p\mathbb{Z}_p$ is the unique maximal ideal in $\mathbb{Z}_p$, and $\mathbb{Z}_p/p\mathbb{Z}_p \cong \mathbb{F}_p$. This quotient is called the *residue field* of $\mathbb{Z}_p$.

Field extensions. Let $f : K \to L$ be a ring homomorphism between fields. Any such map is injective, so we can consider K as a subfield of L. Thus the study of *field extension* is fundamental to the theory. The notation L/K will indicate that L is a field extension of K.

Degree. Let L/K be field extension. Then L can be regarded as a vector space over K; its dimension is called the *degree* of L/K, written:

$$[L : K] = \dim_K(L).$$

The degree may be infinite. For example, $[\mathbb{R} : \mathbb{Q}] = \infty$, while $[\mathbb{Q}(2^{1/n}) : \mathbb{Q}] = n$.

Theorem 8.1 *Given a triple of fields $K \subset L \subset M$, we have*

$$[M : K] = [M : L][L : K].$$

Proof. Let $a_i \in L$ be a basis for L/K, and $b_j \in M$ be a basis for M/L. Clearly the set of products $c_{ij} = a_i b_j$ spans M over K. Moreover, the elements c_{ij} are linearly independent over K, since any relation $\sum k_{ij} a_i b_j = 0$ with $k_{ij} \in K$ can be regrouped to give a relation $\sum_j (\sum_i k_i j a_i) b_j = 0$ between the basis elements (b_j), with coefficients in L. ∎

Algebraic numbers. We say $t \in L$ is *algebraic* over K if it the zero of an irreducible, monic polynomial $p(x) \in K[x]$. In this case $p(x)$ is the *minimal polynomial* of t. The smallest extension of K containing t is denoted by $K(t) \subset L$; and we have

$$K(t) \cong K[x]/(p(x)).$$

Degree of an algebraic number. We let

$$\deg_K(t) = \deg p(x) = [K(t) : K].$$

By consider the tower of fields $L/K(t)/K$, we find:

Corollary 8.2 *If $t \in L$ then $\deg_K(t)$ divides $[K : L]$.*

Corollary 8.3 *If $[L : K]$ is prime, then $K(s) = L$ for any $s \in L$ that is not in K.*

Different roots. Suppose an irreducible polynomial $p(x)$ has two different roots, s and t in L. We then have

$$K(t) \cong K(s)$$

over K, because both fields are isomorphic to $K[x]/(p(x))$. For example, if ω is a primitive cube root of 1, then

$$K = \mathbb{Q}(2^{1/3}) \cong \mathbb{Q}(\omega \cdot 2^{1/3}),$$

as both are roots of the irreducible polynomial $p(x) = x^3 - 2$.

In the first case K is a subfield of $\mathbb{R}$; in the second case, it contains complex numbers. Thus K itself does not 'know' if it is a real or complex field.

Bases and inverses. If $\deg_K(t) = d$, then $1, t, t^2, \ldots, t^{d-1}$ is a basis for $K(t)$ over K. Thus given any $q(x) \in K(x)$ such that $q(t) \neq 0$, there must be another polynomial $r(x)$ such that

$$1/q(t) = 1/(a_0 + a_1 t + \cdots a_{d-1} t^{d-1}) = r(t) = b_0 + b_1 + \cdots + b_{d-t} t^{d-1}.$$

How can one find $r(t)$? We use the fact that $p(x)$ and $q(x)$ are relatively prime in $K[x]$; thus there exist $r(x)$ and $s(x)$ such that

$$q(x)r(x) + p(x)s(x) = 1.$$

This factorization can be found using the Euclidean algorithm. Then clearly $r(x)q(x) = 1$ in $K[x]/(p(x))$.

Example of inversion. Suppose $t \in \mathbb{C}$ is a root of $p(x) = x^3 - x - 1$. We can then find $(1+t)^{-1}$ by noting that, by long division:

$$(t+1)(t-1)t = (t^3 - t - 1) + 1,$$

and hence $1/(t+1) = t(t-1) = t^2 - t$.

The quadratic case. We remark that inversion is easier in quadratic fields such as $K = \mathbb{Q}(\sqrt{d})$, for in this case one can 'rationalize the denominator' to obtain:

$$\frac{1}{a + b\sqrt{d}} = \frac{a - b\sqrt{d}}{a^2 - db^2}.$$

New fields from polynomials. Let us now shift perspective and an issue that arises already with $\mathbb{Q}$. Suppose we have a polynomial $p(x) \in \mathbb{Q}[x]$, with no root in $\mathbb{Q}$. We could like to extend the rationals to a large field where $p(x)$ *does* have a root; for example, we might want to adjoin $2^{1/3}$ or a root of $x^3 = x + 1$. Historically, the need to solve polynomials was one of the primary motivations for constructing fields larger than $\mathbb{Q}$.

Kronecker showed that such an extension is always possible.

Theorem 8.4 *Let K be a field and let $p \in K[x]$ an irreducible polynomial. Then $L = K[x]/(p(x)$ is an extension field of K, and $p(x)$ has a root in L.*

Proof. Since p is irreducible, $(p(x))$ is a maximal ideal, and hence L is a field extending K. The image of x in L gives a root of p. ∎

From this perspective, fields grow upwards from a given field K, by adding solutions to algebraic equations.

The matrix model. Here is a more explicit description of the new field L built from K and $p(x) \in K[x]$. Given an irreducible polynomial

$$p(x) = a_0 + a_1 x + \cdots + \cdots + x^d$$

in $K[x]$, its *companion matrix* $M(p) \in \mathrm{M}_d(K)$ is given by

$$A(p) = \begin{pmatrix} 0 & 0 & 0 & \dots & -a_0 \\ 1 & 0 & 0 & \dots & -a_1 \\ 0 & 1 & 0 & \dots & -a_2 \\ & & \dots & & \\ 0 & 0 & 0 & \dots & -a_{d-1} \end{pmatrix}.$$

There is a natural map $K \hookrightarrow M_d(K)$ sending s to sI, where I is the identity matrix. We then have:

Theorem 8.5 *The field $K[x]/(p(x))$ is isomorphic to the subring of $M_d(K)$ generated by $K \cdot I$ and the companion matrix $A(p)$.*

Example. The companion matrix of $p(x) = x^2 - 2$ is given by $A(p) = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$; note that $A(p)^2 = 2$. From this it is not hard to check that

$$\mathbb{Q}(\sqrt{2}) \cong \mathbb{Q}(A(p)) \subset \mathrm{M}_2(\mathbb{Q}).$$

Proof of Theorem 8.5. Note that $t \in L = K[x]/(p(x)))$ acts by a K–linear mapping on the d–dimensional vector space L/K. With respect to the basis $(1, t, \dots, t^{d-1})$, this mapping has matrix $A(p)$. Indeed, the first $d-1$ columns reflect the fact that $t \cdot t^i = t^{i+1}$, and the last column records the fact that

$$t^d = -a_0 - a_1 t - \dots - a_{d-1} t^{d-1}.$$

■

Example. We note that $p(x)$ can be recovered from $A(p)$ as its characteristic polynomial:

$$p(x) = \det(xI - A(p)).$$

Similarly one can find the minimal polynomials for other elements of $K(t)$ by factoring the characteristic polynomial of the corresponding matrix.

What is the minimal polynomial for $s = 2^{1/3} + 3 \cdot 2^{2/3} + 1$? To answer this question we work in the field $\mathbb{Q}(t)$, $t = 2^{1/3}$, with minimal polynomial $p(x) = x^3 - 2$ and

$$A(p) = \begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Then $s = 3t^2 + t + 1$ corresponds to the matrix

$$B = 3A(p)^2 + A(p) + I = \begin{pmatrix} 1 & 6 & 2 \\ 1 & 1 & 6 \\ 3 & 1 & 1 \end{pmatrix},$$

and its characteristic polynomial

$$\det(xI - B) = x^3 - 3x^2 - 15x - 93$$

gives the characteristic polynomial for s over $\mathbb{Q}$.

Norm and trace. Let $\rho : K(t) \to \mathrm{M}_d(K)$ as above. The quantities $N(s) = \det \rho(s)$ and $\mathrm{Tr}(s) = \mathrm{Tr}\, \rho(s)$ generalize the norm and trace we have used previously in the case of quadratic fields. In particular, $N(t)$ is the product of the roots of $p(x)$, and $\mathrm{Tr}(t)$ is their sum.

Splitting fields. A monic polynomial $p(x) \in K[x]$ *splits* in an extension L/K if it factors into linear terms in $L[x]$. If, in addition, L is generated by the roots of $p(x)$, then L is called the *splitting field* of p.

For $K = \mathbb{Q}$, one can take

$$L = \mathbb{Q}(t_1, \ldots, t_d) \subset \mathbb{C},$$

where (t_i) are the roots of $p(x)$ in $\mathbb{C}$.

Theorem 8.6 *Every monic polynomial $p(x) \in K[x]$ has a splitting field L of finite degree over K. Any two splitting fields for $p(x)$ are isomorphic over K.*

Proof. The proof is by induction on $\deg p$. For $\deg p = 1$ we take $L = K$. For $\deg p > 1$, let q be an irreducible factor of p; then p has a root $t \in L = K[x]/q(x)$ (see Theorem 8.4). Removing the factor $(x - t)$ from $p(x) \in L[x]$, we obtain a polynomial of lower degree and can proceed by induction.

The proof of uniqueness is also inductive. Suppose L and L' are two splitting fields for $p(x)$. Choose $t \in L$ and $a \in L'$, roots of an irreducible factor $q(x)$ of $p(x)$. We then have an isomorphism

$$K(t) \cong K[x]/(q(x)) \cong K(t')$$

between subfields of L and L'. Identifying these two subfields, we can now regard L and L' as two splitting fields for a polynomial of smaller degree over $K(t) \cong K(t')$. It follows by induction that $L \cong L'$ over K. ∎

Remarks. Since the degree of p decreases by at least one at each stage, the proof shows:

The degree of the splitting field of p divides $\deg(p)!$.

This is related to the fact that the Galois group of L over K (to be discussed later) is a subgroup of the symmetric S_d.

A glimpse of Galois theory: the cube root of two. The field $\mathbb{Q}(2^{1/3})$ is *not* the splitting field of $p(x) = x^3 - 2$, since $p(x)$ has two complex roots. Rather, its splitting field is given by

$$L = \mathbb{Q}(\omega, 2^{1/3}),$$

where $\omega^3 = 1$. We have $[L : \mathbb{Q}] = 6 = 3!$ in this case.

There are 4 interesting subfields of L. First, there are the fields generated by the three roots of $p(x)$:

$$\mathbb{Q}(2^{1/3}) \cong \mathbb{Q}(\omega 2^{1/3}) \cong \mathbb{Q}(\omega^2 2^{1/3}).$$

Second, there is the field $\mathbb{Q}(\omega)$, which has degree 2 over $\mathbb{Q}$ (and contains no root of $p(x)$. Together with L and $\mathbb{Q}$, we obtain 6 fields altogether.

As we will see, Galois theory provides a remarkable way to understand these subfields. They correspond to the subgroups of $S_3 = \mathrm{Gal}(L/K)$. More precisely, $\mathbb{Q}$ corresponds to S_3; the three roots of $p(x)$ correspond to the cyclic groups generated by (12), (23) and (13); the field $\mathbb{Q}(\omega)$ corresponds to A_3; and L itself corresponds to (e).

Complex conjugation gives an element of order 2 in $\mathrm{Gal}(L/\mathbb{Q})$, but where does the element of order 3 come from? To see it, let $K = \mathbb{Q}(\omega) \subset L$. Then $x^3 - 2$ is still irreducible over K. But it *splits* in $K(t)$, for any root of $x^3 - 2$, since the other roots are ωt and $\omega^2 t$. This shows in particular that $K(t)$ and $K(\omega t)$ are isomorphic over K. But both of these fields are L; thus there is an automorphism of L/K that sends to t to ωt, and this automorphism has order 3.

Even more concretely, there is an automorphism of $K[x]$ given by $\alpha(q(x)) = q(\omega x)$. Then $\alpha(x^3 - 2) = (x^3 - 2)$, so α descends to an automorphism of $K[x]/(p(x)) \cong K(2^{1/3})$.

Splitting with one root. Here is the general principle at play in the previous example. For concreteness, let K be a subfield of $\mathbb{C}$, $p(x) \in K[x]$ irreducible as usual, and let $t_1, \ldots, t_d$ be the roots of $p(x)$ in $\mathbb{C}$.

As we have already seen for each t_i there exists an isomorphism of fields,

$$\alpha : K(t_1) \to K(t_i).$$

This is an isomorphism *over* K, means that α is K–linear, or equivalently that $\alpha|K = \mathrm{id}$. It is uniquely determined by the property that

$$\alpha(t_1) = t_i.$$

Now, suppose $L = K(t_1)$ is actually the *splitting field* of $p(x)$. Equivalently, suppose that $t_1, t_2, \ldots, t_d \in L$. Then α belongs to the Galois group $\mathrm{Gal}(L/K)$; it gives an automorphism of L since we have

$$L = K(t_1) \xrightarrow{\alpha} K(t_i) = L.$$

Summing up, we have:

Theorem 8.7 *Suppose $L = K(t) \subset \mathbb{C}$ is the splitting field of an irreducible polynomial $p(x) \in K[x]$ with a root $t \in L$. Then $\mathrm{Gal}(L/K)$ acts simply transitively on the roots $t_1, \ldots, t_d$ of $p(x)$ in L.*

In particular, $|\mathrm{Gal}(L/K)| = [L/K]$.

We will later see that this last equality characterizes splitting fields, and that it suffices to assume that K has characteristic zero.

Trigonometry. In general, the splitting field of an irreducible cubic has degree 3 or 6. An example of degree 3 is provided by $p(x) = x^3 - 3x - 1$. Its splitting field is given by $L = \mathbb{Q}(t)$, where t is any root of $p(x)$.

In fact, $x^3 - 3x - 1$ is the minimal polynomial for $t = 2\cos(\pi/9)$; thus we can take

$$L = \mathbb{Q}(\cos(\pi/9)).$$

The other roots of $p(x)$ are $2\cos(2\pi/9)$ and $2\cos(4\pi/9)$. Since $\cos(2\theta) = 2\cos^2(\theta) - 1$, the other roots are rational polynomial functions of t, and hence $p(x)$ factors completely in $\mathbb{Q}(t)$.

We will later give a general criterion, using the discriminant, to determine if the splitting field of a cubic polynomial has degree 3 or degree 6.

Roots of unity. The roots of unity are close relatives of the values of sine and cosine at rational multiples of π. Indeed, if we let $\zeta_n = \exp(2\pi i/n)$, so that $\zeta_n^n = 1$, then

$$2\cos(2\pi k/n) = \zeta_n^k + \zeta_n^{-k}.$$

Since the roots of $x^n - 1 = 0$ are all powers of ζ_n, we have:

Theorem 8.8 *The splitting field of* $p(x) = x^n - 1$ *is* $\mathbb{Q}(\zeta_n)$.

Note that this polynomial is reducible when $n > 1$.

Constructions with ruler and compass. The elementary theory of fields just discussed has striking applications to geometry.

The set of *constructible figures* F is of the smallest set of points, lines and circles in $\mathbb{C}$ such that:

1. The points 0 and 1 belong to F;
2. The line through any two distinct points of F lies in F;
3. The unique circle with center $c \in F$ passing through a point $p \in F$ is in F;
4. The intersection points of any two lines or circles in F are in F, provided the intersection is finite.

The last condition just rules out, for example, the intersection of a circle with itself.

The points in F form the *constructible numbers* $K \subset \mathbb{C}$. These are the points that can be located, starting with 0 and 1, using just a ruler and a compass. (A ruler is also called a straightedge, to emphasize that it has no markings for measurement; it can only be used to construct a line through two given points.)

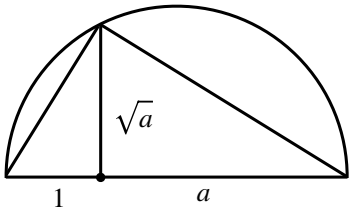


Figure 4. Constructing $\sqrt{a}$.

The classical problem of impossible constructions. The study of constructions with ruler and compass goes back to the period of Greek mathematics around 7 BC. Many elegant constructions were discovered; for example, the Greek proved (in modern terminology) that K is a field, that it

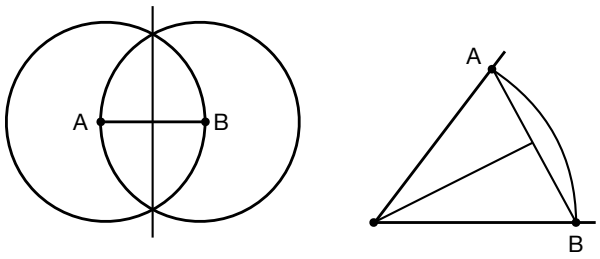


Figure 5. Bisecting an angle.

