Algebraic Number Theory

Keith Conrad

Last updated: September 27, 2010

INTRODUCTION

These are some notes on algebraic number theory for students in a course at MCCME in Fall, 2010. They are based on course notes originally prepared by Ben Salisbury with the help of Lucas David-Roesler. Christine McMeekin and Andrew Phillips spotted a lot of points of possible confusion in the text.

The reader is assumed to be familiar with modules over a PID, prime and maximal ideals, Galois theory, finite fields, and the trace and norm in a finite extension of fields. The first and last topics are reviewed in appendices.

A note about the exercises: their arrangement is based not on increasing order of difficulty, but rather along the lines of development of the material in the chapter.

Keith Conrad
September, 2010

NOTATION AND TERMINOLOGY

We write $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{C}$ for the integers, rational numbers, real numbers, and complex numbers. A finite field of size q is denoted $\mathbf{F}_q$.

Rings will generally be denoted as A or R. The units groups are $A^\times$ or $R^\times$. The $n \times n$ matrix ring over A is $\mathrm{M}_n(A)$. Rings should be assumed to be commutative unless indicated otherwise; they all contain a multiplicative identity and a ring homomorphism satisfies $f(1) = 1$.

We write $p^e || n$ to mean p^e is the highest power of a prime p that divides n.

We write $d = \square$ or $a \equiv \square \bmod p$ to mean d or a is a square.

Ideals will usually be written as fraktur letters: $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$, but sometimes I or J. Primes ideals are $\mathfrak{p}$, $\mathfrak{q}$, $\mathfrak{P}$, and $\mathfrak{Q}$.

For an R-module M, its dual module $\mathrm{Hom}_R(M, R)$ is denoted $M^\vee$.

The abbreviations PID and UFD stand for "principal ideal domain" and "unique factorization domain." The undecorated label "domain" means "integral domain."

The algebraic closure of a field F is $\overline{F}$.

CONTENTS

CHAPTER 1

NUMBER FIELDS

We introduce the basic objects of algebraic number theory: number fields and the algebraic integers inside them. The special case of quadratic fields and their integers and units are explored. As an application of units in quadratic fields, we will describe how to find the integral solutions (x, y) to $x^2 - dy^2 = n$.

1.1 Algebraic Integers

Definition 1.1. A *number field* is a finite extension of $\mathbf{Q}$.

Example 1.2. The fields $\mathbf{Q}(\sqrt{2})$, $\mathbf{Q}(i)$, $\mathbf{Q}(\sqrt[3]{2})$ and $\mathbf{Q}(\sqrt{2}, \sqrt{3})$ are all number fields.

Nonexample 1.3. The field $\mathbf{R}$ is not a number field, as it's uncountable and thus not a finite extension of $\mathbf{Q}$.

Any number field is algebraic over $\mathbf{Q}$, since finite extensions are algebraic extensions, but algebraic extensions need not be finite extensions.

Nonexample 1.4. The field $\mathbf{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}, \ldots, \sqrt{p}, \ldots)$ is algebraic over $\mathbf{Q}$, but it is not finite over $\mathbf{Q}$, so it is not a number field.

Definition 1.5. An element α of a field extension of $\mathbf{Q}$ is called an *algebraic number* if it is a root of a nonzero polynomial in $\mathbf{Q}[T]$.

Example 1.6. The numbers $\frac{1}{2}$, $\sqrt{2}$, $\frac{1}{\sqrt{2}}$, i, and $\sqrt[3]{2}$ are all algebraic numbers.

Nonexample 1.7. The real numbers e and π are not algebraic numbers. This was proved by Hermite (1873) for e and Lindemann (1882) for π.

To say that a number is a root of a nonzero polynomial in $\mathbf{Q}[T]$ is equivalent to saying it's the root of a monic polynomial in $\mathbf{Q}[T]$ because a nonzero polynomial in $\mathbf{Q}[T]$ can be scaled to a monic polynomial in $\mathbf{Q}[T]$ by dividing by its leading coefficient (this does not change its roots). However, a nonzero polynomial in $\mathbf{Z}[T]$ usually can't be scaled to be monic in $\mathbf{Z}[T]$: dividing by the leading coefficient usually doesn't keep the coefficients as integers. It is a genuinely stronger condition for a number to be the root of a monic polynomial in $\mathbf{Z}[T]$ than to be the root of a nonzero polynomial in $\mathbf{Z}[T]$. This leads to the following central concept.

Definition 1.8. An element α of a field extension of $\mathbf{Q}$ is called an *algebraic integer* if it is a root of a monic polynomial in $\mathbf{Z}[T]$.

Example 1.9. Since $\sqrt{2}$ is a root of T^2-2 and $\sqrt{2}+\sqrt{3}$ is a root of T^4-10T^2+1, they are both algebraic integers. Similarly, 5 is an algebraic integer since it is a root of $T-5$.

Nonexample 1.10. The number $\frac{1}{2}$ is a root of $2T-1$, which is not monic. This suggests that $\frac{1}{2}$ is not an algebraic integer, although it does not constitute a proof (maybe there is another polynomial with $\frac{1}{2}$ as a root that is monic in $\mathbf{Z}[T]$). There is a proof as a special case of the next theorem, which you may recognize as a disguised form of the rational roots theorem.

Theorem 1.11. *If $r \in \mathbf{Q}$ is an algebraic integer, then $r \in \mathbf{Z}$.*

Proof. Say $r = \frac{a}{b}$ with $(a,b)=1$ in $\mathbf{Z}$ and $r^n + c_{n-1}r^{n-1} + \cdots + c_1 r + c_0 = 0$ with $c_i \in \mathbf{Z}$. So

$$\frac{a^n}{b^n} + c_{n-1}\frac{a^{n-1}}{b^{n-1}} + c_{n-2}\frac{a^{n-2}}{b^{n-2}} + \cdots + c_1\frac{a}{b} + c_0 = 0,$$

and thus

$$a^n + c_{n-1}a^{n-1}b + c_{n-2}a^{n-2}b^2 + \cdots + c_1 ab^{n-1} + c_0 b^n = 0.$$

Factoring out b yields

$$a^n + b(\text{integer}) = 0.$$

This implies $b \mid a^n$, so $b = \pm 1$ since $(a, b) = 1$. Thus $r = \frac{a}{b} = \pm a \in \mathbf{Z}$. ■

For a number field K, we write the algebraic integers of K as $\mathcal{O}_K$.[1] For example, $\mathcal{O}_{\mathbf{Q}} = \mathbf{Z}$ by Theorem 1.11 and $\mathcal{O}_{\mathbf{Q}(i)} = \mathbf{Z}[i]$. Let's check part of this last equation. Any $\alpha = a + bi \in \mathbf{Z}[i]$ is a root of

$$(T - \alpha)(T - \overline{\alpha}) = T^2 - 2aT + a^2 + b^2 \in \mathbf{Z}[T],$$

so $\mathbf{Z}[i] \subset \mathcal{O}_{\mathbf{Q}(i)}$. We will see the reverse inclusion in Section 1.2.

Example 1.12. The algebraic integers of $\mathbf{Q}(\sqrt{5})$ contain $\mathbf{Z}[\sqrt{5}]$ since any $\alpha = a + b\sqrt{5} \in \mathbf{Z}[\sqrt{5}]$ is a root of the monic polynomial

$$(T - \alpha)(T - \overline{\alpha}) = T^2 - 2aT + a^2 - 5b^2 \in \mathbf{Z}[T].$$

However, there are more algebraic integers in $\mathbf{Q}(\sqrt{5})$ than $\mathbf{Z}[\sqrt{5}]$: $\frac{1+\sqrt{5}}{2}$ is not an element of $\mathbf{Z}[\sqrt{5}]$, but it is a root of $T^2 - T - 1$, so it is an algebraic integer in $\mathbf{Q}(\sqrt{5})$. This shows that if $K = \mathbf{Q}(\gamma)$ with some "natural" algebraic integer γ, then $\mathcal{O}_{\mathbf{Q}(\gamma)}$ might be larger than $\mathbf{Z}[\gamma]$.

Every rational number is a ratio of integers, and the next theorem says every algebraic number is a ratio of algebraic integers; we can even choose the denominator to be an ordinary integer.

Theorem 1.13. *If K is a number field and $\alpha \in K$, then $\alpha = \beta/d$ where $\beta \in \mathcal{O}_K$ and $d \in \mathbf{Z}$.*

Proof. We know α is the root of some monic in $\mathbf{Q}[T]$, so

$$\alpha^n + c_{n-1}\alpha^{n-1} + \cdots + c_1\alpha + c_0 = 0,$$

with $c_i \in \mathbf{Q}$. Write $c_i = b_i/d$ with $b_i, d \in \mathbf{Z}$ and $d \neq 0$. Then

$$\alpha^n + \frac{b_{n-1}}{d}\alpha^{n-1} + \frac{b_{n-2}}{d}\alpha^{n-2} + \cdots + \frac{b_1}{d}\alpha + \frac{b_0}{d} = 0.$$

Multiply through by d^n (not just d):

$$d^n\alpha^n + b_{n-1}d^{n-1}\alpha^{n-1} + b_{n-2}d^{n-1}\alpha^{n-2} + \cdots + b_1d^{n-1}\alpha + b_0d^{n-1} = 0.$$

[1] The letter $\mathcal{O}$ is taken from the German word Ordnung, which means "order" and is essentially Dedekind's label for a ring before rings were defined in general.

Group (some of) the d's and α's:

$$(d\alpha)^n + b_{n-1}(d\alpha)^{n-1} + b_{n-2}d(d\alpha)^{n-2} + \cdots + b_1 d^{n-2}(d\alpha) + b_0 d^{n-1} = 0.$$

Set $\beta = d\alpha$, so $\beta \in \mathcal{O}_K$ and $\alpha = \beta/d$. ■

Since every number field has the form $\mathbf{Q}(\alpha)$ for some algebraic number α, Theorem 1.13 tells us we can always arrange for α to be an algebraic integer.

When $[K : \mathbf{Q}] = 2$, we call K a *quadratic field*. Each quadratic field has the form $\mathbf{Q}(\sqrt{d})$ for exactly one squarefree integer d. If we don't insist d is squarefree in $\mathbf{Z}$ then the uniqueness of d is no longer correct.

Example 1.14. $\mathbf{Q}(\sqrt{24}) = \mathbf{Q}(\sqrt{6})$ and $\mathbf{Q}(\sqrt{5/3}) = \mathbf{Q}(\sqrt{(5/3)\cdot 9}) = \mathbf{Q}(\sqrt{15})$.

When $d > 0$, so $\mathbf{Q}(\sqrt{d})$ can be embedded in $\mathbf{R}$, we call $\mathbf{Q}(\sqrt{d})$ a *real quadratic field*, and for $d < 0$ we call $\mathbf{Q}(\sqrt{d})$ an *imaginary quadratic field*. There are many number-theoretic differences between real and imaginary quadratic fields.

When d is squarefree, any $\alpha = a + b\sqrt{d}$ in $\mathbf{Z}[\sqrt{d}]$ is an algebraic integer because it is a root of the monic polynomial

$$(T - \alpha)(T - \overline{\alpha}) = T^2 - 2aT + a^2 - db^2 \in \mathbf{Z}[T].$$

It is *not* true in general that $\mathcal{O}_{\mathbf{Q}(\sqrt{d})} = \mathbf{Z}[\sqrt{d}]$, as we saw with $d = 5$ in Example 1.12, but sometimes it is true (like $d = -1$). A formula for $\mathcal{O}_{\mathbf{Q}(\sqrt{d})}$ will be given in Section 1.2.

Our next goal is to prove for any number field K that $\mathcal{O}_K$ is a ring. Showing sums and products of algebraic integers are algebraic integers is not easy directly from the definition. It is proved using an important reformulation of the concept of algebraic integer in terms of modules (so it is a *linearization* of the concept, meaning it is more amenable to techniques from linear algebra). If

$$\alpha^n + c_{n-1}\alpha^{n-1} + \cdots + c_1\alpha + c_0 = 0, \qquad c_i \in \mathbf{Z}, \tag{1.1}$$

then

$$\mathbf{Z}[\alpha] = \mathbf{Z} + \mathbf{Z}\alpha + \cdots + \mathbf{Z}\alpha^{n-1}. \tag{1.2}$$

We don't have to go past the power α^{n-1} because (1.1) lets us express α^n as a $\mathbf{Z}$-linear combination of lower powers, and by induction α^m for all $m \geqslant n$ can be written in terms of lower powers and thus ultimately in terms of $1, \alpha, \ldots, \alpha^{n-1}$.

So the *ring* $\mathbf{Z}[\alpha]$ is a *finitely generated* $\mathbf{Z}$*-module.* Compare this to the ring

$$\mathbf{Z}\left[\frac{1}{2}\right] = \sum_{i\geqslant 0} \mathbf{Z}\frac{1}{2^i} = \left\{\frac{a}{2^k} : a \in \mathbf{Z}, k \geqslant 0\right\},$$

which is not finitely generated as a $\mathbf{Z}$-module: the $\mathbf{Z}$-span of a finite number of rationals of the form $a/2^k$ will not allow arbitrarily high powers of 2 in the denominator, so there isn't a finite $\mathbf{Z}$-linear spanning set.

Theorem 1.15. *Let K be a number field and $\alpha \in K$. The following are equivalent:*

(*a*) $\alpha \in \mathcal{O}_K$*;*

(*b*) $\mathbf{Z}[\alpha]$ *is a finitely generated* $\mathbf{Z}$*-module;*

(*c*) *there is a ring in K containing α that is finitely generated as a* $\mathbf{Z}$*-module.*

Proof. $(a) \Rightarrow (b)$: See (1.2).

$(b) \Rightarrow (c)$: Use $\mathbf{Z}[\alpha]$ as the ring.

$(c) \Rightarrow (a)$: We have some ring $R \subset K$ such that $\alpha \in R$ and

$$R = \mathbf{Z}x_1 + \cdots + \mathbf{Z}x_n.$$

(We are not assuming the x_i's are a basis. They just span.) The x_i's are not all 0 since $1 \in R$ and $1 \neq 0$. Multiplication by α sends R back to R, so we can write

$$\alpha x_i = a_{i1}x_1 + \cdots + a_{in}x_n, \qquad \text{for some } a_{ij} \in \mathbf{Z}.$$

We collect the equations over all i into a vector-matrix equation

$$\begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Thus

$$\alpha \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

which implies the matrix $\alpha I_n - (a_{ij})$ kills a nonzero vector in R^n:

$$(\alpha I_n - (a_{ij}))\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{1.3}$$

Viewing the equation (1.3) in K^n shows the matrix $\alpha I_n - (a_{ij})$ is not invertible in $\mathrm{M}_n(K)$, so its determinant is 0:

$$\det(\alpha I_n - (a_{ij})) = 0.$$

Expanding the determinant on the left side gives us a monic polynomial expression in α (of degree n) with $\mathbf{Z}$-coefficients, so the equation shows α is an algebraic integer. ■

Example 1.16. We put the ideas in the proof of Theorem 1.15 to work in an example. Let $K = \mathbf{Q}(\sqrt{2})$, $R = \mathbf{Z} + \mathbf{Z}\sqrt{2}$, and $\alpha = 7 + 3\sqrt{2}$. Using $x_1 = 1$ and $x_2 = \sqrt{2}$, we multiply this spanning set by α to get

$$\begin{aligned} \alpha \cdot 1 &= 7 + 3\sqrt{2}, \\ \alpha \cdot \sqrt{2} &= 6 + 7\sqrt{2}, \end{aligned}$$

so

$$\alpha\begin{pmatrix} 1 \\ \sqrt{2} \end{pmatrix} = \begin{pmatrix} 7 & 3 \\ 6 & 7 \end{pmatrix}\begin{pmatrix} 1 \\ \sqrt{2} \end{pmatrix}.$$

Therefore

$$\det\begin{pmatrix} \alpha - 7 & -3 \\ -6 & \alpha - 7 \end{pmatrix} = 0,$$

so $\alpha^2 - 14\alpha + 31 = 0$. This shows α is an algebraic integer.

(Note: The matrix $\left(\begin{smallmatrix} 7 & 3 \\ 6 & 7 \end{smallmatrix}\right)$ which we found here is the *transpose* of the matrix representation for multiplication by $7+3\sqrt{2}$ on $\mathbf{Q}(\sqrt{2})$ with respect to the basis $\{1, \sqrt{2}\}$. See Example 8.2.)

In PARI, the command which produces the irreducible integral polynomial of least degree up to n that is most likely to have a specified complex number z as a root is `algdep(z,n)`. For instance, entering `algdep(7+3*sqrt(2),2)` returns the answer `x2̂ - 14*x + 31`, which is what we found above.

Corollary 1.17. *For any number field K, $\mathcal{O}_K$ is a ring and K is the fraction field of $\mathcal{O}_K$.*

Proof. Let $\alpha, \beta \in \mathcal{O}_K$. Then

$$\mathbf{Z}[\alpha] = \mathbf{Z} + \mathbf{Z}\alpha + \cdots + \mathbf{Z}\alpha^{m-1}$$

and

$$\mathbf{Z}[\beta] = \mathbf{Z} + \mathbf{Z}\beta + \cdots + \mathbf{Z}\beta^{n-1}$$

for some positive integers m and n. So the *ring* $\mathbf{Z}[\alpha, \beta]$ equals

$$\sum_{\substack{0 \leqslant i \leqslant m-1 \\ 0 \leqslant j \leqslant n-1}} \mathbf{Z}\alpha^i\beta^j,$$

which is a finitely generated $\mathbf{Z}$-module in K containing $\alpha \pm \beta$ and $\alpha\beta$. By Theorem 1.15, $\alpha \pm \beta$ and $\alpha\beta$ are in $\mathcal{O}_K$.

Since $\mathcal{O}_K$ is a ring, Theorem 1.13 says K is the fraction field of $\mathcal{O}_K$. ■

We can detect algebraic integers using their minimal polynomials over $\mathbf{Q}$.

Theorem 1.18. *For a number field K and $\alpha \in K$, α lies in $\mathcal{O}_K$ if and only if its minimal polynomial over $\mathbf{Q}$ is in $\mathbf{Z}[T]$.*

Example 1.19. $\frac{1}{\sqrt{2}}$ is *not* an algebraic integer since its minimal polynomial over $\mathbf{Q}$ is $T^2 - \frac{1}{2} \not\in \mathbf{Z}[T]$. (Another reason it is not an algebraic integer is that algebraic integers form a ring and $(\frac{1}{\sqrt{2}})^2 = \frac{1}{2}$ is not an algebraic integer.)

Proof. We will check the "only if" direction (the other direction is easy). Say α is the root of a monic $f(T) \in \mathbf{Z}[T]$. Enlarging K by a suitable finite extension, we can split $f(T)$ over that field:

$$f(T) = (T - \alpha_1)(T - \alpha_2) \cdots (T - \alpha_n).$$

Each α_i is an algebraic integer (they are all roots of $f(T)$). The minimal polynomial $m(T) \in \mathbf{Q}[T]$ of α is a monic factor of $f(T)$ in $\mathbf{Q}[T]$, so its split factorization is, say,

$$m(T) = (T - \alpha_{i_1}) \cdots (T - \alpha_{i_d}).$$

The coefficients of $m(T)$ are sums of products of the α_{i_j}'s, so the coefficients of $m(T)$ are algebraic integers because algebraic integers in any number field form a ring. Also $m(T) \in \mathbf{Q}[T]$, so the coefficients of $m(T)$ are algebraic integers in $\mathbf{Q}$ and therefore the coefficients are in $\mathbf{Z}$ (Theorem 1.11). ■

A number field K and its ring of integers $\mathcal{O}_K$ determine each other: $\mathcal{O}_K$ is a canonically defined subring of K and K is the fraction field of $\mathcal{O}_K$. In practice, concepts about $\mathcal{O}_K$ are often referred to as concepts about K. For example, a "unit of K" means a unit of $\mathcal{O}_K$ and a "prime ideal of K" means a prime ideal of $\mathcal{O}_K$.[2]

1.2 The Integers in a Quadratic Field

To compute the integers of a quadratic field K, first write $K = \mathbf{Q}(\sqrt{d})$ for a unique squarefree $d \in \mathbf{Z}$, $d \neq 1$. (Note $\mathbf{Q}(\sqrt{12}) = \mathbf{Q}(\sqrt{3})$, but $\mathbf{Z}[\sqrt{12}] = \mathbf{Z}[2\sqrt{3}] \subsetneqq \mathbf{Z}[\sqrt{3}]$: changing d by a square factor doesn't change the quadratic field but does change the ring generated by its square root. Don't forget this!) It should be assumed that whenever we write a quadratic field as $\mathbf{Q}(\sqrt{d})$ that d is a squarefree integer unless we say otherwise.

Theorem 1.20. *For squarefree $d \in \mathbf{Z}$ with $d \neq 1$,*

$$\mathcal{O}_{\mathbf{Q}(\sqrt{d})} = \begin{cases} \mathbf{Z}[\sqrt{d}], & \text{if } d \not\equiv 1 \bmod 4, \\ \mathbf{Z}[\frac{1+\sqrt{d}}{2}], & \text{if } d \equiv 1 \bmod 4, \end{cases} = \begin{cases} \mathbf{Z} + \mathbf{Z}\sqrt{d}, & \text{if } d \not\equiv 1 \bmod 4, \\ \mathbf{Z} + \mathbf{Z}\frac{1+\sqrt{d}}{2}, & \text{if } d \equiv 1 \bmod 4. \end{cases}$$

Proof. For $d \equiv 1 \bmod 4$, $\alpha := \frac{1+\sqrt{d}}{2}$ is an algebraic integer since

$$(T-\alpha)(T-\overline{\alpha}) = T^2 - (\alpha+\overline{\alpha})T + \alpha\overline{\alpha} = T^2 - T - \frac{d-1}{4} \in \mathbf{Z}[T].$$

So the explicit rings in the theorem are contained in the algebraic integers of $\mathbf{Q}(\sqrt{d})$. To show every algebraic integer of $\mathbf{Q}(\sqrt{d})$ lies in the indicated explicit rings, suppose $\alpha = x + y\sqrt{d} \in \mathcal{O}_{\mathbf{Q}(\sqrt{d})}$ with $x, y \in \mathbf{Q}$. If $y = 0$ then $\alpha = x \in \mathbf{Q}$ implies $\alpha \in \mathbf{Z}$ because rational algebraic integers are ordinary integers. So we're done in that case. Now take $y \neq 0$, so $\alpha \notin \mathbf{Q}$. By Theorem 1.18, α is an algebraic integer if and only if its minimal polynomial over $\mathbf{Q}$ has integral coefficients. That minimal polynomial is

$$(T-\alpha)(T-\overline{\alpha}) = T^2 - (\alpha - \overline{\alpha})T + \alpha\overline{\alpha} = T^2 - 2xT + x^2 - dy^2.$$

[2] Fields have a trivial concept of unit and prime ideal, so there is no content in talking literally about units or prime ideals of K itself.

So if $x+y\sqrt{d}$ is an algebraic integer, $2x \in \mathbf{Z}$ and $x^2-dy^2 \in \mathbf{Z}$. Set $x' = 2x \in \mathbf{Z}$ so $x = x'/2$. This is either an integer (if x' is even) or half an odd integer (if x' is odd).

Case 1: If x' is even then $x \in \mathbf{Z}$, so $dy^2 \in \mathbf{Z}$. Since $y \in \mathbf{Q}$ and d is *squarefree*, having $dy^2 \in \mathbf{Z}$ implies $y \in \mathbf{Z}$ (check!). Then $\alpha = x+y\sqrt{d} \in \mathbf{Z}+\mathbf{Z}\sqrt{d} = \mathbf{Z}[\sqrt{d}]$, which is a subset of both explicit rings in the theorem.

Case 2: If x' is odd, let

$$n = x^2 - dy^2 = \frac{x'^2}{4} - dy^2 \in \mathbf{Z}.$$

Clearing the denominator, $4n = x'^2 - d(2y)^2$. Then $d(2y)^2 \in \mathbf{Z}$ and, since d is squarefree, $2y \in \mathbf{Z}$. Set $y' = 2y \in \mathbf{Z}$, so $y = y'/2$.

Now we have $4n = x'^2 - dy'^2$, so

$$dy'^2 \equiv x'^2 \bmod 4 \equiv 1 \bmod 4$$

because x' is odd (all odds square to 1 mod 4). This tells us d and y' are both odd. Then $y'^2 \equiv 1 \bmod 4$, so the displayed congruence becomes $d \equiv 1 \bmod 4$. Thus if $d \not\equiv 1 \bmod 4$ then Case 2 does not happen, so we're done if $d \not\equiv 1 \bmod 4$.

When $d \equiv 1 \bmod 4$, continuing with Case 2,

$$\alpha = x + y\sqrt{d} = \frac{x'}{2} + \frac{y'}{2}\sqrt{d} = \frac{x'-y'}{2} + y'\frac{1+\sqrt{d}}{2},$$

which lies in $\mathbf{Z}+\mathbf{Z}\frac{1+\sqrt{d}}{2} = \mathbf{Z}[\frac{1+\sqrt{d}}{2}]$ since x' and y' are odd. ■

Example 1.21. Here is the ring of integers of $\mathbf{Q}(\sqrt{d})$ for small values of $|d|$:

- For $K = \mathbf{Q}(i)$, $\mathcal{O}_K = \mathbf{Z}[i]$.
- For $K = \mathbf{Q}(\sqrt{2})$, $\mathcal{O}_K = \mathbf{Z}[\sqrt{2}]$.
- For $K = \mathbf{Q}(\sqrt{-2})$, $\mathcal{O}_K = \mathbf{Z}[\sqrt{-2}]$.
- For $K = \mathbf{Q}(\sqrt{3})$, $\mathcal{O}_K = \mathbf{Z}[\sqrt{3}]$.
- For $K = \mathbf{Q}(\sqrt{-3})$, $\mathcal{O}_K = \mathbf{Z}[\frac{1+\sqrt{-3}}{2}]$ ($d = -3 \equiv 1 \bmod 4$).

Remark 1.22. When $d \equiv 1 \bmod 4$, check in Exercise 1.5 that

$$\mathbf{Z}\left[\frac{1+\sqrt{d}}{2}\right] = \left\{\frac{a+b\sqrt{d}}{2} : a, b \in \mathbf{Z}, a \equiv b \bmod 2\right\}.$$

This is a useful alternate description of the ring of integers of $\mathbf{Q}(\sqrt{d})$.

Definition 1.23. If K is a quadratic field, any subring of $\mathcal{O}_K$ other than $\mathbf{Z}$ is called a *quadratic ring*.

Examples of quadratic rings include therings $\mathbf{Z}[\sqrt{d}]$ for any nonsquare integer d, and they also include $\mathbf{Z}[\frac{1+\sqrt{d}}{2}]$ for any nonsquare integer $d \equiv 1 \bmod 4$. A conceptual definition of quadratic rings is in Exercise 1.8c. Rings of the form $\mathbf{Z}[\sqrt{d}]$, where $d \in \mathbf{Z}$ and $d \neq \square$, are called *pure quadratic rings*. The case that d is squarefree is the most important, but d with a square factor is allowed too. Examples are $\mathbf{Z}[2i] = \mathbf{Z}[\sqrt{-4}]$ and $\mathbf{Z}[\sqrt{18}] = \mathbf{Z}[3\sqrt{2}]$.

1.3 Units in a Quadratic Field

There are only four units in $\mathbf{Z}[i]$: ± 1 and $\pm i$. But $\mathbf{Z}[\sqrt{2}]$ has infinitely many units, such as powers of $1+\sqrt{2}$. We want to find all the units in $\mathbf{Z}[\sqrt{2}]$.

For nonsquare integers d, the field $\mathbf{Q}(\sqrt{d})$ has a norm function mapping it multiplicatively to $\mathbf{Q}$: for $\alpha = a + b\sqrt{d}$ in $\mathbf{Q}(\sqrt{d})$, set its conjugate to be[3] $\overline{\alpha} = a - b\sqrt{d}$ and set its *norm* to be

$$\mathrm{N}(\alpha) = \alpha\overline{\alpha} = (a + b\sqrt{d})(a - b\sqrt{d}) = a^2 - db^2.$$

The norm is multiplicative: $\mathrm{N}(\alpha\beta) = \mathrm{N}(\alpha)\,\mathrm{N}(\beta)$. For $d > 0$, the norm has positive and negative values, and for $d < 0$, the norm has only nonnegative values. If $\alpha \in \mathbf{Z}[\sqrt{d}]$ then its norm is in $\mathbf{Z}$.

Theorem 1.24. *The units in* $\mathbf{Z}[\sqrt{d}]$ *are*

$$\mathbf{Z}[\sqrt{d}]^\times = \left\{u \in \mathbf{Z}[\sqrt{d}] : \mathrm{N}(u) = \pm 1\right\} = \left\{a + b\sqrt{d} : a^2 - db^2 = \pm 1\right\}.$$

Proof. If $uv = 1$ in $\mathbf{Z}[\sqrt{d}]$, then $\mathrm{N}(u)\,\mathrm{N}(v) = \mathrm{N}(1) = 1$ in $\mathbf{Z}$, so $\mathrm{N}(u) = \pm 1$. If $\mathrm{N}(u) = \pm 1$, then $u\overline{u} = \pm 1$, so $u(\pm\overline{u}) = 1$ in $\mathbf{Z}[\sqrt{d}]$. ■

Example 1.25. In $\mathbf{Z}[\sqrt{2}]$, $1+\sqrt{2}$ is a unit: $\mathrm{N}(1+\sqrt{2}) = 1^2 - 2\cdot 1^2 = -1$ and $(1+\sqrt{2})(-1+\sqrt{2}) = 1$. Unlike in $\mathbf{Z}[i]$, where there are only four units, there are infinitely many units in $\mathbf{Z}[\sqrt{2}]$ since $1+\sqrt{2}$ has infinitely many different

[3] This is the complex conjugate if $d < 0$, but for $d > 0$ both α and $\overline{\alpha}$ are real numbers and this has nothing to do with complex conjugation. In all cases it is a conjugate in the sense of Galois theory for $\mathbf{Q}(\sqrt{d})/\mathbf{Q}$.

powers. The first few powers of $1+\sqrt{2}$ are

$$1+\sqrt{2},\quad 3+2\sqrt{2},\quad 7+5\sqrt{2},\quad 17+12\sqrt{2},\quad 41+29\sqrt{2},\quad 99+70\sqrt{2}. \tag{1.4}$$

We will show in Section 1.3 that every unit in $\mathbf{Z}[\sqrt{2}]$ is, up to sign, an integral power of $1+\sqrt{2}$, so (1.4) is the first six units greater than 1 in $\mathbf{Z}[\sqrt{2}]$.

Example 1.26. In $\mathbf{Z}[\sqrt{3}]$, units come from solving $a^2-3b^2=\pm 1$. The unit $2+\sqrt{3}$ has norm 1 and its first few powers are

$$2+\sqrt{3},\quad 7+4\sqrt{3},\quad 26+15\sqrt{3},\quad 97+56\sqrt{3},\quad 362+209\sqrt{3},\quad 1351+780\sqrt{3}.$$

Unlike in $\mathbf{Z}[\sqrt{2}]$ there are no units with norm -1 in $\mathbf{Z}[\sqrt{3}]$:

$$a^2-3b^2=-1 \Longrightarrow a^2\equiv -1 \bmod 3 \Longrightarrow -1\equiv \Box \bmod 3,$$

but -1 is not a square mod 3 (the only squares mod 3 are 0 and 1).

Example 1.27. If $d\leqslant -2$ then $\mathbf{Z}[\sqrt{d}]^\times=\{\pm 1\}$ since, writing $-d=D$ with $D\geqslant 2$, the only integral solutions to $x^2+Dy^2=\pm 1$ are $(x,y)=(\pm 1,0)$.

The equation

$$x^2-dy^2=1 \tag{1.5}$$

with $d>0$ and $d\neq\Box$ is called *Pell's equation.* Its integral solutions were studied long before its connection to units in the ring $\mathbf{Z}[\sqrt{d}]$ was known. One reason for interest in the equation is that if $x^2-dy^2=1$ then $(x/y)^2=d+1/y^2\approx d$ if y is large, so integral solutions to Pell's equation tend to give particularly good rational approximations to $\sqrt{d}$. (For instance, from (1.4) the ratio $99/70\approx 1.414285$ is surprisingly close to $\sqrt{2}\approx 1.414213$ considering the size of the denominator.) Lagrange was the first mathematician to prove the equation $x^2-dy^2=1$ has an integral solution (x,y) other than the trivial solutions $(\pm 1,0)$. His proof used a constructive method with continued fractions.

For any $u\in\mathbf{Z}[\sqrt{2}]^\times$, we have the units u, $1/u$, $-u$, and $-1/u$, and one of these is greater than 1. So up to a change in sign and inversion, any unit in $\mathbf{Z}[\sqrt{2}]$ can be turned into a unit greater than 1.

Lemma 1.28. *If $a+b\sqrt{2}$ is a unit in $\mathbf{Z}[\sqrt{2}]$ which, as a real number, is greater than 1 then the integer coefficients a and b are both positive. In particular, $1+\sqrt{2}$ is the smallest unit greater than 1 in $\mathbf{Z}[\sqrt{2}]$.*

It is crucial that we are referring to *units*, as we can't get lower bounds on coefficients of general elements in $\mathbf{Z}[\sqrt{2}]$ that are greater than 1: $31 - 14\sqrt{2} \approx 11.2 > 1$, but its coefficients are not both positive integers.

Proof. By Theorem 1.24, a unit $u = a + b\sqrt{2}$ satisfies $u\overline{u} = \pm 1$. We will break this into two cases. Suppose first that $u\overline{u} = 1$. So $u^{-1} = \overline{u} = a - b\sqrt{2}$. From $u > 1$, $0 < \frac{1}{u} < 1$. Thus $1 < a + b\sqrt{2}$ and $0 < a - b\sqrt{2} < 1$. Adding these two inequalities, we have $1 < 2a$. So $a > \frac{1}{2}$. Then $a \geqslant 1$ since $a \in \mathbf{Z}$, so $b\sqrt{2} > a - 1 \geqslant 0$. Thus $b > 0$, so $b \geqslant 1$ since $b \in \mathbf{Z}$.

Now suppose $u\overline{u} = -1$. Then $u^{-1} = -\overline{u} = -a + b\sqrt{2}$. From $u > 1$, $0 < \frac{1}{u} < 1$. Thus $1 < a + b\sqrt{2}$ and $0 < -a + b\sqrt{2} < 1$. Adding these two inequalities, $1 < 2b\sqrt{2}$, so $b > 0$, which implies $b \geqslant 1$ since $b \in \mathbf{Z}$. Then $a > b\sqrt{2} - 1 \geqslant \sqrt{2} - 1 > 0$, so $a \geqslant 1$. ■

Theorem 1.29. *Every unit in $\mathbf{Z}[\sqrt{2}]$ is a power of $1 + \sqrt{2}$ up to sign:*

$$\mathbf{Z}[\sqrt{2}]^\times = \pm(1 + \sqrt{2})^{\mathbf{Z}}.$$

Proof. Let $u \in \mathbf{Z}[\sqrt{2}]^\times$. Since one of u, $\frac{1}{u}$, $-u$, and $-\frac{1}{u}$ is greater than 1, it suffices to show if $u > 1$ that u is a power of $1 + \sqrt{2}$.

Since $u > 1$ and the nonnegative powers of $1 + \sqrt{2}$ tend monotonically to ∞ starting from 1, u lies between two such powers:

$$(1 + \sqrt{2})^k \leqslant u < (1 + \sqrt{2})^{k+1}$$

for some integer $k \geqslant 0$. Dividing throughout by $(1 + \sqrt{2})^k$,

$$1 \leqslant u(1 + \sqrt{2})^{-k} < 1 + \sqrt{2}.$$

The product $u(1+\sqrt{2})^{-k}$ is a unit in $\mathbf{Z}[\sqrt{2}]$, and since we know that the smallest unit greater than 1 is $1 + \sqrt{2}$, these inequalities tell us $u(1 + \sqrt{2})^{-k} = 1$, so $u = (1 + \sqrt{2})^k$. ■

Let's now treat the general pure quadratic ring $\mathbf{Z}[\sqrt{d}]$, with $d > 0$. Its units comes from integral solutions to $x^2 - dy^2 = \pm 1$ (Theorem 1.24).

Lemma 1.30. *Let $d > 0$ not be a perfect square. For any unit $a+b\sqrt{d}$ in $\mathbf{Z}[\sqrt{d}]$ which is greater than 1, the integer coefficients a and b are both positive.*

Proof. Set $u = a + b\sqrt{d}$. We have $u > 1$ and $u\overline{u} = \pm 1$ by Theorem 1.24, so $1/u = \pm\overline{u}$. Unlike the proof for $\mathbf{Z}[\sqrt{2}]$, we will give a uniform proof that a and b are positive, not relying on the separate cases $u\overline{u} = 1$ and $u\overline{u} = -1$.

Since $-1 < \overline{u} < 1$, $\overline{u} > -1$. Adding the inequalities $u > 1$ and $\overline{u} > -1$ gives $2a > 0$, so $a > 0$. Since $u > 1$ and $-\overline{u} > -1$, adding the inequalities gives $2b\sqrt{d} > 0$, so $b > 0$. ■

Theorem 1.31. *Let $d > 0$ not be a perfect square. If the unit group $\mathbf{Z}[\sqrt{d}]^\times$ is not $\{\pm 1\}$, there is a least unit $\varepsilon > 1$ and $\mathbf{Z}[\sqrt{d}]^\times = \pm\varepsilon^{\mathbf{Z}}$.*

Proof. Since any unit $a + b\sqrt{d}$ which is greater than 1 has $a \geqslant 1$ and $b \geqslant 1$, there is a least unit $\varepsilon > 1$. The rest of the proof is like the proof of Theorem 1.29. ■

The hypothesis in Theorem 1.31 is always satisfied: $\mathbf{Z}[\sqrt{d}]^\times \neq \{\pm 1\}$. We'll prove this in Theorem 1.38 by showing Pell's equation $x^2 - dy^2 = 1$ always has an integral solution (x, y) with $y \neq 0$. (The negative Pell equation $x^2 - dy^2 = -1$ may or may not have an integral solution.)

The least unit greater than 1 in the ring $\mathbf{Z}[\sqrt{d}]$ is called the *fundamental unit* of $\mathbf{Z}[\sqrt{d}]$. For example, the fundamental unit in $\mathbf{Z}[\sqrt{2}]$ is $1+\sqrt{2}$. If $a+b\sqrt{d}$ is the fundamental unit in $\mathbf{Z}[\sqrt{d}]$ then a and b are positive integers, so the coefficients a_n and b_n in the powers

$$a_n + b_n\sqrt{d} = (a + b\sqrt{d})^n$$

for $n \geqslant 1$ are increasing: $a = a_1 < a_2 < a_3 < \cdots$ and $b = b_1 < b_2 < b_3 < \cdots$. Therefore we have a very crude algorithm to find the fundamental unit: starting with $y = 1$, compute $dy^2 \pm 1$ with increasing y until you reach a perfect square x^2. Then $x + y\sqrt{d}$ is a unit greater than 1 and it must be the least one since the units greater than 1 increase along with their coefficient of $\sqrt{d}$. (If you are only interested in units with norm 1, just look at $dy^2 + 1$.)

In practice, this method could take a very long time to work, because the size of the coefficients in the fundamental unit of $\mathbf{Z}[\sqrt{d}]$ varies irregularly with d. For example, the fundamental units in $\mathbf{Z}[\sqrt{60}]$, $\mathbf{Z}[\sqrt{61}]$, and $\mathbf{Z}[\sqrt{62}]$ are

$$31 + 4\sqrt{60}, \quad 29718 + 3805\sqrt{61}, \quad 63 + 8\sqrt{62}. \tag{1.6}$$

The first and third units in (1.6) have norm 1 and the second has norm -1.

The smallest unit in $\mathbf{Z}[\sqrt{61}]$ with norm 1 having positive coefficients is

$$(29718 + 3805\sqrt{61})^2 = 1766319049 + 226153980\sqrt{61}.$$

Fermat (1657) challenged English mathematicians to solve $x^2 - dy^2 = 1$ for general nonsquare $d \in \mathbf{Z}^+$. He wrote that if this was too hard, they should at least try $x^2 - 61y^2 = 1$, where he chose a small coefficient of y^2 "pour ne vous donner pas trop de peine" (so you don't have too much work).

Theorem 1.32. *In any quadratic field K, an algebraic integer α is a unit in $\mathcal{O}_K$ if and only if* $\mathrm{N}(\alpha) = \alpha\overline{\alpha} = \pm 1$.

This generalizes Theorem 1.24, since $\mathcal{O}_K$ need not have the form $\mathbf{Z}[\sqrt{d}]$.

Proof. First we show for $\alpha \in \mathcal{O}_K$ that $\mathrm{N}(\alpha) \in \mathbf{Z}$. The conjugate $\overline{\alpha}$ is an algebraic integer so $\alpha\overline{\alpha}$ is an algebraic integer. It is also rational, so it is in $\mathbf{Z}$.

If α is a unit in $\mathcal{O}_K$, then $\alpha\beta = 1$ for some $\beta \in \mathcal{O}_K$. Taking norms of both sides, $\mathrm{N}(\alpha)\,\mathrm{N}(\beta) = 1$ in $\mathbf{Z}$, so $\mathrm{N}(\alpha)$ is 1 or -1. Conversely, if $\mathrm{N}(\alpha) = \pm 1$ then $\alpha\overline{\alpha} = \pm 1$, so $\alpha(\pm\overline{\alpha}) = 1$, which makes α a unit in $\mathcal{O}_K$. ∎

Finding the units in the integers of $\mathbf{Q}(\sqrt{d})$ is equivalent to solving

$$\mathrm{N}(x + y\sqrt{d}) = x^2 - dy^2 = \pm 1$$

in $\mathbf{Z}$ when $d \not\equiv 1 \bmod 4$. What if $d \equiv 1 \bmod 4$? For $\alpha = x + y\frac{1+\sqrt{d}}{2} \in \mathbf{Q}(\sqrt{d})$ we have

$$\begin{aligned}
\mathrm{N}(\alpha) &= \alpha\overline{\alpha} \\
&= \left(x + y\frac{1+\sqrt{d}}{2}\right)\left(x + y\frac{1-\sqrt{d}}{2}\right) \\
&= \left(x + \frac{y}{2} + \frac{y}{2}\sqrt{d}\right)\left(x + \frac{y}{2} - \frac{y}{2}\sqrt{d}\right) \\
&= \left(x + \frac{y}{2}\right)^2 - d\left(\frac{y}{2}\right)^2 \\
&= x^2 + xy - \frac{d-1}{4}y^2,
\end{aligned}$$

so units in the integers of $\mathbf{Q}(\sqrt{d})$ correspond to $\mathbf{Z}$-solutions (x, y) of

$$x^2 + xy - \frac{d-1}{4}y^2 = \pm 1.$$

This looks messy. If instead we write $\alpha = \frac{a+b\sqrt{d}}{2}$ with $a \equiv b \bmod 2$ (Remark 1.22), then the condition $\mathrm{N}(\alpha) = \pm 1$ becomes

$$a^2 - db^2 = \pm 4, \tag{1.7}$$

which is a close relative of Pell's equation. Reducing (1.7) modulo 2 turns it into $a \equiv b \bmod 2$, since any number is congruent mod 2 to its square, so the parity constraint on a and b in the formula for α is forced by (1.7). Therefore finding units in $\mathbf{Q}(\sqrt{d})$ for $d \equiv 1 \bmod 4$ is equivalent to solving (1.7) in $\mathbf{Z}$.

Theorem 1.33. *Let $d > 0$ not be a perfect square with $d \equiv 1 \bmod 4$. If the unit group $\mathbf{Z}[\frac{1+\sqrt{d}}{2}]^\times$ is not $\{\pm 1\}$, there is a least unit $\varepsilon > 1$ and $\mathbf{Z}[\frac{1+\sqrt{d}}{2}]^\times = \pm\varepsilon^{\mathbf{Z}}$.*

Proof. This is similar to the proof for $\mathbf{Z}[\sqrt{d}]$. The key point is to show any unit $\frac{a+b\sqrt{d}}{2}$ which is greater than 1 has positive a and b. It is done as in the proof of Lemma 1.30. ■

Example 1.34. The units in $\mathbf{Z}[\frac{1+\sqrt{5}}{2}]$ are $\frac{a+b\sqrt{5}}{2}$ where $a^2 - 5b^2 = \pm 4$. One solution is $(a, b) = (1, 1)$. This gives us the unit $\frac{1+\sqrt{5}}{2}$, with norm -1, and every unit in $\mathbf{Z}[\frac{1+\sqrt{5}}{2}]$ is $\pm(\frac{1+\sqrt{5}}{2})^k$ for some $k \in \mathbf{Z}$.

Example 1.35. In $\mathbf{Z}[\frac{1+\sqrt{17}}{2}]$, the first solution to $a^2 - 17b^2 = \pm 4$ with positive a and b is $(a, b) = (8, 2)$, which gives us the unit $\frac{8+2\sqrt{17}}{2} = 4 + \sqrt{17}$. In particular, all the units of $\mathbf{Z}[\frac{1+\sqrt{17}}{2}]$ lie in $\mathbf{Z}[\sqrt{17}]$.

Since the integers of any quadratic field contain a subring of the form $\mathbf{Z}[\sqrt{d}]$, the nontrivial solvability of Pell's equation (coming up in Theorem 1.38) tells us $\mathcal{O}_K^\times$ is infinite when K is real quadratic. The least unit greater than 1 in $\mathcal{O}_K^\times$ is called the fundamental unit of the field K. The fundamental units of $\mathbf{Q}(\sqrt{5})$ and $\mathbf{Q}(\sqrt{17})$ are $\frac{1+\sqrt{5}}{2}$ and $4 + \sqrt{17}$.

In (1.6), the fundamental unit $u = 29718 + 3805\sqrt{61}$ in $\mathbf{Z}[\sqrt{61}]$ is not the fundamental unit of $\mathbf{Q}(\sqrt{61})$, whose ring of integers is $\mathbf{Z}[\frac{1+\sqrt{61}}{2}]$. The fundamental unit of this field is $\frac{39}{2} + \frac{5}{2}\sqrt{61}$ (whose cube is u.) An analogue of (1.6) where the units of three "consecutive" quadratic fields are fundamental for the full ring of integers is

$$1015 + 52\sqrt{381}, \quad 164998439999 + 8442054600\sqrt{382}, \quad 18768 + 959\sqrt{383}.$$

What are the units in an imaginary quadratic field?

Theorem 1.36. *The unit group of an imaginary quadratic field is finite. More precisely,* $\mathbf{Q}(i)$ *has 4 units,* $\mathbf{Q}(\sqrt{-3})$ *has 6 units, and any other imaginary quadratic field has only the two units* ± 1*.*

Proof. We already know the case of $\mathbf{Q}(i)$, so write the field as $\mathbf{Q}(\sqrt{-D})$, where $D \geqslant 2$. First suppose $-D \not\equiv 1 \bmod 4$. The norm of $x + y\sqrt{-D}$ is $x^2 + Dy^2$. This is never negative, so finding units in $\mathbf{Q}(\sqrt{-D})$ is the same as solving $x^2 + Dy^2 = 1$ in integers. Since $x^2 + Dy^2 \geqslant x^2 + 2y^2$, a unit has $y = 0$ and $x = \pm 1$, so $x + y\sqrt{-D} = \pm 1$.

Now suppose $-D \equiv 1 \bmod 4$, so $D \in \{3, 7, 11, 15, \dots\}$. Writing an algebraic integer in $\mathbf{Q}(\sqrt{-D})$ as $\frac{a+b\sqrt{-D}}{2}$ with $a \equiv b \bmod 2$, finding units is the same as solving $a^2 + Db^2 = 4$. When $D = 3$ there are 6 solutions: $(a, b) \in \{(\pm 2, 0), (\pm 1, \pm 1)\}$. When $D \geqslant 7$ the only two solutions are $(a, b) = (\pm 2, 0)$ and then $\frac{a+b\sqrt{-D}}{2} = \pm 1$. ■

There is nothing mysterious about the 6 units in $\mathbf{Q}(\sqrt{-3})$: they are the sixth roots of unity. The number $\frac{-1+\sqrt{-3}}{2}$ is a nontrivial cube root of unity and its negative is a nontrivial sixth root of unity.

Returning to the real quadratic case, we will prove there is a nontrivial solution to Pell's equation by a nonconstructive method based on the following lemma about approximations.

Lemma 1.37. *For positive nonsquare* $d \in \mathbf{Z}$*, there are infinitely many positive integers* x *and* y *such that* $|x - y\sqrt{d}| < \frac{1}{y}$*.*

The point here is not just that there are integral multiples of $\sqrt{d}$ close to integers, but the distance can be controlled by the multiplier on $\sqrt{d}$ (infinitely often).

Proof. We use the pigeonhole principle. For any integer $m \geqslant 2$ consider the $m + 1$ numbers

$$0,\ \sqrt{d},\ 2\sqrt{d},\ \dots,\ m\sqrt{d}.$$

The fractional parts of these numbers are in the half-open interval $[0, 1)$. Break up $[0, 1)$ into m half-open intervals $[i/m, (i + 1)/m)$ for $i = 0, 1, \dots, m - 1$. By the pigeonhole principle, two of the $m + 1$ numbers above, say $a\sqrt{d}$ and $a'\sqrt{d}$, must have fractional parts in the same interval:

$$a\sqrt{d} = A + \varepsilon, \quad a'\sqrt{d} = A' + \varepsilon',$$

where $A, A' \in \mathbf{Z}$ and ε and ε' lie in a common interval $[i/m, (i+1)/m)$, so

$$|\varepsilon - \varepsilon'| < \frac{1}{m}.$$

(We have a strict inequality here since we are using half-open intervals.) Thus

$$\left|(a\sqrt{d} - A) - (a'\sqrt{d} - A')\right| < \frac{1}{m} \Longrightarrow \left|(A' - A) - (a' - a)\sqrt{d}\right| < \frac{1}{m}.$$

Set $x = A' - A$ and $y = a' - a$, so x and y are integers, $0 < y \leqslant m$, and

$$|x - y\sqrt{d}| < \frac{1}{m} \leqslant \frac{1}{y}. \tag{1.8}$$

Since x is within 1 of $y\sqrt{d}$, x is at least $y\sqrt{d} - 1 \geqslant \sqrt{d} - 1 > 0$, so x is positive.

Having found a positive integer solution (x, y) to $|x - y\sqrt{d}| < 1/y$, to get a second such solution choose a positive integer m' such that $1/m' < |x - y\sqrt{d}|$. (There is such an m' because $x - y\sqrt{d} \neq 0$, as $\sqrt{d}$ is irrational.) Run through the argument above with m' in place of m to find x' and y' in $\mathbf{Z}^+$ satisfying $|x' - y'\sqrt{d}| < 1/m'$ with $y' \leqslant m'$, so $|x' - y'\sqrt{d}| < 1/y'$. From the inequalities

$$|x' - y'\sqrt{d}| < \frac{1}{m'} < |x - y\sqrt{d}|, \tag{1.9}$$

the pair (x, y) is obviously not the same as the pair (x', y'). By repeating this argument again to get a smaller $|x'' - y''\sqrt{d}|$, and so on, we are done. ■

Essentially the only property we needed of $\sqrt{d}$ in Lemma 1.37 is that it is a real irrational number. For any real irrational α, the same argument shows the inequality $|x - y\alpha| < 1/y$ has infinitely many integral solutions (x, y) with $y > 0$ (we have to give up on $x > 0$ if α is negative).

Theorem 1.38. *For $d \in \mathbf{Z}^+$ with $d \neq \square$, the equation $x^2 - dy^2 = 1$ has an integral solution x, y where $y \neq 0$.*

Proof. To start, we will show for x and y in $\mathbf{Z}^+$ satisfying $|x - y\sqrt{d}| < 1/y$, $|x^2 - dy^2|$ is bounded above independently of x and y. First we bound x in terms of y:

$$x = x - y\sqrt{d} + y\sqrt{d} < \frac{1}{y} + y\sqrt{d} \leqslant 1 + y\sqrt{d}.$$

Then

$$|x^2 - dy^2| = (x + y\sqrt{d})|x - y\sqrt{d}| < (1 + y\sqrt{d} + y\sqrt{d})\frac{1}{y} = \frac{1}{y} + 2\sqrt{d} \leqslant 1 + 2\sqrt{d}.$$

The bound $1 + 2\sqrt{d}$ does not involve x or y.

Since infinitely many positive integers x and y satisfy $|x - y\sqrt{d}| < 1/y$, by Lemma 1.37, the inequality $|x^2 - dy^2| < 1 + 2\sqrt{d}$ has infinitely many solutions in positive integers (x, y). Therefore the pigeonhole principle tells us that there is an $M \in \mathbf{Z}$ with $|M| < 1 + 2\sqrt{d}$ such that

$$x^2 - dy^2 = M$$

is solvable for infinitely many pairs of positive integers (x, y). (The value of M here is nonconstructive.) Since $\sqrt{d}$ is irrational, $M \neq 0$.

Reduce these pairs mod M. The infinitely many $(x \bmod M, y \bmod M)$ must have a repetition infinitely often, since there are only finitely many pairs of integers mod M. So we can choose two different integral solutions (x_1, y_1) and (x_2, y_2) to the equation $x^2 - dy^2 = M$ such that $x_1 \equiv x_2 \bmod M$ and $y_1 \equiv y_2 \bmod M$. Because there is an infinite repetition mod M, we can even assume $(x_1, y_1) \neq \pm(x_2, y_2)$, which will be handy in a moment.

Write $x_1 = x_2 + Ma$ and $y_1 = y_2 + Mb$, where a and b are in $\mathbf{Z}$. Then

$$x_1 + y_1\sqrt{d} = x_2 + y_2\sqrt{d} + M(a + b\sqrt{d}).$$

Since $M = x_2^2 - dy_2^2 = (x_2 + y_2\sqrt{d})(x_2 - y_2\sqrt{d})$,

$$x_1 + y_1\sqrt{d} = (x_2 + y_2\sqrt{d})(1 + (x_2 - y_2\sqrt{d})(a + b\sqrt{d})).$$

Write the second factor on the right as u, so $x_1 + y_1\sqrt{d} = (x_2 + y_2\sqrt{d})u$ and $u \in \mathbf{Z}[\sqrt{d}]$. Taking the norm of both sides, $M = M\,\mathrm{N}(u)$, so $\mathrm{N}(u) = 1$. Since $(x_1, y_1) \neq \pm(x_2, y_2)$, $u \neq \pm 1$, so u is a nontrivial unit in $\mathbf{Z}[\sqrt{d}]$ and its coefficients provide a solution to $x^2 - dy^2 = 1$ with $y \neq 0$. ■

1.4 Application: Solving $x^2 - dy^2 = n$, I

Using units in $\mathbf{Z}[\sqrt{d}]$, we describe a procedure for finding all the integral solutions of $x^2 - dy^2 = n$ when n is not 0 or 1. This includes the possibility of finding there are no integral solutions.

Before we jump into the details, let's answer the question: why should we care about solving $x^2 - dy^2 = n$?

1. Historically, this is where algebraic number theory came from. Quadratic equations of the form $ax^2 + bxy + cy^2 = n$ were studied in special cases by Fermat and Euler, and in general by Lagrange and Gauss. Their ideas eventually developed into the theory of quadratic fields, which remain an important testing ground in algebraic number theory. While $x^2 - dy^2 = n$ is only a special instance of these quadratic equations, it is not far from the general case (Exercise 1.21).

2. If you prefer abstract mathematical structures (groups, rings, modules,...) to the concreteness of finding integral solutions of equations, there is still an important reason to care about $x^2 - dy^2 = n$: its lack of solutions for particular n can be used to prove that $\mathbf{Z}[\sqrt{d}]$ does not have unique factorization (Exercise 1.26c) and that particular ideals in $\mathbf{Z}[\sqrt{d}]$ are not principal (Exercise 1.27c).

The Diophantine equation $x^2 - dy^2 = n$ is the same as $\mathrm{N}(x + y\sqrt{d}) = n$. If $\mathrm{N}(x + y\sqrt{d}) = n$ and $\mathrm{N}(u) = 1$ then $\mathrm{N}((x + y\sqrt{d})u) = n$, so the coefficients of $x' + y'\sqrt{d} := (x + y\sqrt{d})u$ satisfy $x'^2 - dy'^2 = n$. We will show all the integral solutions to $x^2 - dy^2 = n$ can be generated from a finite number of integral solutions in this way.

Theorem 1.39. *Fix a unit u of $\mathbf{Z}[\sqrt{d}]$ with norm 1 and $u > 1$. For $n \in \mathbf{Z} - \{0\}$, any $\alpha \in \mathbf{Z}[\sqrt{d}]$ satisfying $\mathrm{N}(\alpha) = n$ has a unit multiple $x + y\sqrt{d}$ where $|x| \leqslant \sqrt{|n|u}$ and $|y| \leqslant \sqrt{|n|u}/\sqrt{d}$.*

The bounds on $|x|$ and $|y|$ are expressed in terms of one unit of norm 1 greater than 1. So being able to solve Pell's equation effectively lets us solve any equation of the form $x^2 - dy^2 = n$ effectively.

Proof. We use absolute values and logarithms. Define $L\colon \mathbf{Z}[\sqrt{d}] - \{0\} \to \mathbf{R}^2$ by $L(\alpha) = (\log|\alpha|, \log|\overline{\alpha}|)$. This function satisfies $L(\alpha\beta) = L(\alpha) + L(\beta)$.

Since $u\overline{u} = 1$, $\log|u| + \log|\overline{u}| = 0$, so

$$L(u) = (\log|u|, \log|\overline{u}|) = (\log|u|)(1, -1) = (\log u)(1, -1).$$

This is linearly independent of $(1,1)$, so $(1,1)$ and $L(u)$ form a basis of $\mathbf{R}^2$. Therefore for any nonzero $\alpha \in \mathbf{Z}[\sqrt{d}]$ we can write $L(\alpha) = c_1(1,1) + c_2 L(u)$ for some real c_1 and c_2. Writing out the coordinates on both sides,

$$(\log|\alpha|, \log|\overline{\alpha}|) = (c_1 + c_2 \log u, c_1 - c_2 \log u)$$

for some real numbers c_1 and c_2. Adding the coordinates we can solve for c_1:

$$c_1 = \frac{1}{2}(\log|\alpha| + \log|\overline{\alpha}|) = \frac{1}{2}\log|\mathrm{N}(\alpha)| = \frac{1}{2}\log|n|.$$

Thus

$$L(\alpha) = \frac{1}{2}\log|n|(1,1) + c_2(\log u)(1,-1).$$

Let k be the integer nearest to c_2, so $c_2 = k + \delta$ with $|\delta| \leqslant \frac{1}{2}$. Then

$$L(\alpha) = \frac{1}{2}\log|n|(1,1) + k(\log u)(1,-1) + \delta(\log u)(1,-1).$$

Since $k(\log u)(1,-1) = L(u^k)$, we have

$$L(\alpha u^{-k}) = \frac{1}{2}\log|n|(1,1) + \delta(\log u)(1,-1).$$

Set $\alpha' = \alpha u^{-k}$. This is a unit multiple of α with the same norm. Writing $\alpha' = x + y\sqrt{d}$, the bound $|\delta| \leqslant \frac{1}{2}$ gives us bounds on the coordinates of $L(\alpha')$:

$$\log|x + y\sqrt{d}| = \frac{1}{2}\log|n| + \delta\log u \leqslant \frac{1}{2}\log|n| + \frac{1}{2}\log u \Longrightarrow |x + y\sqrt{d}| \leqslant \sqrt{|n|u}$$

and

$$\log|x - y\sqrt{d}| = \frac{1}{2}\log|n| - \delta\log u \leqslant \frac{1}{2}\log|n| + \frac{1}{2}\log u \Longrightarrow |x - y\sqrt{d}| \leqslant \sqrt{|n|u}.$$

Thus

$$|x| = \left|\frac{x + y\sqrt{d} + x - y\sqrt{d}}{2}\right| \leqslant \sqrt{|n|u}$$

and

$$|y| = \left|\frac{(x + y\sqrt{d}) - (x - y\sqrt{d})}{2\sqrt{d}}\right| \leqslant \frac{\sqrt{|n|u}}{\sqrt{d}}.$$

■

Example 1.40. We will completely solve $x^2 - 15y^2 = 34$ in integers. One solution is $(7, 1)$.

The fundamental unit in $\mathbf{Z}[\sqrt{15}]$ is $4 + \sqrt{15}$, which has norm 1. We search for solutions of $x^2 - 15y^2 = 34$ where

$$|x| \leqslant \sqrt{34u} \approx 16.3, \quad |y| \leqslant \frac{\sqrt{34u}}{\sqrt{15}} \approx 4.2.$$

The bound on y is smaller, so we will focus on y. Let y run through integers from 0 up to 4 and see when $15y^2 + 34$ is a perfect square. It occurs for $y = 1$ and $y = 3$ with corresponding values $x = 7$ and $x = 13$. Therefore any element of $\mathbf{Z}[\sqrt{15}]$ with norm 34 is a unit multiple of one of the numbers

$$7 + \sqrt{15}, \; 7 - \sqrt{15}, \; 13 + 3\sqrt{15}, \; 13 - 3\sqrt{15}.$$

There are unit multiples among these: $13 + 3\sqrt{15} = (7 - \sqrt{15})(4 + \sqrt{15})$ and $13 - 3\sqrt{15} = (7 + \sqrt{15})(4 - \sqrt{15})$, so every integral solution of $x^2 - 15y^2 = 34$ can be found by multiplying $7 + \sqrt{15}$ or $7 - \sqrt{15}$ by a unit with norm 1. For example, the solution $(97, 25)$ arises from

$$(7 - \sqrt{15})(4 + \sqrt{15})^2 = 97 + 25\sqrt{15}.$$

Example 1.41. We search for integral solutions of $x^2 - 82y^2 = 31$. A fundamental unit of $\mathbf{Z}[\sqrt{82}]$ is $9 + \sqrt{82}$, which has norm -1, so its square $u := 163 + 18\sqrt{82}$ has norm 1.

If there is a solution at all, then there is one where $|x| \leqslant \sqrt{31u} \approx 100.5$ and $|y| \leqslant \sqrt{31u}/\sqrt{82} \approx 11.1$. For $0 \leqslant y \leqslant 11$, $82y^2 + 31$ is not a perfect square, so there is no integral solution to $x^2 - 82y^2 = 31$.

Incidentally, the equation $x^2 - 82y^2 = 31$ has infinitely many rational solutions, one being $(\frac{19}{3}, \frac{1}{3})$.

Although Theorem 1.39 provides a completely general method of deciding when $x^2 - dy^2 = n$ has an integral solution, if there are no solutions this can often be proved more simply using congruences. For instance, $x^2 - 5y^2 = 2$ has no integral solutions because if it did then $x^2 \equiv 2 \bmod 5$, which is impossible. But congruence methods do not always suffice to prove there are no solutions! The equation $x^2 - 82y^2 = 31$ illustrates this. We saw it has no integral solutions in Example 1.41, but for every $m \geqslant 2$ it turns out that the congruence $x^2 - 82y^2 \equiv 31 \bmod m$ is solvable (Exercise 1.19).

Another approach to solving $x^2 - dy^2 = n$ is through continued fractions.

(A reader who has never studied continued fractions can learn about them in [8, Chap. 15].) We will assume for the rest of this section that the reader has experience solving Pell's equation $x^2 - dy^2 = 1$ using continued fractions. The basic theorem linking Pell's equation to continued fractions is: if x and y are positive integers satisfying $x^2 - dy^2 = 1$, then x/y is a convergent of the continued fraction for $\sqrt{d}$. The proof of that theorem remains valid for solving $x^2 - dy^2 = n$ when $|n| < \sqrt{d}$: x/y is a convergent of the continued fraction for $\sqrt{d}$ (the fraction x/y might not be in reduced form if n is not squarefree). If P/Q runs through the convergents of the (periodic) continued fraction for $\sqrt{d}$ then the values of $P^2 - dQ^2$ repeat after at most the first two periods, so we can decide effectively, with continued fractions, if $x^2 - dy^2 = n$ is solvable in positive integers when $|n| < \sqrt{d}$.

If $|n| > \sqrt{d}$ then solvability of $x^2 - dy^2 = n$ can be connected to solvability of $x^2 - dy^2 = n'$ for some integer n' where $|n'| < |n|$. Iterating this, eventually the case $|n| < \sqrt{d}$ is reached and that case can be settled using the continued fraction of $\sqrt{d}$. Details of the reduction process are in [54, pp. 208–213]. We will illustrate it for $x^2 - 82y^2 = 31$. If this has a solution then $x^2 \equiv 82y^2 \equiv (12y)^2 \bmod 31$, so $x \equiv \pm 12y \bmod 31$. Write $x = \pm 12y + 31z$. Then

$$(\pm 12y + 31z)^2 - 82y^2 = 31 \Longrightarrow 2y^2 \pm 24yz + (31z^2 - 1) = 0.$$

Viewing the right side as a quadratic polynomial in y, the discriminant $(24z)^2 - 4 \cdot 2 \cdot (31z^2 - 1) = 4(82z^2 + 8)$ has to be a perfect square, so $82z^2 + 8 = t^2$ for some $t \in \mathbf{Z}$. We have reduced solvability of $x^2 - 82y^2 = 31$ to solvability of $t^2 - 82z^2 = 8$, and the greatest common divisor of t and z is 1 or 2 (in fact it must be 2). Since $|8| < \sqrt{82}$, we check the convergents P/Q of the first two periods of the continued fraction for $\sqrt{82}$, which is $[9, 18, 18, 18, \dots]$. The different values of $P^2 - 82Q^2$ are ± 1, so $(2P)^2 - 82(2Q)^2 = \pm 4$. We don't get 8 as a value, so $x^2 - 82y^2 = 31$ is not solvable in integers.

1.5 Overview

For any number field K, here are some general features of $\mathcal{O}_K$ that we will prove later on:

1. $\mathcal{O}_K$ is a free $\mathbf{Z}$-module of rank $n = [K : \mathbf{Q}]$:

$$\mathcal{O}_K = \mathbf{Z}e_1 \oplus \mathbf{Z}e_2 \oplus \cdots \oplus \mathbf{Z}e_n.$$

2. $\mathcal{O}_K^\times$ is finitely generated as an abelian group: for a finite set of units $u_1, \ldots, u_m$ in $\mathcal{O}_K$,
$$\mathcal{O}_K^\times = u_1^{\mathbf{Z}} u_2^{\mathbf{Z}} \cdots u_m^{\mathbf{Z}}.$$
This is called Dirichlet's unit theorem. Some of the u_i's could be roots of unity. (The theorem provides a formula for the smallest value of m, but we don't discuss that now.)

3. $\mathcal{O}_K$ may or may not be a unique factorization domain, but it always has unique factorization of (nonzero) ideals as products of prime ideals.

We have seen the first two features when K is a quadratic field. The third feature, that the ring of integers in a number field has unique factorization of ideals, was historically the setting in which the concept of an ideal was first defined by Dedekind in 1871, clarifying earlier work of Kummer on "ideal numbers" in the period 1844–1847.

To see how nonunique factorization of elements in $\mathbf{Z}[\sqrt{-5}]$ can give way to unique factorization of ideals, we first review addition and multiplication of ideals.

Definition 1.42. Let $\mathfrak{a}$ and $\mathfrak{b}$ be ideals in a commutative ring A. Their sum is $\mathfrak{a} + \mathfrak{b} = \{\alpha + \beta : \alpha \in \mathfrak{a},\ \beta \in \mathfrak{b}\}$ and their product is

$$\mathfrak{a}\mathfrak{b} = \{\alpha_1\beta_1 + \cdots + \alpha_r\beta_r : r \geqslant 1,\ \alpha_k \in \mathfrak{a},\ \beta_k \in \mathfrak{b}\}.$$

Both $\mathfrak{a} + \mathfrak{b}$ and $\mathfrak{a}\mathfrak{b}$ are ideals. The definition of $\mathfrak{a} + \mathfrak{b}$ is natural, but the definition of $\mathfrak{a}\mathfrak{b}$ might look a bit strange. Why do we need the sums of products and not just the products? Whatever $\mathfrak{a}\mathfrak{b}$ might mean, it should include the products $\alpha\beta$ where $\alpha \in \mathfrak{a}$ and $\beta \in \mathfrak{b}$, but the set of products is generally not an ideal (it is not usually closed under addition). This is why we set $\mathfrak{a}\mathfrak{b}$ to be the set of finite sums of such products: that is automatically an additive group. The product $\mathfrak{a}\mathfrak{b}$ is the smallest ideal containing all products $\alpha\beta$ where $\alpha \in \mathfrak{a}$ and $\beta \in \mathfrak{b}$.

Both $\mathfrak{a}$ and $\mathfrak{b}$ are subsets of $\mathfrak{a} + \mathfrak{b}$, so addition *enlarges* ideals. Since $\mathfrak{a}\mathfrak{b} \subset \mathfrak{a}$ and $\mathfrak{a}\mathfrak{b} \subset \mathfrak{b}$, multiplication *shrinks* ideals. (For example, $2\mathbf{Z} \cdot 5\mathbf{Z} = 10\mathbf{Z}$ and $10\mathbf{Z} \subset 2\mathbf{Z}$. Remember this example.) In terms of containments, we have

$$\mathfrak{a}\mathfrak{b} \subset \mathfrak{a} \cap \mathfrak{b} \subset \mathfrak{a} \subset \mathfrak{a} + \mathfrak{b}, \quad \mathfrak{a}\mathfrak{b} \subset \mathfrak{a} \cap \mathfrak{b} \subset \mathfrak{b} \subset \mathfrak{a} + \mathfrak{b}.$$

(While $\mathfrak{a} \cap \mathfrak{b}$ is an ideal, $\mathfrak{a} \cup \mathfrak{b}$ generally is not, even in $\mathbf{Z}$.)

In $\mathbf{Z}$, where all ideals are principal, $a\mathbf{Z}+b\mathbf{Z}=(a,b)\mathbf{Z}$, $a\mathbf{Z}\cdot b\mathbf{Z}=ab\mathbf{Z}$, and $a\mathbf{Z}\cap b\mathbf{Z}=[a,b]\mathbf{Z}$.

Theorem 1.43. *Let A be a commutative ring.*

(*a*) *For ideals $\mathfrak{a}$ and $\mathfrak{b}$ in A, $\mathfrak{ab}=\mathfrak{ba}$.*

(*b*) *If $\mathfrak{a}=(\alpha)=\alpha A$ is principal and $\mathfrak{b}$ is any ideal, $\mathfrak{ab}=\alpha\mathfrak{b}=\{\alpha\beta:\beta\in\mathfrak{b}\}$.*

(*c*) *For ideals $\mathfrak{a},\mathfrak{b},\mathfrak{c}$ in A, ideal multiplication is associative and is distributive over ideal addition:*

$$(\mathfrak{ab})\mathfrak{c}=\mathfrak{a}(\mathfrak{bc}),\quad \mathfrak{a}(\mathfrak{b}+\mathfrak{c})=\mathfrak{ab}+\mathfrak{ac}.$$

Proof. Exercise 1.22. For the equations in part c, show each side contains the other. Be careful for the distributive law not to assume elements from $\mathfrak{ab}$ and $\mathfrak{ac}$ that are to be added together have the same number of terms in the sums (or is that something you can always arrange? Does it even matter?). ∎

Taking $\alpha=1$ in part b shows $A=(1)$ is an identity for multiplication of ideals in A: $A\mathfrak{a}=\mathfrak{a}A=\mathfrak{a}$ for every ideal $\mathfrak{a}$. Part b also shows multiplication of principal ideals behaves as you would hope: $(\alpha)(\alpha')=(\alpha\alpha')$. Ideals do *not* form a ring under addition and multiplication since there are no additive inverses: we can't have $\mathfrak{a}+\mathfrak{b}=(0)$ if $\mathfrak{a}$ or $\mathfrak{b}$ is not (0).