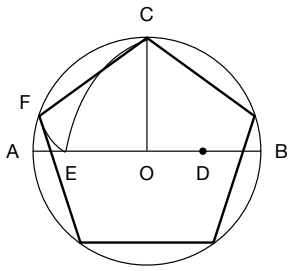


Figure 6. Constructing a regular pentagon. Here D is the midpoint of OB, and the circular arcs CE and EF are centered at D and C respectively.

is closed under taking square–roots, that angles can be bisected, and that a regular pentagon is constructible (see Figures 4, 5 and 6). At the same time 3 challenges emerged:

1. *Trisecting an angle:* Given two lines meeting at angle θ, construct a pair of lines meeting at angle $\theta/3$.

2. *Squaring the circle:* Given a circle, construct a square that encloses the same area.

3. *Doubling the cube:* Given a cube, construct another cube with twice the volume.

The general, the problem of proving a particular mathematical construction *cannot* be done requires a new idea, and a higher level of abstraction. For example, how could one ever show that Euclid's 5th postulate *cannot* be deduced from the remaining axioms? A series of fruitless attempts at deducing it provides little insight into a proof that it cannot be done. For a proof, one needs instead to construct a *new model* for geometry, such as spherical geometry, in which all the other axioms *do* hold and the 5th postulate does not. This shows no such deduction is possible.

To prove that circle cannot be squared, it suffices to show that π is transcendental.

This proof lies beyond the present discussion, but in brief: Lindemann showed in 1882 that if $a \neq 0$ is an algebraic number, then e^a is transcendental. So for example, not only is e transcendental, but so is $e^{\sqrt{2}}$, e^{ζ_n}, etc. Since $e^{\pi i} = -1$ is *not* transcendental, the number πi cannot be algebraic, so π is transcendental.

The field of constructible numbers. However the last two problems are easily proved to be impossible using field theory. First, we will show:

Theorem 8.9 *The field K of constructible numbers is the smallest subfield of $\mathbb{C}$ containing $\mathbb{Q}$ and closed under taking square–roots.*

Corollary 8.10 *If t is a constructible number, then its minimal polynomial has degree 2^n for some n.*

Proof. By the theorem, we can construct K as follows. First, the set $\overline{\mathbb{Q}}$ of all algebraic numbers in $\mathbb{C}$ is countable. Thus we can put them in a list $(a_1, a_2, a_3, \ldots)$.

Now define an increasing sequence of fields inductively as follows: $K_0 = \mathbb{Q}$, and $K_{n+1} = K_n(\sqrt{a_i})$, where $i = i(n)$ is the *smallest index* such that $a_i \in K_n$ but $\sqrt{a_i} \notin K_n$. Let $K = \bigcup K_n$.

We claim that K is closed under square–root. To see this, we first note that $K \subset \overline{\mathbb{Q}}$. Thus any element of K is given by a_j for some j. We also note that $i(n) \to \infty$, since a given square–root is only adjoined once. Thus we can find an n such that $i(n) > j$. Choose n still larger, we can also assume that $a_j \in K_n$. Then, since $i(n) > j$, we must already have $\sqrt{a_j} \in K_n$, otherwise we could adjoin it instead.

By construction, $[K_{n+1} : K_n] = 2$, so $[K_n : \mathbb{Q}] = 2^n$ for all n. Now if $t \in K$ then $t \in K_n$ for some n; thus the degree of its minimal polynomial $p(x)$ divides 2^n, and hence $\deg(p) = 2^m$ for some m. ∎

Theorem 8.11 *There is no trisection of the angle $\pi/3$, and the unit cube cannot be doubled.*

Proof. It is easy to construct an equilateral triangle, and hence two lines meeting at an angle of $\pi/3$. A trisection of this angle would give a proof that $2\cos(\pi/9)$ has even degree over $\mathbb{Q}$; but in fact its minimal polynomial is $x^3 - 3x - 1$, a contradiction.

Similarly, if one could double the unit cube, we would have $t = 2^{1/3} \in K$, contradicting the fact that its minimal polynomial is $x^3 - 2$. ∎

Proof of Theorem 8.9. Let $K \subset \mathbb{C}$ be the constructible numbers, and let $L \subset \mathbb{C}$ be the smallest field containing $\mathbb{Q}$ and closed under taking square–roots. We wish to show that $L = K$.

As remarked, the Greeks showed that L is contained in the constructible numbers; that is, $L \subset K$. For example, the construction of $\sqrt{a}$ when a is real is shown in Figure 4. For complex numbers, one also needs to bisect the angle $\arg(a)$.

It remains to show that $K \subset L$. For this it suffices to show that L is closed under geometric constructions involving lines and circles.

We first remark that $a + ib \in L$ if and only if $a, b \in \operatorname{Re} L$, since L contains i and is closed under complex conjugation.

Next, we note that a line or circle constructed in $\mathbb{C} \cong \mathbb{R}^2$ using points in L is defined by a linear or quadratic equation in $(\operatorname{Re} L)[x, y]$.

It is clear then that the intersection of two lines through points in L is again in L. Similarly, the intersection of a circle with a line leads to quadratic

equations with coefficients in $\operatorname{Re} L$; since these equations can be solved using square–roots, these points also lie in L.

Finally, suppose we have two circles defined using $\operatorname{Re} L$ and meeting at points p and q. Then the line through p and q is also defined by an equation with coefficients in $\operatorname{Re} L$. In fact, the circles are defined by equations of the form $x^2 + y^2 +$ (linear terms) $= 0$, and the difference of these two equations defines the desired line.

Thus the case of two circles reduces to the case of a circle and a line, which we have already checked. ∎

Constructible polygons. A *Fermat prime* is a prime of the form $p = 2^n + 1$. The only know Fermat primes are

$$p = 3, 5, 17, 257 \quad \text{and} \quad 65537.$$

It can be shown that if $2^n + 1$ is prime, then n must be a power of two. This and the known density of primes provides a good heuristic argument that there are no other Fermat primes.

We have seen, using Eisenstein's criterion, that the minimal polynomial for $\zeta_p = \exp(2\pi i/p)$ is simply $x^{p-1} + x^{p-2} + \cdots + 1$. Using this fact, it is straightforward to see:

Theorem 8.12 *A regular polygon with a prime number of sides p is constructible only if p is a Fermat prime.*

Der Koffer. In fact, as shown by Gauss, the converse is also true. An explicit construction of the regular pentagon was known to the Greeks (Euclid's *Elements*, book IV.11). The regular polygon with 17 sides was constructed geometrically by Gauss (1796). Simplifications of both these constructions were obtained by Richmond.

For the regular polygon with 257 sides, an explicit ruler and compass construction was given by Paucker (1822) and Richelot (1832). Finally, the regular polygon with 65,537 sides was constructed by Hermes in his thesis (1894), which is available for inspection in the library of Göttingen University.

For more details, see Coxeter, *Introduction to Geometry*, 1980.

Products of primes. Note that the regular $5 \cdot 17$–gon is also constructible, since $\zeta_5 \cdot \zeta_{17}$ is a constructible number. However it is not true in general that $\zeta_p \in K \implies \zeta_{p^2} \in K$; e.g. we have seen this fails for $p = 3$.

In fact, it is known that

$$\deg(\zeta_n/\mathbb{Q}) = \phi(n) = |(\mathbb{Z}/n)^*|$$

for all n, and in particular the degree of ζ_{p^n} over $\mathbb{Q}$ is $p^{n-1}(p-1)$, which is odd whenever $n > 1$ and p is an odd prime. Thus the regular p^n is not constructible, unless $n = 1$ or $p = 2$. (Of course the regular 2^n-gon is constructible by repeated bisection.) By similar reasoning, one can show:

Theorem 8.13 *The regular n–gon is constructible iff we can write*

$$n = 2^m p_1 \cdots p_k$$

for distinct primes $p_1, \ldots, p_k$, and $p_i - 1$ is a power of 2 for all i.

As we have seen above, there are only 5 choices for the primes p_i.

9 Characteristic zero

Let K be a field of characteristic zero. This means $\mathbb{Q}$ is a subfield of K.

In this section we briefly discuss two properties of such fields: first, in characteristic zero, irreducible polynomials have no multiple roots; and second, the theorem of the primitive element.

Simple roots. The first result is elementary.

Theorem 9.1 *Let $p(x) \in K[x]$ be an irreducible polynomial over a field of characteristic zero. Then $p(x)$ has simple roots in its splitting field.*

Proof. Let L/K be the splitting field of $p(x)$. Then $p(x)$ factors into linear terms in $L[x]$. Suppose it has a repeated root; then we can write

$$p(x) = (x-a)^n q(x),$$

with $a \in L$ and $n > 1$. Now the *derivate* $p'(x) \in K[x]$ *also* has a as a root, since

$$p'(x) = n(x-a)^{n-1} q(x) + (x-a)^n p(x).$$

In characteristic zero, we also know that $p'(x)$ is not zero, since its leading term is dx^{d-1}, $d = \deg(p)$. It follows that $\gcd(p(x), p'(x)) = r(x)$ in $K[x]$ has positive degree, and hence $p(x)$ was not irreducible. ∎

Corollary 9.2 *In characteristic zero, an irreducible polynomial of degree d has d distinct roots in its splitting field.*

Primitive elements. We have seen that an irreducible polynomial determines a finite field extension, $K[x]/(p(x))$. The next result shows that, conversely, every finite field extension comes from a polynomial. This is called the theorem of the *primitive element*.

Theorem 9.3 *Let L/K be a finite field extension in characteristic zero. Then there exists an $a \in L$ such that $L = K(a)$.*

Proof. The the proof is a little tricky.

Since $[L : K]$ is finite, it suffices to show that whenever $L = K(a, b)$, we can find $c \in L$ such that $L = K(c)$.

Let $p(x)$ and $q(x)$ in $K[x]$ be the minimal polynomials of a and b. Changing notation, let L be the splitting field of $p(x)q(x)$. Since we are in characteristic zero, the polynomials $p(x)$ and $q(x)$ have simple roots $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_m)$ respectively, and we can assume that $a = a_1$ and $b = b_1$.

Let $c = a + tb$, with $t \in K$, so $c \in K(a, b)$. Since K is an infinite field, we can choose t so that the only indices such that

$$c = a_i + tb_j$$

are $i = j = 1$. Indeed, the graphs of the linear functions $t \mapsto a_i + tb_j$ cross at only finitely many values of t, and we can simply exclude these values when choosing t.

We will now show that $K(a, b) = K(c)$. The main idea is to consider the polynomial

$$p(c - tx) \in K(c)[x].$$

Like $q(x)$, it vanishes when $x = b$. In fact, this is the only root the two polynomials have in common: since the zeros of $q(x)$ are (b_j) and those of $p(x)$ are (a_i), any common zero satisfies $x = b_j$ for some j, and

$$c - tx = c - tb_j = a_i$$

for some i, which by our choice of t implies $i = j = 1$ and in particular $x = b$.

Thus, the greatest common divisor of $p(c-tx)$ and $q(x)$ is $x-b$. But both of these polynomials have coefficients in $K(c)$, and hence $b \in K(c)$. Then $a = c - tb \in K(c)$ as well, as $K(c) = K(a, b)$. ∎

Simplicity. We used the fact that K has characteristic zero at two points. First, to know that $|K| = \infty$, so we could choose t correctly. And second, to know that $\gcd(p, q) = (x - b)$ and *not* $(x - b)^n$ for some n. The latter would not allow us to conclude that $b \in K(c)$.

Warning: Positive characteristic. It is possible, in positive characteristic, for $p'(x)$ to vanish identically, even when $p(x)$ is not constant. For example, in $\mathbb{F}_7[x]$, the polynomial

$$p(x) = x^{14} + ax^7 + b$$

has $p'(x) = 0$ no matter what a and b are. So in this case, all roots of $p(x)$ are multiple roots. In fact,

$$\textit{If } \operatorname{char} K = p > 0, \textit{ then } (x + y)^p = x^p + y^p \textit{ for all } x, y \in K.$$

Thus we can write

$$p(x) = (x^2 + ax + b)^7.$$

This example may seem contrived, since $p(x)$ is reducible. But if we take $K = \mathbb{F}_p(t)$ to be the field of rational functions, then $p(x) = x^p - t \in K[x]$ is irreducible and $p'(x) = 0$ in this case as well. The resulting splitting field L/K is said to be an *inseparable extension* of K.

When we study Galois theory, we will generally assume $\operatorname{char} K = 0$ to skirt this type of issue.

Absence of primitive elements. It is trickier to give examples of finite field extensions where there is no primitive element.

Theorem 9.4 *The field extension $L = \mathbb{F}_p(x, y)$ over $\mathbb{F}_p(x^p, y^p)$ has no primitive element.*

Proof. We have $[L : K] = p^2$, but for any $f \in L$, we have $f(x, y)^p = f(x^p, y^p) \in K$, and thus $[K(f) : K] \leq p$. ∎

10 Function fields

The field $\mathbb{C}$ is algebraic closed. Thus it has no algebraic extensions. The simplest extension field is $\mathbb{C}(x)$, the field of rational functions in one complex variable.

In this section we will give a glimpse of the geometry underlying $\mathbb{C}(x)$, namely its connection to the Riemann sphere. The field $K = \mathbb{C}(x)$ is no longer algebraically closed, and *its* finite extensions L/K correspond to *algebraic curves* V, or equivalently to compact Riemann surfaces, each equipped with a map $\pi : V \to \widehat{\mathbb{C}}$ that gives rise, dually, to a map

$$\mathbb{C}(x) \to L = \mathbb{C}(V).$$

The target is the field of rational functions on V.

Let $\mathrm{PSL}_2(\mathbb{C}) = \mathrm{SL}_2(\mathbb{C})/(\pm I)$. We will sketch some ideas around the following results:

Theorem 10.1 *The Galois group of* $\mathbb{C}(x)$ *over* $\mathbb{C}$ *is* $\mathrm{PSL}_2(\mathbb{C})$.

Theorem 10.2 *Every rational function* $f(x)$ *of degree* $d > 0$ *determines a subfield of* $\mathbb{C}(x)$ *with*

$$[\mathbb{C}(x) : \mathbb{C}(f(x))] = d.$$

Moreover, $\mathbb{C}(f(x)) \cong \mathbb{C}(x)$.

Theorem 10.3 *There is a natural bijection between degree* d *field extensions* K *of* $\mathbb{C}(x)$*, and compact Riemann surfaces* X *equipped with degree* d *maps* $\pi : X \to \widehat{\mathbb{C}}$.

The Riemann sphere. A *Riemann surface* X is a connected, 1–dimensional complex manifold.

The simplest compact Riemann surface is the *Riemann sphere*, $\widehat{\mathbb{C}} = \mathbb{C} \cup \infty$. A chart near ∞ is provided by $1/x$.

Rational functions. Every rational function $f(x) \in \mathbb{C}(x)$ determines a well–defined map $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$. For example, if $f(x) = p(x)/q(x)$ where the polynomials p and q have no common zeros, then $f(x) = \infty$ at the zeros of $q(x)$. Moreover, if

$$f(x) = \frac{ax^d + \cdots}{bx^e + \cdots},$$

then if $d = e$ we have

$$f(\infty) = \lim_{x \to \infty} f(x) = b/a;$$

otherwise $f(\infty) = 0$ or ∞, depending on whether $d > e$ or $d < e$.

It can be shown that conversely, any holomorphic map $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is given by a rational function. For a general Riemann surface X, we let $\mathbb{C}(X)$ denote the field of rational functions on X, or equivalently the field of holomorphic maps $f : X \to \widehat{\mathbb{C}}$ other than $f(x) = \infty$. Thus

$$\mathbb{C}(\widehat{\mathbb{C}}) \cong \mathbb{C}(x).$$

The degree of a rational function. The degree of $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ can be defined in several equivalent ways.

- Let $f(x) = p(x)/q(x)$ with $\gcd(p, q) = 1$. Then

$$\deg(f) = \max(\deg(p), \deg(q)).$$

- Given a typical point $a \in \widehat{\mathbb{C}}$, we define

$$\deg(f) = |\{x \in \widehat{\mathbb{C}} \,:\, f(x) = a\}.$$

  This answer is always the same, provided a avoids the critical values of f, or if points are counted with multiplicities.