Example 1.44. In the ring $\mathbf{Z}[\sqrt{-5}]$, let $\mathfrak{p}=(2,1+\sqrt{-5})$. Then the elements of $\mathfrak{p}^2$ are finite sums of products of the form

$$\begin{aligned}(2\alpha+(1+\sqrt{-5})\beta)(2\gamma+(1+\sqrt{-5})\delta) &= 4\alpha\gamma+2(1+\sqrt{-5})\alpha\delta\\ &\quad +2(1+\sqrt{-5})\beta\gamma+2(-2+\sqrt{-5})\beta\delta,\end{aligned}$$

where $\alpha,\beta,\gamma,\delta$ are in $\mathbf{Z}[\sqrt{-5}]$. Each of the four terms is divisible by 2, so $\mathfrak{p}^2\subset(2)$. The reverse inclusion holds as well since $2\in\mathfrak{p}^2$:

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

Although 2 is irreducible in $\mathbf{Z}[\sqrt{-5}]$, the ideal (2) is not prime: $(1+\sqrt{-5})^2\in(2)$ and $1+\sqrt{-5}\not\in(2)$. Don't fall into the trap of thinking "if π is irreducible then the ideal (π) is prime." That is true in a UFD but not more generally. If a domain[4] is not a UFD then *an irreducible element need not generate a prime ideal.*

Remark 1.45. When we are given generators of two ideals, the product of the ideals is the ideal generated by all the pairwise products. For example, when $\mathfrak{p} = (2, 1+\sqrt{-5})$ in $\mathbf{Z}[\sqrt{-5}]$,

$$\begin{aligned} \mathfrak{p}^2 &= (2, 1+\sqrt{-5})(2, 1+\sqrt{-5}) \\ &= (4, 2(1+\sqrt{-5}), (1+\sqrt{-5})^2) \\ &= (4, 2+2\sqrt{-5}, -4+2\sqrt{-5}) \\ &= (2)(2, 1+\sqrt{-5}, -2+\sqrt{-5}). \end{aligned}$$

The ideal $(2, 1+\sqrt{-5}, -2+\sqrt{-5})$ contains 2 and $1+\sqrt{-5}-(-2+\sqrt{-5}) = 3$, so it contains 1 and is thus the unit ideal. This shows $\mathfrak{p}^2 = (2)(1) = (2)$ in a simpler way than our computation in Example 1.44.

Now we are ready to turn a nonunique irreducible factorization in $\mathbf{Z}[\sqrt{-5}]$ into a unique prime ideal factorization in $\mathbf{Z}[\sqrt{-5}]$.

Example 1.46. In $\mathbf{Z}[\sqrt{-5}]$,

$$6 = 2 \cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5}). \tag{1.10}$$

These are different irreducible factorizations of 6. An equation of elements implies an equation of the principal ideals they generate: if $x = yz$ as elements of $\mathbf{Z}[\sqrt{-5}]$ then $(x) = (yz) = (y)(z)$ as principal ideals. So

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

Although 2, 3, $1+\sqrt{-5}$, and $1-\sqrt{-5}$ are irreducible as elements of $\mathbf{Z}[\sqrt{-5}]$, the principal ideals they generate factor nontrivially. Set

$$\mathfrak{p} = (2, 1+\sqrt{-5}), \quad \mathfrak{q} = (3, 1+\sqrt{-5}), \quad \mathfrak{q}' = (3, 1-\sqrt{-5}).$$

We have seen before that $(2) = \mathfrak{p}^2$. It turns out that

$$(3) = \mathfrak{q}\mathfrak{q}', \quad (1+\sqrt{-5}) = \mathfrak{p}\mathfrak{q}, \quad (1-\sqrt{-5}) = \mathfrak{p}\mathfrak{q}', \tag{1.11}$$

[4] We always use “domain” to mean “integral domain”.

so the two *different* principal ideal factorizations

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

decompose further into the *same* ideal factorizations:

$$(6) = \mathfrak{p}\mathfrak{p} \cdot \mathfrak{q}\mathfrak{q}' = \mathfrak{p}\mathfrak{q} \cdot \mathfrak{p}\mathfrak{q}'. \tag{1.12}$$

To verify (1.11), let's check $\mathfrak{p}\mathfrak{q} = (1+\sqrt{-5})$:

$$\begin{aligned} \mathfrak{p}\mathfrak{q} &= (2, 1+\sqrt{-5})(3, 1+\sqrt{-5}) \\ &= (6, 2+2\sqrt{-5}, 3+3\sqrt{-5}, (1+\sqrt{-5})(1+\sqrt{-5})). \end{aligned}$$

Since $6 = (1+\sqrt{-5})(1-\sqrt{-5})$, each of the 4 generators of $\mathfrak{p}\mathfrak{q}$ is a multiple of $1+\sqrt{-5}$. Therefore $\mathfrak{p}\mathfrak{q} \subset (1+\sqrt{-5})$. For the reverse inclusion we take the difference of the second and third generators of $\mathfrak{p}\mathfrak{q}$:

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

Therefore $(1+\sqrt{-5}) \subset \mathfrak{p}\mathfrak{q}$, so $\mathfrak{p}\mathfrak{q} = (1+\sqrt{-5})$. Checking the first and third equations in (1.11) is left to you as part of Exercise 1.23. (The ideal $\mathfrak{p} = (2, 1+\sqrt{-5})$ can be written as $\mathfrak{p} = (2, 1-\sqrt{-5})$, so the computation of $\mathfrak{p}\mathfrak{q}'$ can be carried out in an identical, and not merely analogous, manner to the computation of $\mathfrak{p}\mathfrak{q} = (2, 1+\sqrt{-5})(3, 1+\sqrt{-5})$ by replacing $\sqrt{-5}$ with $-\sqrt{-5}$ everywhere.)

The ideals $\mathfrak{p}$, $\mathfrak{q}$, and $\mathfrak{q}'$ are all prime. We will check this by showing they are maximal ideals (any maximal ideal is a prime ideal). By definition,

$$\mathfrak{p} = \mathbf{Z}[\sqrt{-5}] \cdot 2 + \mathbf{Z}[\sqrt{-5}](1+\sqrt{-5})$$

and

$$\mathfrak{q} = \mathbf{Z}[\sqrt{-5}] \cdot 3 + \mathbf{Z}[\sqrt{-5}](1+\sqrt{-5}), \quad \mathfrak{q}' = \mathbf{Z}[\sqrt{-5}] \cdot 3 + \mathbf{Z}[\sqrt{-5}](1-\sqrt{-5}),$$

but these generators using $\mathbf{Z}[\sqrt{-5}]$-coefficients even work with integral coefficients:

$$\mathfrak{p} = \mathbf{Z}\cdot 2 + \mathbf{Z}(1+\sqrt{-5}), \quad \mathfrak{q} = \mathbf{Z}\cdot 3 + \mathbf{Z}(1+\sqrt{-5}), \quad \mathfrak{q}' = \mathbf{Z}\cdot 3 + \mathbf{Z}(1-\sqrt{-5}). \tag{1.13}$$

Obviously the right side of each equation in (1.13) is in the left side. Checking the left sides are in the right sides is Exercise 1.24. Lest you think there is some general principle that the generators of an ideal in $\mathbf{Z}[\sqrt{-5}]$ will always span the ideal using $\mathbf{Z}$-coefficients, notice

$$3\sqrt{-5} \in (3, 1+2\sqrt{-5}), \quad 3\sqrt{-5} \not\in \mathbf{Z}\cdot 3 + \mathbf{Z}(1+2\sqrt{-5}).$$

So there really is something to check in (1.13). In any case, once you have verified (1.13), the index of the ideals in $\mathbf{Z}[\sqrt{-5}] = \mathbf{Z} + \mathbf{Z}(1+\sqrt{-5})$ can easily be computed using the $\mathbf{Z}$-basis $\{1, 1+\sqrt{-5}\}$ for $\mathbf{Z}[\sqrt{-5}]$ in place of the usual $\mathbf{Z}$-basis $\{1, \sqrt{-5}\}$:

$$\mathbf{Z}[\sqrt{-5}]/\mathfrak{p} = (\mathbf{Z} \oplus \mathbf{Z}(1+\sqrt{-5}))/(\mathbf{Z}\cdot 2 \oplus \mathbf{Z}(1+\sqrt{-5})) \cong \mathbf{Z}/2\mathbf{Z},$$

and similarly

$$\mathbf{Z}[\sqrt{-5}]/\mathfrak{q} \cong \mathbf{Z}/3\mathbf{Z}, \quad \mathbf{Z}[\sqrt{-5}]/\mathfrak{q}' \cong \mathbf{Z}/3\mathbf{Z}.$$

Rings of size 2 or 3 must be fields, so $\mathfrak{p}$, $\mathfrak{q}$, and $\mathfrak{q}'$ are maximal ideals in $\mathbf{Z}[\sqrt{-5}]$. This completes the proof that passing to ideals turns the different irreducible factorizations of the element 6 in (1.10) into the same prime ideal factorization of the principal ideal (6) in (1.12).

1.6 Exercises

1. The following table shows in the second row fraktur versions of several letters in the first row. Write each fraktur letter 25 times, or at least until you can comfortably reproduce the letter. Note $\mathfrak{P}$ is a capital P, *not* B (the capital B is $\mathfrak{B}$).

a	b	c	d	p	q	P	Q
$\mathfrak{a}$	$\mathfrak{b}$	$\mathfrak{c}$	$\mathfrak{d}$	$\mathfrak{p}$	$\mathfrak{q}$	$\mathfrak{P}$	$\mathfrak{Q}$

2. Determine if the following numbers are algebraic integers:

a) $\dfrac{\sqrt{5}+\sqrt{17}}{\sqrt{2}}$.

b) $\dfrac{5+\sqrt{-7}}{\sqrt{3}}$.

c) $\dfrac{1+\sqrt[3]{17}}{\sqrt{3}}$.

d) $\dfrac{\sqrt{10+2\sqrt{5}}-2}{\sqrt{5}+1}$.

3. Suppose $\alpha^3 - \alpha - 2 = 0$. Express α^m in the form $a + b\alpha + c\alpha^2$ $(a, b, c \in \mathbf{Z})$ for $m = 3, 4, 5, 6, 7$.

4. a) Show $\mathbf{Z} + \mathbf{Z}\sqrt[3]{12} + \mathbf{Z}\sqrt[3]{18}$ is a ring.

   b) Find a monic polynomial in $\mathbf{Z}[T]$ with $\sqrt[3]{12} + \sqrt[3]{18}$ as a root.

   c) Find a monic polynomial in $\mathbf{Z}[T]$ with $\sqrt[3]{2} + \sqrt[3]{3}$ as a root.

5. Let d be a squarefree integer.

   a) If $d \equiv 1 \bmod 4$, show

$$\mathbf{Z}\left[\frac{1+\sqrt{d}}{2}\right] = \left\{\frac{a+b\sqrt{d}}{2} : a \equiv b \bmod 2\right\}.$$

   b) For all d, show the ring of integers of $\mathbf{Q}(\sqrt{d})$ is $\mathbf{Z} + \mathbf{Z}\frac{d+\sqrt{d}}{2}$.

6. Let $K = \mathbf{Q}(\sqrt{-3})$. Inside K is $\zeta_3 = (-1+\sqrt{-3})/2$, a nontrivial cube root of unity. (Note ζ_3 is *not* $(1+\sqrt{-3})/2 = 1 + \zeta_3 = -\zeta_3^2$ and $\mathbf{Z}[\frac{1+\sqrt{-3}}{2}] = \mathbf{Z}[1+\zeta_3] = \mathbf{Z}[\zeta_3]$.) The ring $\mathbf{Z}[\zeta_3]$ is the full ring of integers in K and is called the ring of *Eisenstein integers*. It contains $\mathbf{Z}[\sqrt{-3}] = \mathbf{Z}[2\zeta_3]$ with index 2.

   The norm formulas from K to $\mathbf{Q}$ when elements are written using the $\mathbf{Q}$-bases $\{1, \sqrt{-3}\}$ and $\{1, \zeta_3\}$ are

$$\mathrm{N}(x + y\sqrt{-3}) = (x + y\sqrt{-3})(x - y\sqrt{-3}) = x^2 + 3y^2$$

   and

$$\mathrm{N}(a + b\zeta_3) = (a + b\zeta_3)(a + b\zeta_3^{-1}) = a^2 - ab + b^2$$

   for rational x, y, a, and b.

   a) Show the units of $\mathbf{Z}[\zeta_3]$ are $\{\pm 1, \pm\zeta_3, \pm\zeta_3^2\}$.

   b) Show $\mathbf{Z}[\zeta_3]$ is Euclidean with respect to the norm, so $\mathbf{Z}[\zeta_3]$ is a UFD (unlike $\mathbf{Z}[\sqrt{-3}]$).

   c) Explain why the equation

$$21 = 3 \cdot 7 = (3 + 2\sqrt{-3})(3 - 2\sqrt{-3})$$

is not an example of nonunique factorization in $\mathbf{Z}[\zeta_3]$.

d) Although $\mathbf{Z}[\sqrt{-3}]$ is a proper subset of $\mathbf{Z}[\zeta_3]$, show the norm values on both rings are the same: for any $\alpha \in \mathbf{Z}[\zeta_3]$ there is a $\beta \in \mathbf{Z}[\sqrt{-3}]$ such that $\mathrm{N}(\alpha) = \mathrm{N}(\beta)$. (The other way is automatic: for every $\beta \in \mathbf{Z}[\sqrt{-3}]$ there is an $\alpha \in \mathbf{Z}[\zeta_3]$ such that $\mathrm{N}(\alpha) = \mathrm{N}(\beta)$ because we can use $\alpha = \beta$.) In simpler terms, you're being asked to show for any a and b in $\mathbf{Z}$ that $a^2 - ab + b^2 = x^2 + 3y^2$ for some x and y in $\mathbf{Z}$. (Hint: Consider the norm of $(a + b\zeta_3)u$ for $u = 1, \zeta_3$, or ζ_3^2. At least one product is in $\mathbf{Z}[\sqrt{-3}]$.)

e) Use previous parts of this exercise to show for any prime $p \neq 2$ that

$$-3 \equiv \square \bmod p \Longleftrightarrow p = x^2 + 3y^2 \text{ for some } x, y \in \mathbf{Z}.$$

(This equivalence would follow from $\mathbf{Z}[\sqrt{-3}]$ being a UFD, but it is not a UFD. The result is nevertheless correct. Somewhere in your argument there should be a step which relies on $p \neq 2$.)

7. Use parts a and b of the previous exercise to show a prime number can be written in the form $x^2 + 3y^2$, with x and y in $\mathbf{Z}$, in at most one way up to the choice of signs on x and y. (Hint: $x \not\equiv y \bmod 2$.)

8. Let K be a quadratic field.

a) For each integer $c \geqslant 1$ show $\mathbf{Z} + c\mathcal{O}_K$ is the unique subring of $\mathcal{O}_K$ with index c. Writing $\mathcal{O}_K = \mathbf{Z} + \mathbf{Z}\omega = \mathbf{Z}[\omega]$ for some ω (a choice for ω is $\sqrt{d}$ or $\frac{1+\sqrt{d}}{2}$ depending on $d \bmod 4$ in the usual notation), check

$$\mathbf{Z} + c\mathcal{O}_K = \mathbf{Z} + \mathbf{Z}c\omega = \mathbf{Z}[c\omega].$$

b) Show any subring of $\mathcal{O}_K$ other than $\mathbf{Z}$ has finite index in $\mathcal{O}_K$, so part a describes all subrings of $\mathcal{O}_K$ other than $\mathbf{Z}$. (Your argument has to distinguish between subrings and subgroups, since there are many subgroups of $\mathcal{O}_K$ other than $\mathbf{Z}$ that don't have finite index: $\mathbf{Z}\alpha$ for any α in $\mathcal{O}_K$.)

c) Show a domain of characteristic 0 which has rank 2 as a $\mathbf{Z}$-module is a quadratic ring (Definition 1.23), and conversely.

9. a) In $\mathbf{Z}[\sqrt{6}]$, $2 \cdot 3 = \sqrt{6}^2$ is a square and 2 and 3 have no common factors except units (after all, their difference is 1). Are 2 and 3 unit multiples of squares in $\mathbf{Z}[\sqrt{6}]$?

b) In $\mathbf{Z}[\sqrt{-6}]$, $2\cdot(-3) = \sqrt{-6}^2$ is a square and 2 and -3 have no common factors except units (their sum is -1). Are 2 and -3 unit multiples of squares in $\mathbf{Z}[\sqrt{-6}]$?

10. a) Find the fundamental unit of $\mathbf{Z}[\sqrt{d}]$ for $d = 3, 5, 6, 7, 8, 10, 12$, and 145.

    b) Describe all the units in $\mathbf{Z}[\sqrt{3}]$ and $\mathbf{Z}[\sqrt{6}]$.

    c) Find the first four units greater than 1 in $\mathbf{Z}[\sqrt{2}]$, $\mathbf{Z}[\sqrt{3}]$, and $\mathbf{Z}[\sqrt{12}]$.

11. a) Verify the following are fundamental units in $\mathbf{Z}[\sqrt{d}]$ subject to the indicated constraint. (The values of d in the table may not be squarefree, but they are never perfect squares.)

d	Fundamental Unit	Constraint
n^2+1	$n+\sqrt{n^2+1}$	$n \geqslant 1$
n^2-1	$n+\sqrt{n^2-1}$	$n \geqslant 2$
n^2+2	$n^2+1+n\sqrt{n^2+2}$	$n \geqslant 1$
n^2-2	$n^2-1+n\sqrt{n^2-2}$	$n \geqslant 3$

    b) Verify the following are fundamental units in $\mathbf{Z}[\frac{1+\sqrt{d}}{2}]$ subject to the indicated constraint.

d	Fundamental Unit	Constraint
n^2+4	$\frac{n+\sqrt{n^2+4}}{2}$	$n \geqslant 1$
n^2-4	$\frac{n+\sqrt{n^2-4}}{2}$	$n \geqslant 4$

    Note. Knowing the unit group of $\mathbf{Z}[\sqrt{n^2-1}]$ is important in one of the proofs that $x^2 - y^q = 1$ has no solution in nonzero integers x and y when q is a prime at least 5. See [49, pp. 14–15].

12. Let d be a squarefree positive integer. Suppose $d \equiv 1 \bmod 4$, so $d \equiv 1 \bmod 8$ or $d \equiv 5 \bmod 8$.

    a) If $d \equiv 1 \bmod 8$, show every *unit* in $\mathbf{Z}[\frac{1+\sqrt{d}}{2}]$ is in $\mathbf{Z}[\sqrt{d}]$. (Hint: It is convenient to write $\mathbf{Z}[\frac{1+\sqrt{d}}{2}]$ as $\{\frac{a+b\sqrt{d}}{2} : a \equiv b \bmod 2\}$, as in Exercise 1.5a.)

    b) The first few squarefree $d \equiv 5 \bmod 8$ are $5, 13, 21, 29, 37, and 53$. For these d, determine when the fundamental unit of $\mathbf{Z}[\frac{1+\sqrt{d}}{2}]$ is in $\mathbf{Z}[\sqrt{d}]$.

    c) If $d \equiv 5 \bmod 8$ show every $\alpha \in \mathbf{Z}[\frac{1+\sqrt{d}}{2}]$ (not just units) satisfies $\alpha^3 \in \mathbf{Z}[\sqrt{d}]$, and check this explicitly for the fundamental units found in part

b. (It is still convenient, as in part a, to write $\mathbf{Z}[\frac{1+\sqrt{d}}{2}] = \{\frac{a+b\sqrt{d}}{2} : a \equiv b \bmod 2\}$.)

13. a) For a prime $p \equiv 3 \bmod 4$, show each unit in $\mathbf{Z}[\sqrt{p}]$ has norm 1. (That is, no unit has norm -1.)

b) For a prime $p \equiv 1 \bmod 4$, show some unit in $\mathbf{Z}[\sqrt{p}]$ has norm -1. (Hint: Let $x + y\sqrt{p} > 1$ be the least unit greater than 1 with norm 1. Then x and y are positive integers. Show x is odd, y is even, and then from the equation $py^2 = x^2 - 1 = (x+1)(x-1)$ show there is an equation $m^2 - pn^2 = 1$ where $1 < m < x$. This can't continue forever, so some unit must have norm -1.)

14. Let F be a field not of characteristic 2 and $f(X) \in F[X]$ be nonconstant and not a perfect square. The ring $F[X, \sqrt{f(X)}] = F[X] + F[X]\sqrt{f(X)}$ is analogous to $\mathbf{Z}[\sqrt{d}]$.

a) For $a(X)$ and $b(X)$ in $F[X]$, show $a(X) + b(X)\sqrt{f(X)}$ is a unit in $F[X, \sqrt{f(X)}]$ if and only if $a(X)^2 - f(X)b(X)^2 \in F^\times$ (analogue of Theorem 1.24) and conclude that $F[X, \sqrt{f(X)}]^\times = F^\times$ if $\deg f(X)$ is odd or if the leading coefficient of $f(X)$ is not a square in F.

b) For an indeterminate U, show every unit in the ring $F[U, 1/U] = \{a(U)/U^n : a(U) \in F[U], n \geqslant 0\}$ has the form cU^k for $c \in F^\times$ and $k \in \mathbf{Z}$. Briefly, $F[U, 1/U]^\times = F^\times U^{\mathbf{Z}}$.

c) Show $F[X, \sqrt{X^2+1}]^\times = \{c(X + \sqrt{X^2+1})^k : c \in F^\times, k \in \mathbf{Z}\}$ and $F[X, \sqrt{X^2-1}]^\times = \{c(X + \sqrt{X^2-1})^k : c \in F^\times, k \in \mathbf{Z}\}$. Note the similarity to the fundamental units in the table in Exercise 1.11. (Hint: Show $F[X, \sqrt{X^2+1}] = F[U, 1/U]$ when $U = X + \sqrt{X^2+1}$.)

15. Continue the notation from the previous exercise: $f(X)$ is nonconstant and not a perfect square in $F[X]$. Assume that F does not have characteristic 2. This exercise develops (following a method of Dubickas and Steuding [16]) an $F[X]$-analogue of Theorem 1.31: if the unit group $F[X, \sqrt{f(X)}]^\times$ is larger than $F^\times$ then it equals $F^\times\varepsilon^{\mathbf{Z}}$ for some unit $\varepsilon \notin F^\times$.

a) Assume there are units not in F. Let $\varepsilon = a + b\sqrt{f}$ be such a unit with $\deg(b)$ minimal (necessarily $b(X) \neq 0$). For any unit $u = A + B\sqrt{f}$ not in

F,

$$\begin{aligned}\varepsilon u &= (aA + fbB) + (aB + bA)\sqrt{f},\\ \overline{\varepsilon} u &= (aA - fBb) + (aB - bA)\sqrt{f},\end{aligned}$$

where $\overline{\varepsilon} = a - b\sqrt{f}$. Show aB and bA have the same degree and their leading coefficients are equal up to sign, so if the coefficients of $\sqrt{f}$ in εu and $\overline{\varepsilon} u$ are both nonzero then the coefficients have different degrees.

b) Let $s = aB + bA$ and $t = aB - bA$ (the coefficients of $\sqrt{f}$). If s and t are nonzero, show $st \in F^\times B^2 + F^\times b^2$ and conclude that s or t has degree less than $\deg(B)$. (Recall $\deg(B) \geqslant \deg(b)$ by minimality of $\deg(b)$.)

c) Prove by induction on $\deg(B)$ that every unit in $F[X, \sqrt{f(X)}]^\times$ is an element of $F^\times$ times a power of ε.

d) We call any unit in $F[X, \sqrt{f(X)}]$ which, together with $F^\times$, generates all units a fundamental unit of this ring. Show $F[X, \sqrt{X^2 + X}]$ has fundamental unit $2X + 1 + 2\sqrt{X^2 + X}$ and $F[X, \sqrt{X^4 + X}]$ has fundamental unit $2X^3 + 1 + 2X\sqrt{X^4 + X}$.

e) If F has characteristic 3, show $F[X, \sqrt{X^4 + X^3}]$ has fundamental unit $(X + 2)^3 + (X + 1)\sqrt{X^4 + X^3}$.

f) If F has characteristic 0, show the only units in $F[X, \sqrt{X^4 + X^3}]$ are elements of $F^\times$. (Hint: Write $F[X, \sqrt{X^4 + X^3}] = F[X, X\sqrt{X^2 + X}]^\times$ and use part d.)

16. Let $\mathbf{F}$ be a *finite* field with odd characteristic and $f(X)$ be a nonsquare in $\mathbf{F}[X]$ with positive degree. If there is an integer N such that $\deg(g^2 - fh^2) \leqslant N$ for infinitely many pairs (g, h) in $\mathbf{F}[X]$, show there is a nontrivial solution (g, h) to Pell's equation $g^2 - fh^2 = 1$ in $\mathbf{F}[X]$, and therefore $\mathbf{F}[X, \sqrt{f(X)}]$ has units outside of $\mathbf{F}$.

17. Show all the integral solutions to $x^2 - 10y^2 = 6$ are given by the formulas $x + y\sqrt{10} = (4 + \sqrt{10})u^k$ and $x + y\sqrt{10} = (4 - \sqrt{10})u^k$, where $u = 19 + 6\sqrt{10}$ and $k \in \mathbf{Z}$.

18. a) Show all the integral solutions to $x^2 - 6y^2 = 10$ are given by the formulas $x + y\sqrt{6} = (4 + \sqrt{6})u^k$ and $x + y\sqrt{6} = (4 - \sqrt{6})u^k$, where $u = 5 + 2\sqrt{6}$ and $k \in \mathbf{Z}$.

b) Use part a to show any integral solution of $x^2 - 6y^2 = 10$ has $x \equiv 0 \bmod 4$ and y odd. Then draw this conclusion using congruence arguments without part a.

c) A triangular number is an integer of the form $T_n = 1 + 2 + \cdots + n = \frac{1}{2}n(n+1)$. The sum $T_{14} + T_{15} + T_{16}$ is $361 = 19^2$. Show the task of finding $n \geqslant 1$ such that $T_{n-1} + T_n + T_{n+1}$ is a perfect square is the same as finding solutions to $x^2 - 6y^2 = 10$ in positive integers x and y (excluding small y, which lead to $n \leqslant 0$). Then use part a to find two sets of three consecutive triangular numbers besides T_{14}, T_{15}, and T_{16} whose sum is a perfect square.

19. a) By Example 1.41, $x^2 - 82y^2 = 31$ has no integral solution. It has the rational solutions $(\frac{19}{3}, \frac{1}{3})$ and $(\frac{67}{11}, \frac{3}{11})$. Use this to show the congruence $x^2 - 82y^2 \equiv 31 \bmod m$ has a solution for all $m \geqslant 2$. Thus the lack of integral solutions can't be proved by congruence methods. (Hint: The first rational solution makes sense mod m if $(3, m) = 1$ and the second makes sense mod m if $(11, m) = 1$. Now use the Chinese remainder theorem.)

    b) The equation $x^2 - 34y^2 = -1$ has the rational solutions $(\frac{5}{3}, \frac{1}{3})$ and $(\frac{3}{5}, \frac{1}{5})$. Use this to show the congruence $x^2 - 34y^2 \equiv -1 \bmod m$ has a solution for all $m \geqslant 2$ and then show the equation $x^2 - 34y^2 = -1$ does not have a solution in $\mathbf{Z}$.

    c) The equations $2x^2 + 7y^2 = 1$, $3x^2 + 13y^2 = 1$, and $5x^2 + 11y^2 = 1$ obviously have no integral solutions. Show each equation can be solved as a congruence mod m for every m using a suitable pair of rational solutions for each equation.

    d) In $\mathbf{F}_3(X)$, the equation $g^2 - (X^3 + X^2 + 2)h^2 = X^4 + X + 2$ has solutions $(\frac{X^3+2X^2+1}{X+2}, \frac{1}{X+2})$ and $(\frac{X^4+2X^3+2X^2+X+2}{X^2+2X+2}, \frac{X+2}{X^2+2X+2})$. Show this equation has no solution in $\mathbf{F}_3[X]$ but the congruence $g^2 - (X^3 + X^2 + 2)h^2 \equiv X^4 + X + 2 \bmod m(X)$ has a solution for all nonconstant $m(X) \in \mathbf{F}_3[X]$.

20. Determine if the following equations have any integral solutions (x, y), and if so describe all solutions without repetition.

    a) $x^2 - 2y^2 = 31$.

    b) $x^2 - 2y^2 = 55$.

    c) $x^2 - 2y^2 = 119$.

    d) $x^2 - 2y^2 = 146$.

e) $x^2 - 65y^2 = 49$.

f) $x^2 - 65y^2 = 536$.

21. a) For integers a, b, and c such that $d := b^2 - 4ac$ is not a perfect square (so $a \neq 0$), show the integral solutions (x, y) of $ax^2 + bxy + cy^2 = n$ correspond to the integral solutions (X, Y) of $X^2 - dY^2 = 4an$ such that $X \equiv bY \bmod 2a$.

    b) Determine the integral solutions to $x^2 + 12xy + y^2 = 401$.

    c) Determine the integral solutions to $2x^2 + 11xy + 10y^2 = 39$.

    d) Determine the integral solutions to $2x^2 - 17y^2 = 1$.

    e) Determine the integral solutions to $2x^2 - 17y^2 = -1$.

22. Show multiplication of ideals in any commutative ring is commutative, associative, and distributes over addition of ideals.

23. In $\mathbf{Z}[\sqrt{-5}]$, verify the following formulas for products of ideals (remember that the only ideal containing 1 is (1)):

$$\begin{aligned} (3, 1+\sqrt{-5})(3, 1-\sqrt{-5}) &= (3), \\ (2, 1+\sqrt{-5})(3, 1-\sqrt{-5}) &= (1-\sqrt{-5}), \\ (2, 1+\sqrt{-5})(7, 3+\sqrt{-5}) &= (3+\sqrt{-5}), \\ (3, 1+\sqrt{-5})^2 &= (2-\sqrt{-5}). \end{aligned}$$

24. Verify (1.13).

25. Let d be a nonsquare integer with $d \equiv 1 \bmod 4$.

    a) In $\mathbf{Z}[\sqrt{d}]$ (which is not the full ring of integers of $\mathbf{Q}(\sqrt{d})$), show the ideal $(2, 1+\sqrt{d})$ has $\mathbf{Z}$-basis $\{2, 1+\sqrt{d}\}$. That is, $\mathbf{Z}[\sqrt{d}] \cdot 2 + \mathbf{Z}[\sqrt{d}](1+\sqrt{d}) = \mathbf{Z} \cdot 2 \oplus \mathbf{Z}(1+\sqrt{d})$.

    b) Use part a to show $(2, 1+\sqrt{d})$ has index 2 in $\mathbf{Z}[\sqrt{d}]$, so $(2, 1+\sqrt{d})$ is a maximal ideal.

    c) In $\mathbf{Z}[\sqrt{d}]$, show the ideals (2) and $(2, 1+\sqrt{d})$ are distinct, yet

$$(2, 1+\sqrt{d})(2, 1+\sqrt{d}) = (2)(2, 1+\sqrt{d}).$$

Thus, cancellation of ideals is not generally valid in $\mathbf{Z}[\sqrt{d}]$. Why does this problem disappear if we use the ideals with the same generators in the full ring of integers $\mathbf{Z}[\frac{1+\sqrt{d}}{2}]$?

26. Let $d \in \mathbf{Z}$ not be a perfect square.

a) In $\mathbf{Z}[\sqrt{d}]$, if $\mathrm{N}(x+y\sqrt{d}) = \pm p$ for some prime number p, show $x+y\sqrt{d}$ is irreducible. (Recall the characterization of units in $\mathbf{Z}[\sqrt{d}]$ in Theorem 1.24.)

b) If $\mathbf{Z}[\sqrt{d}]$ has unique factorization, show the following conditions on a prime number p are equivalent:

- p is reducible in $\mathbf{Z}[\sqrt{d}]$,
- $\pm p = x^2 - dy^2$ for some integers x and y (and some choice of sign on the left),
- $d \equiv \square \bmod p$.

Only the implication from the third condition to the first should use unique factorization.

c) When $d = -10$, 10, 34, and 79, find a prime p such that $d \equiv \square \bmod p$ and neither p nor $-p$ has the form $x^2 - dy^2$. Therefore by part b, such rings $\mathbf{Z}[\sqrt{d}]$ do not have unique factorization, and this has been proved without finding examples of nonunique factorization. Instead, you showed these rings fail to satisfy a consequence of unique factorization.

27. Let $d \in \mathbf{Z}$ with $d \neq \square$.

a) For nonzero $n \in \mathbf{Z}$, show $\#(\mathbf{Z}[\sqrt{d}]/(n)) = n^2$.

b) For nonzero $a + b\sqrt{d} \in \mathbf{Z}[\sqrt{d}]$, show

$$\#(\mathbf{Z}[\sqrt{d}]/(a+b\sqrt{d})) = |a^2 - db^2|$$

by considering the chain of ideals $(\mathrm{N}(\alpha)) \subset (\alpha) \subset \mathbf{Z}[\sqrt{d}]$, where $\alpha = a + b\sqrt{d}$, and the indices of these ideals as subgroups of one another.

c) In $\mathbf{Z}[\sqrt{82}]$, let $\mathfrak{a}$ be the ideal $(31, \sqrt{82}+12)$. Show $\#(\mathbf{Z}[\sqrt{82}]/\mathfrak{a}) = 31$ and use part b and Example 1.41 to show $\mathfrak{a}$ is not a principal ideal.

28. Let d be a nonsquare integer. In $\mathbf{Z}[\sqrt{d}]$, let $\mathfrak{a}$ be the ideal $(a, b+c\sqrt{d})$, where a, b, and c are integers and a and c are not 0. So as a $\mathbf{Z}[\sqrt{d}]$-module,

$$\mathfrak{a} = \mathbf{Z}[\sqrt{d}]a + \mathbf{Z}[\sqrt{d}](b+c\sqrt{d}),$$

while as a $\mathbf{Z}$-module

$$\mathfrak{a} = \mathbf{Z}a + \mathbf{Z}a\sqrt{d} + \mathbf{Z}(b+c\sqrt{d}) + \mathbf{Z}(cd+b\sqrt{d}).$$

It is natural to ask: does $\mathfrak{a} = \mathbf{Z}a + \mathbf{Z}(b+c\sqrt{d})$?

a) Show $\mathfrak{a} = \mathbf{Z}a + \mathbf{Z}(b+c\sqrt{d})$, that is, the given $\mathbf{Z}[\sqrt{d}]$-module generators are also $\mathbf{Z}$-module generators, if and only if the following three conditions are all satisfied: $c \mid a$, $c \mid b$, and $d \equiv (b/c)^2 \bmod a/c$. (In particular, when $\mathfrak{a} = (a, b \pm \sqrt{d})$, we have $\mathfrak{a} = \mathbf{Z}a + \mathbf{Z}(b \pm \sqrt{d})$ if and only if $d \equiv b^2 \bmod a$.)

b) In $\mathbf{Z}[\sqrt{-5}]$ let $\mathfrak{p} = (2, 1+\sqrt{-5})$, $\mathfrak{q} = (3, 1+\sqrt{-5})$, and $\mathfrak{q}' = (3, 1-\sqrt{-5})$. Show the indicated ideal generators are also $\mathbf{Z}$-module generators: $\mathfrak{p} = \mathbf{Z}\cdot 2 + \mathbf{Z}(1+\sqrt{-5})$, $\mathfrak{q} = \mathbf{Z}\cdot 3 + \mathbf{Z}(1+\sqrt{-5})$, and $\mathfrak{q}' = \mathbf{Z}\cdot 3 + \mathbf{Z}(1-\sqrt{-5})$.

c) Find an element of the ideal $(3, 1+2\sqrt{-5})$ which is not a $\mathbf{Z}$-linear combination of 3 and $1+2\sqrt{-5}$. (So 3 and $1+2\sqrt{-5}$ span the ideal as a $\mathbf{Z}[\sqrt{-5}]$-module but not as a $\mathbf{Z}$-module.) Then find a pair of elements which generates the ideal $(3, 1+2\sqrt{-5})$ as both a $\mathbf{Z}[\sqrt{-5}]$-module and as a $\mathbf{Z}$-module. Do the same for the ideal $(7, 2+3\sqrt{-5})$.

CHAPTER 2

SOME COMMUTATIVE ALGEBRA

Before we study general rings of algebraic integers $\mathcal{O}_K$ closely in later chapters, we describe the concept of "integral element" for any extension of commutative rings. This helps to clarify properties of algebraic integers that are not specific to number theory.

2.1 Integral Ring Extensions

Let B/A be an extension of commutative rings. (The notation B/A just means $B \supset A$, as in field theory; it is not a quotient object.) We call an element $b \in B$ *integral* over A if

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

for some $n \geqslant 1$ with $a_i \in A$. Such an equation is called an equation of integral dependence of b over A. The key point is that the coefficient of b^n is 1. We say B/A is an *integral extension* if all elements in B are integral over A.

Example 2.1. Take $A = \mathbf{Z}$ and $B = \mathcal{O}_K$ for some number field K. Then $\mathcal{O}_K/\mathbf{Z}$ is an integral extension.

Example 2.2. The ring extension $\mathbf{Z}[\sqrt{d}]/\mathbf{Z}$ is integral.

Example 2.3. The ring $A[\varepsilon]$, where $\varepsilon^2 = 0$ (this is defined rigorously as $A[X]/(X^2)$) is an integral extension of A called the ring of *dual numbers* over A. It is not a domain, even if A is.

The three equivalent characterizations of algebraic integers in Theorem 1.15 generalize to integral elements in any extension of commutative rings, but the proof will need a new idea.

Theorem 2.4. *When B/A is an extension of commutative rings and $b \in B$, the following are equivalent:*

(*a*) *b is integral over A;*

(*b*) *the ring $A[b]$ is a finitely generated A-module;*

(*c*) *there is a subring of B containing A and b that is a finitely generated A-module.*

Proof. (*a*) ⇒ (*b*): Say

$$b^n + a_{n-1}b^{n-1} + \cdots + a_1 b + a_0 = 0.$$

Then

$$b^n \in A + Ab + \cdots + Ab^{n-1}.$$

So

$$b^{n+1} \in Ab + Ab^2 + \cdots + Ab^n \subset A + Ab + \cdots + Ab^{n-1}.$$

By induction, $b^m \in A + Ab + \cdots + Ab^{n-1}$ for all $m \geqslant n$, so

$$A[b] = \sum_{i \geqslant 0} Ab^i = A + Ab + \cdots + Ab^{n-1}.$$

(*b*) ⇒ (*c*): Use $A[b]$.

(*c*) ⇒ (*a*): Say R is a ring where $A \subset R \subset B$ with $b \in R$, and

$$R = Ax_1 + Ax_2 + \cdots + Ax_n.$$

The x_i's are not all 0, since they span R, $1 \in R$, and $1 \neq 0$. (We're bypassing the trivial case that A is the zero ring, whose only ring extension is itself. The theorem is obvious in that case.) Multiplication by b sends R back to R, so

$$bx_i = a_{i1}x_1 + \cdots + a_{in}x_n, \qquad a_{ij} \in A.$$

We collect the equations over all i into a vector-matrix equation

$$\begin{pmatrix} bx_1 \\ bx_2 \\ \vdots \\ bx_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Thus

$$b\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

which implies

$$(bI_n - (a_{ij}))\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{2.1}$$

This equation is a generalization of (1.3). Since the vector $(x_1, \ldots, x_n)$ is not zero, (2.1) says the matrix $bI_n - (a_{ij}) \in \mathrm{M}_n(B)$ is not one-to-one on B^n. Looking back at the proof of Theorem 1.15, it is natural at this step to say the determinant of the matrix is 0. However, matters are a little subtle. While a square matrix over a field that is not injective as a linear map has determinant 0, over a general commutative ring this is no longer true: the determinant of a square matrix over a commutative ring that is not injective (that is, it kills a nonzero vector) need not be 0 (see Exercise 2.6 for an example). Is $\det(bI_n - (a_{ij}))$, which is a monic polynomial in b with coefficients in A, equal to 0? Yes, but we will need a result from linear algebra over rings to explain this.

In linear algebra, there is a "universal" formula for inverting a square matrix M: $M^{-1} = \frac{1}{\det M}\mathrm{adj}(M)$, where $\mathrm{adj}(M)$, called the classical adjoint of M, has (i, j) entry equal to the determinant of the matrix M with row j and column i removed, multiplied by $(-1)^{i+j}$. (For example, $\mathrm{adj}\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$.) This universal inverse formula doesn't always make sense, since the scalar $\det M$ might not be invertible, but $\mathrm{adj}(M)$ always makes sense since its entries are just polynomials in the entries of M. Multiplying through the inverse matrix formula by $\det M$ and by M gives the algebraic identity

$$\operatorname{adj}(M)M = (\det M)I_n.$$

This formula is valid for all square matrices over all commutative rings. It says we can multiply any square matrix M by a particular second matrix to produce a diagonal matrix with the determinant of M on the diagonal.

Let $D(b) = \det(bI_n - (a_{ij}))$ and multiply both sides of (2.1) on the left by $\operatorname{adj}(bI_n - (a_{ij}))$ to get

$$D(b)I_n \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Longrightarrow D(b) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Comparing coordinates,

$$D(b)x_i = 0 \text{ for all } i.$$

Since $R = Ax_1 + \cdots + Ax_n$,

$$D(b)r = 0 \text{ for all } r \in R.$$

Use $r = 1$: $D(b) = 0$. Since $D(b)$ is a monic polynomial in b with coefficients in A, we are done. ■

We call the set of all elements of B integral over A the *integral closure* of A in B. Using the third condition in Theorem 2.4, the integral closure is a ring. The proof is identical to our proof that $\mathcal{O}_K$ is a ring and is omitted.

Example 2.5. In a number field K, $\mathcal{O}_K$ is the integral closure of $\mathbf{Z}$.

Nonexample 2.6. An integral closure in a noncommutative ring need not be a ring. In $\mathrm{M}_2(\mathbf{Q})$, set $M_1 = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $M_2 = \left(\begin{smallmatrix} 0 & 0 \\ 1/2 & 0 \end{smallmatrix}\right)$. Since $M_1^2 - 2M_1 + I_2 = O$ and $M_2^2 = O$, M_1 and M_2 are integral over $\mathbf{Z}$. (We view $\mathbf{Z}$ inside $\mathrm{M}_2(\mathbf{Q})$ as integral diagonal matrices $\left(\begin{smallmatrix} a & 0 \\ 0 & a \end{smallmatrix}\right)$.) The sum $M_1 + M_2 = \left(\begin{smallmatrix} 1 & 1 \\ 1/2 & 1 \end{smallmatrix}\right)$ and product $M_1M_2 = \left(\begin{smallmatrix} 1/2 & 0 \\ 1/2 & 0 \end{smallmatrix}\right)$ have characteristic polynomials $T^2 - 2T + \frac{1}{2}$ and $T^2 - \frac{1}{2}T$, so they have eigenvalues which are not algebraic integers ($1 + \frac{1}{\sqrt{2}}$ for $M_1 + M_2$ and $\frac{1}{2}$ for M_1M_2). Any polynomial that vanishes at a square matrix has all eigenvalues of the matrix as roots, so no polynomial vanishing at $M_1 + M_2$ and

M_1M_2 can be monic in $\mathbf{Z}[T]$. Therefore $M_1 + M_2$ and M_1M_2 are not integral over $\mathbf{Z}$.

Integrality is useful in the noncommutative setting, but we will stick to commutative rings.

If A is a domain with fraction field F, we say A is *integrally closed* if the integral closure of A in F is A: all $x \in F$ integral over A are already in A.

Theorem 2.7. *Any* UFD *is integrally closed.*

Proof. Suppose A is a UFD and a/b is integral over A where $a, b \in A$ and $b \neq 0$:

$$\left(\frac{a}{b}\right)^n + c_{n-1}\left(\frac{a}{b}\right)^{n-1} + \cdots + c_1\frac{a}{b} + c_0 = 0$$

with $c_i \in A$. We want to show $a/b \in A$. Canceling the greatest common divisor of a and b in the fraction a/b (this is where we use the UFD property; the gcd is only defined up to unit multiple), we may assume a and b share no common factors except units. Clear the denominator by multiplying through by b^n:

$$a^n + c_{n-1}a^{n-1}b + \cdots + c_1ab^{n-1} + c_0b^n = 0.$$

Every term except a^n is divisible by b, so $b \mid a^n$. Since a and b are relatively prime in a UFD, the divisibility relation implies b is a unit in A, so $a/b \in A$. ■

Example 2.8. Any PID, such as $\mathbf{Z}$ or $\mathbf{Z}[i]$, is integrally closed.

Nonexample 2.9. A domain that is *not* integrally closed is $\mathbf{Z}[\sqrt{5}]$, since $\frac{1+\sqrt{5}}{2}$ is in the fraction field and is a root of $T^2 - T - 1 \in \mathbf{Z}[T] \subset \mathbf{Z}[\sqrt{5}][T]$ but it is not in $\mathbf{Z}[\sqrt{5}]$. (This shows that $\mathbf{Z}[\sqrt{5}]$ is not a UFD, because all UFDs are integrally closed. So we know $\mathbf{Z}[\sqrt{5}]$ is not a UFD without having an example of nonunique factorization in it!)

Nonexample 2.10. For any field L, the ring $A = L[T^2, T^3]$ is not integrally closed: its fraction field $L(T)$ contains T, and T is integral over A (being a root of $X^2 - T^2 \in A[X]$) but T is not in A.

Let's show the integral closure of the ring $A = L[T^2, T^3]$ in $L(T)$ is $L[T]$. Both A and $L[T]$ have fraction field $L(T)$, and $A \subset L[T]$. Thus any element of the integral closure of A in $L(T)$ is integral over $L[T]$, so it is in $L[T]$ since $L[T]$ (a UFD) is integrally closed. In the other direction, $L[T]$ is in the integral closure of A since X is integral over A.

Although "integral extension" means "algebraic extension" in the setting of fields (because all nonzero polynomials over a field can be scaled to monic polynomials), integrally closed does not mean algebraically closed for fields! Any field is integrally closed since it is its own fraction field, but being algebraically closed is a substantial property for a field. The point is that the meanings of "closed" in "integrally closed" and "algebraically closed" are not the same: integrally closed refers to what is happening in the fraction field and algebraically closed refers to what is happening in all algebraic extensions.

Theorem 2.11 (Transitivity). *If $A \subset B \subset C$ and B/A and C/B are integral, then C/A is integral.*

Proof. Pick $c \in C$. Using Theorem 2.4, we seek a subring of C containing c and A that is a finitely generated A-module. Since c is integral over B we can write

$$c^n + b_{n-1}c^{n-1} + \cdots + b_1 c + b_0 = 0, \qquad b_i \in B.$$

Consider the ring $R = A[b_0, \ldots, b_{n-1}, c] = A[b_0, \ldots, b_{n-1}][c]$. Using the equation of integral dependence of c over B above,

$$R = \sum_{i=0}^{n-1} A[b_0, \ldots, b_{n-1}]c^i. \tag{2.2}$$

Each of $b_0, \ldots, b_{n-1}$ is integral over A. Let b_k have an equation of integral dependence over A with degree d_k, so

$$A[b_0, \ldots, b_{n-1}] = \sum_{j_0 \leqslant d_0 - 1, \ldots, j_{n-1} \leqslant d_{n-1}-1} A b_0^{j_0} \cdots b_{n-1}^{j_{n-1}}.$$

This is a finitely generated A-module. Feeding this into (2.2) shows R is a finitely generated A-module. ■

Corollary 2.12. *If $A \subset B$, A' denotes the integral closure of A in B and $A'' = (A')'$ denotes the integral closure of A' in B, then $A'' = A'$.*

Proof. We have $A \subset A' \subset A'' \subset B$ with A'/A integral, and A''/A' integral. By Theorem 2.11, A''/A is integral, so $A'' \subset A'$. Thus $A'' = A'$. ■

This says taking integral closures of one ring in another doesn't produce anything new after it is done once.

Example 2.13. For any number field K, $\mathcal{O}_K$ is integrally closed. This follows from Corollary 2.12 using $A = \mathbf{Z}$, $B = K$, and $A' = \mathcal{O}_K$.

The next theorem generalizes the theorem that the minimal polynomial in $\mathbf{Q}[T]$ of an algebraic integer must be in $\mathbf{Z}[T]$ (Theorem 1.18). Notice the integrally closed hypothesis that occurs.

Theorem 2.14. *Let A be integrally closed with fraction field F and E/F be an algebraic extension of F. An element of E is integral over A if and only if its minimal polynomial over F is in $A[T]$.*

Proof. This is the same as the proof of Theorem 1.18, where we assumed the top field K is a finite extension of $\mathbf{Q}$, but all that proof needs is that K is an algebraic extension of $\mathbf{Q}$. Briefly, an element of E integral over A has F-conjugates that are integral over A, so the coefficients of its minimal polynomial in $F[T]$ are integral over A and lie in F, thus are in A since A is integrally closed. ■

Nonexample 2.15. Take $A = \mathbf{Z}[\sqrt{5}]$ and $F = \mathbf{Q}(\sqrt{5})$. The ring A is not integrally closed, and it provides a counterexample to Theorem 2.14 without the integrally closed hypothesis: $\frac{1+\sqrt{5}}{2}$ is integral over A since it is a root of $T^2 - T - 1$, but its minimal polynomial in $F[T]$ is $T - \frac{1+\sqrt{5}}{2} \not\in A[T]$.

If we drop the integrally closed condition in Theorem 2.14, and take A to be any domain with fraction field F, an element of E is integral over A if and only if its minimal polynomial in $F[T]$ is in $A'[T]$, where A' is the integral closure of A in F. Saying A is integrally closed is equivalent to $A' = A$.

By Theorem 2.7,

$$\text{PID} \Longrightarrow \text{UFD} \Longrightarrow \text{integrally closed domain.}$$

The converses are both *false*. For example, $\mathbf{Z}[T]$ is a UFD but not a PID; $(2, T)$ is a nonprincipal ideal in $\mathbf{Z}[T]$. And $\mathbf{Z}[\sqrt{-5}]$ is an integrally closed domain but not a UFD.

Remark 2.16. The purely algebraic notion of being integrally closed has a geometric interpretation. Let $f(X, Y)$ be irreducible in $\mathbf{C}[X, Y]$. Since $\mathbf{C}[X, Y]$ is a UFD, the ideal (f) in $\mathbf{C}[X, Y]$ is prime. It can be shown that the domain $\mathbf{C}[X, Y]/(f)$ is integrally closed if and only if the curve $f(X, Y) = 0$ in $\mathbf{C}^2$ is smooth (*i.e.*, there are no solutions to $f(x, y) = 0$ at which both partial derivatives of f vanish). Therefore being an integrally closed domain is an algebraic analogue of smoothness.

For instance, the curve $Y^2 = X^3$ has a singularity (a cusp) at the origin, as in Figure 2.1, so $\mathbf{C}[X,Y]/(Y^2 - X^3)$ is not integrally closed. In fact, this ring is isomorphic to $\mathbf{C}[T^2,T^3]$ if we map $X \mapsto T^2$ and $Y \mapsto T^3$, and we saw in Nonexample 2.10 that $\mathbf{C}[T^2,T^3]$ is not integrally closed and its integral closure in its fraction field is $\mathbf{C}[T] \cong \mathbf{C}[X,Y]/(Y - X)$. The curve $Y = X$ is smooth.

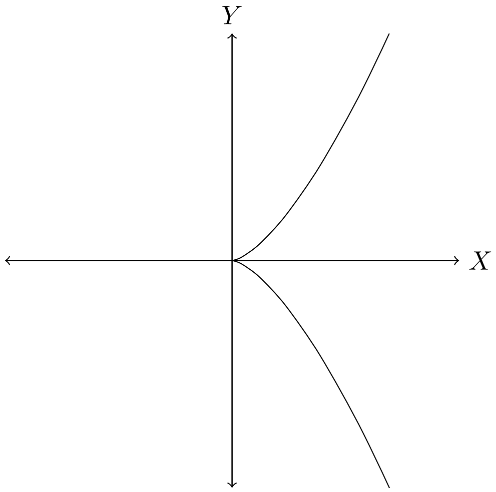


Figure 2.1: The real points on the curve $Y^2 = X^3$.

The algebraic process of enlarging a nonintegrally closed domain to its integral closure in its fraction field, like enlarging $\mathbf{C}[T^2,T^3]$ to $\mathbf{C}[T]$ (or, apparently less geometrically, $\mathbf{Z}[\sqrt{5}]$ to $\mathbf{Z}[\frac{1+\sqrt{5}}{2}]$), can be regarded as an algebraic analogue of "resolving singularities" or "smoothing" a singular curve.

2.2 Ideals in Integral Extensions of a Domain

We will prove a few theorems about ideals in an integral extension B/A where both A and B are domains. The key property is that for nonzero $b \in B$, the constant term of any "minimal" equation of integral dependence must be nonzero: if

$$b^n + a_{n-1}b^{n-1} + \cdots + a_1 b + a_0 = 0, \qquad a_i \in A,$$

with $n \geqslant 1$ minimal and $a_0 = 0$ then $b(b^{n-1} + a_{n-1}b^{n-2} + \cdots + a_1) = 0$, which implies $b^{n-1} + a_{n-1}b^{n-2} + \cdots + a_1 = 0$, violating the minimality of n.

Theorem 2.17. *Let B/A be an integral extension of domains. If $\mathfrak{b}$ is a nonzero ideal in B then $\mathfrak{b} \cap A \neq (0)$.*

Proof. Let $b \in \mathfrak{b}$ with $b \neq 0$. Let

$$b^n + a_{n-1}b^{n-1} + \cdots a_1 b + a_0 = 0, \qquad a_i \in A,$$

with $n \geqslant 1$ minimal. Then $a_0 \neq 0$ and $a_0 \in bB \subset \mathfrak{b}$, so $\mathfrak{b} \cap A \neq (0)$. ■

Example 2.18. In $\mathbf{Z}[\sqrt{d}]$ with an ideal $\mathfrak{b} \neq (0)$, choose $\beta \in \mathfrak{b}$ with $\beta \neq 0$. Then $\mathrm{N}(\beta) \neq 0$ and $\mathrm{N}(\beta) = \beta\overline{\beta} \in \mathfrak{b} \cap \mathbf{Z}$.

Nonexample 2.19. Consider $\mathbf{Z} \subset \mathbf{Z}[\varepsilon]$, where $\varepsilon^2 = 0$. This is an integral ring extension, but the top ring is not a domain so we can't expect Theorem 2.17 to apply and in fact it breaks down: $(\varepsilon) \cap \mathbf{Z} = \{0\}$.

Theorem 2.20. *Let $A \subset B$ be an integral extension of domains.*

(*a*) *A is a field if and only if B is a field.*

(*b*) *If every nonzero prime ideal in A is maximal, this also holds in B.*

Proof. (a) Let A be a field and pick a nonzero $b \in B$. Choose an equation of integral dependence for b over A, say

$$b^n + a_{n-1}b^{n-1} + \cdots + a_1 b + a_0 = 0, \qquad a_i \in A,$$

with n minimal. Then

$$b(b^{n-1} + a_{n-1}b^{n-2} + \cdots + a_1) = -a_0$$

and $a_0 \neq 0$. Let a_0' be the inverse of a_0 in A. Then multiplying through by $-a_0'$,

$$b(b^{n-1} + a_{n-1}b^{n-2} + \cdots + a_1)(-a_0') = 1,$$

so $b \in B^\times$.

Now suppose B is a field. Let a be a nonzero element of A. It has an inverse $b \in B$; $ab = 1$. Since B/A is integral, we have

$$b^n + a_{n-1}b^{n-1} + a_{n-2}b^{n-2} + \cdots + a_1 b + a_0 = 0, \qquad a_i \in A.$$

Multiply by a^{n-1}, so we have

$$b + a_{n-1} + a_{n-2}a + \cdots + a_1 a^{n-2} + a_0 a^{n-1} = 0.$$

Every term on the left side other than b is in A, so $b \in A$.

(b) Let $\mathfrak{P}$ be a nonzero prime ideal in B,[1] so $B/\mathfrak{P}$ is a domain. We want to show $\mathfrak{P}$ is maximal. Consider

$$\mathfrak{p} = \mathfrak{P} \cap A = \ker(A \xrightarrow{\text{inclusion}} B \xrightarrow{\text{reduction}} B/\mathfrak{P}) \neq (0).$$

Since there is a natural embedding $A/\mathfrak{p} \hookrightarrow B/\mathfrak{P}$ and $B/\mathfrak{P}$ is a domain, $A/\mathfrak{p}$ is a domain. Then $\mathfrak{p}$ is a nonzero prime ideal in A, so $\mathfrak{p}$ is a maximal ideal in A by assumption. Since B is integral over A, $B/\mathfrak{P}$ is integral over $A/\mathfrak{p}$. Now we have an integral extension of domains $A/\mathfrak{p} \subset B/\mathfrak{P}$ where the bottom ring is a field. Then $B/\mathfrak{P}$ is a field by part a, so $\mathfrak{P}$ is maximal. ■

Nonexample 2.21. Note that $\mathbf{Z} \subset \mathbf{Q}$ and $\mathbf{Q}$ is a field but $\mathbf{Z}$ is not. This doesn't violate the theorem since $\mathbf{Q}$ is not an integral extension of $\mathbf{Z}$.

Nonexample 2.22. The ring $\mathbf{Q}[\varepsilon]$ with $\varepsilon^2 = 0$ is an integral extension of $\mathbf{Q}$ and is not a field. This doesn't violate the theorem since $\mathbf{Q}[\varepsilon]$ is not a domain.

Nonexample 2.23. Consider $\mathbf{Z} \subset \mathbf{Z}[\varepsilon]$, where $\varepsilon^2 = 0$. Then

$$\mathbf{Z}[\varepsilon]/(\varepsilon) \cong (\mathbf{Z}[X]/(X^2))/((X)/(X^2)) \cong \mathbf{Z},$$

so (ε) is a nonzero prime ideal in $\mathbf{Z}[\varepsilon]$ that is not maximal. This doesn't violate the theorem since $\mathbf{Z}[\varepsilon]$ is not a domain.

Example 2.24. For any number field K, all nonzero prime ideals in $\mathcal{O}_K$ are maximal since it's true in $\mathbf{Z}$ and $\mathcal{O}_K$ is a domain and an integral extension of $\mathbf{Z}$. For the same reason, all nonzero prime ideals in $\mathbf{Z}[\sqrt{d}]$ are maximal, whether or not $\mathbf{Z}[\sqrt{d}]$ is the full ring of algebraic integers in $\mathbf{Q}(\sqrt{d})$.

How do we know there are any nonzero prime ideals in $\mathcal{O}_K$? We can't simply use ideals $p\mathcal{O}_K$ where p is a prime number, since this often is *not* a prime ideal! For instance, $2\mathbf{Z}[i]$ is not a prime ideal in $\mathbf{Z}[i]$ since $2 = -i(1+i)^2$.[2] It is a general theorem that every nonzero commutative ring has a maximal ideal (see,

[1] Such an ideal does exist in B: Zorn's lemma implies B has a maximal ideal, which is therefore a prime ideal, and (0) is not maximal in B since B is not a field by part a.

[2] In $\mathbf{Z}[i]$ there are prime ideals of the form $p\mathbf{Z}[i]$, such as $3\mathbf{Z}[i]$, but there are number fields K such that $p\mathcal{O}_K$ is not a prime ideal for any prime number p, such as $\mathbf{Q}(\sqrt{2}, \sqrt{3})$.

for instance, [2, pp. 3–4]), so $\mathcal{O}_K$ has maximal ideals and maximal ideals are prime ideals. We will see a more concrete proof that $\mathcal{O}_K$ has maximal ideals (and therefore prime ideals) in Corollary 3.5.

The two parts of Theorem 2.20 can be put into a common setting using the concept of dimension for commutative rings. The *Krull dimension* of a nonzero commutative ring A is the largest $n \geqslant 0$ such that there is an ascending tower of prime ideals

$$\mathfrak{p}_0 \subsetneqq \mathfrak{p}_1 \subsetneqq \cdots \subsetneqq \mathfrak{p}_n$$

in A, and we set $\dim A = \infty$ if such chains of prime ideals can be arbitrarily long.[3]

Example 2.25. A field has dimension 0.

Example 2.26. For a prime p and integer $r \geqslant 1$, the only prime ideal in $\mathbf{Z}/p^r\mathbf{Z}$ is $(p) = p\mathbf{Z}/p^r\mathbf{Z}$, so $\mathbf{Z}/p^r\mathbf{Z}$ has dimension 0 (and is not a field or even a domain when $r > 1$).

Example 2.27. The dimension of $\mathbf{Z}$ is 1, using a tower of prime ideals $(0) \subset (p)$. More generally, any PID, such as $\mathbf{Q}[T]$, has dimension 1.

Example 2.28. Any $\mathcal{O}_K$ has dimension 1 since any nonzero prime ideal in $\mathcal{O}_K$ is maximal and since maximal ideals in $\mathcal{O}_K$ exist (Example 2.24). More generally, the integral closure of any PID in an algebraic extension of its fraction field has dimension 1.

Example 2.29. The dimension of $\mathbf{Z}[T]$ is 2. An example of a tower of prime ideals in $\mathbf{Z}[T]$ is $(0) \subset (T) \subset (2, T)$.

With this new terminology, a *domain* has dimension 0 if and only if it is a field and a *domain* has dimension 1 if and only if every nonzero prime ideal is maximal. Theorem 2.20 says for an integral extension of domains B/A that $\dim A = 0$ if and only if $\dim B = 0$, and if $\dim A = 1$ then $\dim B = 1$. It is true more generally that for any integral extension of rings B/A (neither one necessarily a domain), the Krull dimension is preserved: $\dim B = \dim A$. This is one of the most basic properties of integral extensions, although we will not need such a broad result (a proof is in [33, p. 47]). The special case in Theorem 2.20 is adequate.

[3] The polynomial ring $\mathbf{Q}[T_1, T_2, T_3, \ldots]$ has infinite dimension due to $(0) \subset (T_1) \subset (T_1, T_2) \subset (T_1, T_2, T_3) \subset \ldots$.

To illustrate the usefulness of Theorem 2.20 beyond rings of interest in number theory, here is an application to polynomial rings.

Corollary 2.30. *The maximal ideals in* $\mathbf{C}[X,Y]$ *are of the form* $(X-\alpha, Y-\beta)$ *for some* α *and* β *in* $\mathbf{C}$.

Proof. Any ideal $(X-\alpha, Y-\beta)$ is maximal, since $\mathbf{C}[X,Y]/(X-\alpha,Y-\beta) \cong \mathbf{C}$.

To show any maximal ideal $\mathfrak{m}$ has this special form, the idea is to show $\mathfrak{m} \cap \mathbf{C}[X] \neq \{0\}$, thereby reducing us to knowledge of the maximal ideals in $\mathbf{C}[X]$, which all are some $(X-\alpha)$.

Write $x = X \bmod \mathfrak{m}$ and $y = Y \bmod \mathfrak{m}$, so $\mathbf{C}[X,Y]/\mathfrak{m} = \mathbf{C}[x,y]$. Pick a nonzero $f(X,Y)$ in $\mathfrak{m}$, so $f(x,y) = 0$. This is a nontrivial algebraic relation between x and y over $\mathbf{C}$. If $f(X,Y)$ is monic as a polynomial in Y, then the equation $f(x,y) = 0$ shows y is integral over $\mathbf{C}[x]$, so $\mathbf{C}[x,y]$ is an integral extension of $\mathbf{C}[x]$. In fact, if $f(X,Y)$ as a polynomial in Y has a leading coefficient in $\mathbf{C}^\times$ then y is still integral over $\mathbf{C}[x]$. Since $\mathbf{C}[x,y]$ is a field, Theorem 2.20 tells us $\mathbf{C}[x]$ must be a field. Since $\mathbf{C}[x] \cong \mathbf{C}[X]/(g(X))$ where $g(X)$ is monic and $g(x) = 0$, $\mathbf{C}[x]$ being a field forces $g(X) = X - \alpha$ for some $\alpha \in \mathbf{C}$. Then $x - \alpha = 0$ in $\mathbf{C}[x]$, so $X - \alpha \in \mathfrak{m}$. In a similar way, $Y - \beta \in \mathfrak{m}$ for some $\beta \in \mathbf{C}$, so $(X-\alpha, Y-\beta) \subset \mathfrak{m}$. The ideal $(X-\alpha,Y-\beta)$ is maximal so the inclusion is an equality.

This argument assumed $f(X,Y)$ as a polynomial in Y is monic, or more generally has a leading coefficient in $\mathbf{C}^\times$. If the leading coefficient is nonconstant in $\mathbf{C}[X]$, we can make it constant with a linear change of variables in X. Set $X = X' + cY$, for $c \in \mathbf{C}$ to be determined. Let N be the degree of $f(X,Y)$ (maximal degree of a nonzero monomial in it), so $f(X,Y) = \sum_{d=0}^{N}\sum_{i=0}^{d} a_{i,d-i}X^iY^{d-i}$, where some $a_{i,N-i}$ is nonzero. Then

$$\begin{aligned} f(X'+cY,Y) &= \sum_{d=0}^{N}\sum_{i=0}^{d} a_{i,d-i}(X'+cY)^iY^{d-i} \\ &= \left(\sum_{i=0}^{N} a_{i,N-i}c^i\right)Y^N + \text{ lower degree terms in } Y. \end{aligned}$$

To see what's happening, if $f = XY^3 - 5X^2Y^2 + 4X^2Y + 1$ then

$$\begin{aligned} f(X,Y) &= (X'+cY)Y^3 - 5(X'+cY)^2Y^2 + 4(X'+cY)^2Y + 1 \\ &= (c-5c^2)Y^4 + ((1-10c)X' + 4c^2)Y^3 + \text{ lower degree in } Y. \end{aligned}$$

Provided $c \neq 0$ or $1/5$, the coefficient of Y^4 is in $\mathbf{C}^\times$. (If we are careful we can choose c to make the leading coefficient 1, but that is not crucial.)

Back in the general case, the coefficient of Y^N in $f(X' + cY, Y)$ is the value at c of some nonzero polynomial (with degree at most N). Since $\mathbf{C}$ is an infinite field, we can choose c so that $f(X' + cY, Y)$ as a polynomial in Y has leading coefficient in $\mathbf{C}^\times$. Reducing mod $\mathfrak{m}$, y is integral over $\mathbf{C}[x']$, so $\mathbf{C}[x, y] = \mathbf{C}[x', y]$ is integral over $\mathbf{C}[x']$. Therefore there are α and β in $\mathbf{C}$ such that $\mathfrak{m} = (X' - \alpha, Y - \beta) = (X - cY - \alpha, Y - \beta) = (X - (c\beta + \alpha), Y - \beta)$. ■

The linear change of variables argument in the proof is not an *ad hoc* trick; it is taken from a proof of the Noether normalization lemma [33, p. 49], which is an important foundational result in commutative algebra and algebraic geometry.

2.3 Trace and Norm

Associated to any finite extension of fields E/F are two canonical maps $E \to F$, the trace $\mathrm{Tr}_{E/F}$ and the norm $\mathrm{N}_{E/F}$. These are the trace and determinant of the multiplication maps $m_\alpha \colon E \to E$ for $\alpha \in E$, and are reviewed in Section 8.1 along with the characteristic polynomial $\chi_{E/F,\alpha}(T) \in F[T]$. Read that section, particularly the examples, if you are not familiar with these terms. Writing out the characteristic polynomial, we can find the trace and norm among its coefficients, up to sign:

$$\chi_{E/F,\alpha}(T) = T^n - \mathrm{Tr}_{E/F}(\alpha)T^{n-1} + \cdots + (-1)^n \, \mathrm{N}_{E/F}(\alpha).$$

We will discuss here how the trace, norm, and characteristic polynomial generalize from fields to integrally closed domains. The general diagram to keep in mind is the following one. The ring A is an integrally closed domain with fraction field F, E/F is a finite field extension, and B is the integral closure of A in E.

$$\begin{tikzcd}
 & E \arrow[dd, dash, "n"] \\
B \arrow[ur, dash] \arrow[dd, dash] & \\
 & F \\
A \arrow[ur, dash] &
\end{tikzcd}$$

Theorem 2.31. *With notation as above, an element of E is integral over A if and only if its characteristic polynomial is in $A[T]$.*

Proof. Pick $x \in E$. By Theorem 8.12, $\chi_{E/F,x}(T) = f_x(T)^{n/d}$ where $f_x(T)$ is the minimal polynomial of x in $F[T]$ and $d = [F(x) : F]$. If x is integral over A then $f_x(T) \in A[T]$ by Theorem 2.14, so $\chi_{E/F,x}(T) \in A[T]$. Conversely, if $\chi_{E/F,x}(T) \in A[T]$ then x is integral over A since $\chi_{E/F,x}(T)$ is monic. ■

In particular, for a number field K, a number α in K is an algebraic integer if and only if the *characteristic* polynomial $\chi_{K/\mathbf{Q},\alpha}(T)$ has coefficients in $\mathbf{Z}$. We saw earlier (Theorem 1.2) that $\alpha \in \mathcal{O}_K$ if and only if its *minimal* polynomial over $\mathbf{Q}$ has coefficients in $\mathbf{Z}$, but in practice the characteristic polynomial is easier to compute than the minimal polynomial since we don't have to worry about checking irreducibility.

The following theorem tells us about the trace and norm on an integral closure.

Theorem 2.32. *With notation as above, let $b \in B$, so b is integral over A.*

(*a*) $\mathrm{Tr}_{E/F}(b) \in A$ *and* $\mathrm{N}_{E/F}(b) \in A$.

(*b*) $\mathrm{N}_{E/F}(b) = b\widetilde{b}$ *for some* $\widetilde{b} \in A[b]$. *In particular,*

$$b \in B^\times \Longleftrightarrow \mathrm{N}_{E/F}(b) \in A^\times,$$

in which case $b^{-1} \in A[b]$. *Thus* $B^\times \cap A[b] = A[b]^\times$ *and* $B^\times \cap F = A^\times$.

Proof. (a): By Theorem 2.31, $\chi_{E/F,b}(T) \in A[T]$, and the trace and norm of b are, up to sign, coefficients of this polynomial.

(b): Let $\chi_{E/F,b}(T) = T^n + c_{n-1}T^{n-1} + \cdots + c_1T + c_0 \in A[T]$. Then

$$0 = \chi_{E/F,b}(b) = b^n + c_{n-1}b^{n-1} + \cdots + c_1b + c_0,$$

where $c_0 = \pm\,\mathrm{N}_{E/F}(b)$. So

$$\begin{aligned}\mathrm{N}_{E/F}(b) &= \pm(b^n + c_{n-1}b^{n-1} + \cdots + c_1b)\\ &= b\widetilde{b} \text{ for some } \widetilde{b} \in A[b].\end{aligned}$$

If $b \in B^\times$, let $bb' = 1$ in B. Then $\mathrm{N}_{E/F}(b)\,\mathrm{N}_{E/F}(b') = \mathrm{N}_{E/F}(1) = 1$ in A. Thus $\mathrm{N}_{E/F}(b) \in A^\times$ by part a. If $\mathrm{N}_{E/F}(b) \in A^\times$, let $\mathrm{N}_{E/F}(b)a' = 1$. Then $b\widetilde{b}a' = 1$, so $b \in B^\times$ and $b^{-1} = \widetilde{b}a' \in A[b]$ so $b \in A[b]^\times \subset B^\times$.