- A rational map f determines a *subfield* $\mathbb{C}(f(x))$ in $\mathbb{C}(x)$, and we define

$$\deg(f) = [\mathbb{C}(x) : \mathbb{C}(f(x))].$$

These three definitions coincide. For example, to find the preimage of $a \in \mathbb{C}$ amounts to solving the equation

$$p(x) - aq(x) = 0,$$

which is (typically) a polynomial of degree $\max(\deg(p), \deg(q))$ in x.

It is easy to see from the second definition that

$$\deg(f \circ g) = \deg(f) \circ \deg(g).$$

In particular, a nonconstant rational function $f(x)$ satisfies no polynomial $p(x) \in \mathbb{C}[x]$, since this would give $p(f(x)) = 0$. Thus $\mathbb{C}(f(x))$ is a transcendental extension of $\mathbb{C}$, and hence

$$\mathbb{C}(x) \cong \mathbb{C}(f(x)).$$

Thus $K = \mathbb{C}(x)$ gives an example of a field with many proper subfields isomorphic to K.

The Galois group. Now consider an automorphism

$$\alpha : \mathbb{C}(x) \to \mathbb{C}(x)$$

that is constant on $\mathbb{C}$, i.e. an element $\alpha \in \mathrm{Gal}(\mathbb{C}(x)/\mathbb{C})$. Then α is uniquely determined by $f(x) = \alpha(x)$. In fact, we have $\alpha(x^n) = f(x)^n$ and, more generally,

$$\alpha(g(x)) = g(f(x)).$$

By the formula for degree as an index, if $\mathbb{C}(x) = \mathbb{C}(f(x))$, then we must have $\deg(f) = 1$. Put differently, if $\alpha^{-1}(x) = g(x)$, then

$$x = \alpha(\alpha^{-1}(x)) = g(f(x)),$$

so $\deg(f) = \deg(g) = 1$. This means

$$f(x) = \frac{ax + b}{cx + d}.$$

Moreover, these polynomials are relatively prime, which means $ad - bc \neq 0$. Rescaling, we can assume that $ad - bc = 1$. In other words, we have a map

$$\mathrm{SL}_2(\mathbb{C}) \to \mathbb{C}(x)$$

whose image is the *group* of rational functions of degree one, with the group law composition. The kernel of this map is $\pm I$. Since the Galois group can also be identitifed with the possible values for $\alpha(x)$, we obtain Theorem 10.1.

Algebraic functions and Riemann surfaces. Finally we provide a glimpse of Theorem 10.3.

Consider a general finite extension L of the field $K = \mathbb{C}(x)$. Every such extension has the form

$$L = K[y]/(p(y)),$$

where $p(y)$ is an irreducible polynomial in y whose coefficients are rational functions of x. To simplify matters, we can always clear denominators and make the coefficients polynomials; we then have $p(y) = p(x, y) \in \mathbb{C}[x, y]$.

The zero locus of $p(x, y)$ defines an *algebraic curve* $V \subset \mathbb{C}^2$. Its ring of polynomial functions, considered before, is

$$\mathcal{O}(V) = \mathbb{C}[x, y]/(p(x, y)),$$

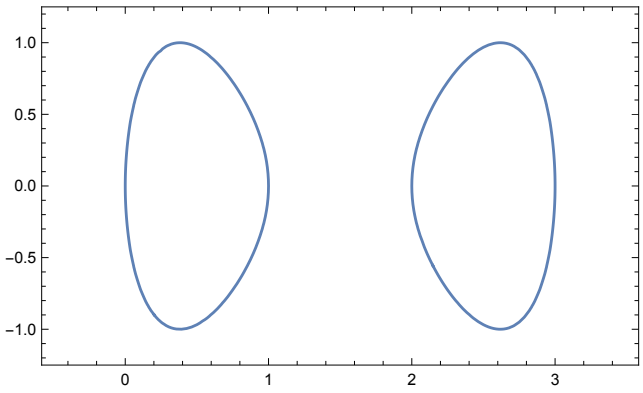


Figure 7. The elliptic curve $y^2 = x(1-x)(2-x)(3-x)$.

and the field of fractions of $\mathcal{O}(V)$ is $\mathbb{C}(V) \cong L$.

Example. Let $p(x) = x(1-x)(2-x)(3-x)$ be a polynomial with simple zeros at $0, 1, 2, 3 \in \mathbb{R}$. Note that $p(x) \geq 0$ on the intervals $[0, 1]$ and $[2, 3]$, so the curve

$$y^2 = p(x)$$

consists of two ovals in $\mathbb{R}^2$. These ovals arise from the real cross-section of the corresponding Riemann surface X, which is a torus.

11 Finite fields

In this section we provide a brief introduction to the theory of finite fields. The main result is:

Theorem 11.1 *Let K be a finite field. Then $|K| = p^n$ is a power of a prime. Conversely, there exists a unique finite field $\mathbb{F}_q$ of order $q = p^n$ for each prime power.*

By *unique* we mean that if $|K| = q$ then $K \cong \mathbb{F}_q$; however there may be many choices for this isomorphism.

We will also show that the theorem of the primitive element holds for finite fields.

Basics of finite fields. The simplest finite fields are the fields $\mathbb{F}_p = \mathbb{Z}/p$, where p is prime. Any finite field K has prime characteristic, so it can be

regarded as an extension field of $\mathbb{F}_p$, $p = \operatorname{char} K$. This shows

$$|K| = p^n, \quad n = [K : \mathbb{F}_p].$$

It is important to note that for $n > 1$, the ring $\mathbb{Z}/p^n$ is *not a field*, since it has zero divisors and even nilpotent elements, such as p itself.

We recall from Theorem 6.14 a pivotal property of finite fields:

The unit group $K^\times$ of a finite field K is cyclic.

Corollary 11.2 *If $|K| = q$ then $x^q = x$ for all $x \in K$.*

The Frobenius. To illuminate the theory of finite fields more completely, we introduce the *Frobenius endomorphism*. This defined, on a field K of characteristic p, by

$$\sigma(x) = x^p.$$

It is a field endomorphism because, using the fact that $\binom{p}{k} = 0 \bmod p$ for $1 \le k \le p-1$, we have

$$(x+y)^p = x^p + y^p.$$

Provided K is *finite*, σ is an automorphism.

To use this map effectively, we note:

Theorem 11.3 *Let $\alpha : K \to K$ be an automorphism of a field. Then $K^\alpha = \{x \in K \,:\, \alpha(x) = x\}$ is a subfield of K.*

The proof is straightforward.

Construction of finite fields. We can now show that there exist finite fields of every possible order.

Theorem 11.4 *Let $q = p^n$, and let L be the splitting field of $p(x) = x^q - x$ over $\mathbb{F}_p$. Then $|L| = q$.*

Proof. Let $\sigma : L \to L$ be the automorphism defined by $\sigma(x) = x^p$, and let $K \subset L$ be the sub–field fixed by $\sigma^n(x) = x^q$. Then K coincides with the roots of $p(x)$. Since $p'(x) = -1 \neq 0$, the roots of $p(x)$ are distinct, and hence $|K| = q$. Thus K is a finite field of the desired order. Since the splitting field of $p(x)$ is generated by its roots, we have $K = L$. ∎

Corollary 11.5 *The field of order $q = p^n$ is unique up to isomorphism.*

Proof. The same argument shows any field with $|L| = q$ is a splitting field for $p(x) = x^q - x$ over $\mathbb{F}_p$. Since the splitting field is unique up to isomorphism (Theorem 8.6), any two fields of order q are isomorphic. ∎

Theorem 11.6 *Let $q = p^n$. Then $\mathbb{F}_q$ contains a unique subfield of order p^d for each integer $d|n$, and no other subfields.*

Proof. Let $r = p^d$. Since $r|q$, the condition $x^r = x$ implies $x^q = x$. Thus $x^r - x$ dividies $x^q - x$, and therefore splits in $\mathbb{F}_q$, and its splitting field K has order r. Conversely, any field of order r is contained in the zero set of $x^r - x$, and hence K is unique. Finally any subfield K of $\mathbb{F}_q$ has order p^r for some r, and if $[\mathbb{F}_q : K] = m$, then $|\mathbb{F}_q| = p^n = (p^r)^e$ and hence $r|n$. ∎

Primitive elements. As we saw in Theorem 9.4, the theorem of the primitive element can fail for *infinite fields* of positive characteristic. However it holds for finite fields.

Corollary 11.7 *Every finite field extension of $\mathbb{F}_p$ is generated by a single element; that is, $\mathbb{F}_q = \mathbb{F}_p[a]$ for some $a \in \mathbb{F}_q$.*

Proof. Just take a to be a generator of the cyclic group $\mathbb{F}_q^\times$. Alternatively one can use a counting argument: summing over the proper subfields K of $\mathbb{F}_q$, we have

$$\sum |K| = \sum_{d|n, d<n} p^d \leq \sum_{i=0}^{n-1} p^i = \frac{p^n - 1}{p - 1} < p^n = |\mathbb{F}_q|,$$

and any element a not in such a K generates $\mathbb{F}_q$. ∎

Finding all irreducible polynomials. The results above have an important constructive consequence.

Theorem 11.8 *The irreducible polynomials in $\mathbb{F}_p[x]$ of degree $d|n$ coincide with the irreducible factors of the polynomial $p(x) = x^q - x$, $q = p^n$.*

Moreover, there is at least one factor of each degree $d|n$.

Proof. Let $p(x)$ be an irreducible factor of $x^q - x$. Then $p(a) = 0$ for some $a \in \mathbb{F}_q$, and thus

$$\deg(p) = [\mathbb{F}_p[a] : \mathbb{F}_p] \text{ divides } n = [\mathbb{F}_q : \mathbb{F}_p].$$

Conversely, if an irreducible polynomial $p(x)$ has degree $d|n$, then

$$\mathbb{F}_p[x]/(p(x)) \cong \mathbb{F}_r,$$

where $r = p^d$; and since $\mathbb{F}_r \subset \mathbb{F}_q$, $p(x)$ also has a root $a \in \mathbb{F}_q$. But then a is a zero of $x^q - x$, and hence its minimal polynomial $p(x)$ is a divisor of $x^q - x$.

The final remark follows from the fact that the extension $\mathbb{F}_r/\mathbb{F}_p$ is generated by a single element. ∎

Configuration of subfields. Fix $q = p^n$. There for each divisor d of n, there is a unique field $K(d)$ with

$$\mathbb{F}_p \subset K(d) \subset \mathbb{F}_q$$

and $|K(d)| = p^d$. The partial ordering of these subfields by inclusion coincides with the partial ordering of the divisor of n by divisibility; that is,

$$K(d) \subset K(e) \iff d|e.$$

Examples of finite fields. To construct a finite field of order $\mathbb{F}_q$, $q = p^n$, explicitly, it suffices to find an irreducible polynomial $p(x)$ of degree n in $\mathbb{F}_p[x]$; then $\mathbb{F}_q \cong \mathbb{F}_p[x]/(p(x))$. The polynomial $p(x)$ can be found by factoring $x^q - x$ into irreducibles.

The simplest finite field other than those of the form $\mathbb{F}_p$ is

$$\mathbb{F}_4 \cong \mathbb{F}_2[x]/(x^2 + x + 1).$$

This extension is defined using the *unique* irreducible quadratic polynomial in $\mathbb{F}_2[x]$.

We also recognize $x^2 + x + 1$ as the minimal polynomial for ω in $\mathbb{Z}[x]$; this reflects the fact that 2 is a prime in $\mathbb{Z}[\omega]$, and

$$\mathbb{Z}[\omega]/(2) \cong \mathbb{F}_4.$$

We can similarly construct fields of orders 8 and 16 using the factors of highest degree of

$$x^8 - x = x(x+1)(x^3+x+1)(x^3+x^2+1)$$

and

$$x^{16} - x = x(x+1)(x^2+x+1)(x^4+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1),$$

and construct $\mathbb{F}_9$ using any of the quadratic factors of

$$x^9 - x = x(x+1)(x+2)(x^2+1)(x^2+x+2)(x^2+2x+2).$$

Note that the number of irreducible quadratic polynomials in $\mathbb{F}_p[x]$ is given by

$$|\mathbb{F}_{p^2} - \mathbb{F}_p|/2 = p(p-1)/2,$$

since each has two roots in $\mathbb{F}_{p^2} - \mathbb{F}_p$.

12 Galois theory

Galois theory is the capstone of our discussion of fields.

It is a remarkable insight that the theory of symmetry groups of fields is very rich, and it opens a completely new avenue to exploring the secrets of polynomials and their roots.

Throughout this section we make the standing assumption:

> *All fields under consideration will have characteristic zero.*

Thus we may also assume:

> *All irreducible polynomials have simple roots, and any finite field extension is generated by a single element.*

Little will be lost by assuming that all fields in question are number fields in $\mathbb{C}$. In particular we will sometimes refer the a primitive nth root of unity as ζ_n, which usually denotes a complex number.

Alternatively, one can imagine we are working in the algebraic closure $\overline{K}$ of some base field K of characteristic zero. For a general field, ζ_n denotes *any* primitive nth root of unity.

Roots and polynomials. Let K be a field (of characteristic zero), and let

$$p(x) \in K[x]$$

be a monic polynomial of degree d. Let L/K be a field extension in which $p(x)$ splits into linear factors:

$$p(x) = \prod_{i=1}^{n}(x - a_i).$$

Here $a_i \in L$ are the roots of $p(x)$, repeated with multiplicity.

The set of roots (a_i) is uniquely determined by the coefficients of $p(x)$. Consequently any function of the roots $F(a)$, which treats all roots equally, actually depends just on the coefficients of $p(x)$, not the roots themselves.

Let us make this precise in the case of a polynomial $F(a)$.

Symmetric functions. In general, when a group G acts by automorphisms of a ring A, its invariant subring is denoted by

$$A^G = \{x \in A \ : \ g \cdot x = x \ \ \forall g \in G\}.$$

The discussion of polynomials and their roots can be formulated in terms of the action of the symmetric group S_n on the ring

$$A = K[a_1, \ldots, a_n].$$

This action of S_n permutes the variables, i.e. $a_i^\sigma = a_{\sigma \cdot i}$.

For brevity of notation we write $a = (a_1, \ldots, a_n)$. We say $F(a) \in A$ is a *symmetric polynomial* if it does not depend on the ordering of its variables; i.e. if $F(a)^\sigma = F(a)$ for all $\sigma \in \S_n$.

From roots to coefficients. Now let us consider the polynomial in x whose roots are a_i, defined by

$$p(x) = \prod_{1}^{n}(x - a_i) = x^n - s_1 x^{n-1} + s_2 x^{n-2} + \cdots + (-1)^n s_n.$$

Since the coefficients s_i of $p(x)$ do not depend on the ordering of its roots,

they are examples of symmetric polynomials. In fact, we have

$$\begin{aligned} s_1 &= \sum a_i, \\ s_2 &= \sum_{i<j} a_i a_j, \\ s_3 &= \sum_{i<j} a_i a_j a_k, \quad \text{and} \\ s_n &= \prod a_i. \end{aligned}$$

These polynomials are called the *elementary symmetric functions*. We can now state the main result.

Theorem 12.1 *Every symmetric function is a polynomial in $s = (s_1, \ldots, s_n)$. Equivalently, we have*

$$K[a_1, \ldots, a_n]^{S_n} = K[s_1, \ldots, s_n].$$

Remark on degrees. A polynomial $F(a)$ is *homogeneous* of degree d if $F(\lambda u) = \lambda^d F(a)$ for all $\lambda \in K$.

Note that the degree of $F_1(a)F_2(a)$ is the sum of their degrees. Every monomial $a_1^{e_1} \cdots a_n^{e_m}$ is homogeneous; its degree is given by the sum of its exponents, $\sum e_i$. A polynomial homogeneous polynomial $F(a)$ of degree d only involves monomials of degree d. We remark that:

The symmetric polynomial $s_i(a)$ is homogeneous of degree i.

When a homogeneous symmetric function $F(a)$ is expressed as a polynomial in $s_1, \ldots, s_i$, each term is homogeneous of the same degree as F.

Cubic example. Before proceeding to the proof, we work out a simple example. When $n = \deg(p) = 2$, there are two symmetric functions, the trace and the norm:

$$s_1 = a_1 + a_2 \quad \text{and} \quad s_2 = a_1 a_2.$$

Any other symmetric function, such as $a_1^d + a_2^d$, can be expressed in terms of these. For example, we have

$$s_1^3 = a_1^3 + 3a_1^2 a_2 + 3a_1 a_2^2 + a_2^3 = a_1^3 + a_2^3 + 3s_1 s_2,$$

and thus

$$a_1^3 + a_2^3 = s_1^3 - 3s_1s_2.$$

Proof of Theorem 12.1. Let $F(a)$ be a symmetric polynomial. The proof will be by induction on the number of variables n, as well as $\deg F$.