If $b \in B^\times \cap F$ then $\mathrm{N}_{E/F}(b) = b^n \in A^\times$. Since $b \in F$ and A is integrally closed, we get $b \in A$, so $b \in A^\times$. ■

Theorem 2.33. *For any number field K,*

$$\mathcal{O}_K^\times = \left\{\alpha \in \mathcal{O}_K : \mathrm{N}_{K/\mathbf{Q}}(\alpha) = \pm 1\right\}.$$

In particular, any $a \in \mathbf{Z} - \{0\}$ besides ± 1 is not a unit in $\mathcal{O}_K$.

Proof. By Theorem 2.32, $\mathcal{O}_K^\times = \left\{\alpha \in \mathcal{O}_K : \mathrm{N}_{K/\mathbf{Q}}(\alpha) \in \mathbf{Z}^\times\right\}$. For $a \in \mathbf{Z} - \{0\}$, $\mathrm{N}_{K/\mathbf{Q}}(a) = a^{[K:\mathbf{Q}]} = \pm 1$ only when $a = \pm 1$. ∎

Nonexample 2.34. Although $\frac{3}{5} + \frac{4}{5}i \in \mathbf{Q}(i)$ has norm 1 down to $\mathbf{Q}$, it is not in $\mathbf{Z}[i]^\times$. This doesn't contradict Theorem 2.33 since $\frac{3}{5} + \frac{4}{5}i$ is not an algebraic integer. More generally, the units of $\mathcal{O}_K$ are *not* the α in K satisfying $\mathrm{N}_{K/\mathbf{Q}}(\alpha) = \pm 1$ but the α in $\mathcal{O}_K$ satisfying that condition.

Example 2.35. Let d be an integer which is not a perfect cube. For rational x, y, and z,

$$\mathrm{N}_{\mathbf{Q}(\sqrt[3]{d})/\mathbf{Q}}(x + y\sqrt[3]{d} + z\sqrt[3]{d^2}) = x^3 + dy^3 + d^2z^3 - 3dxyz. \tag{2.3}$$

Therefore Theorem 2.33 tells us the units in $\mathbf{Z}[\sqrt[3]{d}]$ correspond to integral solutions (x, y, z) of

$$x^3 + dy^3 + d^2z^3 - 3dxyz = \pm 1.$$

Whether or not $\mathbf{Z}[\sqrt[3]{d}]$ is the full ring of integers of $\mathbf{Q}(\sqrt[3]{d})$, Theorem 2.32b tells us a number in $\mathbf{Z}[\sqrt[3]{d}]$ that is invertible as an algebraic integer has its inverse in $\mathbf{Z}[\sqrt[3]{d}]$, so the concepts "invertible in $\mathbf{Z}[\sqrt[3]{d}]$" and "number in $\mathbf{Z}[\sqrt[3]{d}]$ that is invertible in $\mathcal{O}_{\mathbf{Q}(\sqrt[3]{d})}$" mean the same thing. Taking $d = 2$, we discover the unit $1 + \sqrt[3]{2} + \sqrt[3]{4}$ in $\mathbf{Z}[\sqrt[3]{2}]$. Its inverse is $-1 + \sqrt[3]{2}$ (factor $2 - 1 = \sqrt[3]{2}^3 - 1$).

2.4 Exercises

1. If $f(T) \in \mathbf{Q}(T)$ and $f(T)^n \in \mathbf{Z}[T]$ for some $n \geqslant 1$, show $f(T) \in \mathbf{Z}[T]$.

2. (Continuation of Exercise 1.8)

   a) Inside the product ring $\mathbf{Q}^2 = \mathbf{Q} \times \mathbf{Q}$, show the integral closure of $\mathbf{Z}$ is $\mathbf{Z}^2 = \mathbf{Z} \times \mathbf{Z}$.

   b) For each integer $c \geqslant 1$, show the unique subring of $\mathbf{Z}^2$ with index c is

$$\mathbf{Z} + c\mathbf{Z}^2 = \mathbf{Z} + \mathbf{Z}c(0,1) = \{(a, b) \in \mathbf{Z}^2 : a \equiv b \bmod c\}.$$

Here we view $\mathbf{Z}$ inside $\mathbf{Z}^2$ as a subring: $\mathbf{Z} = \mathbf{Z}(1,1)$.

c) Show any subring of $\mathbf{Z}^2$ other than $\mathbf{Z}$ has finite index in $\mathbf{Z}^2$, so part b describes all subrings of $\mathbf{Z}^2$ other than $\mathbf{Z}$.

3. a) Let A be a (nonzero) commutative ring. Prove there is unique division with remainder in $A[T]$ by monic polynomials: if $f(T)$ and $g(T)$ are in $A[T]$ with $g(T)$ monic, there are unique $q(T)$ and $r(T)$ in $A[T]$ such that $f(T) = g(T)q(T) + r(T)$ with $r(T) = 0$ or $\deg r < \deg g$. (If $g(T)$ is not monic this might not be possible: we can't divide T^2 by $2T + 1$ in $\mathbf{Z}[T]$, for instance.)

   b) Suppose A is a domain with fraction field F. Let f and g be in $A[T]$ with f monic. If $f \mid g$ in $F[T]$, use part a to show $f \mid g$ in $A[T]$. Then give an example of $f(T)$ and $g(T)$ in $\mathbf{Z}[T]$ where f is not monic, $f \mid g$ in $\mathbf{Q}[T]$ and $f \nmid g$ in $\mathbf{Z}[T]$.

   c) Suppose A is a domain with fraction field F. If $f(T) \in A[T]$ is monic and irreducible in $F[T]$, with root α in some extension field, use part b to show the evaluation map $A[T] \to A[\alpha]$ that sends T to α induces a ring isomorphism $A[T]/(f(T)) \cong A[\alpha]$. (For example, $\mathbf{Z}[i] \cong \mathbf{Z}[T]/(T^2 + 1)$. Not all ideals in $\mathbf{Z}[T]$ are principal, which makes this isomorphism slightly more interesting than the isomorphism $\mathbf{Q}[i] \cong \mathbf{Q}[T]/(T^2 + 1)$.)

4. Let A be integrally closed with fraction field F.

   a) Prove an analogue of Gauss's lemma from $\mathbf{Z}[T]$: if $f(T)$ is *monic* in $A[T]$ and $f(T) = g(T)h(T)$ in $F[T]$ then there are constants $a, b \in F^\times$ such that $f(T) = (ag(T))(bh(T))$ where $ag(T)$ and $bh(T)$ are in $A[T]$. (It is not necessary that $g(T)$ and $h(T)$ are monic, only that they are in $F[T]$. And it is generally not the case that $f(T)$ splits into linear factors in $F[T]$.)

   b) Show a *monic* $f(T) \in A[T]$ is irreducible in $A[T]$ if and only if it is irreducible in $F[T]$, in which case the ideal $fA[T]$ in $A[T]$ is prime. (Hint: The natural ring homomorphism $A[T] \to F[T]/fF[T]$ has kernel containing $fA[T]$. Show the kernel is exactly $fA[T]$ using the previous exercise.)

   c) Show $A[T]$ is integrally closed. (Hint: The rings $A[T]$ and $F[T]$ have the same fraction field and $F[T]$ is integrally closed since it is a UFD, so reduce to showing an element of $F[T]$ integral over $A[T]$ is in $A[T]$. If

$f(T) \in F[T]$ is integral over $A[T]$ and $H(X) \in A[T][X]$ is monic in X with $H(f(T)) = 0$, show $H(X + T^d)$ has a constant term that is monic in $A[T]$ for $d \gg 0$. Use Gauss's lemma as in part a to conclude that $T^d - f(T)$ is monic in $A[T]$, so $f(T) \in A[T]$.)

d) Show $T^2 - T - 1$ is irreducible in $\mathbf{Z}[\sqrt{5}][T]$, is reducible in $\mathbf{Q}(\sqrt{5})[T]$, and the ideal $(T^2 - T - 1)$ in $\mathbf{Z}[\sqrt{5}][T]$ is not prime. Therefore parts a and b break down for $A = \mathbf{Z}[\sqrt{5}]$, which is not integrally closed.

e) A polynomial $f(T) = a_nT^n + a_{n-1}T^{n-1} + \cdots + a_1T + a_0 \in A[T]$ is called *primitive* when its coefficients generate the unit ideal: $(a_0, a_1, \ldots, a_n) = (1)$. Any monic polynomial is primitive, so primitive polynomials generalize monic polynomials. Show $f(T) = (1 + \sqrt{-5})T^2 - \sqrt{-5}T + (1 + \sqrt{-5})$ is primitive and irreducible in $\mathbf{Z}[\sqrt{-5}][T]$, is reducible in $\mathbf{Q}(\sqrt{-5})[T]$, and the ideal $(f(T))$ in $\mathbf{Z}[\sqrt{-5}][T]$ is not prime. Therefore parts a and b are not true for primitive polynomials in place of monic polynomials when $A = \mathbf{Z}[\sqrt{-5}]$, which is integrally closed. (When A is a PID, parts a and b are true for primitive polynomials in place of monic polynomials.)

5. Let A be a domain and $\mathfrak{p}$ be a prime ideal in A. A monic polynomial

$$T^n + c_{n-1}T^{n-1} + \cdots + c_1T + c_0 \in A[T]$$

is called *Eisenstein* at $\mathfrak{p}$ if $c_i \in \mathfrak{p}$ for all i and $c_0 \notin \mathfrak{p}^2$. (More generally, Eisenstein polynomials are allowed a leading coefficient not in $\mathfrak{p}$, but we take them to be monic for simplicity.)

a) Show an Eisenstein polynomial in $A[T]$ is irreducible in $A[T]$.

b) If A is integrally closed with fraction field F, show an Eisenstein polynomial in $A[T]$ is irreducible in $F[T]$. (See the previous exercise.)

c) In $A = \mathbf{Z}[2i]$, whose fraction field is $\mathbf{Q}(i)$, show the ideal $\mathfrak{p} = (2, 2i)$ in A is prime and $T^2 - 2i \in A[T]$ is Eisenstein at $\mathfrak{p}$ but is reducible in $\mathbf{Q}(i)[T]$. (Note $\mathbf{Z}[2i]$ is not integrally closed, so this doesn't violate part b.)

6. In a ring A, suppose $ab = 0$ with $a \neq 0$ and $b \neq 0$. Construct a 2×2 matrix over A which is not injective as a linear map $A^2 \to A^2$ but its determinant is a rather than 0.

7. Show any finite-dimensional algebra over a field has Krull dimension 0: every prime ideal is maximal. (Hint: Theorem 2.20.)

8. Let $f(X,Y) = 3X^4Y^2 - 11X^5Y - 4X^6 - 9X^2 + 1729$. Find all $c \in \mathbf{C}$ such that $f(X+cY, Y)$ is *not* monic as a polynomial in Y. (The answer is a finite set.)

9. Verify (2.3) and find a unit $\neq \pm 1$ in $\mathbf{Z}[\sqrt[3]{3}]$ and in $\mathbf{Z}[\sqrt[3]{5}]$. (Hint: There are answers with all small coefficients if you allow some coefficients to be negative or 0.)

10. (Continuation of Exercise 1.14) Let F be a field not of characteristic 2.

    a) Show every quadratic extension of $F(X)$ has the form $F(X, \sqrt{f(X)})$ where $f(X) \in F[X]$ and $f(X)$ is squarefree.[4]

    b) If $a \in F^\times$ and a is not a perfect square, show the integral closure of $F[X]$ in $F(X, \sqrt{a})$ is $F(\sqrt{a})[X] = F[X] + F[X]\sqrt{a}$.

    c) For the rest of this exercise, suppose $f(X) \in F[X]$ is *nonconstant* and squarefree. Show the integral closure of $F[X]$ in $F(X, \sqrt{f(X)})$ is $F[X, \sqrt{f(X)}] = F[X] + F[X]\sqrt{f(X)}$. This is analogous to knowing the ring of integers in a quadratic field, as in the diagram below. For example, the ring $\mathbf{C}[X, \sqrt{X^3 - X}]$ is integrally closed.

$$\begin{tikzcd}
 & K \arrow[dd, dash, "2"] \\
\mathcal{O}_K \arrow[ur, dash] \arrow[dd, dash] & \\
 & \mathbf{Q} \\
\mathbf{Z} \arrow[ur, dash] &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & F(X, \sqrt{f(X)}) \arrow[dd, dash, "2"] \\
F[X, \sqrt{f(X)}] \arrow[ur, dash] \arrow[dd, dash] & \\
 & F(X) \\
F[X] \arrow[ur, dash] &
\end{tikzcd}$$

    d) Let $\pi(X)$ be a monic irreducible factor of $f(X)$ in $F[X]$. (They exist since $f(X)$ is nonconstant.) In $F[X, \sqrt{f(X)}]$, set $\mathfrak{p}_\pi = (\pi(X), \sqrt{f(X)})$. Show this is a maximal ideal in $F[X, \sqrt{f(X)}]$ and $\mathfrak{p}_\pi^2 = (\pi(X))$.

    e) Let $\overline{F}$ be an algebraic closure of F. For any $\alpha, \beta \in \overline{F}$ such that $\beta^2 = f(\alpha)$ (that is, the point (α, β) lies on the curve $y^2 = f(x)$ and has

[4] Squarefree in $F[X]$ means no irreducible factor appears more than once. It has nothing to do with *constant* factors that may be squares: $4X(X+1)$ is considered squarefree in $F[X]$.

coordinates algebraic over F), let $\mathfrak{p}_{(\alpha,\beta)}$ be the elements of $F[X, \sqrt{f(X)}]$ vanishing at (α, β):

$$\left\{a(X) + b(X)\sqrt{f(X)} : a(X), b(X) \in F[X] \text{ and } a(\alpha) + b(\alpha)\beta = 0\right\}.$$

Show $\mathfrak{p}_{(\alpha,\beta)}$ is a maximal ideal in $F[X, \sqrt{f(X)}]$ and the ideal $\mathfrak{p}_\pi$ in part d has the form $\mathfrak{p}_{(\alpha,\beta)}$ for some $(\alpha, \beta) \in \overline{F}^2$ on the curve $y^2 = f(x)$.

11. This exercise is the characteristic 2 analogue of the previous one. Square roots (roots of $T^2 - c$) are replaced by roots of $T^2 - T - c$.

    If k has characteristic 2 and $T^2 - T - c$ is irreducible in $k[T]$, with root α in an extension of k, the second root is $\alpha + 1$ and $k(\alpha)/k$ is Galois. Artin–Schreier theory says every quadratic *Galois* extension of k arises in this way, and that moreover $k(\alpha)$ determines c up to addition by an element of the form $b^2 - b$ where $b \in k$. (That is, $k(\alpha)$ is a splitting field over k of $T^2 - T - c'$ if and only if $c' = c + b^2 - b$ for some $b \in k$. The "if" direction is easy since $\alpha + b$ is a root.)

    Let F be a field of characteristic 2, and choose $f(X) \in F(X)$ such that $T^2 - T - f(X)$ is irreducible over $F(X)$. Let α be a root: $\alpha^2 - \alpha - f(X) = 0$.

    a) If $f(X) \in F[X]$, show the integral closure of $F[X]$ in $F(X, \alpha)$ is $F[X][\alpha] = F[X] + F[X]\alpha$. (Hint: Since $F[X]$ is integrally closed, Theorem 2.31 can be used.)

    b) Let $K = F(X, \alpha)$ where $\alpha^2 - \alpha - X^3 = 0$. Show $[K : F(X)] = 2$ and the integral closure of $F[X]$ in K is $F[X][\alpha]$.

    c) If $F(X, \alpha) = F(X, \beta)$ where $\alpha^2 - \alpha - f(X) = 0$ with $f(X) \in F(X)$ and $\beta^2 - \beta - g(X) = 0$ for some *polynomial* $g(X) \in F[X]$, show every irreducible factor of the denominator of $f(X)$ has even multiplicity. This is a nontrivial constraint, for instance it is not true when $f(X) = 1/X$, so it shows we can't generally change α to make $f(X)$ a polynomial. (Hint: Write $g(X) = f(X) + h(X)^2 - h(X)$ for some $h(X) \in F(X)$.)

    d) Let $K_d = F(X, \alpha_d)$, where $\alpha_d^2 - \alpha_d - 1/X^d = 0$ for an odd positive integer d. Although α_d is not integral over $F[X]$, show $\beta_d := X^{(d+1)/2}\alpha_d$ is. Does the integral closure of $F[X]$ in K_d equal $F[X] + F[X]\beta_d$? If you find this too hard, at least try the special case $d = 1$.

12. Let A be a ring. A sequence $a_1, a_2, a_3, \ldots$ in A satisfies an r-term linear

recursion when there are constants $c_1, c_2, \dots, c_r$ in A such that

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_r a_{n-r}$$

for all $n > r$. Without mentioning r, the sequence is called linearly recursive.

The set of all sequences in A obviously is a ring under termwise addition and multiplication. Remarkably, the linearly recursive sequences in A are a subring. This will be proved using ideas similar to the proof that integral elements over a ring are closed under addition and multiplication (Theorem 2.4).

a) Show the squares of the Fibonacci numbers, F_n^2, satisfy the linear recursion $F_n^2 = 2F_{n-1}^2 + 2F_{n-2}^2 - F_{n-3}^2$. More generally, if a sequence $a_1, a_2, a_3, \dots$ satisfies the 2-term linear recursion $a_n = a_{n-1} + a_{n-2}$, show the sequence $b_n := a_n^2$ satisfies the 3-term linear recursion $b_n = 2b_{n-1} + 2b_{n-2} - b_{n-3}$. (Admittedly the recursion for F_n^2 is coming out of nowhere.)

b) Let Seq(A) denote the set of all sequences in A. We think of them as infinite-length vectors: $\mathbf{a} = (a_1, a_2, a_3, \dots)$. Define the shift operator $S\colon \mathrm{Seq}(A) \to \mathrm{Seq}(A)$ by $S(a_1, a_2, a_3, \dots) = (a_2, a_3, a_4, \dots)$. It drops the first term and moves all other terms back by one position. Show S is a ring homomorphism.

c) Show $(S^2 - S - I)(\mathcal{F}) = \mathbf{0}$, where $\mathcal{F} = (1, 1, 2, 3, 5, \dots)$ is the Fibonacci sequence, and more generally a sequence $\mathbf{a}$ in A satisfies a linear recursion if and only if $f(S)(\mathbf{a}) = \mathbf{0}$ for some *monic* polynomial $f(T) \in A[T]$. This is the link between linear recursions and integrality.

d) Use part c to show the sum of two linearly recursive sequences is linearly recursive.

e) A subset M of Seq(A) is called shift-stable if $S(M) \subset M$. Show the following conditions on a sequence $\mathbf{a}$ in Seq(A) are equivalent:

1. $\mathbf{a}$ is linearly recursive,
2. the A-module $\sum_{n \geqslant 0} AS^n(\mathbf{a}) = A\mathbf{a} + AS(\mathbf{a}) + AS^2(\mathbf{a}) + \cdots$ is finitely generated and shift-stable,
3. $\mathbf{a}$ is contained in a finitely generated shift-stable A-submodule of Seq(A).

As in the proof of Theorem 2.4, the hardest part is going from the last condition to the first one.

f) Deduce that the product of two linearly recursive sequences is linearly recursive and use your solution to derive the recursion in part a for F_n^2 in a systematic way.

CHAPTER 3

BASES AND DISCRIMINANTS

Our main goal in this chapter is to show $\mathcal{O}_K$ has a **Z**-basis and see how a **Z**-basis can be computed using discriminants, which are an algebraic type of volume. There are several kinds of discriminants: for a basis, for a lattice, and for a polynomial. Knowing how different concepts of discriminant are related is particularly important for computational work.

3.1 The Integers of $\mathbf{Q}(\sqrt[3]{2})$.

It is not generally true that the ring of integers of a number field $\mathbf{Q}(\sqrt[n]{a})$ is $\mathbf{Z}[\sqrt[n]{a}]$. In the quadratic case, for instance, $\mathbf{Z}[\sqrt{d}]$ is not always the full ring of integers in $\mathbf{Q}(\sqrt{d})$ (squarefree d). For the first cubic field that comes to mind, though, what you think is the ring of integers is in fact the right answer.

Theorem 3.1. *The ring of integers of* $\mathbf{Q}(\sqrt[3]{2})$ *is* $\mathbf{Z}[\sqrt[3]{2}]$.

Proof. Set $K = \mathbf{Q}(\sqrt[3]{2})$. Obviously $\mathbf{Z}[\sqrt[3]{2}] \subset \mathcal{O}_K$. Using $\{1, \sqrt[3]{2}, \sqrt[3]{4}\}$ as a **Q**-basis of K, any $\alpha \in K$ can be written as

$$\alpha = x + y\sqrt[3]{2} + z\sqrt[3]{4}$$

for rational x, y, z. The question is: if $\alpha \in \mathcal{O}_K$, are $x, y, z \in \mathbf{Z}$?

First we gain control over the denominators of the coefficients: if $\alpha \in \mathcal{O}_K$, then $x, y, z \in \frac{1}{3}\mathbf{Z}$. We will prove this by a lot of trace computations. We begin with the trace of α:

$$\begin{aligned}\mathrm{Tr}_{K/\mathbf{Q}}(\alpha) &= \mathrm{Tr}_{K/\mathbf{Q}}(x + y\sqrt[3]{2} + \sqrt[3]{4}) \\ &= x\,\mathrm{Tr}_{K/\mathbf{Q}}(1) + y\,\mathrm{Tr}_{K/\mathbf{Q}}(\sqrt[3]{2}) + z\,\mathrm{Tr}_{K/\mathbf{Q}}(\sqrt[3]{4}).\end{aligned}$$

The characteristic polynomials of $\sqrt[3]{2}$ and $\sqrt[3]{4}$ for the extension $K/\mathbf{Q}$ are $T^3 - 2$ and $T^3 - 4$, where T^2 has coefficient 0, so $\mathrm{Tr}_{K/\mathbf{Q}}(\sqrt[3]{2})$ and $\mathrm{Tr}_{K/\mathbf{Q}}(\sqrt[3]{4})$ are both 0. The trace of 1 is $[K : \mathbf{Q}] = 3$, so

$$\mathrm{Tr}_{K/\mathbf{Q}}(\alpha) = \mathrm{Tr}_{K/\mathbf{Q}}(x + y\sqrt[3]{2} + z\sqrt[3]{4}) = 3x.$$

Since $\alpha \in \mathcal{O}_K$, $\mathrm{Tr}_{K/\mathbf{Q}}(\alpha) \in \mathbf{Z}$ (Theorem 2.32 with $A = \mathbf{Z}$), so $3x \in \mathbf{Z}$.

Similarly, from

$$\alpha\sqrt[3]{2} = 2z + x\sqrt[3]{2} + y\sqrt[3]{4} \in \mathcal{O}_K,$$

we get $\mathrm{Tr}_{K/\mathbf{Q}}(\alpha\sqrt[3]{2}) = 6z \in \mathbf{Z}$ and from

$$\alpha\sqrt[3]{4} = 2y + 2z\sqrt[3]{2} + x\sqrt[3]{4} \in \mathcal{O}_K$$

we get $\mathrm{Tr}_{K/\mathbf{Q}}(\alpha\sqrt[3]{4}) = 6y \in \mathbf{Z}$. We have $3x, 6y, 6z \in \mathbf{Z}$ but we want $3x, 3y, 3z \in \mathbf{Z}$. The product

$$3\alpha = 3x + 3y\sqrt[3]{2} + 3z\sqrt[3]{4} \tag{3.1}$$

is in $\mathcal{O}_K$ and $3x \in \mathbf{Z}$, so (since $\mathcal{O}_K$ is a ring)

$$3y\sqrt[3]{2} + 3z\sqrt[3]{4} \in \mathcal{O}_K.$$

Multiply by $\sqrt[3]{2}$:

$$3y\sqrt[3]{4} + 6z \in \mathcal{O}_K,$$

so $3y\sqrt[3]{4} \in \mathcal{O}_K$ since $6z \in \mathbf{Z}$. Taking norms yields $(3y)^3 \cdot 4 \in \mathbf{Z}$. Rewrite this as $\frac{(6y)^3}{2} \in \mathbf{Z}$. This implies that $(6y)^3$ is even, which means $6y$ is even (we already know $6y \in \mathbf{Z}$), so $3y \in \mathbf{Z}$. By (3.1), $3z\sqrt[3]{4} \in \mathcal{O}_K$, so $3z \in \mathbf{Z}$ by the same argument used to show $3y \in \mathbf{Z}$.

From $3x, 3y, 3z \in \mathbf{Z}$ we have $3\mathcal{O}_K \subset \mathbf{Z}[\sqrt[3]{2}]$, so

$$\mathbf{Z}[\sqrt[3]{2}] \subset \mathcal{O}_K \subset \frac{1}{3}\mathbf{Z}[\sqrt[3]{2}].$$

Having placed $\mathcal{O}_K$ between $\mathbf{Z}[\sqrt[3]{2}]$ and $\frac{1}{3}\mathbf{Z}[\sqrt[3]{2}]$, we want to make a list of coset representatives for $\frac{1}{3}\mathbf{Z}[\sqrt[3]{2}]/\mathbf{Z}[\sqrt[3]{2}]$ and see which representatives are in $\mathcal{O}_K$. The main issue here is coming to grips with what $\frac{1}{3}\mathbf{Z}[\sqrt[3]{2}]/\mathbf{Z}[\sqrt[3]{2}]$ means.

All algebra students are comfortable with cosets in $\mathbf{Z}/m\mathbf{Z}$, but $\frac{1}{m}\mathbf{Z}/\mathbf{Z}$ might look strange. It isn't, as long as you think only additively: $\mathbf{Z}$ is an additive group, $\frac{1}{m}\mathbf{Z}$ is a larger additive group (inside $\mathbf{Q}$), and $\frac{1}{m}\mathbf{Z}/\mathbf{Z}$ is an additive quotient group. It is the fractions having denominator m, considered up to addition by integers. What $\frac{1}{m}\mathbf{Z}/\mathbf{Z}$ is not is a ring: it makes no sense to multiply in $\frac{1}{m}\mathbf{Z}/\mathbf{Z}$ if $m > 1$. So only have your additive hat on when you are contemplating $\frac{1}{m}\mathbf{Z}/\mathbf{Z}$. If it makes you more comfortable, scaling by m turns $\frac{1}{m}\mathbf{Z}$ into $\mathbf{Z}$ and $\mathbf{Z}$ into $m\mathbf{Z}$, so $\frac{1}{m}\mathbf{Z}/\mathbf{Z} \cong \mathbf{Z}/m\mathbf{Z}$ as additive groups. Standard coset representatives in $\mathbf{Z}/m\mathbf{Z}$ are $0, 1, 2, \ldots, m-1$, so dividing by m gives us coset representatives for $\frac{1}{m}\mathbf{Z}/\mathbf{Z}$: $0, \frac{1}{m}, \frac{2}{m}, \ldots, \frac{m-1}{m}$. This makes sense: fractions with denominator m all equal some integer plus $\frac{i}{m}$, where $0 \leqslant i \leqslant m-1$.

Returning to $\frac{1}{3}\mathbf{Z}[\sqrt[3]{2}]/\mathbf{Z}[\sqrt[3]{2}]$, coset representatives are

$$\frac{a}{3} + \frac{b}{3}\sqrt[3]{2} + \frac{c}{3}\sqrt[3]{4}, \qquad \text{where } 0 \leqslant a, b, c \leqslant 2. \tag{3.2}$$

Every $\alpha \in \frac{1}{3}\mathbf{Z}[\sqrt[3]{2}]$ is one of these coset representatives plus an element of $\mathbf{Z}[\sqrt[3]{2}]$. Since $\mathbf{Z}[\sqrt[3]{2}] \subset \mathcal{O}_K$, saying α is in $\mathcal{O}_K$ is the same as saying its coset representative from (3.2) is in $\mathcal{O}_K$. The numbers in (3.2) besides the one where $a = b = c = 0$ are *not* algebraic integers since, by computations, their characteristic polynomials are not in $\mathbf{Z}[T]$ (Theorem 2.31). In fact, using the formula

$$\mathrm{N}_{\mathbf{Q}(\sqrt[3]{2})/\mathbf{Q}}(x + y\sqrt[3]{2} + z\sqrt[3]{4}) = x^3 + 2y^3 + 4z^3 - 6xyz$$

with rational x, y, and z, the norms of the nonzero coset representatives in (3.2) are not in $\mathbf{Z}$ and that needs fewer computations than the characteristic polynomials. Either way, we are done. ■

3.2 Z-basis of $\mathcal{O}_K$

In our computation of the ring of integers of $\mathbf{Q}(\sqrt[3]{2})$, we picked the $\mathbf{Q}$-basis $\{1, \sqrt[3]{2}, \sqrt[3]{4}\}$ of $\mathbf{Q}(\sqrt[3]{2})$ and for an algebraic integer α in the field the traces of α, $\alpha\sqrt[3]{2}$, and $\alpha\sqrt[3]{4}$ told us something about the coefficients of α in that basis. This idea of multiplying by a basis and forming the traces will now be used to prove every $\mathcal{O}_K$ has a $\mathbf{Z}$-basis.

Theorem 3.2. *If $[K:\mathbf{Q}]=n$, then $\mathcal{O}_K$ has a $\mathbf{Z}$-basis of rank n: there are $e_1,\dots,e_n$ in $\mathcal{O}_K$ such that $\mathcal{O}_K=\bigoplus_{i=1}^n \mathbf{Z}e_i$.*

Proof. Pick a $\mathbf{Q}$-basis $\{\alpha_1,\dots,\alpha_n\}$ of K. Every algebraic number is an algebraic integer divided by an ordinary integer (Theorem 1.13), so after scaling the terms of our basis by suitable nonzero integers, we may suppose $\alpha_1,\dots,\alpha_n\in\mathcal{O}_K$. Then

$$\sum_{i=1}^n \mathbf{Z}\alpha_i=\bigoplus_{i=1}^n \mathbf{Z}\alpha_i\subset\mathcal{O}_K,$$

so $\mathcal{O}_K$ contains a subgroup that is free of rank n. There is definitely no reason to expect $\sum_{i=1}^n \mathbf{Z}\alpha_i=\mathcal{O}_K$. Using linear algebra, we will use the free abelian group of rank n contained in $\mathcal{O}_K$ to find a free abelian group of rank n containing $\mathcal{O}_K$.

For any $x\in\mathcal{O}_K$, write it in terms of the basis for $K/\mathbf{Q}$:

$$x=c_1\alpha_1+\cdots+c_n\alpha_n,\qquad c_i\in\mathbf{Q}.$$

Most likely the c_i's are not integers. But we will find, from x being in $\mathcal{O}_K$, that the c_i's have a universal common denominator as fractions, independent of the choice of x in $\mathcal{O}_K$. Multiply x by each basis member α_i:

$$\alpha_i x=c_1\alpha_i\alpha_1+\cdots+c_n\alpha_i\alpha_n,$$

and now take the trace down to $\mathbf{Q}$ of both sides to get

$$\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i x)=c_1\,\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i\alpha_1)+\cdots+c_n\,\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i\alpha_n).$$

We can write these equations for all i as a single matrix equation:

$$\begin{pmatrix}\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_1 x)\\ \vdots\\ \mathrm{Tr}_{K/\mathbf{Q}}(\alpha_n x)\end{pmatrix}=\begin{pmatrix}\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_1\alpha_1) & \cdots & \mathrm{Tr}_{K/\mathbf{Q}}(\alpha_1\alpha_n)\\ \vdots & \ddots & \vdots\\ \mathrm{Tr}_{K/\mathbf{Q}}(\alpha_n\alpha_1) & \cdots & \mathrm{Tr}_{K/\mathbf{Q}}(\alpha_n\alpha_n)\end{pmatrix}\begin{pmatrix}c_1\\ \vdots\\ c_n\end{pmatrix}.$$

Since $x\in\mathcal{O}_K$, each $\alpha_i x$ is in $\mathcal{O}_K$ so its trace is in $\mathbf{Z}$: the vector on the left side is in $\mathbf{Z}^n$. Each $\alpha_i\alpha_j$ is an algebraic integer, so the matrix on the right side is in $\mathrm{M}_n(\mathbf{Z})$. Notice the matrix depends solely on the basis, not on x. By Cramer's rule, we can write $c_i=a_i/d$, where $a_i\in\mathbf{Z}$ and

$$d=\det(\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i\alpha_j))\in\mathbf{Z}. \tag{3.3}$$

Applying Cramer's rule requires $d \neq 0$. Could the determinant d be 0?

Saying $d = 0$ is the same saying the matrix $(\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i\alpha_j))$ is not invertible in $\mathrm{M}_n(\mathbf{Q})$, which is the same as saying its null space is not just the vector $\mathbf{0}$: there exist $b_1, \dots, b_n$ in $\mathbf{Q}$, not all zero, such that

$$\begin{pmatrix} \mathrm{Tr}_{K/\mathbf{Q}}(\alpha_1\alpha_1) & \cdots & \mathrm{Tr}_{K/\mathbf{Q}}(\alpha_1\alpha_n) \\ \vdots & \ddots & \vdots \\ \mathrm{Tr}_{K/\mathbf{Q}}(\alpha_n\alpha_1) & \cdots & \mathrm{Tr}_{K/\mathbf{Q}}(\alpha_n\alpha_n) \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

This is true if and only if

$$b_1 \,\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i\alpha_1) + \cdots + b_n \,\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i\alpha_n) = 0$$

for all i, which is equivalent to

$$\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i(b_1\alpha_1 + \cdots + b_n\alpha_n)) = 0 \tag{3.4}$$

for all i. Since the $\mathbf{Q}$-span of the α_i's is K, (3.4) for all i is the same as

$$\mathrm{Tr}_{K/\mathbf{Q}}(\alpha(b_1\alpha_1 + \cdots + b_n\alpha_n)) = 0 \text{ for all } \alpha \in K. \tag{3.5}$$

The number $b_1\alpha_1 + \cdots + b_n\alpha_n$ is not 0 since some b_i is not 0 and the α_i's are a basis, so the K-multiples of this number fill up K. Therefore (3.5) says $\mathrm{Tr}_{K/\mathbf{Q}}$, as a function from K to $\mathbf{Q}$, is identically 0. But that is absurd: $\mathrm{Tr}_{K/\mathbf{Q}}(1) = [K : \mathbf{Q}] \neq 0$. So d in (3.3) is not 0. Returning to our general number x in $\mathcal{O}_K$, we have

$$x = \frac{1}{d}(a_1\alpha_1 + \cdots + a_n\alpha_n) \Longrightarrow \mathcal{O}_K \subset \mathbf{Z}\frac{\alpha_1}{d} + \cdots + \mathbf{Z}\frac{\alpha_n}{d},$$

so $d = \det(\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i\alpha_j))$ is a common denominator for *all* algebraic integers of K when expressed in the basis $\{\alpha_1, \dots, \alpha_n\}$:

$$\boxed{\bigoplus_{i=1}^{n} \mathbf{Z}\alpha_i \subset \mathcal{O}_K \subset \bigoplus_{i=1}^{n} \mathbf{Z}\frac{\alpha_i}{d}, \quad d = \det(\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i\alpha_j)).} \tag{3.6}$$

From the structure of subgroups of finite free abelian groups (specifically, Corollary 8.29), any abelian group that both contains and is contained in abelian groups that are free of rank n is itself free of rank n, so $\mathcal{O}_K \cong \mathbf{Z}^n$ as abelian groups, which means $\mathcal{O}_K$ has a **Z**-basis of size n. ■

Remember that this proof that $\mathcal{O}_K$ has a **Z**-basis is indirect. We squeeze $\mathcal{O}_K$ between two free **Z**-modules of rank n and general theorems then tell us $\mathcal{O}_K$ is a free **Z**-module of rank n. Practical ways to find a **Z**-basis are discussed in Section 3.5.

While the primitive element theorem from field theory implies $K = \mathbf{Q}(\alpha) = \mathbf{Q}[\alpha]$ for some α, it is *not* true that every $\mathcal{O}_K$ has a power basis $\{1, \alpha, \ldots, \alpha^{n-1}\}$ over **Z**. That is, $\mathcal{O}_K$ need not have the form $\mathbf{Z}[\alpha]$, although we saw it does when K is a quadratic field and when $K = \mathbf{Q}(\sqrt[3]{2})$. We will see an example of a number field whose ring of integers has no power basis over **Z** in Section 4.5.

In PARI, the command to find a **Z**-basis of a number field is `nfbasis`. For example, to find a **Z**-basis of $\mathbf{Q}(\sqrt[3]{2})$, typing `nfbasis(x^3 - 2)` returns the answer `[1,x,x^2]`, which means $\{1, \sqrt[3]{2}, \sqrt[3]{4}\}$ is a **Z**-basis. Here are a few more examples.

- `nfbasis(x^3+5*x+2): [1, x, 1/2*x^2 + 1/2*x]`,
- `nfbasis(x^4+8*x+12): [1, x, 1/2*x^2, 1/4*x^3 + 1/2*x]`,
- `nfbasis(x^5+9*x+9): [1, x, x^2, 1/3*x^3, 1/3*x^4]`.

These examples and more like them naturally suggest the following result.

Corollary 3.3. *There is a* **Z***-basis of* $\mathcal{O}_K$ *which includes the number* 1*.*

Proof. Let $\{e_1, \ldots, e_n\}$ be a **Z**-basis of $\mathcal{O}_K$. Write $1 = c_1e_1 + \cdots + c_ne_n$ with $c_i \in \mathbf{Z}$. We can say almost nothing about the c_i's, since we don't know what the e_i's are. But because this combination is equal to 1, the integers $c_1, \ldots, c_n$ are relatively prime as an n-tuple: if $c \in \mathbf{Z}$ is a common factor of the c_i's then we can factor c out of each coefficient to see $c \in \mathcal{O}_K^\times$. Then $1/c \in \mathcal{O}_K \cap \mathbf{Q} = \mathbf{Z}$, so $c = \pm 1$.

From the structure theorem for finite free **Z**-modules, a linear combination of a basis having coefficients that are relatively prime is itself part of some (other) basis. Applying this result to 1 shows 1 belongs to a **Z**-basis of $\mathcal{O}_K$.

For completeness, we give the deduction from the structure theorem. Suppose $\{v_1, \ldots, v_n\}$ is a basis of a finite free **Z**-module M and $v := c_1v_1 + \cdots + c_nv_n$ with $(c_1, \ldots, c_n) = 1$. By the relative primality, we can write $a_1c_1 + \cdots + a_nc_n = 1$ for some $a_i \in \mathbf{Z}$. Define $\varphi\colon M \to \mathbf{Z}$ by $\varphi(x_1v_1 + \cdots + x_nv_n) = a_1x_1 + \cdots + a_nx_n$, so φ is **Z**-linear and $\varphi(v) = 1$. Then $M = \mathbf{Z}v + \ker\varphi$: for any $w \in M$, write

$\varphi(w) = b$, so $\varphi(bv) = b = \varphi(w)$. Thus $w = bv + (w - bv)$ with $w - bv \in \ker \varphi$. The sum decomposition of M is direct since φ is injective on $\mathbf{Z}v$ and identically 0 on $\ker \varphi$. As a submodule of a finite free $\mathbf{Z}$-module, $\ker \varphi$ is finite free. A basis of $\ker \varphi$ together with v gives us a basis of M. ■

3.3 The Ideal Norm

The finiteness of $\mathbf{Z}/m\mathbf{Z}$ is important in number theory, since many of the basic elementary arithmetic functions count something about it, *e.g.*, $\varphi(m)$ is the number of its units and $d(m)$ (the number of positive divisors of m) is the number of its subgroups. The finiteness of $\mathbf{Z}/m\mathbf{Z}$ extends to residue rings of $\mathcal{O}_K$, as a consequence of $\mathcal{O}_K$ having a (finite) $\mathbf{Z}$-basis.

Theorem 3.4. *For any nonzero ideal $\mathfrak{a}$ in $\mathcal{O}_K$, $\mathcal{O}_K/\mathfrak{a}$ is finite and $\mathfrak{a} \cong \mathbf{Z}^n$ as abelian groups.*

Proof. Since $\mathfrak{a}$ is nonzero, $\mathfrak{a} \cap \mathbf{Z} \neq \{0\}$ (Theorem 2.17). Pick a nonzero $a \in \mathfrak{a} \cap \mathbf{Z}$. Write $\mathcal{O}_K = \mathbf{Z}e_1 \oplus \cdots \oplus \mathbf{Z}e_n$. Then $a\mathcal{O}_K \subset \mathfrak{a}$ and

$$a\mathcal{O}_K = \mathbf{Z}ae_1 \oplus \cdots \oplus \mathbf{Z}ae_n,$$

so $a\mathcal{O}_K \cong \mathbf{Z}^n$ as additive groups. In $\mathcal{O}_K/a\mathcal{O}_K$, the coefficients of $\overline{e}_1, \ldots, \overline{e}_n$ only matter modulo a, so $\mathcal{O}_K/a\mathcal{O}_K \cong (\mathbf{Z}/a\mathbf{Z})^n$ as additive groups and $\#(\mathcal{O}_K/a\mathcal{O}_K) = |a|^n$.

Since $a\mathcal{O}_K \subset \mathfrak{a} \subset \mathcal{O}_K$, and $\mathcal{O}_K/a\mathcal{O}_K$ is finite, $\mathcal{O}_K/\mathfrak{a}$ is finite. The ideal $\mathfrak{a}$ lies between two finite free $\mathbf{Z}$-modules of rank n, $\mathcal{O}_K$ and $a\mathcal{O}_K$, so $\mathfrak{a}$ is also a finite free $\mathbf{Z}$-module of rank n (Corollary 8.29). ■

Corollary 3.5. *Every proper ideal in $\mathcal{O}_K$ is contained in a maximal ideal of $\mathcal{O}_K$. In particular, there are maximal ideals in $\mathcal{O}_K$.*

Proof. An example of a proper nonzero ideal in $\mathcal{O}_K$ is a principal ideal $\alpha\mathcal{O}_K$ where α is any nonzero nonunit in $\mathcal{O}_K$, like an integer besides 0 and ± 1 (Theorem 2.33).

Let $\mathfrak{a}$ be any proper nonzero ideal in $\mathcal{O}_K$. The ideal $\mathfrak{a}$ has finite index greater than 1 in $\mathcal{O}_K$, so there are only finitely many proper ideals of $\mathcal{O}_K$ which contain $\mathfrak{a}$. (Ideals between $\mathfrak{a}$ and $\mathcal{O}_K$ correspond naturally to ideals in $\mathcal{O}_K/\mathfrak{a}$, which is a finite ring.) One of these finitely many ideals, say $\mathfrak{m}$, is contained in none of the others. Any ideal of $\mathcal{O}_K$ containing $\mathfrak{m}$ also contains $\mathfrak{a}$, so by the definition of $\mathfrak{m}$ a proper ideal of $\mathcal{O}_K$ containing $\mathfrak{m}$ has to be $\mathfrak{m}$. Thus $\mathfrak{m}$ is a maximal ideal in $\mathcal{O}_K$. ■

Corollary 3.6. *Every nonzero prime ideal in $\mathcal{O}_K$ is a maximal ideal.*

This is just like in $\mathbf{Z}$, where the nonzero prime ideals $p\mathbf{Z}$ are maximal ideals.

Proof. We already saw this before (Example 2.24), but now we get a different proof. When $\mathfrak{p}$ is a nonzero prime ideal in $\mathcal{O}_K$, $\mathcal{O}_K/\mathfrak{p}$ is a finite domain. Any finite domain is a field: if D is a finite domain and $a \neq 0$ in D then the function $f(x) = ax$ on D is injective and therefore surjective, so $ax = 1$ for some x. Thus $\mathcal{O}_K/\mathfrak{p}$ is a field, so $\mathfrak{p}$ is maximal. ■

Definition 3.7. For a nonzero ideal $\mathfrak{a}$ in $\mathcal{O}_K$, we call $\#(\mathcal{O}_K/\mathfrak{a})$ the *ideal norm* of $\mathfrak{a}$ and write $\mathrm{N}(\mathfrak{a}) = \#(\mathcal{O}_K/\mathfrak{a})$.

Example 3.8. If $n = [K : \mathbf{Q}]$ and $a \in \mathbf{Z} - \{0\}$, then $\mathrm{N}(a\mathcal{O}_K) = |a|^n$ from the proof of Theorem 3.4.

Example 3.9. From the end of Example 1.46, the ideal $\mathfrak{p} = (2, 1 + \sqrt{-5})$ in $\mathbf{Z}[\sqrt{-5}]$ has index 2, so $\mathrm{N}(\mathfrak{p}) = 2$.

To compute the norm of an ideal, which is the index of the ideal in the ring of integers, we can use the following determinant formula for the index of one finite free $\mathbf{Z}$-module in another.

Theorem 3.10. *Let V be a $\mathbf{Q}$-vector space of dimension n, M be a finite free $\mathbf{Z}$-module in V with rank n, and M' be a submodule of M with rank n. The index $[M : M']$ equals $|\det \varphi|$, where $\varphi \colon V \to V$ is any $\mathbf{Q}$-linear map such that $\varphi(M) = M'$.*

Proof. Our proof will be an application of aligned bases in a finite free abelian group and finite-index subgroup (Theorem 8.33).

Choose a $\mathbf{Z}$-basis of M, say

$$M = \mathbf{Z}x_1 \oplus \cdots \oplus \mathbf{Z}x_n.$$

Then

$$M' = \varphi(M) = \mathbf{Z}\varphi(x_1) \oplus \cdots \oplus \mathbf{Z}\varphi(x_n).$$

These two direct sum decompositions are usually *not* aligned with each other: although $M' \subset M$, we usually won't have $\mathbf{Z}\varphi(x_i) \subset \mathbf{Z}x_i$. By Theorem 8.33, there is a set of aligned bases for M and M' over $\mathbf{Z}$:

$$M = \mathbf{Z}e_1 \oplus \cdots \oplus \mathbf{Z}e_n, \quad M' = \mathbf{Z}a_1e_1 \oplus \cdots \oplus \mathbf{Z}a_ne_n,$$

where the a_i's are nonzero integers. So

$$M/M' \cong \bigoplus_{i=1}^{n} \mathbf{Z}/a_i\mathbf{Z},$$

which tells us $[M : M'] = |a_1a_2\cdots a_n|$. Any $\mathbf{Z}$-linearly independent set in V is $\mathbf{Q}$-linearly independent, so all four sets $\{x_i\}$, $\{\varphi(x_i)\}$, $\{e_i\}$ and $\{a_ie_i\}$ are $\mathbf{Q}$-bases for V. For any two $\mathbf{Q}$-bases of V there is a $\mathbf{Q}$-linear map $V \to V$ taking one basis to the other. The $\mathbf{Q}$-linear map $V \to V$ taking x_i to $\varphi(x_i)$ is φ. Consider the diagram of $\mathbf{Q}$-linear maps

$$\begin{tikzcd}
V \arrow[r, "\varphi"] \arrow[d, "f"'] & V & & x_i \arrow[r, mapsto, "\varphi"] \arrow[d, mapsto, "f"'] & \alpha x_i \\
 & & \text{where} & & \\
V \arrow[r, "g"'] & V \arrow[u, "h"'] & & e_i \arrow[r, mapsto, "g"'] & a_ie_i \arrow[u, mapsto, "h"']
\end{tikzcd}$$

This diagram commutes: $\varphi = h \circ g \circ f$. Taking determinants of these $\mathbf{Q}$-linear maps $V \to V$,

$$\det(\varphi) = \det(h)\det(g)\det(f). \tag{3.7}$$

Using the $\mathbf{Q}$-basis $\{e_i\}$ of V, the matrix $[g]$ is diagonal with a_i's along its main diagonal, so

$$\det(g\colon V \to V) = a_1a_2\cdots a_n.$$

What about $\det(f)$? The map f identifies two $\mathbf{Z}$-bases of the *same* $\mathbf{Z}$-module $M = \mathbf{Z}x_1 \oplus \cdots \oplus \mathbf{Z}x_n = \mathbf{Z}e_1 \oplus \cdots \oplus \mathbf{Z}e_n$:

$$f(c_1x_1 + \cdots + c_nx_n) = c_1e_1 + \cdots + c_ne_n, \qquad c_i \in \mathbf{Z}.$$

Therefore f is invertible as a $\mathbf{Z}$-linear map of M to itself. Using $\{x_i\}$ as the basis in which a matrix for f is computed, the matrices of $f\colon V \to V$ over $\mathbf{Q}$ and $f\colon M \to M$ over $\mathbf{Z}$ are the same. Since an invertible $\mathbf{Z}$-linear map has determinant ± 1,

$$\det(f\colon V \to V) = \det(f\colon M \to M) = \pm 1.$$

A similar argument shows

$$\det(h\colon V \to V) = \det(h\colon M' \to M') = \pm 1$$

since h identifies two $\mathbf{Z}$-bases of the free $\mathbf{Z}$-module M'. Feeding these determinant formulas into the right side of (3.7),

$$\det(\varphi) = \pm a_1 a_2 \cdots a_n.$$

Thus $|\det(\varphi)| = |a_1 a_2 \cdots a_n| = [M : M']$. ■

Example 3.11. Let $K = \mathbf{Q}(\sqrt{10})$. Let $\mathfrak{a} = (2 + 5\sqrt{10}, 4 + 7\sqrt{10})$, an ideal in $\mathcal{O}_K = \mathbf{Z}[\sqrt{10}]$. We will compute $\mathrm{N}(\mathfrak{a})$ by finding $\mathbf{Z}$-bases for $\mathcal{O}_K$ and $\mathfrak{a}$ and then computing the determinant of the matrix expressing the second basis in terms of the first.

A $\mathbf{Z}$-basis for $\mathcal{O}_K$ is $\{1, \sqrt{10}\}$. A $\mathbf{Z}$-basis for $\mathfrak{a}$ is $\{2 + 5\sqrt{10}, 4 + 7\sqrt{10}\}$, but this requires verification because it is a stronger condition to generate $\mathfrak{a}$ as a $\mathbf{Z}$-module than to generate it as an ideal (see Exercise 1.28): the initial definition of $\mathfrak{a}$ tells us that

$$\begin{aligned}\mathfrak{a} &= \mathbf{Z}[\sqrt{10}](2 + 5\sqrt{10}) + \mathbf{Z}[\sqrt{10}](4 + 7\sqrt{10}) \\ &= (\mathbf{Z} + \mathbf{Z}\sqrt{10})(2 + 5\sqrt{10}) + (\mathbf{Z} + \mathbf{Z}\sqrt{10})(4 + 7\sqrt{10}) \\ &= \mathbf{Z}(2 + 5\sqrt{10}) + \mathbf{Z}(50 + 2\sqrt{10}) + \mathbf{Z}(4 + 7\sqrt{10}) + \mathbf{Z}(70 + 4\sqrt{10}).\end{aligned}$$

For $\mathfrak{a}$ to be spanned over $\mathbf{Z}$ by $2 + 5\sqrt{10}$ and $4 + 7\sqrt{10}$, we need to write the other two $\mathbf{Z}$-module generators in terms of them. After some linear algebra, we can do this:

$$\begin{aligned}50 + 2\sqrt{10} &= -57(2 + 5\sqrt{10}) + 41(4 + 7\sqrt{10}), \\ 70 + 4\sqrt{10} &= -79(2 + 5\sqrt{10}) + 57(4 + 7\sqrt{10}).\end{aligned}$$

Therefore $\mathfrak{a} = \mathbf{Z}(2 + 5\sqrt{10}) + \mathbf{Z}(4 + 7\sqrt{10})$ and the numbers $2 + 5\sqrt{10}$ and $4 + 7\sqrt{10}$ are obviously $\mathbf{Z}$-linearly independent, so they are a $\mathbf{Z}$-basis of $\mathfrak{a}$. Since

$$\begin{pmatrix} 2 & 5 \\ 4 & 7 \end{pmatrix} \begin{pmatrix} 1 \\ \sqrt{10} \end{pmatrix} = \begin{pmatrix} 2 + 5\sqrt{10} \\ 4 + 7\sqrt{10} \end{pmatrix},$$

Theorem 3.10 tells us that $\mathrm{N}(\mathfrak{a}) = |\det(\begin{smallmatrix} 2 & 5 \\ 4 & 7 \end{smallmatrix})| = 6$. We will learn another way to compute $\mathrm{N}(\mathfrak{a})$ in Example 4.43.

The ideal norm of a principal ideal in $\mathcal{O}_K$ is related to the field norm on any generator of the ideal.

Theorem 3.12. *For nonzero $\alpha \in \mathcal{O}_K$,*

$$\mathrm{N}((\alpha)) = \left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right|.$$

Proof. Apply Theorem 3.10 with $V = K$, $M = \mathcal{O}_K$, $M' = \alpha\mathcal{O}_K$, and $\varphi = m_\alpha$. The index $[M : M'] = [\mathcal{O}_K : \alpha\mathcal{O}_K]$ is $\mathrm{N}((\alpha))$ and $\det\varphi = \mathrm{N}_{K/\mathbf{Q}}(\alpha)$. ■

Example 3.13. $\#(\mathbf{Z}[i]/(a+bi)) = a^2 + b^2$ for integers a and b not both 0.

Example 3.14. $\#(\mathbf{Z}[\sqrt{14}]/(a+b\sqrt{14})) = |a^2 - 14b^2|$ for $a, b \in \mathbf{Z}$ not both 0. Do not forget the absolute value signs. For example, if we try to find a principal ideal $(a + b\sqrt{14})$ with norm 5 by solving the equation $a^2 - 14b^2 = 5$, there is no solution since that equation implies $a^2 \equiv 5 \bmod 7$, but $5 \bmod 7$ is not a square (the nonzero squares mod 7 are $1, 2, 4$). This does not mean there isn't a principal ideal of norm 5, because the equation $a^2 - 14b^2 = -5$ has a solution: $a = 3, b = 1$. So $\mathrm{N}((3 + \sqrt{14})) = |-5| = 5$.

Remark 3.15. The most naive attempt at extending Theorem 3.12 to ideals with two generators is wrong: $\mathrm{N}((\alpha, \beta))$ is usually not $|(\mathrm{N}_{K/\mathbf{Q}}(\alpha), \mathrm{N}_{K/\mathbf{Q}}(\beta))|$. For instance, $(1 + 2i, 1 - 2i) = (1)$ as ideals in $\mathbf{Z}[i]$, so the norm of this ideal is 1, but $(\mathrm{N}_{\mathbf{Q}(i)/\mathbf{Q}}(1+2i), \mathrm{N}_{\mathbf{Q}(i)/\mathbf{Q}}(1-2i)) = 5$.

Theorem 3.16. *The norm of a nonzero prime ideal $\mathfrak{p}$ in $\mathcal{O}_K$ is the power of a prime number. More precisely, $\mathfrak{p} \cap \mathbf{Z} = p\mathbf{Z}$ for some prime number p, $\mathcal{O}_K/\mathfrak{p}$ has characteristic p, and $\mathrm{N}(\mathfrak{p})$ is a power of p.*

Proof. The natural composite of ring homomorphisms

$$\mathbf{Z} \to \mathcal{O}_K \to \mathcal{O}_K/\mathfrak{p} = \text{finite field}$$

has kernel $\mathfrak{p} \cap \mathbf{Z}$, which is a prime ideal. This intersection is nonzero by Theorem 2.17 (or because $\mathbf{Z}$ is infinite and $\mathcal{O}_K/\mathfrak{p}$ is finite). Therefore $\mathfrak{p} \cap \mathbf{Z} = p\mathbf{Z}$ for some prime number p, so $p = 0$ in $\mathcal{O}_K/\mathfrak{p}$, which means $\mathcal{O}_K/\mathfrak{p}$ has characteristic p. Viewing $\mathcal{O}_K/\mathfrak{p}$ as a vector space over $\mathbf{Z}/p\mathbf{Z}$, its size is p^d, where d is the dimension. ■

Example 3.17. The ideal (3) in $\mathbf{Z}[i]$ has norm 9 and is prime since $\mathbf{Z}[i]/(3)$ is a field, which can be checked by finding an inverse for each nonzero element of this quotient ring.

Although nonzero prime ideals in $\mathcal{O}_K$ have prime power norm, an ideal with prime power norm *need not* be a prime ideal. For instance, in $\mathbf{Z}[i]$ the ideal (5) has norm 25 but it is not prime since $5 = (1+2i)(1-2i)$. If the norm is a prime number, then the ideal is prime: if $\mathcal{O}_K/\mathfrak{a}$ has prime size then it is a field, so $\mathfrak{a}$ is maximal and thus prime.

Definition 3.18. For a nonzero prime $\mathfrak{p}$ in a number field K, the finite field $\mathcal{O}_K/\mathfrak{p}$ is called the *residue field* at $\mathfrak{p}$.

The residue fields at primes in $\mathbf{Q}$ are the fields $\mathbf{Z}/p\mathbf{Z}$, which all have prime size and different characteristics as p varies. In a general number field K, the residue fields $\mathcal{O}_K/\mathfrak{p}$ need not have prime size or different characteristics: $\mathbf{Z}[i]/(1+2i)$ and $\mathbf{Z}[i]/(1-2i)$ each have size 5 and $\mathbf{Z}[i]/(3)$ has size 9. The family of finite fields $\mathcal{O}_K/\mathfrak{p}$ is the most basic reason that algebraic number theory requires a familiarity with general finite fields and not only the fields of prime size.

Since $\mathrm{N}_{K/\mathbf{Q}}$ is multiplicative on K, Theorem 3.12 tells us that the ideal norm is multiplicative on nonzero principal ideals: $\mathrm{N}((\alpha)(\beta)) = \mathrm{N}((\alpha))\,\mathrm{N}((\beta))$. In other words, $[\mathcal{O}_K : (\alpha\beta)] = [\mathcal{O}_K : (\alpha)][\mathcal{O}_K : (\beta)]$. It is natural to wonder if this index formula could be explained conceptually by an isomorphism of the rings $\mathcal{O}_K/(\alpha\beta)$ and $\mathcal{O}_K/(\alpha) \times \mathcal{O}_K/(\beta)$, but generally these rings are not isomorphic. For instance, $\mathcal{O}_K/(\alpha^2)$ and $\mathcal{O}_K/(\alpha) \times \mathcal{O}_K/(\alpha)$ have the same size but they are not isomorphic as rings when α is not a unit (Exercise 4.23).

That $\mathcal{O}_K$ is the full ring of integers of K is not essential in the proof of Theorem 3.12. What matters is that $\mathcal{O}_K$ is a subring of K which is also a free $\mathbf{Z}$-module of rank n, where $n = [K : \mathbf{Q}]$. For example, Theorem 3.12 applies to $\mathbf{Z}[\sqrt{d}]$, whether or not it is the full ring of integers of $\mathbf{Q}(\sqrt{d})$: if the integers a and b are not both 0, then

$$\#(\mathbf{Z}[\sqrt{d}]/(a+b\sqrt{d})) = |a^2 - db^2|.$$

(This formula, for the index of a principal ideal in a pure quadratic ring, is established in a much simpler way in Exercise 1.27.)

3.4 Discriminant of a Basis and a Polynomial

In the proof that $\mathcal{O}_K$ has a $\mathbf{Z}$-basis, we found elements of $\mathcal{O}_K$ have the common denominator $\det(\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i\alpha_j))$ when they are written in terms of a $\mathbf{Q}$-basis $\{\alpha_1, \dots, \alpha_n\}$ of algebraic integers. This determinant construction can be applied to any basis of a finite extension field.

Definition 3.19. If E/F is a finite extension of fields and $\{\alpha_1, \dots, \alpha_n\}$ is a basis of E/F, the *discriminant* of $\{\alpha_1, \dots, \alpha_n\}$ is

$$\mathrm{disc}_{E/F}(\alpha_1, \dots, \alpha_n) = \det(\mathrm{Tr}_{E/F}(\alpha_i\alpha_j)) \in F.$$

Example 3.20. In $\mathbf{Q}(\sqrt[3]{2})$,

$$\mathrm{disc}_{\mathbf{Q}(\sqrt[3]{2})/\mathbf{Q}}(1, \sqrt[3]{2}, \sqrt[3]{4}) = \det\begin{pmatrix} \mathrm{Tr}(1) & \mathrm{Tr}(\sqrt[3]{2}) & \mathrm{Tr}(\sqrt[3]{4}) \\ \mathrm{Tr}(\sqrt[3]{2}) & \mathrm{Tr}(\sqrt[3]{4}) & \mathrm{Tr}(2) \\ \mathrm{Tr}(\sqrt[3]{4}) & \mathrm{Tr}(2) & \mathrm{Tr}(2\sqrt[3]{2}) \end{pmatrix},$$

where $\mathrm{Tr} = \mathrm{Tr}_{\mathbf{Q}(\sqrt[3]{2})/\mathbf{Q}}$. Since $\mathrm{Tr}_{\mathbf{Q}(\sqrt[3]{2})/\mathbf{Q}}(a + b\sqrt[3]{2} + c\sqrt[3]{4}) = 3a$, the matrix is

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 6 \\ 0 & 6 & 0 \end{pmatrix},$$

which has determinant $3(-36) = -108$. Thus $\mathrm{disc}_{\mathbf{Q}(\sqrt[3]{2})/\mathbf{Q}}(1, \sqrt[3]{2}, \sqrt[3]{4}) = -108$.

Of course when you first see the word discriminant you think about $b^2 - 4ac$ from quadratic polynomials. Higher-degree polynomials also have a discriminant: if $f(T) \in F[T]$ is monic[1] with degree $n \geqslant 2$, its discriminant is defined in terms of its roots by

$$f(T) = (T - \alpha_1)\cdots(T - \alpha_n) \Longrightarrow \mathrm{disc}(f(T)) := \prod_{i<j}(\alpha_j - \alpha_i)^2. \tag{3.8}$$

When $n = 1$, the product over $i < j$ is empty and we define $\mathrm{disc}(f(T)) = 1$. In the quadratic case $f(T) = T^2 + bT + c = (T - \alpha)(T - \beta)$, $(\beta - \alpha)^2 = (\alpha + \beta)^2 - 4\alpha\beta = b^2 - 4c$, which connects (3.8) to the high school definition. The polynomial discriminant doesn't look like the discriminant of a basis for a

[1] Discriminants of nonmonic polynomials can be defined, using the leading coefficient, but we have no need for that. See [35, p. 204].

field extension, but it turns out to be the discriminant of a special basis. We'll see this in Theorem 3.25. Here is the case of degree 2.

Example 3.21. Let $E = F(\alpha)$ be a quadratic extension with $\alpha^2 + b\alpha + c = 0$ $(b, c \in F)$. Then $\mathrm{Tr}_{E/F}(1) = 2$ and $\mathrm{Tr}_{E/F}(\alpha) = -b$, so $\mathrm{Tr}_{E/F}(\alpha^2) = \mathrm{Tr}_{E/F}(-b\alpha - c) = b^2 - 2c$. We have

$$\begin{aligned} \mathrm{disc}_{E/F}(1, \alpha) &= \det \begin{pmatrix} \mathrm{Tr}_{E/F}(1) & \mathrm{Tr}_{E/F}(\alpha) \\ \mathrm{Tr}_{E/F}(\alpha) & \mathrm{Tr}_{E/F}(\alpha^2) \end{pmatrix} \\ &= \det \begin{pmatrix} 2 & -b \\ -b & b^2 - 2c \end{pmatrix} \\ &= b^2 - 4c, \end{aligned}$$

which is the discriminant of $T^2 + bT + c$. For another basis,

$$\begin{aligned} \mathrm{disc}_{E/F}(1+\alpha, 1-\alpha) &= \det \begin{pmatrix} \mathrm{Tr}_{E/F}((1+\alpha)^2) & \mathrm{Tr}_{E/F}(1-\alpha^2) \\ \mathrm{Tr}_{E/F}(1-\alpha^2) & \mathrm{Tr}_{E/F}((1-\alpha)^2) \end{pmatrix} \\ &= \det \begin{pmatrix} 2 - 2b + b^2 - 2c & 2 - b^2 + 2c \\ 2 - b^2 + 2c & 2 + 2b + b^2 - 2c \end{pmatrix} \\ &= 4(b^2 - 4c). \end{aligned}$$

The two bases of E/F have discriminants differing by a square factor. This is a general phenomenon, as we'll see below (Equation (3.10)).

How are discriminants of two bases of E/F related to one another? Let $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\beta_1, \ldots, \beta_n\}$ be bases of E/F. Write each member of the second basis as an F-linear combination of the first basis and combine these equations into the single matrix equation

$$\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}, \qquad c_{ij} \in F. \tag{3.9}$$

Then

$$\beta_k \beta_\ell = \Bigl(\sum_{j=1}^{n} c_{kj}\alpha_j\Bigr)\Bigl(\sum_{j'=1}^{n} c_{\ell j'}\alpha_{j'}\Bigr) = \sum_{j,j'=1}^{n} c_{kj}c_{\ell j'}\alpha_j\alpha_{j'},$$

so

$$\begin{aligned}\operatorname{Tr}_{E/F}(\beta_k\beta_\ell) &= \sum_{j,j'} c_{kj}c_{\ell j'}\operatorname{Tr}_{E/F}(\alpha_j\alpha_{j'})\\ &= \sum_{j,j'} c_{kj}\operatorname{Tr}_{E/F}(\alpha_j\alpha_{j'})c_{\ell j'}.\end{aligned}$$

The indices in the threefold product on the right appear as $kj, jj', \ell j'$, which is almost like what we see in multiplication of 3 matrices, except we should have $j'\ell$ instead of $\ell j'$. Turning $\ell j'$ into $j'\ell$ requires a transpose, so

$$(\operatorname{Tr}_{E/F}(\beta_k\beta_\ell)) = C(\operatorname{Tr}_{E/F}(\alpha_j\alpha_{j'}))C^\top,$$

where $C = (c_{ij})$ is the change-of-basis matrix from (3.9). Taking determinants on both sides yields

$$\operatorname{disc}_{E/F}(\beta_1,\ldots,\beta_n) = (\det C)^2\operatorname{disc}_{E/F}(\alpha_1,\ldots,\alpha_n). \tag{3.10}$$

Therefore changing bases changes the discriminant by a nonzero square factor. In particular, all bases of E/F have nonzero discriminant or all have discriminant 0. We saw in the proof of Theorem 3.2 that bases of a number field have nonzero discriminant over $\mathbf{Q}$. Other field extensions might have discriminant 0.

How can a basis of E/F have discriminant 0? Reviewing the proof that $d \neq 0$ in the proof of Theorem 3.2, the same argument shows $\operatorname{disc}_{E/F}(\alpha_1,\ldots,\alpha_n) = 0$ if and only if the trace function $\operatorname{Tr}_{E/F}\colon E \to F$ is identically 0. In the number field case this can't happen since $\operatorname{Tr}_{K/\mathbf{Q}}(1) = [K : \mathbf{Q}] \neq 0$, but there are field extensions E/F whose trace map is identically 0. In fact, Corollary 8.21 tells us $\operatorname{Tr}_{E/F}$ is identically 0 if and only if E/F is an inseparable extension. (Inseparable means not separable: some element of E has a minimal polynomial over F which is not separable.) So we get the following theorem.

Theorem 3.22. *For separable E/F, every basis has nonzero discriminant. For inseparable E/F, every basis has discriminant* 0*.*

This is why the discriminant is essentially useless when E/F is inseparable.

The formula (3.10) justifies something that was implicitly assumed in our notation $\operatorname{disc}_{E/F}(\alpha_1,\ldots,\alpha_n)$: the discriminant only depends on the basis as a set, rather than as an ordered list. The definition of the discriminant uses an ordered basis (knowing which basis vector is called the first one, and so on, in order to know how to fill in the matrix $(\mathrm{Tr}_{E/F}(e_ie_j)))$, but two orderings of the same basis are linked by a change-of-basis matrix which is a permutation matrix, and permutation matrices have determinant ±1 (namely the sign of the permutation). The square of that determinant is 1, so discriminants of the same basis ordered in two different ways yield the same value.

The discriminant has a geometric analogue in Euclidean space. For a basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of $\mathbf{R}^n$, the "box"

$$\{a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n : 0 \leqslant a_i \leqslant 1\}$$

has n-dimensional volume $|\det(\mathbf{v}_1 \cdots \mathbf{v}_n)|$, where $(\mathbf{v}_1 \cdots \mathbf{v}_n)$ is the $n \times n$ matrix with the vectors $\mathbf{v}_i$ as its columns. There is a different formula for this volume, using a matrix of dot products which resembles the discriminant.

Theorem 3.23. *If $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbf{R}^n$, the box $\{a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n : 0 \leqslant a_i \leqslant 1\}$ has volume $\sqrt{|\det(\mathbf{v}_i \cdot \mathbf{v}_j)|}$.*

Proof. This box is the image of the unit n-cube $I^n = [0, 1]^n$ in $\mathbf{R}^n$ under the matrix with the $\mathbf{v}_i$'s as its columns:

$$M = \begin{pmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{pmatrix}.$$

The volume of the box $M(I^n)$ is

$$|\det M| \cdot \mathrm{vol}(I^n) = |\det M|.$$

That is the volume formula mentioned before the theorem. To give a different formula, we compute

$$M^\top M = \begin{pmatrix} — & \mathbf{v}_1 & — \\ & \vdots & \\ — & \mathbf{v}_n & — \end{pmatrix} \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix} = (\mathbf{v}_i \cdot \mathbf{v}_j). \tag{3.11}$$

Taking determinants,

$$(\det M)^2 = \det(\mathbf{v}_i \cdot \mathbf{v}_j),$$

so

$$\sqrt{|\det(\mathbf{v}_i \cdot \mathbf{v}_j)|} = |\det M| = \mathrm{vol}(M(I^n)).$$

■

Think about $\mathbf{v} \cdot \mathbf{w}$ on $\mathbf{R}^n$ and $\mathrm{Tr}_{E/F}(xy)$ on E as analogous constructions: the dot product

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v} \cdot \mathbf{w}$$

is a bilinear map $\mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ and the *trace pairing*

$$\langle x, y \rangle = \mathrm{Tr}_{E/F}(xy)$$

is a bilinear map $E \times E \to F$. We call the matrix $(\mathrm{Tr}_{E/F}(\alpha_i \alpha_j))$ a *trace pairing matrix*. The two determinants

$$\det(\mathbf{v}_i \cdot \mathbf{v}_j) \qquad \text{and} \qquad \det(\mathrm{Tr}_{E/F}(\alpha_i \alpha_j))$$

for a set of n vectors $\{\mathbf{v}_i\}$ and a set of n numbers $\{\alpha_i\}$ are analogous. For instance, the box spanned by any n linearly dependent vectors in $\mathbf{R}^n$ has n-dimensional volume 0 and the discriminant of n linearly dependent numbers in E, defined in the same way as the discriminant of a basis, is 0 (Exercise 3.10).

Theorem 3.23 says $\mathrm{vol}(M(I^n))^2 = |\det(\mathbf{v}_i \cdot \mathbf{v}_j)|$, which suggests we interpret $\mathrm{disc}_{E/F}(\alpha_1, \ldots, \alpha_n)$ as the "squared volume" of the "box" with the α_i's as the edges, in some sense. That is the geometric intuition behind discriminants: it's a kind of abstract squared volume.

For computational purposes, it is important that the discriminant of a *power basis* for a number field can be computed as the discriminant of a polynomial and as a norm. (See Theorem 3.25.) To make these connections we use the following discriminant formula that involves all the embeddings of a number field into $\mathbf{C}$.

Lemma 3.24. *Let $n = [K : \mathbf{Q}]$. If $\{\alpha_1, \ldots, \alpha_n\}$ is a basis of $K/\mathbf{Q}$ then*

$$\mathrm{disc}_{K/\mathbf{Q}}(\alpha_1, \ldots, \alpha_n) = \det(\sigma_i(\alpha_j))^2,$$

where $\sigma_1, \ldots, \sigma_n$ are the different field embeddings of K into $\mathbf{C}$.