Let $F^*(a)$ denote the polynomial obtained by setting $a_n = 0$, and similarly for s_i^*. By induction on n, we can write

$$F^*(a) = S(s_1^*, \ldots, s_{n-1}^*)$$

for some polynomial S. Consider the symmetric polynomial

$$G(a) = F(a) - S(s).$$

By construction, $G(a)$ vanishes if we set $a_n = 0$, so $G(a)$ is divisible by a_n. In fact, since $G(a)$ is a symmetric polynomial, it is divisible by $a_1 \cdots a_n = s_n$. Thus we have shown that

$$F(a) = S(s) + s_n H(a),$$

where $H(a)$ is a symmetric polynomial of degree at most $\deg(F) - n$. The proof is completed by induction on $\deg(F)$ and on n. ∎

Cubic example, continued. Let us express $F(a) = a_1^3 + a_2^3 + a_3^3$ in terms of s_1, s_2 and s_3. We have already solved to $n = 2$ case, so the argument above shows:

$$F(a) = s_1^3 - 3s_1s_2 + as_3$$

for some constant a. Setting $a = (1, 1, 1)$ gives $F(a) = 3$ and $(s_1, s_2, s_3) = (3, 3, 1)$, so $3 = 3^3 - 3^3 + a$, which gives $a = 3$.

We note that the same formula gives $\sum_1^n a_i^3$ for every $n \geq 3$, we just need to use the symmetric functions of n variables for s_1, s_2 and s_3.

Splitting fields, revisited. The next result allows us to adopt a new perspective on splitting fields.

Theorem 12.2 *Let L/K be a splitting field. Then every irreducible polynomial $q(x) \in K[x]$ that has at least one root in L, splits in L.*

Proof. Let L be the splitting field of a degree n polynomial $p(x) \in K[x]$, with roots $a = (a_1, \ldots, a_n)$. Suppose $q(x) \in K[x]$ is an irreducible polynomial with a root $b \in L$. Since L is generated by the roots of $p(x)$, we can write $b = f(a_1, \ldots, a_n)$ for some polynomial f. Note that $f(a^\sigma) \in L$ for each $\sigma \in S_n$. Thus we can form the product of linear terms

$$g(x) = \prod_{\sigma \in S_n} (x - f(a)^\sigma).$$

Then each *coefficient* of $g(x)$ is a symmetric function of $(a_1, \ldots, a_n)$, and hence it can be expressed in terms of the coefficients of $p(x)$. Consequently $g(x) \in K[x]$, and it splits in L. Since $g(b) = g(f(a)) = 0$, $g(x)$ is divisible by $q(x)$, and hence $q(x)$ also splits in L. ∎

New definition of a splitting field. In view of this result, we will simply say that L/K is a *splitting field* if *every* irreducible polynomial in $K[x]$ with a root in L splits in L.

By the theorem of the primitive element, there is at least one such polynomial whose roots generate L, so this is consistent with our previous terminology.

The Galois group. We now come to the main object of interest, the *Galois group* of L/K. This group is defined by:

$$\mathrm{Gal}(L/K) = \{\alpha : L \to L \ : \ \alpha | K = \mathrm{id}\},$$

where α ranges over all automorphisms of L as a field. For brevity we will write $\alpha(x)$ by x^α; thus

$$(xy)^\alpha = x^\alpha y^\alpha \quad \text{and} \quad (x+y)^\alpha = x^\alpha + y^\alpha.$$

Use of a primitive element. How big can the Galois group be? Using the theorem of the primitive element, we can write

$$L = K(t) \cong K[x]/(p(x)),$$

where $t \in K$ and $p(x)$ is its minimal polynomial, of degree $d = [L : K]$. Then any element α of $\mathrm{Gal}(L/K)$ is uniquely determined by the value $u = t^\alpha$ that it takes on the generator t; and u must be a root of $p(x)$, since α fixes the coefficients of p.

Conversely, for any root u of $p(x)$, there is an automorphism of L/K with $t^\alpha = u$, since $K(u) \cong K[x]/(p(x))$ as well. Summing up, we have:

Theorem 12.3 *Consider the field extension L/K, where*

$$L \cong K[x]/(p(x)).$$

Then $|\operatorname{Gal}(L/K)|$ is the same as the number of roots of $p(x)$ in L.

Corollary 12.4 *We have $|\operatorname{Gal}(L/K)| = [L : K]$ if and only if L/K is a splitting field.*

In this case we also say that L is a *Galois extension* of K.

Cubic example. Let us revisit the field extension $L = \mathbb{Q}(2^{1/3}, \omega)$ of $\mathbb{Q}$. This degree 6 extension is the splitting field of $x^3 - 2$, so it is also a Galois extension. As we have discussed before, its Galois group is S_3. On the other hand, $K = \mathbb{Q}(2^{1/3}) \subset L$ is a cubic extension of $\mathbb{Q}$, but *not* a splitting field. Since only one root of $x^3 - 2$ has been adjoined, its Galois group is *trivial*.

Intermediate fields. One of the main goals of Galois theory is to provide a clear picture of fields M with $K \subset M \subset L$. To this end we note:

Theorem 12.5 *If L/K is Galois, then L/M is also Galois, for any field with $K \subset M \subset L$.*

Proof. Suppose L/K is Galois; then it is the splitting field for some $p(x) \in K[x]$. Since $K[x] \subset M[x]$, L is also the splitting field for $p(x)$ over M. ∎

Fixed fields. Any subgroup $H \subset \operatorname{Gal}(L/K)$ determines a subfield of L, defined by

$$L^H = \{x \in L \,:\, x^\alpha = x \;\forall \alpha \in H\}.$$

Here is a pivotal result in Galois theory. *Note that we do not assume that L/K is Galois!*

Theorem 12.6 *Let L/K be a finite extension, let H be a subgroup of $\operatorname{Gal}(L/K)$, and let $M = L^H$. Then L/M is a Galois extension, and*

$$\operatorname{Gal}(L/M) = H.$$

Corollary 12.7 *For any subgroup $H \subset \operatorname{Gal}(L/K)$, we have $[L : L^H] = |H|$.*

Proof. To see that L/M is a Galois extension, write $L = M(t)$, and note that L is the splitting field, over M, of the polynomial

$$p(x) = \prod_{\alpha \in H} (x - t^\alpha) \in M[x].$$

The coefficients of $p(x)$ lie in M since they are symmetric functions of the orbit of t under the H, and hence themselves H–invariant. This polynomial is also irreducible, since the minimal polynomials of t and t^α in $M[x]$ are the same for all $\alpha \in H$. This shows that

$$[L : M] = |H|.$$

Clearly H is a subgroup of $\mathrm{Gal}(L/M)$; in fact these groups coincide, since

$$|H| \leq |\,\mathrm{Gal}(L/M)| \leq [L : M] = |H|.$$

■

Corollary 12.8 *An extension L/K is Galois iff the fixed field of* $\mathrm{Gal}(L/K)$ *is K.*

The main result of Galois theory. We can now establish:

Theorem 12.9 (Galois) *Let L/K be a splitting field, with Galois group $G = \mathrm{Gal}(L/K)$. Then the map*

$$i : H \mapsto L^H = M$$

establishes a bijection between the set of subgroups $H \subset G$, and the set of fields with $K \subset M \subset L$. Its inverse is given by

$$j : M \mapsto \mathrm{Gal}(L/M) = H.$$

Under this correspondence, $|H| = [L : M]$.

Proof. Let us first show that $i \circ j(M) = M$. This amounts to showing that M is the fixed field of $\mathrm{Gal}(L/M)$, which follows from the fact that L/M is Galois by Theorem 12.5.

Similarly, to show $j \circ i(H) = H$ amounts to showing that $\mathrm{Gal}(L/L^H) = H$. This follows from Theorem 12.6.

The index formula is immediate from the fact that L/M is Galois and $H = \mathrm{Gal}(L/M)$. ■

Corollary 12.10 *Let L/K be a finite extension. Then there are only finitely many fields M between K and L.*

Splitting fields and normal subgroups. Suppose one of the intermediate fields M for L/K is itself a splitting field over K. That is, it splits some polynomial $p(x) \in K[x]$. Then, since K is fixed by $\mathrm{Gal}(L/K)$, so is $p(x)$, and therefore M is as well.

This means that if we write $M = L^H$, then H is a normal subgroup of $\mathrm{Gal}(L/K)$. Indeed, given $\alpha \in G$, the subgroup fixing $\alpha(M)$ is $\alpha H \alpha^{-1}$, so if $\alpha(M) = M$, then $\alpha H \alpha^{-1} = H$.

Conversely, if H is normal, then $M = L^H$ is invariant under $G = \mathrm{Gal}(L/K)$. The subgroup H acts trivially on M, so we obtain a map $G/H \to \mathrm{Gal}(M/K)$. It follows that

$$|G/H| = |G|/|H| \leq |\,\mathrm{Gal}(M/K)| \leq [M:K] = [L:K]/[L:M] = |G/H|,$$

and thus equality holds throughout. In summary, we have shown:

Theorem 12.11 *The field $M = L^H$ is a Galois extension of K if and only if H is a normal subgroup of $\mathrm{Gal}(L/K)$, in which case*

$$\mathrm{Gal}(M/K) \cong \mathrm{Gal}(L/K)/H.$$

Corollary 12.12 *If the Galois group of L/K is abelian, then every intermediate subfield M is Galois over K.*

Example: $2^{1/3}$ revisited. As we have seen previously, for $K = \mathbb{Q}$ and $p(x) = x^3 - 2$, the splitting field is $L = \mathbb{Q}(\omega, 2^{1/3})$ with $t_i = 2^{1/3} \cdot \omega^i$ the roots of p, $i = 1, 2, 3$; and the Galois group satisfies

$$G = \mathrm{Gal}(L/\mathbb{Q}) \cong S_3,$$

the isomorphism coming from the permutation of the roots (t_1, t_2, t_3).

The subgroup $A_3 = \langle (123) \rangle \subset S_3$ is *normal*, and it corresponds to the Galois extension $\mathbb{Q}(\omega)/\mathbb{Q}$. The $\mathbb{Z}/2$ subgroups of S_3, corresponding to its three transpositions, are not normal, and they correspond to the non–Galois extensions $\mathbb{Q}(t_i)$, $i = 1, 2, 3$.

Once we know the Galois group action this explicitly, it is easy to find primitive elements. For example, we have

$$L = \mathbb{Q}(\omega + 2^{1/3}).$$

To see this, we observe that the orbit of $\omega + 2^{1/3}$ under the Galois group consists of its orbit under A_3, which is

$$\omega + 2^{1/3}\omega^i, i = 1, 2, 3;$$

and the rest of its orbit is obtained by applying complex conjugation to these three points. The resulting six points are distinct, and it follows that $\omega + 2^{1/3}$ must generate the whole field L. (If it generated a proper subfield L^H, then it would be fixed by the elements of H.)

Field extensions $K(t) \subset L$. The Galois group allows one to easily find the degree of a field extension $K(t)$ and all the roots of an irreducible polynomial as well.

Here is a typical result in this direction. Let L/K be a Galois extension with $G = \mathrm{Gal}(L/K)$. Given $t \in L$, let $t^G = \{t^\alpha \ : \ \alpha \in G\})$.

Theorem 12.13 *Suppose $t \in L$ has irreducible polynomial $p(x) \in K[x]$. Then:*

1. *The roots of $p(x)$ coincide with the Galois orbit t^G of t.*
2. *The field $K(t) = L^H$, where $H = \{\alpha \in G \ : \ t^\alpha = t\}$.*
3. *We have $\deg(t/K) = [G : H] = |t^G|$.*

The proof is straightforward from the main theorem of Galois theory, using the fact that $\mathrm{Gal}(L/K(t)) = H$. The roots of $p(x)$ are often called the *Galois conjugates* of t.

Corollary 12.14 *We have $L = K(t)$ if and only if the only $\alpha \in \mathrm{Gal}(L/K)$ such that $t^\alpha = t$ is the identity element.*

Function fields. Let us now take a brief look at the Galois theory of function fields. As we have seen, we have:

Theorem 12.15 *The Galois group of $\mathbb{C}(t)$ over $\mathbb{C}$ is isomorphic to $\mathrm{PSL}_2(\mathbb{C})$.*

These automorphisms come from degree one rational maps.

More generally, every nonconstant rational map $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ determines a finite field L/K, where $L \cong K \cong \mathbb{C}(t)$, and

$$[L : K] = [\mathbb{C}(t) : \mathbb{C}(f(t))] = \deg f.$$

This extension is *rarely Galois*. When it is, the Galois group can be realized as a finite subgroup of $\mathrm{Aut}(\widehat{\mathbb{C}}) \cong \mathrm{PSL}_2(\mathbb{C})$, in such a way we can write

$$f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}/G \cong \widehat{\mathbb{C}}.$$

Equivalently, $f(x) = f(y)$ if and only if $x \in G \cdot y$.

Examples. The classification of finite subgroups of $\mathrm{PSL}_2(\mathbb{C})$ is well–known, and hence all the Galois rational maps are also known. Here is the list.

1. The map $f(t) = t^d$ has Galois group $\mathbb{Z}/d$. Here G is generated by $t \mapsto \zeta_d t$.

2. The map $f(t) = t^d + t^{-d}$ has Galois group D_{2d}, the dihedral group of order $2d$. Here D_{2d} is generated by $t \mapsto \zeta_d t$ and $t \mapsto 1/d$.

3. There are 3 more complicated rational maps, which we will not write down, that have Galois groups A_4, S_4 and A_5. Klein showed that one can solve quintic polynomials using radicals and the inverse of the degree 60 rational map with Galois group A_5.

Cyclic extensions. A large supply of cyclic extensions of $\mathbb{C}(x)$ are provided by the algebraic curves associated to equations of the form

$$y^n = p(x) \in \mathbb{C}[x].$$

Here we should take care to insure that $y^n - p(x)$ is irreducible; this amounts to avoiding the cases where $p(x) = q(x)^d$, and $d|n$. Then, up to automorphisms of $\mathbb{C}(x)$, all cyclic extensions have this form.

When $n = 2$, and $\deg p = d$ is even, the resulting Riemann surface has genus $g = (d-2)/2$.

13 Solving polynomials

In this section we will use Galois theory to study the behavior of polynomials and their roots.

Let $p(x) \in K[x]$ be an irreducible polynomial of degree n, with splitting field L/K. We can then regard $G = \mathrm{Gal}(L/K)$ as the *Galois group of the polynomial* $p(x)$. One of our goals will be to develop the following famous theorem.

Theorem 13.1 (Galois) *The polynomial $p(x)$ is solvable by radicals if and only if its Galois group is a solvable group.*

Along the way, we will examine a variety of different and concrete examples.

Galois groups as permutation groups. Let L/K be the splitting field of $p(x)$, and let $T = \{t_1, \ldots, t_n\} \subset L$ denote the set of roots of p. We then have a natural faithful representation,

$$\mathrm{Gal}(L/K) \to \mathrm{Sym}(T) \cong S_n.$$

The image of this representation is also called the Galois group of p. In this setting we have several general results.

Theorem 13.2 *The Galois group of $p(x)$ acts transitivity on its roots.*

Proof. The polynomial

$$q(x) = \prod_{\alpha \in \mathrm{Gal}(L/K)} (x - t_1^\alpha)$$

lies in $K[x]$, and all its roots are also roots of the irreducible polynomial $p(x)$. It follows that $q(x)$ is a power of $p(x)$, and in particular every root of $p(x)$ is in the orbit of the Galois group. ∎

Corollary 13.3 *The stabilizers of the roots t_i of $p(x)$ are conjugate subgroups $H_i \subset \mathrm{Gal}(L/K)$.*

Corollary 13.4 *The fields $K(t_i) \subset L$ obtained by adjoining a root of $p(x)$ are permuted transitively by $\mathrm{Gal}(L/K)$.*

Corollary 13.5 *If t is a root of $p(x)$, then $\mathrm{Gal}(L/K(t))$ is isomorphic to the intersection of the Galois group of p with S_{n-1}.*

Corollary 13.6 *The splitting field of $p(x)$ is generated by a single root of p iff the $\mathrm{Gal}(L/K)$ acts simply transitively on the roots of p; equivalently, iff $|\mathrm{Gal}(L/K)| = \deg(p)$.*

The discriminant. Let $A_n \subset S_n$ be the alternating group. It is natural to ask if the Galois group of $p(x)$ lies in this subgroup of the symmetric group, i.e. if all permutations have even parity.

To answer this, we define the *discriminant* of $p(x)$ in terms of its roots by

$$D(p) = \prod_{i<j} (t_i - t_j)^2.$$

We have $D(p) \in K$ since it is invariant under permutations of the t_i. In fact $D(p)$ is a symmetric function of the roots, so there is a universal formula for $D(p)$ in terms of the coefficients of p. This formula is given in terms of resultants by

$$D(p) = (-1)^{n(n-1)/2} \operatorname{Res}(p, p').$$

Note that some care must be taken with the sign, which plays a critical role, since negative numbers in $\mathbb{Q}$ cannot be squares.

Theorem 13.7 *The Galois group of $p(x)$ lies in A_n if and only if the discriminant $D(p)$ is a square in K.*

Proof. Let $\widetilde{D}(p) = \prod_{i<j}(t_i - t_j) \in L$. Then $\widetilde{D}(p)$ is one of the two square roots of $D(p)$, and it is invariant exactly under the subgroup $H = \mathrm{Gal}(L/K) \cap A_n$. Thus

$$\widetilde{D}(p) \in K \iff \mathrm{Gal}(L/K) \subset A_n.$$

■

Examples of Galois groups of polynomials. We now turn to examples. Let $p(x) \in K[x]$ be an irreducible polynomial. We describe several special cases.

1. Quadratic polynomials. It is well–known that every quadratic polynomial can be solved by extracting a square–root; thus if $\deg(p) = 2$, we have

$$\mathrm{Gal}(L/K) \cong S_2 = \mathbb{Z}/2.$$

2. Roots of unity. Let $\zeta_n = \exp(2\pi i/n) \in \mathbb{C}$. This number is algebraic, since it is a root of $p(x) = x^n - 1$. In fact we have:

Theorem 13.8 *The field extension $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ is the splitting field for $x^n = -1$. Its Galois group is abelian.*

Proof. The first statement follows from the fact that ζ_n generates a cyclic subgroup μ_n of $\mathbb{C}^*$, isomorphic to $\mathbb{Z}/n$, that coincides with the nth roots of unity.

For the second, observe that every $\alpha \in \mathrm{Gal}(K/\mathbb{Q})$, $K = \mathbb{Q}(\zeta_n)$, must send ζ_n to another generator of μ_n; thus

$$\alpha(\zeta_n) = \zeta_n^k, \text{ with } k \in (\mathbb{Z}/n)^\times.$$

Since α is determined by its value on ζ_n, the map $\alpha \mapsto k$ exhibits $\mathrm{Gal}(K/\mathbb{Q})$ as a subgroup of the abelian group $(\mathbb{Z}/n)^\times$, so the Galois group itself is also abelian. ∎

Corollary 13.9 *Every subfield of $\mathbb{Q}(\zeta_n)$ is a splitting field.*

Cyclotomic numbers. Let us define the *cyclotomic numbers* by

$$\mathcal{C} = \bigcup_n \mathbb{Q}(\zeta_n) \subset \mathbb{C}.$$

Since $\zeta_a, \zeta_b \in \zeta_{ab}$, it is easy to see that any two elements of $\mathcal{C}$ lie in a cyclotomic field. Thus, in view of the Corollary above, we have:

Theorem 13.10 *The set of cyclotomic numbers $\mathcal{C}$ is a subfield of $\mathbb{C}$, and for any $t \in \mathcal{C}$, the field $L = \mathbb{Q}(t)$ is Galois over $K = \mathbb{Q}$ and $\mathrm{Gal}(K/L)$ is abelian.*

Examples of cyclotomic numbers include $\cos(\pi p/q)$ and $\sin(\pi p/q)$ for any rational p/q. As we have seen, $\sqrt{5}$ is also a cyclotomic number, since $2\cos(\pi/5) = (1+\sqrt{5})/2$. It is a nontrival fact that $\sqrt{n} \in \mathcal{C}$ for all $n \in \mathbb{Z}$.

Cyclotomic polynomial. The minimal polynomial for ζ_n is called the *cyclotomic polynomial* of order n. There is an obvious candidate for this polynomial, namely:

$$\Phi_n(x) = \prod_{k \in (\mathbb{Z}/n)^\times} (x - \zeta_n^k) \in \mathbb{Z}[x].$$

In fact we have:

Theorem 13.11 (Gauss) *The polynomial $\Phi_n(x)$ is irreducible.*

Corollary 13.12 *We have* $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q}) \cong (\mathbb{Z}/n)^*$*, and hence*

$$[\mathbb{Q}(\zeta_n)/\mathbb{Q}] = \deg \Phi_n = \phi(n).$$

Here the Euler ϕ–function, $\phi(n)$, gives the number of $k \in \mathbb{Z}/n$ relatively prime to n.

Abelian and cyclotomic extensions. In a sense, we have now seen all the abelian extensions of $\mathbb{Q}$. More precisely, we have the following striking result:

Theorem 13.13 (Kronecker–Weber) *Any finite abelian Galois extension $L/\mathbb{Q}$ is contained in a cyclotomic extension.*

Corollary 13.14 *An algebraic number $t \in \overline{\mathbb{Q}}$ is a cyclotomic number iff $\mathbb{Q}(t)$ is a Galois extension of $\mathbb{Q}$ with abelian Galois group.*

For example, $2^{1/3}$ is *not* a cyclotomic number, since $\mathbb{Q}(2^{1/3})$ is not Galois. The same is true for an irrational number of the form $a^{1/p}$, where $a \in \mathbb{Z}$ and $p > 2$ is prime.

Example: $\mathbb{Q}(\zeta_p)$. Let $p > 2$ be an odd prime, and let $\zeta = \zeta_p$.

Then $L = \mathbb{Q}(\zeta)$ is a Galois extension of $K = \mathbb{Q}$ with Galois group

$$G = (\mathbb{Z}/p)^* = \{1, 2, \ldots, p-1\} \cong \mathbb{Z}/(p-1).$$

This group is multiplicative and *cyclic*. (In general it is hard to say what a generator is!)

A basis for $L/\mathbb{Q}$ is given simply by $(\zeta, \zeta^2, \ldots, \zeta^{p-1})$. The action of the Galois group in this basis is by *permutations* of the basis elements. Indeed, if $\alpha \in G$ corresponds to $k \in (\mathbb{Z}/p)^*$, then

$$(\zeta^i)^\alpha = \zeta^{ik},$$

and if $a_i \in \mathbb{Q}$, then

$$(\sum a_i \zeta^i)^\alpha = \sum a_i \zeta^{ik}. \tag{13.1}$$

The real subfield. The action of complex conjugation sends ζ^i to ζ^{-i}, and its fixed field is given by

$$R = \mathbb{Q}(\zeta) \cap \mathbb{R} = \mathbb{Q}(\cos(2\pi/p)).$$

In fact, a basis over $\mathbb{Q}$ for this field is given by

$$e_k = \zeta^k + \zeta^{-k} = 2\cos(2\pi k/p)$$

$k = 1, 2, \ldots, (p-1)/2$, and we have:

Theorem 13.15 *The degree of* $\cos(2\pi/p)$ *over* $\mathbb{Q}$ *is* $(p-1)/2$*, and its Galois group is isomorphic to* $(\mathbb{Z}/p)^*/(\pm 1)$*.*

The quadratic subfield. Since $G \cong \mathbb{Z}/(p-1)$ has even order, it contains a unique subgroup of index two, namely $2G$. By Galois theory this shows:

Theorem 13.16 *There is a unique quadratic extension* $\mathbb{Q}(\sqrt{d})$ *of* $\mathbb{Q}$ *contained in* $\mathbb{Q}(\zeta_p)$*.*

It can be shown, using e.g. Gauss sums, that $d = \sqrt{p}$ when $p = 1 \bmod 4$ and $d = \sqrt{-p}$ when $p = 3 \bmod 4$. Let us be more concrete.

Under the isomorphism $G \cong (\mathbb{Z}/p)^*$, the group $2G$ is the group of *squares*. Thus the quadratic extension above is given by $\mathbb{Q}(u)$ where

$$u = \frac{1}{2} \sum_{1}^{p-1} \zeta^{k^2}.$$

(We have included $1/2$ because every square occurs twice.)

For example, when $p = 5$, we find that $u = \zeta + \zeta^4$. Using the fact that $1 + \zeta + \zeta^2 + \zeta^3 + \zeta^4 = 0$, we find that

$$u^2 = \zeta^2 + \zeta^3 + 2 = 1 - u,$$

and thus $u^2 + u - 1 = 0$. Since the discriminant of this polynomials is 5, we see that

$$\mathbb{Q}(\sqrt{5}) \subset \mathbb{Q}(\zeta_5).$$

This shows that neither i nor ω belong to $\mathbb{Q}(\zeta_5)$. From this we find that

$$\sin(2\pi/5) = \frac{\zeta_5 - \zeta_5^{-1}}{2i} \notin \mathbb{Q}(\zeta_5).$$

However, $\zeta_{20}^5 = i$, so we have

$$\sin(2\pi/5) \in \mathbb{Q}(\zeta_{20}).$$

In fact, $\sin(2\pi/5)$ has degree 4 over $\mathbb{Q}$, which is the same as the degree of the real subfield of $\mathbb{Q}(\zeta_{20})$, which suggests that

$$\mathbb{Q}(\sin(2\pi/5)) = \mathbb{Q}(\cos(\pi/10).$$

In fact more is true: $\sin(2\pi/5) = \cos(\pi/10)$!

Stabilizer of a cyclotomic number. In view of equation (13.1), it is straightforward to compute the subgroup of $G \cong (\mathbb{Z}/p)^*$ fixing any given element $t \in \mathbb{Q}(\zeta)$, and hence compute the degree d of $\mathbb{Q}(t)/\mathbb{Q}$. Moreover, the Galois group of this extension is $\mathbb{Z}/d$.

As a particuarly simple case, we note the following application of Theorem 12.13:

Theorem 13.17 *Given $E \subset (\mathbb{Z}/p)^*$, let*

$$u = \sum_{e \in E} \zeta^e.$$

Let $H = \{k \in (\mathbb{Z}/p)^* \,:\, kE = E\}$. Then

$$\mathbb{Q}(u) = \mathbb{Q}(\zeta_p)^H,$$

$d = \deg(u/\mathbb{Q}) = [G : H]$, and $\mathrm{Gal}(\mathbb{Q}(u)/\mathbb{Q}) \cong \mathbb{Z}/d$.

For example:

If $p > 3$, then $\zeta + \zeta^2$ is a primitive element for $\mathbb{Q}(\zeta)$ over $\mathbb{Q}$.

Indeed, we just need to observe that if $\zeta^k + \zeta^{2k} = \zeta + \zeta^2$, with $k \in (\mathbb{Z}/p)^*$, then either $k = 1$ or $k = 2$ and $2k = 4 = 1 \bmod p$; in the latter case, $p = 3$.

When $p = 1 \bmod 4$, we can find an element of degree $(p-1)/4$: for example, if $p = 17$, then

$$H = \langle 4 \rangle = \{1, 4, 13, 16\} \cong \mathbb{Z}/4$$

is a subgroup of $(\mathbb{Z}/17)^* \cong \mathbb{Z}/16$, and

$$t = \zeta + \zeta^4 + \zeta^{13} + \zeta^{16}$$

is invariant exactly under H, so it has degree 4 over $\mathbb{Q}$ and $\mathrm{Gal}(\mathbb{Q}(t)/\mathbb{Q}) = \mathbb{Z}/4$.

3. Cubic polynomials. By a preliminary change of variables $x \mapsto x + c$, $c \in K$, any monic cubic polynomial $p(x) \in K[x]$ can be put into the form

$$p(x) = x^3 + ax + b.$$

As we have previously computed with resultants, we have

$$D(p) = -4a^3 - 27b^2,$$

and there are only two possibilities for the Galois group, S_3 and $A_3 \cong \mathbb{Z}/3$.

Theorem 13.18 *A cubic polynomial $p(x) \in K[x]$ has Galois group $\mathbb{Z}/3$ if its discriminant is a square in K, and S_3 otherwise.*

In the first case, only a cubic root is necessary to solve $p(x)$, provided a cube root of unity already lies in K. In the second case, a square–root and a cube–root are sufficient to solve $p(x)$, by which we mean find all of its roots.

Examples. The discriminant of $x^3 - 2$ is negative, so it is not a square in $\mathbb{Q}$ and the Galois group is S_3. This is the typical case; its splitting field has degree 6.

The discriminant of $x^3 - 3x + 1$, however, is $(4-1) \cdot 27 = 81 = 9^2$, so its Galois group is $\mathbb{Z}/3$. In this case for any root t, the splitting field is $\mathbb{Q}(t)$ which has degree 3 over $\mathbb{Q}$.

How to solve cubics. Let us see this principle at work by illustrating one of the best ways to solve cubic polynomials $p(x)$.

First, one can make a preliminary change of variables of the form $x \mapsto x+t$ to arrange that the sum of the roots of p is zero; that is, it takes the form

$$p(x) = x^3 + ax + b,$$

with discriminant $D(p) = -4a^3 - 27b^2$. Making a further change of variables of the form $x \mapsto sx$, we can arrange that $a = 3$. Then we are solving the equation

$$p(x) = x^3 - 3x = -b. \tag{13.2}$$

Now comes an elegant insight, related to the multiple angle formula for $\cos(3x)$: if we solve for u such that

$$u + u^{-1} = -b,$$

and then solve $v^3 = u$, then

$$(v + 1/v)^3 - 3(v + 1/v) = v^3 + 1/v^3 = -b.$$

So $p(v + 1/v) = 0$, and $x = v + 1/v$ solves the cubic (13.2).

What was the discriminant? In the course of the solution, we had to solve the quadratic equation

$$q(u) = u^2 + bu + 1 = 0,$$

with discriminant $D(q) = b^2 - 4$. Note that, since $a = -3$, this discriminant is proportional to

$$D(p) = -4a^3 - 27b^2 = 27(4 - b^2).$$

Thus $\sqrt{D(p)}$ and $\sqrt{D(q)}$ differ by a factor of $\sqrt{-27} = 3\sqrt{-3}$.

Let us assume that the cube root of unity ω lies in our ground field K. Equivalently, $\sqrt{-3} \in K$. This assumption allows us to find all the roots of $p(x)$ at once, since once we have one solution to $v^3 = u$, the others are given by ωv and $\omega^2 v$. Moreover, it implies that

$$D(p) \in K^2 \iff D(q) \in K^2,$$

and so the Galois group of $p(x)$ is S_3 exactly when the quadratic $q(u)$ has no roots in K.

4. Radical extensions. A field extension L/K is *radical* if it has the form

$$L = K(b), \text{ where } a = b^n \in K \text{ for some } n \geq 1.$$

Informally, one often writes $L = K(a^{1/n})$.

Theorem 13.19 *Let $L = K(a^{1/n})$, with a and ζ_n in K. Then L is the splitting field of $x^n - a$, and* $\mathrm{Gal}(L/K)$ *is cyclic.*

Proof. The polynomial $x^n - a$ splits because its roots are given by $\zeta_n^i a^{1/n}$, $i = 0, 1, 2, \ldots, n-1$; and the Galois group injects into $\mathbb{Z}/n$, by sending α to the unique i such that $\alpha(a^{1/n}) = \zeta^i a^{1/n}$. ∎

Remarks. We have not required that $x^n - a$ is irreducible, and it need not be; for example, the theorem applies to $a = 2$ with $n = 4$ and $K = \mathbb{Q}(i)$.

Note an important distinction here: we have realized the Galois group as a subgroup of $(\mathbb{Z}/n)$, not $(\mathbb{Z}/n)^*$. This result says little in the case where $a = 1$, since we are assuming that $\zeta_n \in K$.

Since $\zeta_n \in \mathbb{C}$, one might ask if we must assume that K is a subfield of $\mathbb{C}$. We need not be. A more invariant version of this hypothesis is that K contains the splitting field of $x^n - 1$; then ζ_n can be any *primitive* nth root of unity in K.

Example: A quartic extension. A chain of Galois extensions is not necessarily Galois. Consider the example:

$$F = \mathbb{Q} \subset \mathbb{Q}(\sqrt{2}) \subset \mathbb{Q}(\sqrt[4]{2}) = E$$

Each extension is quadratic, hence Galois, but E/F is *not Galois*. For example, the polynomial $X^4 - 2$ has a root in E but does not split completely there, since it lacks its complex roots.

To understand this polynomial in more detail, consider its splitting field $K = \mathbb{Q}(i, \sqrt[4]{2})$. Since K/E is quadratic, clearly $[K : F] = 8$. The Galois group acts effectively by permutations of the 4 roots of $X^4 - 2$.

If we think of the roots as forming a diamond in the complex plane, then the action is by the dihedral group D_4 of order 8.