Proof. There are n field embeddings $\sigma_i \colon K \to \mathbf{C}$ since we can write $K = \mathbf{Q}(x)$ (primitive element theorem) and then each embedding is determined by sending x to one of the complex roots of its minimal polynomial over $\mathbf{Q}$. The minimal polynomial, whose degree is n, has as many roots as its degree since it is separable.

For any $\alpha \in K$, $\mathrm{Tr}_{K/\mathbf{Q}}(\alpha) = \sum_{k=1}^{n} \sigma_k(\alpha)$ by Theorem 8.18, so

$$\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i\alpha_j) = \sum_{k=1}^{n} \sigma_k(\alpha_i)\sigma_k(\alpha_j) = (\sigma_1(\alpha_i), \dots, \sigma_n(\alpha_i)) \cdot (\sigma_1(\alpha_j), \dots, \sigma_n(\alpha_j)).$$

In the same way we found that

$$\det(\mathbf{v}_i \cdot \mathbf{v}_j) = \det\begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix}^2$$

in the proof of Theorem 3.23,

$$\det(\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i\alpha_j)) = \det\begin{pmatrix} | & & | \\ \sigma_i(\alpha_1) & \cdots & \sigma_i(\alpha_n) \\ | & & | \end{pmatrix}^2 = \det(\sigma_i(\alpha_j))^2. \quad \blacksquare$$

Theorem 3.25. *If $f(T) \in \mathbf{Q}[T]$ is monic irreducible of degree $n \geqslant 1$ and α is a root, then*

$$\mathrm{disc}_{\mathbf{Q}(\alpha)/\mathbf{Q}}(1, \alpha, \dots, \alpha^{n-1}) = \mathrm{disc}(f(T)) = (-1)^{\frac{n(n-1)}{2}} \mathrm{N}_{\mathbf{Q}(\alpha)/\mathbf{Q}}(f'(\alpha)).$$

Proof. When $n = 1$, the terms we want to compare are all equal to 1. Let $n = \deg f \geqslant 2$ and $\sigma_1, \dots, \sigma_n$ be the field embeddings $\mathbf{Q}(\alpha) \to \mathbf{C}$. Since $f(T)$ is irreducible over $\mathbf{Q}$, $f(T) = \prod_{i=1}^{n}(T - \sigma_i(\alpha))$, so

$$\mathrm{disc}(f(T)) = \prod_{i<j}(\sigma_j(\alpha) - \sigma_i(\alpha))^2.$$

By Lemma 3.24,

$$\mathrm{disc}_{\mathbf{Q}(\alpha)/\mathbf{Q}}(1, \alpha, \dots, \alpha^{n-1}) = \det(\sigma_i(\alpha^{j-1}))^2.$$

Writing $\sigma_i(\alpha^{j-1})$ as $\sigma_i(\alpha)^{j-1}$, the determinant is given by Vandermonde's formula:

$$\det\begin{pmatrix} 1 & \sigma_1(\alpha) & \cdots & \sigma_1(\alpha)^{n-1} \\ 1 & \sigma_2(\alpha) & \cdots & \sigma_2(\alpha)^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \sigma_n(\alpha) & \cdots & \sigma_n(\alpha)^{n-1} \end{pmatrix} = \prod_{i<j}(\sigma_j(\alpha) - \sigma_i(\alpha)).$$

Square this and we get $\operatorname{disc}(f(T))$.

To show

$$(-1)^{n(n-1)/2}\,\mathrm{N}_{\mathbf{Q}(\alpha)/\mathbf{Q}}(f'(\alpha)) \overset{?}{=} \prod_{1\leqslant i<j\leqslant n}(\sigma_j(\alpha)-\sigma_i(\alpha))^2$$

we will rearrange the terms on the right. From the product rule for derivatives,

$$f(T)=(T-\sigma_1(\alpha))\cdots(T-\sigma_n(\alpha)) \Longrightarrow f'(\sigma_i(\alpha))=\prod_{j\neq i}(\sigma_j(\alpha)-\sigma_i(\alpha)).$$

Multiplying these over all i,

$$\prod_{i=1}^{n}\prod_{j\neq i}(\sigma_j(\alpha)-\sigma_i(\alpha))=\prod_{i=1}^{n}f'(\sigma_i(\alpha)).$$

The product of $\sigma_j(\alpha)-\sigma_i(\alpha)$ runs over sets of distinct indices i and j. To rewrite this product over index pairs where $i<j$, collect $\sigma_j(\alpha)-\sigma_i(\alpha)$ and $\sigma_i(\alpha)-\sigma_j(\alpha)$ together as $-(\sigma_j(\alpha)-\sigma_i(\alpha))^2$. There are $\binom{n}{2}=\frac{n(n-1)}{2}$ such pairs, so

$$\prod_{i<j}(\sigma_j(\alpha)-\sigma_i(\alpha))^2=(-1)^{n(n-1)/2}\prod_{i=1}^{n}f'(\sigma_i(\alpha)).$$

The product of derivatives is $\mathrm{N}_{\mathbf{Q}(\alpha)/\mathbf{Q}}(f'(\alpha))$ by Corollary 8.16. ■

Example 3.26. Returning to Example 3.20, we compute the discriminant more easily as

$$\operatorname{disc}_{\mathbf{Q}(\sqrt[3]{2})/\mathbf{Q}}(1,\sqrt[3]{2},\sqrt[3]{4})=-\,\mathrm{N}_{\mathbf{Q}(\sqrt[3]{2})/\mathbf{Q}}\left(3\sqrt[3]{2}^{2}\right)=-27\cdot 4=-108.$$

Here are formulas for discriminants of trinomial polynomials. For $n\geqslant 2$,

$$\operatorname{disc}(T^n+aT+b)=(-1)^{n(n-1)/2}((-1)^{n-1}(n-1)^{n-1}a^n+n^nb^{n-1}). \quad (3.12)$$

Some particular cases are

$$\begin{aligned}
\operatorname{disc}(T^2+aT+b) &= a^2-4b,\\
\operatorname{disc}(T^3+aT+b) &= -(4a^3+27b^2),\\
\operatorname{disc}(T^4+aT+b) &= -27a^4+256b^3,\\
\operatorname{disc}(T^5+aT+b) &= 256a^5+3125b^4.
\end{aligned}$$

These are trinomials where the middle term has degree 1. What about trinomials with a middle term of higher degree? When $0 < m < n$ and $(m, n) = 1$,

$$\operatorname{disc}(T^n + aT^m + b) = (-1)^{n(n-1)/2} b^{m-1}((-1)^{n-1} m^m (n-m)^{n-m} a^n + n^n b^{n-m}).$$

When $0 < m < n$ and $d = (m, n)$ is not necessarily 1, $\operatorname{disc}(T^n + aT^m + b)$ equals

$$(-1)^{n(n-1)/2} b^{m-1}((-1)^{n/d-1} m^{m/d} (n-m)^{(n-m)/d} a^{n/d} + n^{n/d} b^{(n-m)/d})^d.$$

This last formula subsumes all the other ones, and it is proved in [58, Theorem 2]. In examples we often only need (3.12).

Example 3.27. Say $K = \mathbf{Q}(\beta)$ with $\beta^5 - \beta - 1 = 0$. The polynomial $T^5 - T - 1$ is irreducible mod 5, hence irreducible over $\mathbf{Q}$, so $[K : \mathbf{Q}] = 5$ and

$$\operatorname{disc}_{K/\mathbf{Q}}(1, \beta, \beta^2, \beta^3, \beta^4) = \operatorname{disc}(T^5 - T - 1) = 256(-1)^5 + 3125(-1)^4 = 2869.$$

Another application of Lemma 3.24 is the determination of the sign of the discriminant of a $\mathbf{Q}$-basis of K. The discriminants of any two bases differ by a square factor, so they have the same sign. What is this common sign?

Theorem 3.28 (Brill, 1877). *Let $K = \mathbf{Q}(\alpha)$ and let $f(T)$ be the minimal polynomial of α over $\mathbf{Q}$. The sign of the discriminant of any $\mathbf{Q}$-basis of K is $(-1)^P$, where P is the number of pairs of complex conjugate roots of $f(T)$.*

Any non-real root $z \in \mathbf{C}$ of $f(T)$ comes together with a second root $\overline{z}$. They form a pair of conjugate roots and P is the number of these pairs.

Proof. Let $\{\alpha_1, \ldots, \alpha_n\}$ be a $\mathbf{Q}$-basis of K and D be its discriminant, so $D = \det(\sigma_i(\alpha_j))^2$. This is a real number, so $\det(\sigma_i(\alpha_j))$ is a real number if $D > 0$ and is a pure imaginary number if $D < 0$. We will show the two possibilities are related to $(-1)^P$.

Split up the field embeddings $\sigma_1, \ldots, \sigma_n \colon K \to \mathbf{C}$ into two sets: those with image in $\mathbf{R}$ and the rest. The first set is those σ_i sending α to a real root of $f(T)$ (if there are any). When $\sigma_i(K) \not\subset \mathbf{R}$, the complex-conjugate embedding $\overline{\sigma}_i$ is another non-real embedding of K. There are P pairs $\sigma_i, \overline{\sigma}_i$, corresponding to the choices of complex-conjugate roots of $f(T)$. Applying complex conjugation to the matrix $(\sigma_i(\alpha_j))$ permutes the rows for σ_i and $\overline{\sigma}_i$, so there are P row swaps. Therefore $\overline{\det(\sigma_i(\alpha_j))} = (-1)^P \det(\sigma_i(\alpha_j))$. Multiplying both sides by

$\det(\sigma_i(\alpha_j))$,

$$|\det(\sigma_i(\alpha_j))|^2 = (-1)^P D.$$

The left side is positive, so D must have sign $(-1)^P$. ■

Example 3.29. The polynomial $T^3 - 2$ has two non-real roots, forming one pair, so $\operatorname{disc}_{\mathbf{Q}(\sqrt[3]{2})/\mathbf{Q}}(1, \sqrt[3]{2}, \sqrt[3]{4})$ has sign $(-1)^1 = -1$. We know its discriminant is -108.

Example 3.30. The polynomial $T^5 - T - 1$ has four non-real roots, which are two pairs of complex-conjugate roots, so if β is a root then the sign of $\operatorname{disc}_{\mathbf{Q}(\beta)/\mathbf{Q}}(1, \beta, \beta^2, \beta^3, \beta^4)$ is $(-1)^2 = 1$. We computed the discriminant as 2869 in Example 3.27.

Why does Theorem 3.28 matter? Sometimes $\operatorname{disc}(K)$ is computed by finding its prime factors and their multiplicities, which leaves $\operatorname{disc}(K)$ known up to a sign. Theorem 3.28 pins down the sign.

3.5 Discriminant of a Number Field

Definition 3.31. The *discriminant* of a number field K is defined to be the discriminant of any $\mathbf{Z}$-basis of $\mathcal{O}_K$:

$$\operatorname{disc}(K) := \operatorname{disc}_{K/\mathbf{Q}}(x_1, \ldots, x_n) = \det(\operatorname{Tr}_{K/\mathbf{Q}}(x_i x_j)) \in \mathbf{Z} - \{0\},$$

where $\{x_1, \ldots, x_n\}$ is any $\mathbf{Z}$-basis of $\mathcal{O}_K$. It is an integer since the traces $\operatorname{Tr}_{K/\mathbf{Q}}(x_i x_j)$ are integers, and it is not 0 by Theorem 3.22.

Different bases of a field extension have discriminants differing by a nonzero square factor. But different $\mathbf{Z}$-bases of $\mathcal{O}_K$ have the same discriminant. If $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_n\}$ are both $\mathbf{Z}$-bases of $\mathcal{O}_K$, then by (3.10),

$$\operatorname{disc}_{K/\mathbf{Q}}(x_1, \ldots, x_n) = (\det C)^2 \operatorname{disc}_{K/\mathbf{Q}}(y_1, \ldots, y_n)$$

where C is the change-of-basis matrix from the y_i's to the x_i's. Since the x_i's and y_i's have the same $\mathbf{Z}$-span, the change of basis matrix between them is an invertible integral matrix, so of determinant ± 1, whose square is 1.

Example 3.32. For a squarefree integer d, use a $\mathbf{Z}$-basis of the integers of $\mathbf{Q}(\sqrt{d})$ to verify from Example 3.21 that

$$\operatorname{disc}(\mathbf{Q}(\sqrt{d})) = \begin{cases} d, & \text{if } d \equiv 1 \bmod 4, \\ 4d, & \text{if } d \equiv 2, 3 \bmod 4. \end{cases}$$

In particular, when K is a quadratic field with discriminant D, $K = \mathbf{Q}(\sqrt{D})$.

Example 3.33. Since the ring of integers of $\mathbf{Q}(\sqrt[3]{2})$ is $\mathbf{Z}[\sqrt[3]{2}]$, which has $\mathbf{Z}$-basis $\{1, \sqrt[3]{2}, \sqrt[3]{4}\}$, $\operatorname{disc}(\mathbf{Q}(\sqrt[3]{2})) = -108$ by Example 3.20 or 3.26.

Remark 3.34. The discriminant is divisible in towers: if $\mathbf{Q} \subset K \subset L$ then $\operatorname{disc}(K) \mid \operatorname{disc}(L)$. (In fact, $\operatorname{disc}(K)^{[L:K]} \mid \operatorname{disc}(L)$.) This divisibility relation is analogous to the volume of a box being divisible by the area of one of its sides.

The discriminant is a tool for computing $\mathcal{O}_K$. To use it in this way it is convenient to introduce discriminants of “big” subgroups of K called $\mathbf{Z}$-lattices.

Definition 3.35. A $\mathbf{Z}$-*lattice* in a number field K of degree n is a subgroup $M \subset K$ which is isomorphic to $\mathbf{Z}^n$.[2]

Example 3.36. The ring of integers $\mathcal{O}_K$ is a $\mathbf{Z}$-lattice in K. This is the most important example.

Example 3.37. If $K = \mathbf{Q}(\alpha)$ and α is an algebraic integer, the ring $\mathbf{Z}[\alpha]$ is a $\mathbf{Z}$-lattice.

Unlike $\mathcal{O}_K$, a general $\mathbf{Z}$-lattice in K need not be a ring.

Example 3.38. Nonzero ideals in $\mathcal{O}_K$ are $\mathbf{Z}$-lattices, as is $\frac{1}{2}\mathcal{O}_K$.

Nonexample 3.39. In $\mathbf{Q}$, the ring $\mathbf{Z}[\frac{1}{2}]$ is not a $\mathbf{Z}$-lattice since it is not finitely generated as a $\mathbf{Z}$-module.

Definition 3.40. The *discriminant* of a $\mathbf{Z}$-lattice M in K is defined to be

$$\operatorname{disc}(M) = \det(\operatorname{Tr}_{K/\mathbf{Q}}(e_i e_j)) \in \mathbf{Q}^\times,$$

where $\{e_1, \ldots, e_n\}$ is any $\mathbf{Z}$-basis of M.

[2] There is no subgroup of K isomorphic to $\mathbf{Z}^m$ with $m > n$, since $\mathbf{Z}$-linear independence in K implies $\mathbf{Q}$-linear independence, and a $\mathbf{Q}$-linearly independent subset of K has at most n members.

In the notation of Definition 3.35, what we defined to be $\operatorname{disc}(K)$ is really $\operatorname{disc}(\mathcal{O}_K)$. Since M may not be a ring, generally $\operatorname{disc}(M)$ is in $\mathbf{Q}$ rather than $\mathbf{Z}$.

Example 3.41. In $\mathbf{Q}(i)$, the $\mathbf{Z}$-lattice $\mathbf{Z}\frac{1}{3} + \mathbf{Z}(1+i)$ has discriminant

$$\det\begin{pmatrix} \operatorname{Tr}_{\mathbf{Q}(i)/\mathbf{Q}}(\frac{1}{9}) & \operatorname{Tr}_{\mathbf{Q}(i)/\mathbf{Q}}(\frac{1}{3}+\frac{i}{3}) \\ \operatorname{Tr}_{\mathbf{Q}(i)/\mathbf{Q}}(\frac{1}{3}+\frac{i}{3}) & \operatorname{Tr}_{\mathbf{Q}(i)/\mathbf{Q}}(2i) \end{pmatrix} = \det\begin{pmatrix} 2/9 & 2/3 \\ 2/3 & 0 \end{pmatrix} = -\frac{4}{9}.$$

The well-definedness of the discriminant of a $\mathbf{Z}$-lattice (its independence of the $\mathbf{Z}$-basis used) is due to the same reason as the well-definedness of $\operatorname{disc}(K)$: two $\mathbf{Z}$-bases of a $\mathbf{Z}$-lattice have change-of-basis matrix with determinant ±1, whose square is 1.

For a general $\mathbf{Z}$-lattice in K, the only way to compute its discriminant is with a trace pairing matrix. For $\mathbf{Z}$-lattices of the form $\mathbf{Z}[\alpha]$, with power basis $\{1, \alpha, \ldots, \alpha^{n-1}\}$, Theorem 3.25 tells us its discriminant can be computed in two other ways:

$$\operatorname{disc}(\mathbf{Z}[\alpha]) = \operatorname{disc}(f(T)) = (-1)^{\frac{n(n-1)}{2}} \operatorname{N}_{\mathbf{Q}(\alpha)/\mathbf{Q}}(f'(\alpha)), \tag{3.13}$$

where $f(T)$ is the minimal polynomial of α over $\mathbf{Q}$. We will use this in many examples.

Theorem 3.42. *If $M_1 \subset M_2$ are two $\mathbf{Z}$-lattices in K, then*

$$\operatorname{disc}(M_1) = [M_2 : M_1]^2 \operatorname{disc}(M_2).$$

The appearance of $[M_2 : M_1]^2$, and not just $[M_2 : M_1]$, matches the intuition of discriminants as squared volume. In Euclidean space, a linear transformation alters volumes by (the absolute value of) its determinant, but discriminant computations get altered by square factors.

Notice the smaller lattice M_1 has the larger discriminant, which is analogous to the geometric fact that the box for a sublattice of a lattice in $\mathbf{R}^n$ has a larger volume. (A lattice in $\mathbf{R}^n$ is the $\mathbf{Z}$-span of a basis of $\mathbf{R}^n$, like $\mathbf{Z}^n$.)

Proof. Let $n = [K : \mathbf{Q}]$. Both M_1 and M_2 are finite free $\mathbf{Z}$-modules of rank n. We use aligned $\mathbf{Z}$-bases for M_1 and M_2 (Theorem 8.33):

$$M_2 = \bigoplus_{i=1}^{n} \mathbf{Z}e_i \qquad \text{and} \qquad M_1 = \bigoplus_{i=1}^{n} \mathbf{Z}a_i e_i$$

with $a_i \in \mathbf{Z} - \{0\}$. So

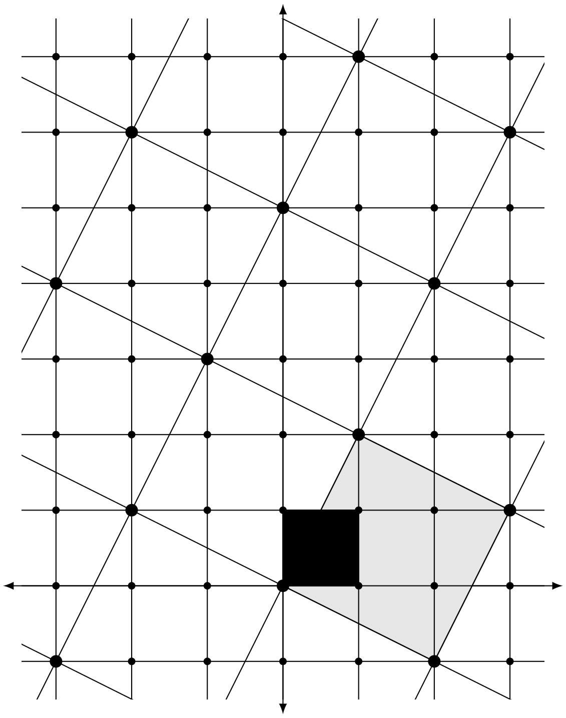
$$\begin{aligned}\operatorname{disc}(M_1) &= \det(\operatorname{Tr}_{K/\mathbf{Q}}(a_i e_i a_j e_j))\\ &= \det(a_i a_j \operatorname{Tr}_{K/\mathbf{Q}}(e_i e_j))\\ &= a_1^2 a_2^2 \cdots a_n^2 \det(\operatorname{Tr}_{K/\mathbf{Q}}(e_i e_j))\\ &= (a_1 \cdots a_n)^2 \operatorname{disc}(M_2),\end{aligned}$$

and

$$M_2/M_1 = \bigoplus_{i=1}^{n} \mathbf{Z} e_i \Big/ \bigoplus_{i=1}^{n} \mathbf{Z} a_i e_i \cong \bigoplus_{i=1}^{n} \mathbf{Z}/a_i\mathbf{Z},$$

so $[M_2 : M_1] = |a_1 \cdots a_n|$. ■

Example 3.43. Figure 3.1 shows the $\mathbf{Z}$-lattice $\mathbf{Z}[i]$ in $\mathbf{Q}(i)$ and its sublattice the ideal $\mathfrak{a} = (1 + 2i)$.



$$\mathbf{Z}[i] = \mathbf{Z}(1 + 2i) + \mathbf{Z}i, \quad (1 + 2i) = \mathbf{Z}(1 + 2i) + \mathbf{Z}(-2 + i)$$

Figure 3.1: A sublattice of $\mathbf{Z}[i]$.

Using the indicated $\mathbf{Z}$-bases of the lattices to make computations, check $\operatorname{disc}(\mathbf{Z}[i]) = -4$ and $\operatorname{disc}(\mathfrak{a}) = -100$. In Figure 3.1 the bases for $\mathbf{Z}[i]$ and the ideal $\mathfrak{a}$ do not lie along the same lines, so the boxes for the two lattices don't mesh with each other and it's not clear from the picture what $[\mathbf{Z}[i] : (1+2i)]$ is. By Theorem 3.10, we can compute $[\mathbf{Z}[i] : (1+2i)]$ by writing the basis $1+2i$ and $-2+i$ of $\mathfrak{a}$ in terms of the basis 1 and i of $\mathbf{Z}[i]$ and computing the determinant of their transition matrix:

$$\begin{pmatrix} 1+2i \\ -2+i \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ i \end{pmatrix} \Longrightarrow [\mathbf{Z}[i] : (1+2i)] = \left| \det \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \right| = 5.$$

This is consistent with $\operatorname{disc}(\mathfrak{a})/\operatorname{disc}(\mathbf{Z}[i]) = -100/(-4) = 25 = 5^2$. Aligned bases for $\mathbf{Z}[i]$ and $\mathfrak{a}$ in Figure 3.2, which lie along the same lines, make it visually clear the index is 5: five copies of the parallelogram having the new basis of $\mathbf{Z}[i]$ as edges fit in the parallelogram with the new basis of $\mathfrak{a}$ as edges.

In practice we will focus on $\mathbf{Z}$-lattices M inside $\mathcal{O}_K$, so $\operatorname{disc}(M) \in \mathbf{Z}$ and

$$\operatorname{disc}(M) = [\mathcal{O}_K : M]^2 \operatorname{disc}(K) = [\mathcal{O}_K : M]^2 \operatorname{disc}(K). \tag{3.14}$$

This equation tells us $\operatorname{disc}(M)$ is a multiple of $[\mathcal{O}_K : M]^2$. Let s^2 be the largest square factor of $\operatorname{disc}(M)$, so $[\mathcal{O}_K : M] \mid s$. Since $[\mathcal{O}_K : M]\mathcal{O}_K \subset M$, also $s\mathcal{O}_K \subset M$. Therefore

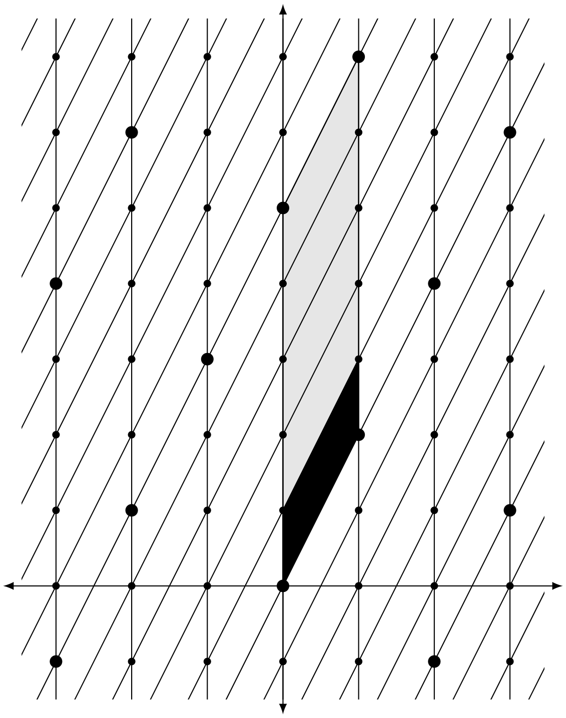
$$\boxed{M \subset \mathcal{O}_K \subset \frac{1}{s}M, \quad \text{where } s^2 \text{ is the largest square factor of } \operatorname{disc}(M),} \tag{3.15}$$

which is a sharper version of (3.6). This leads to a method of computing $\mathcal{O}_K$. Pick any $\mathbf{Z}$-lattice M in $\mathcal{O}_K$ (such as $\mathbf{Z}[\alpha]$ for a specific α). We can compute $\operatorname{disc}(M)$ *without* knowing $\mathcal{O}_K$. Let s^2 be the largest square factor of $\operatorname{disc}(M)$ and run through a list of coset representatives x_i for $\frac{1}{s}M/M$ and see which ones are algebraic integers by computing their characteristic polynomials for $K/\mathbf{Q}$ (Theorem 2.31). The cosets $x_i + M$ with $x_i \in \mathcal{O}_K$ fill up $\mathcal{O}_K$.

Corollary 3.44. *Let K be a number field of degree n and M be a $\mathbf{Z}$-lattice inside $\mathcal{O}_K$. If $\operatorname{disc}(M)$ is squarefree then $\mathcal{O}_K = M$.*

Proof. Since $\operatorname{disc}(M) = [\mathcal{O}_K : M]^2 \operatorname{disc}(K)$ and the largest square factor of $\operatorname{disc}(M)$ is 1, $[\mathcal{O}_K : M] = 1$. ■

We will now use discriminants to compute a lot of rings of integers.



$$\mathbf{Z}[i] = \mathbf{Z}(1+2i) + \mathbf{Z}i, \quad (1+2i) = \mathbf{Z}(1+2i) + \mathbf{Z}\cdot 5i$$

Figure 3.2: Aligning $\mathbf{Z}[i]$ and a sublattice.

Example 3.45. Let $K = \mathbf{Q}(\alpha)$ where $\alpha^3 - \alpha - 1 = 0$. Since $T^3 - T - 1$ is irreducible mod 3, it is irreducible over $\mathbf{Q}$, so $[K : \mathbf{Q}] = 3$ and

$$\operatorname{disc}(\mathbf{Z}[\alpha]) = \operatorname{disc}_{K/\mathbf{Q}}(1, \alpha, \alpha^2) = \operatorname{disc}(T^3 - T - 1) = -23,$$

which is squarefree, so $\mathcal{O}_K = \mathbf{Z}[\alpha] = \mathbf{Z} + \mathbf{Z}\alpha + \mathbf{Z}\alpha^2$.

Example 3.46. Let $K = \mathbf{Q}(\beta)$ with $\beta^5 - \beta - 1 = 0$. By Example 3.27, $\operatorname{disc}(T^5 - T - 1) = 2869 = 19 \cdot 151$, which is squarefree, so $\mathcal{O}_K = \mathbf{Z}[\beta]$.

Many number fields do not have squarefree discriminant. For example, we have already seen $\operatorname{disc}(\mathbf{Q}(\sqrt[3]{2})) = -108 = -4 \cdot 27$.

Example 3.47. Let $K = \mathbf{Q}(\gamma)$ with $\gamma^3 - \gamma - 4 = 0$. This is a cubic field since

$T^3 - T - 4$ is irreducible over $\mathbf{Q}$ (look mod 3) and

$$\operatorname{disc}(\mathbf{Z}[\gamma]) = \operatorname{disc}_{K/\mathbf{Q}}(1, \gamma, \gamma^2) = \operatorname{disc}(T^3 - T - 4) = -428 = -4 \cdot 107.$$

The largest square factor is 4, so $[\mathcal{O}_K : \mathbf{Z}[\gamma]]^2 \mid 4$. Therefore $\mathbf{Z}[\gamma]$ has index 1 or 2 in $\mathcal{O}_K$, which means $2\mathcal{O}_K \subset \mathbf{Z}[\gamma]$, so

$$\mathbf{Z}[\gamma] \subset \mathcal{O}_K \subset \frac{1}{2}\mathbf{Z}[\gamma].$$

Coset representatives for $\mathbf{Z}[\gamma]$ in $\frac{1}{2}\mathbf{Z}[\gamma]$ are $\frac{a}{2} + \frac{b}{2}\gamma + \frac{c}{2}\gamma^2$ where $a, b, c \in \{0, 1\}$. (This is analogous to the coset representatives for $\frac{1}{3}\mathbf{Z}[\sqrt[3]{2}]/\mathbf{Z}[\sqrt[3]{2}]$ in (3.2.)) If $\mathcal{O}_K \neq \mathbf{Z}[\gamma]$ one of the nonzero coset representatives has to be in $\mathcal{O}_K$ (and in fact only one can be, because then $[\mathcal{O}_K : \mathbf{Z}[\gamma]] = 2$). The matrix for multiplication by $\frac{a}{2} + \frac{b}{2}\gamma + \frac{c}{2}\gamma^2$ on $\mathbf{Q}(\gamma)$ with respect to the $\mathbf{Q}$-basis $\{1, \gamma, \gamma^2\}$ is

$$\begin{pmatrix} a/2 & 2c & 2b \\ b/2 & a/2 + c/2 & b/2 + 2c \\ c/2 & b/2 & a/2 + c/2 \end{pmatrix}. \tag{3.16}$$

The characteristic polynomials in $\mathbf{Q}(\gamma)/\mathbf{Q}$ of the coset representatives are in Table 3.1. For $a = 0$ and $b = c = 1$ we get an algebraic integer: $\frac{\gamma+\gamma^2}{2}$ has characteristic polynomial $T^3 - T^2 - 3T - 2$.

(a, b, c)	Characteristic polynomial of (3.16)
$(1, 0, 0)$	$T^3 - \frac{3}{2}T^2 + \frac{3}{4}T - \frac{1}{8}$
$(0, 1, 0)$	$T^3 - \frac{1}{4}T - \frac{1}{2}$
$(0, 0, 1)$	$T^3 - T^2 + \frac{1}{4}T - 2$
$(1, 1, 0)$	$T^3 - \frac{3}{2}T^2 + \frac{1}{2}T - \frac{1}{2}$
$(1, 0, 1)$	$T^3 - \frac{5}{2}T^2 + 2T - \frac{5}{2}$
$(0, 1, 1)$	$T^3 - T^2 - 3T - 2$
$(1, 1, 1)$	$T^3 - \frac{5}{2}T^2 - \frac{5}{4}T - \frac{7}{8}$

Table 3.1: Characteristic polynomials of nonzero coset representatives.

Now that we have found (by systematic work, not random guessing) the algebraic integer $\frac{\gamma+\gamma^2}{2}$, we must have $[\mathcal{O}_K : \mathbf{Z}[\gamma]] = 2$ and

$$\mathbf{Z}[\gamma] \subsetneqq \mathbf{Z} + \mathbf{Z}\gamma + \mathbf{Z}\frac{\gamma + \gamma^2}{2} \subset \mathcal{O}_K,$$

so

$$\mathcal{O}_K = \mathbf{Z} + \mathbf{Z}\gamma + \mathbf{Z}\frac{\gamma+\gamma^2}{2}$$

and $\operatorname{disc}(K) = -107$.

Example 3.48. Let $K = \mathbf{Q}(\alpha)$ where α is a root of $T^3 - 12T + 2$ (Eisenstein at 2, so irreducible). We have

$$\operatorname{disc}(\mathbf{Z}[\alpha]) = \operatorname{disc}(T^3 - 12T + 2) = 6804 = 2^2 \cdot 3^5 \cdot 7 = [\mathcal{O}_K : \mathbf{Z}[\alpha]]^2 \operatorname{disc}(K),$$

so $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ divides $2 \cdot 3^2$ and 7 divides $\operatorname{disc}(K)$.

Coset representatives for $\frac{1}{2}\mathbf{Z}[\alpha]/\mathbf{Z}[\alpha]$ have the form $\frac{a}{2} + \frac{b}{2}\alpha + \frac{c}{2}\alpha^2$ where $a, b, c \in \{0, 1\}$. The nonzero coset representatives do not have characteristic polynomial in $\mathbf{Z}[T]$, so $2 \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$ and $4 \mid \operatorname{disc}(K)$. Among the nonzero coset representatives for $\frac{1}{3}\mathbf{Z}[\alpha]/\mathbf{Z}[\alpha]$ there is an algebraic integer: $\beta = \frac{1}{3}(1 + \alpha + \alpha^2)$. To show this is an algebraic integer, we compute its characteristic polynomial in the extension $K/\mathbf{Q}$. The matrix for multiplication by β on K with respect to the basis $\{1, \alpha, \alpha^2\}$ is

$$\begin{pmatrix} 1/3 & -2/3 & -2/3 \\ 1/3 & 13/3 & 10/3 \\ 1/3 & 1/3 & 13/3 \end{pmatrix},$$

whose characteristic polynomial is $T^3 - 9T^2 + 21T - 7$. This is monic with integral coefficients, so $\beta \in \mathcal{O}_K$.

Since $[K : \mathbf{Q}]$ is prime and $\beta \notin \mathbf{Q}$, $K = \mathbf{Q}(\beta)$.[3] Might $\mathcal{O}_K = \mathbf{Z}[\beta]$? Let's compute $\operatorname{disc}(\mathbf{Z}[\beta])$. This ring discriminant is the polynomial discriminant of $T^3 - 9T^2 + 21T - 7$, but this polynomial is not a trinomial so our explicit polynomial discriminant formulas don't directly apply to it. However, if we replace T with $T + c$ in the polynomial we don't change the polynomial's discriminant (Exercise 3.5) and at $c = 3$ the polynomial becomes $T^3 - 6T + 2$, so

$$\operatorname{disc}(\mathbf{Z}[\beta]) = \operatorname{disc}(T^3 - 6T + 2) = 756 = 2^2 \cdot 3^3 \cdot 7 = [\mathcal{O}_K : \mathbf{Z}[\beta]]^2 \operatorname{disc}(K).$$

We already know $\operatorname{disc}(K)$ is divisible by 4 and 7, so $[\mathcal{O}_K : \mathbf{Z}[\beta]]$ is 1 or 3. A computation of characteristic polynomials shows no nonzero coset representative of $\frac{1}{3}\mathbf{Z}[\beta]/\mathbf{Z}[\beta]$ is an algebraic integer, so $\mathcal{O}_K = \mathbf{Z}[\beta]$ and $\operatorname{disc}(K) = 2^2 \cdot 3^3 \cdot 7 = 756$.

[3] This implies we must be able to write α in terms of β. Explicitly, $\alpha = \beta^2 - 5\beta + 2$.

Example 3.49. The polynomial $f(T) = T^4 + 2T^2 + 3T + 1$ is irreducible in $\mathbf{Q}[T]$ (check mod 2). Let $K = \mathbf{Q}(\alpha)$, where $f(\alpha) = 0$. What is $\operatorname{disc}(\mathbf{Z}[\alpha]) = \operatorname{disc}(f(T))$? We can't get $f(T+c)$ to be a trinomial for any c, so we compute $\operatorname{disc}(\mathbf{Z}[\alpha])$ using the norm formula in Theorem 3.25:

$$\operatorname{disc}(\mathbf{Z}[\alpha]) = (-1)^{4\cdot 3/2}\operatorname{N}_{K/\mathbf{Q}}(f'(\alpha)) = \operatorname{N}_{K/\mathbf{Q}}(4\alpha^3 + 4\alpha + 3).$$

To find this norm, we compute the matrix for multiplication by $f'(\alpha)$ with respect to the basis $\{1, \alpha, \alpha^2, \alpha^3\}$. It is

$$\begin{pmatrix} 3 & -4 & 0 & 4 \\ 4 & -9 & -4 & 12 \\ 0 & -4 & -9 & 4 \\ 4 & 0 & -4 & -9 \end{pmatrix},$$

whose determinant is $117 = 3^2 \cdot 13 = [\mathcal{O}_K : \mathbf{Z}[\alpha]]^2 \operatorname{disc}(K)$. Thus $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ is 1 or 3. No nonzero coset representative of $\frac{1}{3}\mathbf{Z}[\alpha]/\mathbf{Z}[\alpha]$ is an algebraic integer, so $\mathcal{O}_K = \mathbf{Z}[\alpha]$ and $\operatorname{disc}(K) = 3^2 \cdot 13 = 117$.

Example 3.50. Let $K = \mathbf{Q}(\sqrt[3]{10})$, so

$$\operatorname{disc}(\mathbf{Z}[\sqrt[3]{10}]) = -\operatorname{N}_{K/\mathbf{Q}}(3\sqrt[3]{10}^2) = -2700 = -2^2 \cdot 3^3 \cdot 5^2.$$

No nonzero coset representatives for $\frac{1}{2}\mathbf{Z}[\sqrt[3]{10}]/\mathbf{Z}[\sqrt[3]{10}]$ or $\frac{1}{5}\mathbf{Z}[\sqrt[3]{10}]/\mathbf{Z}[\sqrt[3]{10}]$ are algebraic integers, so 4 and 25 divide $\operatorname{disc}(K)$. Looking at nonzero coset representatives for $\frac{1}{3}\mathbf{Z}[\sqrt[3]{10}]/\mathbf{Z}[\sqrt[3]{10}]$, we find the algebraic integer $\alpha = \frac{1}{3} + \frac{1}{3}\sqrt[3]{10} + \frac{1}{3}\sqrt[3]{100}$, whose characteristic polynomial is $T^3 - T^2 - 3T - 3$. Since

$$\operatorname{disc}(\mathbf{Z}[\alpha]) = -300 = -2^2 \cdot 3 \cdot 5^2 = [\mathcal{O}_K : \mathbf{Z}[\alpha]]^2 \operatorname{disc}(K),$$

which is divisible by 3 just once, $3 \mid \operatorname{disc}(K)$, so $\mathcal{O}_K = \mathbf{Z}[\alpha]$ and $\operatorname{disc}(K) = -2^2 \cdot 3 \cdot 5^2 = -300$.

Example 3.51. Let $K = \mathbf{Q}(\sqrt{d})$ for a squarefree integer d. We will use the discriminant to compute $\mathcal{O}_K$. Compare the method with that of Theorem 1.20.

By Example 3.21, $\operatorname{disc}(\mathbf{Z}[\sqrt{d}]) = \operatorname{disc}_{K/\mathbf{Q}}(1, \sqrt{d}) = 4d$. (This can also be computed as $\operatorname{disc}(T^2 - d)$.) Since d is squarefree, the largest square factor of $4d$ is 4 (even if d is even), so $[\mathcal{O}_K : \mathbf{Z}[\sqrt{d}]] \mid 2$. Therefore $\mathbf{Z}[\sqrt{d}]$ has index 1 or 2 in

$\mathcal{O}_K$, which implies $2\mathcal{O}_K \subset \mathbf{Z}[\sqrt{d}]$, so

$$\mathbf{Z}[\sqrt{d}] \subset \mathcal{O}_K \subset \frac{1}{2}\mathbf{Z}[\sqrt{d}].$$

Coset representatives for $\frac{1}{2}\mathbf{Z}[\sqrt{d}]/\mathbf{Z}[\sqrt{d}]$ are $\frac{a}{2} + \frac{b}{2}\sqrt{d}$ where $a, b \in \{0, 1\}$. The three nonzero coset representatives are $\frac{1}{2}$, $\frac{1}{2}\sqrt{d}$, and $\frac{1+\sqrt{d}}{2}$. If $\mathcal{O}_K \neq \mathbf{Z}[\sqrt{d}]$ one of these coset representatives is an algebraic integer.

The first and second numbers are not algebraic integers. We are left to check $\frac{1+\sqrt{d}}{2}$. Its norm to $\mathbf{Q}$ is $\frac{1-d}{4}$, which is not an integer if $d \not\equiv 1 \bmod 4$, so for $d \not\equiv 1 \bmod 4$ we get $\mathcal{O}_K = \mathbf{Z}[\sqrt{d}]$. If $d \equiv 1 \bmod 4$ then $\frac{1+\sqrt{d}}{2}$ is an algebraic integer (we've already seen that), so $[\mathcal{O}_K : \mathbf{Z}[\sqrt{d}]] = 2$. Therefore $\mathbf{Z}[\sqrt{d}] \subsetneq \mathbf{Z} + \mathbf{Z}\frac{1+\sqrt{d}}{2} \subset \mathcal{O}_K$, so $\mathcal{O}_K = \mathbf{Z} + \mathbf{Z}\frac{1+\sqrt{d}}{2}$.

In Table 3.2 we list all the number fields whose discriminants we have computed. The even discriminants are all multiples of 4 and the odd discriminants are all 1 mod 4.

Example	3.33	3.45	3.46	3.47	3.48	3.49	3.50
disc(K)	-108	-23	2869	-107	756	117	-300
mod 4	0	1	1	1	0	1	0

Table 3.2: Discriminants mod 4

Theorem 3.52 (Stickelberger, 1897). *For any number field K,* $\operatorname{disc}(K) \equiv 0, 1 \bmod 4$.

Proof. (Schur) Let $\alpha_1, \ldots, \alpha_n$ be a $\mathbf{Z}$-basis of $\mathcal{O}_K$, so $\operatorname{disc}(K) = \det(\sigma_i(\alpha_j))^2$ in the notation of Lemma 3.24. The determinant $\det(\sigma_i(\alpha_j))$ can be computed as a sum over permutations in S_n: $\det(\sigma_i(\alpha_j)) = P - N$ where P is a sum over even permutations π and N is a sum (without overall sign) over odd permutations π of products $\sigma_1(\alpha_{\pi(1)}) \cdots \sigma_n(\alpha_{\pi(n)})$. Squaring,

$$\operatorname{disc}(K) = (P - N)^2 = (P + N)^2 - 4PN.$$

The sum $P + N$ and product PN are symmetric in $\alpha_1, \ldots, \alpha_n$, so they are rational by Galois theory. They are both algebraic integers, so they are in $\mathbf{Z}$. Therefore $\operatorname{disc}(K)$ mod 4 is congruent to a square and the squares mod 4 are 0 and 1. ■

Stickelberger's theorem is a good way to remember the discriminant of a quadratic field (Example 3.32). For squarefree d, pretend the discriminant of $\mathbf{Q}(\sqrt{d})$ wants to be d. When $d \equiv 1 \bmod 4$, this is consistent with Stickelberger's theorem and it's right. When $d \equiv 2, 3 \bmod 4$ the discriminant could not be d because that would violate Stickelberger's theorem, so use $4d$ to be consistent with the theorem.

Example 3.53. Let $K = \mathbf{Q}(\alpha)$, where $\alpha^3 + \alpha + 4 = 0$. Then

$$\operatorname{disc}(\mathbf{Z}[\alpha]) = \operatorname{disc}(T^3 + T + 4) = -436 = -4 \cdot 109.$$

If $\mathcal{O}_K \neq \mathbf{Z}[\alpha]$ then $\operatorname{disc}(K) = -109$, but $-109 \equiv 3 \bmod 4$. Therefore $\mathcal{O}_K = \mathbf{Z}[\alpha]$.

Remark 3.54. There is a more profound theorem in algebraic number theory that is also called Stickelberger's theorem. See [26, Chap. 13].

The discriminant is important for both a computational reason and a conceptual reason. Computationally, the discriminant of a $\mathbf{Z}$-lattice in K helps us compute the ring of integers of K, as we have already seen. Conceptually, the prime factors of $\operatorname{disc}(K)$ provide important information about prime ideals in $\mathcal{O}_K$, which we'll discuss in Chapter 6.

The PARI commands for finding the discriminant of a polynomial and the discriminant of a number field (generated over $\mathbf{Q}$ by the root of a polynomial) are `poldisc` and `nfdisc`. For example, `poldisc(x^3-x - 4)` has answer `-428` and `nfdisc(x^3 - x - 4)` has answer `-107`, which we found in Example 3.47. (**Warning**. If you compute discriminants in a computer algebra system other than PARI, be sure you know the meaning of the discriminant command you choose. You don't want to think you're computing a number field discriminant but actually be computing a polynomial discriminant.)

3.6 Discriminant over a PID

Consider the field $K = \mathbf{Q}(i, \sqrt{-5})$. What is $\mathcal{O}_K$?

$$\begin{tikzcd} & K \arrow[dl, dash] \arrow[dr, dash] & \\ \mathbf{Q}(i) \arrow[dr, dash] & & \mathbf{Q}(\sqrt{-5}) \arrow[dl, dash] \\ & \mathbf{Q} & \end{tikzcd} \qquad \begin{tikzcd} & \mathcal{O}_K \arrow[dl, dash] \arrow[dr, dash] & \\ \mathbf{Z}[i] \arrow[dr, dash] & & \mathbf{Z}[\sqrt{-5}] \arrow[dl, dash] \\ & \mathbf{Z} & \end{tikzcd}$$

Methods we have discussed so far would suggest thinking of $\mathcal{O}_K$ as a $\mathbf{Z}$-module and trying to determine a $\mathbf{Z}$-basis, which has 4 elements in it. However, we can consider K not only as a quartic extension of $\mathbf{Q}$, but as a quadratic extension of either $\mathbf{Q}(i)$ or $\mathbf{Q}(\sqrt{-5})$ (or $\mathbf{Q}(\sqrt{5})$). Since $\mathcal{O}_K$ contains $\mathbf{Z}[i]$ and $\mathbf{Z}[\sqrt{-5}]$, we can view $\mathcal{O}_K$ as a $\mathbf{Z}[i]$-module or as a $\mathbf{Z}[\sqrt{-5}]$-module, and can anticipate $\mathcal{O}_K$ should have something like rank 2 over these rings. If we could get this idea to work it might simplify calculations a lot, since a rank 2 module is easier to investigate than a rank 4 module.

Our study of the additive structure of rings of integers exploited properties of finite free abelian groups. Almost the whole theory of finite free abelian groups carries over to finite free modules over a PID. Since $\mathbf{Z}[i]$ is a PID and $\mathbf{Z}[\sqrt{-5}]$ is not, it's simpler to think about $\mathcal{O}_K$ as a $\mathbf{Z}[i]$-module. Is it a free module?

This question motivates us to consider the setup in the diagram below.

$$\begin{tikzcd} & E \arrow[dd, dash, "n"] \\ B \arrow[ur, dash] \arrow[dd, dash] & \\ & F \\ A \arrow[ur, dash] & \end{tikzcd} \tag{3.17}$$

Here A is a PID, F is the fraction field of A, E/F is a finite *separable* extension of degree n, and B is the integral closure of A in E. We will use discriminants to show B is a free A-module of rank n and find a squarefree criterion for an F-basis of E in B to be an A-basis of B, generalizing Corollary 3.44. The application we have in mind is to the top part of the following diagram.

$$\begin{tikzcd}
 & K \arrow[ld, dash] \arrow[dd, dash, "2"] \\
\mathcal{O}_K \arrow[dd, dash] & \\
 & \mathbf{Q}(i) \arrow[ld, dash] \arrow[dd, dash, "2"] \\
\mathbf{Z}[i] \arrow[dd, dash] & \\
 & \mathbf{Q} \arrow[ld, dash] \\
\mathbf{Z} &
\end{tikzcd}$$

Studying number fields over an intermediate subfield and not directly over $\mathbf{Q}$ is called the relative case. Working over $\mathbf{Q}$ is called the absolute case. This section will sketch some ideas for the relative case and then apply them to this example.

When E/F is separable, for any F-basis $\{e_1, \dots, e_n\}$ of E we have

$$\operatorname{disc}_{E/F}(e_1, \dots, e_n) = \det(\operatorname{Tr}_{E/F}(e_i e_j)) \in F - \{0\}.$$

Every element of E is a ratio b/a where $b \in B$ and $a \in A$ (generalization of Theorem 1.13), so by scaling there is an F-basis $\{e_1, \dots, e_n\}$ of E inside of B. Since A is integrally closed, $\operatorname{Tr}_{E/F}(e_i e_j) \in A$, so $\operatorname{disc}_{E/F}(e_1, \dots, e_n) \in A - \{0\}$.

Since $\sum_{i=1}^n Ae_i = \bigoplus_{i=1}^n Ae_i \subset B$, B contains an A-module that is finite free of rank n. We also want to place B inside a finite free A-module of rank n. For this we use discriminants, just like when $A = \mathbf{Z}$. For any $b \in B$, write $b = c_1 e_1 + \cdots + c_n e_n$ $(c_i \in F)$. Then $\operatorname{Tr}_{E/F}(e_i b) = c_1 \operatorname{Tr}_{E/F}(e_i e_1) + \cdots + c_n \operatorname{Tr}_{E/F}(e_i e_n)$ for all i, which is a system of n equations in $c_1, \dots, c_n$. Using Cramer's rule as in the proof of Theorem 3.2, we can write $c_i = a_i/d$ where $a_i \in A$ and $d = \det(\operatorname{Tr}_{E/F}(e_i e_j))$. The number d is nonzero because E/F is separable (Theorem 3.22). Therefore

$$\boxed{\sum_{i=1}^n Ae_i \subset B \subset \sum_{i=1}^n A\frac{e_i}{d}, \quad d = \det(\operatorname{Tr}_{E/F}(e_i e_j)),}$$

so Corollary 8.29 tells us $B \cong A^n$ as an A-module, which means B has a basis as an A-module. Just as in the case $A = \mathbf{Z}$, $\operatorname{disc}_{E/F}(e_1, \dots, e_n)$ is a common denominator for all numbers in B when written as F-linear combinations of the e_i's.

Returning to the intended application in $K = \mathbf{Q}(i, \sqrt{-5})$, we are now assured that $\mathcal{O}_K \cong \mathbf{Z}[i]^2$ as a $\mathbf{Z}[i]$-module. That is, we can write $\mathcal{O}_K = \mathbf{Z}[i]x_1 \oplus \mathbf{Z}[i]x_2$, but the general theory doesn't tell us x_1 and x_2. To find x_1 and x_2 we will generalize discriminants of $\mathbf{Z}$-lattices to discriminants of lattices over a PID.

Definition 3.55. Let A be a PID with fraction field F and V be a finite-dimensional F-vector space. An *A-lattice* in V is a free A-module in V of rank $\dim_F(V)$.[4]

Concretely, an A-lattice is simply the A-span (rather than the F-span) of some basis of V, but our definition is a basis-free description. We will be concerned with A-lattices in field extensions of F, where we can multiply and not just form linear combinations. (For an interesting aspect of lattices in a pure vector space, see Exercise 3.18.)

Example 3.56. Some $\mathbf{Z}[i]$-lattices in $K = \mathbf{Q}(i, \sqrt{-5})$ are $\mathcal{O}_K$ and $\mathbf{Z}[i, \sqrt{-5}] = \mathbf{Z}[i] + \mathbf{Z}[i]\sqrt{-5}$, which is a $\mathbf{Z}[i]$-sublattice of $\mathcal{O}_K$.

Definition 3.57. When E is a finite extension of F, with degree n, the *discriminant* (or *A-discriminant*) of an A-lattice M in E is

$$\operatorname{disc}_A(M) = \operatorname{disc}_{E/F}(e_1, \ldots, e_n) = \det(\operatorname{Tr}_{E/F}(e_i e_j)) \in F,$$

where $\{e_1, \ldots, e_n\}$ is any A-basis of M.

What we wrote before as $\operatorname{disc}(M)$ for a $\mathbf{Z}$-lattice M is $\operatorname{disc}_{\mathbf{Z}}(M)$ in the notation of this new definition.

Equation (3.10), when applied to two A-bases of M, tells us their discriminants differ by the square of the determinant of the change-of-basis matrix between the A-bases. That determinant is in $A^\times$, so $\operatorname{disc}_A(M)$ is well-defined up to multiplication by a unit square in A. It is nonzero when E/F is separable.

To use A-discriminants in order to compute rings of integers, we want a generalization of the formula in Theorem 3.42:

$$\operatorname{disc}(M_1) = [M_2 : M_1]^2 \operatorname{disc}(M_2)$$

where M_1 and M_2 are $\mathbf{Z}$-lattices in a number field and $M_1 \subset M_2$. We know how to generalize $\operatorname{disc}(M_i)$ to discriminants of A-lattices. What is the A-lattice analogue of the index of one $\mathbf{Z}$-lattice inside another?

[4] A-linear independence in V implies F-linear independence since an F-linear dependence relation can be scaled to an A-linear dependence relation, so an A-module in V can have at most $\dim_F(V)$ linearly independent elements over A.

For **Z**-lattices M_1 and M_2, the index $[M_2 : M_1]$ is the size of the quotient group M_2/M_1. When M_1 and M_2 are A-lattices, we want a concept of "A-size" for the quotient module M_2/M_1. The A-module analogue of a finite abelian group is a finitely generated torsion A-module. (Finitely generated torsion **Z**-modules are the same thing as finite abelian groups.) Let's check M_2/M_1 is such an A-module in cases of interest.

Theorem 3.58. *If M_1 and M_2 are finite free A-modules of the same rank with $M_1 \subset M_2$, then M_2/M_1 is a finitely generated torsion A-module.*

Proof. Using determinants, we will find a single $d \in A - \{0\}$ such that $dM_2 \subset M_1$, making M_2/M_1 a torsion A-module. Let $\{x_1, \dots, x_n\}$ be an A-basis of M_1 and $\{y_1, \dots, y_n\}$ be an A-basis of M_2, so $\sum_{i=1}^n Ax_i \subset \sum_{j=1}^n Ay_j$. Write $x_i = \sum_{j=1}^n a_{ij}y_j$. View M_2 inside F^n by identifying the y_j's with the standard basis of F^n. By Cramer's rule, $y_j \in \sum_{i=1}^n Ax_i/d$, where $d = \det(a_{ij}) \in A$. This determinant d is not 0 since the x_i's and y_j's are both bases of F^n. Since $dy_j \in M_1$ for all j, $dM_2 \subset M_1$.

This theorem can also be proved using the alignment theorem for a finite free module over a PID and a submodule (Theorem 8.33): we can write $M_2 = \bigoplus_{i=1}^n Ae_i$ and $M_1 = \bigoplus_{i=1}^n Aa_ie_i$ for some basis $\{e_i\}$ and nonzero scalars $\{a_i\}$. Then obviously $a_1 \cdots a_n M_2 \subset M_1$. This is essentially the same proof as the first one, since $a_1 \cdots a_n = d$ up to unit multiple (generalize Theorem 3.10, which identifies an index with a determinant up to sign when $A = \mathbf{Z}$). ■

Definition 3.59. Let T be a finitely generated torsion A-module. By the structure theorem for finitely generated torsion modules over a PID (Corollary 8.36),

$$T \cong A/a_1A \oplus \cdots \oplus A/a_kA$$

where the a_i's are nonzero. The *A-cardinality* of T is defined to be the principal ideal

$$\mathrm{card}_A(T) = (a_1a_2 \cdots a_k).$$

If M_1 and M_2 are two finite free A-modules with equal rank and $M_1 \subset M_2$, we set the *A-index* of M_1 in M_2 to be the A-cardinality of their quotient module:

$$[M_2 : M_1]_A = \mathrm{card}_A(M_2/M_1).$$

If $A = \mathbf{Z}$ then T is a finite abelian group and $\mathrm{card}_A(T) = \#T\mathbf{Z}$, so it is essentially the size of T. Since the A-cardinality of T is an ideal in A, a generator of it does not have a combinatorial meaning when $A \neq \mathbf{Z}$ (even when a positive integer is a generator).

Example 3.60. Let $A = \mathbf{Z}[i]$,

$$M_1 = \mathbf{Z}[i]^2 = \mathbf{Z}[i]\begin{pmatrix}1\\0\end{pmatrix} + \mathbf{Z}[i]\begin{pmatrix}0\\1\end{pmatrix},$$

and

$$M_2 = \mathbf{Z}[i]\begin{pmatrix}3\\0\end{pmatrix} + \mathbf{Z}[i]\begin{pmatrix}0\\1+2i\end{pmatrix} = \mathbf{Z}[i]3\begin{pmatrix}1\\0\end{pmatrix} + \mathbf{Z}[i](1+2i)\begin{pmatrix}0\\1\end{pmatrix}.$$

Since $M_2/M_1 \cong \mathbf{Z}[i]/(3) \oplus \mathbf{Z}[i]/(1+2i)$, $[M_2 : M_1]_{\mathbf{Z}[i]} = (3(1+2i))$.

Using different bases for these two modules,

$$M_1 = \mathbf{Z}[i]\begin{pmatrix}3\\1+2i\end{pmatrix} + \mathbf{Z}[i]\begin{pmatrix}1-2i\\2\end{pmatrix}$$

and

$$M_2 = \mathbf{Z}[i]\begin{pmatrix}3\\1+2i\end{pmatrix} + \mathbf{Z}[i]3(1+2i)\begin{pmatrix}1-2i\\2\end{pmatrix},$$

so $M_2/M_1 \cong \mathbf{Z}[i]/(3(1+2i))$. Thus again we compute $[M_2 : M_1]_{\mathbf{Z}[i]} = (3(1+2i))$.

The following theorem addresses the well-defineness of $\mathrm{card}_A(T)$.

Theorem 3.61. *If $T \cong A/a_1A \oplus \cdots \oplus A/a_kA$, the product $a_1 \cdots a_k$ as a function of T is well-defined up to unit multiple, so the ideal $\mathrm{card}_A(T)$ is well-defined.*

Proof. A prime π divides $a_1 \cdots a_k$ if and only if $\pi T \neq T$, so the prime factors of $a_1 \cdots a_k$ are determined (up to unit multiple) by T. For any prime π, write $a_i = \pi^{e_i} a_i'$ with $e_i \geqslant 0$ and a_i' not divisible by π. The individual e_i's are not intrinsic to T, but the power of π in $a_1 \cdots a_k$ has multiplicity $e_1 + \cdots + e_k$ and this exponent is intrinsic to T since you can check in Exercise 3.19 that

$$e_1 + \cdots + e_k = \sum_{j \geqslant 0} \dim_{A/\pi A}(\pi^j T/\pi^{j+1} T).$$

The sum is finite since $\pi^j T = 0$ for $j \geqslant \max(e_1, \ldots, e_k)$. ■

We have $\mathrm{card}_A(T) = (1)$ only when each a_i is a unit, which means $T = 0$.

It is sometimes convenient to work with $\mathrm{card}_A(T)$ as an element of A. Then $\mathrm{card}_A(T)$ is well-defined only up to unit multiple. For example, in Example 3.60 we could say $[M_2 : M_1]_{\mathbf{Z}[i]} = \mathrm{card}_{\mathbf{Z}[i]}(M_2/M_1) = 3(1+2i)$. Here is a theorem where A-cardinalities are elements.

Theorem 3.62. *If $M_1 \subset M_2$ are two A-lattices in E, then*

$$\operatorname{disc}_A(M_1) = [M_2 : M_1]_A^2 \operatorname{disc}_A(M_2),$$

where $[M_2 : M_1]_A = \operatorname{card}_A(M_2/M_1)$ is an element of A and the equation is true up to unit multiple.

Proof. Proceed as in the special case $A = \mathbf{Z}$ (Theorem 3.42). Use aligned bases for a finite free A-module and submodule of the same rank (Theorem 8.33). ■

Corollary 3.63. *With the notation of (3.17), if $\{e_1, \ldots, e_n\}$ is a basis for E/F lying in B and $\operatorname{disc}_A(\sum_{i=1}^n Ae_i)$ is squarefree*[5]*, then*

$$B = \sum_{i=1}^n Ae_i = \bigoplus_{i=1}^n Ae_i.$$

Proof. Since $Ae_1 + \cdots + Ae_n \subset B$, Theorem 3.62 implies

$$\operatorname{disc}_A\left(\sum_{i=1}^n Ae_i\right) = \left[B : \sum_{i=1}^n Ae_i\right]_A^2 \operatorname{disc}_A(B),$$

so the squarefree assumption implies $[B : Ae_1 + \cdots + Ae_n]_A$ is a unit. Thus $B/(Ae_1 + \cdots + Ae_n) = 0$, so

$$B = \sum_{i=1}^n Ae_i. \qquad ■$$

Let's apply these ideas to $K = \mathbf{Q}(i, \sqrt{-5})$ as an extension of $\mathbf{Q}(i)$. Since $\sqrt{-5} \in \mathcal{O}_K$, is $\{1, \sqrt{-5}\}$ a $\mathbf{Z}[i]$-basis of $\mathcal{O}_K$? Set

$$M = \mathbf{Z}[i] + \mathbf{Z}[i]\sqrt{-5} \subset \mathcal{O}_K$$

and compute its $\mathbf{Z}[i]$-discriminant. In the basis $\{1, \sqrt{-5}\}$ of $K/\mathbf{Q}(i)$,

$$[m_{x+y\sqrt{-5}}] = \begin{pmatrix} x & -5y \\ y & x \end{pmatrix} \Longrightarrow \operatorname{Tr}_{K/\mathbf{Q}(i)}(x + y\sqrt{-5}) = 2x.$$

[5] An element of A is called squarefree when it has no multiple prime factors. Divisibility by unit squares doesn't matter.

Adopting Tr as shorthand for $\mathrm{Tr}_{K/\mathbf{Q}(i)}$,

$$\begin{aligned}\mathrm{disc}_{\mathbf{Z}[i]}(\mathbf{Z}[i]+\mathbf{Z}[i]\sqrt{-5}) &= \det\left(\begin{smallmatrix}\mathrm{Tr}(1) & \mathrm{Tr}(\sqrt{-5})\\ \mathrm{Tr}(\sqrt{-5}) & \mathrm{Tr}(-5)\end{smallmatrix}\right)\\ &= \det\left(\begin{smallmatrix}2 & 0\\ 0 & -10\end{smallmatrix}\right)\\ &= -20\\ &= (1+i)^4(1+2i)(1-2i).\end{aligned}$$

We factored -20 into nonassociate primes in $\mathbf{Z}[i]$, not $\mathbf{Z}$. We know by Theorem 3.62 that

$$\mathrm{disc}_{\mathbf{Z}[i]}(\mathbf{Z}[i]+\mathbf{Z}[i]\sqrt{-5}) = [\mathcal{O}_K : M]^2_{\mathbf{Z}[i]}\,\mathrm{disc}_{\mathbf{Z}[i]}(\mathcal{O}_K),$$

so

$$(1+i)^4(1+2i)(1-2i) = [\mathcal{O}_K : M]^2_{\mathbf{Z}[i]}\,\mathrm{disc}_{\mathbf{Z}[i]}(\mathcal{O}_K).$$

By unique factorization in $\mathbf{Z}[i]$, $[\mathcal{O}_K : M]_{\mathbf{Z}[i]}$ divides $(1+i)^2$, so $(1+i)^2(\mathcal{O}_K/M) = 0$,[6] which means $(1+i)^2\mathcal{O}_K \subset M$. Since $(1+i)^2 = 2i$, we can write this more simply as $2\mathcal{O}_K \subset M$, so

$$M \subset \mathcal{O}_K \subset \frac{1}{2}M.$$

To determine how much larger $\mathcal{O}_K$ is than M (if it is larger at all), we list coset representatives for $\frac{1}{2}M/M$ and see which nonzero representatives are algebraic integers. Explicitly,

$$\frac{1}{2}M/M = \left(\mathbf{Z}[i]\frac{1}{2}+\mathbf{Z}[i]\frac{\sqrt{-5}}{2}\right)/(\mathbf{Z}[i]+\mathbf{Z}[i]\sqrt{-5})$$

has coset representatives $\frac{\alpha}{2}+\frac{\beta\sqrt{-5}}{2}$ where $\alpha,\beta \in \{0,1,i,1+i\}$ (representatives for $\mathbf{Z}[i]/(2)$). An algebraic integer occurs at $\alpha = i$, $\beta = 1$: $\frac{i+\sqrt{-5}}{2}$ is a root of $T^2 - iT + 1$, so it is integral over $\mathbf{Z}[i]$ and over $\mathbf{Z}$ (transitivity of integrality).

Let's now try instead $\{1, \frac{i+\sqrt{-5}}{2}\}$ as a new potential $\mathbf{Z}[i]$-basis of $\mathcal{O}_K$: setting

$$N = \mathbf{Z}[i] + \mathbf{Z}[i]\frac{i+\sqrt{-5}}{2} \subset \mathcal{O}_K,$$

[6] For a finitely generated torsion A-module T, $\mathrm{card}_A(T)t = 0$ for all $t \in T$ by the definition of $\mathrm{card}_A(T)$. Therefore $[M_2 : M_1]_A(M_2/M_1) = 0$, so $[M_2 : M_1]_A M_2 \subset M_1$.

the $\mathbf{Z}[i]$-discriminant of N is

$$\begin{aligned}\operatorname{disc}_{\mathbf{Z}[i]}(N) &= \det\begin{pmatrix} \operatorname{Tr}(1) & \operatorname{Tr}(\frac{i+\sqrt{-5}}{2}) \\ \operatorname{Tr}(\frac{i+\sqrt{-5}}{2}) & \operatorname{Tr}((\frac{i+\sqrt{-5}}{2})^2)\end{pmatrix} \\ &= \left(\begin{smallmatrix} 2 & i \\ i & -3\end{smallmatrix}\right) \\ &= -6 - i^2 \\ &= -5 \\ &= -(1+2i)(1-2i),\end{aligned}$$

which is squarefree (!) in $\mathbf{Z}[i]$, so Corollary 3.63 says

$$\mathcal{O}_K = N = \mathbf{Z}[i] + \mathbf{Z}[i]\frac{i+\sqrt{-5}}{2} = \mathbf{Z} + \mathbf{Z}i + \mathbf{Z}\frac{i+\sqrt{-5}}{2} + \mathbf{Z}\frac{-1+i\sqrt{-5}}{2}.$$

By a relative computation over $\mathbf{Z}[i]$ we have computed a $\mathbf{Z}[i]$-basis of $\mathcal{O}_K$ and then expanded that out to a $\mathbf{Z}$-basis.

Using the $\mathbf{Z}$-basis of $\mathcal{O}_K$, check as Exercise 3.11 that $\operatorname{disc}(K) = 400$ and $\operatorname{disc}(\mathbf{Z}[\frac{i+\sqrt{-5}}{2}]) = 400$, so $\mathcal{O}_K = \mathbf{Z}[\frac{i+\sqrt{-5}}{2}]$.

3.7 Discriminant of a Ring Extension with Basis

Section 3.6 showed the usefulness of having a discriminant concept over PIDs besides $\mathbf{Z}$. Let's formulate this concept over more general commutative rings. (This material will not be needed until Chapter 6.)

Characteristic polynomials, traces, and norms make sense not just for a finite field extension E/F but for any ring extension B/A where B is a finite free A-module: using an A-basis of B, for each $b \in B$ the multiplication-by-b map $m_b \colon B \to B$ is A-linear and can be turned into a matrix whose characteristic polynomial, trace, and determinant are independent of the basis. We set

$$\chi_{B/A,b}(T) = \det(TI_n - [m_b]) \in A[T], \quad \operatorname{Tr}_{B/A}(b) = \operatorname{Tr}(m_b), \quad \mathrm{N}_{B/A}(b) = \det(m_b).$$

The trace $\operatorname{Tr}_{B/A} \colon B \to A$ is A-linear and the norm $\mathrm{N}_{B/A} \colon B \to A$ is multiplicative.

Three basic examples of this situation other than field extensions are

(1) $\mathcal{O}_K$ as a $\mathbf{Z}$-module,

(2) $\mathcal{O}_K/m\mathcal{O}_K$ as a $\mathbf{Z}/m\mathbf{Z}$-module for $m \geqslant 2$,

(3) $F[X]/(f(X))$ as an F-vector space, where $f(X)$ is nonconstant in $F[X]$.

In the third example, one basis of $F[X]/(f(X))$ over F is the power basis $\{1, X, \ldots, X^{n-1}\}$, where $n = \deg f$. We allow reducible $f(X)$.

Example 3.64. The ring $\mathbf{Z}[i] = \mathbf{Z} + \mathbf{Z}i$ is a free $\mathbf{Z}$-module with basis $\{1, i\}$. In this basis, $[m_{2+3i}] = \left(\begin{smallmatrix} 2 & -3 \\ 3 & 2 \end{smallmatrix}\right)$, so

$$\chi_{\mathbf{Z}[i]/\mathbf{Z},2+3i}(T) = T^2 - 4T + 13, \quad \mathrm{Tr}_{\mathbf{Z}[i]/\mathbf{Z}}(2+3i) = 4, \text{ and } \mathrm{N}_{\mathbf{Z}[i]/\mathbf{Z}}(2+3i) = 13.$$

These are the same computations we would find for $2+3i$ in the field extension $\mathbf{Q}(i)/\mathbf{Q}$.

Example 3.65. The ring $F[\varepsilon]$ of dual numbers over F, where $\varepsilon^2 = 0$, has F-basis $\{1, \varepsilon\}$. Using this basis, $[m_{1+\varepsilon}] = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$, so

$$\chi_{F[\varepsilon]/F,1+\varepsilon}(T) = (T-1)^2, \quad \mathrm{Tr}_{F[\varepsilon]/F}(1+\varepsilon) = 2, \text{ and } \mathrm{N}_{F[\varepsilon]/F}(1+\varepsilon) = 1.$$

Since $F[\varepsilon] \cong F[X]/(X^2)$, this is an instance of the third basic example where $f(X)$ is reducible.

Theorem 3.66. *When K is a number field and $\alpha \in \mathcal{O}_K$,*

$$\chi_{K/\mathbf{Q},\alpha}(T) = \chi_{\mathcal{O}_K/\mathbf{Z},\alpha}(T), \quad \mathrm{Tr}_{K/\mathbf{Q}}(\alpha) = \mathrm{Tr}_{\mathcal{O}_K/\mathbf{Z}}(\alpha), \quad \mathrm{N}_{K/\mathbf{Q}}(\alpha) = \mathrm{N}_{\mathcal{O}_K/\mathbf{Z}}(\alpha).$$

Proof. A $\mathbf{Z}$-basis of $\mathcal{O}_K$ is also a $\mathbf{Q}$-basis of K, so a matrix representation of $m_\alpha \colon \mathcal{O}_K \to \mathcal{O}_K$ with respect to a $\mathbf{Z}$-basis is also a matrix representation of the $\mathbf{Q}$-linear map $m_\alpha \colon K \to K$. Therefore $\chi_{K/\mathbf{Q},\alpha}(T) = \chi_{\mathcal{O}_K/\mathbf{Z},\alpha}(T)$, and the equality of traces and norms follows by equating appropriate coefficients. ■

Remark 3.67. If B/A and B'/A are ring extensions and finite free A-modules and there is an A-linear ring isomorphism $\varphi \colon B \to B'$, then $[m_x] = [m_{\varphi(x)}]$ for all $x \in B$ if we use an A-basis of B and its φ-values as the A-basis of B'. Therefore $\chi_{B/A,x}(T) = \chi_{B'/A,\varphi(x)}(T)$, $\mathrm{Tr}_{B/A}(x) = \mathrm{Tr}_{B'/A}(\varphi(x))$, and $\mathrm{N}_{B/A}(x) = \mathrm{N}_{B'/A}(\varphi(x))$.