To see this, note that the extension $K/\mathbb{Q}(i)$ has degree 4 and is generated by $\sqrt[4]{2}$. Thus the polynomial x^4-2 is irreducible over $\mathbb{Q}(i)$, and so their exists an automorphism ϕ of $K/\mathbb{Q}(i)$ sending $\sqrt[4]{2}$ to $i\sqrt[4]{2}$. Since this automorphism commutes with multiplication by i, it cyclically permutes the 4 roots of f.

Complex conjugation, on the other hand, provides another automorphism ψ of K that fixes two roots of f and exchanges the other two. The automorphisms ϕ and ψ generate D_4, a group of order 8, which must coincide with the full Galois group since $8 = [K : \mathbb{Q}]$.

The subfield E is fixed by complex conjugation; that is, $G^E = \mathbb{Z}/2$. This subgroup is *not normal*, which explains why E/F is not Galois.

This example is also instructive for the construction of automorphisms of a splitting field. Let $L = \mathbb{Q}(\sqrt{2})$. There is an automorphism ϕ of L/F given by $\sqrt{2} \mapsto -\sqrt{2}$. But this automorphism does *not* extend to E. The point is that E/L is the splitting field of $g(x) = x^2 - \sqrt{2}$. We have $\phi(g) = x^2 + \sqrt{2}$ which does *not* split in E. But g and $\phi(g)$ are both factors of $f(x) = x^4-2$, so $\phi(g)$ *does* split in the splitting field K, allowing the automorphism to extend.

5. Cyclic extensions. What can we say in general, when we know the Galois group is cyclic as above? The answer is simplest for cyclic extensions of prime order p, with $\zeta_p \in K$.

Theorem 13.20 *Let L/K be a Galois extension, with* $\mathrm{Gal}(L/K) = \mathbb{Z}/p$*, p prime, and suppose $\zeta_p \in K$. Then L/K is radical extension.*

Proof. Let $\alpha : L \to L$ be a generator of $\mathrm{Gal}(L/K)$. Then α is a linear operator on a p—dimensional vector space over K, satisfying $\alpha^p = 1$, or equivalently

$$\prod_{i=0}^{p-1}(\alpha - \zeta_p^i I) = 0.$$

Since $\mathrm{Ker}(\alpha - I) = K \subset L$, one of the other factors must have a nontrivial kernel. That is, we have an eigenvector $a \in L$ such that

$$\alpha(a) = \zeta_p^i a.$$

Then $b = a^p \in K$ and $L = K(a)$, since $a \notin K$ and $[K(a) : K]$ must divide p. ∎

The same result holds with p replaced by an integer $n > 0$, but the proof is more subtle and the result above will suffice for our purposes.

A cautionary tale. It is true that radical and cyclic extensions are closely related, but they are not the same, as the following two examples show.

1. Let $K = \mathbb{Q}(2^{1/3})/\mathbb{Q}$. Then K is a radical extension, but $K/\mathbb{Q}$ is not even Galois, let alone cyclic. If it were, $x^3 - 2$ would split completely in K, since it has one root in K; but its complex roots are missing from K.

   In fact, the splitting field L of $x^3 - 2$ has Galois group isomorphic to S_3. The field K is the fixed field of the involution $\tau(z) = \overline{z}$; thus $\mathrm{Gal}(L/K)$ is one of the three non-normal subgroups of S_3 generated by a transposition.

2. Let $p(x) = x^3 - 3x - 1$. Then its splitting field $L/\mathbb{Q}$ has cyclic Galois group $\mathbb{Z}/3$, but it is not a radical extension.

   The first statement follows from the fact that the discriminant of $D(p) = -4a^3 - 27b^2 = 4 \cdot 27 - 27 = 81$ is a square. As for the second, if L were a radical extension, it would have to split a polynomial of the form $x^3 - b$; but then we would have $\zeta_3 \in L$. Since ζ_3 has degree 2 over $\mathbb{Q}$, this contradicts the fact that $[L : \mathbb{Q}] = 3$.

These examples also show that the theory founders around issues of roots of unity, and in fact the two types of extensions L/K are the same if enough roots of unity lie in K.

6. Solvable extensions. We say L/K is a *solvable* field extension if L/K is Galois and $\mathrm{Gal}(L/K)$ is a solvable group. For example, the splitting field of any polynomial $p(x)$ of degree ≤ 4 is solvable, but a typical irreducible polynomial of degree 5 or higher does *not* have a solvable Galois group.

Solvable groups. Recall that a finite group G is *solvable* if we can find a nested sequence of subgroups

$$(e) = G_1 \subset G_2 \subset \cdots \subset G_n = G,$$

each normal in the next, such that G_{i+1}/G_i is abelian for all i. Such a sequence can always be refined so that $G_{i+1}/G_i \cong \mathbb{Z}/p_i$, where p_i is a prime dividing $|G|$.

Any subgroup or quotient group of a solvable group is solvable. If N is a normal subgroup of G, and G/N and N are solvable, so is G.

Examples. The symmetric group S_n is solvable iff $n < 5$. The proof of the 'insolvability of the quintic' will pivot on this transition.

Let us check this statement. Clearly S_3 is solvable, since $S_3/A_3 \cong \mathbb{Z}/2$ and $A_3 \cong \mathbb{Z}/4$. There is a normal subgroup $N \cong \mathbb{Z}/2 \times /zed/2$ in S_4 such that $S_4/N \cong S_3$, so S_4 is solvable. (Geometrically, S_4 is the symmetries group of a cube, and the map to S_3 is given by its action on pairs of opposite faces.)

On the other hand, the alternating groups A_n are simple for $n \geq 5$, hence they are not solvable, It follows that S_n is not solvable for $n \geq 5$.

From radical to solvable. By adjoining a suitable root of unity, we can always insure that a radical extension is solvable.

Theorem 13.21 *For any $a \in K$, the extension $L = K(\zeta_n, a^{1/n})$ is solvable.*

Proof. First, this extension is the splitting field of $(x^n - 1)(x^n - a)$, so it is Galois. As we have seen, the intermediate field $M = K(\zeta_n)$ is abelian over K, and L/M has cyclic Galois group, so altogether the extension is solvable. ∎

7. Composite extensions. It is useful to have a way of putting two field extensions together.

Let L_p/K denote the splitting field of a polynomial $p \in K[x]$. Given a pair of polynomials p and q, we can form the *composite* splitting field L_{pq}, and regard L_p and L_q as the subfields of generated by the roots of p and q respectively.

To describe how their Galois groups interact, we formulate:

Lemma 13.22 *Let N_1, N_2 be normal subgroups of a group G, with $N_1 \cap N_2 = (e)$. Then we have a natural injective map*

$$G \to (G/N_1) \times (G/N_2),$$

and natural injective maps

$$N_1 \to G/N_2 \quad \textit{and} \quad N_2 \to G/N_1.$$

If one of these three maps is an isomorphism, so are the others.

Proof for finite groups. Surjectivity in all 3 cases is equivalent to $|N_1| \cdot |N_2| = |G|$. ∎

(This lemma is also true for infinite groups.)

Theorem 13.23 *Given any pair of polynomials $p, q \in K[x]$, we have an injective map*

$$\mathrm{Gal}(L_{pq}/K) \to \mathrm{Gal}(L_p/K) \times \mathrm{Gal}(L_q/K),$$

and an injective map

$$\mathrm{Gal}(L_{pq}/L_q) \to \mathrm{Gal}(L_p/K). \tag{13.3}$$

Proof. Let $G = \mathrm{Gal}(L_{pq}/K)$, and let N_1 and N_2 be the normal subgroups with fixed fields L_q and L_p respectively. Any element of $N_1 \cap N_2$ fixes the all the roots of pq, and hence it this intersection is the trivial group. The result then follows from the Lemma above, since

$$G/N_1 \cong \mathrm{Gal}(L_p/K),\ G/N_2 \cong \mathrm{Gal}(L_q/K), \quad \text{and} \quad N_2 \cong \mathrm{Gal}(L_{pq}/L_q).$$

∎

Corollary 13.24 *If L_p/K and L_q/K are solvable extensions of K, then so is L_{pq}.*

Equation (13.3) captures the intuitively clear statement that the Galois group of p can only become simple when the field K is enlarged.

8. Towers of radical extensions. We can now begin the proof of Galois's Theorem 13.1.

We say a polynomial $p(x) \in K[x]$ can be *solved by radicals* if we have a tower of field extensions,

$$L = K_n \supset K_{n-1} \supset \cdots \supset K_1 = K,$$

such that K_{i+1}/K_i is a radical extension for each i, and $p(x)$ has a root in L. This means there is a formula for a root of p involving only nested radicals and numbers in K.

Theorem 13.25 *If the Galois group of $p(x) \in K[x]$ is solvable, then $p(x)$ is solvable by radicals.*

Proof. Let L_p/K be the splitting field of p, with Galois group $G = \mathrm{Gal}(L_p/K)$. Let

$$d = [L_p : K] = |\,\mathrm{Gal}(L_p/K)|,$$

and let N be the product of the primes dividing d. Let $q(x) = x^N - 1$, and let

$$L_q = K(\zeta_N)$$

be its splitting field. Then L_q/K is a radical extension.

As we have seen, we can regard L_p and L_q as subfields of the compositum, L_{pq}, and we have an injective map,

$$G = \mathrm{Gal}(L_{pq}/L_q) \hookrightarrow \mathrm{Gal}(L_p/K).$$

Since $\mathrm{Gal}(L_p/K)$ is solvable, so is G; and $|G|$ divides d.

Choose a nested sequence of subgroups

$$(e) = G_1 \subset G_2 \subset \cdots \subset G_n = G,$$

each normal in the next, such that $G_i/G_{i+1} \cong \mathbb{Z}/p_i$ for all i, where p_i is a prime divisor of d and hence N.

Taking the corresponding fixed fields, we obtain a nested sequence of fields

$$L_q = K_1 \subset K_2 \subset \cdots K_n = L_{pq},$$

such that K_{i+1}/K_i is a cyclic Galois extension of degree p_i, for each i.

Since $\zeta_N \in L_q$, we also have $\zeta_{p_i} \in K_i$ for all i. Thus by Theorem 13.20, K_{i+1}/K_i is a radical extension for each i. Of course $K_1/K = K(\zeta_N)/K$ is also a radical extension, since $\zeta_N^N = 1 \in K$.

Thus $p(x)$ is solvable by radicals over K, since the field L_{pq} is obtained from K by a chain of radical extensions, and $p(x)$ has a root (in fact all its roots) in L_{pq}. ∎

Since any subgroup of S_n, $n < 5$, is solvable, we have:

Corollary 13.26 *Any polynomial of degree less than 5 in $K[x]$ is solvable by radicals.*

Examples of solvable numbers. Let us say an algebraic number $t \in \mathbb{C}$ is solvable if its Galois group is solvable. Then all the numbers that can be constructed by ruler and compass are solvable. So are all the cycltomic numbers.

However, it can be shown that a 'typical' algebraic number of degree d has Galois group S_d (as big as possible), and hence most numbers of degree $d \geq 5$ over $\mathbb{Q}$ are not solvable.

Aside: Why radicals? One might ask why solution by radicals was classically considered particularly explicit or elementary.

One reason is that the roots of $x^n = a$ can be easily expressed in terms of *logarithms*. Name, they are given in terms of a fixed choice of $\log(a)$ by

$$x_k = \exp\left(\frac{2\pi i k + \log(a)}{n}\right),$$

$k = 0, 1, 2, \ldots, n-1$.

Another reason is that the roots of $p(x) = x^n - a$ can be rapidly and reliably found by iterating Newton's method,

$$F(x) = x - \frac{p(x)}{p'(x)} = x \cdot \left(\frac{(n-1) + a/x^n}{n}\right).$$

In fact when a is real, this iteration converges to $\sqrt[n]{a}$ for any initial guess $x > 0$, and the number of digits of accuracy doubles with each iteration. For complex values of a, convergence is still guaranteed, provided the initial guess is chosen at random.

9. Finale: The quintic is unsolvable. We have just seen that all solvable extensions of K can be constructed by radicals. The converse statement is:

Theorem 13.27 *Suppose $p(x)$ is solvable by radicals. Then its Galois group is solvable.*

The proof of this result is not difficult, but it requires a fair amount of book keeping. The main issue is that the top field in a chain of radical extensions $K_1 \subset K_2 \subset \cdots \subset K_n$ need not be Galois over K_1. To use Galois theory effectively, one must enlarge each K_i to a splitting field over K_1, and then verify that $\mathrm{Gal}(K_n/K_1)$ is solvable.

The case of A_5. Instead, we will give a more focused proof of:

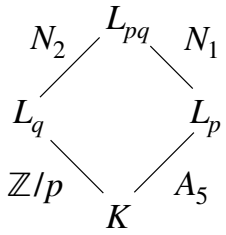
Theorem 13.28 *Let $p(x) \in K[x]$ be an irreducible quintic polynomial with Galois group S_5 or A_5. Then $p(x)$ cannot be solved by radicals.*

We begin by showing that prime degree extensions of K cannot improve the Galois group of p. For simplicity we assume the Galois group is A_5. This is the key step in the argument.

Lemma 13.29 *Let L_p/K be the splitting field of a quintic $p(x) \in K[x]$, with Galois group A_5. Let L_q/K be a Galois extension of prime degree p. Then*

$$\mathrm{Gal}(L_{pq}/L_p) = A_5.$$

Proof. We appeal to Lemma 13.22 as usual, with $G = \mathrm{Gal}(L_{pq})$, and N_1 and N_2 the normal subgroups whose fixed fields are L_p and L_q. Then $N_1 \cap N_2 = (e)$, $G/N_1 \cong A_5$ and $G/N_2 \cong \mathbb{Z}/p$. See the diagram below.



Consider the injective map

$$N_1 \to G/N_2 \cong \mathbb{Z}/p.$$

If this map is an isomorphism, then by the Lemma, so is the map

$$N_2 = \mathrm{Gal}(L_{pq}/L_p) \to G/N_1 \cong A_5,$$

and we are done.

Otherwise, N_1 must be the trivial group. But then $G = G/N_1 = A_5$, and N_2 is a normal subgroup of G with $G/N_2 = \mathbb{Z}/p$. This contradicts the simplicity of A_5. ∎

Lemma 13.30 *Suppose the Galois group of $p(x) \in K[x]$ is A_5, and L/K is a solvable, Galois extension. Then the Galois group of $p(x)$ over L is still A_5.*

Proof. Construct L/K using a tower of Galois extensions of prime order, and then apply the Lemma above to each stage in the tower. ∎

Lemma 13.31 *Similarly, if M/K is a radical extension, then the Galois group of p over M remains A_5.*

Proof. Suppose M is obtained by adding a root of $q(x) = x^n - a$. Let $r(x) = x^n - 1$, and let L_{qr}/K be the splitting field of qr. Since L_r/K is abelian and L_q/L_r is cyclic, the extension L_{qr}/K is solvable. Moreover M is a subfield of L_{qr}. Since the Galois group of $p(x)$ remains A_5 over L_{qr}, it also remains A_5 over M. ∎

Proof of Theorem 13.28. Let $p(x) \in K[x]$ be an irreducible quintic with Galois group S_5 or A_5. Adjoining the square–root of $D(p)$ to K, we can assume its Galois group is A_5. The preceding Lemma shows that a radical extension of K does not change the Galois group of p. By induction, the Galois group remains A_5 over any field L/K obtained by a sequence of radical extensions. But if p were to have a root in L, its Galois group would be reduced to a subgroup of S_4. Since this is not the case, $p(x)$ cannot be solved by radicals. ∎

An unsolvable quintic. It remains to show that there *exists* an irreducible quintic polynomial $p(x) \in \mathbb{Q}[x]$ with Galois group S_5.

A nice example is provided by

$$p(x) = x^5 - 4x - 2.$$

This polynomial is irreducible by Eisenstein's criterion, and it has exactly three real roots (see Figure 8). Thus its Galois group $G \subset S_5$ contains a transposition, provided by complex conjugation. We also know that G acts transitively on the 5 roots of p; these two facts together are enough to show that $G = S_5$.

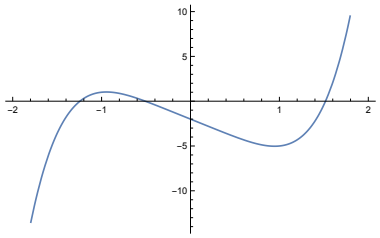


Figure 8. The quintic polynomial $x^5 - 4x - 2$ has exactly three real roots.

14 Problems

1. Let G be the free group $\langle a, b, c\rangle$. How many reduced words of length n are contained in G?

2. Let $G = \langle x, y \ : \ xyx = yxy\rangle$. Prove that G is an infinite group. (Hint: find a surjective homomorphism $f : G \to \mathbb{Z}$.)