Definition 3.68. For any A-basis $\{x_1, \ldots, x_n\}$ of B, its *discriminant* is defined to be

$$\mathrm{disc}_{B/A}(x_1, \ldots, x_n) = \det(\mathrm{Tr}_{B/A}(x_i x_j)) \in A. \tag{3.18}$$

How does the discriminant vary with the basis? If $\{y_1, \dots, y_n\}$ is a second A-basis of B, then

$$\operatorname{disc}_{B/A}(y_1, \dots, y_n) = (\det(c_{ij}))^2 \operatorname{disc}_{B/A}(x_1, \dots, x_n), \tag{3.19}$$

where (c_{ij}) is the matrix expressing the y_i's in terms of the x_i's. The proof of this formula is identical to that of (3.10), which is the special case of a finite field extension. A change-of-basis matrix for a finite free A-module is an invertible matrix over A, and invertible matrices over A have determinant in $A^\times$. So the discriminants of any two A-bases of B differ by a unit square in A. We define $\operatorname{disc}_A(B)$ to mean the discriminant of any A-basis of B; it is well-defined up to a unit square in A. For example, whether or not $\operatorname{disc}_A(B) = 0$ or $\operatorname{disc}_A(B)$ is a square in A are meaningful questions to ask.

Equation (3.19) tells us that the discriminant doesn't depend on the ordering of the basis: a rearrangement of the terms of a basis is described by a change-of-basis matrix that is a permutation matrix, whose determinant is ±1.

Remark 3.69. If B/A and B'/A are ring extensions and finite free A-modules and there is an A-linear ring isomorphism $\varphi\colon B \to B'$, $\operatorname{disc}_{B/A}(x_1, \dots, x_n) = \operatorname{disc}_{B'/A}(\varphi(x_1), \dots, \varphi(x_n))$ since $\operatorname{Tr}_{B/A}(x) = \operatorname{Tr}_{B'/A}(\varphi(x))$ for all $x \in B$.

We have met many contexts for discriminants. It may be convenient to recall them in one place to compare all the definitions.

1. If E/F is a finite extension of fields, the discriminant of an F-basis $\alpha_1, \dots, \alpha_n$ of E is
$$\det(\operatorname{Tr}_{E/F}(\alpha_i, \alpha_j)) \in F.$$
If the basis changes, the discriminant changes by a square in $F^\times$.

2. If $f(T) = (T - \alpha_1) \cdots (T - \alpha_n)$ is a monic polynomial then its discriminant is
$$\prod_{i<j} (\alpha_j - \alpha_i)^2.$$

3. In a number field K, a $\mathbf{Z}$-lattice M with basis $e_1, \dots, e_n$ has discriminant
$$\operatorname{disc}(M) = \det(\operatorname{Tr}_{K/\mathbf{Q}}(e_i e_j)) \in \mathbf{Q}.$$
This is independent of the choice of $\mathbf{Z}$-basis. If $M \subset \mathcal{O}_K$ then $\operatorname{disc}(M) \in \mathbf{Z}$ and we define $\operatorname{disc}(K) := \operatorname{disc}(\mathcal{O}_K)$.

4. Let A be a PID with fraction field F and E/F be a finite field extension. An A-lattice M in E has discriminant
$$\operatorname{disc}_A(M) = \operatorname{disc}_{E/F}(e_1, \ldots, e_n) \in F,$$
where $e_1, \ldots, e_n$ is an A-basis of M. This is well-defined up to multiplication by a unit square in A.

5. Let B/A be an extension of commutative rings such that B is a finite free A-module. If $x_1, \ldots, x_n$ is an A-basis of B, its discriminant is
$$\det(\operatorname{Tr}_{B/A}(x_i x_j)) \in A.$$
Changing the basis changes this by a unit square in A.

These do not exhaust all the variations on discriminants. (See Exercise 3.21, for instance.) Some of these discriminants are visibly special cases of others, but one which looks different from the others is the discriminant of a polynomial. We saw in Theorem 3.25 that the discriminant of a monic irreducible polynomial in $\mathbf{Q}[T]$ is the discriminant of a power basis for the field generated over $\mathbf{Q}$ by a root. The next theorem, whose proof takes up the rest of the section, identifies the discriminant of any (nonconstant) monic polynomial over any field with the discriminant of a power basis of a ring extension, thus putting our last concept of discriminant to work.

Theorem 3.70. *If F is a field and $f(T) \in F[T]$ is monic of degree $n \geqslant 1$,*
$$\begin{aligned}\operatorname{disc}_{(F[T]/(f))/F}(1, T, \ldots, T^{n-1}) &= \operatorname{disc}(f(T)) \\ &= (-1)^{n(n-1)/2}\,\mathrm{N}_{(F[T]/(f))/F}(f'(T)).\end{aligned}$$

When $f(T)$ is reducible or inseparable we can't parametrize the roots of $f(T)$ by embeddings of $F[T]/(f(T))$ into a field, so the proof of Theorem 3.25 doesn't work in this context and we will need new ideas.

Proof. When $n = 1$, the numbers we want to compute all equal 1, so we can assume $n \geqslant 2$. If we replace F with a larger field E, $\{1, T, \ldots, T^{n-1}\}$ is also an E-basis of $E[T]/(f(T))$ and the matrix for multiplication by T^i on $F[T]/(f(T))$ with respect to this basis is also the matrix for multiplication by T^i on $E[T]/(f(T))$ with respect to the same basis viewed over E. Therefore

$\mathrm{Tr}_{(F[T]/(f))/F}(T^i) = \mathrm{Tr}_{(E[T]/(f))/E}(T^i)$ for all $i \geqslant 0$, so

$$\mathrm{disc}_{(F[T]/(f))/F}(1, T, \ldots, T^{n-1}) = \mathrm{disc}_{(E[T]/(f))/E}(1, T, \ldots, T^{n-1}).$$

By replacing F with a splitting field of $f(T)$ over F, we can assume $f(T) = (T - \alpha_1) \cdots (T - \alpha_n)$ in $F[T]$. This greatly simplifies the structure of the ring $F[T]/(f(T))$, and that will help us compute the discriminant.

First suppose the roots α_i are distinct. By the Chinese remainder theorem,

$$F[T]/(f(T)) = F[T]/((T - \alpha_1) \cdots (T - \alpha_n)) \cong \prod_{i=1}^{n} F[T]/(T - \alpha_i) \cong F^n,$$

where F^n on the right side is a product of n copies of F with componentwise ring operations. The isomorphism $F[T]/(f(T)) \to F^n$ is evaluation at the α_i's: $h(T) \bmod f(T) \mapsto (h(\alpha_1), \ldots, h(\alpha_n))$. This isomorphism matches discriminants of corresponding bases (Remark 3.69), so

$$\mathrm{disc}_{(F[T]/(f))/F}(1, T, \ldots, T^{n-1}) = \mathrm{disc}_{F^n/F}(v_1, \ldots, v_n),$$

where $v_i = (\alpha_1^{i-1}, \ldots, \alpha_n^{i-1}) \in F^n$ corresponds to $T^{i-1} \in F[T]/(f(T))$. If we write the v_i's in terms of the standard basis $\{e_1, \ldots, e_n\}$ of F^n, then by (3.19)

$$\mathrm{disc}_{F^n/F}(v_1, \ldots, v_n) = (\det C)^2 \, \mathrm{disc}_{F^n/F}(e_1, \ldots, e_n), \tag{3.20}$$

where C is the matrix expressing the v_i's in terms of the e_i's. That matrix is

$$\begin{pmatrix} 1 & \alpha_1 & \cdots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \cdots & \alpha_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_n & \cdots & \alpha_n^{n-1} \end{pmatrix}.$$

The determinant of such a matrix is given by Vandermonde's formula:

$$\det \begin{pmatrix} 1 & \alpha_1 & \cdots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \cdots & \alpha_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_n & \cdots & \alpha_n^{n-1} \end{pmatrix} = \prod_{i<j} (\alpha_j - \alpha_i).$$

Feeding this into (3.20),

$$\operatorname{disc}_{F^n/F}(v_1,\ldots,v_n) = \prod_{i<j}(\alpha_j-\alpha_i)^2 \cdot \operatorname{disc}_{F^n/F}(e_1,\ldots,e_n).$$

In the ring F^n, $e_ie_j = 0$ if $i \neq j$, $e_i^2 = e_i$, and $\operatorname{Tr}_{F^n/F}(e_i) = 1$ since the matrix for multiplication by e_i on F^n with respect to the standard basis of F^n has a 1 in the (i,i)-entry and 0 everywhere else. Therefore

$$\operatorname{disc}_{F^n/F}(e_1,\ldots,e_n) = \det(\operatorname{Tr}_{F^n/F}(e_ie_j)) = 1,$$

so $\operatorname{disc}_{F^n/F}(v_1,\ldots,v_n) = \prod_{i<j}(\alpha_j-\alpha_i)^2 = \operatorname{disc}(f(T))$.

Now suppose the roots α_i are not all distinct, so $\operatorname{disc}(f(T))$ is 0. The argument above breaks down since the ring $F[T]/(f(T))$ is not isomorphic to F^n, so we need a new idea. We will directly show the basis discriminant is also 0. Let α_1 denote one of the multiple roots of $f(T)$, and write $f(T) = (T-\alpha_1)^e g(T)$ where $e \geqslant 2$. Then $(T-\alpha_1)g(T) \not\equiv 0 \bmod f(T)$ but $((T-\alpha_1)g(T))^e \equiv 0 \bmod f(T)$, so $(T-\alpha_1)g(T)$ is a nonzero nilpotent element of $F[T]/(f(T))$. Multiplication by a nonzero nilpotent element in a ring is a nilpotent linear operator (if $a^k = 0$ then $m_a^k = m_{a^k} = m_0 = 0$), and the trace of a nilpotent operator is 0 since its only eigenvalues are 0. Build an F-basis of $F[T]/(f(T))$ that includes the chosen nonzero nilpotent element (any nonzero member of a vector space can be extended to a basis). The trace pairing matrix for this basis has a column of 0's since a nilpotent element times anything is nilpotent and thus has trace 0. Therefore the discriminant of this basis is 0 and that implies the discriminant of any basis is 0.

To show

$$\operatorname{disc}(f(T)) = (-1)^{n(n-1)/2}\, \mathrm{N}_{(F[T]/(f))/F}(f'(T)), \tag{3.21}$$

we use α_i in place of the non-existent $\sigma_i(\alpha)$ from the proof of Theorem 3.25:

$$f'(\alpha_i) = \prod_{j\neq i}(\alpha_j-\alpha_i)$$

by the product rule for derivatives, so

$$\prod_{i=1}^n f'(\alpha_i) = \prod_{i=1}^n\prod_{j\neq i}(\alpha_j-\alpha_i) = (-1)^{n(n-1)/2}\prod_{i<j}(\alpha_j-\alpha_i)^2.$$

Therefore (3.21) is equivalent to

$$\mathrm{N}_{(F[T]/(f))/F}(f'(T)) \overset{?}{=} \prod_{i=1}^{n} f'(\alpha_i). \tag{3.22}$$

(We can't use Corollary 8.16 to prove (3.22), as we do in the proof of the similar part of Theorem 3.25, because $F[T]/(f)$ may not be a field.) Verifying (3.22) is a good example of the principle of proving a theorem by proving a stronger theorem: we will show for all $g(T) \in F[T]$, not just for $f'(T)$, that

$$\mathrm{N}_{(F[T]/(f))/F}(g(T)) = \prod_{i=1}^{n} g(\alpha_i). \tag{3.23}$$

We still have f as a modulus on the left side of (3.23). Only f' has been replaced.

Both sides of (3.23) are multiplicative in g. If $g(T) = c \in F$, each side of (3.23) is c^n. So we may now assume $g(T)$ is nonconstant and monic. With respect to the basis $\{1, T, \dots, T^{n-1}\}$ of $F[T]/(f(T))$, the matrix for multiplication by $g(T) \bmod f(T)$ depends on the remainders when we reduce each $T^i g(T) \bmod f(T)$, and these remainders are unchanged when F is replaced by a larger field. Replacing F with a larger field in which *both* $g(T)$ and $f(T)$ split into linear factors, we may assume in $F[T]$ that $f(T) = (T - \alpha_1)\cdots(T - \alpha_n)$ and $g(T) = (T - \beta_1)\cdots(T - \beta_d)$. Since both sides of (3.23) are multiplicative in g, we may now assume $g(T) = T - \beta$. That is, we want to check for $\beta \in F$ that

$$\mathrm{N}_{(F[T]/(f))/F}(T - \beta) \overset{?}{=} \prod_{i=1}^{n} (\alpha_i - \beta).$$

Using the linear change of variables $T \mapsto T + \beta$ on $F[T]$,

$$\mathrm{N}_{(F[T]/(f(T)))/F}(T - \beta) = \mathrm{N}_{(F[T+\beta]/(f(T+\beta)))/F}(T) = \mathrm{N}_{(F[T]/(f(T+\beta)))/F}(T).$$

Since $f(T + \beta) = \prod_{i=1}^{n}(T + \beta - \alpha_i) = \prod_{i=1}^{n}(T - (\alpha_i - \beta))$, renaming $f(T + \beta)$ as $f(T)$ reduces us to checking

$$\mathrm{N}_{(F[T]/(f(T)))/F}(T) \overset{?}{=} \prod_{i=1}^{n} \alpha_i$$

where $f(T) = \prod_{i=1}^{n}(T - \alpha_i)$.

By definition, $\mathrm{N}_{(F[T]/(f(T)))/F}(T)$ is the determinant of any matrix representing multiplication by T. Using the basis $\{1, T, \ldots, T^{n-1}\}$ for $F[T]/(f(T))$, the matrix for multiplication by T is

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -c_{n-1} \end{pmatrix},$$

where $f(T) = T^n + c_{n-1}T^{n-1} + \cdots + c_1T + c_0$. Computing the determinant by a cofactor expansion along the top row, the determinant is $(-1)^{n-1}(-c_0) = (-1)^n c_0 = (-1)^n f(0) = \prod_{i=1}^n \alpha_i$. ■

Example 3.71. Let $f(T) = T^n + c$, a binomial polynomial. By Theorem 3.70,

$$\begin{aligned} \operatorname{disc}(T^n + c) &= \operatorname{disc}_{(F[T]/(T^n+c))/F}(1, T, \ldots, T^{n-1}) \\ &= (-1)^{n(n-1)/2}\,\mathrm{N}_{(F[T]/(T^n+c))/F}(nT^{n-1}) \\ &= (-1)^{n(n-1)/2}n^n\,\mathrm{N}_{(F[T]/(T^n+c))/F}(T)^{n-1} \\ &= (-1)^{n(n-1)/2}n^n((-1)^n c)^{n-1} \\ &= (-1)^{n(n-1)/2}n^n c^{n-1} \end{aligned}$$

since $(-1)^{n(n-1)} = 1$. For example, $\operatorname{disc}(T^2+c) = -4c$ and $\operatorname{disc}(T^3+c) = -27c^2$.

3.8 Exercises

1. Find the ideal norms of $(6, 2 + 7\sqrt{-5})$, $(3, 1 + 2\sqrt{-5})$, and $(7, 2 + 3\sqrt{-5})$ in $\mathbf{Z}[\sqrt{-5}]$. (First find a $\mathbf{Z}$-basis of the ideal, which in none of these cases is the given pair of generators for the ideal. See Exercise 1.28.)

2. Let α be an algebraic integer with minimal polynomial $f(T) \in \mathbf{Z}[T]$ of degree n. Set $K = \mathbf{Q}(\alpha)$.

   a) Show α is part of a $\mathbf{Z}$-basis of $\mathcal{O}_K$ if and only if $\alpha \notin p\mathcal{O}_K$ for all primes p.

   b) If the non-leading coefficients of $f(T)$ are relatively prime, show α fits the hypothesis of part a.

c) If $\alpha \in p\mathcal{O}_K$ for some prime p, then show $p^n \mid \mathrm{N}_{K/\mathbf{Q}}(\alpha)$, which limits the possible choices of p.

d) Show $\{1, \alpha\}$ is part of a $\mathbf{Z}$-basis of $\mathcal{O}_K$ if and only if $\alpha \notin \mathbf{Z} + p\mathcal{O}_K$ for all primes p.

e) If $\alpha \in a + p\mathcal{O}_K$ for some $a \in \mathbf{Z}$ and prime p, show $p^{n(n-1)} \mid \mathrm{disc}(f(T))$ and $f(T) \equiv (T-a)^n \bmod p$. This limits the choices for p and $a \bmod p$. (For random polynomials it is very unlikely that $\mathrm{disc}(f(T))$ will have a prime factor with multiplicity $n(n-1)$ when $n \geqslant 3$, so when $f(T)$ is picked at random $\{1, \alpha\}$ is probably part of a $\mathbf{Z}$-basis.)

f) Let α be a root of $T^3 - 3T + 25$. Show α belongs to a $\mathbf{Z}$-basis of $\mathbf{Q}(\alpha)$ (don't construct it, however) while $\{1, \alpha\}$ does not.

3. Generalize Corollary 3.3 to ideals: if $\mathfrak{a}$ is a nonzero ideal in $\mathcal{O}_K$ and we write $\mathfrak{a} \cap \mathbf{Z} = a\mathbf{Z}$, then $\mathfrak{a}$ has a $\mathbf{Z}$-basis that includes a.

4. a) Verify $\mathrm{disc}(T^3 + aT + b) = -4a^3 - 27b^2$.

   b) Verify $\mathrm{disc}(T^3 + aT^2 + b) = -4a^3b - 27b^2$.

   c) Verify $\mathrm{disc}(T^4 + aT^2 + b) = 16b(a^2 - 4b)^2$.

5. a) For monic $f(T) \in F[T]$ and $c \in F$, show $f(T+c)$ and $f(T)$ have the same discriminant.

   b) Let $f(T) = T^3 - T^2 - 3T - 3$. Find $c \in \mathbf{Q}$ such that $f(T+c)$ has no T^2 term and use that to verify $\mathrm{disc}(f(T)) = -300$.

6. Compute a $\mathbf{Z}$-basis and the discriminant of the following cubic fields:

   a) $\mathbf{Q}(\alpha)$, $\alpha^3 - 10\alpha + 1 = 0$.

   b) $\mathbf{Q}(\alpha)$, $\alpha^3 + \alpha + 8 = 0$. (Hint: Stickelberger's theorem.)

   c) $\mathbf{Q}(\alpha)$, $\alpha^3 + \alpha^2 + 8 = 0$. (Hint: Stickelberger's theorem.)

7. Let $K = \mathbf{Q}(\alpha)$ be a cubic field where $\alpha^3 + b\alpha + c = 0$ for integers b and c.

   a) Show for $x, y \in \mathbf{Z}$ not both 0, $[\mathbf{Z}[\alpha] : \mathbf{Z}[x\alpha + y\alpha^2]] = |x^3 + bxy^2 + cy^3|$.

   b) If $\alpha^3 + a\alpha^2 + b\alpha + c = 0$ and $x, y \in \mathbf{Z}$ are not both 0, does the index $[\mathbf{Z}[\alpha] : \mathbf{Z}[x\alpha + y\alpha^2]]$ equal $|x^3 + ax^2y + bxy^2 + cy^3|$?

   c) For integers x and y which are not both 0, show $\mathbf{Z}[x\sqrt[3]{2} + y\sqrt[3]{4}] = \mathbf{Z}[\sqrt[3]{2}]$ if and only if $x^3 - 2y^3 = \pm 1$.

d) A theorem of Nagell and Delaunay says for noncube integers d that the equation $x^3 - dy^3 = 1$ has at most one integral solution other than $(1, 0)$. Use this and the index formula in part a to determine all power bases of $\mathbf{Z}[\sqrt[3]{2}]$.

8. The previous exercise connected integral solutions of $x^3 - 2y^3 = 1$ with power bases of $\mathbf{Z}[\sqrt[3]{2}]$. Here you will connect them with integral solutions of $y^2 = x^3 + 1$.

a) If $x^3 - 2y^3 = 1$, check $y'^2 = x'^3 + 1$ when $x' = 2xy$ and $y' = 4y^3 + 1$. Note $y' \equiv 1 \bmod 4$.

b) If $y'^2 = x'^3 + 1$ in $\mathbf{Z}$ and y' is odd, change signs on y' if necessary to assume $y' \equiv 1 \bmod 4$. Find integers x and y such that $x^3 - 2y^3 = 1$ with $2xy = x'$ and $4y^3 + 1 = y'$. (Hint: $(\frac{x'}{2})^3 = \frac{y'+1}{2}\frac{y'-1}{4}$.)

c) If $y'^2 = x'^3 + 1$ and y' is even, so x' is odd, show $y' + 1$ and $y' - 1$ are both cubes. Conclude $y' = 0$ and $x' = -1$.

9. Let K be a quadratic field. By Exercise 1.8, the unique quadratic ring with index c in $\mathcal{O}_K$ is

$$\mathbf{Z} + c\mathcal{O}_K = \mathbf{Z} + \mathbf{Z}c\omega = \mathbf{Z}[c\omega],$$

where $\mathcal{O}_K = \mathbf{Z}[\omega]$.

a) Show the discriminant of any quadratic ring is not a perfect square and is 0 or 1 mod 4.

b) Show the discriminant of a quadratic ring can be used to reconstruct the quadratic ring (up to isomorphism), so nonisomorphic quadratic rings have different discriminants.

c) For any integer D that is not a perfect square and satisfies $D \equiv 0$ or $1 \bmod 4$, show there is a quadratic ring with discriminant D. (This ring is unique up to isomorphism by part b.)

d) Determine explicitly the quadratic rings with the following discriminants: -4, 45, 28, and -28. (Make sure your answer is a ring and not just an additive group, *e.g.*, it must contain 1.)

e) Show every perfect square is the discriminant of a unique subring of the product ring $\mathbf{Z}^2$ (Exercise 2.2).

10. If $[E : F] = n$ and $\alpha_1, \ldots, \alpha_n$ is a linearly dependent set in E, show $\det(\mathrm{Tr}_{E/F}(\alpha_i\alpha_j)) = 0$.

11. Let $K = \mathbf{Q}(i, \sqrt{-5})$.

    a) Show $\mathcal{O}_K = \mathbf{Z}[\frac{i+\sqrt{-5}}{2}]$ and $\mathrm{disc}(K) = 400$.

    b) Show $\mathcal{O}_K = \mathbf{Z}[\sqrt{-5}] + \mathbf{Z}[\sqrt{-5}]\frac{i+\sqrt{-5}}{2}$. (Note: Since $\mathbf{Z}[\sqrt{-5}]$ is not a PID, *a priori* we have no reason to expect $\mathcal{O}_K$ is a free $\mathbf{Z}[\sqrt{-5}]$-module.)

    c) Perhaps you are intrigued that $\mathrm{disc}(K)$ comes out to be a perfect square. Show the discriminant of any quartic field $K/\mathbf{Q}$ with $\mathrm{Gal}(K/\mathbf{Q}) \cong \{\pm1\} \times \{\pm1\}$ is a perfect square. (Hint: Use the theorem from Galois theory which explains when the Galois group of an irreducible polynomial in $\mathbf{Q}[T]$ naturally lies in the alternating group.)

12. For a squarefree integer $d \neq \pm1$, determine a $\mathbf{Z}[i]$-basis for the ring of integers of $\mathbf{Q}(i, \sqrt{d})$ and use this to compute $\mathrm{disc}(\mathbf{Q}(i, \sqrt{d}))$.

13. a) Show $\mathbf{Q}(\sqrt{2}, \sqrt{3}) = \mathbf{Q}(\sqrt{2} + \sqrt{3})$.

    b) Show $\mathbf{Z}[\sqrt{2}, \sqrt{3}] \neq \mathbf{Z}[\sqrt{2}+\sqrt{3}]$ by checking $[\mathbf{Z}[\sqrt{2}, \sqrt{3}] : \mathbf{Z}[\sqrt{2}+\sqrt{3}]] = 4$.

    c) Show $\alpha := \frac{\sqrt{2}+\sqrt{6}}{2}$ is an algebraic integer and compute its minimal polynomials over $\mathbf{Q}(\sqrt{2})$ and $\mathbf{Q}(\sqrt{3})$.

    d) Let $K = \mathbf{Q}(\sqrt{2}, \sqrt{3})$. Use part c to show $\mathrm{disc}_{K/\mathbf{Q}(\sqrt{2})}(1, \alpha) = 6$ and conclude that $\mathbf{Z}[\sqrt{2}]+\mathbf{Z}[\sqrt{2}]\alpha \subset \mathcal{O}_K \subset \frac{1}{\sqrt{2}}(\mathbf{Z}[\sqrt{2}]+\mathbf{Z}[\sqrt{2}]\alpha)$. (Hint: What is the largest square factor of 6 in $\mathbf{Z}[\sqrt{2}]$?)

    e) Use coset representatives for $\frac{1}{\sqrt{2}}(\mathbf{Z}[\sqrt{2}]+\mathbf{Z}[\sqrt{2}]\alpha)/(\mathbf{Z}[\sqrt{2}]+\mathbf{Z}[\sqrt{2}]\alpha)$ to show $\mathcal{O}_K = \mathbf{Z}[\sqrt{2}] + \mathbf{Z}[\sqrt{2}]\alpha$.

    f) Use part e to compute a $\mathbf{Z}$-basis of $\mathcal{O}_K$, the index $[\mathcal{O}_K : \mathbf{Z}[\sqrt{2}, \sqrt{3}]]$, and $\mathrm{disc}(K)$.

    g) Is $\{1, \sqrt{2}, \sqrt{3}, \alpha\}$ a $\mathbf{Z}$-basis of $\mathcal{O}_K$? What about $\{1, \sqrt{2}, \sqrt{6}, \alpha\}$? Or $\{1, \sqrt{3}, \sqrt{6}, \alpha\}$?

    h) Find a $\mathbf{Z}[\sqrt{3}]$-basis of $\mathcal{O}_K$. (Make sure your basis is actually in $\mathcal{O}_K$.)

14. Let $F = \mathbf{Q}(\sqrt{3})$ and

$$E = F(i) = \mathbf{Q}(\sqrt{3}, i) = \mathbf{Q}(\sqrt{3}, \zeta_3),$$

where $\zeta_3 = (-1+\sqrt{-3})/2$ is a nontrivial cube root of unity. Since $\mathcal{O}_F = \mathbf{Z}[\sqrt{3}]$ is a PID, $\mathcal{O}_E$ is a free $\mathcal{O}_F$-module with rank $[E:F]=2$.

a) Prove the subrings $\mathcal{O}_F[\zeta_3]$ and $\mathcal{O}_F[i]$ of $\mathcal{O}_E$ do not contain each other. In particular, neither is the ring of integers in E.

b) Compute $\operatorname{disc}_{E/F}(1,\zeta_3)$.

c) Compute $\operatorname{disc}_{E/F}(1,i)$.

d) Let $\zeta_{12} = i\zeta_3 = -(\sqrt{3}+i)/2$, which is a root of unity of order 12. It does not lie in the (real) field F. Check ζ_{12} is a root of $X^2+\sqrt{3}X+1 \in F[X]$ and compute $\operatorname{disc}_{E/F}(1,\zeta_{12})$.

e) Give an example of an $\mathcal{O}_F$-basis of $\mathcal{O}_E$, and justify your answer.

f) Use part e to give a $\mathbf{Z}$-basis of $\mathcal{O}_E$.

15. For any field F, let $K = F(X,\alpha)$, where α is a root of T^3+XT+X. Let R be the integral closure of $F[X]$ in K.

a) Show the largest square factor of $\operatorname{disc}_{K/F(X)}(1,\alpha,\alpha^2)$ is X^2, so

$$F[X]+F[X]\alpha+F[X]\alpha^2 \subset R \subset \frac{1}{X}(F[X]+F[X]\alpha+F[X]\alpha^2).$$

The cases when F has characteristic 2 or 3 may need a separate consideration.

b) Show coset representatives for the additive group

$$\frac{1}{X}(F[X]+F[X]\alpha+F[X]\alpha^2)/(F[X]+F[X]\alpha+F[X]\alpha^2)$$

are $(a+b\alpha+c\alpha^2)/X$, where $a,b,c \in F$.

c) With notation as in part b, compute $\operatorname{Tr}_{K/F(X)}((a+b\alpha+c\alpha^2)/X)$ and conclude that if $(a+b\alpha+c\alpha^2)/X \in R$ then $a=0$ as long as F does not have characteristic 3.

d) With notation as in part b, compute $\mathrm{N}_{K/F(X)}((a+b\alpha+c\alpha^2)/X)$ and conclude that $(a+b\alpha+c\alpha^2)/X \in R$ only if $a=b=c=0$, so $R = F[X][\alpha]$. (This bypasses part c and in particular is valid with no constraints on the characteristic of F.)

16. (Continuation of Exercise 2.11.) For a field F with characteristic 2 and an odd positive integer d, let α_d be a root of T^2-T-1/X^d over $F(X)$, so

$F(X, \alpha_d)/F(X)$ is a quadratic extension. Let $\mathcal{O}_d$ be the integral closure of $F[X]$ in $F(X, \alpha_d)$.

Setting $\beta_d = X^{(d+1)/2}\alpha_d$, compute $\operatorname{disc}_{F[X]}(1, \beta_d)$ and conclude that

$$F[X] + F[X]\beta_d \subset \mathcal{O}_d \subset \frac{1}{X^{(d+1)/2}}(F[X] + F[X]\beta_d).$$

If you had trouble finding an $F[X]$-basis of $\mathcal{O}_d$ in Exercise 2.11, try again (at least for small odd d).

17. We want to show if $f(T) \in \mathbf{Z}[T]$ is monic and $g(T)$ is a monic factor of $f(T)$ in $\mathbf{Z}[T]$ then the integer $\operatorname{disc} g$ is a factor of $\operatorname{disc} f$.

    Let F be a field and $g(T)$ and $h(T)$ be nonconstant monic polynomials in $F[T]$ with respective degrees m and n. Factor them in a splitting field as

    $$g(T) = (T - \alpha_1)\cdots(T - \alpha_m), \quad h(T) = (T - \beta_1)\cdots(T - \beta_n).$$

    Define the resultant of g and h to be

    $$\operatorname{Res}(f, g) = \prod_{i,j}(\alpha_i - \beta_j).$$

    This is not generally symmetric: $\operatorname{Res}(g, f) = (-1)^{mn}\operatorname{Res}(f, g)$. (The resultant can be defined for nonmonic or constant polynomials by a slightly more general formula, but we don't need that.)

    a) Show $\operatorname{Res}(g, h) = \mathrm{N}_{(F[T]/(g))/F}(h(T))$, so $\operatorname{Res}(g, h) \in F$.

    b) Show $\operatorname{disc}(gh) = \operatorname{disc}(g)\operatorname{disc}(h)\operatorname{Res}(g, h)^2$.

    c) If $g(T)$ and $h(T)$ are monic in $\mathbf{Z}[T]$, show $\operatorname{Res}(g, h) \in \mathbf{Z}$ and conclude that $\operatorname{disc}(g)$ is a factor of $\operatorname{disc}(gh)$.

18. This exercise puts aligned bases to use in order to define the index of one $\mathbf{Z}$-lattice in another when neither one is inside the other.

    Let V be a finite-dimensional $\mathbf{Q}$-vector space, L and L' be two $\mathbf{Z}$-lattices in V, and $f\colon V \to V$ and $g\colon V \to V$ be two $\mathbf{Q}$-linear maps such that $f(L) = L'$ and $g(L) = L'$. Concretely, f and g are each identifications of a $\mathbf{Z}$-basis of L with one of L' and then extended $\mathbf{Q}$-linearly to all of V.

    a) Show f and g are injective and surjective as maps from V to itself. (Just review the definition of a lattice over a PID as a "big" subgroup of V.)

b) Show $\det f = \pm \det g$ by adapting the proof of Theorem 3.10 to this situation. The determinants here are as maps from V to V. (It would make no sense to take the determinant of a map from L to L'.)

c) By part b, $|\det f|$ is independent of the choice of $\mathbf{Q}$-linear map $f \colon V \to V$ satisfying $f(L) = L'$. When $L' \subset L$, $[L : L'] = |\det f|$ by Theorem 3.10. This suggests defining the index $[L : L']$ of any two lattices in V as $|\det f|$ where $f(L) = L'$. Now we may speak about $[L : L']$ without requiring that $L' \subset L$, but $[L : L']$ may be rational. As an example in $\mathbf{Q}^2$, let $L = \mathbf{Z}^2$ and $L' = \mathbf{Z}\binom{2}{1} + \mathbf{Z}\binom{1/3}{1}$. Compute $[L : L']$ in two ways (two choices of basis in $\mathbf{Q}^2$) and check your answers agree. Give an example where $[L : L'] = 1$ but $L \neq L'$.

d) Show there is an $m \in \mathbf{Z}^+$ such that $mL' \subset L$ and express $[L : L']$ (rational index) in terms of $[L : mL']$ (integral index). Check your answer with the examples in part c.

e) For any lattices L, L', and L'' in V, show $[L : L'][L' : L''] = [L : L'']$. (In particular, using $L'' = L$ gives us $[L' : L] = 1/[L : L']$, and for any lattice L'' contained in L and L' we have $[L : L'] = [L : L'']/[L' : L'']$, where the indices on the right are ordinary subgroup indices. This could be taken as the definition of the generalized lattice index $[L : L']$, once it is shown to be independent of the choice of L'', and a similar definition could be given using a lattice containing L and L'.)

19. Let A be a PID and T be a finitely generated torsion A-module. Write $T \cong A/a_1A \oplus \cdots \oplus A/a_kA$. Recall $\operatorname{card}_A(T) = (a_1 \cdots a_k)$. Let π be a prime factor of $a_1 \cdots a_k$. Write $a_i = \pi^{e_i} a_i'$ with $e_i \geqslant 0$ and a_i' not divisible by π.

a) Prove an analogue of Cauchy's theorem from group theory: there is some $t \in T$ with "order" π: the annihilator ideal $\operatorname{Ann}_A(t) = \{a \in A : at = 0\}$ is πA.

b) Assume $k = 2$, so $T \cong A/a_1A \oplus A/a_2A$ and we choose the indexing so that $e_1 \leqslant e_2$. Show

$$\pi^j T/\pi^{j+1} T \cong \begin{cases} (A/\pi A)^2, & \text{if } j < e_1, \\ A/\pi A, & \text{if } e_1 \leqslant j < e_2, \\ 0, & \text{if } j \geqslant e_2. \end{cases}$$

Conclude that $\sum_{j\geqslant 0}\dim_{A/\pi A}(\pi^j T/\pi^{j+1}T) = e_1 + e_2$.

c) In the general case, show

$$e_1 + \cdots + e_k = \sum_{j\geqslant 0}\dim_{A/\pi A}(\pi^j T/\pi^{j+1}T).$$

20. Let A be a PID with fraction field F, V be a finite-dimensional F-vector space, and L and L' be A-lattices in V (Definition 3.55).

a) Show $L \cap L'$ and $L + L'$ are A-lattices in V.

b) Generalize Theorem 3.10: if $L \supset L'$, show $[L : L']_A = (\det\varphi)$ as ideals in A, where $\varphi\colon V \to V$ is any F-linear map such that $\varphi(L) = L'$.

c) If $L \supset L' \supset L''$, show $[L : L'']_A = [L : L']_A[L' : L'']_A$.

d) Generalize Theorem 3.12. Let A be a PID with fraction field F, E/F be a finite extension, and B be the integral closure of A in E. Assume B is a finite free A-module (*e.g.*, , this is true when E/F is a separable field extension by Section 3.6). For any nonzero $b \in B$, show $[B : bB]_A = \mathrm{N}_{E/F}(b)A$.

21. Let A be a PID with fraction field F and V be a finite-dimensional vector space over F. Fix a symmetric bilinear form b on V. (That means $b\colon V \times V \to F$ is linear in each component when the other one is fixed and $b(v, w) = b(w, v)$ for all v and w in V.) An example is the dot product if we identify V with F^n. If $V = E$ is a finite extension field of F then another example of a symmetric bilinear form is the trace pairing $b(x, y) = \mathrm{Tr}_{E/F}(xy)$.

a) If $\{e_1, \ldots, e_n\}$ is a basis of V, set its b-discriminant to be

$$\mathrm{disc}_b(e_1, \ldots, e_n) = \det(b(e_i, e_j)) \in F.$$

If b is non-degenerate (that means if $b(v, w) = 0$ for all w then $v = 0$) show $\mathrm{disc}_b(e_1, \ldots, e_n) \neq 0$.

b) For an A-lattice M in V, set its b-discriminant to be

$$\mathrm{disc}_b(M) = \det(b(e_i, e_j))$$

where $\{e_1, \ldots, e_n\}$ is any A-basis of M. Show $\mathrm{disc}_b(M)$ is well-defined up to multiplication by a unit square in A.

c) If M_1 and M_2 are lattices in V with $M_1 \subset M_2$, show

$$\operatorname{disc}_b(M_1) = [M_2 : M_1]_A^2 \operatorname{disc}_b(M_2).$$

CHAPTER 4

FACTORIZATION OF IDEALS

Let K be a number field. Now that we know what $\mathcal{O}_K$ looks like additively, our next goal is to prove the main multiplicative property of $\mathcal{O}_K$: every nonzero proper ideal $\mathfrak{a}$ in $\mathcal{O}_K$ has a prime ideal factorization, say $\mathfrak{a} = \mathfrak{p}_1\mathfrak{p}_2\cdots\mathfrak{p}_r$ (some $\mathfrak{p}_i$'s could be equal), which is unique up to the order of the terms: if also $\mathfrak{a} = \mathfrak{q}_1\mathfrak{q}_2\cdots\mathfrak{q}_s$ with nonzero prime ideals $\mathfrak{q}_i$, then $r = s$ and $\mathfrak{p}_i = \mathfrak{q}_i$ after a suitable reindexing. The first proof we will give uses induction on the norm of an ideal, which is a special finiteness property of $\mathcal{O}_K$ (in most other rings, nonzero ideals do not generally have finite index). Using other finiteness conditions we will prove unique factorization of ideals in integral closures of $F[X]$ and then in a broad class of rings called Dedekind domains.

4.1 Divisibility of Ideals

We saw in Section 1.5 how to multiply ideals (Definition 1.42). From multiplication we get divisibility.

Definition 4.1. For nonzero ideals $\mathfrak{a}$ and $\mathfrak{b}$ in $\mathcal{O}_K$, write $\mathfrak{a} \mid \mathfrak{b}$ if $\mathfrak{b} = \mathfrak{a}\mathfrak{c}$ for some ideal $\mathfrak{c}$ in $\mathcal{O}_K$.

Note that if $\mathfrak{a} \mid \mathfrak{b}$, then $\mathfrak{a} \supset \mathfrak{b}$, or equivalently $\mathfrak{b} \subset \mathfrak{a}$. Ideal factors are *larger* (in terms of containment) and ideal multiples are *smaller*. (This is obvious already in $\mathbf{Z}$: $2 \mid 10$ but $10\mathbf{Z} \subset 2\mathbf{Z}$.) There is nothing special about $\mathcal{O}_K$ here: in any commutative ring we can multiply ideals, define divisibility for ideals, and the condition $\mathfrak{a} \mid \mathfrak{b}$ implies $\mathfrak{a} \supset \mathfrak{b}$. The converse is generally false in most rings, but it is true in $\mathcal{O}_K$: if $\mathfrak{a} \supset \mathfrak{b}$ then $\mathfrak{a} \mid \mathfrak{b}$. We will prove this as Theorem 4.23.

While passing from elements to their principal ideals doesn't always turn irreducible elements into prime ideals (*e.g.*, in $\mathbf{Z}[\sqrt{-5}]$, 3 is irreducible but the ideal (3) is not prime since $1+\sqrt{-5}$ and $1-\sqrt{-5}$ have product $6 \in (3)$ but neither number is in (3)), let's show passage to principal ideals has no effect on divisibility relations.

Theorem 4.2. *For nonzero $\alpha, \beta \in \mathcal{O}_K$, $\alpha \mid \beta$ as elements if and only if $(\alpha) \mid (\beta)$ as ideals.*

Proof. If $\alpha \mid \beta$, then $\beta = \alpha\gamma$ for some $\gamma \in \mathcal{O}_K$. So $(\beta) = (\alpha\gamma) = (\alpha)(\gamma)$. Thus $(\alpha) \mid (\beta)$. Conversely, say $(\alpha) \mid (\beta)$, so $(\beta) = (\alpha)\mathfrak{c}$ for some ideal $\mathfrak{c}$. Since $(\alpha)\mathfrak{c} = \alpha\mathfrak{c}$ and $\beta \in (\beta)$, $\beta = \alpha\gamma$ for some $\gamma \in \mathfrak{c}$. Thus $\alpha \mid \beta$. ■

Theorem 4.3. *Let A be a commutative ring. If $\mathfrak{p}$ is a prime ideal of A, for any ideals $\mathfrak{a}$ and $\mathfrak{b}$ of A where $\mathfrak{p} \supset \mathfrak{a}\mathfrak{b}$, we have $\mathfrak{p} \supset \mathfrak{a}$ or $\mathfrak{p} \supset \mathfrak{b}$.*

Proof. We prove the contrapositive. Say $\mathfrak{p} \not\supset \mathfrak{a}$ and $\mathfrak{p} \not\supset \mathfrak{b}$, and pick $\alpha \in \mathfrak{a}$ and $\beta \in \mathfrak{b}$ such that $\alpha \notin \mathfrak{p}$ and $\beta \notin \mathfrak{p}$. Then since $\mathfrak{p}$ is prime, $\alpha\beta \notin \mathfrak{p}$. Thus $\mathfrak{p} \not\supset \mathfrak{a}\mathfrak{b}$. ■

Remark 4.4. The condition "$\mathfrak{p} \supset \mathfrak{a}\mathfrak{b}$ implies $\mathfrak{p} \supset \mathfrak{a}$ or $\mathfrak{p} \supset \mathfrak{b}$" is not just a property of prime ideals $\mathfrak{p}$ in any commutative ring but is essentially an alternate way to define such ideals. The usual definition is: $\mathfrak{p}$ is prime when $\mathfrak{p} \neq A$ and if $xy \in \mathfrak{p}$ then $x \in \mathfrak{p}$ or $y \in \mathfrak{p}$ (more briefly: $\mathfrak{p}$ is prime when $A/\mathfrak{p}$ is a domain). If $\mathfrak{p} \neq A$ and for all ideals $\mathfrak{a}$ and $\mathfrak{b}$ in A we have $\mathfrak{p} \supset \mathfrak{a}\mathfrak{b} \Rightarrow \mathfrak{p} \supset \mathfrak{a}$ or $\mathfrak{p} \supset \mathfrak{b}$, then

$$xy \in \mathfrak{p} \Longrightarrow \mathfrak{p} \supset (xy) = (x)(y) \Longrightarrow \mathfrak{p} \supset (x) \text{ or } (y) \Longrightarrow x \in \mathfrak{p} \text{ or } y \in \mathfrak{p},$$

so $\mathfrak{p}$ is a prime ideal.

Corollary 4.5. *If $\mathfrak{p}$ is prime and $\mathfrak{p} \supset \mathfrak{a}_1 \cdots \mathfrak{a}_r$, then $\mathfrak{p} \supset \mathfrak{a}_i$ for some i. In $\mathcal{O}_K$, if $\mathfrak{p} \supset \mathfrak{p}_1 \cdots \mathfrak{p}_r$ with $\mathfrak{p}_i$ nonzero prime ideals, then $\mathfrak{p} = \mathfrak{p}_i$ for some i.*

Proof. Use induction on r for the first part. Theorem 4.3 is the case $r = 2$. For the second part, from $\mathfrak{p} \supset \mathfrak{p}_i$ we get $\mathfrak{p} = \mathfrak{p}_i$ because nonzero prime ideals in $\mathcal{O}_K$ are maximal (Corollary 3.6). ■

To prove unique factorization of ideals in $\mathcal{O}_K$, we will use multiplicative inverses for nonzero prime ideals in $\mathcal{O}_K$. As in $\mathbf{Z}$, where integers other than ±1 do not have multiplicative inverses in $\mathbf{Z}$, proper ideals in $\mathcal{O}_K$ do not have multiplicative inverses among the ideals in $\mathcal{O}_K$: for a proper ideal $\mathfrak{a}$ and any ideal $\mathfrak{b}$, $\mathfrak{a}\mathfrak{b} \subset \mathfrak{a}$, so $\mathfrak{a}\mathfrak{b} \neq \mathcal{O}_K$. To find multiplicative inverses for ideals we will go beyond ideals in $\mathcal{O}_K$ and look at $\mathcal{O}_K$-modules in K.

Example 4.6. In $\mathbf{Q}$, two $\mathbf{Z}$-modules not in $\mathbf{Z}$ are $\frac{2}{3}\mathbf{Z}$ and $\mathbf{Z}[\frac{1}{2}] = \sum_{k \geqslant 0} \mathbf{Z}\frac{1}{2^k}$; the second example is not finitely generated as a $\mathbf{Z}$-module.

Example 4.7. In $\mathbf{Q}(\sqrt{-5})$,

$$\mathbf{Z}[\sqrt{-5}] + \frac{1+\sqrt{-5}}{3}\mathbf{Z}[\sqrt{-5}] = \frac{1}{3}(3\mathbf{Z}[\sqrt{-5}] + (1+\sqrt{-5})\mathbf{Z}[\sqrt{-5}])$$

is a $\mathbf{Z}[\sqrt{-5}]$-module not in $\mathbf{Z}[\sqrt{-5}]$. It is an ideal in $\mathbf{Z}[\sqrt{-5}]$ divided by 3.

Let M and N be $\mathcal{O}_K$-modules in K. Their product is defined to be

$$MN = \Big\{ \sum_{i=1}^{r} x_i y_i : r \geqslant 1, x_i \in M,\ y_i \in N \Big\},$$

which is an $\mathcal{O}_K$-module. Multiplication of $\mathcal{O}_K$-modules in K is commutative, associative, and distributes over addition of $\mathcal{O}_K$-modules. The multiplicative identity is $\mathcal{O}_K$. An $\mathcal{O}_K$-module in K is called *principal* if it has the form $\alpha\mathcal{O}_K = \{\alpha x : x \in \mathcal{O}_K\}$. We do not insist α be in $\mathcal{O}_K$, like $\frac{2}{3}\mathbf{Z}$ as a $\mathbf{Z}$-module in $\mathbf{Q}$. The principal $\mathcal{O}_K$-modules that lie in $\mathcal{O}_K$ are the principal ideals of $\mathcal{O}_K$. Multiplication of an $\mathcal{O}_K$-module in K by a principal $\mathcal{O}_K$-module is the same as scaling: $(\alpha\mathcal{O}_K)M = \alpha M$.

Among the $\mathcal{O}_K$-modules in K, the ideals of $\mathcal{O}_K$ can be described in a trivial but important way:

Theorem 4.8. *The ideals of $\mathcal{O}_K$ are the $\mathcal{O}_K$-modules in K that happen to lie in $\mathcal{O}_K$.*

Proof. Any ideal in $\mathcal{O}_K$ is an $\mathcal{O}_K$-module since it is preserved by multiplication by $\mathcal{O}_K$. Conversely, an $\mathcal{O}_K$-module in K that happens to lie in $\mathcal{O}_K$ is an additive subgroup of $\mathcal{O}_K$ that is preserved by multiplication by any element of $\mathcal{O}_K$, so it is an ideal in $\mathcal{O}_K$ by the definition of ideals. ■

This theorem is not special to rings of integers. For any domain A with fraction field F, the ideals of A are the A-modules in F that lie in A.

4.2 Proof of Unique Factorization

4.2.1 Step 1: Multiplicative Inverses

We want to show every nonzero prime ideal $\mathfrak{p}$ in $\mathcal{O}_K$ has a multiplicative inverse as $\mathcal{O}_K$-modules:

$$\mathfrak{p}\widetilde{\mathfrak{p}} = \mathcal{O}_K$$

for some $\mathcal{O}_K$-module $\widetilde{\mathfrak{p}}$ in K.

Example 4.9. The inverse of $2\mathbf{Z}$ as a $\mathbf{Z}$-module is $\frac{1}{2}\mathbf{Z}$.

Example 4.10. In $\mathbf{Q}(\sqrt{-5})$, the ideal $(3, 1+\sqrt{-5})$ in $\mathbf{Z}[\sqrt{-5}]$ has $\mathbf{Z}[\sqrt{-5}]$-module inverse $\mathbf{Z}[\sqrt{-5}] + \frac{1-\sqrt{-5}}{3}\mathbf{Z}[\sqrt{-5}]$ because

$$(3, 1+\sqrt{-5})(3, 1-\sqrt{-5}) = 3\mathbf{Z}[\sqrt{-5}]$$

and multiplying by $\frac{1}{3}$ gives

$$(3, 1+\sqrt{-5})\left(\mathbf{Z}[\sqrt{-5}] + \frac{1-\sqrt{-5}}{3}\mathbf{Z}[\sqrt{-5}]\right) = \mathbf{Z}[\sqrt{-5}].$$

Let's accept for now that we can always find such an $\mathcal{O}_K$-module $\widetilde{\mathfrak{p}}$. (We'll see what $\widetilde{\mathfrak{p}}$ is in Subsection 4.2.4.) One thing that follows is that we can cancel $\mathfrak{p}$ in products of ideals:

$$\begin{aligned} \mathfrak{p}\mathfrak{a} = \mathfrak{p}\mathfrak{b} &\Longrightarrow \widetilde{\mathfrak{p}}\mathfrak{p}\mathfrak{a} = \widetilde{\mathfrak{p}}\mathfrak{p}\mathfrak{b} \\ &\Longrightarrow \mathcal{O}_K\mathfrak{a} = \mathcal{O}_K\mathfrak{b} \\ &\Longrightarrow \mathfrak{a} = \mathfrak{b}. \end{aligned}$$

Another important consequence is that we can show containment implies divisibility for a *prime* ideal: if $\mathfrak{a} \subset \mathfrak{p}$, then $\mathfrak{p} \mid \mathfrak{a}$ as ideals. Indeed, from $\mathfrak{a} \subset \mathfrak{p}$ we get $\widetilde{\mathfrak{p}}\mathfrak{a} \subset \mathcal{O}_K$, so $\widetilde{\mathfrak{p}}\mathfrak{a}$ is an $\mathcal{O}_K$-module in $\mathcal{O}_K$, which means $\widetilde{\mathfrak{p}}\mathfrak{a}$ is an ideal (Theorem 4.8). Set $\mathfrak{b} = \widetilde{\mathfrak{p}}\mathfrak{a}$, which is an ideal, and $\mathfrak{p}\mathfrak{b} = \mathfrak{a}$.

4.2.2 Step 2: Existence

To show each nonzero proper ideal $\mathfrak{a}$ is a product of primes, let $\mathfrak{a}$ be a counterexample with least index in $\mathcal{O}_K$. There are finitely many proper ideals of $\mathcal{O}_K$ containing $\mathfrak{a}$. At least one of them is not contained in the others, so it is a maximal ideal. Call this maximal ideal $\mathfrak{p}$. From $\mathfrak{a} \subset \mathfrak{p}$, we get $\mathfrak{a} = \mathfrak{p}\mathfrak{b}$ for some nonzero ideal $\mathfrak{b}$ (a consequence of Step 1). Since $\mathfrak{a}$ is *not* a product of primes, $\mathfrak{a} \neq \mathfrak{p}$, so $\mathfrak{b} \neq (1)$. Since $\mathfrak{a} \subset \mathfrak{b} \subset \mathcal{O}_K$, we *expect* $\mathfrak{b}$ to have smaller index in $\mathcal{O}_K$ than $\mathfrak{a}$. If so, then $\mathfrak{b}$ is a product of prime ideals, so $\mathfrak{a} = \mathfrak{p}\mathfrak{b}$ is also a product of primes. Contradiction!

How do we show $[\mathcal{O}_K : \mathfrak{b}] < [\mathcal{O}_K : \mathfrak{a}]$? That is, why is $\mathfrak{p}\mathfrak{b} \subsetneqq \mathfrak{b}$? Suppose $\mathfrak{p}\mathfrak{b} = \mathfrak{b}$. Then

$$\mathfrak{b} = \mathfrak{p}\mathfrak{b} = \mathfrak{p}(\mathfrak{p}\mathfrak{b}) = \mathfrak{p}^2\mathfrak{b} = \cdots = \mathfrak{p}^k\mathfrak{b}$$

for all $k \geqslant 1$. Since $\mathfrak{p}^k\mathfrak{b} \subset \mathfrak{p}^k$, we have $\mathfrak{b} \subset \mathfrak{p}^k$, so

$$\mathcal{O}_K \supset \mathfrak{p} \supset \cdots \supset \mathfrak{p}^k \supset \cdots \supset \mathfrak{b}.$$

The ideal $\mathfrak{b}$ has finite index in $\mathcal{O}_K$, so at some point the chain of powers of $\mathfrak{p}$ must stabilize: $\mathfrak{p}^{k+1} = \mathfrak{p}^k$ some k. We know $\mathfrak{p}$ can be cancelled from both sides, and doing this k times gives us $\mathfrak{p} = \mathcal{O}_K$. Contradiction! Thus $\mathfrak{p}\mathfrak{b} \subsetneqq \mathfrak{b}$ for any nonzero ideal $\mathfrak{b}$, and this completes the proof of existence of prime factorization for all nonzero proper ideals in $\mathcal{O}_K$.[1]

4.2.3 Step 3: Uniqueness

Say

$$\mathfrak{p}_1 \cdots \mathfrak{p}_r = \mathfrak{q}_1 \cdots \mathfrak{q}_s$$

where each $\mathfrak{p}_i$ and $\mathfrak{q}_j$ is a nonzero prime ideal. We want $r = s$ and $\mathfrak{p}_i = \mathfrak{q}_i$ after a suitable reindexing. Without loss of generality, suppose $r \leqslant s$. We will proceed by induction on r.

When $r = 1$, $\mathfrak{p}_1 = \mathfrak{q}_1 \cdots \mathfrak{q}_s$. If $s > 1$ we seek a contradiction. From $\mathfrak{p}_1 = \mathfrak{q}_1 \cdots \mathfrak{q}_s \subset \mathfrak{q}_1$ we get $\mathfrak{p}_1 = \mathfrak{q}_1$ since nonzero prime ideals in $\mathcal{O}_K$ are maximal. Then $\mathfrak{p}_1 = \mathfrak{p}_1\mathfrak{q}_2 \cdots \mathfrak{q}_s$, so by cancellation $(1) = \mathfrak{q}_2 \cdots \mathfrak{q}_s \subset \mathfrak{q}_2 \subsetneqq (1)$, which is a contradiction.

When $r \geqslant 2$, $\mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_r = \mathfrak{q}_1\mathfrak{q}_2 \cdots \mathfrak{q}_s$ implies that $\mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_r \subset \mathfrak{q}_1$, so $\mathfrak{q}_1 = \mathfrak{p}_i$ for some i (Corollary 4.5). We may relabel $\mathfrak{p}_i$ as $\mathfrak{p}_1$: $\mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_r = \mathfrak{p}_1\mathfrak{q}_2 \cdots \mathfrak{q}_s$. Cancelling $\mathfrak{p}_1$ on both sides, $\mathfrak{p}_2 \cdots \mathfrak{p}_r = \mathfrak{q}_2 \cdots \mathfrak{q}_s$. By induction $r - 1 = s - 1$ (so $r = s$) and we can write $\mathfrak{p}_i = \mathfrak{q}_i$ for $2 \leqslant i \leqslant r$ by relabelling.

[1] This proof is not constructive, and neither is the proof of the existence of prime factorization in the positive integers.

4.2.4 Back to Step 1

The gap in our proof of unique factorization is showing every nonzero prime ideal $\mathfrak{p}$ in $\mathcal{O}_K$ is invertible as an $\mathcal{O}_K$-module. We seek an $\mathcal{O}_K$-module $\widetilde{\mathfrak{p}}$ in K such that $\mathfrak{p}\widetilde{\mathfrak{p}} = \mathcal{O}_K$. If the $\mathcal{O}_K$-module $\widetilde{\mathfrak{p}}$ exists, then

$$\widetilde{\mathfrak{p}} \subset \{x \in K : x\mathfrak{p} \subset \mathcal{O}_K\}.$$

The right side is itself an $\mathcal{O}_K$-module, and we will use it as our definition of $\widetilde{\mathfrak{p}}$. We will make the definition in any domain, not just in $\mathcal{O}_K$.

Definition 4.11. Let A be a domain with fraction field F. For a nonzero A-module $\mathfrak{a} \subset F$, set

$$\widetilde{\mathfrak{a}} := \{x \in F : x\mathfrak{a} \subset A\}.$$

Example 4.12. In $\mathbf{Z}$ if $\mathfrak{a} = 2\mathbf{Z}$, then $\widetilde{\mathfrak{a}} = \frac{1}{2}\mathbf{Z}$.

Example 4.13. In $\mathbf{Z}[\sqrt{-5}]$, if $\mathfrak{p}$ is the ideal $(3, 1+\sqrt{-5})$ in $\mathbf{Z}[\sqrt{-5}]$, then from $(3, 1+\sqrt{-5})(3, 1-\sqrt{-5}) = (3)$ we have

$$\widetilde{\mathfrak{p}} = \frac{1}{3}(3, 1-\sqrt{-5}) = \mathbf{Z}[\sqrt{-5}] + \frac{1-\sqrt{-5}}{3}\mathbf{Z}[\sqrt{-5}].$$

Lemma 4.14. *If $\mathfrak{a}$ has an A-module inverse in F, the inverse must be $\widetilde{\mathfrak{a}}$.*

Proof. Suppose $\mathfrak{a}\mathfrak{a}' = A$. Then $\mathfrak{a}' \subset \widetilde{\mathfrak{a}}$. Multiplying both sides by $\mathfrak{a}$, we get $A \subset \mathfrak{a}\widetilde{\mathfrak{a}} \subset A$, so $\mathfrak{a}\widetilde{\mathfrak{a}} = A$. Multiplying both sides by $\mathfrak{a}'$, $\widetilde{\mathfrak{a}} = \mathfrak{a}'$. ∎

We return to the number field setting. By the definition of $\widetilde{\mathfrak{p}}$, $\mathfrak{p}\widetilde{\mathfrak{p}} \subset \mathcal{O}_K$. Easily $\mathcal{O}_K \subset \widetilde{\mathfrak{p}}$, and multiplying both sides by $\mathfrak{p}$ gives $\mathfrak{p} \subset \mathfrak{p}\widetilde{\mathfrak{p}}$, so $\mathfrak{p} \subset \mathfrak{p}\widetilde{\mathfrak{p}} \subset \mathcal{O}_K$. Since $\mathfrak{p}$ is a maximal ideal, $\mathfrak{p}\widetilde{\mathfrak{p}}$ is either $\mathfrak{p}$ or $\mathcal{O}_K$. The *key* to proving $\mathfrak{p}\widetilde{\mathfrak{p}} = \mathcal{O}_K$ is showing $\mathcal{O}_K \subsetneqq \widetilde{\mathfrak{p}}$. Assuming this strict inclusion, we will show $\mathfrak{p}\widetilde{\mathfrak{p}} = \mathcal{O}_K$.

Pick $x \in \widetilde{\mathfrak{p}}$ with $x \notin \mathcal{O}_K$. Then

$$x\mathfrak{p} \subset \mathcal{O}_K \Longrightarrow \mathfrak{p} \subset \mathfrak{p} + x\mathfrak{p} \subset \mathcal{O}_K.$$

By maximality of $\mathfrak{p}$, $\mathfrak{p} + x\mathfrak{p} = \mathfrak{p}$ or $\mathfrak{p} + x\mathfrak{p} = \mathcal{O}_K$. We will eliminate the first choice by contradiction. Suppose $\mathfrak{p} + x\mathfrak{p} = \mathfrak{p}$. Then $x\mathfrak{p} \subset \mathfrak{p}$. Since $\mathfrak{p}$ is a finitely generated $\mathbf{Z}$-module, our earlier linear characterization of integrality (Theorem 1.15) can be applied even though now we don't have a ring finitely generated as a $\mathbf{Z}$-module containing x but instead an ideal $\mathfrak{p}$ finitely generated as a $\mathbf{Z}$-module that is preserved by multiplication by x; the integrality of x over $\mathbf{Z}$ follows in the same way. That is, $x\mathfrak{p} \subset \mathfrak{p} \Rightarrow x \in \mathcal{O}_K$. However, we chose $x \notin \mathcal{O}_K$, so this is a contradiction. Thus $\mathfrak{p} + x\mathfrak{p} = \mathcal{O}_K$, so $\mathfrak{p}(\mathcal{O}_K + x\mathcal{O}_K) = \mathcal{O}_K$, which shows $\mathcal{O}_K + x\mathcal{O}_K \subset \widetilde{\mathfrak{p}}$. Then multiplication of both sides by $\mathfrak{p}$ yields $\mathcal{O}_K \subset \mathfrak{p}\widetilde{\mathfrak{p}}$. We saw earlier $\mathfrak{p}\widetilde{\mathfrak{p}} \subset \mathcal{O}_K$, so $\mathfrak{p}\widetilde{\mathfrak{p}} = \mathcal{O}_K$.

From $\mathfrak{p}\widetilde{\mathfrak{p}} = \mathcal{O}_K$ and $\mathfrak{p}(\mathcal{O}_K + x\mathcal{O}_K) = \mathcal{O}_K$ we get $\mathfrak{p}\widetilde{\mathfrak{p}} = \mathfrak{p}(\mathcal{O}_K + x\mathcal{O}_K)$. Multiplication by $\widetilde{\mathfrak{p}}$ gives

$$\widetilde{\mathfrak{p}} = \mathcal{O}_K + x\mathcal{O}_K$$

for any $x \in \widetilde{\mathfrak{p}} - \mathcal{O}_K$. So we have a "formula" for $\widetilde{\mathfrak{p}}$ in terms of a choice of any $x \in \widetilde{\mathfrak{p}} - \mathcal{O}_K$.

To complete Step 1 of the proof that $\mathcal{O}_K$ has unique factorization of ideals, it remains to show for each nonzero prime ideal $\mathfrak{p}$ in $\mathcal{O}_K$, that $\mathcal{O}_K \subsetneqq \widetilde{\mathfrak{p}}$: some $x \in K$ with $x \notin \mathcal{O}_K$ satisfies $x\mathfrak{p} \subset \mathcal{O}_K$.

Lemma 4.15. *In $\mathcal{O}_K$, any nonzero ideal $\mathfrak{a}$ contains a product of nonzero prime ideals:*

$$\mathfrak{a} \supset \mathfrak{p}_1 \cdots \mathfrak{p}_n.$$

Proof. If false, let $\mathfrak{a}$ be a counterexample with least index. Then $[\mathcal{O}_K : \mathfrak{a}] \geqslant 2$ since the lemma is true for $\mathfrak{a} = \mathcal{O}_K$. Since $\mathfrak{a}$ is not prime, there are $x, y \in \mathcal{O}_K$ such that $xy \in \mathfrak{a}$ and $x \notin \mathfrak{a}$ and $y \notin \mathfrak{a}$. Consider the ideals $(x) + \mathfrak{a}$ and $(y) + \mathfrak{a}$. They each strictly contain $\mathfrak{a}$, and larger ideals have smaller index, so they each contain a product of nonzero prime ideals:

$$(x) + \mathfrak{a} \supset \mathfrak{p}_1 \cdots \mathfrak{p}_k, \quad (y) + \mathfrak{a} \supset \mathfrak{q}_1 \cdots \mathfrak{q}_\ell.$$

Then

$$\begin{aligned}((x) + \mathfrak{a})((y) + \mathfrak{a}) &= (x)(y) + (x)\mathfrak{a} + (y)\mathfrak{a} + \mathfrak{a}^2 \\ &= (xy) + x\mathfrak{a} + y\mathfrak{a} + \mathfrak{a}^2 \\ &\subset \mathfrak{a},\end{aligned}$$

so $\mathfrak{a} \supset \mathfrak{p}_1 \cdots \mathfrak{p}_k \mathfrak{q}_1 \cdots \mathfrak{q}_\ell$, which is a contradiction. ■

Although we have not yet proved containment is the same as divisibility, the way to think about the lemma intuitively is that it says "$\mathfrak{a} \mid \mathfrak{p}_1 \cdots \mathfrak{p}_n$." (In Step 2 we prove $\mathfrak{a}$ equals a product of prime ideals, but right now we are justifying Step 1.)

To show $\mathcal{O}_K \subsetneqq \widetilde{\mathfrak{p}}$, pick nonzero $\alpha \in \mathfrak{p}$. By Lemma 4.15,

$$\mathfrak{p} \supset (\alpha) \supset \mathfrak{p}_1 \cdots \mathfrak{p}_n$$

for some nonzero prime ideals $\mathfrak{p}_i$. Choose n to be as small as possible. By Corollary 4.5, $\mathfrak{p} = \mathfrak{p}_i$ for some i, and without loss of generality we can take $i = 1$: $\mathfrak{p} = \mathfrak{p}_1$. Then

$$\mathfrak{p} \supset (\alpha) \supset \mathfrak{p}\mathfrak{p}_2 \cdots \mathfrak{p}_n.$$

If $n = 1$, then $\mathfrak{p} = (\alpha)$, which implies $\widetilde{\mathfrak{p}} = \frac{1}{\alpha}\mathcal{O}_K$ and $\frac{1}{\alpha} \notin \mathcal{O}_K$ since $\alpha \notin \mathcal{O}_K^\times$ (because $(\alpha) = \mathfrak{p}$). So $\frac{1}{\alpha} \in \widetilde{\mathfrak{p}}$ and $\frac{1}{\alpha} \notin \mathcal{O}_K$. If $n > 1$, then

$$(\alpha) \not\supset \mathfrak{p}_2 \cdots \mathfrak{p}_n$$

by the minimality of n. Choose $\beta \in \mathfrak{p}_2 \cdots \mathfrak{p}_n$ such that $\beta \notin (\alpha)$. Since $(\alpha) = \alpha\mathcal{O}_K$, $\beta/\alpha \notin \mathcal{O}_K$. Also,

$$\frac{\beta}{\alpha}\mathfrak{p} \subset \frac{1}{\alpha}\mathfrak{p}_2 \cdots \mathfrak{p}_n\mathfrak{p} \subset \frac{1}{\alpha}(\alpha) = \mathcal{O}_K.$$

Hence $\beta/\alpha \in \widetilde{\mathfrak{p}}$ and $\beta/\alpha \notin \mathcal{O}_K$. Our proof that $\mathcal{O}_K$ has unique factorization of ideals is complete.

In Steps 1 and 2, we used finiteness of $\mathcal{O}_K/\mathfrak{a}$ to make arguments based on the index of an ideal (*e.g.*, choosing a counterexample of least index). Let's review why $\#(\mathcal{O}_K/\mathfrak{a})$ is finite. First, as a $\mathbf{Z}$-module $\mathcal{O}_K \cong \mathbf{Z}^n$ because choosing a $\mathbf{Q}$-basis $\{e_1, \ldots, e_n\}$ of K in $\mathcal{O}_K$ leads to

$$\sum_{i=1}^n \mathbf{Z}e_i \subset \mathcal{O}_K \subset \frac{1}{d}\sum_{i=1}^n \mathbf{Z}e_i = \sum_{i=1}^n \mathbf{Z}\frac{e_i}{d},$$

where $d = \operatorname{disc}_{K/\mathbf{Q}}(e_1, \ldots, e_n) \in \mathbf{Z} - \{0\}$. This puts $\mathcal{O}_K$ between two finite free $\mathbf{Z}$-modules of rank n, so it too is finite free of rank n. Then for any nonzero ideal $\mathfrak{a}$, picking $a \in \mathfrak{a} \cap \mathbf{Z} - \{0\}$ we have $a\mathcal{O}_K \subset \mathfrak{a} \subset \mathcal{O}_K$, which puts $\mathfrak{a}$ between two finite free $\mathbf{Z}$-modules of rank n. Thus $\mathfrak{a} \cong \mathbf{Z}^n$ as a $\mathbf{Z}$-module. Since $\mathcal{O}_K/a\mathcal{O}_K \cong (\mathbf{Z}/a\mathbf{Z})^n$ is finite, also $\mathcal{O}_K/\mathfrak{a}$ is finite.

4.3 Inverses and Greatest Common Divisors

For any nonzero prime ideal $\mathfrak{p}$ in $\mathcal{O}_K$, we have found an inverse $\widetilde{\mathfrak{p}}$. Let's show there is an inverse for any nonzero ideal $\mathfrak{a}$ in $\mathcal{O}_K$. Set

$$\widetilde{\mathfrak{a}} = \{x \in K : x\mathfrak{a} \subset \mathcal{O}_K\}.$$

Writing $\mathfrak{a} = \mathfrak{p}_1\mathfrak{p}_2\cdots\mathfrak{p}_r$, where each $\mathfrak{p}_i$ is a nonzero prime ideal, $x\mathfrak{a} \subset \mathcal{O}_K$ if and only if $x\mathfrak{p}_1\mathfrak{p}_2\cdots\mathfrak{p}_r \subset \mathcal{O}_K$. This containment is equivalent to $x\mathfrak{p}_1\cdots\mathfrak{p}_{r-1} \subset \widetilde{\mathfrak{p}}_r$, and repeating this gets us to $x \in \widetilde{\mathfrak{p}}_1\widetilde{\mathfrak{p}}_2\cdots\widetilde{\mathfrak{p}}_r$. So $\widetilde{\mathfrak{a}} = \widetilde{\mathfrak{p}}_1\widetilde{\mathfrak{p}}_2\cdots\widetilde{\mathfrak{p}}_r$, and therefore

$$\mathfrak{a}\widetilde{\mathfrak{a}} = \mathfrak{p}_1\cdots\mathfrak{p}_r\widetilde{\mathfrak{p}}_1\cdots\widetilde{\mathfrak{p}}_r = (\mathfrak{p}_1\widetilde{\mathfrak{p}}_1)\cdots(\mathfrak{p}_r\widetilde{\mathfrak{p}}_r) = \mathcal{O}_K.$$

We henceforth write $\widetilde{\mathfrak{a}}$ as $\mathfrak{a}^{-1}$ and call it the inverse of $\mathfrak{a}$.

Remark 4.16. The inverse $\mathfrak{a}^{-1}$ is *not* the set of inverses of the (nonzero) elements of $\mathfrak{a}$, which is not even an additive group. For example, in $\mathbf{Q}$ we have $(2\mathbf{Z})^{-1} = \frac{1}{2}\mathbf{Z}$, and the only $\frac{1}{2k}$ in $\frac{1}{2}\mathbf{Z}$ is $\pm\frac{1}{2}$.

Set $\mathfrak{a}^n = (\mathfrak{a}^{-1})^{|n|}$ for $n < 0$. Now we can speak about arbitrary integral powers of a nonzero ideal $\mathfrak{a}$ in $\mathcal{O}_K$. Verify for all integers m and n that

$$\mathfrak{a}^m\mathfrak{a}^n = \mathfrak{a}^{m+n}, \quad (\mathfrak{a}^m)^n = \mathfrak{a}^{mn}, \quad (\mathfrak{a}\mathfrak{b})^n = \mathfrak{a}^n\mathfrak{b}^n. \tag{4.1}$$

For a nonzero proper ideal $\mathfrak{a}$ in $\mathcal{O}_K$, its inverse

$$\mathfrak{a}^{-1} = \{x \in K : x\mathfrak{a} \subset \mathcal{O}_K\}$$

does not lie in $\mathcal{O}_K$: $\mathfrak{a} \subsetneqq \mathcal{O}_K$ implies $\mathcal{O}_K \subsetneqq \mathfrak{a}^{-1}$. What kind of object is $\mathfrak{a}^{-1}$? It is an $\mathcal{O}_K$-module in K with the key property that it can be scaled to an ideal in $\mathcal{O}_K$: for any nonzero $\alpha \in \mathfrak{a}$, $\alpha\mathfrak{a}^{-1} \subset \mathcal{O}_K$, so $\alpha\mathfrak{a}^{-1}$ is an ideal in $\mathcal{O}_K$. We consider α to be a common denominator for $\mathfrak{a}^{-1}$ as an $\mathcal{O}_K$-module.