3. Let $G = \langle x, y \ : \ x^2 = y^2 = e, xyx = yxy\rangle$. Prove that G is isomorphic to the symmetric group S_3. (Hint: to show $|G| \leq 6$, first show that $(xy)^3 = e$.)

4. Does every finite group have a finite presentation?

5. Is $\mathbb{Q}$ a finitely generated group?

6. What are the possible signatures (p, q) for a symmetric matrix $A \in \mathrm{M}_n(\mathbb{R})$ if $\det(A) > 0$?

7. Consider the conic in $\mathbb{R}^2$ defined by $x^2 + Axy + y^2 + x = 2$. For which values of A is this conic an ellipse? For which values of A is the conic empty?

8. What type of real quadric is the surface defined by $z^2 + xy = 1$? By $z^2 + xy = -1$? By $x^2 + y^2 + z^2 - xy = 1$? (Possible answers: ellipsoid, 1–sheeted hyperboloid, 2–sheeted hyperboloid.)

9. Consider the quadratic forms defined by the matrices $A_1 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $A_2 = \left(\begin{smallmatrix} 4 & 1 \\ 1 & 6 \end{smallmatrix}\right)$. Show that A_1 and A_2 are equivalent over $\mathbb{R}$, but not over $\mathbb{Q}$. (Equivalent over k means $B^t A_1 B = A_2$ for some $B \in \mathrm{GL}_2(k)$.)

10. (Bonus problem.) Prove that the free groups $F_2 = \langle a, b\rangle$ and $F_3 = \langle a, b, c\rangle$ are not isomorphic. (A brief, correct answer is acceptable.)

11. (i) What is the multiplicative inverse of 7 in the ring $\mathbb{Z}/101$? (ii) Find integers $r, s \in \mathbb{Z}$ such that $7r + 101s = \gcd(7, 101) = 1$.

12. Consider the evaluation map

$$E : \mathbb{R}[x] \to \mathbb{R}[x]/(x^2 + 1) \cong \mathbb{C}.$$

Does $E(p) = p(i)$ or $p(-i)$? Discuss this seemingly paradoxical question, and resolve it.

13. Let A be a ring. We say $x \in A$ is *nilpotent* if $x^k = 0$ for some $k \geq 0$.

(i) Prove that if x is nilpotent, then $1 + x$ is a unit. (ii) Show that the set N of all nilpotent elements in A is an ideal. (iii) Show that A/N has no nilpotent elements, other than zero.

14. Let $I \subset \mathbb{Z}[i]$ be an ideal, $I \neq (0)$. Prove that I contains a nonzero integer $n \in \mathbb{Z}$.

15. Let $A = \mathbb{Z}[i]/(1 + 3i)$. (i) Prove that 10 belongs to the principal ideal $(1 + 3i)$. (ii) Prove that A is isomorphic, as a ring, to $\mathbb{Z}/10$.

16. Show that $\mathbb{R}[x]/(x^2 - 1)$ is isomorphic to the product ring, $\mathbb{R} \times \mathbb{R}$.

17. Determine the automorphisms of the ring $\mathbb{Z}[x]$. That is, find all bijective ring homomorphisms $f : \mathbb{Z}[x] \to \mathbb{Z}[x]$.

18. Let E be a set, and let Let $\mathcal{P}(E)$ be the collection of all subsets of E. Given $A, B \in \mathcal{P}(E)$, define $A \cdot B = A \cap B$ and $A + B = (A \cup B) - (A \cap B)$.

(i) Prove that with these operations, $\mathcal{P}(E)$ is a ring. (ii) Given $x \in E$, show that $I = \{A \subset E \ : \ x \notin A\}$ is a maximal ideal. What is the field $\mathcal{P}(E)/I$? (iii) Show that $\mathcal{P}(E) \cong (\mathbb{Z}/2)^E$, the ring of functions $f : E \to \mathbb{Z}/2$.

19. (Bonus problem.) Show that ring $\mathcal{P}(\mathbb{N})$ contains a maximal ideal I that is not of the form $I = \{A \subset E \ : \ x \notin A\}$.

20. Let I and J be ideals in a ring A, and let $IJ = \{xy \,:\, x \in I, y \in J\}$. (i) Show that if I and J are principal ideals, then so is IJ. (ii) Give an example of ideals I, J in $\mathbb{R}[x, y]$ such that IJ is not an ideal. (iii) Identify, in your example, the smallest ideal containing IJ.

21. Let S be a subset of a ring A such that $AS \subset S$. Show that the additive group $I \subset A$ generated by S is an ideal.

22. Let A be a subring of a ring B. Given $b \in B$, let

$$I = \{a \in A \,:\, b|a \text{ in } b\}.$$

(This means $bc = a$ for some $c \in B$.) Prove that I is an ideal in A.

23. Consider the polynomials $p, q \in \mathbb{Q}[x]$ given by

$$p(x) = x^4 + 2x^3 + x^2 + 3x + 2 \quad \text{and} \quad q(x) = 2x^4 + x^3 + 2x^2 + 3x + 1.$$

Using long division, find the $\gcd(p, q)$. That is, find the unique monic polynomial $r(x)$ such that $(p, q) = (r)$ as ideals in $\mathbb{Q}[x]$.

24. Suppose two monic, irreducible polynomials $p, q \in \mathbb{Q}[x]$ have a common zero $z \in \mathbb{C}$. Show that $p = q$. (Hint: what are the possibilities for the ideal (p, q)?)

25. Prove that a finite integral domain is a field.

26. Show that $I = (2, x)$ is not a principal ideal in $\mathbb{Z}[x]$.

27. Is the ideal $I = (3, x^2 + x + 1)$ maximal in $\mathbb{Z}[x]$? Why or why not?

28. What is the quotient ring $\mathbb{Q}[x]/(x^2 + 1, x^2 + x + 1)$?

29. Let I be an ideal in a ring A, and suppose $fg \in I$ but $f \notin I$ and $g \notin I$. (i) Show that $J = \langle I, f \rangle \neq A$. (Here J is the smallest ideal in A containing I and f.) (ii) Using (i), prove that every maximal ideal in A is prime.

30. Let $p(x) = x^3 + 5x + 1$. (i) Prove that $\mathbb{Z}[x]/(p)$ is an integral domain. (ii) Find the inverse y of x in $\mathbb{Z}[x]/(p)$. (Express y as a polynomial in x.)

31. (Bonus problem.) Suppose that $f(x, y) \in \mathbb{C}[x, y]$ vanishes on the zero locus $V \subset \mathbb{C}^2$ of an irreducible polynomial $p(x, y) \in \mathbb{C}[x, y]$. Prove that $p(x, y)$ divides $f(x, y)$.

32. Let $p(x) = x^3 + x + 1$. (i) Show that $\mathbb{Q}[x]/(p)$ is a field. (ii) Find the inverse y of $x + 1$ in $\mathbb{Q}[x]/(p)$. (Express y as a polynomial in x.)

33. Prove that $A = \mathbb{Z}[\sqrt{3}]$ is a Euclidean domain, with $\sigma(a + b\sqrt{3}) = |a^2 - 3b^2|$. (Hint: first check that $\sigma(xy) = \sigma(x)\sigma(y)$ for all $x, y \in A$.)

34. Let I, J be ideals in A, and suppose $I + J = A$. (i) Prove that the natural map $A \to (A/I) \times (A/J)$ is surjective. (This is a generalization of the Chinese remainder theorem.)

    (ii) Let $A = \mathbb{Q}[x]$, $I = (x^2 + 1)$, $J = (x^2 - 2)$. Find a polynomial $p \in A$ such that p maps to $(1, 0)$ in $A/I \times A/J$.

35. Prove that for any field K, there are infinitely many distinct monic, irreducible polynomials in $K[x]$. (Hint: mimic the proof that there are infinitely many prime numbers in $\mathbb{Z}$.)

36. (i) Prove that determinant polynomial,

$$q(a, b, c, d) = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc,$$

    is irreducible in the unique factorization domain $\mathbb{R}[a, b, c, d]$.

    (ii) What is the signature of $q(a, b, c, d)$ as a quadratic form on $\mathbb{R}^4$?

37. Let $n = a_0 a_1 \dots a_n$ be a prime number written in base 10, with $a_0 \geq 2$ and $1 \leq a_i \leq 9$ for $i = 1, \dots, n$, and let

$$p(x) = a_0 x^n + x_1 x^{n-1} + \cdots + a_n \in \mathbb{Z}[x].$$

    (Thus $p(x) = 3x^3 + 1$ if $n = 3001$.) Prove that $p(x)$ is an irreducible polynomial.

38. Let $A = K[t, t^{-1}]$ be the ring of Laurent polynomials over a field K. (Its elements have the form $\sum_{-n}^{n} a_i t^i$, $a_i \in K$.)

    (i) What are the units in A? (ii) Show that A is a PID.

39. Let $p(x) = x^d + a_1 x^{d-1} + \cdots a_d$ be a polynomial in $\mathbb{Z}[x]$. Suppose $y \in \mathbb{Q}$ is a root of $p(x)$. Show that in fact $y \in \mathbb{Z}$, and y is a divisor of a_d.

40. Compute the discriminant of $p(x) = ax^2 + bx + c$ using resultants.

41. (Bonus problem.) Let $p(x)$ and $q(x)$ be monic polynomials in $\mathbb{C}[x]$, with roots (a_i) and (b_j) respectively. (i) Prove that the resultant of p and q satisfies

$$\mathrm{Res}(p, q) = c \prod_{i,j} (a_i - b_j),$$

where c depends only on the degrees of p and q. (ii) Find the value of c.

42. (Bonus problem.) Let $C(\mathbb{R})$ denote the ring of real–valued continuous functions $f : \mathbb{R} \to \mathbb{R}$, and let $A \subset C(\mathbb{R})$ be the subring generated by $\sin(t)$ and $\cos(t)$.

(i) Show that $A \cong \mathbb{R}[x, y]/(x^2 + y^2 - 1)$.

(ii) Show that A is not a UFD.

(iii) Show that $B = \mathbb{C}[x, y]/(x^2 + y^2 - 1)$ is a PID, and hence a UFD.

(iv) Find the units in B.

(Hint: show that B is isomorphic to the ring of Laurent polynomials, $\mathbb{C}[u, u^{-1}]$. It may be useful to think of u as $\exp(it)$.)

43. Let $\overline{\mathbb{Q}} \subset \mathbb{C}$ denote the field of algebraic numbers.

(i) Prove that if $K \subset \mathbb{C}$ is a number field and $a \in \overline{\mathbb{Q}}$, then $K(a)$ is a number field (meaning $[K(a) : \mathbb{Q}]$ is finite).

(ii) Prove that $\overline{\mathbb{Q}}$ is *algebraically closed*: that is, any nonconstant polynomial $p(x) \in \overline{\mathbb{Q}}[x]$ has a root in $\overline{\mathbb{Q}}$. (You may use the fact that $\mathbb{C}$ is algebraically closed.)

44. (i) Show that for any integer $n > 0$, $2\cos(\pi/n)$ and $2\sin(\pi/n)$ are algebraic integers.

(ii) Show that if $2\cos(\pi/n) \in \mathbb{Q}$, then $n = 1, 2$ or 3.

(iii) Show that $2\cos(\pi/5)$ is the golden ratio, $(1 + \sqrt{5})/2$.

(iii) Show that $2\sin(\pi/5) = \sqrt{5 - \sqrt{5}}\sqrt{2}$. (Hint: compute the norm and trace of its square.)

45. Factor $n = 34$ into primes in $\mathbb{Z}[i]$.

46. Let $p(x) = x^2 + ax + b$ be the minimal polynomial of $t \in \mathbb{Z}[\sqrt{-5}]$.

    (i) Show that $N(t) = b$.

    (ii) Show that $N(t) \neq 2$ or $3 \bmod 5$.

    (iii) More generally, show that $p(n) \neq 2$ or $3 \bmod 5$ for all $n \in \mathbb{Z}$.

47. (i) Give an algebraic equation relating $\mathrm{tr}(m)$, $\mathrm{tr}(m^2)$ and $\det(m)$ that is valid for all matrices $m \in \mathrm{M}_2(\mathbb{R})$.

    (ii) Give an algebraic equation relating $\mathrm{tr}(a)$, $\mathrm{tr}(a^2)$ and $N(a)$, that is valid for all a in the quadratic field $\mathbb{Q}(\sqrt{d})$.

48. Recall that the Eisenstein integers $\mathbb{Z}[\omega]$ are a UFD. Let p be a prime.

    (i) Show that $\mathbb{F}_p$ contains a nontrivial cube–root of unity if and only if $p = 1 \bmod 3$.

    (ii) Show that if $p > 3$, then p is a prime in $\mathbb{Z}[\omega]$ if and only if $p = 2 \bmod 3$.

    (iii) Factor 3 into primes in $\mathbb{Z}[\omega]$.

49. Let $M \subset \mathbb{C}[x, y]$ be the maximal ideal consisting of all polynomials p that vanish at the origin in $\mathbb{C}^2$. Prove that $M \cdot M$ consists of all p whose first derivatives, dp/dx and dp/dy, also vanish at the origin.

50. Let π be a prime in $\mathbb{Z}[i]$, with $N(\pi) = p$ an ordinary prime. Prove that π and $\overline{\pi}$ are associates if and only if $p = 2$. (Keep in the mind that there are only four units in $\mathbb{Z}[i]$.)

51. How many ways can one write $a^2 + b^2 = n$, with $(a, b) \in \mathbb{Z}^2$, when (i) $n = 2$, (ii) $n = 17$, (iii) $n = 19$, and (iv) $n = 117$?

    (Note that you are counting the number of integral points on a circle of a given radius.)

52. (Bonus problem.) Let p be a prime, with $p = 1 \bmod 4$. Justify the following algorithm to solve the equation $p = a^2 + b^2$.

    (i) Show that if $x \in \mathbb{F}_p$ is not a square $\bmod p$, then $y = x^{(p-1)/4}$ satisfies $y^2 = -1 \bmod p$.

    (ii) Given y as in (i), let $\gcd(p, i + y) = a + bi$, the gcd taken in the Euclidean domain $\mathbb{Z}[i]$. Show that $a^2 + b^2 = p$.

53. Consider the ring $A = \mathbb{Z}[10i] \subset K = \mathbb{Q}(i)$, $i = \sqrt{-1}$. (i) Find an ideal $I \subset A$ such that $I \cdot \overline{I}$ is not principal. (ii) Show geometrically that, as a lattice in $\mathbb{C}$, I is not similar to A.

    (Here $\overline{I} = \{a - bi : a + bi \in I\}$. This exercise illustrates the importance of working with the full ring of integers $\mathcal{O}_K = \mathbb{Z}[i]$.)

54. Let K be a quadratic number field. Prove that for any $a \in \mathcal{O}_K$, $a \neq 0$, we have $|\mathcal{O}_K/(a)| = |N(a)|$.

55. Consider the ideals

$$I = (2, 1 + \sqrt{5}) \quad \text{and} \quad J = (3, 1 + \sqrt{5})$$

    in the ring of integers

$$\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}] \subset K = \mathbb{Q}(\sqrt{-5}).$$

    (i) Show that I and J represent the same ideal class. That is, find $a \in K$ such that $aI = J$.

    (ii) Compute $N(I)$ and $N(J)$, and conclude that I and J are prime.

    (iii) Factor (6) into a product of prime ideals in $\mathcal{O}_K$.

56. (Bonus problem.) Factor (11) into a product of prime ideals in $\mathbb{Z}[\sqrt{-5}]$.

57. Let $\mathcal{O}_K$ be the ring of integers in $K = \mathbb{Q}(\sqrt{-7})$. (i) Find all the normalized lattices L containing $\mathcal{O}_K$. (Normalized means that $1 \in L$ and $|L| = 1$). (ii) Which of these are fractional ideals for $\mathcal{O}_K$? (iii) Conclude that K has class number one.

58. Prove that $K = \mathbb{Q}(\sqrt{-10})$ has class number $h(K) = 2$. Give explicit representatives of both ideal classes.

59. Let $K = \mathbb{Q}(\sqrt{-10})$. Is the ideal (p) prime in $\mathcal{O}_K$, when $p = 17$?

60. Consider the ideal $I = (5, \sqrt{-10}) \subset \mathbb{Z}[\sqrt{-10}]$. Is I prime? Principal? Justify your answer.

61. Prove that $K = \mathbb{Q}(\sqrt{-67})$ has class number one.

62. Let $P \subset \mathcal{O}_K$ be a prime ideal in a quadratic imaginary field. Prove that its norm $N(P)$ is given by p or p^2, for some ordinary prime p.

63. Let K be a field. An *order* $A \subset K$ is a subring such that (i) $\mathbb{Q} \cdot A = K$ and (ii) all elements of A are algebraic integers.