Definition 4.17. A *fractional ideal* in K is a nonzero $\mathcal{O}_K$-module $\mathfrak{a} \subset K$ admitting a common denominator: there is some nonzero $d \in \mathcal{O}_K$ such that $d\mathfrak{a} \subset \mathcal{O}_K$. A fractional ideal is called *principal* when it has the form $\mathcal{O}_K x$ for some $x \in K^\times$.

Example 4.18. One fractional ideal in $\mathbf{Q}$ is $\frac{2}{3}\mathbf{Z}$.

Example 4.19. Any nonzero ideal in $\mathcal{O}_K$ is a fractional ideal. Use $d = 1$.

Example 4.20. In $\mathbf{Z}[\sqrt{-5}]$, $(3, 1+\sqrt{5})(3, 1-\sqrt{-5}) = (3)$, so

$$(3, 1+\sqrt{-5})^{-1} = \frac{1}{3}(3, 1-\sqrt{-5}) = \mathbf{Z}[\sqrt{-5}] + \mathbf{Z}[\sqrt{-5}]\frac{1-\sqrt{-5}}{3}.$$

This is a fractional ideal with common denominator 3.

Nonexample 4.21. The field K is not a fractional ideal. While it is an $\mathcal{O}_K$-module, there is no $d \in \mathcal{O}_K - \{0\}$ such that $dK \subset \mathcal{O}_K$ since $dK = K$.

If we need to emphasize the distinction between nonzero ideals in $\mathcal{O}_K$ and the larger set of fractional ideals, which are not ideals unless they lie inside $\mathcal{O}_K$, we may call nonzero ideals in $\mathcal{O}_K$ *integral ideals.*

Since, in the notation of Definition 4.17, $\mathfrak{a} = \frac{1}{d} \cdot d\mathfrak{a}$ and $d\mathfrak{a}$ is an ideal in $\mathcal{O}_K$, fractional ideals in K are the $\mathcal{O}_K$-modules of the form $\frac{1}{d}I$ where $d \in \mathcal{O}_K - \{0\}$ and I is a nonzero ideal in $\mathcal{O}_K$. Concretely, a fractional ideal is just an ideal (not 0) divided by a number in $\mathcal{O}_K$. The principal fractional ideals in K are principal ideals divided by an element of $\mathcal{O}_K - \{0\}$, such as $\frac{2}{3}\mathbf{Z}$.

Theorem 4.22. *The fractional ideals in K are the nonzero finitely generated $\mathcal{O}_K$-modules in K.*

Proof. If $M \subset K$ is a nonzero finitely generated $\mathcal{O}_K$-module, then

$$M = \sum_{i=1}^{r} \mathcal{O}_K x_i$$

where $x_i \in K$ and some $x_i \neq 0$. Write $x_i = \alpha_i/d_i$ with $\alpha_i \in \mathcal{O}_K$ and $d_i \in \mathcal{O}_K - \{0\}$. Since there a finite number of d_i's, without loss of generality we can use a common denominator (such as the product of all d_i's) and thereby assume all d_i's equal, say $d_i = d$ for all i. Then

$$M = \sum_{i=1}^{r} \mathcal{O}_K \frac{\alpha_i}{d} = \frac{1}{d}\underbrace{\Big(\sum_{i=1}^{r} \mathcal{O}_K \alpha_i\Big)}_{\text{ideal in } \mathcal{O}_K}$$

and

$$dM = \sum_{i=1}^{r} \mathcal{O}_K \alpha_i \subset \mathcal{O}_K.$$

Conversely, any fractional ideal is $\frac{1}{d}I$ for some $d \in \mathcal{O}_K - \{0\}$ and ideal I. Since I is finitely generated as a $\mathbf{Z}$-module, it is finitely generated as an $\mathcal{O}_K$-module, so $\frac{1}{d}I$ is a finitely generated $\mathcal{O}_K$-module too. ■

Let's look at fractional ideals from the viewpoint of prime ideals. Any nonzero ideal I and principal ideal (d) in $\mathcal{O}_K$ can be factored as

$$I = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}, \quad (d) = \mathfrak{q}_1^{f_1} \cdots \mathfrak{q}_s^{f_s},$$

where $e_i, f_j \geqslant 0$. Then

$$\frac{1}{d} I = \frac{1}{d} \mathcal{O}_K \cdot I = (d)^{-1} I = \mathfrak{q}_1^{-f_1} \cdots \mathfrak{q}_s^{-f_s} \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}.$$

So while the nonzero ideals in $\mathcal{O}_K$ are $\{\mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r} : \mathfrak{p}_i \text{ prime}, e_i \geqslant 0\}$, the fractional ideals in K are

$$\{\mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r} : \mathfrak{p}_i \text{ prime}, e_i \in \mathbf{Z}\} = \left\{ IJ^{-1} : I, J \subset \mathcal{O}_K \text{ nonzero ideals} \right\},$$

which is a group under multiplication of $\mathcal{O}_K$-modules and is freely generated by the prime ideals (no multiplicative relations by unique factorization).

Using inverses of ideals, we get a crucial equivalence between divisibility and containment of ideals in $\mathcal{O}_K$.

Theorem 4.23. *For nonzero ideals $\mathfrak{a}$ and $\mathfrak{b}$ in $\mathcal{O}_K$, $\mathfrak{a} \supset \mathfrak{b}$ if and only if $\mathfrak{a} \mid \mathfrak{b}$.*

Proof. The direction ($\Leftarrow$) is true in all commutative rings. For ($\Rightarrow$), set $\mathfrak{c} = \mathfrak{a}^{-1}\mathfrak{b} \subset \mathfrak{a}^{-1}\mathfrak{a} = \mathcal{O}_K$, so $\mathfrak{c}$ is an $\mathcal{O}_K$-module in $\mathcal{O}_K$ and thus an ideal. Easily $\mathfrak{a}\mathfrak{c} = \mathfrak{b}$. ■

In words: "to contain is to divide" for nonzero ideals in $\mathcal{O}_K$.

Nonexample 4.24. Consider $\mathbf{Z}[\sqrt{d}]$ with nonsquare $d \equiv 1 \bmod 4$. This is not the ring of integers of $\mathbf{Q}(\sqrt{d})$ (it has index 2 in it if d is squarefree and has higher index if d has a square factor). Let $\mathfrak{p} = (2, 1 + \sqrt{d})$, so $(2) \subset \mathfrak{p}$. We will show $\mathfrak{p} \nmid (2)$. Check, using aligned $\mathbf{Z}$-bases, successive ideals below have index 2:

$$\mathfrak{p}^2 \underset{2}{\subsetneq} (2) \underset{2}{\subsetneq} \mathfrak{p} \underset{2}{\subsetneq} \mathbf{Z}[\sqrt{d}].$$

(Here we need $d \equiv 1 \bmod 4$. If d were even then $\mathfrak{p} = (1)$, and if $d \equiv 3 \bmod 4$ then $\mathfrak{p}^2 = (2)$.) Suppose $\mathfrak{p} \mid (2)$, so $(2) = \mathfrak{p}\mathfrak{b}$ where $\mathfrak{b}$ is a nonzero proper ideal. Let $\mathfrak{b} \subset \mathfrak{q}$, where $\mathfrak{q}$ is some maximal ideal. Then $\mathfrak{p}^2 \subset (2) \subset \mathfrak{b} \subset \mathfrak{q}$, so $\mathfrak{q} = \mathfrak{p}$ (Corollary 4.5). So $(2) = \mathfrak{p}\mathfrak{b} \subset \mathfrak{p}\mathfrak{p} = \mathfrak{p}^2 \subsetneqq (2)$, which is a contradiction. Hence $\mathfrak{p} \nmid (2)$.

In fact, (2) has no prime ideal factorization in $\mathbf{Z}[\sqrt{d}]$. It doesn't even have a prime ideal factor: if $\mathfrak{p}'$ is a prime factor of (2) then $\mathfrak{p}' \supset (2) \supset \mathfrak{p}^2$, so $\mathfrak{p}' = \mathfrak{p}$ (Corollary 4.5 again), but we already saw $\mathfrak{p} \nmid (2)$. (By Exercise 1.25, $\mathfrak{p}^2 = 2\mathfrak{p}$, so $\mathfrak{p}$ has no inverse as a $\mathbf{Z}[\sqrt{d}]$-module in $\mathbf{Q}(\sqrt{d})$ since otherwise $\mathfrak{p} = (2)$.)

That containment is the same as divisibility in $\mathcal{O}_K$ has a lot of basic consequences.

Corollary 4.25. *Every nonzero ideal in $\mathcal{O}_K$ is a factor of a nonzero principal ideal.*

Proof. Let $\mathfrak{a} \neq (0)$ and let α be any nonzero element of $\mathfrak{a}$. Then $(\alpha) \subset \mathfrak{a}$, so $\mathfrak{a} \mid (\alpha)$. ■

That every nonzero ideal in $\mathcal{O}_K$ can be found as a factor of a principal ideal ties all ideals in $\mathcal{O}_K$ to principal ideals.

Example 4.26. In $\mathcal{O}_K$, any nonzero prime ideal $\mathfrak{p}$ can be found as the factor of a principal ideal $(p) = p\mathcal{O}_K$ where p is a prime number. Indeed, $\mathfrak{p} \cap \mathbf{Z} = p\mathbf{Z}$ for some prime number p. Then $p \in \mathfrak{p}$ so $\mathfrak{p} \mid (p)$. Finding all prime ideals in $\mathcal{O}_K$ amounts to knowing how the particular ideals $p\mathcal{O}_K$ factor as p runs through the prime numbers. In Section 4.5, we will see how this works in practice.

Corollary 4.27. *For nonzero ideals $\mathfrak{a}$ and $\mathfrak{b}$ in $\mathcal{O}_K$,* $\mathrm{N}(\mathfrak{a}\mathfrak{b}) = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{b})$.

This is called total multiplicativity of the ideal norm. (The label "total" emphasizes that we have no constraints on $\mathfrak{a}$ and $\mathfrak{b}$. Many functions in number theory are multiplicative only on products of relatively prime numbers or relatively prime ideals.)

Proof. We may suppose $\mathfrak{a}$ is a proper ideal and, by ideal factorization, that $\mathfrak{b}$ is a prime ideal $\mathfrak{p}$. We have

$$\mathfrak{a}\mathfrak{p} \underset{\neq}{\subset} \mathfrak{a} \underset{\neq}{\subset} \mathcal{O}_K.$$

By definition $[\mathcal{O}_K : \mathfrak{a}\mathfrak{p}] = \mathrm{N}(\mathfrak{a}\mathfrak{p})$ and $[\mathcal{O}_K : \mathfrak{a}] = \mathrm{N}(\mathfrak{a})$. To say $\mathrm{N}(\mathfrak{a}\mathfrak{p}) = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{p})$ is therefore the same as saying $[\mathfrak{a} : \mathfrak{a}\mathfrak{p}] = \mathrm{N}(\mathfrak{p})$. Let's look at $\mathfrak{a}/\mathfrak{a}\mathfrak{p}$. This $\mathcal{O}_K$-module is killed by $\mathfrak{p}$, so it's an $(\mathcal{O}_K/\mathfrak{p})$-vector space. We want to show its size is $\#(\mathcal{O}_K/\mathfrak{p})$, *i.e.*, $\dim_{\mathcal{O}_K/\mathfrak{p}}(\mathfrak{a}/\mathfrak{a}\mathfrak{p}) = 1$.

What are the $(\mathcal{O}_K/\mathfrak{p})$-subspaces of $\mathfrak{a}/\mathfrak{a}\mathfrak{p}$? Since multiplication by $\mathfrak{p}$ kills every element of $\mathfrak{a}/\mathfrak{a}\mathfrak{p}$, the $(\mathcal{O}_K/\mathfrak{p})$-subspaces of $\mathfrak{a}/\mathfrak{a}\mathfrak{p}$ are the same as the $\mathcal{O}_K$-submodules of $\mathfrak{a}/\mathfrak{a}\mathfrak{p}$. From the general correspondence between submodules of a module and a quotient module, the $\mathcal{O}_K$-submodules of $\mathfrak{a}/\mathfrak{a}\mathfrak{p}$ are all uniquely of the form $\mathfrak{b}/\mathfrak{a}\mathfrak{p}$ where $\mathfrak{b}$ is an $\mathcal{O}_K$-module with $\mathfrak{a}\mathfrak{p} \subset \mathfrak{b} \subset \mathfrak{a}$. Such $\mathfrak{b}$ are ideals in $\mathcal{O}_K$, and these containments are the same as $\mathfrak{a} \mid \mathfrak{b} \mid \mathfrak{a}\mathfrak{p}$, which tells us that $\mathfrak{b} = \mathfrak{a}$ or $\mathfrak{b} = \mathfrak{a}\mathfrak{p}$ from unique factorization. Therefore $\mathfrak{b}/\mathfrak{a}\mathfrak{p}$ equals $\mathfrak{a}/\mathfrak{a}\mathfrak{p}$ or $\{\overline{0}\}$. We have shown the only $(\mathcal{O}_K/\mathfrak{p})$-subspaces of $\mathfrak{a}/\mathfrak{a}\mathfrak{p}$ are the whole space and the zero subspace. Therefore $\mathfrak{a}/\mathfrak{a}\mathfrak{p}$ is 1-dimensional over $\mathcal{O}_K/\mathfrak{p}$. ■

Corollary 4.28. *Let $\mathfrak{a}$ be an ideal in $\mathcal{O}_K$ and p be a prime number. The ideal $\mathfrak{a}$ is divisible by some prime factor of (p) if and only if* $\mathrm{N}(\mathfrak{a})$ *is divisible by p.*

Proof. Write $\mathfrak{a} = \mathfrak{p}_1^{a_1} \cdots \mathfrak{p}_r^{a_r}$ with distinct $\mathfrak{p}_i$'s. Since $\mathrm{N}(\mathfrak{a})$ is the product of every $\mathrm{N}(\mathfrak{p}_i)^{a_i}$, $p \mid \mathrm{N}(\mathfrak{a})$ if and only if some $\mathrm{N}(\mathfrak{p}_i)$ is divisible by p. By Theorem 3.16, $p \mid \mathrm{N}(\mathfrak{p}_i)$ if and only if $\mathcal{O}_K/\mathfrak{p}_i$ has characteristic p, which is equivalent to $p \equiv 0 \bmod \mathfrak{p}_i$, and

$$p \equiv 0 \bmod \mathfrak{p}_i \iff (p) \subset \mathfrak{p}_i \iff \mathfrak{p}_i \mid (p).$$

■

Example 4.29. In $\mathbf{Z}[\sqrt{10}]$, let $\mathfrak{a} = (2 + 5\sqrt{10}, 4 + 7\sqrt{10})$. In Example 3.11, we saw $\mathrm{N}(\mathfrak{a}) = 6$. Since 6 is not a prime power, $\mathfrak{a}$ can't be a prime ideal. Thus $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$ where $\mathfrak{b}$ and $\mathfrak{c}$ are proper ideals in $\mathcal{O}_K$: their norms are greater than 1. By multiplicativity of the ideal norm, $6 = \mathrm{N}(\mathfrak{b})\,\mathrm{N}(\mathfrak{c})$. Therefore $\mathfrak{b}$ and $\mathfrak{c}$ have norms 2 and 3 (in some order), which are prime numbers, so $\mathfrak{b}$ and $\mathfrak{c}$ are prime ideal factors of (2) and (3). The factorization $\mathfrak{b}\mathfrak{c}$ is the prime ideal factorization of $\mathfrak{a}$. We know $\mathfrak{b}$ and $\mathfrak{c}$ exist, but we have no formula for them. They will be found in Example 4.43.

Factor two nonzero ideals $\mathfrak{a}$ and $\mathfrak{b}$ in $\mathcal{O}_K$ as $\prod \mathfrak{p}_i^{a_i}$ and $\prod \mathfrak{p}_i^{b_i}$ with nonnegative a_i and b_i. Define

$$\gcd(\mathfrak{a}, \mathfrak{b}) = \prod_i \mathfrak{p}_i^{\min(a_i, b_i)} \qquad \text{and} \qquad \operatorname{lcm}(\mathfrak{a}, \mathfrak{b}) = \prod_i \mathfrak{p}_i^{\max(a_i, b_i)}.$$

The first ideal is a common ideal divisor of $\mathfrak{a}$ and $\mathfrak{b}$ which all other common ideal divisors divide, and the second ideal is a common ideal multiple of $\mathfrak{a}$ and $\mathfrak{b}$ which all other common ideal multiples of $\mathfrak{a}$ and $\mathfrak{b}$ are multiples of. So calling these two ideals the gcd and lcm of $\mathfrak{a}$ and $\mathfrak{b}$ is reasonable.

Corollary 4.30. *For any nonzero ideals $\mathfrak{a}$ and $\mathfrak{b}$ in $\mathcal{O}_K$,* $\gcd(\mathfrak{a}, \mathfrak{b}) = \mathfrak{a} + \mathfrak{b}$ *and* $\operatorname{lcm}(\mathfrak{a}, \mathfrak{b}) = \mathfrak{a} \cap \mathfrak{b}$.

Proof. For any nonzero ideal $\mathfrak{d}$ in $\mathcal{O}_K$, comparing its prime factorization with those of $\mathfrak{a}$ and $\mathfrak{b}$ shows $\mathfrak{d} \mid \mathfrak{a}$ and $\mathfrak{d} \mid \mathfrak{b}$ if and only if $\mathfrak{d} \mid \gcd(\mathfrak{a}, \mathfrak{b})$. At the same time, because containment is the same as divisbility,

$$\mathfrak{d} \mid \mathfrak{a} \text{ and } \mathfrak{d} \mid \mathfrak{b} \Longleftrightarrow \mathfrak{d} \supset \mathfrak{a} \text{ and } \mathfrak{d} \supset \mathfrak{b} \Longleftrightarrow \mathfrak{d} \supset \mathfrak{a} + \mathfrak{b} \Longleftrightarrow \mathfrak{d} \mid (\mathfrak{a} + \mathfrak{b}).$$

Thus $\mathfrak{d} \mid \gcd(\mathfrak{a}, \mathfrak{b})$ if and only if $\mathfrak{d} \mid (\mathfrak{a} + \mathfrak{b})$. Taking $\mathfrak{d} = \gcd(\mathfrak{a}, \mathfrak{b})$ shows $\gcd(\mathfrak{a}, \mathfrak{b})$ divides $\mathfrak{a} + \mathfrak{b}$, and taking $\mathfrak{d} = \mathfrak{a} + \mathfrak{b}$ shows $(\mathfrak{a} + \mathfrak{b}) \mid \gcd(\mathfrak{a}, \mathfrak{b})$. Therefore $\gcd(\mathfrak{a}, \mathfrak{b})$ and $\mathfrak{a} + \mathfrak{b}$ contain each other, so they are equal. The argument for the least common multiple is left as Exercise 4.39. ■

For nonzero principal ideals (α) and (β) in $\mathcal{O}_K$, $\gcd((\alpha), (\beta)) = (\alpha) + (\beta)$, which is (α, β) by the definition of what the ideal (α, β) means. Following the notation of elementary number theory, we will usually write $\gcd(\mathfrak{a}, \mathfrak{b})$ as $(\mathfrak{a}, \mathfrak{b})$, so $(\mathfrak{a}, \mathfrak{b}) = \mathfrak{a} + \mathfrak{b}$. In particular, on principal ideals $((\alpha), (\beta)) = (\alpha, \beta)$ and this is not a tautology.

Nonzero ideals $\mathfrak{a}$ and $\mathfrak{b}$ in $\mathcal{O}_K$ are called *relatively prime* when $(\mathfrak{a}, \mathfrak{b}) = (1)$, which is the same[2] as $\mathfrak{a} + \mathfrak{b} = (1)$, which is equivalent to $x + y = 1$ for some $x \in \mathfrak{a}$ and $y \in \mathfrak{b}$. We say an element $x \in \mathcal{O}_K - \{0\}$ is relatively prime to a nonzero ideal $\mathfrak{a}$ when $((x), \mathfrak{a}) = (1)$. For example,

$$\begin{aligned}(\mathcal{O}_K/\mathfrak{a})^\times &= \{x \bmod \mathfrak{a} : xy \equiv 1 \bmod \mathfrak{a} \text{ for some } y \in \mathcal{O}_K\} \\ &= \{x \bmod \mathfrak{a} : (x) + \mathfrak{a} = (1)\} \\ &= \{x \bmod \mathfrak{a} : ((x), \mathfrak{a}) = (1)\},\end{aligned}$$

so the units mod $\mathfrak{a}$ are the cosets represented by elements relatively prime to the modulus.

Theorem 4.31 (Chinese remainder theorem). *Let A be a commutative ring. If $\mathfrak{a}$ and $\mathfrak{b}$ are ideals in A such that $\mathfrak{a} + \mathfrak{b} = (1)$, then $A/\mathfrak{a}\mathfrak{b} \cong A/\mathfrak{a} \times A/\mathfrak{b}$ as commutative rings by the map*

$$x \bmod \mathfrak{a}\mathfrak{b} \mapsto (x \bmod \mathfrak{a},\ x \bmod \mathfrak{b}).$$

In particular, for relatively prime ideals $\mathfrak{a}$ and $\mathfrak{b}$ in $\mathcal{O}_K$, $\mathcal{O}_K/\mathfrak{a}\mathfrak{b} \cong \mathcal{O}_K/\mathfrak{a} \times \mathcal{O}_K/\mathfrak{b}$ as commutative rings.

[2] In $\mathbf{Z}[T]$, 2 and T are relatively prime elements while $(2) + (T) = (2, T)$ is not (1), so relatively prime elements in a general UFD don't necessarily generate the unit ideal. It goes through for PIDs, though.

Proof. Since $\mathfrak{ab}$ is a subset of $\mathfrak{a}$ and $\mathfrak{b}$, the indicated mapping $A/\mathfrak{ab} \to A/\mathfrak{a} \times A/\mathfrak{b}$ is well-defined and it is easy to see it's a ring homomorphism.

Let $\alpha + \beta = 1$ for some $\alpha \in \mathfrak{a}$ and $\beta \in \mathfrak{b}$. Then $\alpha \equiv 0 \bmod \mathfrak{a}$ and $\alpha = 1 - \beta \equiv 1 \bmod \mathfrak{b}$, so $\alpha \bmod \mathfrak{ab} \mapsto (0,1)$. Similarly, $\beta \bmod \mathfrak{ab} \mapsto (1,0)$, so $b\alpha + a\beta \mapsto (a,b)$. Thus the ring homomorphism is surjective.

If $x \bmod \mathfrak{ab} \mapsto (0,0)$ then $x \in \mathfrak{a} \cap \mathfrak{b}$. So proving injectivity amounts to showing $\mathfrak{a} \cap \mathfrak{b} = \mathfrak{ab}$. The inclusion $\supset$ is clear. For the reverse inclusion, write any $x \in \mathfrak{a} \cap \mathfrak{b}$ in the form

$$x = x \cdot 1 = x \cdot (\alpha + \beta) = x\alpha + x\beta \in \mathfrak{ba} + \mathfrak{ab} = \mathfrak{ab}. \quad \blacksquare$$

The Chinese remainder theorem is true for more than 2 ideals: if $\mathfrak{a}_1, \ldots, \mathfrak{a}_r$ are ideals in A such that $\mathfrak{a}_i + \mathfrak{a}_j = A$ for $i \neq j$ then $\bigcap_{i=1}^{r} \mathfrak{a}_i = \mathfrak{a}_1 \cdots \mathfrak{a}_r$ and the natural ring homomorphism $A/\mathfrak{a}_1 \cdots \mathfrak{a}_r \to A/\mathfrak{a}_1 \times \cdots \times A/\mathfrak{a}_r$ is an isomorphism. This is proved by induction on r, with the base case $r = 2$ being Theorem 4.31. Applications of the Chinese remainder theorem will often use this more general version (*e.g.*, Theorems 4.88 and 4.89).

For nonzero ideals $\mathfrak{a}$ and $\mathfrak{b}$ in $\mathcal{O}_K$, the Chinese remainder theorem implies $\mathrm{N}(\mathfrak{ab}) = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{b})$ when $\mathfrak{a}$ and $\mathfrak{b}$ are relatively prime, but it's important to remember the ideal norm is multiplicative on any two nonzero ideals in $\mathcal{O}_K$ (Corollary 4.27). If $(\mathfrak{a}, \mathfrak{b}) \neq (1)$, the rings $\mathcal{O}_K/\mathfrak{ab}$ and $\mathcal{O}_K/\mathfrak{a} \times \mathcal{O}_K/\mathfrak{b}$ are *not* isomorphic, but they have equal size by total multiplicativity of the ideal norm. (As an example of what can go wrong, $12 = 6 \cdot 2$ and $\mathbf{Z}/12\mathbf{Z} \not\cong \mathbf{Z}/6\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$ as rings since they're not even isomorphic as additive groups: the left side is cyclic of order 12 while on the right side every element is killed by 6.)

Corollary 4.32. *When* $(\mathfrak{a}, \mathfrak{b}) = (1)$, *the natural map* $(\mathcal{O}_K/\mathfrak{ab})^\times \to (\mathcal{O}_K/\mathfrak{a})^\times \times (\mathcal{O}_K/\mathfrak{b})^\times$ *is a group isomorphism.*

Proof. The natural map $\mathcal{O}_K/\mathfrak{ab} \to \mathcal{O}_K/\mathfrak{a} \times \mathcal{O}_K/\mathfrak{b}$ is a ring isomorphism. Restrict it to the unit groups. $\blacksquare$

Example 4.33. For a number field K and nonzero ideal $\mathfrak{a}$ in $\mathcal{O}_K$, set $\varphi_K(\mathfrak{a}) = \#(\mathcal{O}_K/\mathfrak{a})^\times$. This generalizes Euler's φ-function, which is the case $K = \mathbf{Q}$. By group theory, $\alpha^{\varphi_K(\mathfrak{a})} \equiv 1 \bmod \mathfrak{a}$ if $((\alpha), \mathfrak{a}) = (1)$.

By Corollary 4.32, $\varphi_K(\mathfrak{ab}) = \varphi_K(\mathfrak{a})\varphi_K(\mathfrak{b})$ if $\mathfrak{a}$ and $\mathfrak{b}$ are relatively prime, so computing $\varphi_K(\mathfrak{a})$ reduces to the case that $\mathfrak{a} = \mathfrak{p}^r$ is a prime power with $r \geqslant 1$, just like in $\mathbf{Z}$. Since $\mathcal{O}_K/\mathfrak{p}$ is a field, $\varphi_K(\mathfrak{p}) = \mathrm{N}(\mathfrak{p}) - 1$. For higher powers, we compute $\varphi_K(\mathfrak{p}^r)$ by looking for nonunits mod $\mathfrak{p}^r$. To say x mod $\mathfrak{p}^r$ is not a unit is the same as saying $((x), \mathfrak{p}^r) \neq (1)$, which is equivalent to $\mathfrak{p} \mid (x)$, so

$$\begin{aligned}\varphi_K(\mathfrak{p}^r) &= \#(\mathcal{O}_K/\mathfrak{p}^r) - \#(\mathfrak{p}/\mathfrak{p}^r) \\ &= \mathrm{N}(\mathfrak{p}^r) - \mathrm{N}(\mathfrak{p}^r)/\,\mathrm{N}(\mathfrak{p}) \\ &= \mathrm{N}(\mathfrak{p})^r - \mathrm{N}(\mathfrak{p})^{r-1} \\ &= \mathrm{N}(\mathfrak{p}^r)\left(1 - \frac{1}{\mathrm{N}(\mathfrak{p})}\right),\end{aligned}$$

so

$$\varphi_K(\mathfrak{a}) = \mathrm{N}(\mathfrak{a}) \prod_{\mathfrak{p}|\mathfrak{a}} \left(1 - \frac{1}{\mathrm{N}(\mathfrak{p})}\right),$$

a nice generalization of the classical formula $\varphi(m) = m\prod_{p|m}(1 - 1/p)$.

While the formula for $\#(\mathcal{O}_K/\mathfrak{a})^\times$ is an exact parallel to that for $\#(\mathbf{Z}/m\mathbf{Z})^\times$, there is a big difference in the *structure* of the unit groups: for any odd prime power p^r, $(\mathbf{Z}/p^r\mathbf{Z})^\times$ is a cyclic group, but the groups $(\mathcal{O}_K/\mathfrak{p}^r)^\times$ are often noncyclic when $r > 1$. (They are cyclic for $r = 1$ since $(\mathcal{O}_K/\mathfrak{p})^\times$ is the multiplicative group of a finite field, which is well-known to be cyclic.) See Exercise 4.26.

4.4 Finding Primes in $\mathbf{Z}[\sqrt{-5}]$

To illustrate how prime ideals can be discovered, we will work out examples in $\mathbf{Z}[\sqrt{-5}]$. Corollary 4.25 assures us that factoring principal ideals will reveal all prime ideals to us. Actually, any nonzero prime ideal $\mathfrak{p}$ in $\mathcal{O}_K$ occurs as the factor of a principal ideal of the form (p) for p a prime number (Example 4.26), but we will focus in this section on finding prime ideals in $\mathbf{Z}[\sqrt{-5}]$ from factoring somewhat general principal ideals. In Section 4.5 we will see how to find prime ideals by factoring the specific principal ideals (p).

The degree of a number field bounds the exponent in the norm of a prime ideal. When $n = [K : \mathbf{Q}]$, $\mathrm{N}((p)) = p^n$, so for any prime ideal $\mathfrak{p}$ dividing (p) we have $\mathrm{N}(\mathfrak{p}) \mid p^n$, so $\mathrm{N}(\mathfrak{p}) \in \{p, p^2, \ldots, p^n\}$. For example, if $\mathfrak{p} \mid (p)$ in $\mathbf{Z}[\sqrt{-5}]$ then $\mathrm{N}(\mathfrak{p}) = p$ or p^2. We're now ready to begin.

Since $\mathrm{N}((1 + \sqrt{-5})) = 6 = 2 \cdot 3$, the prime ideal factors of $(1 + \sqrt{-5})$ must have norm 2 and 3; there is no such thing as a prime ideal with norm 6 because a prime ideal has prime power norm (or use Corollary 4.28). Therefore

$$(1 + \sqrt{-5}) = \mathfrak{p}_2\mathfrak{p}_3, \text{ where } \mathrm{N}(\mathfrak{p}_2) = 2 \text{ and } \mathrm{N}(\mathfrak{p}_3) = 3.$$

Notice we know $\mathfrak{p}_2$ and $\mathfrak{p}_3$ exist without having explicit formulas for these ideals: a norm computation showed these prime ideals exist, and unique factorization shows they are uniquely determined as the prime ideal factors of $(1+\sqrt{-5})$. We have $2 \in \mathfrak{p}_2$ and $3 \in \mathfrak{p}_3$ ($p \in \mathfrak{p}$ when $\mathrm{N}(\mathfrak{p})$ is a power of p, because $p = 0$ in $\mathcal{O}_K/\mathfrak{p}$), so $\mathfrak{p}_2 \mid (2)$ and $\mathfrak{p}_3 \mid (3)$.

Consider next $\mathrm{N}((1-\sqrt{-5})) = 6 = 2 \cdot 3$. So $(1-\sqrt{-5})$ is a product of a prime of norm 2 and a prime of norm 3, but are these primes $\mathfrak{p}_2$ and $\mathfrak{p}_3$, or are they new prime ideals? We figure this out by working modulo $\mathfrak{p}_2$ and $\mathfrak{p}_3$. We will determine if $1-\sqrt{-5} \equiv 0 \bmod \mathfrak{p}_2$ or if $1-\sqrt{-5} \equiv 0 \bmod \mathfrak{p}_3$. Since $\mathfrak{p}_2 \mid (1+\sqrt{-5})$, we have $1+\sqrt{-5} \equiv 0 \bmod \mathfrak{p}_2$, so $\sqrt{-5} \equiv -1 \equiv 1 \bmod \mathfrak{p}_2$ (since $2 \in \mathfrak{p}_2$). Thus $1-\sqrt{-5} \equiv 2 \equiv 0 \bmod \mathfrak{p}_2$, which means $(1-\sqrt{-5}) \subset \mathfrak{p}_2$, so $\mathfrak{p}_2 \mid (1-\sqrt{-5})$. Since $1+\sqrt{-5} \equiv 0 \bmod \mathfrak{p}_3$, $1-\sqrt{-5} \equiv 1-(-1) \equiv 2 \not\equiv 0 \bmod \mathfrak{p}_3$ because $\mathbf{Z}[\sqrt{-5}]/\mathfrak{p}_3$ has characteristic 3. Thus $\mathfrak{p}_3 \nmid (1-\sqrt{-5})$, so

$$(1-\sqrt{-5}) = \mathfrak{p}_2\mathfrak{p}_3', \text{ where } \mathrm{N}(\mathfrak{p}_3') = 3.$$

(Primes decorated in a new way are usually understood to be new primes: $\mathfrak{p}_3' \neq \mathfrak{p}_3$.)

Multiply the factorizations of $(1+\sqrt{-5})$ and $(1-\sqrt{-5})$:

$$\begin{aligned}(1+\sqrt{-5})(1-\sqrt{-5}) &= \mathfrak{p}_2\mathfrak{p}_3\mathfrak{p}_2\mathfrak{p}_3' \\ (6) &= \mathfrak{p}_2^2\mathfrak{p}_3\mathfrak{p}_3' \\ (2)(3) &= \mathfrak{p}_2^2\mathfrak{p}_3\mathfrak{p}_3'.\end{aligned}$$

The ideals (2) and (3) are relatively prime since 2 and 3 are already relatively prime in $\mathbf{Z}$. So from $\mathfrak{p}_2 \mid (2)$, $\mathfrak{p}_3 \mid (3)$, and $\mathfrak{p}_3' \mid (3)$, we must have

$$(2) = \mathfrak{p}_2^2, \quad (3) = \mathfrak{p}_3\mathfrak{p}_3'.$$

Although we found the factorizations

$$\boxed{(1+\sqrt{-5}) = \mathfrak{p}_2\mathfrak{p}_3 \text{ and } (1-\sqrt{-5}) = \mathfrak{p}_2\mathfrak{p}_3',}$$

we *don't know what these prime ideals are*, but from the additional factorizations $\boxed{(2) = \mathfrak{p}_2^2 \text{ and } (3) = \mathfrak{p}_3\mathfrak{p}_3'}$ we see that the prime ideals are gcds of two explicit principal ideals:

$$\boxed{\mathfrak{p}_2 = \gcd((2), (1+\sqrt{-5})) = (2, 1+\sqrt{-5}) = (2, 1-\sqrt{-5})}$$

and

$$\boxed{\mathfrak{p}_3 = (3, 1+\sqrt{-5}) \quad \text{and} \quad \mathfrak{p}_3' = (3, 1-\sqrt{-5}).}$$

These ideals $\mathfrak{p}_2$, $\mathfrak{p}_3$, and $\mathfrak{p}_3'$ are not principal since they have norm 2 and 3 but there are no solutions to $x^2 + 5y^2 = 2$ or 3 in $\mathbf{Z}$. Since $\mathfrak{p}_2$ is the unique prime factor of (2), any ideal with even norm must be divisible by $\mathfrak{p}_2$. On the other hand, an ideal whose norm is divisible by 3 will be divisible by either $\mathfrak{p}_3$ or $\mathfrak{p}_3'$, but you can't say which one without a closer look.

For any ideal $\mathfrak{a}$ in $\mathbf{Z}[\sqrt{-5}]$, its conjugate $\overline{\mathfrak{a}} = \{\overline{\alpha} : \alpha \in \mathfrak{a}\}$ is also an ideal. For example, $\overline{\mathfrak{p}}_3 = \mathfrak{p}_3'$ and $\overline{\mathfrak{p}_3'} = \mathfrak{p}_3$. But notice $\overline{\mathfrak{p}}_2 = \mathfrak{p}_2$, so $\mathfrak{p}_2$ is its own complex conjugate without having a set of real generators. (Real numbers in $\mathbf{Z}[\sqrt{-5}]$ are in $\mathbf{Z}$, and an ideal with generators in $\mathbf{Z}$ is principal, which $\mathfrak{p}_2$ is not.) Galois theory would vaguely suggest "anything" fixed by complex conjugation should be defined over the reals, but it just isn't true for ideals.

Now consider $\mathrm{N}((2+\sqrt{-5})) = 9$. The possibilities for a prime ideal factorization of $(2+\sqrt{-5})$ are $\mathfrak{p}_3^2$, $\mathfrak{p}_3\mathfrak{p}_3'$ or $\mathfrak{p}_3'^2$, since an ideal with norm 9 can only be a product of prime ideals dividing (3) (having 3-power norm, that is), and those three products are the only combinations of $\mathfrak{p}_3$ and $\mathfrak{p}_3'$ with norm 9. We already know $\mathfrak{p}_3\mathfrak{p}_3' = (3)$, and $(2+\sqrt{-5}) \neq (3)$ since $2+\sqrt{-5} \neq 3u$ for any $u \in \mathbf{Z}[\sqrt{-5}]^\times = \{\pm 1\}$, so $(2+\sqrt{-5})$ has to be either $\mathfrak{p}_3^2$ or $\mathfrak{p}_3'^2$. Which is it?

The answer depends on whether $\mathfrak{p}_3 \mid (2+\sqrt{-5})$ or $\mathfrak{p}_3' \mid (2+\sqrt{-5})$, *i.e.*, whether $2+\sqrt{-5} \equiv 0 \bmod \mathfrak{p}_3$ or $2+\sqrt{-5} \equiv 0 \bmod \mathfrak{p}_3'$. Since $1+\sqrt{-5} \equiv 0 \bmod \mathfrak{p}_3$, $\sqrt{-5} \equiv -1 \bmod \mathfrak{p}_3$. Therefore $2+\sqrt{-5} \equiv 1 \not\equiv 0 \bmod \mathfrak{p}_3$, so we *must* have $2+\sqrt{-5} \equiv 0 \bmod \mathfrak{p}_3'$. Thus $\boxed{(2+\sqrt{-5}) = \mathfrak{p}_3'^2.}$ While $\mathfrak{p}_3'$ is not principal, we now see that its square is principal. We already saw $\mathfrak{p}_2^2 = (2)$ is principal.

Squaring $(3) = \mathfrak{p}_3\mathfrak{p}_3'$ gives us $(9) = \mathfrak{p}_3^2\mathfrak{p}_3'^2 = \mathfrak{p}_3^2(2+\sqrt{-5})$, so $\mathfrak{p}_3^2 = (\alpha)$ for $\alpha = 9/(2+\sqrt{-5}) = 2-\sqrt{-5}$: $\boxed{(2-\sqrt{-5}) = \mathfrak{p}_3^2.}$

It is easy to find primes dividing (5), since $(5) = (\sqrt{-5})^2$. From $\mathrm{N}((\sqrt{-5})) = 5$, the ideal

$$\boxed{\mathfrak{p}_5 = (\sqrt{-5})}$$

is prime and $\boxed{(5) = \mathfrak{p}_5^2,}$ so $\mathfrak{p}_5$ is the only prime factor of (5) and it is principal.

How do we find a prime $\mathfrak{p}$ such that $\mathfrak{p} \mid (7)$? We look for (α) where $\mathrm{N}((\alpha))$ is divisible by 7. From $\mathrm{N}((3+\sqrt{-5})) = 14$ and $\mathrm{N}((1+2\sqrt{-5})) = 21$, $\boxed{(3+\sqrt{-5}) = \mathfrak{p}_2\mathfrak{p}_7,}$ where $\mathrm{N}(\mathfrak{p}_7) = 7$ and $(1+2\sqrt{-5})$ is the product of a prime of norm 3 and a prime of norm 7, but we don't know which ideals of norms 3 and 7 they are. (Having already defined $\mathfrak{p}_7$ as the prime of norm 7 dividing $(3+\sqrt{-5})$, we have to be careful about whether or not this is the same prime of norm 7 in

$(1+2\sqrt{-5})$.) The ring $\mathbf{Z}[\sqrt{-5}]/\mathfrak{p}_7 \cong \mathbf{Z}/(7)$ has $\sqrt{-5} \equiv -3 \equiv 4 \bmod \mathfrak{p}_7$. So

$$1+2\sqrt{-5} \equiv 1+2\cdot 4 = 9 \not\equiv 0 \bmod \mathfrak{p}_7.$$

Thus $\mathfrak{p}_7 \nmid (1+2\sqrt{-5})$. In $\mathbf{Z}[\sqrt{-5}]/\mathfrak{p}_3$, $\sqrt{-5} \equiv -1 \equiv 2 \bmod \mathfrak{p}_3$, so

$$1+2\sqrt{-5} \equiv 1+2\cdot 2 \equiv 5 \not\equiv 0 \bmod \mathfrak{p}_3.$$

Thus $\mathfrak{p}_3 \nmid (1+2\sqrt{-5})$. So $\boxed{(1+2\sqrt{-5}) = \mathfrak{p}_3'\mathfrak{p}_7',}$ where $\mathfrak{p}_7'$ is a new prime of norm 7. Since the prime ideals $\mathfrak{p}_7$ and $\mathfrak{p}_7'$ each divide (7) and are not equal, their product divides (7): $(7) = \mathfrak{p}_7\mathfrak{p}_7'\mathfrak{a}$. Taking norms, $49 = 7\cdot 7\cdot \mathrm{N}(\mathfrak{a})$. Thus $\mathrm{N}(\mathfrak{a}) = 1$, so $\mathfrak{a} = (1)$: $\boxed{(7) = \mathfrak{p}_7\mathfrak{p}_7'.}$ Now we can express $\mathfrak{p}_7$ and $\mathfrak{p}_7'$ as gcds:

$$\mathfrak{p}_7 = (7, 3+\sqrt{-5}) \text{ and } \mathfrak{p}_7' = (7, 1+2\sqrt{-5}).$$

These two formulas don't look as symmetric as our gcd formulas for $\mathfrak{p}_3$ and $\mathfrak{p}_3'$. That's because they come from factoring the "unrelated" numbers $3+\sqrt{-5}$ and $1+2\sqrt{-5}$. If you look at their complex conjugates you can check that $(3-\sqrt{-5}) = \mathfrak{p}_2\mathfrak{p}_7'$ and $(1-2\sqrt{-5}) = \mathfrak{p}_3\mathfrak{p}_7$, so

$$\boxed{\mathfrak{p}_7 = (7, 3+\sqrt{-5}) \text{ and } \mathfrak{p}_7' = (7, 3-\sqrt{-5})}$$

and

$$\boxed{\mathfrak{p}_7 = (7, 1-2\sqrt{-5}) \text{ and } \mathfrak{p}_7' = (7, 1+2\sqrt{-5}).}$$

The factorizations of $(3+\sqrt{-5})$ and $(3-\sqrt{-5})$ give another proof of how (7) factors, since

$$(3+\sqrt{-5})(3-\sqrt{-5}) = \mathfrak{p}_2^2\mathfrak{p}_7\mathfrak{p}_7'$$

implies

$$(14) = (2)\mathfrak{p}_7\mathfrak{p}_7' = 2\mathfrak{p}_7\mathfrak{p}_7' \Longrightarrow \mathfrak{p}_7\mathfrak{p}_7' = (7).$$

Looking back at the equation $(3+\sqrt{-5}) = \mathfrak{p}_2\mathfrak{p}_7$, we already know $\mathfrak{p}_2^2$ is principal, so we can square to see $\mathfrak{p}_7^2$ is principal:

$$(3+\sqrt{-5})^2 = \mathfrak{p}_2^2\mathfrak{p}_7^2 \Longrightarrow (4+6\sqrt{-5}) = (2)\mathfrak{p}_7^2 \Longrightarrow \boxed{\mathfrak{p}_7^2 = (2+3\sqrt{-5}).}$$

Similarly, $\boxed{\mathfrak{p}_7'^2 = (2-3\sqrt{-5})}$.

Which prime ideal $\mathfrak{p}$ divides (11)? We must have $\mathrm{N}(\mathfrak{p}) = 11$ or 121. Following our success with primes of norm 3 and 7, we could try looking for a principal ideal (α) where α has norm divisible by 11 exactly once, letting us peel off a prime of norm 11 as a factor. But that's a waste of time because it's impossible: no element of $\mathbf{Z}[\sqrt{-5}]$ has norm divisible by 11 just once. If $x^2+5y^2 \equiv 0 \bmod 11$ then $x^2 \equiv -5y^2 \bmod 11$. This forces $y \equiv 0 \bmod 11$ since $-5 \not\equiv \Box \bmod 11$. (If $y \not\equiv 0 \bmod 11$ then $-5 \equiv (xy^{-1})^2 \bmod 11$, but by explicit computation $-5 \not\equiv \Box \bmod 11$.) Once y vanishes mod 11, also $x \equiv 0 \bmod 11$, so x^2+5y^2 is divisible by 11^2. Norms of elements which are multiples of 11 are automatically multiples of 121.

Let $\mathfrak{p}$ be a prime factor of (11). For nonzero $\alpha \in \mathfrak{p}$, $\mathfrak{p} \mid (\alpha)$ so $\mathrm{N}(\mathfrak{p}) \mid \mathrm{N}((\alpha))$. Since $\mathrm{N}(\mathfrak{p})$ is 11 or 11^2, $11 \mid (x^2+5y^2)$, where $\alpha = x + y\sqrt{-5}$. We saw just above that if $x^2+5y^2 \equiv 0 \bmod 11$ then $x \equiv 0 \bmod 11$ and $y \equiv 0 \bmod 11$. So $11 \mid x$ and $11 \mid y$ in $\mathbf{Z}$, which implies $\alpha \in (11)$. We get $\mathfrak{p} \subset (11)$, so $\mathfrak{p} = (11)$ since $\mathfrak{p}$ is maximal. The ideal (11) in $\mathbf{Z}[\sqrt{-5}]$ is prime!

Theorem 4.34. *If p is a prime number and $-5 \not\equiv \Box \bmod p$ then (p) is prime in $\mathbf{Z}[\sqrt{-5}]$.*

Proof. Replace 11 by p in the argument above. ■

Theorem 4.35. *If p is a prime and $-5 \equiv \Box \bmod p$, then $(p) = \mathfrak{p}\mathfrak{p}'$ where $\mathfrak{p}$ and $\mathfrak{p}'$ are prime ideals in $\mathbf{Z}[\sqrt{-5}]$ with norm p.*

The ideals $\mathfrak{p}$ and $\mathfrak{p}'$ here might be equal. (This is different from our usual convention up to now that a decorated prime ideal like $\mathfrak{p}'$ means a prime not equal to $\mathfrak{p}$.) For $p=2$ we have $\mathfrak{p} = \mathfrak{p}' = \mathfrak{p}_2$, for $p=5$ we have $\mathfrak{p} = \mathfrak{p}' = (\sqrt{-5})$, and for $p=3$ and $p=7$ we have $\mathfrak{p} \neq \mathfrak{p}'$.

Proof. Write $-5 \equiv c^2 \bmod p$, so $p \mid (c^2+5)$ in $\mathbf{Z}$. Thus $p \mid (c+\sqrt{-5})(c-\sqrt{-5})$ in $\mathbf{Z}[\sqrt{-5}]$.

Recall in any domain that $\alpha \mid \beta$ if and only if $(\alpha) \mid (\beta)$ as ideals. Passing to principal ideals in our divisibility relation, $(p) \mid (c+\sqrt{-5})(c-\sqrt{-5})$. If (p) were a prime ideal, then $(p) \mid (c+\sqrt{-5})$ or $(p) \mid (c-\sqrt{-5})$ as ideals. This would imply $p \mid (c+\sqrt{-5})$ or $p \mid (c-\sqrt{-5})$ as elements of $\mathbf{Z}[\sqrt{-5}]$. But this is impossible because it implies $p \mid 1$ or $p \mid -1$ in $\mathbf{Z}$. So (p) is *not* a prime ideal: $(p) = \mathfrak{a}\mathfrak{b}$ for some proper (nonzero) ideals $\mathfrak{a}$ and $\mathfrak{b}$. Taking norms, $p^2 = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{b})$, where both norms on the right side are greater than 1, so $\mathrm{N}(\mathfrak{a}) = \mathrm{N}(\mathfrak{b}) = p$. Thus $\mathfrak{a}$ and $\mathfrak{b}$ are prime. Rewriting $\mathfrak{a} = \mathfrak{p}$ and $\mathfrak{b} = \mathfrak{p}'$, we are done. ■

4.5 Dedekind–Kummer Theorem

We saw in Section 4.4 that the ideal (p) in $\mathbf{Z}[\sqrt{-5}]$ factors if and only if $-5 \equiv \square \bmod p$, which is equivalent to saying $T^2 + 5 \bmod p$ factors.

The next theorem is a big generalization of this pattern. It says that if $\mathcal{O}_K = \mathbf{Z}[\alpha]$, where α has minimal polynomial $f(T)$ in $\mathbf{Z}[T]$, then for every prime p the way the ideal (p) in $\mathcal{O}_K$ factors can be read off from the way the mod p reduction $\overline{f(T)}$ in $\mathbf{F}_p[T]$ factors. (We now write $\mathbf{F}_p$ often instead of $\mathbf{Z}/p\mathbf{Z}$.)

Theorem 4.36 (Kummer). *Suppose $\mathcal{O}_K = \mathbf{Z}[\alpha]$ and α has minimal polynomial $f(T)$ in $\mathbf{Z}[T]$. For any prime p, the way the ideal $(p) = p\mathcal{O}_K$ and polynomial $\overline{f} = f \bmod p$ factor match: $(p) = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$ and $\overline{f} = \overline{\pi}_1^{e_1} \cdots \overline{\pi}_r^{e_r}$, where the $\mathfrak{p}_i$'s are distinct prime ideals and the $\overline{\pi}_i$'s are distinct monic irreducibles in $\mathbf{F}_p[T]$, and $\mathrm{N}(\mathfrak{p}_i) = p^{\deg \overline{\pi}_i}$. (That is, $\#\mathcal{O}_K/\mathfrak{p}_i = \#\mathbf{F}_p[T]/(\overline{\pi}_i)$.) Moreover, $\mathfrak{p}_i = (p, \pi_i(\alpha))$ for any monic lift of $\overline{\pi}_i$ to π_i in $\mathbf{Z}[T]$.*

We describe the similar factorization of (p) and $\overline{f}$ by saying the "shape" of the factorizations of (p) and $\overline{f}$ are the same. By the shape of a factorization we mean the parameters $r, e_1, \ldots, e_r$, and the norms of the $\mathfrak{p}_i$'s or degrees of the $\overline{\pi}_i$'s (same as the degrees of the π_i's since they are monic).[3] The shape does not include explicit formulas for the prime ideals $\mathfrak{p}_i$, although the end of Kummer's theorem indicates how that information is transferred from the factorization of $\overline{f}$ to the factorization of (p).

Example 4.37. Let $K = \mathbf{Q}(i)$, so $\mathcal{O}_K = \mathbf{Z}[i]$. Using $\alpha = i$ and $f(T) = T^2 + 1$, Kummer's theorem says the way p factors in $\mathbf{Z}[i]$ matches how $T^2 + 1 \bmod p$ factors. See Table 4.1, which reflects what we know about when $-1 \equiv \square \bmod p$, which is essentially the same as knowing when p is a sum of two squares. In the second row of the table, the linear factors $T + r$ and $T - r$ are different and the prime ideals $(x + yi)$ and $(x - yi)$ are different.

p	$T^2 + 1 \bmod p$	$(p) = p\mathbf{Z}[i]$
2	$(T+1)^2$	$(1+i)^2$
1 mod 4	$(T+r)(T-r)$	$(x+yi)(x-yi)$
3 mod 4	irreducible	(p)

Table 4.1: Factoring primes in $\mathbf{Z}[i]$.

[3] This could also be called the decomposition type or factorization pattern of (p).

Example 4.38. Let $K = \mathbf{Q}(\sqrt[3]{2})$, so $\mathcal{O}_K = \mathbf{Z}[\sqrt[3]{2}]$. Use $\alpha = \sqrt[3]{2}$ and $f(T) = T^3 - 2$. Table 4.2 gives the factorization of (p) from that of $T^3 - 2 \bmod p$ for small p. The norms of prime ideals appear as subscripts, and the exponent of the prime power in the subscript comes from the degree of the corresponding irreducible polynomial mod p. For instance, $\mathfrak{p}_{25}$ corresponds to the quadratic irreducible factor of $T^3 - 2 \bmod 5$.

p	$T^3 - 2 \bmod p$	$(p) = p\mathbf{Z}[\sqrt[3]{2}]$
2	T^3	$(\sqrt[3]{2})^3$
3	$(T-2)^3$	$\mathfrak{p}_3^3$
5	$(T-3)(T^2+3T+4)$	$\mathfrak{p}_5\mathfrak{p}_{25}$
7	irreducible	(7)
11	$(T-7)(T^2+7T+5)$	$\mathfrak{p}_{11}\mathfrak{p}_{121}$

Table 4.2: Factoring primes by shape in $\mathbf{Z}[\sqrt[3]{2}]$.

Kummer's theorem will come from the isomorphism $\mathbf{Z}[T]/(f(T)) \cong \mathbf{Z}[\alpha] = \mathcal{O}_K$ given by evaluation at α (which sends $f(T)$ to 0). Reducing the isomorphism mod p, we get

$$\mathbf{F}_p[T]/(\overline{f}(T)) \cong \mathbf{Z}[T]/(p, f(T)) \cong \mathbf{Z}[\alpha]/p\mathbf{Z}[\alpha] = \mathcal{O}_K/p\mathcal{O}_K,$$

where the second isomorphism is evaluation at α. We will see how the factorization of $p\mathcal{O}_K$ in $\mathcal{O}_K$ can be read off from the ring structure of $\mathcal{O}_K/p\mathcal{O}_K$ and the factorization of $\overline{f}(T)$ in $\mathbf{F}_p[T]$ can be read off from the ring structure of $\mathbf{F}_p[T]/(\overline{f})$, so the isomorphism of these rings will tell us the factorizations of $p\mathcal{O}_K$ and $\overline{f}(T)$ have the same shape.

Example 4.39. The ideal $p\mathcal{O}_K$ is prime in $\mathcal{O}_K$ if and only if $\overline{f}(T)$ is irreducible in $\mathbf{F}_p[T]$ because these are equivalent to saying the isomorphic rings $\mathcal{O}_K/p\mathcal{O}_K$ and $\mathbf{F}_p[T]/(\overline{f}(T))$ are fields. So $p\mathcal{O}_K$ is not prime if and only if $\overline{f}(T)$ is reducible in $\mathbf{F}_p[T]$.

Before we see how the way (p) factors is encoded in the ring $\mathcal{O}_K/p\mathcal{O}_K$, let's look at the simpler setting of $\mathbf{Z}$: how an integer $m > 1$ factors can be read off from the structure of the ring $\mathbf{Z}/m\mathbf{Z}$ for $m > 1$. Say $m = p_1^{e_1} \cdots p_r^{e_r}$ for distinct primes p_i. Then

$$\mathbf{Z}/m\mathbf{Z} \cong \mathbf{Z}/p_1^{e_1}\mathbf{Z} \times \cdots \times \mathbf{Z}/p_r^{e_r}\mathbf{Z}$$

by the Chinese remainder theorem. What are the prime ideals of $\mathbf{Z}/m\mathbf{Z}$? The ideals of $\mathbf{Z}/m\mathbf{Z}$ are $d\mathbf{Z}/m\mathbf{Z}$ for $d \mid m$ and $(\mathbf{Z}/m\mathbf{Z})/(d\mathbf{Z}/m\mathbf{Z}) \cong \mathbf{Z}/d\mathbf{Z}$. So the prime ideals are

$$\{p\mathbf{Z}/m\mathbf{Z} : p \text{ a prime}, \ p \mid m\} = \{p_1\mathbf{Z}/m\mathbf{Z}, \dots, p_r\mathbf{Z}/m\mathbf{Z}\}.$$

Therefore the number r of prime factors of m is the number of prime ideals in $\mathbf{Z}/m\mathbf{Z}$.

In $\mathbf{Z}/m\mathbf{Z}$, if $p \mid m$ the powers of the ideal $p\mathbf{Z}/m\mathbf{Z}$ are

$$\begin{aligned}(p\mathbf{Z}/m\mathbf{Z})^k &= \begin{cases} p^k\mathbf{Z}/m\mathbf{Z}, & \text{if } p^k \mid m, \\ (p^k\mathbf{Z} + m\mathbf{Z})/m\mathbf{Z}, & \text{for all } k \end{cases} \\ &= (p^k, m)\mathbf{Z}/m\mathbf{Z}.\end{aligned}$$

These eventually stabilize: if p^e is the largest power of p dividing m, then we get

$$p\mathbf{Z}/m\mathbf{Z} \supsetneqq p^2\mathbf{Z}/m\mathbf{Z} \supsetneqq \cdots \supsetneqq p^e\mathbf{Z}/m\mathbf{Z} = (p\mathbf{Z}/m\mathbf{Z})^e = (p\mathbf{Z}/m\mathbf{Z})^{e+1} = \cdots.$$

So the multiplicity e of a prime p in m is the number of different powers of the ideal $p\mathbf{Z}/m\mathbf{Z}$ in the ring $\mathbf{Z}/m\mathbf{Z}$.

Remark 4.40. Writing $m = p^e m'$ such that $p \nmid m'$, $\mathbf{Z}/m\mathbf{Z} \cong \mathbf{Z}/p^e\mathbf{Z} \times \mathbf{Z}/m'\mathbf{Z}$ by the Chinese remainder theorem, so we get the correspondence of ideals

$$p\mathbf{Z}/m\mathbf{Z} \longleftrightarrow p\mathbf{Z}/p^e\mathbf{Z} \times \mathbf{Z}/m'\mathbf{Z}.$$

Taking powers,

$$p^k\mathbf{Z}/m\mathbf{Z} \longleftrightarrow p^k\mathbf{Z}/p^e\mathbf{Z} \times \mathbf{Z}/m'\mathbf{Z}$$

for $k \leqslant e$ and $p^k\mathbf{Z}/m\mathbf{Z} \leftrightarrow \{0\} \times \mathbf{Z}/m'\mathbf{Z}$ for $k \geqslant e$.

Now we begin the proof of Theorem 4.36. If it seems a bit confusing at first, you can probably skip the proof without affecting your understanding of later results, provided you carefully read the worked examples so you know how to apply the theorem.

Proof. The rings $\mathcal{O}_K/p\mathcal{O}_K = \mathbf{Z}[\alpha]/p\mathbf{Z}[\alpha]$ and $\mathbf{F}_p[T]/(\overline{f})$ are isomorphic through the intermediate ring $\mathbf{Z}[T]/(p, f(T))$:

$$\begin{tikzcd}
& \mathbf{Z}[T]/(p, f(T)) \arrow[dl, "T\mapsto\alpha"', "\cong"] \arrow[dr, "h(T)\mapsto\overline{h}(T)", "\cong"'] & \\
\mathcal{O}_K/p\mathcal{O}_K & & \mathbf{F}_p[T]/(\overline{f})
\end{tikzcd}$$

We will check separately how the factorization of $p\mathcal{O}_K$ and $\overline{f}$ can be seen through the structure of $\mathcal{O}_K/p\mathcal{O}_K$ and $\mathbf{F}_p[T]/(\overline{f})$, and then compare the two rings through the isomorphism between them in the diagram above. Although we write $p\mathcal{O}_K = \mathfrak{p}_1^{e_1}\cdots\mathfrak{p}_r^{e_r}$ and $\overline{f} = \overline{\pi}_1^{e_1}\cdots\overline{\pi}_r^{e_r}$, part of the proof includes showing the same parameters (*e.g.*, number r of prime factors) really agree in both.

Reading off the factorization of m from $\mathbf{Z}/m\mathbf{Z}$ applies to the factorization of $\overline{f}$ from $\mathbf{F}_p[T]/(\overline{f})$. The prime ideals of $\mathbf{F}_p[T]/(\overline{f})$ are $\left\{(\overline{\pi}_i)/(\overline{f}) : i = 1, \ldots, r\right\}$. So r is the number of prime ideals in $\mathbf{F}_p[T]/(\overline{f})$ and each prime ideal $(\overline{\pi}_i)/(\overline{f})$ of $\mathbf{F}_p[T]/(\overline{f})$ has e_i different powers:

$$(\overline{\pi}_i)/(f) \underset{\neq}{\supseteq} \left((\overline{\pi}_i)/(f)\right)^2 \underset{\neq}{\supseteq} \cdots \underset{\neq}{\supseteq} \left((\overline{\pi}_i)/(f)\right)^{e_i} = \left((\overline{\pi}_i)/(f)\right)^{e_i+1}.$$

Now we look at $\mathcal{O}_K/p\mathcal{O}_K$. From $p\mathcal{O}_K = \mathfrak{p}_1^{e_1}\cdots\mathfrak{p}_r^{e_r}$, the ideals of $\mathcal{O}_K/p\mathcal{O}_K$ are

$$\{\mathfrak{a}/p\mathcal{O}_K : \mathfrak{a} \supset p\mathcal{O}_K\} = \{\mathfrak{a}/p\mathcal{O}_K : \mathfrak{a} \mid p\mathcal{O}_K\}$$

and since $(\mathcal{O}_K/p\mathcal{O}_K)/(\mathfrak{a}/p\mathcal{O}_K) \cong \mathcal{O}_K/\mathfrak{a}$, the prime ideals among these are

$$\{\mathfrak{p}/p\mathcal{O}_K : \mathfrak{p} \mid p\mathcal{O}_K,\ \mathfrak{p} \text{ prime}\} = \{\mathfrak{p}_1/p\mathcal{O}_K, \ldots, \mathfrak{p}_r/p\mathcal{O}_K\}.$$

Thus $\mathcal{O}_K/p\mathcal{O}_K$ has r prime ideals, where r is the number of prime ideal factors of $p\mathcal{O}_K$.

If $\mathfrak{p}$ is prime and $\mathfrak{p}^e$ is the largest power of $\mathfrak{p}$ dividing $p\mathcal{O}_K$, we can read off e from $\mathcal{O}_K/p\mathcal{O}_K$ as the number of powers of $\mathfrak{p}/p\mathcal{O}_K$ in $\mathcal{O}_K/p\mathcal{O}_K$. Why? For any $k \geqslant 1$,

$$(\mathfrak{p}/p\mathcal{O}_K)^k = (\mathfrak{p}^k + p\mathcal{O}_K)/p\mathcal{O}_K = \gcd(\mathfrak{p}^k, p\mathcal{O}_K)/p\mathcal{O}_K.$$

For $k \leqslant e$, this power is $\mathfrak{p}^k/p\mathcal{O}_K$ and they are distinct for different k. When $k \geqslant e$, $\gcd(\mathfrak{p}^k, p\mathcal{O}_K) = \mathfrak{p}^e$, so we have

$$\mathfrak{p}/p\mathcal{O}_K \supsetneqq (\mathfrak{p}/p\mathcal{O}_K)^2 \supsetneqq \cdots \supsetneqq (\mathfrak{p}/p\mathcal{O}_K)^e = (\mathfrak{p}/p\mathcal{O}_K)^{e+1}.$$

(It was not essential here that $p\mathcal{O}_K$ is a principal ideal. We can replace $p\mathcal{O}_K$ with any nonzero ideal $\mathfrak{a}$, and see in the same way that the factorization of $\mathfrak{a}$ can be read off from the strucutre of the ring $\mathcal{O}_K/\mathfrak{a}$.)

It remains to show the primes $\mathfrak{p}_i$ appearing in $p\mathcal{O}_K$ and the monic irreducibles π_i appearing in $\overline{f}$ can be paired off so that $\mathrm{N}(\mathfrak{p}_i) = p^{\deg \overline{\pi}_i}$. We will use the isomorphisms at the start of the proof to see what a prime ideal in $\mathbf{F}_p[T]/(\overline{f})$ corresponds to in $\mathcal{O}_K/p\mathcal{O}_K$.

Under the reduction isomorphism $\mathbf{Z}[T]/(p, f(T)) \to \mathbf{F}_p[T]/(\overline{f})$, a prime ideal $(\overline{\pi}_i)/(\overline{f})$ in $\mathbf{F}_p[T]/(\overline{f})$ has inverse image $(p, \pi_i(T))/(p, f(T))$ for any monic lift $\pi_i(T)$ of $\overline{\pi}_i(T)$ to $\mathbf{Z}[T]$. Under the "evaluation" isomorphism $\mathbf{Z}[T]/(p, f(T)) \cong \mathbf{Z}[\alpha]/p\mathbf{Z}[\alpha]$, the ideal $(p, \pi_i(T))/(p, f(T))$ has image $(p, \pi_i(\alpha))/p\mathcal{O}_K$ in $\mathcal{O}_K/p\mathcal{O}_K$. The prime ideals of $\mathcal{O}_K/p\mathcal{O}_K$ are $\{\mathfrak{p}_1/p\mathcal{O}_K, \ldots, \mathfrak{p}_r/p\mathcal{O}_K\}$, so there is a labeling such that $\boxed{\mathfrak{p}_i = (p, \pi_i(\alpha)).}$ So

$$\mathrm{N}(\mathfrak{p}_i) = \#\mathcal{O}_K/\mathfrak{p}_i = \#\mathbf{Z}[\alpha]/(p, \pi_i(\alpha)) = \#\mathbf{F}_p[T]/(\overline{\pi}_i(T)) = p^{\deg \overline{\pi}_i}. \qquad \blacksquare$$

Example 4.41. We return to $\mathbf{Z}[\sqrt[3]{2}]$. Table 4.3 provides explicit factorizations that refine the information in Table 4.2. Just because an ideal in Table 4.3 is given by two generators doesn't mean it can't be a principal ideal! We don't get any information about that by the theorem.

p	$T^3 - 2 \bmod p$	$p\mathbf{Z}[\sqrt[3]{2}]$
2	T^3	$(\sqrt[3]{2})^3$
3	$(T-2)^3$	$(3, \sqrt[3]{2} - 2)^3$
5	$(T-3)(T^2+3T+4)$	$(5, \sqrt[3]{2} - 3)(5, \sqrt[3]{4} + 3\sqrt[3]{2} + 4)$
7	irreducible	(7)
11	$(T-7)(T^2+7T+5)$	$(11, \sqrt[3]{2} - 7)(11, \sqrt[3]{4} + 7\sqrt[3]{2} + 5)$

Table 4.3: Factoring primes explicitly in $\mathbf{Z}[\sqrt[3]{2}]$.

Example 4.42. Let $K = \mathbf{Q}(\alpha)$, where α is a root of $T^4 + 2T^2 + 3T + 1$. From Example 3.49, $\mathcal{O}_K = \mathbf{Z}[\alpha]$, so every prime p factors in $\mathcal{O}_K$ the way $T^4 + 2T^2 + 3T + 1$ factors in $\mathbf{F}_p[T]$. Small primes are factored in Table 4.4.

p	$T^4+2T^2+3T+1 \bmod p$	(p)
2	irreducible	(2)
3	$(T^2+1)^2$	$(3,\alpha^2+1)^2$
5	irreducible	(5)
7	$(T-1)(T-4)(T^2+5T+2)$	$(7,\alpha-1)(7,\alpha-4)(7,\alpha^2+5\alpha+2)$
11	irreducible	(11)

Table 4.4: Factoring primes explicitly in $\mathbf{Z}[\alpha]$, $\alpha^4+2\alpha^2+3\alpha+1=0$.

Example 4.43. In Example 3.11 we showed the ideal $\mathfrak{a}=(2+5\sqrt{10},4+7\sqrt{10})$ in $\mathbf{Z}[\sqrt{10}]$ has norm 6 using a determinant formula for the norm. In Example 4.29, we found an abstract prime factorization of $\mathfrak{a}$ from knowing the norm. Now we will use Kummer's theorem to work this out again from scratch and also find generators of the prime ideal factors.

Since $(2+5\sqrt{10})$ and $(4+7\sqrt{10})$ are in $\mathfrak{a}$, $\mathfrak{a}\mid(2+5\sqrt{10})$ and $\mathfrak{a}\mid(4+7\sqrt{10})$. Taking ideal norms, $\mathrm{N}(\mathfrak{a})\mid 246$ and $\mathrm{N}(\mathfrak{a})\mid 474$. Therefore $\mathrm{N}(\mathfrak{a})$ is a factor of $(246,474)=6$. So a prime ideal factor of $\mathfrak{a}$ must have norm 2 or 3.

What are the primes dividing (2) and (3) in $\mathbf{Z}[\sqrt{10}]$? Mod 2, $T^2-10=T^2$, so $(2)=\mathfrak{p}_2^2$ where $\mathfrak{p}_2=(2,\sqrt{10})$. Mod 3, $T^2-10=(T+1)(T-1)$, so $(3)=\mathfrak{p}_3\mathfrak{p}_3'$ where $\mathfrak{p}_3=(3,\sqrt{10}+1)$ and $\mathfrak{p}_3'=(3,\sqrt{10}-1)$.

At this point you could try multiplying $\mathfrak{p}_2$ with $\mathfrak{p}_3$ or $\mathfrak{p}_3'$ to see if you get $\mathfrak{a}$ (or maybe $\mathfrak{a}$ is just one of the prime ideals alone or it is (1)). But this is going to be a mess:

$$\mathfrak{p}_2\mathfrak{p}_3=(2,\sqrt{10})(3,\sqrt{10}+1)=(6,2\sqrt{10}+2,3\sqrt{10},10+\sqrt{10}).$$

To compare this with $\mathfrak{a}$, we need to see if the two generators of $\mathfrak{a}$ are $\mathbf{Z}[\sqrt{10}]$-linear combinations of the four generators of $\mathfrak{p}_2\mathfrak{p}_3$. It is a lot easier to abandon this method and separately check which prime ideals are factors of $\mathfrak{a}$ by checking which prime ideals *contain* $\mathfrak{a}$.

Is $\mathfrak{a}\subset\mathfrak{p}_2$? The generators $2+5\sqrt{10}$ and $4+7\sqrt{10}$ of $\mathfrak{a}$ are obviously linear combinations of 2 and $\sqrt{10}$, so $\mathfrak{a}\subset\mathfrak{p}_2$. So $\mathfrak{p}_2\mid\mathfrak{a}$.

Is $\mathfrak{a}\subset\mathfrak{p}_3$? We need to check if $2+5\sqrt{10}$ and $4+7\sqrt{10}$ are linear combinations of 3 and $\sqrt{10}+1$. They are:

$$2+5\sqrt{10}=5(\sqrt{10}+1)-3,\quad 4+7\sqrt{10}=7(\sqrt{10}+1)-3.$$

So $\mathfrak{p}_3 \mid \mathfrak{a}$. (Strictly speaking, we should be allowing for $\mathbf{Z}[\sqrt{10}]$-linear combinations of 3 and $\sqrt{10}+1$, because we are using generators of an ideal, but we try $\mathbf{Z}$-coefficients first and it works out. Moreover, the ideal $(3, \sqrt{10}+1)$ is spanned by 3 and $\sqrt{10}+1$ as an abelian group, so $\mathbf{Z}$-coefficients truly are enough to use. If we had run into a problem then we should ask if $\mathfrak{a} \subset \mathfrak{p}_3'$ instead.)