    Find all the orders in $\mathbb{Q}(\sqrt{5})$. What is the maximal (largest) order?

64. Determine, for all ordinary primes $p \leq 11$, if the equation $|a^2 - 2b^2| = p$ has a solution $(a, b) \in \mathbb{Z}^2$. Find a solution when one exists. (Hint: use arithmetic in the UFD $\mathbb{Z}[\sqrt{2}]$.)

65. (Bonus problem.)

    (i) It is known that the complex quadratic field $K = \mathbb{Q}(\sqrt{-2})$ has class number one. Show that whenever an ordinary prime p splits in K, there exists an integral solution to the equation $a^2 + 2b^2 = p$. Give an explicit solution for $p = 19$.

    (ii) It is known that the real quadratic field $K = \mathbb{Q}(\sqrt{7})$ has class number one. Verify that the prime $p = 19$ splits in K, but there is no integral solution to the equation $a^2 - 7b^2 = 19$. Explain how this is possible.

66. (i) Draw a portion of the lattice in $\mathbb{R}^2$ generated by $(1, 1)$ and $(\sqrt{3}, -\sqrt{3})$, as well as the hyperbolas given by $|xy| = 1$.

    (ii) Prove that $\epsilon = 2 + \sqrt{3}$ is the fundamental unit of $K = \mathbb{Q}(\sqrt{3})$. (I.e. prove that ϵ is the smallest element of $\mathcal{O}_K$ such that $\epsilon > 1$ and $|N(\epsilon)| = 1$.)

67. (Bonus problem.) Let $L \subset \mathbb{C}$ be a lattice such that $1 \in L$ and $|z|^2 \in \mathbb{Z}$ for all $z \in L$. Prove that L is contained in a quadratic imaginary field.

    Formulate and prove an analogous theorem for lattices $L \subset \mathbb{R}^2$ with $xy \in \mathbb{Z}$ for all $(x, y) \in L$.

68. Given a field extension L/K, the group $\mathrm{Aut}(L/K)$ consists of automorphisms of L (as a ring) that fix K pointwise.

    (i) Prove that $\mathrm{Aut}(\mathbb{R}/\mathbb{Q})$ is trivial.

    (ii) Prove that $\mathrm{Aut}(\mathbb{C}/\mathbb{R})$ is $\mathbb{Z}/2$, with generator $\alpha(z) = \overline{z}$.

69. (i) What is a basis for $\mathbb{Q}(2^{1/3})$ over $\mathbb{Q}$?

    (ii) Using division of polynomials, find $q(x) \in \mathbb{Z}[x]$ and $r \in \mathbb{Z}$ such that $(x + 1)q(x) = (x^3 - 2) + r$.

(iii) Express $1/(2^{1/3}+1)$ in the form $a+b\cdot 2^{1/3}+c\cdot 2^{2/3}$, with $a, b, c \in \mathbb{Q}$.

70. (i) Prove that for any angle θ, we have

$$(2\cos\theta)^3 = (2\cos 3\theta)^3 + 3(2\cos\theta).$$

(Hint: use the fact that $2\cos\theta = z + 1/z$, where $z = \exp(i\theta)$.

(ii) Prove that $t = 2\cos\pi/9$ is a root of $p(x) = x^3 - 3x - 1$.

(iii) Prove that $p(x)$ is irreducible in $\mathbb{Q}[x]$.

71.

72. (i) What is a basis for $\mathbb{Q}(2^{1/3})$ over $\mathbb{Q}$?

(ii) Using division of polynomials, find $q(x) \in \mathbb{Z}[x]$ and $r \in \mathbb{Z}$ such that $(x+1)q(x) = (x^3 - 2) + r$.

(iii) Express $1/(2^{1/3}+1)$ in the form $a+b\cdot 2^{1/3}+c\cdot 2^{2/3}$, with $a, b, c \in \mathbb{Q}$.

73. Prove that if $p = 2^n + 1$ is prime, and $n > 0$, then n is a power of 2.

74. Prove that for any prime $p > 2$ and $n > 0$, the unit group of $\mathbb{Z}/p^n$ is cyclic of order $(p-1)p^{n-1}$.

75. Prove that the construction shown in Figure 6 really does produce a regular pentagon. (Hint: it may be useful to refer to Problem 44.)

76. Let $L = \mathbb{C}(t)$ be the field of rational functions over $\mathbb{C}$, and let $K = \mathbb{C}(t^2) \subset L$. Prove that L is isomorphic to K. What is $[L : K]$?

77. Given a square, is it possible to construct (with ruler and compass) a pentagon with the same area?

78. Suppose we are given the points $\mathbb{Z} \cup i\mathbb{Z} \subset \mathbb{C}$. What are all the points we can then construct, using only a straightedge?

79. Let $K = \mathbb{F}_p(t)$ be the field of rational functions over $\mathbb{F}_p$, and consider the field endomorphism $\sigma : K \to K$ defined by $\sigma(x) = x^p$. Prove that σ is injective but not surjective.

80. (Bonus.) Let p be a prime and let $L = \mathbb{F}_p[x, y]$ be a field extension of $K = \mathbb{F}_p[x^p, y^p]$.

(i) Prove that $[L : K] = p^2$.

(ii) Prove that for any $f \in L$ we have $f^p \in K$.

(iii) Prove that $[K(f) : K] = 1$ or p for any $f \in L$. (Thus L has no primitive element.)

(iv) Given $c \neq 0 \in K$, let $M_c = K(x + cy)$. Prove that $M_c \neq M_d$ whenever $c \neq d$.

(Hint: if $M = M_c = M_d$ then we have $(c-d)y \in M$ and hence $M = L$, contradicting (iii)).

(v) Conclude that there are infinitely many distinct fields M with $K \subset M \subset L$. This example illustrates the failure of Galois theory in characteristic p.

81. How many irreducible polynomials of degree 6 are there in $\mathbb{F}_3[x]$?

82. Let $q = p^n$, $n > 1$. Prove or disprove: if $\mathbb{F}_q = \mathbb{F}_p[a]$, then a generates the cyclic group $\mathbb{F}_q^\times$.

83. Let $q = p^n$, and let

$$\mathbb{F}_q^* = \{a \in \mathbb{F}_q \ : \ \mathbb{F}_q = \mathbb{F}_p[a]\}$$

denote the set of generators for $\mathbb{F}_q$ over $\mathbb{F}_p$. Show that

$$|\mathbb{F}_q^*| = \sum_{d|n} \mu(n/d)p^d.$$

(Hint: first show p^n is the sum over $d|n$ of the number of generators of the field with p^d elements.)

84. Prove that every irreducible polynomial in $\mathbb{F}_p[x]$ is separable.

In the Galois theory problems that follow, all fields are assumed to have characteristic zero.

85. Let $K = \mathbb{Q}(\sqrt{2}, \sqrt{5})$.

(i) Prove that $\sqrt{5} \notin \mathbb{Q}(\sqrt{2})$, and hence $[K : \mathbb{Q}] = 4$.

(ii) Find an explicit (primitive) element $t \in K$, such that $K = \mathbb{Q}(t)$.

(iii) Find the minimal polynomial $p(x) \in \mathbb{Q}[x]$ for t.

86. Let $K = \mathbb{Q}(i)$ and let $p(x) = x^4 - 5$.

(i) Prove that $p(x)$ is irreducible in $K[x]$.

(ii) Prove that $L = K(5^{1/4})$ is the splitting field of $p(x)$ over K.

(iii) Prove there is an automorphism $\alpha \in \mathrm{Gal}(L/K)$ such that $\alpha(5^{1/4}) = i \cdot 5^{1/4}$.

(iv) Prove that $\mathrm{Gal}(L/K)$ is isomorphic to $\mathbb{Z}/4$.

87. (i) Prove that for any angle θ, we have

$$(2\cos\theta)^3 = (2\cos 3\theta)^3 + 3(2\cos\theta).$$

(Hint: use the fact that $2\cos\theta = z + 1/z$, where $z = \exp(i\theta)$.

(ii) Prove that $t = 2\cos\pi/9$ is a root of $p(x) = x^3 - 3x - 1$.

(iii) Prove that $p(x)$ is irreducible in $\mathbb{Q}[x]$.

88. An algebraic integer $t \in \mathbb{C}$ is an algebraic *unit* if $1/t$ is also an algebraic integer.

Prove that t is an algebraic unit if and only if its minimal polynomial $p(x) \in \mathbb{Z}[x]$ satisfies $|p(0)| = 1$. Assuming this, find a polynomial $q(x) \in \mathbb{Z}[x]$ such that $q(t) = 1/t$.

89. (Bonus problem.) Let $\zeta_n = \exp(2\pi i/n)$.

(i) Prove that whenever $n > 1$ is odd, $1 + \zeta_n$ is an algebraic unit.

(ii) Prove that $2\cos(\pi/n)$ is also an algebraic unit.

90. Let $L = K(a_1, \ldots, a_n)$ be a finite extension of K. Prove that $L = K[a_1, \ldots, a_n]$; that is, every element of L can be expressed as a *polynomial* in $(a_1, \ldots, a_n)$ with coefficients in K.

91. Show that quadratic extension L/K is a splitting field, and hence Galois. (Quadratic means $[L : K] = 2$.) What is $\mathrm{Gal}(L/K)$?

92. Prove that the extension $L = \mathbb{Q}(\sqrt{2}, \sqrt{5})$ of $K = \mathbb{Q}$ is a splitting field, and that $\mathrm{Gal}(L/\mathbb{Q}) \cong \mathbb{Z}/2 \times \mathbb{Z}/2$.

(Hint: since $\deg(L/K) = 4$, we have $4 \geq |\mathrm{Gal}(L/K)|$ and equality holds iff L/K is a splitting field. To find some elements of the Galois group, consider the quadratic extensions L/M where $M = \mathbb{Q}(\sqrt{2})$ or $\mathbb{Q}(\sqrt{5})$.)

93. Let $p(x)$ be a polynomial of degree d in $\mathbb{C}[x]$.

(i) Show that $|p^{-1}(a)| = d$ for all $a \in \mathbb{C}$ with at most $d-1$ exceptions.

(ii) Give, for each d, an example where there are $d-1$ exceptions.

94. (Bonus problem.) Let

$$L = \mathbb{C}(x), K_1 = \mathbb{C}(x^2) \quad \text{and} \quad K_2 = \mathbb{C}((x-1)^2).$$

Show that $[L : K_1] = [L : K_2] = 2$, but $[L : K_1 \cap K_2] = \infty$.

(Hint: show that if $f(x) \in K_1 \cap K_2$, then $f(x) = f(-x) = f(2-x)$, and from this conclude that $f(x)$ is constant.)

95. Let $p(x) = x^5 + 5x^3 + 1$ have roots $a_i \in \mathbb{C}$. Compute the value of $\sum_1^5 a_i^2$.

96. Let $s_i(a)$ denote the ith elementary symmetric function of $(a_1, \ldots, a_n)$.

(i) Prove that the formula $\sum a_i^2 = s_1^2 - 2s_2$ holds for all $n \geq 2$.

(ii) Given $d > 0$, prove that there is a single polynomial $F(s)$ such that $\sum_1^n a_i^d = F(s_1, \ldots, s_d)$ for *all* $n \geq d$.

97. Prove that if K is a field and $p(x) \in K[x]$ is a separable polynomial, then $K[x]/(p(x))$ is isomorphic to a product of finitely many fields, $F_1 \times \cdots \times F_n$.

98. Let $p(x) \in \mathbb{Q}[x]$ be a monic, irreducible cubic polynomial with only one real root. Prove that its Galois group is isomorphic to S_3.

99. Let L/K be the splitting field of an irreducible polynomial $p(x) \in K[x]$. Prove that the following are equivalent:

(a) One of the roots t of $p(x)$ in L is a primitive element; that is, $L = K(t)$.

(b) Every root of $p(x)$ in L is a primitive element.

(c) We have $|\operatorname{Gal}(L/K)| = \deg p(x)$.

(d) The group $\operatorname{Gal}(L/K)$ acts simply transitively on the roots $(t_1, \ldots, t_d)$ of $p(x)$. That is, for any i, j there is a unique $\alpha \in \operatorname{Gal}(L/K)$ such that $t_i^\alpha = t_j$.

100. Let G be a subgroup of S_5 that acts transitively and contains a transposition. Prove that $G = S_5$.

101. Prove the group theory result in Lemma 13.22, without any finiteness assumptions.

102. Let $p > 2$ be an odd prime, and let $\zeta_p = \exp(2\pi i/p)$.

(i) Show that any field $K \subset \mathbb{Q}(\zeta_p)$ is a Galois extension of $\mathbb{Q}$ and $\operatorname{Gal}(K/\mathbb{Q})$ is cyclic.

(ii) What is the Galois group of $K = \mathbb{Q}(\cos 2\pi/p)$ over $\mathbb{Q}$?

(iii) Let $p(x) \in \mathbb{Q}[x]$ be the minimal polynomial for $\cos 2\pi/p$. What are its roots?

103. Suppose the Galois group of $K/\mathbb{Q}$ is A_5. Prove that K does not contain any irrational number of the form $\sqrt{d}$, $d \in \mathbb{Z}$.

104. Let $\zeta = \zeta_7 = \exp(2\pi i/7)$.

(i) Prove that $K = \mathbb{Q}(\zeta + \zeta^2 + \zeta^4)$ is a quadratic extension of $\mathbb{Q}$.

(ii) Find an integer d such that $K = \mathbb{Q}(\sqrt{d})$.

(iii) Prove that $x^2 + 1$ does not split in $\mathbb{Q}(\zeta)$.

(iv) What is the degree of $\sin(2\pi/7)$ over $\mathbb{Q}$?

105. Let $\zeta = \zeta_{13} = \exp(2\pi i/13)$. Determine the degree over $\mathbb{Q}$ of the following algebraic numbers:

(i) $\zeta + \zeta^{12}$; (ii) $\zeta + \zeta^5$; (iii) $\zeta + \zeta^3 + \zeta^9$; (iv) $\zeta + \zeta^3 + \zeta^4 + \zeta^9 + \zeta^{10} + \zeta^{12}$.

106. Let $n > 0$ be an odd integer.

(i) Prove that $\mathbb{Q}(\zeta_n) = \mathbb{Q}(\zeta_{2n})$.

(ii) Express $\cos(\pi/7)$ as a polynomial in $\cos(2\pi/7)$.

107. Let L/K be a Galois extension, let H be a subgroup of $G = \mathrm{Gal}(L/K)$, and let $M = L^H$.

(i) Prove that if M/K is Galois, then every element of $\mathrm{Gal}(M/K)$ extends to an automorphism of L/K.

(ii) Prove the same is true even if M/K is not Galois. (Hint: M/K' is Galois, where K' is the fixed field of $\mathrm{Gal}(M/K)$; replace L/K with L/K'.)

(iii) Prove that $\mathrm{Gal}(M/K) \cong N(H)/H$, where

$$N(H) = \{g \in G \,:\, gHg^{-1} = H\}$$

is the *normalizer* of H.

108. Let $p > 2$ be an odd prime.

(i) Prove that $L = \mathbb{Q}(\zeta_p) \subset \mathbb{C}$ has degree two over $K = L \cap \mathbb{R}$.

(ii) Find the minimal polynomial for ζ_5 over $\mathbb{Q}(\sqrt{5})$.

109. Let $L \supset M \supset K$ be a tower of finite extensions. Prove or give a counterexample to each statement below:

(i) If L/K is Galois, then L/M is Galois.

(ii) If L/K is Galois, then M/K is Galois.

(iii) If L/M and M/K are Galois, then so is L/K.

110. Let $S(f) = \deg(L/K)$ denote the degree of the splitting field L of a monic polynomial f in $K[x]$.

(i) Show that if $f(x) = g(x)h(x)$, then $S(f)$ divides $S(g)S(h)$.

(ii) Give an example where $S(f) = S(g)S(h) > 3$.

(iii) Give an example where g and h have different splitting fields, and yet

$$\max(S(g), S(h)) < S(f) < S(g)S(h).$$

111. Let $f(x) = x^3 + ax^2 + bx + c$ with $a, b, c \in \mathbb{C}$. Let u_1, u_2 be the two roots of $f'(x) = 0$.

(i) Show that $D(f) = f(u_1)f(u_2)$ can be expressed as a polynomial in a, b and c.

(ii) Show that $D(f) = 0$ iff $f(x)$ has a multiple root in $\mathbb{C}$.

(iii) Suppose that $a, b, c \in \mathbb{R}$ and $D(f) \neq 0$. Show that $D(f) > 0$ iff $f(x)$ has exactly one root in $\mathbb{R}$.