Since $\mathfrak{p}_2$ and $\mathfrak{p}_3$ each divide $\mathfrak{a}$ and they are different prime ideals, $\mathfrak{a} = \mathfrak{p}_2\mathfrak{p}_3\mathfrak{a}'$ for some ideal $\mathfrak{a}'$. The norm of $\mathfrak{a}$ divides 6 and $\mathfrak{p}_2\mathfrak{p}_3$ has norm 6, so $\mathrm{N}(\mathfrak{a}') = 1$. Therefore $\mathfrak{a}' = (1)$ (the only ideal of norm 1 is (1), by the definition of the ideal norm as an index) and $\mathfrak{a} = \mathfrak{p}_2\mathfrak{p}_3$. This concludes the example.

If $\mathrm{N}(\mathfrak{a}) = p_1 \cdots p_r$ where the p_i's are distinct primes, then $\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_r$ where $\mathrm{N}(\mathfrak{p}_i) = p_i$. But if $\mathrm{N}(\mathfrak{a}) = 12$, say, it does not follow that the prime ideal factorization of $\mathfrak{a}$ is $\mathfrak{p}^2\mathfrak{q}$ where $\mathrm{N}(\mathfrak{p}) = 2$ and $\mathrm{N}(\mathfrak{q}) = 3$. That is one possibility, but there are two others: $\mathfrak{a} = \mathfrak{p}\mathfrak{p}'\mathfrak{q}$ where $\mathfrak{p}$ and $\mathfrak{p}'$ are distinct primes of norm 2 and $\mathfrak{q}$ has norm 3, and $\mathfrak{a} = \mathfrak{p}\mathfrak{q}$ where $\mathrm{N}(\mathfrak{p}) = 4$ and $\mathrm{N}(\mathfrak{q}) = 3$. Don't forget that (nonzero) prime ideals in $\mathcal{O}_K$ need not have prime norm, but only prime power norm.

What we called Kummer's theorem was only worked out by Kummer for restricted types of number fields, essentially the cyclotomic fields and Kummer extensions of them. Dedekind found a generalization of Kummer's theorem which tells us how to factor all but a finite number of prime numbers in any $\mathcal{O}_K$, even if we don't know an α for which $\mathcal{O}_K = \mathbf{Z}[\alpha]$.

Theorem 4.44 (Dedekind). *Let $K = \mathbf{Q}(\alpha)$ where $\alpha \in \mathcal{O}_K$ and let $f(T)$ be the minimal polynomial of α in $\mathbf{Z}[T]$. For any prime p such that $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$, the factorizations of (p) in $\mathcal{O}_K$ and $\overline{f}$ in $\mathbf{F}_p[T]$ have the same shape and we can find generators for the prime factors of (p) from the irreducible factors of $\overline{f}$.*

When $\mathcal{O}_K = \mathbf{Z}[\alpha]$, $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ is 1 and we recover Kummer's theorem. If you skipped the proof of Kummer's theorem, you probably should skip the proof of Dedekind's theorem, but be sure to look at the examples of how it is used.

Proof. The key point is that we will show

$$p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]] \Longrightarrow \mathcal{O}_K/p\mathcal{O}_K \cong \mathbf{F}_p[T]/(\overline{f}(T)) \tag{4.2}$$

in a natural way. If $\mathcal{O}_K = \mathbf{Z}[\alpha]$ then this isomorphism holds for all p and was the key to proving Kummer's factorization theorem. The point is that even if $\mathcal{O}_K \neq \mathbf{Z}[\alpha]$, such an isomorphism at least remains true when $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$.

The composite ring homomorphism $\mathbf{Z}[\alpha] \to \mathcal{O}_K \to \mathcal{O}_K/p\mathcal{O}_K$ kills p, so it induces a ring homomorphism

$$\mathbf{Z}[\alpha]/p\mathbf{Z}[\alpha] \to \mathcal{O}_K/p\mathcal{O}_K. \tag{4.3}$$

In both rings $p = 0$, so we can view both sides as $\mathbf{F}_p$-vector spaces. Let $n = \deg f = [K : \mathbf{Q}]$. Both $\mathbf{Z}[\alpha]$ and $\mathcal{O}_K$ have $\mathbf{Z}$-bases of size n, and the mod p reduction of a $\mathbf{Z}$-basis will be an $\mathbf{F}_p$-basis of $\mathbf{Z}[\alpha]/p\mathbf{Z}[\alpha]$ and $\mathcal{O}_K/p\mathcal{O}_K$. Since the rings in (4.3) have equal dimension n over $\mathbf{F}_p$, to check the $\mathbf{F}_p$-linear map in (4.3) is an isomorphism it suffices to check it is either injective or surjective. We will show it is surjective. This is where we will use $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$.

The additive group $\mathcal{O}_K/p\mathcal{O}_K$ has p-power size. Set $m = [\mathcal{O}_K : \mathbf{Z}[\alpha]]$. Since $p \nmid m$, scaling by m on $\mathcal{O}_K/p\mathcal{O}_K$ is onto. At the same time, $m\mathcal{O}_K \subset \mathbf{Z}[\alpha]$. Therefore $\mathcal{O}_K/p\mathcal{O}_K$ is represented by $\mathbf{Z}[\alpha]$. This shows (4.3) is onto. (Explicitly, writing $px + my = 1$ in $\mathbf{Z}$, for any $\beta \in \mathcal{O}_K$,

$$\begin{aligned} \beta &= px\beta + my\beta \\ &\equiv my\beta \bmod p\mathcal{O}_K \\ &\equiv \underbrace{y(m\beta)}_{\text{in } \mathbf{Z}[\alpha]} \bmod p\mathcal{O}_K.) \end{aligned}$$

Thus the natural ring homomorphism $\mathbf{Z}[\alpha]/p\mathbf{Z}[\alpha] \to \mathcal{O}_K/p\mathcal{O}_K$ is an isomorphism. We already know the shape of the factorization of $p\mathcal{O}_K$ is determined by the structure of the ring $\mathcal{O}_K/p\mathcal{O}_K$, and from $\mathcal{O}_K/p\mathcal{O}_K \cong \mathbf{Z}[\alpha]/p\mathbf{Z}[\alpha] \cong \mathbf{F}_p[T]/(\overline{f})$ we once again can read off the shape of (p) from the shape of $\overline{f}$. Moreover, following a prime ideal through the sequence of isomorphisms

$$\mathbf{F}_p[T]/(\overline{f}) \to \mathbf{Z}[T]/(p, f(T)) \to \mathbf{Z}[\alpha]/p\mathbf{Z}[\alpha] \to \mathcal{O}_K/p\mathcal{O}_K$$

shows the prime ideals in $\mathcal{O}_K$ dividing (p) all have the form as given in the proof of Theorem 4.36: in the notation of that proof, $\mathfrak{p}_i = (p, \pi_i(\alpha))$, where $\pi_i(T)$ is any monic lift of $\overline{\pi}_i(T)$ to $\mathbf{Z}[T]$. That still works here. ∎

Example 4.45. Let $K = \mathbf{Q}(\sqrt{5})$. An obvious subring of $\mathcal{O}_K$ is $\mathbf{Z}[\sqrt{5}]$. Pretend we do not know $\mathcal{O}_K = \mathbf{Z}[\frac{1+\sqrt{5}}{2}]$. From

$$\operatorname{disc}(T^2 - 5) = -20 = [\mathcal{O}_K : \mathbf{Z}[\sqrt{5}]]^2 \operatorname{disc}(K),$$

we see $[\mathcal{O}_K : \mathbf{Z}[\sqrt{5}]] = 1$ or 2. So for *every* prime $p \neq 2$, Theorem 4.44 says the way $(p) = p\mathcal{O}_K$ factors can be read off from how $T^2 - 5 \bmod p$ factors. In Table 4.5 the first few $p \neq 2$ are factored in $\mathcal{O}_K$ and we can give generators for the prime ideal factors of (p) by substituting $\sqrt{5}$ for T in the irreducible factors of $T^2 - 5 \bmod p$. Note the generators are for ideals in the "unknown" $\mathcal{O}_K$, not in $\mathbf{Z}[\sqrt{5}]$.

p	$T^2 - 5 \bmod p$	$p\mathcal{O}_K$	Prime Ideal Factors of $p\mathcal{O}_K$
2	$(T-1)^2$	?	?
3	irreducible	(3)	(3)
5	T^2	$\mathfrak{p}_5^2$	$(5, \sqrt{5}) = (\sqrt{5})$
7	irreducible	(7)	(7)
11	$(T+4)(T-4)$	$\mathfrak{p}_{11}\mathfrak{p}_{11}'$	$(11, \sqrt{5}+4), (11, \sqrt{5}-4)$

Table 4.5: Factoring prime numbers in $K = \mathbf{Q}(\sqrt{5})$, except 2.

Although $T^2 - 5 \equiv (T-1)^2 \bmod 2$, this does *not* justify saying $(2) = \mathfrak{p}_2^2$ in $\mathcal{O}_K$, and in fact that isn't how (2) factors. Because we know $\mathcal{O}_K = \mathbf{Z}[\frac{1+\sqrt{5}}{2}]$ and $\frac{1+\sqrt{5}}{2}$ is a root of $T^2 - T - 1$, we can factor all primes, including 2, by using $T^2 - T - 1$ instead of $T^2 - 5$. See Table 4.6. In particular, $T^2 - T - 1$ is irreducible mod 2, so (2) is a prime ideal rather than the square of a prime ideal. In Table 4.6 we recover the same factorization for primes $p \neq 2$ that we have in Table 4.5, although we don't get the same generators for prime ideals. It is not a surprise that $T^2 - 5$ and $T^2 - T - 1$ factor mod p in the same way for $p \neq 2$ because $T^2 - T - 1 = (T - \frac{1}{2})^2 - \frac{5}{4} = \frac{1}{4}((2T-1)^2 - 5)$ is the same as $T^2 - 5$ up to a change of variables that is defined in any field where $2 \neq 0$.

p	$T^2 - T - 1 \bmod p$	$p\mathcal{O}_K$	Prime Ideal Factors
2	irreducible	(2)	(2)
3	irreducible	(3)	(3)
5	$(T-3)^2$	$\mathfrak{p}_5^2$	$(5, \frac{1+\sqrt{5}}{2} - 3) = (\sqrt{5})$
7	irreducible	(7)	(7)
11	$(T-4)(T-8)$	$\mathfrak{p}_{11}\mathfrak{p}_{11}'$	$(11, \frac{1+\sqrt{5}}{2} - 4), (11, \frac{1+\sqrt{5}}{2} - 8)$

Table 4.6: Factoring prime numbers in $K = \mathbf{Q}(\sqrt{5})$.

In Tables 4.5 and 4.6, the prime ideals dividing (11) can be matched up by a computation: $(11, \sqrt{5}+4) = (11, \frac{1+\sqrt{5}}{2} - 4)$ and $(11, \sqrt{5}-4) = (11, \frac{1+\sqrt{5}}{2} - 8)$. Since these ideals are maximal, a containment in either direction implies equality and it is easy to show the left side is in the right side for both: $\sqrt{5}+4 = 11+2(\frac{1+\sqrt{5}}{2}-4)$ and $\sqrt{5}-4 = 11+2(\frac{1+\sqrt{5}}{2}-8)$.

Example 4.46. In Table 4.7 the decomposition rules for prime numbers in a quadratic field $\mathbf{Q}(\sqrt{d})$, d a squarefree integer, are summarized. Since $[\mathcal{O}_K : \mathbf{Z}[\sqrt{d}]]$ is 1 or 2, for $p \neq 2$ we can read off the factorization of (p) from that of $T^2 - d \bmod p$. The decomposition of 2 has to be treated more carefully, looking at $T^2 - d \bmod 2$ when $d \equiv 2, 3 \bmod 4$ and $T^2 - T - \frac{d-1}{4} \bmod 2$ when $d \equiv 1 \bmod 4$.

Condition on d	$T^2 - d \bmod p$	(p), $p \neq 2$	Condition on d	(2)
$d \not\equiv \square \bmod p$	irreducible	(p)	$d \equiv 5 \bmod 8$	(2)
$d \equiv \square \bmod p$	$(T-r)(T-r')$	$\mathfrak{p}\mathfrak{p}'$	$d \equiv 1 \bmod 8$	$\mathfrak{p}\mathfrak{p}'$
$p \mid d$	T^2	$\mathfrak{p}^2$	$d \equiv 2, 3 \bmod 4$	$\mathfrak{p}^2$

Table 4.7: Factoring primes in $\mathbf{Q}(\sqrt{d})$, d squarefree.

To apply Theorem 4.44 to a particular prime p, we need to be able to check that $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$. How can we do that without knowing $\mathcal{O}_K$? There are two answers to that question:

(1) If you're content to miss a finite set of primes to which the theorem might work, we can get enough information from the equation

$$\operatorname{disc}(\mathbf{Z}[\alpha]) = [\mathcal{O}_K : \mathbf{Z}[\alpha]]^2 \operatorname{disc}(K). \tag{4.4}$$

The left side is computable without knowing $\mathcal{O}_K$, because

$$\operatorname{disc}(\mathbf{Z}[\alpha]) = \operatorname{disc}(f(T))$$

by (3.13), so any prime not dividing $\operatorname{disc}(f(T))$ also does not divide $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$. By using $p \nmid \operatorname{disc}(f(T))$ instead of the sharper $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$, (4.4) shows any prime dividing $\operatorname{disc}(K)$ but not $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ will be overlooked in this way, even though Theorem 4.44 is valid for such primes; we wouldn't notice them.

(2) We can use a little group theory to find the p dividing $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ without computing $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$. First compute $\operatorname{disc}(\mathbf{Z}[\alpha])$. Its prime factors include all p dividing $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$, by (4.4). So the prime factors of $\operatorname{disc}(\mathbf{Z}[\alpha])$ are all we need to look at and decide among them which divide the index

and which do not. Since $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ is the size of the group $\mathcal{O}_K/\mathbf{Z}[\alpha]$, Cauchy's theorem tells us that a prime p divides $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ if and only if $\mathcal{O}_K/\mathbf{Z}[\alpha]$ has an element of additive order p. To say $x \in \mathcal{O}_K$ has order p in $\mathcal{O}_K/\mathbf{Z}[\alpha]$ means $px \in \mathbf{Z}[\alpha]$ and $x \notin \mathbf{Z}[\alpha]$, *i.e.*, x is nonzero in $\frac{1}{p}\mathbf{Z}[\alpha]/\mathbf{Z}[\alpha]$. Thus we can find out if there are any such x by running through a list of coset representatives for $\frac{1}{p}\mathbf{Z}[\alpha]/\mathbf{Z}[\alpha]$ and seeing if some nonzero coset representative is an algebraic integer. (To decide if a number in K is an algebraic integer, compute its characteristic polynomial for $K/\mathbf{Q}$ using any field basis at all and see if the polynomial is in $\mathbf{Z}[T]$.) If an algebraic integer is found then $p \mid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$. If not then $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$.

Example 4.47. Let $K = \mathbf{Q}(\alpha)$ where α is a root of $T^3 - 12T + 2$. Since $\operatorname{disc}(T^3 - 12T + 2) = 2^2 \cdot 3^5 \cdot 7$, only 2 and 3 could divide $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$. Therefore any prime $p \neq 2$ or 3 can be factored in $\mathcal{O}_K$ by factoring $T^3 - 12T + 2$ in $\mathbf{F}_p[T]$. From Example 3.48, $2 \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$, so we can also factor 2 using this polynomial. Since $T^3 - 12T + 2 \equiv T^3 \bmod 2$, $(2) = \mathfrak{p}_2^3$.

How do we determine the factorization of 3 in $\mathcal{O}_K$? We found in Example 3.48 that $\mathcal{O}_K = \mathbf{Z}[\beta]$, where $\beta = \frac{1}{3}(1 + \alpha + \alpha^2)$ is a root of $T^3 - 9T^2 + 21T - 7$. So the way 3 factors in $\mathcal{O}_K$ matches the way $T^3 - 9T^2 + 21T - 7$ factors in $\mathbf{F}_3[T]$. The factorization is $(T-1)^3$, so $3\mathcal{O}_K = \mathfrak{p}_3^3$. We are not justified in applying Theorem 4.44 to $T^3 - 12T + 2$ to factor $3\mathcal{O}_K$, although if we tried it anyway we would get the right answer: $T^3 - 12T + 2 \equiv (T-1)^3 \bmod 3$.

Example 4.48. Let $K = \mathbf{Q}(\sqrt[3]{10})$. Since $\operatorname{disc}(T^3 - 10) = -2700 = -2^2 \cdot 3^3 \cdot 5^2$, any prime p other than 2, 3, or 5 can be factored in $\mathcal{O}_K$ by factoring $T^3 - 10$ in $\mathbf{F}_p[T]$. We saw in Example 3.50 that 2 and 5 do not divide $[\mathcal{O}_K : \mathbf{Z}[\sqrt[3]{10}]]$, so we can also factor 2 and 5 in $\mathcal{O}_K$ by factoring $T^3 - 10$ modulo 2 and 5. This shows $(2) = \mathfrak{p}_2^3$ and $(5) = \mathfrak{p}_5^3$. What about (3)?

By Example 3.50, $\mathcal{O}_K = \mathbf{Z}[\alpha]$, where $\alpha = \frac{1}{3} + \frac{1}{3}\sqrt[3]{10} + \frac{1}{3}\sqrt[3]{100}$ is a root of $T^3 - T^2 - 3T - 3$. The way 3 factors in $\mathcal{O}_K$ matches the way $T^3 - T^2 - 3T - 3$ factors modulo 3: the polynomial modulo 3 is $T^2(T-1)$, so $3\mathcal{O}_K = \mathfrak{p}_3^2\mathfrak{p}_3'$. We are not justified in factoring 3 in $\mathcal{O}_K$ by factoring $T^3 - 10 \bmod 3$, and now we see it would have predicted the wrong factorization, since $T^3 - 10 \equiv (T-1)^3 \bmod 3$.

Dedekind initially thought, as the examples above suggest, that for any number field K and prime p there should be an $\alpha_p \in \mathcal{O}_K$ such that $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha_p]]$. Then Theorem 4.44 tells us how to find the factorization of $p\mathcal{O}_K$. This would reduce the task of factoring all primes in number fields to factoring polynomials over finite fields. Dedekind spent years trying to show there is an α_p

for each p, but eventually he discovered "Dedekind's field" $K = \mathbf{Q}(\gamma)$, where $\gamma^3 - \gamma^2 - 2\gamma - 8 = 0$. (This is a cubic field since $T^3 - T^2 - 2T - 8$ is irreducible mod 3.)

Theorem 4.49 (Dedekind, 1878). *Let $K = \mathbf{Q}(\gamma)$, where $\gamma^3 - \gamma^2 - 2\gamma - 8 = 0$. For every $\alpha \in \mathcal{O}_K - \mathbf{Z}$, $2 \mid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$. In particular, $\mathcal{O}_K \neq \mathbf{Z}[\alpha]$ for any $\alpha \in \mathcal{O}_K - \mathbf{Z}$.*

Proof. First we compute a $\mathbf{Z}$-basis for $\mathcal{O}_K$. Using a 3×3 trace pairing matrix or a polynomial discriminant formula for general cubics, $\operatorname{disc}(\mathbf{Z}[\gamma]) = -2012 = -4 \cdot 503$, so $[\mathcal{O}_K : \mathbf{Z}[\gamma]]$ is 1 or 2.

Coset representatives for $\frac{1}{2}\mathbf{Z}[\gamma]/\mathbf{Z}[\gamma]$ are $\frac{a}{2} + \frac{b}{2}\gamma + \frac{c}{2}\gamma^2$ where $a, b, c \in \{0, 1\}$. When we compute the characteristic polynomial for the 7 nonzero coset representatives to see which are algebraic integers, an algebraic integer occurs when $b = 1$ and $c = 1$: the matrix for multiplication by $\frac{1}{2}\gamma + \frac{1}{2}\gamma^2$ on $\mathbf{Q}(\gamma)/\mathbf{Q}$ with respect to the basis $\{1, \gamma, \gamma^2\}$ is

$$\begin{pmatrix} 0 & 4 & 8 \\ 1/2 & 1 & 6 \\ 1/2 & 1 & 2 \end{pmatrix},$$

whose characteristic polynomial is $T^3 - 3T^2 - 10T - 8$. Therefore $[\mathcal{O}_K : \mathbf{Z}[\gamma]] = 2$, $\operatorname{disc}(K) = -2012/4 = -503$, and

$$\mathbf{Z}[\gamma] \subsetneqq \mathbf{Z} + \mathbf{Z}\gamma + \mathbf{Z}\frac{\gamma + \gamma^2}{2} \subset \mathcal{O}_K,$$

so

$$\mathcal{O}_K = \mathbf{Z} + \mathbf{Z}\gamma + \mathbf{Z}\frac{\gamma + \gamma^2}{2}.$$

Set $\gamma' = \frac{1}{2}\gamma + \frac{1}{2}\gamma^2$. Check the multiplication rules for the $\mathbf{Z}$-basis $\{1, \gamma, \gamma'\}$ of $\mathcal{O}_K$ are

$$\gamma^2 = -\gamma + 2\gamma', \quad \gamma\gamma' = 4 + 2\gamma', \quad \gamma'^2 = 6 + 2\gamma + 3\gamma'. \tag{4.5}$$

For any $\alpha \in \mathcal{O}_K - \mathbf{Z}$, we compute $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ by finding the determinant of a transition matrix expressing $\{1, \alpha, \alpha^2\}$ in terms of $\{1, \gamma, \gamma'\}$ (Theorem 3.10). Write $\alpha = a + b\gamma + c\gamma'$ with integers a, b, and c. Since $\mathbf{Z}[\alpha] = \mathbf{Z}[\alpha - a]$, we may assume $a = 0$ for the purpose of computing $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$. Then, after expressing

$\alpha^2 = (b\gamma + c\gamma')^2$ in the basis $\{1, \gamma, \gamma'\}$ using the formulas (4.5),

$$\begin{pmatrix} 1 \\ \alpha \\ \alpha^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & b & c \\ 6b^2 + 8bc & -b^2 + 2c^2 & 2b^2 + 4bc + 3c^2 \end{pmatrix} \begin{pmatrix} 1 \\ \gamma \\ \gamma^2 \end{pmatrix}.$$

The matrix, when reduced mod 2 (and using $x^2 \equiv x \bmod 2$ for any x), is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & b & c \\ 0 & b & c \end{pmatrix} \bmod 2,$$

whose determinant is 0 mod 2. Therefore $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ is even. ■

This example was quite important in Dedekind's thinking, particularly in his development of ideal theory [43, pp. 85, 115]. Only after finding it did Dedekind abandon trying to study prime factorization in number fields in terms of rings of the form $\mathbf{Z}[\alpha]$, and develop other methods.

When $\mathcal{O}_K = \mathbf{Z}[\alpha]$, we say $\mathcal{O}_K$ is *monogenic*, meaning it is generated as a ring over $\mathbf{Z}$ with one generator. The arithmetic properties of a monogenic $\mathcal{O}_K$ essentially reduce to the behavior of a polynomial (the minimal polynomial of α). It is important to remember $\mathcal{O}_K$ may not be monogenic, so the arithmetic of a number field generally can't be reduced to the study of a polynomial.

The trick of finding a common prime factor in every index $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ in order to prove a ring $\mathcal{O}_K$ doesn't have the form $\mathbf{Z}[\alpha]$ is not always applicable: sometimes $\mathcal{O}_K$ is not $\mathbf{Z}[\alpha]$ but the indices $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ have no common prime factor. See Example 6.47. A different method of showing a ring $\mathcal{O}_K$ has no power basis is described in Exercise 4.9.

4.6 Application: Solving $x^2 - dy^2 = n$, II

In Section 1.4 we saw how to find all integral solutions to $x^2 - dy^2 = n$ effectively once we know a nontrivial unit in $\mathbf{Z}[\sqrt{d}]$. The effectiveness was illustrated in Examples 1.40 and 1.41. Using ideal factorizations, we will solve these examples in a second way.

Theorem 4.50. *All* $\mathbf{Z}$*-solutions to* $x^2 - 15y^2 = 34$ *are generated from the two solutions* $(7, 1)$ *and* $(7, -1)$ *using units in* $\mathbf{Z}[\sqrt{15}]$*. That is, the* $\mathbf{Z}$*-solutions* (x, y)

arise as

$$x + y\sqrt{15} = \pm(7+\sqrt{15})u^k \text{ or } \pm(7-\sqrt{15})u^k,$$

where $u = 4+\sqrt{15}$ *and* $k \in \mathbf{Z}$.

Proof. In $\mathbf{Z}[\sqrt{15}]$, the ring of integers of $\mathbf{Q}(\sqrt{15})$,

$$34 = (7+\sqrt{15})(7-\sqrt{15}).$$

A nontrivial factor of $7 \pm \sqrt{15}$ would have norm ± 2 or ± 17. Since ± 2 and ± 17 are *not* of the form $x^2 - 15y^2$ (any x^2-15y^2 is a square mod 5, but ± 2 and ± 17 are not), $7+\sqrt{15}$ and $7-\sqrt{15}$ are irreducible in $\mathbf{Z}[\sqrt{15}]$. So *if* $\mathbf{Z}[\sqrt{15}]$ were a UFD then

$$x^2 - 15y^2 = 34 \Longrightarrow (x+y\sqrt{15})(x-y\sqrt{15}) = (7+\sqrt{15})(7-\sqrt{15}),$$

with $x \pm y\sqrt{15}$ not a unit (its norm is 34, not ± 1), so

$$x + y\sqrt{15} = (7 \pm \sqrt{15}) \cdot (\text{unit of norm } 1). \tag{4.6}$$

The unit group of $\mathbf{Z}[\sqrt{15}]$ is $\pm(4+\sqrt{15})^{\mathbf{Z}}$, with $\mathrm{N}(4+\sqrt{15}) = 1$. Feeding this into (4.6) would complete the proof, except for one problem: $\mathbf{Z}[\sqrt{15}]$ is not a UFD! In fact, the equation

$$2 \cdot 17 = (7+\sqrt{15})(7-\sqrt{15})$$

demonstrates this since any nontrivial factor of the four numbers here would have norm ± 2 or ± 17 and those are not norm-values. (We also need that the factors on the right aren't unit multiples of the factors on the left, and that's true since the factors on the left have norm 4 and 49 while those on the right have norm 34.)

Despite this mistake (of assuming $\mathbf{Z}[\sqrt{15}]$ is a UFD), the proof can be saved using unique factorization of ideals in $\mathbf{Z}[\sqrt{15}]$. If $x^2 - 15y^2 = 34$ then the principal ideal $(x+y\sqrt{15})$ has ideal norm 34, so

$$(x+y\sqrt{15}) = \mathfrak{p}_2\mathfrak{p}_{17}.$$

To see how (2) and (17) factor in $\mathbf{Z}[\sqrt{15}]$ we factor $T^2 - 15$ modulo 2 and 17:

$$T^2 - 15 \equiv (T-1)^2 \bmod 2, \quad T^2 - 15 \equiv (T+7)(T-7) \bmod 17,$$

so $(2) = \mathfrak{p}_2^2$ and $(17) = \mathfrak{p}_{17}\mathfrak{p}_{17}'$.

Therefore the only ideals of norm 34 in $\mathbf{Z}[\sqrt{15}]$ are $\mathfrak{p}_2\mathfrak{p}_{17}$ and $\mathfrak{p}_2\mathfrak{p}_{17}'$. Since $(7+\sqrt{15})$ and $(7-\sqrt{15})$ are ideals of norm 34 and are different (the ratio $(7+\sqrt{15})/(7-\sqrt{15})$ is not in $\mathbf{Z}[\sqrt{15}]$), the set of ideals with norm 34 can be described in two ways:

$$\{\mathfrak{p}_2\mathfrak{p}_{17}, \mathfrak{p}_2\mathfrak{p}_{17}'\} = \{(7+\sqrt{15}), (7-\sqrt{15})\}.$$

Thus

$$(x+y\sqrt{15}) = (7+\sqrt{15}) \text{ or } (7-\sqrt{15}),$$

which implies $x+y\sqrt{15} = (7\pm\sqrt{15})\cdot(\text{unit of norm 1})$. This is the same conclusion we drew before from the false assumption of unique factorization in $\mathbf{Z}[\sqrt{15}]$, but now it is justified. Go back to the list of units of norm 1 to finish the proof. ■

In this proof, we never needed generators for $\mathfrak{p}_2$, $\mathfrak{p}_{17}$, and $\mathfrak{p}_{17}'$. Existence of these prime ideals was sufficient.

Theorem 4.51. *The equation $x^2 - 82y^2 = 31$ has no integral solutions.*

Proof. We will work in $\mathbf{Z}[\sqrt{82}]$, the ring of integers of $\mathbf{Q}(\sqrt{82})$. The theorem is saying there is no principal ideal with norm 31. (Technically, to have a principal ideal $(x+y\sqrt{82})$ of norm 31 should allow for $x^2-82y^2 = -31$, but this is really the same problem since there is a unit in $\mathbf{Z}[\sqrt{82}]$ with norm -1, such as the fundamental unit $9+\sqrt{82}$. So 31 is the norm of an element if and only if -31 is the norm of an element.)

Since $T^2 - 82 \equiv (T+12)(T-12) \bmod 31$, $(31) = \mathfrak{p}_{31}\mathfrak{p}_{31}'$, where

$$\mathfrak{p}_{31} = (31, \sqrt{82}+12), \quad \mathfrak{p}_{31}' = (31, \sqrt{82}-12).$$

We want to show $\mathfrak{p}_{31}$ and $\mathfrak{p}_{31}'$ are not principal. We will work with $\mathfrak{p}_{31}$; the argument for $\mathfrak{p}_{31}'$ is similar.

To show $\mathfrak{p}_{31}$ is not principal, we will first show its square $\mathfrak{p}_{31}^2$ *is* principal, with an explicit generator. The square of $\mathfrak{p}_{31}$ is

$$\mathfrak{p}_{31}^2 = (961, 372+31\sqrt{82}, 226+24\sqrt{82})$$

and has norm $31^2 = 961$. If this were going to be principal, say $(x+y\sqrt{82})$, we would need a solution to $x^2-82y^2 = 961$ besides the obvious choices $(\pm 31, 0)$, and there is one: $x = 33$ and $y = 5$. Actually we could use $x = \pm 33$ and $y = \pm 5$, and those sign alternatives are important because a computation shows every generator of $\mathfrak{p}_{31}^2$ above is a multiple of $33 - 5\sqrt{82}$. So

$$\mathfrak{p}_{31}^2 = (33 - 5\sqrt{82}).$$

Having found a power of $\mathfrak{p}_{31}$ that is principal, we prove $\mathfrak{p}_{31}$ is not principal by contradiction. If $\mathfrak{p}_{31} = (\alpha)$ then $(\alpha^2) = (33 - 5\sqrt{82})$, so

$$\alpha^2 = (33 - 5\sqrt{82})u \tag{4.7}$$

where $u \in \mathbf{Z}[\sqrt{82}]^\times$. The solvability of this in α only depends on u up to multiplication by squares of units, since unit squares can be absorbed into α. Since $\mathbf{Z}[\sqrt{82}]^\times = \pm(9 + \sqrt{82})^{\mathbf{Z}}$, every unit is a unit square times one of 1, -1, $9 + \sqrt{82}$, or $-(9 + \sqrt{82})$. Since $\alpha^2 > 0$ and $33 - 5\sqrt{82} < 0$ we need $u < 0$ so we just need to check (4.7) can't be solved for α when $u = -1$ and $u = -(9+\sqrt{82})$. (Notice we need to be paying attention to signs and not sloppily saying the units of $\mathbf{Z}[\sqrt{82}]$ are the powers of $9 + \sqrt{82}$.) We want to show the equations

$$\alpha^2 = 5\sqrt{82} - 33 \quad \text{and} \quad \alpha^2 = (5\sqrt{82} - 33)(9 + \sqrt{82}) = 113 + 12\sqrt{82} \tag{4.8}$$

have no solution in $\mathbf{Z}[\sqrt{82}]$.

Since $5\sqrt{82} - 33$ is a root of $T^2 + 66T - 961$, $\pm\sqrt{5\sqrt{82} - 33}$ is a root of $T^4 + 66T^2 - 961$. Similarly, $\pm\sqrt{113 + 12\sqrt{82}}$ is a root of $T^4 - 226T^2 + 961$. PARI says both of these quartic polynomials are irreducible over $\mathbf{Q}$ (*e.g.*, the PARI command `polisirreducible(x^4+66*x^2-961)` has affirmative answer `1`), so solutions to the equations in (4.8) don't lie in the quadratic field $\mathbf{Q}(\sqrt{82})$.

It's important to know there is a different way to show the equations in (4.8) are not solvable in $\mathbf{Z}[\sqrt{82}]$, using reduction modulo a prime ideal: if $a \in \mathbf{Z}[\sqrt{82}]$ and $\alpha^2 = a$ for some $\alpha \in \mathbf{Z}[\sqrt{82}]$, then $\alpha^2 \equiv a \bmod \mathfrak{p}$ for any prime ideal $\mathfrak{p}$. Therefore if there is a prime $\mathfrak{p}$ such that $a \bmod \mathfrak{p}$ is not a square in $\mathbf{Z}[\sqrt{82}]/\mathfrak{p}$ then a is not a square in $\mathbf{Z}[\sqrt{82}]$. A nonsquare in $\mathbf{Z}[\sqrt{82}]$ is a nonsquare modulo $\mathfrak{p}$ for half of all $\mathfrak{p}$, in a suitable sense, so we expect to find $\mathfrak{p}$ such that $a \not\equiv \square \bmod p$ fairly quickly if there any at all. We will now reduce the right sides of (4.8) modulo suitable $\mathfrak{p}$ to see they are not squares in $\mathbf{Z}[\sqrt{82}]/\mathfrak{p}$, and thus are not squares in $\mathbf{Z}[\sqrt{82}]$. (The prime ideal used for each equation need not be the same.)

We won't reduce (4.8) modulo prime ideals dividing (2), since every number in a finite field of characteristic 2 is a square and we have no chance of finding a contradiction that way. Let's look at prime ideals dividing (3). Since $T^2 - 82 \equiv (T+1)(T-1) \bmod 3$,

$$(3) = \mathfrak{p}_3\mathfrak{p}_3', \quad \text{where} \quad \mathfrak{p}_3 = (3, \sqrt{82}+1) \quad \text{and} \quad \mathfrak{p}_3' = (3, \sqrt{82}-1).$$

In $\mathbf{Z}[\sqrt{82}]/\mathfrak{p}_3 \cong \mathbf{F}_3$, $\sqrt{82}$ reduces to -1 and

$$5\sqrt{82} - 33 \equiv 5(-1) \equiv 1 \bmod 3, \quad (5\sqrt{82}-33)(9+\sqrt{82}) \equiv -1 \bmod 3.$$

Since $-1 \bmod 3$ is not a square, we have shown the second equation in (4.8) has no solution in $\mathbf{Z}[\sqrt{82}]$ since it has no solution modulo $\mathfrak{p}_3$. In $\mathbf{Z}[\sqrt{82}]/\mathfrak{p}_3' \cong \mathbf{F}_3$, $\sqrt{82}$ reduces to 1 and

$$5\sqrt{82} - 33 \equiv 5 \equiv 2 \not\equiv \square \bmod 3,$$

so we have a contradiction to the first equation in (4.8) by working mod $\mathfrak{p}_3'$.

If we hadn't been prescient enough to remove the choices $u = 1$ and $u = 9 + \sqrt{82}$ in (4.7) on the ground of sign problems by looking at (4.7) in $\mathbf{R}$, we could have eliminated them by congruence arguments: there is no solution to (4.7) reduced mod $\mathfrak{p}_3$ when $u = 1$ and there is no solution to (4.7) reduced modulo one of the primes dividing (13) when $u = 9 + \sqrt{82}$. Check that as Exercise 4.18. ■

Obviously the treatment of $x^2 - 15y^2 = 34$ and $x^2 - 82y^2 = 31$ here is more involved than in Section 1.4, but it is a good illustration of how prime ideal factorizations can be used to explore the solvability of Diophantine equations. We also see the importance of knowing the structure of the unit group in order to prove ideals are not principal.

4.7 Replacing Z with $F[X]$

We return to the theme of unique factorization of ideals, and want to extend that result to more rings than the integral closure of $\mathbf{Z}$ in a number field. The polynomial ring $F[X]$, for F a field, is a basic example of a PID besides $\mathbf{Z}$. Its fraction field is $F(X)$.

Definition 4.52. Let F be a field. A field that is isomorphic to $F(X)$, where X is an indeterminate (intrinsically, X is transcendental over F), is called a *rational function field over F.* A finite extension of a rational function field over F is called a *function field over F.*

Example 4.53. One function field over F is $K = F(x, y)$ where x and y are transcendental over F and satisfy the algebraic relation $x^2 + y^2 = 1$. Assuming F does not have characteristic 2, K is a quadratic extension of the rational function field $F(x)$ since y is a root of the polynomial $Y^2 - (1 - x^2) \in F(x)[Y]$, which is irreducible over $F(x)$ since $1 - x^2 = (1 + x)(1 - x)$ is not a square in $F(x)$.

Although K is presented to us as generated by two elements over F, it in fact can be generated by one element over F: $K = F(u)$ where $u = y/(x + 1)$. Verify that

$$x = \frac{1 - u^2}{1 + u^2}, \quad y = \frac{2u}{1 + u^2}. \tag{4.9}$$

Therefore K is a rational function field over F. (The element u is transcendental over F since $x \in F(u)$ and x is transcendental over F.) This realization of K as a rational function field over F has its origins in calculus: if $F = \mathbf{R}$, $x = \cos\theta$, and $y = \sin\theta$ then $u = \tan(\theta/2)$ (see Figure 4.1) and (4.9) becomes the sneaky $\tan(\theta/2)$-substitution integral calculus that converts any integral of a rational function of $\sin\theta$ and $\cos\theta$ into an integral of a rational function in u.

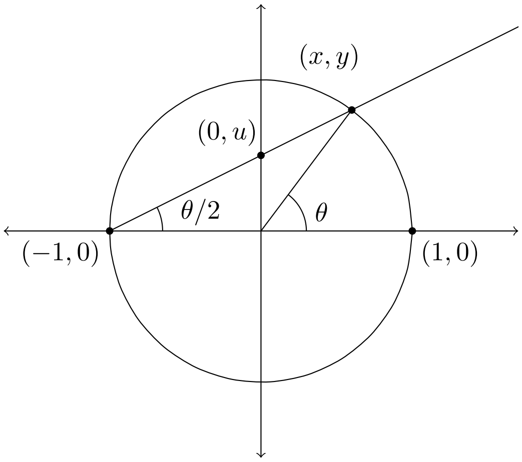


Figure 4.1: Rational parametrization of the unit circle.

Although we can write $K = F(u)$, the subring $F[x, y]$ usually

What happens if F has characteristic 2? Then the equation $x^2 + y^2 = 1$ is the same as $(x+y)^2 = 1$, so $x+y = 1$ and $F(x,y) = F(x,1-x) = F(x)$ is also a rational function field.

Example 4.54. Another function field over F is $K = F(x,y)$ where x and y are transcendental over F and satisfy $y^2 = x^3 - x$. Think of K as the quadratic extension of $F(x)$ by a root of $Y^2 - (x^3-x) \in F(x)[Y]$, which is irreducible over $F(x)$ since $x^3 - x$ is not a square in $F(x)$. Or think of K as the cubic extension of $F(y)$ by a root of $X^3 - X - y^2 \in F(y)[X]$, which is irreducible over $F(y)$ since $1 - y^2$ is not a cube in $F(y)$.

Unlike the previous example, K is not a rational function field over F. If we could write $K = F(u)$ for some $u \in K$ then u is transcendental over F and x and y are nonconstant rational functions of u, say $x = f(u)$ and $y = g(u)$. Then $g(u)^2 = f(u)^3 - f(u)$, so the equation $Y^2 = X^3 - X$ has a solution in nonconstant elements of $F(u)$. When F does not have characteristic 2, the impossibility of such a solution can be proved algebraically [44, pp. 28–29] or geometrically [52, p. 19].

What happens if F has characteristic 2? Then $y^2 = x^3 - x = x(x-1)^2$, and if we set $t = y/(x-1)$ then $t^2 = x$ and $y = t(x-1) = t(t^2-1)$, so $F(x,y) = F(t)$ is a rational function field.

Nonexample 4.55. The field $\mathbf{Q}(\pi)$ is a rational function field over $\mathbf{Q}$ since π is a transcendental number. It is isomorphic to $\mathbf{Q}(X)$ by identifying X with π.

Example 4.56. The field $\mathbf{Q}(\sqrt{2}, X)$ is a function field over $\mathbf{Q}$ and also over $\mathbf{Q}(\sqrt{2})$. To study a function field over F, it is common to replace F with its largest algebraic extension in the function field (necessarily a finite extension; see Exercise 4.36), so the only elements of the function field which are algebraic over F are in F. We then say F is algebraically closed in the function field, but that doesn't mean F is an algebraically closed field; $\mathbf{Q}$ is algebraically closed in $\mathbf{Q}(X)$.

Any field F is algebraically closed in $F(X)$ (Exercise 4.29), so the largest algebraic extension of $\mathbf{Q}$ inside $\mathbf{Q}(\sqrt{2}, X) = \mathbf{Q}(\sqrt{2})(X)$ is $\mathbf{Q}(\sqrt{2})$ and it is natural to study $\mathbf{Q}(\sqrt{2}, X)$ first as a function field over $\mathbf{Q}(\sqrt{2})$ rather than over $\mathbf{Q}$.

Remark 4.57. Function fields over F can be described without any reference to a rational function subfield as the finitely generated field extensions of F with transcendence degree 1 over F. The field $\overline{\mathbf{Q}}(X)$ is not a function field over

Q, even though it has transcendence degree 1 over **Q** since it is not finitely generated over **Q**.

Historically, the first function fields studied were over **C**, as fields of (actual) functions on compact Riemann surfaces. A Riemann surface is a connected one-dimensional complex manifold. These are the spaces where one-variable complex analysis can be carried out. If you are not familiar with compact Riemann surfaces, just keep the simplest example in mind: the Riemann sphere $\mathbf{C} \cup \{\infty\}$, as pictured below.

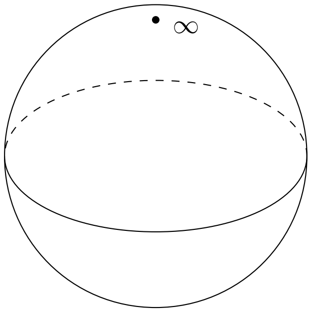


Figure 4.2: The Riemann sphere.

The rational functions $\mathbf{C}(z)$ are natural functions from the Riemann sphere to itself. It is a theorem from complex analysis that every meromorphic function on the Riemann sphere (only poles allowed as singularities) is a rational function.

(Readers who know about elliptic functions know another function field: for any lattice L in **C**, the Weierstrass $\wp$-function $\wp_L(z)$ and its derivative $\wp'_L(z)$ generate the field of meromorphic functions on the torus $\mathbf{C}/L$, a compact Riemann surface, as pictured below, and there is an algebraic relation $\wp'_L(z)^2 = 4\wp_L(z)^3 - g_2(L)\wp_L(z) - g_3(L)$, where the constants $g_2(L)$ and $g_3(L)$ depend on L. Letting $x = \wp_L(z)$ and $y = \wp'_L(z)$, $\mathbf{C}(\wp_L, \wp'_L) = \mathbf{C}(x, y)$, where y is quadratic over $\mathbf{C}(x)$.)

Consider a diagram

$$\begin{tikzcd}
 & K \arrow[d, dash, "n"] \\
R \arrow[ur, dash] \arrow[d, dash] & F(X) \\
F[X] \arrow[ur, dash] &
\end{tikzcd}$$

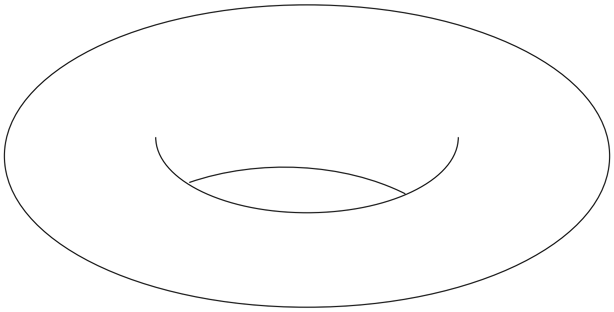


Figure 4.3: A torus.

where $K/F(X)$ is a finite extension of degree n and R is the integral closure of $F[X]$ in K. Does R have unique factorization of ideals? Following the method used over $\mathbf{Z}$, the first thing we want to check is that R has an $F[X]$-basis, *i.e.*, $R \cong F[X]^n$ as an $F[X]$-module. Here we need to make a separability hypothesis.

Theorem 4.58. *Let $K/F(X)$ be a finite separable extension of degree n and R be the integral closure of $F[X]$ in K. Then R, as well as any nonzero ideal in R, is a finite free $F[X]$-module of rank n.*

Proof. When $K/F(X)$ is *separable*, we can bring in a discriminant: picking an $F(X)$-basis $\{e_1, \dots, e_n\}$ of K that lies in R (such a basis can be found by scaling, because K is the fraction field of R),

$$\sum_{i=1}^{n} F[X]e_i \subset R \subset \sum_{i=1}^{n} F[X]\frac{e_i}{d},$$

where $d = \operatorname{disc}_{K/F(X)}(e_1, \dots, e_n) = \det(\operatorname{Tr}_{K/F(X)}(e_ie_j)) \in F[X] - \{0\}$. Having placed R between two finite free $F[X]$-modules of rank n, $R \cong F[X]^n$ as an $F[X]$-module (Corollary 8.29). Any nonzero ideal in R is also free of rank n as an $F[X]$-module by the same argument used for nonzero ideals in $\mathcal{O}_K$ as $\mathbf{Z}$-modules (Theorem 3.4). ■

Theorem 4.59. *For any field F, if the integral closure R of $F[X]$ in a finite extension of $F[X]$ is a finite free $F[X]$-module then R has unique factorization of ideals.*

Proof. The quotient rings of R are infinite if F is infinite (*e.g.*, $\mathbf{C}[X]/(X) = \mathbf{C}$ is infinite), so we can't argue by induction on the size of $R/\mathfrak{a}$. There nevertheless is a finiteness condition that we can exploit: for any nonzero ideal $\mathfrak{a}$ in R, the ring $R/\mathfrak{a}$ is a finite-dimensional vector space over F. To see this, choose aligned $F[X]$-bases for $\mathfrak{a}$ and R (Theorem 8.33):

$$R = F[X]e_1 \oplus \cdots \oplus F[X]e_n, \quad \mathfrak{a} = F[X]f_1e_1 \oplus \cdots \oplus F[X]f_ne_n$$

for some nonzero $f_1, \ldots, f_n \in F[X]$. Then

$$R/\mathfrak{a} \cong \bigoplus_{i=1}^{n}(F[X]/(f_i))\overline{e}_i.$$

Each $F[X]/(f_i)$ is a finite-dimensional F-vector space (the dimension is the degree of f_i), so $R/\mathfrak{a}$ is finite-dimensional as an F-vector space. This is the polynomial analogue of $\mathcal{O}_K/\mathfrak{a}$ being a finite ring. Using $\dim_F(R/\mathfrak{a})$ in place of $\#(\mathcal{O}_K/\mathfrak{a}) = \mathrm{N}(\mathfrak{a})$ from the number field case, the proof that $\mathcal{O}_K$ has unique factorization of ideals can be carried over to show R has unique factorization of ideals. It is left to the reader to check the details as part of Exercise 4.30. ■

Since all finite extensions in characteristic 0 are separable, Theorems 4.58 and 4.59 tell us R has unique factorization of ideals when F has characteristic 0. If F has characteristic p and $K/F(X)$ is inseparable, all bases have discriminant 0, so there is no nonzero discriminant to help us prove R has an $F[X]$-basis (and then conclude that $\dim_F(R/\mathfrak{a}) < \infty$ for all nonzero ideals $\mathfrak{a}$ in R). That does *not* mean R can't be free of rank n as an $F[X]$-module, or that R doesn't have unique factorization of ideals, but our approach definitely breaks down for inseparable $K/F(X)$ since without knowing for sure if $\dim_F(R/\mathfrak{a}) < \infty$ it is not clear how we would, say, prove the existence of prime ideal factorizations in R as in Section 4.2.2 (what would a smallest counterexample mean?).

The following theorem will help us push the unique factorization of ideals through for inseparable extensions when F is a finite field.

Theorem 4.60. *Let K be a function field over the finite field $\mathbf{F}$. There is some $U \in K$ such that $K/\mathbf{F}(U)$ is separable.*

Proof. By assumption, K is a finite extension of some rational function field $\mathbf{F}(X)$. Let p be the characteristic of $\mathbf{F}$. We want to write X as a p^mth power in K with $m \geqslant 0$ as large as possible. Why is there an upper bound on m? If $X = U^{p^m}$ in K, then U has degree p^m over $\mathbf{F}(X)$ since $T^{p^m} - X$ is Eisenstein at X in $\mathbf{F}[X]$, so $\mathbf{F}(X) \subset \mathbf{F}(U) \subset K$ with $[\mathbf{F}(U) : \mathbf{F}(X)] = p^m \leqslant [K : \mathbf{F}(X)]$.

Thus m is bounded above. Writing $X = U^{p^m}$ in K with m as large as possible, we will show that $K/\mathbf{F}(U)$ is separable.

Every finite extension of fields K/F in characteristic p can be expressed as a separable extension E/F of the base field followed by a purely inseparable extension K/E up to the top field. (If you don't know what purely inseparable extensions are, the rest of this proof won't make sense, so skip it or read about purely inseparable extensions in an algebra book and then come back.)

$$\begin{tikzcd} K \arrow[d, dash, "\text{purely inseparable}"'] \\ E \arrow[d, dash, "\text{separable}"'] \\ F \end{tikzcd}$$

For the particular extension $K/\mathbf{F}(U)$, this means there is a field K', where $\mathbf{F}(U) \subset K' \subset K$ with $K'/\mathbf{F}(U)$ separable and K/K' purely inseparable. Set $p^r = [K : K']$. Then $K^{p^r} \subset K' \subset K$.

We now show $[K : K^{p^r}] = p^r$. Consider the following tower of field extensions, where $d = [K : \mathbf{F}(U)]$.

$$\begin{tikzcd} & K \arrow[dl, dash] \arrow[dr, dash, "d"] & \\ K^p \arrow[dr, dash] & & \mathbf{F}(U) \arrow[dl, dash, "p"] \\ & \mathbf{F}(U^p) & \end{tikzcd}$$

The pth power map is an isomorphism $K \cong K^p$ that sends $\mathbf{F}(U)$ to $(\mathbf{F}(U))^p = \mathbf{F}^p(U^p) = \mathbf{F}(U^p)$. Therefore $[K^p : \mathbf{F}(U^p)] = [K : \mathbf{F}(U)] = d$, so $[K : K^p] = p$ from the field diagram. Running through this argument again with K and its subfield $\mathbf{F}(U)$ replaced by K^p and its subfield $\mathbf{F}(U^p)$, $[K^p : K^{p^2}] = p$, so $[K : K^{p^2}] = [K : K^p][K^p : K^{p^2}] = p^2$. Repeating the argument several times, $[K : K^{p^r}] = p^r$.

Since $K^{p^r} \subset K' \subset K$ with $[K : K'] = p^r$ and $[K : K^{p^r}] = p^r$, $K' = K^{p^r}$ by comparing field degrees. Then $\mathbf{F}(U) \subset K^{p^r}$. Since U is not a pth power in K (by maximality of m), we must have $r = 0$, so $K = K'$ is separable over $\mathbf{F}(U)$. ■

Corollary 4.61. *If* $\mathbf{F}$ *is a finite field, the integral closure of* $\mathbf{F}[X]$ *in any finite* *extension of* $\mathbf{F}(X)$ *is a finite free* $\mathbf{F}[X]$*-module and has unique factorization of ideals.*

Proof. Let $K/\mathbf{F}(X)$ be a finite extension and R be the integral closure of $\mathbf{F}[X]$ in K. By the proof of Theorem 4.60, if we write $X = U^{p^m}$ for some $U \in K$ with m as large as possible, then $\mathbf{F}(X) \subset \mathbf{F}(U) \subset K$ and $K/\mathbf{F}(U)$ is separable. The integral closure of $\mathbf{F}[X]$ in $\mathbf{F}(U)$ is $\mathbf{F}[U]$, since $\mathbf{F}[U]$ is integral over $\mathbf{F}[X]$ and is integrally closed (any PID is integrally closed), so the integral closure of $\mathbf{F}[U]$ in K is R. Since $K/\mathbf{F}(U)$ is separable, R is a finite free $\mathbf{F}[U]$-module (Theorem 4.58). The minimal integral relation for U over $\mathbf{F}[X]$ is $U^{p^m} = X$, so $\mathbf{F}[U]$ has $\mathbf{F}[X]$-basis $\{1, X, \dots, X^{p^m} - 1\}$ and therefore R is a finite free $\mathbf{F}[X]$-module. By Theorem 4.59, R has unique factorization of ideals. ■

Where did we need $\mathbf{F}$ to be a finite field in Theorem 4.60? In just one place: the computation $(\mathbf{F}(U))^p = \mathbf{F}(U^p)$, which is due to $\mathbf{F}^p = \mathbf{F}$. Fields F of characteristic p satisfying $F^p = F$, as well as all fields of characteristic 0, are called *perfect* fields. The significance of these fields is that perfect fields are precisely the fields whose finite extensions are all separable. The standard examples of perfect fields in characteristic p are finite fields and algebraically closed fields of characteristic p. (The basic nonperfect field is $k(X)$, where k is any field of characteristic p, since X is not a pth power in $k(X)$.) The proofs of Theorem 4.60 and Corollary 4.61 go through when finite fields are replaced by perfect fields. Theorem 4.60 has counterexamples when $\mathbf{F}$ is replaced by some infinite fields of characteristic p (Exercise 4.35), but that doesn't mean Corollary 4.61 is false when $\mathbf{F}$ is replaced by an infinite field of characteristic p; in fact Corollary 4.61 is true when $\mathbf{F}$ is replaced by any field at all. We will come back to this point in Remark 4.70.

While the fields $\mathbf{Q}$ and $F(X)$ are quite analogous, there are essential differences between number fields (finite extensions of $\mathbf{Q}$) and function fields (finite extensions of $F(X)$). First of all, while there is a unique copy of $\mathbf{Q}$ inside any number field, there is not a unique copy of $F(X)$ inside a function field over F. In Example 4.54, for instance, the function field $F(x, y)$ contains the two rational function fields $F(x)$ and $F(y)$ and it has different degrees over each of them. The moral is that function fields, unlike number fields, have no natural bottom. At the same time, it is a basic theorem of field theory that any field lying strictly between F and $F(X)$ has the form $F(f(X))$ for some nonconstant rational function $f(X)$, which means the only function fields inside a rational function field are themselves rational function fields (example: $F(X^5)$ inside

$F(X)$). Therefore in a sense the "bottom" of any function field is a rational function field, except you need to remember that there isn't a canonical bottom to pick.

Another essential difference between number fields and function fields is that $\mathbf{Q}$ has only the trivial field automorphism while $F(X)$ has many non-identity field automorphisms, like $f(X) \mapsto f(1/X)$, or more generally any linear fractional change of variables $X \mapsto (aX+b)/(cX+d)$ where $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(F)$. (All automorphisms of $F(X)$ fixing F arise in this way; see Exercise 4.29.) This automorphism sends $F[X]$ to $F[(aX+b)/(cX+d)]$. While the rings $F[X]$ and $F[(aX+b)/(cX+d)]$ are isomorphic, their integral closures inside a finite extension of $F(X)$ need not be isomorphic (Exercise 4.34), so there is not just one analogue in function fields of the canonical ring $\mathcal{O}_K$ in a number field K. There can be very interesting mathematics in the integral closure of $F[X]$ in a finite extension of $F(X)$, but the choice of $F[X]$ among all the rings $F[(aX+b)/(cX+d)]$, say, is still a choice. On the geometric side, selecting $\mathbf{C}[X]$ as a preferred subring of $\mathbf{C}(X)$ amounts to marking one point on the Riemann sphere and focusing on the ring of rational functions with no pole away from that point. (Marking the north pole gives $\mathbf{C}[z]$ and marking the south pole gives $\mathbf{C}[1/z]$.) Each point of the Riemann sphere is on the same footing as any other, since the sphere is a homogeneous object.

4.8 Some (More) Commutative Algebra

To develop a general setting where unique factorization of ideals can be proved, the correct replacement for finiteness of $\mathcal{O}_K/\mathfrak{a}$ or finite-dimensionality of $R/\mathfrak{a}$ (R an integral closure of $F[X]$) is the following finiteness condition on ideals.

Definition 4.62. A commutative ring A is called *Noetherian* if all of its ideals $\mathfrak{a}$ are finitely generated; *i.e.*,

$$\mathfrak{a} = Ax_1 + \cdots + Ax_n.$$

Example 4.63. Any PID is Noetherian since each ideal has only one generator.

Example 4.64. Any ring A that is a finitely generated $\mathbf{Z}$-module is Noetherian since any ideal $\mathfrak{a}$ in A is a subgroup of A and subgroups of finitely generated abelian groups are finitely generated abelian groups, so $\mathfrak{a}$ is a finitely generated $\mathbf{Z}$-module, say $\mathfrak{a} = \mathbf{Z}x_1 + \cdots + \mathbf{Z}x_n$. Since $\mathfrak{a}$ is an ideal, the $\mathbf{Z}$-module generators are also ideal generators: $\mathfrak{a} = Ax_1 + \cdots + Ax_n$. Thus $\mathfrak{a}$ is a finitely generated ideal.

Examples of rings that are finitely generated $\mathbf{Z}$-modules are $\mathcal{O}_K$ or $\mathbf{Z}[\alpha]$ for an algebraic integer α. All such rings are Noetherian.

Nonexample 4.65. The most basic non-Noetherian ring is the polynomial ring over a field with countably many indeterminates: in $F[T_1, T_2, \ldots]$, the ideal $\mathfrak{a} = (T_1, T_2, \ldots)$ of polynomials with constant term 0 is not finitely generated. Any finite set of polynomials $f_1, \ldots, f_n \in \mathfrak{a}$ involves only finitely many of the T_i's. Setting those finitely many T_i's equal to 0 makes $f_1, \ldots, f_n$ all vanish and therefore makes all polynomials in the ideal $(f_1, \ldots, f_n)$ vanish. Any other T_j does not vanish and is in $\mathfrak{a}$, so $\mathfrak{a} \neq (f_1, \ldots, f_n)$.

Nonexample 4.66. A less widely-known example of a non-Noetherian ring is the ring of "integral-valued polynomials" $\mathrm{Int}(\mathbf{Z}) = \{f \in \mathbf{Q}[T] : f(\mathbf{Z}) \subset \mathbf{Z}\}$. You might think: isn't that just $\mathbf{Z}[T]$? There is more than that, *e.g.*, $\frac{T(T-1)}{2}$ belongs to the ring. As an abelian group, $\mathrm{Int}(\mathbf{Z})$ has as a $\mathbf{Z}$-basis the binomial coefficient polynomials:

$$\mathrm{Int}(\mathbf{Z}) = \bigoplus_{n \geqslant 0} \mathbf{Z}\binom{T}{n}.$$

An example of an ideal in $\mathrm{Int}(\mathbf{Z})$ that is not finitely generated is the ideal generated by $\{\binom{T}{p^r} : r \geqslant 0\}$ for a fixed prime p.

For greater flexibility, it is important to have the Noetherian concept on modules, not just rings.

Definition 4.67. For any commutative ring A, an A-module M is called *Noetherian* if all of its submodules are finitely generated.

Since ideals in a ring A are precisely the submodules of A viewed as a module over itself, A is a Noetherian ring exactly when it is a Noetherian A-module. Any submodule of a Noetherian A-module M is a Noetherian A-module, since if all submodules of M are finitely generated then all submodules of any submodule of M are finitely generated. The corresponding statement for rings is false: a subring of a Noetherian ring need not be a Noetherian ring (consider $\mathrm{Int}(\mathbf{Z})$ inside $\mathbf{Q}[T]$, or more simply any non-Noetherian domain inside its fraction field). The point is that the Noetherian property for a subring involves a smaller ring of scalars, while the Noetherian property for a submodule does not (same scalar ring as before).

Theorem 4.68. *Let A be a commutative ring.*

(*a*) *For A-modules $N \subset M$, M is Noetherian if and only if N and M/N are Noetherian.*

(*b*) *For A-modules M and N, $M \oplus N$ is Noetherian if and only if M and N are Noetherian.*

Proof. (a) Suppose M is Noetherian. Any submodule of N is a submodule of M and therefore is finitely generated. So N is Noetherian. Any submodule $M' \subset M/N$ can be lifted to a submodule of M by the natural reduction map $M \to M/N$ (take the inverse image of M' in M). The lifting is $M'' = \{m \in M : m \bmod N \in M'\}$. Since M is Noetherian, M'' is finitely generated, say by $m_1, \ldots, m_d$. Then $m_1 \bmod N, \ldots, m_d \bmod N$ generate M', as $M''/N = M'$. So M/N is Noetherian.

Now suppose N and M/N are Noetherian. For any submodule $\widetilde{M}$ of M, its image under the natural map $M \to M/N$ is a submodule of M/N so it is finitely generated. Let $\overline{m}_1, \ldots, \overline{m}_k$ generate the image of $\widetilde{M}$ in M/N for some m_i's in $\widetilde{M}$. The intersection $\widetilde{M} \cap N$ is a submodule of N, so it is finitely generated, say by $n_1, \ldots, n_\ell$. We will show $m_1, \ldots, m_k, n_1, \ldots, n_\ell$ generate $\widetilde{M}$. For any $m \in \widetilde{M}$, write $\overline{m} = \sum_{i=1}^k a_i \overline{m}_i$ for some $a_i \in A$. Then $m - \sum_{i=1}^k a_i m_i \in N \cap \widetilde{M}$, so the difference is a linear combination of $n_1, \ldots, n_\ell$, which makes m a linear combination of $m_1, \ldots, m_k, n_1, \ldots, n_\ell$.

(b) Apply part a to the A-module $M \oplus N$ and its submodule $\{0\} \oplus N \cong N$, with $(M \oplus N)/(\{0\} \oplus N) \cong M$. ■

If A is a Noetherian ring, then it is a Noetherian A-module, so for any $n \geqslant 1$, A^n is a Noetherian A-module by Theorem 4.68(b); use $A^n \cong A \oplus A^{n-1}$ and induction. It follows that any finitely generated A-module is Noetherian: if $M = Ax_1 + \cdots + Ax_n$ then there is a surjection $A^n \to M$ sending the standard basis of A^n to the spanning set $x_1, \ldots, x_n$ of M, so M is isomorphic to a quotient of A^n and thus is a Noetherian A-module. In words: a finitely generated module over a Noetherian ring is a Noetherian module over that ring. The following corollary puts this to use.

Corollary 4.69. *Let A be a Noetherian integrally closed domain with fraction field F, E/F be a finite separable extension, and B be the integral closure of A in E. Then B is a finitely generated A-module and a Noetherian ring.*

Proof. Since A is integrally closed, $\mathrm{Tr}_{E/F}(B) \subset A$. We can choose an F-basis $\{e_1, \ldots, e_n\}$ of E in A. Set $d = \det(\mathrm{Tr}_{E/F}(e_ie_j)) \in A - \{0\}$. (That $d \neq 0$ is where we need E/F separable.) Then

$$\sum_{i=1}^{n} Ae_i \subset B \subset \sum_{i=1}^{n} A\frac{e_i}{d}.$$

This places B inside a finite free A-module,[4] so B is a Noetherian A-module because A^n is a Noetherian A-module. In particular, B is finitely generated as an A-module. Any ideal in B is finitely generated as an A-module, and thus finitely generated as a B-module (can use the same generating set over B as used over A). That shows B is a Noetherian ring. ■

Remark 4.70. If we drop the separability condition on E/F in this corollary, then the whole discriminant argument breaks down (we don't get an embedding of B into A^n) and in fact the result is false: B need not be a finitely generated A-module, even if A is a PID! For an example, see [6, Exer. 11, p. 205]. However, if $A = F[X]$ for any field F, or more generally $A = F[x_1, \ldots, x_m]$ is a finitely generated ring over a field F, then B is a finitely generated A-module [18, pp. 297–298].

The Noetherian property interacts well with formation of quotient rings and polynomial rings.

Theorem 4.71. *If A is a Noetherian ring, so is $A/\mathfrak{a}$ for any ideal $\mathfrak{a}$ and $A[T]$.*

Proof. Ideals in $A/\mathfrak{a}$ are the same thing as A-modules in $A/\mathfrak{a}$, and $A/\mathfrak{a}$ is a Noetherian A-module when A is a Noetherian ring by Theorem 4.68. See [2, p. 81], or really any book that discusses Noetherian rings, for a proof that $A[T]$ is Noetherian when A is. ■

By induction, when A is a Noetherian ring, so is $A[T_1, \ldots, T_n]$. That the Noetherian property is preserved by $A \rightsquigarrow A[T_1, \ldots, T_n]$ is called the *Hilbert basis theorem*. Hilbert proved it to show ideals in polynomial rings over a field are finitely generated.

Example 4.72. All ideals in $\mathbf{C}[T_1, \ldots, T_n]$ are finitely generated.

We will have no use for the Hilbert basis theorem, but it shows that many rings one meets in algebra are Noetherian: any ring that is finitely generated

[4] This does *not* imply B is a finite free A-module, since A is not assumed to be a PID.

over $\mathbf{Z}$ has the form $\mathbf{Z}[a_1, \ldots, a_n]$ for some a_i's in the ring, and such a ring is naturally a homomorphic image of $\mathbf{Z}[T_1, \ldots, T_n]$, so $\mathbf{Z}[a_1, \ldots, a_n]$ is isomorphic to a quotient ring of a Noetherian ring, which must be Noetherian. Similarly, any ring that is finitely generated over a field is Noetherian. By comparison with the Noetherian property on rings, the PID property is not as well preserved. If A is a PID then $A[T]$ is *not* a PID, unless A is a field (and then $A[T, U]$ is not a PID).

Remark 4.73. Using the Krull dimension (Section 2.2), if A is Noetherian then $\dim A[T] = 1 + \dim A$. (A proof is in [38, Theorem 22, p. 83].) Therefore by induction,

$$\dim A[T_1, \ldots, T_n] = n + \dim A.$$

The most general ring which has unique factorization of ideals is called a Dedekind domain.

Definition 4.74. A *Dedekind domain* is a domain that is

(1) a Noetherian ring,

(2) integrally closed,

(3) 1-dimensional: all nonzero prime ideals are maximal.

Example 4.75. Any PID is a Dedekind domain.

Example 4.76. The ring of integers in any number field is a Dedekind domain and these are often not PIDs (such as $\mathbf{Z}[\sqrt{-5}]$).

To show every Dedekind domain has unique factorization of ideals, we use three lemmas.

Lemma 4.77. *For any commutative ring A, an A-module M is Noetherian if and only if any infinite ascending chain of submodules in M stabilizes:*

$$M_1 \subset M_2 \subset M_3 \subset \cdots \Longrightarrow M_i = M_{i+1} \textit{ for large } i.$$

In particular, A is a Noetherian ring if and only if any infinite ascending chain of ideals stabilizes.

Proof. Suppose M is Noetherian and there is an ascending chain of submodules in M:

$$M_1 \subset M_2 \subset M_3 \subset \cdots .$$

The union $\bigcup M_i$ is a submodule of M so it is finitely generated. The generating set is in some M_n, so $\bigcup M_i = M_n$, which implies $M_i = M_n$ for $i \geqslant n$.

Suppose M is not Noetherian. Then there is a submodule $N \subset M$ that is not finitely generated, so any finitely generated submodule of N is strictly smaller than N. We can recursively find a sequence $x_1, x_2, \ldots$ in N such that $x_{n+1} \notin Ax_1 + \cdots + Ax_n$. Set $M_i = Ax_1 + \cdots + Ax_i$, so

$$M_1 \subsetneqq M_2 \subsetneqq M_3 \subsetneqq \cdots$$

is an infinite ascending chain of submodules of M that does not stabilize. ∎

Lemma 4.78. *In a Noetherian domain, any nonzero ideal* $\mathfrak{a}$ *contains a product of nonzero prime ideals:*

$$\mathfrak{a} \supset \mathfrak{p}_1 \cdots \mathfrak{p}_n.$$

Proof. We adapt the proof we gave earlier in the case of $\mathcal{O}_K$ (Lemma 4.15), but we have to make a change since the proof in $\mathcal{O}_K$ appealed to a counterexample with least index and in a general Noetherian domain the ideals don't have finite index. Compare what we do here with with the proof in $\mathcal{O}_K$.

If the result is false, let $\mathfrak{a}$ be a counterexample: $\mathfrak{a}$ is an ideal that does not contain a product of nonzero prime ideals. Then $\mathfrak{a}$ is not the whole ring since any nonzero commutative ring contains a maximal ideal and any maximal ideal is prime. The ideal $\mathfrak{a}$ is also not prime, so there are x and y in the ring such that $xy \in \mathfrak{a}$ and $x \notin \mathfrak{a}$ and $y \notin \mathfrak{a}$. The ideals $(x) + \mathfrak{a}$ and $(y) + \mathfrak{a}$ each strictly contain $\mathfrak{a}$ while their product is in $\mathfrak{a}$:

$$\begin{aligned}((x) + \mathfrak{a})((y) + \mathfrak{a}) &= (x)(y) + (x)\mathfrak{a} + (y)\mathfrak{a} + \mathfrak{a}^2 \\ &= (xy) + x\mathfrak{a} + y\mathfrak{a} + \mathfrak{a}^2 \\ &\subset \mathfrak{a}.\end{aligned}$$

Since $\mathfrak{a}$ doesn't contain a product of nonzero ideals, the inclusion above shows $(x) + \mathfrak{a}$ and $(y) + \mathfrak{a}$ can't both contain a product of nonzero prime ideals. Therefore any ideal not containing a product of nonzero prime ideals is strictly contained in another ideal not containing a product of nonzero prime ideals. Repeating this over and over leads to a strictly ascending chain of ideals

$$\mathfrak{a} = \mathfrak{a}_1 \subsetneqq \mathfrak{a}_2 \subsetneqq \mathfrak{a}_3 \subsetneqq \cdots,$$

but this is false in a Noetherian ring, by Lemma 4.77. So there are no counterexamples. ■

Remark 4.79. In the proof we used the existence of maximal ideals, which is established in general nonzero commutative rings using Zorn's lemma, but for Noetherian rings there is an alternate argument: any proper non-maximal ideal lies inside another proper ideal, which if not maximal lies in yet another proper ideal, and so on. This chain has to stabilize in a proper ideal contained inside no other proper ideal, and that's a maximal ideal.

Lemma 4.80. *Let A be a Noetherian domain in which all nonzero prime ideals are maximal. For a nonzero prime ideal $\mathfrak{p}$ in A, there is some t in the fraction field of A, but not in A, such that $t\mathfrak{p} \subset A$.*

Proof. We proved this at the end of Section 4.2.4 when $A = \mathcal{O}_K$, and the argument goes through with no changes. Let's review it. Pick nonzero $\alpha \in \mathfrak{p}$ and by Lemma 4.78 let $(\alpha) \supset \mathfrak{p}_1 \cdots \mathfrak{p}_n$ for some nonzero prime ideals $\mathfrak{p}_i$, with n chosen as small as possible. Since $\mathfrak{p} \supset (\alpha)$ and the $\mathfrak{p}_i$'s are all maximal, we may take $\mathfrak{p}_1 = \mathfrak{p}$. If $n = 1$ then $\mathfrak{p} = (\alpha)$ and we can use $t = 1/\alpha$. If $n > 1$, then $(\alpha) \not\supset \mathfrak{p}_2 \cdots \mathfrak{p}_n$. Choosing any $\beta \in \mathfrak{p}_2 \cdots \mathfrak{p}_n - (\alpha)$, we can use $t = \beta/\alpha$. ■

Theorem 4.81. *Every Dedekind domain has unique factorization of ideals.*

Proof. Let A be a Dedekind domain with fraction field F. Just as in number fields, we can define multiplication of A-modules M and N in F:

$$MN = \left\{ \sum_{i=1}^{r} x_i y_i : r \geqslant 1,\ x_i \in M,\ y_i \in N \right\}.$$

This includes multiplication of ideals in A as a special case, but within the larger framework of (nonzero) A-modules in F we will pick up inverses.

Step 1: Every nonzero prime ideal $\mathfrak{p}$ has an A-module inverse in F.

Set

$$\widetilde{\mathfrak{p}} = \{x \in F : x\mathfrak{p} \subset A\}.$$

Then $A \subset \widetilde{\mathfrak{p}}$, so $\mathfrak{p} \subset \mathfrak{p}\widetilde{\mathfrak{p}} \subset A$. Thus $\mathfrak{p}\widetilde{\mathfrak{p}} = \mathfrak{p}$ or A by the $\boxed{\text{maximality of } \mathfrak{p}.}$ We want to show $\mathfrak{p}\widetilde{\mathfrak{p}} = A$. Suppose $\mathfrak{p}\widetilde{\mathfrak{p}} = \mathfrak{p}$. For each $x \in \widetilde{\mathfrak{p}}$, $x\mathfrak{p} \subset \mathfrak{p}$. Since $\boxed{A \text{ is a Noetherian ring,}}$ $\mathfrak{p}$ is a finitely generated A-module. So x is integral over A,[5] and since $\boxed{A \text{ is integrally closed}}$ we get $x \in A$. Thus $\widetilde{\mathfrak{p}} \subset A$. But by

[5] This follows by the same method used to prove an algebraic number is integral if it lies inside a ring that is a finitely generated $\mathbf{Z}$-module (Theorem 1.15).

Lemma 4.80 there is some $t \in \widetilde{\mathfrak{p}}$ which is not in A, so we have a contradiction. Thus $\mathfrak{p}\widetilde{\mathfrak{p}} = A$ and we used all three defining conditions of a Dedekind domain in this first step.

A consequence of this is that if $\mathfrak{a} \subset \mathfrak{p}$ with $\mathfrak{a} \neq (0)$ and $\mathfrak{p}$ prime, then $\mathfrak{p} \mid \mathfrak{a}$ as ideals. Indeed, if $\mathfrak{a} \subset \mathfrak{p}$, then $\widetilde{\mathfrak{p}}\mathfrak{a} \subset \widetilde{\mathfrak{p}}\mathfrak{p} = A$, so $\mathfrak{b} := \widetilde{\mathfrak{p}}\mathfrak{a}$ is an ideal in A (it's an A-module in A) and $\mathfrak{p}\mathfrak{b} = \mathfrak{a}$. From $\mathfrak{p}\mathfrak{b} = \mathfrak{a}$ let's show $\mathfrak{a} \subsetneqq \mathfrak{b}$. We just need to show $\mathfrak{a} \neq \mathfrak{b}$ (since obviously $\mathfrak{a} \subset \mathfrak{b}$). The argument made in the $\mathcal{O}_K$-case, which started by writing $\mathfrak{a} = \mathfrak{b}$ as $\mathfrak{p}\mathfrak{b} = \mathfrak{b}$, depended on nonzero ideals having finite index,[6] and that is not usually true in A. Instead, we write $\mathfrak{a} = \mathfrak{b}$ as $\mathfrak{a} = \widetilde{\mathfrak{p}}\mathfrak{a}$. Then for any $x \in \widetilde{\mathfrak{p}}$, $x\mathfrak{a} \subset \mathfrak{a}$, and since $\mathfrak{a}$ is a finitely generated A-module (A is a Noetherian ring) we get that x is integral over A, so $x \in A$ since A is integrally closed. Thus $\widetilde{\mathfrak{p}} \subset A$, which is a contradiction! So $\mathfrak{p}\mathfrak{b} \subsetneqq \mathfrak{b}$.

Step 2: Existence of prime ideal factorization.

Suppose some nonzero proper ideal $\mathfrak{a}$ has no prime ideal factorization. We have $\mathfrak{a} \subset \mathfrak{p}$ for some maximal ideal $\mathfrak{p}$. Then $\mathfrak{p} \mid \mathfrak{a}$, so $\mathfrak{a} = \mathfrak{p}\mathfrak{a}'$ for some ideal $\mathfrak{a}'$, and $\mathfrak{a} \subsetneqq \mathfrak{a}'$. The ideal $\mathfrak{a}'$ is not the unit ideal (otherwise $\mathfrak{a} = \mathfrak{p}$ would be a prime ideal, which it isn't) and $\mathfrak{a}'$ can't have a prime ideal factorization since then $\mathfrak{a}$ would, by the equation $\mathfrak{a} = \mathfrak{p}\mathfrak{a}'$. We've shown any nonzero proper ideal that has no prime ideal factorization is strictly contained in another nonzero proper ideal that has no prime ideal factorization. Thus we get an infinite strictly ascending chain of ideals

$$\mathfrak{a} \subsetneqq \mathfrak{a}' \subsetneqq \mathfrak{a}'' \subsetneqq \cdots \subsetneqq A,$$

which is impossible because A is a Noetherian ring (Lemma 4.77).

Step 3: Uniqueness of prime ideal factorization.

This is shown just like the $\mathcal{O}_K$-case. It depends on all nonzero prime ideals being maximal and existence of inverses for nonzero prime ideals (Step 1). ■

Theorem 4.81 admits a converse: any domain which has unique factorization of ideals and is not a field[7] is a Dedekind domain.[8] Dedekind himself worked with domains defined directly as those having unique factorization of ideals, and Noether found that the three properties we used to define Dedekind domains are an equivalent set of conditions.

[6] When $A = \mathcal{O}_K$, write $\mathfrak{a} = \mathfrak{b}$ as $\mathfrak{p}\mathfrak{b} = \mathfrak{b}$ to get $\mathfrak{p}^k\mathfrak{b} = \mathfrak{b}$ for all k, so $\mathfrak{b} \subset \mathfrak{p}^k$ for all k and by $\mathcal{O}_K \supset \mathfrak{p} \supset \cdots \supset \mathfrak{p}^k \supset \mathfrak{b}$ we get $\mathfrak{p}^{k+1} = \mathfrak{p}^k$ for some k since $\mathfrak{b}$ has finite index in $\mathcal{O}_K$. So $\mathfrak{p} = \mathcal{O}_K$, which is a contradiction.

[7] A field has a vacuous unique factorization of ideals.

[8] The existence of prime ideal factorization already forces a domain to be Dedekind. See [17, Theorem 15, p. 765].

Corollary 4.82. *Let A be Dedekind with fraction field F and E/F be a finite separable extension. Then the integral closure of A in E is Dedekind. In particular, the integral closure of a* PID *in a finite separable extension of its fraction field is a Dedekind domain.*

Proof. Let B be the integral closure of A in E, so B is integrally closed and $\dim A = 1$ implies that $\dim B = 1$ (Theorem 2.20). By Corollary 4.69, B is a Noetherian ring (here we use E/F separable). ■

Example 4.83. The integral closure of $\mathbf{C}[X]$ in any finite extension of $\mathbf{C}(X)$ is a Dedekind domain. For example, the integral closure of $\mathbf{C}[X]$ in $\mathbf{C}(X, \sqrt{X^3 - X})$ is $\mathbf{C}[X, \sqrt{X^3 - X}]$. This ring is Dedekind and it is not a UFD.

The ring $\mathbf{C}[X, \sqrt{X^3 - X}]$ is closely associated with the plane curve $y^2 = x^3 - x$. Integral closures of $\mathbf{C}[X]$ in finite extensions of $\mathbf{C}(X)$ are geometric examples of Dedekind domains.

Nonexample 4.84. The ring of *all* algebraic integers $\overline{\mathbf{Z}}$ inside an algebraic closure $\overline{\mathbf{Q}}$ of the rationals is not a Dedekind domain. It is integrally closed and it has dimension 1 (it is integral over $\mathbf{Z}$, which is 1-dimensional), but it is not Noetherian: the ideal $(2, \sqrt{2}, \sqrt[4]{2}, \sqrt[8]{2}, \dots)$ is not finitely generated.

Remark 4.85. Corollary 4.82 is true if E/F is inseparable too: B is Dedekind when E/F is any finite extension. However, B may *not* be finitely generated as an A-module, so in the course of showing B is Dedekind a new idea is needed to show B is Noetherian. See [27, pp. 633–634] for a proof.

Lemma 4.86. *In a Noetherian domain, any nonzero nonunit is a product of irreducible elements.*

Proof. Let A be a Noetherian domain and we may assume it is not a field (otherwise the lemma is vacuous for A).

Our argument is by contradiction. Suppose there is a nonzero nonunit $\alpha \in A$ which is not a product of irreducible elements. Then α is not irreducible, so $\alpha = \beta\gamma$ where β and γ are (nonzero) nonunits in A. At least one of β or γ is not a product of irreducibles (otherwise α is). Suppose β is not. Since γ is not a unit, $(\alpha) \subsetneqq (\beta)$. Thus any principal ideal in A generated by a nonzero nonunit which is not a product of irreducibles is strictly contained in a principal ideal of the same kind. Repeating this construction indefinitely, we get an infinite strictly ascending chain of ideals, which is impossible in a Noetherian ring. ■

Theorem 4.87. *A Dedekind domain is a unique factorization domain if and only if it is a principal ideal domain.*

Proof. It is a general theorem of algebra that any PID is a UFD. We now show for a Dedekind domain A that if A is a UFD then it is a PID. It suffices to show every prime ideal in A is principal, since any nonzero ideal in A is a product of prime ideals.

For any irreducible π in a UFD, the ideal (π) is prime. Now choose any nonzero prime ideal $\mathfrak{p}$ and pick $\alpha \in \mathfrak{p}$ with $\alpha \neq 0$. Since $(\alpha) \subset \mathfrak{p}$, $\mathfrak{p} \mid (\alpha)$ as ideals. By Lemma 4.86, there is an irreducible factorization of α in A, say $\alpha = \pi_1 \cdots \pi_r$ with the π_i's being irreducible in A. Then $(\alpha) = (\pi_1) \cdots (\pi_r)$ as ideals. Each of the ideals (π_i) is prime, and $\mathfrak{p} \mid (\alpha)$, so by unique prime ideal factorization in A we must have $\mathfrak{p} = (\pi_i)$ for some i. Thus all prime ideals are principal, so A is a PID. ■

UFDs and Dedekind domains each have a unique factorization property, of elements or ideals, and neither class of rings includes the other ($\mathbf{Z}[T]$ is a UFD which is not Dedekind, since (2) is a prime ideal in $\mathbf{Z}[T]$ that is not maximal, and $\mathbf{Z}[\sqrt{-5}]$ is Dedekind but not a UFD). There is a class of domains with a unique factorization-like property, called *Krull rings*, which includes UFDs and Dedekind domains as special cases. We don't give the definition here. See [6, Chap. 3] and [39, Chap. 12] for the definition and some basic properties (in [6] they are called "rings with a theory of divisors.") The overall picture is in Figure 4.4. The intersection of UFDs and Dedekind domains are PIDs which are not fields. Dedekind domains are the 1-dimensional Krull rings. Every Noetherian integrally closed domain is Krull.

If A is Krull then it can be shown that the integral closure of A in any finite extension of its fraction field (separable or inseparable) is Krull and $A[T]$ is Krull. In contrast, the integral closure of a UFD in a finite extension of its fraction field need not be a UFD (*e.g.*, every $\mathcal{O}_K$ is an integral closure of $\mathbf{Z}$ and lots of them are not UFDs) and for any Dedekind domain A, $A[T]$ is not Dedekind (A has dimension 1 while $A[T]$ has dimension 2; see Remark 4.73). So although $\mathcal{O}_K[T]$ is not a Dedekind domain, it is a Krull ring.

As with the rings $\mathcal{O}_K$, for any Dedekind domain A we can find an inverse for any nonzero ideal $\mathfrak{a}$ inside the fraction field F of A:

$$\mathfrak{a}^{-1} = \{x \in F : x\mathfrak{a} \subset A\}.$$

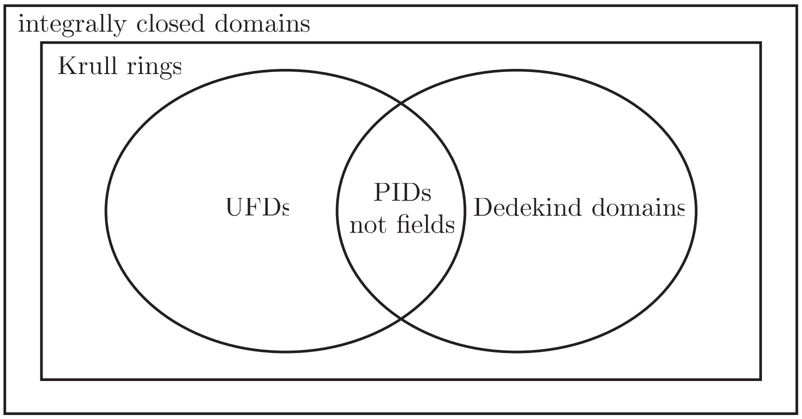


Figure 4.4: Beyond UFDs and Dedekind domains.

This is an A-module in F and verifying $\mathfrak{a}\mathfrak{a}^{-1} = A$ proceeds just as at the start of Section 4.3 for ideals in $\mathcal{O}_K$. Defining $\mathfrak{a}^n$ for $n < 0$ to be $(\mathfrak{a}^{-1})^{|n|}$, just as in $\mathcal{O}_K$, these integral powers of ideals satisfy (4.1). A lot of other properties of nonzero ideals in $\mathcal{O}_K$ carry over to Dedekind domains with the same proofs:

1. $\mathfrak{a} \supset \mathfrak{b}$ if and only if $\mathfrak{a} \mid \mathfrak{b}$. (Theorem 4.23)
2. Every nonzero ideal in A has a nonzero principal ideal multiple. (Corollary 4.25)
3. For a nonzero prime ideal $\mathfrak{p}$, $\mathfrak{a}/\mathfrak{a}\mathfrak{p} \cong A/\mathfrak{p}$ as A-modules. (Corollary 4.27)
4. $\gcd(\mathfrak{a}, \mathfrak{b}) = \mathfrak{a} + \mathfrak{b}$ and $\operatorname{lcm}(\mathfrak{a}, \mathfrak{b}) = \mathfrak{a} \cap \mathfrak{b}$. (Corollary 4.30)

As in the case of $\mathcal{O}_K$, we say ideals $\mathfrak{a}$ and $\mathfrak{b}$ in a Dedekind domain A are relatively prime when $\mathfrak{a} + \mathfrak{b} = A$. That every nonzero ideal in A divides a nonzero principal ideal is a close relation between all ideals in A and principal ideals. Here is another link between all ideals and principal ideals.

Theorem 4.88. *Every nonzero ideal $\mathfrak{a}$ in a Dedekind domain A needs at most two generators: $\mathfrak{a} = (x, y)$ where $x \in \mathfrak{a} - \{0\}$ is arbitrary. Equivalently, $\mathfrak{a}$ is the greatest common divisor of two principal ideals.*

Proof. We can assume that $\mathfrak{a}$ is a proper ideal, as otherwise we can choose $y = 1$. Picking any $x \in \mathfrak{a} - \{0\}$, we have $(x) \subset \mathfrak{a}$, so $\mathfrak{a} \mid (x)$. We seek $y \in A - \{0\}$ such that

$$\mathfrak{a} = (x, y) = (x) + (y) = \gcd((x), (y)).$$

Factor $\mathfrak{a}$ and (x):

$$\mathfrak{a} = \mathfrak{p}_1^{a_1} \cdots \mathfrak{p}_r^{a_r} \qquad \text{and} \qquad (x) = \mathfrak{p}_1^{b_1} \cdots \mathfrak{p}_r^{b_r}, \quad 0 \leqslant a_i \leqslant b_i.$$

We will find y so that

$$(y) = \mathfrak{p}_1^{a_1} \cdots \mathfrak{p}_r^{a_r} \mathfrak{b}, \quad \mathfrak{p}_i \nmid \mathfrak{b} \text{ for all } i.$$

Then $\gcd((x),(y)) = \mathfrak{a}$.

First we find, for each $\mathfrak{p}_i$, an element of A whose principal ideal is divisible by $\mathfrak{p}_i^{a_i}$ and not a higher power of $\mathfrak{p}_i$. Pick any $y_i \in \mathfrak{p}_i^{a_i} - \mathfrak{p}_i^{a_i+1}$. Then $y_i \in \mathfrak{p}_i^{a_i}$ and $y_i \notin \mathfrak{p}_i^{a_i+1}$, so $\mathfrak{p}_i^{a_i} \mid (y_i)$ and $\mathfrak{p}_i^{a_1+1} \nmid (y_i)$. That means the exact power of $\mathfrak{p}_i$ in (y_i) is $\mathfrak{p}_i^{a_i}$. Now we glue this prime-power data together with the Chinese remainder theorem: there is $y \in A$ such that

$$y \equiv y_i \bmod \mathfrak{p}_i^{a_i+1} \text{ for all } i.$$

Then $y = y_i + z_i$ where $z_i \in \mathfrak{p}_i^{a_i+1}$. Since y_i is in $\mathfrak{p}_i^{a_i}$ but not in $\mathfrak{p}_i^{a_i+1}$, while z_i is in both, we get $y \in \mathfrak{p}_i^{a_i}$ and $y \notin \mathfrak{p}_i^{a_i+1}$, so the exact power of $\mathfrak{p}_i$ in (y) is $\mathfrak{p}_i^{a_i}$ for all i. There might be other prime ideals dividing (y), but they don't divide (x) so they don't divide the greatest common divisor of (x) and (y). ■

Theorem 4.88 says every nonzero ideal $\mathfrak{a}$ in a Dedekind domain is the greatest common divisor of two nonzero principal ideals, with one of the two principal ideals being arbitrary inside $\mathfrak{a}$. In the case of ideals in $\mathcal{O}_K$, this reveals a substantial difference between a nonzero ideal as an $\mathcal{O}_K$-module and as an abelian group ($\mathbf{Z}$-module): it has two generators as an $\mathcal{O}_K$-module but requires $n = [K : \mathbf{Q}]$ generators as a $\mathbf{Z}$-module. (For quadratic K, we don't see much of a difference in this aspect, but it is false that generators of an ideal as an $\mathcal{O}_K$-module are also generators of the ideal as a $\mathbf{Z}$-module, even though the number of generators needed in both cases is 2. For instance, in $\mathbf{Z}[\sqrt{-5}]$ the ideal $(7, 2+3\sqrt{-5})$ does not have $\mathbf{Z}$-basis $\{7, 2+3\sqrt{-5}\}$ since the $\mathbf{Z}$-span of that pair doesn't contain $7\sqrt{-5}$.)

Since a nonzero prime ideal $\mathfrak{p}$ in $\mathcal{O}_K$ has $\mathfrak{p} \cap \mathbf{Z} = p\mathbf{Z}$ for some prime number p, Theorem 4.88 says we can use p as one of the two generators:

$$\mathfrak{p} = (p, \alpha)$$

for some $\alpha \in \mathcal{O}_K$. We saw some prime ideals in $\mathbf{Z}[\sqrt{-5}]$ like this already: $(2, 1+\sqrt{-5}), (3, 1+\sqrt{-5})$, and $(3, 1-\sqrt{-5})$.

Every nonzero ideal $\mathfrak{a}$ in a Dedekind domain has a nonzero principal multiple: $\mathfrak{a}\mathfrak{b}$ is principal for some $\mathfrak{b} \neq (0)$. Now we will refine this, as an application of the Chinese remainder theorem.

Theorem 4.89. *Given nonzero ideals $\mathfrak{a}$ and $\mathfrak{a}'$ in a Dedekind domain, $\mathfrak{a}$ has a principal multiple $\mathfrak{a}\mathfrak{b}$ where $(\mathfrak{b}, \mathfrak{a}') = (1)$.*

Proof. Let $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ be all the prime factors appearing in $\mathfrak{a}$ or $\mathfrak{a}'$. Write

$$\mathfrak{a} = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}, \qquad e_i \geqslant 0.$$

Pick $x_i \in \mathfrak{p}_i^{e_i} - \mathfrak{p}_i^{e_i+1}$, so $\mathfrak{p}_i^{e_i} \mid (x_i)$ and $\mathfrak{p}_i^{e_i+1} \nmid (x_i)$. Thus $\mathfrak{p}_i^{e_i} || (x_i)$. (The double bars mean this power of $\mathfrak{p}_i$ is a factor and no higher one.) By the Chinese remainder theorem, there exists x such that

$$x \equiv x_i \bmod \mathfrak{p}_i^{e_i+1}$$

for all i. So $x \in \mathfrak{p}_i^{e_i} - \mathfrak{p}_i^{e_i+1}$, which means $\mathfrak{p}_i^{e_i} || (x)$ for all i. So the prime factorization of (x) is

$$(x) = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r} \mathfrak{b} = \mathfrak{a}\mathfrak{b}$$

for some ideal $\mathfrak{b}$ not divisible by any $\mathfrak{p}_i$. Hence $(\mathfrak{b}, \mathfrak{a}') = (1)$ (and $(\mathfrak{b}, \mathfrak{a}) = 1$). ■

Corollary 4.90. *For nonzero ideals $\mathfrak{a}$ and $\mathfrak{b}$ in a Dedekind domain A, $\mathfrak{a}/\mathfrak{a}\mathfrak{b} \cong A/\mathfrak{b}$ as A-modules.*

This generalizes the isomorphism $\mathfrak{a}/\mathfrak{a}\mathfrak{p} \cong A/\mathfrak{p}$ for prime $\mathfrak{p}$ (essentially Corollary 4.27).

Proof. By Theorem 4.89 there is an ideal $\mathfrak{c}$ relatively prime to $\mathfrak{b}$ such that $\mathfrak{a}\mathfrak{c} = (x_0)$ is principal. Then $A/\mathfrak{b} \cong \mathfrak{a}/\mathfrak{a}\mathfrak{b}$ by $a \bmod \mathfrak{b} \mapsto ax_0 \bmod \mathfrak{a}\mathfrak{b}$ (check!). ■

The concept of fractional ideal in a number field generalizes to all domains.

Definition 4.91. Let A be a domain with fraction field F. A *fractional A-ideal* in F is a nonzero A-module $\mathfrak{a} \subset F$ admitting a common denominator: there is some nonzero $d \in A$ such that $d\mathfrak{a} \subset A$. A fractional ideal is called *principal* when it has the form Ax for some $x \in F^\times$.

We include "A" in the label "fractional A-ideal" because F is the fraction field of other subrings, which have their own corresponding concept of fractional ideal that may not match those for A. For example, in $\mathbf{Q}(i)$, $\mathbf{Z}[2i] = \mathbf{Z} + \mathbf{Z} \cdot 2i$ is trivially a fractional $\mathbf{Z}[2i]$-ideal but it is not a fractional $\mathbf{Z}[i]$-ideal since it is not even a $\mathbf{Z}[i]$-module (it contains $\mathbf{Z}$ but not $\mathbf{Z}i$). In a number field K, where $\mathcal{O}_K$ is a canonical subring, the fractional $\mathcal{O}_K$-ideals are what we called before the fractional ideals of K. Fractional ideals in a number field relative to subrings other than the ring of integers need to have the subring specified.

Any nonzero ideal in a domain A is a fractional A-ideal with common denominator $d = 1$. As in the number field case, a nonzero ideal in A is called an integral ideal if we need to distinguish it from the broader class of fractional A-ideals.

Theorem 4.22 generalizes to fractional A-ideals provided A is Noetherian.

Theorem 4.92. *When A is a Noetherian domain, the fractional A-ideals are the nonzero finitely generated A-modules in F.*

Proof. That any nonzero finitely generated A-module in F is a fractional A-ideal is shown just as in the proof of Theorem 4.22, using a common denominator for the finite generating set. For the converse direction, any fractional A-ideal has the form $\frac{1}{d}I$ for some $d \in A - \{0\}$ and ideal I. Since A is Noetherian, I is a finitely generated A-module, so $\frac{1}{d}I$ is a finitely generated A-module. ■

For us, the most interesting setting for fractional A-ideals is when A is a Dedekind domain. In that case, all the basic properties we have seen so far of fractional ideals in a number field generalize to fractional A-ideals:

1. For any nonzero ideal $\mathfrak{a}$ in A, its inverse $\mathfrak{a}^{-1}$ is a fractional A-ideal.

2. The fractional A-ideals are the nonzero finitely generated A-modules in F (even just when A is Noetherian, by Theorem 4.92).

3. The fractional A-ideals are the group generated by the nonzero prime ideals of A:

$$\{\mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r} : \mathfrak{p}_i \text{ prime}, e_i \in \mathbf{Z}\} = \left\{IJ^{-1} : I, J \subset A \text{ nonzero ideals}\right\}.$$

Since ideals in a Dedekind domain A only need 2 generators (Theorem 4.88), a fractional A-ideal only needs 2 generators as an A-module:

$$I = Ax + Ay \text{ for some } x, y \in A \Longrightarrow \frac{1}{d}I = A\frac{x}{d} + A\frac{y}{d}.$$

4.9 Exercises

1. Let A be a domain with fraction field F.

   a) Show multiplication of A-modules in F is commutative, associative, and distributes over addition of A-modules in F.

   b) If $x \in F$ and M is an A-module in F, show $xA \cdot M = xM$.

   c) Show the product of two finitely generated A-modules in F is a finitely generated A-module.

2. Show the following equations are examples of nonunique irreducible factorization and then find prime ideal factorizations of the principal ideals generated by factors on both sides.

   a) $3 \cdot 3 \cdot 3 \cdot 3 = (5 + 2\sqrt{-14})(5 - 2\sqrt{-14})$ in $\mathbf{Z}[\sqrt{-14}]$.

   b) $2 \cdot 3 = (\frac{1+\sqrt{-23}}{2})(\frac{1-\sqrt{-23}}{2})$ in $\mathbf{Z}[\frac{1+\sqrt{-23}}{2}]$.

   c) $3 \cdot 3 \cdot 3 = (2 + \sqrt{-23})(2 - \sqrt{-23})$ in $\mathbf{Z}[\frac{1+\sqrt{-23}}{2}]$.

   d) $3 \cdot 3 \cdot 3 = (1 + \sqrt{-26})(1 - \sqrt{-26})$ in $\mathbf{Z}[\sqrt{-26}]$.

   e) $5 \cdot 5 \cdot 5 = (6 + \sqrt{-89})(6 - \sqrt{-89})$ in $\mathbf{Z}[\sqrt{-89}]$.

3. Let a and b be relatively prime nonzero integers such that ab is not a square.

   a) In the ring $\mathbf{Z}[\sqrt{ab}]$, verify the equality of ideals

   $$(a, \sqrt{ab})^2 = (a).$$

   In particular, when d is a squarefree integer and p is a prime dividing d, so d/p is relatively prime to p, the ideal (p) in $\mathbf{Z}[\sqrt{d}]$ is the square of an ideal.

   b) For instance, in $\mathbf{Z}[\sqrt{6}]$, $(2) = (2, \sqrt{6})^2$. Is the ideal $(2, \sqrt{6})$ principal? (Warning: Just because $\sqrt{2} \notin \mathbf{Z}[\sqrt{6}]$ does not mean there can't be a principal ideal which squares to (2). What we would need is some $\alpha \in \mathbf{Z}[\sqrt{6}]$ such that α^2 equals 2 *up to a unit multiple*.)

4. In the table below, the primes p such that $T^2 - 2 \bmod p$ is reducible are listed. For each of these primes, factor $p\mathbf{Z}[\sqrt{2}]$ explicitly into a product of two *principal* prime ideals. (Since $\mathbf{Z}[\sqrt{2}]$ is a Euclidean domain, all of its ideals are principal.)

p	$T^2 - 2 \bmod p$	$p\mathbf{Z}[\sqrt{2}]$
2	T^2	
7	$(T+3)(T-3)$	
17	$(T+6)(T-6)$	
23	$(T+5)(T-5)$	
31	$(T+8)(T-8)$	
41	$(T+17)(T-17)$	
47	$(T+7)(T-7)$	

5. Describe the prime ideal factorization in $\mathbf{Z}[\sqrt{-6}]$ of all prime numbers less than 20, not just in terms of the shape of the factorization but also giving explicit generators for each prime ideal that appears.

6. Find all primes $p \leqslant 23$ such that (p) is prime in $\mathbf{Z}[\sqrt{79}]$.

7. In the ring of integers of $\mathbf{Q}(\sqrt{-15})$, show the ideal (2) is not prime. Then describe the factorization of (2) into prime ideals with explicit generators.

8. Verify by explicit calculation the ideal factorization

$$(5) = (5, \sqrt[3]{2} - 3)(5, \sqrt[3]{4} + 3\sqrt[3]{2} + 4)$$

in $\mathbf{Z}[\sqrt[3]{2}]$ and that both ideals on the right are prime (check the residue rings are fields of size 5 and 25).

9. In $\mathcal{O}_K$, suppose there is a prime number p such that the ideal (p) has more than p prime ideal factors with norm p. Show $\mathcal{O}_K \neq \mathbf{Z}[\alpha]$ for any α. (Hint: First assume $(p) = \mathfrak{p}_1 \cdots \mathfrak{p}_g$ where $g > p$ and every $\mathfrak{p}_i$ has norm p, so $\mathcal{O}_K/(p) \cong \prod_{i=1}^{g} \mathcal{O}_K/\mathfrak{p}_i \cong \mathbf{F}_p^g$. Think about why no $\mathbf{F}_p[T]/(f(T))$ can be isomorphic to the product ring $\mathbf{F}_p^g$. The main ideas for the general case are mostly visible in this special case.)

10. Let $K = \mathbf{Q}(\gamma)$, where $\gamma^3 - \gamma^2 - 2\gamma - 8 = 0$. (This is Dedekind's field.)

a) For any nonzero prime $\mathfrak{p}$ in $\mathcal{O}_K$ with $\mathfrak{p} \mid (2)$, show $\mathfrak{p} \mid (\gamma)$ or $\mathfrak{p} \mid (\gamma - 1)$, but not both.

b) Compute $\mathrm{N}_{K/\mathbf{Q}}(\gamma + c)$ for $c \in \mathbf{Z}$ and use this to factor $(\gamma - 1)$ into prime ideals. (Specify the norm of each prime.)

c) Use parts a and b to show the ideal (2) must have at least two prime factors which do not divide $(\gamma - 1)$, and therefore $(2) = \mathfrak{p}_2\mathfrak{p}_2'\mathfrak{p}_2''$ with the prime factors all distinct. Conclude from Exercise 4.9 that $\mathcal{O}_K \neq \mathbf{Z}[\alpha]$ for any $\alpha \in \mathcal{O}_K$. (Hint: Think about prime ideal factors of (γ), $(\gamma - 1)$, and $(\gamma - 2)$.)

d) Factor the ideals (γ), $(\gamma + 1)$, $(\gamma + 2)$, $(\gamma - 2)$, $(\gamma + 3)$, and $(\gamma - 3)$ into primes, specifying the norm of each prime that appears.

11. Let $K = \mathbf{Q}(\gamma)$ where $\gamma^3 - 2\gamma^2 - 9\gamma + 2 = 0$. The polynomial $T^3 - 2T^2 - 9T + 2$ has discriminant $2^2 \cdot 31^2$.

    a) Find a $\mathbf{Z}$-basis for $\mathcal{O}_K$ and compute $\mathrm{disc}(K)$.

    b) Show $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ is even for all $\alpha \in \mathcal{O}_K - \mathbf{Z}$. (Therefore $\mathcal{O}_K \neq \mathbf{Z}[\alpha]$.)

    c) Show $2\mathcal{O}_K = \mathfrak{p}\mathfrak{p}'\mathfrak{p}''$, where each prime ideal factor has norm 2. (Therefore $\mathcal{O}_K \neq \mathbf{Z}[\alpha]$ for any α by Exercise 4.9.)

12. Let $K = \mathbf{Q}(\alpha)$ where $\alpha^3 + 2\alpha + 22 = 0$.

    a) Show $\mathrm{disc}(\mathbf{Z}[\alpha]) = -2^2 \cdot 5^2 \cdot 131$.

    b) Show $2 \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$, so $4 \mid \mathrm{disc}(K)$.

    c) Verify $\beta := \frac{1}{5}(3 + \alpha + \alpha^2)$ is an algebraic integer and $\mathrm{disc}(\mathbf{Z}[\beta]) = -2^2 \cdot 131$.

    d) Show $\mathcal{O}_K = \mathbf{Z}[\beta]$.

    e) Show $5\mathcal{O}_K = \mathfrak{p}\mathfrak{p}'\mathfrak{p}''$, where each prime ideal factor has norm 5. Compare this with the factorization of $T^3 + 2T + 22 \bmod 5$.

13. For $i = 1, 2, 3, 4$, define four cubic fields $K_i = \mathbf{Q}(\alpha_i)$ where α_i is the root of $f_i(T)$:

$$\begin{aligned} f_1(T) &= T^3 - T^2 - 20T - 1, \\ f_2(T) &= T^3 - T^2 - 34T - 24, \\ f_3(T) &= T^3 - T^2 - 52T + 159, \\ f_4(T) &= T^3 - 41T - 95. \end{aligned}$$

    a) Show all four polynomials are irreducible over $\mathbf{Q}$ with three real roots.

b) Show all four number fields have prime discriminant 32009.

c) Find a basis for the ring of integers of each field.

d) Prove the fields are nonisomorphic by finding prime numbers that factor in different ways in each pair of fields.

14. Let $K = \mathbf{Q}(\alpha)$ where $\alpha^4 - 22\alpha - 6 = 0$.

    a) Show $\operatorname{disc}(\mathbf{Z}[\alpha]) = -2^4 \cdot 3^6 \cdot 547$.

    b) Verify $\beta := \frac{1}{3}(\alpha^2 - \alpha)$ is an algebraic integer and $\operatorname{disc}(\mathbf{Z}[\beta]) = -2^4 \cdot 17^2 \cdot 547$.

    c) Show $3\mathcal{O}_K = \mathfrak{p}_4\mathfrak{p}_3'\mathfrak{p}_9$. Compare with the factorization of $T^4 - 22T - 6 \bmod 3$.

15. In the integers of $\mathbf{Q}(\sqrt[p]{p})$, where p is a prime number, explain why the shape of the factorization of any prime $q \neq p$ can be determined from the way $T^p - p \bmod q$ factors. What can you say for $q = p$?

16. Let $\mathfrak{p}_7 = (7, 3 + \sqrt{-5})$ in $\mathbf{Z}[\sqrt{-5}]$. In $\mathbf{Z}[\sqrt{-5}]/\mathfrak{p}_7 \cong \mathbf{F}_7$, find $a \in \mathbf{Z}$ such that $3 + 2\sqrt{-5} \equiv a(1 + \sqrt{-5}) \bmod \mathfrak{p}_7$.

17. In $\mathbf{Z}[\sqrt{-5}]$, the prime ideals $\mathfrak{p}_3 = (3, 1 + \sqrt{-5})$, $\mathfrak{p}_7 = (7, 3 + \sqrt{-5})$, and (11) have norms 3, 7, and 121.

    a) Find an $\alpha \in \mathbf{Z}[\sqrt{-5}]$ satisfying $\alpha \equiv 1 \bmod \mathfrak{p}_3$ and $\alpha \equiv 2 \bmod \mathfrak{p}_7$. Can you choose $\alpha \in \mathbf{Z}$?

    b) Find an $\alpha \in \mathbf{Z}[\sqrt{-5}]$ satisfying $\alpha \equiv 2 \bmod \mathfrak{p}_3$ and $\alpha \equiv \sqrt{-5} \bmod (11)$. Can you choose $\alpha \in \mathbf{Z}$?

18. In $\mathbf{Z}[\sqrt{82}]$, show $33 - 5\sqrt{82} \bmod \mathfrak{p}$ is not a perfect square when $\mathfrak{p}$ is a prime dividing (3) and $(33 - 5\sqrt{82})(9 + \sqrt{82}) \bmod \mathfrak{q}$ is not a perfect square when $\mathfrak{q}$ is a prime dividing (13).

19. a) In $K = \mathbf{Q}(\sqrt{7})$, show the ideal $\mathfrak{p} = (5)$ is prime in $\mathcal{O}_K = \mathbf{Z}[\sqrt{7}]$ and find a generator of the group $(\mathcal{O}_K/\mathfrak{p})^\times$.

    b) Compute the size of the group $(\mathbf{Z}[i]/9\mathbf{Z}[i])^\times$ and show it is not cyclic.

20. Let $K = \mathbf{Q}(\alpha)$ with $\alpha \in \mathcal{O}_K$. Suppose $\mathrm{N}_{K/\mathbf{Q}}(a + b\alpha)$ has a prime factor p, so $(a + b\alpha)$ has a prime ideal factor $\mathfrak{p}$ where $\mathfrak{p} \mid (p)$. If $(a, b) = 1$ and $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$ (for instance, maybe $\mathcal{O}_K = \mathbf{Z}[\alpha]$), show $\mathcal{O}_K/\mathfrak{p}$ is represented by integers, so $\mathrm{N}(\mathfrak{p}) = p$.

The point of this exercise is that it shows if you need to find a principal ideal divisible by a prime ideal with norm p^k where $k > 1$, for the most part you will have to look at norms of polynomials in α with degree greater than 1.

21. a) For any prime p, show $\{x \in \mathbf{Z}/p^e\mathbf{Z} : x^2 = x\} = \{0, 1\}$.

    b) Use part a to show $\#\{x \in \mathbf{Z}/m\mathbf{Z} : x^2 = x\} = 2^r$, where r is the number of prime factors of m. So counting solutions to $x^2 \equiv x \bmod m$, in principle, tells us how many prime factors m has.

    c) Find all solutions to $x^2 \equiv x \bmod 15$ and $x^2 \equiv x \bmod 42$.

    d) For any odd prime p, show $\{x \in \mathbf{Z}/p^e\mathbf{Z} : x^2 = 1\} = \{1, -1\}$. What happens when $p = 2$? Convert this into a method of counting prime factors of m when m is odd and apply it to $m = 15$.

22. Let K be a number field and $\mathfrak{p}$ be a nonzero prime ideal in $\mathcal{O}_K$. In the ring $\mathcal{O}_K/\mathfrak{p}^r$, where $r \geqslant 1$, show the only solutions to $x^2 = x$ are 0 and 1.

23. a) Let a and b be positive integers such that $(a, b) \neq 1$. Show the rings $\mathbf{Z}/ab\mathbf{Z}$ and $\mathbf{Z}/a\mathbf{Z} \times \mathbf{Z}/b\mathbf{Z}$ are not isomorphic, although they have the same size. (Hint: Show their unit groups do not have the same size.)

    b) Let K be a number field and $\mathfrak{a}$ and $\mathfrak{b}$ be two nonzero ideals in $\mathcal{O}_K$ that are not relatively prime. Show the rings $\mathcal{O}_K/\mathfrak{a}\mathfrak{b}$ and $\mathcal{O}_K/\mathfrak{a} \times \mathcal{O}_K/\mathfrak{b}$ are not isomorphic, although they have the same size.

24. For any number field K, let's extend some classical irreducibility tests from $\mathbf{Q}[T]$ to $K[T]$. Set

$$f(T) = T^d + a_{d-1}T^{d-1} + \cdots + a_1T + a_0,$$

    a monic polynomial in $\mathcal{O}_K[T]$.

    a) (Gauss's Lemma) If $f(T)$ is irreducible in $\mathcal{O}_K[T]$, prove $f(T)$ is irreducible in $K[T]$.

    b) (Reduction mod $\mathfrak{p}$) If the reduction $f(T) \bmod \mathfrak{p} \in (\mathcal{O}_K/\mathfrak{p})[T]$ is irreducible for some prime $\mathfrak{p}$ of $\mathcal{O}_K$, prove $f(T)$ is irreducible in $K[T]$.

    c) (Eisenstein) If, for some prime $\mathfrak{p}$ of $\mathcal{O}_K$, $a_i \equiv 0 \bmod \mathfrak{p}$ for all i and $a_0 \not\equiv 0 \bmod \mathfrak{p}^2$, prove $f(T)$ is irreducible in $K[T]$.

d) Use parts b or c to show

$$T^3 + \sqrt{-5}T + 1 - 2\sqrt{-5}$$

and

$$T^4 + (3 + \sqrt{-5})T^2 - 2T + 7 - \sqrt{-5}$$

are irreducible in $\mathbf{Q}(\sqrt{-5})[T]$. Show $T^n - (1 + 3\sqrt{-5})$ is irreducible in $\mathbf{Q}(\sqrt{-5})[T]$ for all $n \geqslant 1$.

25. Let K be a quadratic field and $\mathfrak{a}$ be a nonzero ideal in $\mathcal{O}_K$. Set $\overline{\mathfrak{a}}$ to be the conjugate ideal (the ideal whose elements are the Galois conjugates of the elements of $\mathfrak{a}$).

a) If $\mathfrak{a} = (\alpha, \beta)$, show

$$\mathfrak{a}\overline{\mathfrak{a}} = (\mathrm{N}_{K/\mathbf{Q}}(\alpha), \mathrm{Tr}_{K/\mathbf{Q}}(\alpha\overline{\beta}), \mathrm{N}_{K/\mathbf{Q}}(\beta)).$$

In particular, $\mathfrak{a}\overline{\mathfrak{a}}$ is a principal ideal since its generators are in $\mathbf{Z}$. (This may be harder than it looks.)

b) Generalize the formula for $\mathfrak{a}\overline{\mathfrak{a}}$ in part a if $\mathfrak{a} = (\alpha_1, \ldots, \alpha_m)$.

c) For nonsquare $d \equiv 1 \bmod 4$, let $\mathfrak{p} = (2, 1 + \sqrt{d})$ in $\mathbf{Z}[\sqrt{d}]$. Show the formula in part a is false for the ideal $\mathfrak{p}$. (Since $\mathbf{Z}[\sqrt{d}]$ is not the ring of integers of $\mathbf{Q}(\sqrt{d})$, there is no contradiction.)

26. Let $\mathfrak{p}$ be a prime ideal in $\mathcal{O}_K$ lying over the prime number p and pick $\pi \in \mathfrak{p} - \mathfrak{p}^2$, so $\mathfrak{p} \mid (\pi)$ but $\mathfrak{p}^2 \nmid (\pi)$.

a) For $x \in \mathcal{O}_K$ with $x \not\equiv 0 \bmod \mathfrak{p}$ and $r \geqslant 2$, show $(1 + \pi^{r-1}x)^i \equiv 1 + i\pi^{r-1}x \bmod \mathfrak{p}^r$ for all i. In particular, $1 + \pi^{r-1}x$ has order p in $(\mathcal{O}_K/\mathfrak{p}^r)^\times$.

b) If $\mathrm{N}(\mathfrak{p}) > p$, which means $\#(\mathcal{O}_K/\mathfrak{p}) > p$, let $\overline{x}$ and $\overline{y}$ in $\mathcal{O}_K/\mathfrak{p}$ be linearly independent over $\mathbf{F}_p$. Show $1 + \pi^{r-1}x$ and $1 + \pi^{r-1}y$ generate different subgroups of order p in $(\mathcal{O}_K/\mathfrak{p}^r)^\times$, so $(\mathcal{O}_K/\mathfrak{p}^r)^\times$ is noncyclic for all $r \geqslant 2$. (This contrasts with $(\mathbf{Z}/p^r\mathbf{Z})^\times$, which for odd p is cyclic for all $r \geqslant 2$.)

27. Let K be a number field containing the nth roots of unity (that is, a full set of solutions to $z^n = 1$). Let $\mathfrak{p}$ be a prime in $\mathcal{O}_K$ not dividing n.

a) Show different nth roots of unity in K remain different when reduced into $(\mathcal{O}_K/\mathfrak{p})^\times$ and conclude that $\mathrm{N}(\mathfrak{p}) \equiv 1 \bmod n$.

b) If $\alpha \not\equiv 0 \bmod \mathfrak{p}$, show there is a unique nth root of unity $\zeta \in K$ such that $\alpha^{(\mathrm{N}(\mathfrak{p})-1)/n} \equiv \zeta \bmod \mathfrak{p}$. We write $\zeta = (\frac{\alpha}{\mathfrak{p}})_n$, so $(\frac{\alpha}{\mathfrak{p}})_n$ is the unique nth root of unity in K satisfying

$$\alpha^{(\mathrm{N}(\mathfrak{p})-1)/n} \equiv \left(\frac{\alpha}{\mathfrak{p}}\right)_n \bmod \mathfrak{p}.$$

We call $(\frac{\cdot}{\mathfrak{p}})_n$ the *nth power residue symbol.* When $K = \mathbf{Q}$ and $n = 2$ this is the Legendre symbol from elementary number theory.

c) Compute $(\frac{3}{1+2i})_4$ and $(\frac{1+2i}{3})_4$ in $\mathbf{Z}[i]$ and $(\frac{1+\omega}{2})_3$ and $(\frac{2}{5})_3$ in $\mathbf{Z}[\omega]$.

d) Show the power residue symbol has the following properties.

- $(\frac{\alpha}{\mathfrak{p}})_n = 1$ if and only if $\alpha \equiv x^n \bmod \mathfrak{p}$ for some $x \in \mathcal{O}_K$. (Hint: $(\mathcal{O}_K/\mathfrak{p})^\times$ is cyclic.)
- For $\alpha, \beta \not\equiv 0 \bmod \mathfrak{p}$, $(\frac{\alpha\beta}{\mathfrak{p}})_n = (\frac{\alpha}{\mathfrak{p}})_n(\frac{\beta}{\mathfrak{p}})_n$.
- For any nth root of unity ζ in K, $(\frac{\zeta}{\mathfrak{p}})_n = \zeta^{(\mathrm{N}(\mathfrak{p})-1)/n}$ (this is equality in K, not congruence mod $\mathfrak{p}$).
- The order of $(\frac{\alpha}{\mathfrak{p}})_n$ as a root of unity is the order of $\overline{\alpha}$ in $\mathbf{F}_\mathfrak{p}^\times/(\mathbf{F}_\mathfrak{p}^\times)^n$, where $\mathbf{F}_p = \mathcal{O}_K/\mathfrak{p}$.

28. Let M be a $\mathbf{Z}$-lattice in $\mathcal{O}_K$ with basis $\{e_1, \ldots, e_n\}$ and suppose we know some prime p divides $[\mathcal{O}_K : M]$ by finding an element of order p in $\mathcal{O}_K/M$, say $\overline{x}$. We want to find a lattice in $\mathcal{O}_K$ containing M with index p.

    Write
    $$px = a_1e_1 + \cdots + a_ne_n$$
    where the a_i's are in $\mathbf{Z}$ and at least one is not in $p\mathbf{Z}$. If $a_{i_0} \not\equiv 0 \bmod p$, show there is an x such that $a_{i_0} = 1$ and the lattice $M' = \sum_{i \neq i_0} \mathbf{Z}e_i + \mathbf{Z}x$ satisfies $[M' : M] = p$.

29. Let F be a field. For any nonconstant rational function f/g in $F(X)$, where f and g are in $F[X]$ and are relatively prime, Lüroth's theorem says $[F(X) : F(f/g)] = \max(\deg f, \deg g)$.

    a) Use Lüroth's theorem to show F is algebraically closed in $F(X)$: the only elements of $F(X)$ that are algebraic over F are in F.

    b) If φ is a field automorphism of $F(X)$ which fixes the elements of F, show $\varphi(X) = (aX + b)/(cX + d)$ where $(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}) \in \mathrm{GL}_2(F)$.

30. Let F be a field, K be a finite extension of $F(X)$, and R be the integral closure of $F[X]$ in K. Assume R is a finite free $F[X]$-module (this assumption was proved when $K/F(X)$ is separable in Theorem 4.58 but it does hold in general by Remark 4.70).

    a) Prove R has unique factorization of ideals (Theorem 4.59).

    b) For $\alpha \in R - \{0\}$, show $\operatorname{card}_{F[X]}(R/\alpha R) = (\mathrm{N}_{K/F(X)}(\alpha))$.

    c) For nonzero ideals $\mathfrak{a}$ and $\mathfrak{b}$ in R, show

    $$\operatorname{card}_{F[X]}(R/\mathfrak{a}\mathfrak{b}) = \operatorname{card}_{F[X]}(R/\mathfrak{a}) \operatorname{card}_{F[X]}(R/\mathfrak{b}).$$

    d) Define the degree of a nonzero ideal in $F[X]$ to be the degree of any generator of the ideal. If T if a finitely generated torsion $F[X]$-module, show $\dim_F(T) = \deg(\operatorname{card}_{F[X]}(T))$, so $\dim_F(R/\alpha R) = \deg(\mathrm{N}_{K/F(X)}(\alpha))$ and $\dim_F(R/\mathfrak{a}\mathfrak{b}) = \dim_F(R/\mathfrak{a}) + \dim_F(R/\mathfrak{b})$ in the notation of parts b and c. (See Exercise 3.20.)

31. Let $\mathbf{F}$ be a finite field, K be a finite extension of the rational function field $\mathbf{F}(X)$, and R be the integral closure of $\mathbf{F}[X]$ in K.

    a) For any nonzero ideal $\mathfrak{a}$, show $\mathfrak{a}$ is a finite free $\mathbf{F}[X]$-module and $R/\mathfrak{a}$ is finite.

    b) The ring $R/\mathfrak{a}$ is both a finite ring and a finitely generated torsion $\mathbf{F}[X]$-module. Define two norms for $\mathfrak{a}$, one with value in the positive integers and the other with value in $\mathbf{F}[X]$:

    $$\mathrm{N}(\mathfrak{a}) = \#(R/\mathfrak{a}), \quad \mathrm{N}_{\mathbf{F}[X]}(\mathfrak{a}) = \operatorname{card}_{\mathbf{F}[X]}(R/\mathfrak{a}) = [R:\mathfrak{a}]_{\mathbf{F}[X]}.$$

    Show these are related by the formula

    $$\mathrm{N}(\mathfrak{a}) = \#(\mathbf{F}[X]/\,\mathrm{N}_{\mathbf{F}[X]}(\mathfrak{a})),$$

    so the $\mathbf{F}[X]$-valued norm of $\mathfrak{a}$ determines the $\mathbf{Z}$-valued norm of $\mathfrak{a}$. In particular, using the previous exercise, $\mathrm{N}(\alpha R) = \#(\mathbf{F}[X]/\,\mathrm{N}_{K/\mathbf{F}(X)}(\alpha))$ and $\mathrm{N}(\mathfrak{a}\mathfrak{b}) = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{b})$. (Hint: First check the case $R = \mathbf{F}[X]$. For any finite-dimensional $\mathbf{F}$-vector space W, $\#W = q^{\dim_{\mathbf{F}}(W)}$, where $q = \#\mathbf{F}$.)

    c) For any nonzero ideal $\mathfrak{a}$ in R, the unit group $(R/\mathfrak{a})^\times$ is finite. Define $\varphi_R(\mathfrak{a}) = \#(R/\mathfrak{a})^\times$. Derive the formula $\varphi_R(\mathfrak{a}) = \mathrm{N}(\mathfrak{a})\prod_{\mathfrak{p}|\mathfrak{a}}(1 - 1/\,\mathrm{N}(\mathfrak{p}))$, an exact parallel to a formula in the number field case (Example 4.33).

32. Let A be a PID with fraction field F, E/F be a finite separable extension, and B be the integral closure of A in E. Write $E = F(\beta)$ where $\beta \in B$ and β has minimal polynomial $f(T) \in A[T]$. Using the A-index (from Chapter 3) in place of the group index, we will extend the Dedekind–Kummer factorization method to B.

    a) For an irreducible $\pi \in A$ such that $f(T) \bmod \pi$ is separable over A/π, show $\pi \nmid [B : A[\beta]]_A$.

    b) For any irreducible $\pi \in A$ such that $\pi \nmid [B : A[\beta]]_A$, show the natural ring homomorphism $A[\beta]/\pi A[\beta] \to B/\pi B$ is an isomorphism.

    c) Use part b to explain why πB and $f \bmod \pi \in (A/\pi)[T]$ factor into prime ideals and irreducible polynomials in the same way when $\pi \nmid [B : A[\beta]]_A$. (In particular, if $f(T) \bmod \pi$ is irreducible and separable over A/π then πB is prime in B.)

    d) Let $A = F[X]$ and $E = F(X, \alpha)$, where $\alpha^3 + X\alpha + X = 0$. By Exercise 3.15, $B = F[X, \alpha]$. Find the shape of the prime ideal factorization of πB, where F and π are in the table below. (The shape includes a description of the residue field of each prime ideal factor.)

F	π
$\mathbf{R}$	$X, X+1, X^2+X+1$
$\mathbf{F}_2$	$X, X+1, X^2+X+1$
$\mathbf{C}$	$X, X+1$

33. (The circle ring) The ring $\mathbf{R}[X, \sqrt{1-X^2}] = \mathbf{R}[X,Y]/(X^2+Y^2-1)$ has unit group $\mathbf{R}^\times$ by Exercise 1.14 and it is integrally closed by Exercise 2.10.

    a) Show $(X-1) = (X-1, \sqrt{1-X^2})^2$, and the ideal $(X-1, \sqrt{1-X^2})$ is not principal, so $\mathbf{R}[X, \sqrt{1-X^2}]$ is not a PID.

    b) Show the equation $\sqrt{1-X^2}\sqrt{1-X^2} = (1+X)(1-X)$ is an example of nonunique irreducible factorization in $\mathbf{R}[X, \sqrt{1-X^2}]$. In more down to earth terms, the ring of trigonometric polynomials $\mathbf{R}[\cos\theta, \sin\theta]$ doesn't have unique factorization because of the high school identity $\sin^2\theta = (1+\cos\theta)(1-\cos\theta)$.

34. Let F be a field not of characteristic 2 and $f(X) \in F[X]$ be nonconstant and squarefree. By Exercise 2.10, the integral closure of $F[X]$ in $F(X)(\sqrt{f(X)})$ is $F[X, \sqrt{f(X)}] = F[X] + F[X]\sqrt{f(X)}$ and for any monic

irreducible factor $\pi(X)$ of $f(X)$, the ideal $\mathfrak{p}_\pi = (\pi(X), \sqrt{f(X)})$ is maximal and $\mathfrak{p}_\pi^2 = (\pi(X))$. Write $R = F[X, \sqrt{f(X)}]$, so R is a Dedekind domain. The residue rings $R/\mathfrak{a}$ for nonzero ideals $\mathfrak{a}$ are finite-dimensional F-vector spaces.

a) Show

$$(\sqrt{f(X)}) = \prod_{\pi | f} \mathfrak{p}_\pi,$$

where the product runs over monic irreducible factors $\pi(X)$ of $f(X)$ in $F[X]$.

b) For a nonzero prime ideal $\mathfrak{p}$ in R, let $\mathfrak{p} \cap F[X] = \pi(X)F[X]$, so $\mathfrak{p} \mid \pi(X)R$. Show that if $\dim_F(R/\mathfrak{p}) = 1$, then $\pi(X)$ is linear.

c) For $c \in F$, show $(X - c)$ factors in R as follows:

$$(X - c) = \begin{cases} \mathfrak{p}^2, & \text{if } f(c) = 0, \\ \mathfrak{p}\mathfrak{p}', & \text{if } f(c) = \square \text{ in } F^\times, \text{ with } \mathfrak{p} \neq \mathfrak{p}', \\ \mathfrak{p}, & \text{if } f(c) \neq \square \text{ in } F^\times. \end{cases}$$

Also determine the F-dimension of the residue field at each of these prime ideals.

d) Now we look at a specific example. Set $K = \mathbf{F}_3(X, \sqrt{X^3 + X})$. This is a quadratic extension of $\mathbf{F}_3(X)$ and the integral closure of $\mathbf{F}_3[X]$ in K is $\mathbf{F}_3[X, \sqrt{X^3 + X}]$ since $X^3 + X$ is squarefree. Determine the number of prime ideals in $\mathbf{F}_3[X, \sqrt{X^3 + X}]$ whose residue field is $\mathbf{F}_3$.

e) With K as in part d, let $X' = 1/(X - 1)$. The rings $\mathbf{F}_3[X]$ and $\mathbf{F}_3[X']$ are not equal but are obviously isomorphic to each other and these two rings have the same fraction field: $\mathbf{F}_3(X) = \mathbf{F}_3(X')$. Compute the integral closure of $\mathbf{F}_3[X']$ in K and then show the number of prime ideals in it with residue field $\mathbf{F}_3$ is not the same number as you found in part d, so $\mathbf{F}_3[X]$ and $\mathbf{F}_3[X']$ have nonisomorphic integral closures in K. (Hint: To find the integral closure of $\mathbf{F}_3[X']$ in K, set $Y = \sqrt{X^3 + X}$, $Y' = Y/(X - 1)^2 = YX'^2$, and compute Y'^2.)

f) Repeat part e using $X'' = 1/(X + 1)$ in place of X'.

35. Let $K = \mathbf{F}_p(X^{1/p}, Y^{1/p})$.

a) Show $[K : \mathbf{F}_p(X, Y)] = p^2$ and $K^p = \mathbf{F}_p(X, Y)$.

b) Set $F = \mathbf{F}_p(Y)$, so $[K : F(X)] = p^2$ and F is an infinite field of characteristic p. Show there is no $U \in K$ such that $K/F(U)$ is a finite separable extension. This is a contrast with Theorem 4.60.

36. If the field E is a finitely generated extension of the field F (meaning $E = F(a_1, \ldots, a_n)$), show every intermediate field is also a finitely generated extension of F. In particular, the elements of E algebraic over F form a finite-dimensional extension of F.

37. Show a domain is a UFD if and only if every nonzero principal proper ideal is a unique (finite) product of principal prime ideals.

38. Since the ideal norm is totally multiplicative on nonzero ideals in $\mathcal{O}_K$, it extends uniquely to a multiplicative function on fractional ideals: when $\mathfrak{a} = IJ^{-1}$ for integral ideals I and J, set $\mathrm{N}(\mathfrak{a}) = \mathrm{N}(I)\,\mathrm{N}(J)^{-1}$. This is a positive rational number, so we lose the combinatorial meaning of $\mathrm{N}(\mathfrak{a})$ as $[\mathcal{O}_K : \mathfrak{a}]$. But if we use the generalized lattice index defined in Exercise 3.18, show it is still true that $\mathrm{N}(\mathfrak{a}) = [\mathcal{O}_K : \mathfrak{a}]$ if $\mathfrak{a} \not\subset \mathcal{O}_K$. That is, show $\mathrm{N}(\mathfrak{a}) = |\det f|$, where $f\colon K \to K$ is any $\mathbf{Q}$-linear map such that $f(\mathcal{O}_K) = \mathfrak{a}$.

39. For nonzero ideals $\mathfrak{a}$ and $\mathfrak{b}$ in a Dedekind domain A, show $\mathrm{lcm}(\mathfrak{a}, \mathfrak{b}) = \mathfrak{a} \cap \mathfrak{b}$. That is, if $\mathfrak{a} = \prod_i \mathfrak{p}_i^{a_i}$ and $\mathfrak{b} = \prod_i \mathfrak{p}_i^{b_i}$, then $\mathfrak{a} \cap \mathfrak{b} = \prod_i \mathfrak{p}_i^{\max(a_i, b_i)}$.

40. For nonzero x and y in a Dedekind domain A, show $(x, y)^n = (x^n, y^n)$ for all $n \geqslant 1$. (Hint: Think about (x, y) as the greatest common divisor of the ideals (x) and (y).)

41. Let A be a Dedekind domain with fraction field F. Suppose that every nonzero element α of F can be written in reduced form over F: $\alpha = x/y$ where x and y are relatively prime in A (that is, $(x, y) = (1)$). Show A is a PID, so any Dedekind domain that is not a PID has an example of a ratio in its fraction field that has no reduced form expression. (Hint: any nonzero ideal in A has 2 nonzero generators.)

42. Let A be a nonzero commutative ring which admits a decomposition $A = A_1 \times A_2$ as a direct product of nonzero rings. (The subsets $A_1 \times \{0\}$ and $\{0\} \times A_2$ are *not* subrings of A, but ideals. They are not subrings since they do not contain the multiplicative identity of A, although they do have their own identities, $(1, 0)$ or $(0, 1)$.)

a) Prove the prime ideals of A are the ideals $\mathfrak{p}_1 \times A_2$ or $A_1 \times \mathfrak{p}_2$, where $\mathfrak{p}_i$ is a prime ideal of A_i, and such ideals are maximal precisely when $\mathfrak{p}_i$ is maximal in A_i.

b) An *idempotent* is an element satisfying $x^2 = x$. The trivial idempotents are 0 and 1. Others are called nontrivial. (For example, in $\mathbf{Z}/(6)$, 3 and 4 are nontrivial idempotents.) Prove A has nontrivial idempotents. Conversely, show a ring with a nontrivial idempotent has a decomposition into a direct product of two nonzero rings. (Hint: if x is an idempotent, $1 - x$ is as well.)

c) Let K be a number field. For a prime $\mathfrak{p}$ in $\mathcal{O}_K$ and integer $m \geqslant 1$, prove $\mathcal{O}_K/\mathfrak{p}^m$ does not admit a decomposition into a direct product of two (nonzero) rings. On the other hand, when $\mathfrak{a} \neq (0), (1)$ is not a prime power, prove $\mathcal{O}_K/\mathfrak{a}$ does admit such a decomposition. (Equivalently, there are nontrivial idempotents in $\mathcal{O}_K/\mathfrak{a}$.)

d) Part a concerned prime ideals of a product ring. If we work with general ideals in a product ring, say $A_1 \times A_2$, does every ideal have the form $I_1 \times I_2$ with I_j an ideal of A_j?

CHAPTER 5

IDEAL CLASSES

The nonzero ideals in $\mathcal{O}_K$, considered up to scaling, lead to a fundamental finite abelian group: the ideal class group of K. This group is trivial precisely when $\mathcal{O}_K$ is a PID, which by Theorem 4.87 is the same as $\mathcal{O}_K$ being a UFD. So the ideal class group of K is the obstruction to $\mathcal{O}_K$ having unique factorization of elements. To illustrate the importance of the ideal class group of K in other problems, we will see how it parametrizes both the orbits for a group action and the possible $\mathcal{O}_K$-module structures for the integers in a finite extension of K.

5.1 Ideal Classes

Let A be a domain with fraction field F. Recall the fractional A-ideals are the nonzero A-submodules $\mathfrak{a} \subset F$ such that $d\mathfrak{a} \subset A$ for some nonzero $d \in A$. These are the nonzero ideals in A divided by elements of $F^\times$. Principal fractional A-ideals are $(x) := xA$ for $x \in F^\times$.

Call two fractional A-ideals $\mathfrak{a}$ and $\mathfrak{a}'$ equivalent when they are related by scaling: $\mathfrak{a} = x\mathfrak{a}'$ for some $x \in F^\times$. Write this as $\mathfrak{a} \sim \mathfrak{a}'$. The equivalence class of $\mathfrak{a}$ (the set of all $x\mathfrak{a}$ for $x \in F^\times$) is called the *ideal class* of $\mathfrak{a}$ and we write it as $[\mathfrak{a}]$, so $[\mathfrak{a}] = [\mathfrak{a}']$ is the same thing as $\mathfrak{a} \sim \mathfrak{a}'$. The principal fractional A-ideals form a single ideal class, namely $[(1)] = [A]$. Every ideal class is represented by a nonzero ideal in A since we can write any fractional A-ideal as $\frac{1}{d}I$ for some nonzero ideal I in A, and $\frac{1}{d}I \sim I$. An ideal in A is equivalent to (1) if and only if it is a principal ideal: if $\mathfrak{a} \subset A$ and $\mathfrak{a} = xA$ for some $x \in F^\times$ then $x \in xA = \mathfrak{a} \subset A$.

Since $xAyA = xyA$, it makes sense to multiply ideal classes by multiplying representatives: $[\mathfrak{a}][\mathfrak{a}'] = [\mathfrak{a}\mathfrak{a}']$. Multiplication of ideal classes is obviously commutative and associative, with identity $[(1)] = [A]$, which will usually be written just as 1. We say $[\mathfrak{a}]$ is invertible if $[\mathfrak{a}][\mathfrak{b}] = 1$ for some ideal class $[\mathfrak{b}]$. That means $\mathfrak{a}\mathfrak{b}$ is a principal fractional A-ideal. If an ideal class $[\mathfrak{a}]$ has an inverse ideal class, $\mathfrak{a}$ has an inverse as a fractional A-ideal: if $[\mathfrak{a}][\mathfrak{b}] = 1$ then $\mathfrak{a}\mathfrak{b} = xA$ for some $x \in F^\times$, so $\mathfrak{a} \cdot \frac{1}{x}\mathfrak{b} = A$. Conversely, if $\mathfrak{a}$ is an invertible fractional A-ideal then its ideal class is invertible: $\mathfrak{a}\mathfrak{a}' = A \Rightarrow [\mathfrak{a}][\mathfrak{a}'] = 1$. Some ideal classes might not have inverses, so the ideal classes might not form a group.

Example 5.1. In $\mathbf{Z}[\sqrt{5}]$, let $\mathfrak{p} = (2, 1+\sqrt{5})$. This is a prime ideal (index 2 in $\mathbf{Z}[\sqrt{5}]$) and $\mathfrak{p}^2 = 2\mathfrak{p}$ (Exercise 1.25), so $\mathfrak{p}^2 \sim \mathfrak{p}$. Thus $[\mathfrak{p}]^2 = [\mathfrak{p}]$. If $[\mathfrak{p}]$ had an inverse, then $\mathfrak{p}$ would have an inverse as a fractional $\mathbf{Z}[\sqrt{5}]$-ideal, so the equation $\mathfrak{p}^2 = 2\mathfrak{p}$ would imply $\mathfrak{p} = (2)$ by cancellation. But $1+\sqrt{5} \in \mathfrak{p}$ and $1+\sqrt{5} \notin (2)$.

The ideal classes of $\mathbf{Z}[\sqrt{5}]$ are not a group and can't be embedded in a group, since Example 5.1 shows there would be a contradiction if $[\mathfrak{p}]$ becomes invertible somehow.

Theorem 5.2. *The ideal classes of fractional ideals in any number field form a group.*

Proof. All fractional ideals in a number field have inverses. ■

We call the ideal classes of fractional ideals in K its *ideal class group* or just *class group* and write it as $\mathrm{Cl}(K)$. The elements of this group are ideal classes $[\mathfrak{a}] = \{x\mathfrak{a} : x \in K\}$, with the group law being multiplication of representatives. The group $\mathrm{Cl}(K)$ is abelian, and this group is trivial if and only if all fractional ideals in K are principal, which is equivalent to $\mathcal{O}_K$ being a PID. The groups $\mathrm{Cl}(K)$ are fundamental objects in number theory. Here is the basic finiteness theorem about them.

Theorem 5.3. *Any number field K has a finite ideal class group. That is, there are a finite number of fractional ideals $\mathfrak{a}_1, \ldots, \mathfrak{a}_r$ such that every fractional ideal is $x\mathfrak{a}_i$ for some $x \in K^\times$.*

Proof. The argument has two steps.

Step 1. There is a constant $C > 0$ such that in every nonzero ideal $\mathfrak{a} \subset \mathcal{O}_K$ there is a nonzero α such that $\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| \leqslant C[\mathcal{O}_K : \mathfrak{a}]$.

(For nonzero $\alpha \in \mathfrak{a}$, $(\alpha) \subset \mathfrak{a} \subset \mathcal{O}_K$, so $\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| = [\mathcal{O}_K : (\alpha)] \geqslant [\mathcal{O}_K : \mathfrak{a}]$. Thus we are saying this inequality can be reversed for some α in $\mathfrak{a}$ at the cost of introducing a constant C that is independent of the choice of $\mathfrak{a}$.)

The constant C will depend on a choice of $\mathbf{Z}$-basis of $\mathcal{O}_K$. Write $n = [K : \mathbf{Q}]$ and $\mathcal{O}_K = \mathbf{Z}e_1 \oplus \cdots \oplus \mathbf{Z}e_n$. For only the second time, we will use embeddings of K into the complex numbers,[1] and now we will make *estimates* with these embeddings. There are n field embeddings

$$\sigma_1, \ldots, \sigma_n \colon K \to \mathbf{C},$$

and the norm map $\mathrm{N}_{K/\mathbf{Q}}$ can be expressed in terms of them (Theorem 8.18): for each $x \in K$,

$$\mathrm{N}_{K/\mathbf{Q}}(x) = \sigma_1(x)\sigma_2(x)\cdots\sigma_n(x). \tag{5.1}$$

Writing $x = c_1e_1 + \cdots + c_ne_n$ with $c_i \in \mathbf{Q}$, we get

$$\begin{aligned}\left|\mathrm{N}_{K/\mathbf{Q}}(x)\right| &= \prod_{j=1}^{n} |\sigma_j(x)| \\ &= \prod_{j=1}^{n} \left|\sum_{i=1}^{n} c_i\sigma_j(e_i)\right| \\ &\leqslant \prod_{j=1}^{n} \left(\sum_{i=1}^{n} |c_i|\,|\sigma_j(e_i)|\right) \\ &\leqslant \left(\max|c_i|\right)^n \underbrace{\prod_{j=1}^{n} \left(\sum_{i=1}^{n} |\sigma_j(e_i)|\right)}_{\text{Call this } C}.\end{aligned}$$

For any nonzero ideal $\mathfrak{a}$ in $\mathcal{O}_K$, its index in $\mathcal{O}_K$ lies between nth powers of consecutive integers, say $k^n \leqslant [\mathcal{O}_K : \mathfrak{a}] < (k+1)^n$. The set

$$\left\{\sum_{i=1}^{n} a_ie_i : a_i \in \mathbf{Z},\ 0 \leqslant a_i \leqslant k\right\}$$

[1] The previous time was in the discriminant formula in Lemma 3.24 and its consequences Theorems 3.25, 3.28, and 3.52.

has size $(k+1)^n$, so by the pigeonhole principle we have an instance of

$$\sum_{i=1}^{n} a_i e_i \equiv \sum_{i=1}^{n} a_i' e_i \bmod \mathfrak{a},$$

where $0 \leqslant a_i, a_i' \leqslant k$ and $a_i \neq a_i'$ for some i. Taking the difference $c_i = a_i - a_i'$,

$$\sum_{i=1}^{n} c_i e_i \in \mathfrak{a},$$

and $|c_i| \leqslant k$ with $c_i \neq 0$ for some i. Call this sum α. Then

$$\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| \leqslant \left(\max |c_i|\right)^n C \leqslant k^n C \leqslant [\mathcal{O}_K : \mathfrak{a}]C.$$

Step 2. (Finiteness)

Every fractional ideal class is represented by a nonzero ideal $\mathfrak{a} \subset \mathcal{O}_K$. Pick nonzero α in $\mathfrak{a}$ such that

$$\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| \leqslant C[\mathcal{O}_K : \mathfrak{a}] \tag{5.2}$$

by Step 1. Since $\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| = [\mathcal{O}_K : \alpha\mathcal{O}_K]$ and $\alpha\mathcal{O}_K \subset \mathfrak{a} \subset \mathcal{O}_K$, the inequality (5.2) is equivalent to $[\mathfrak{a} : \alpha\mathcal{O}_K] \leqslant C$. So $[\frac{1}{\alpha}\mathfrak{a} : \mathcal{O}_K] \leqslant C$. Thus every fractional ideal class is represented by a fractional ideal $\frac{1}{\alpha}\mathfrak{a}$ which *contains* $\mathcal{O}_K$ with index bounded above independently of the ideal class.

To prove there are finitely many ideal classes, it suffices to show for each $r \in \mathbf{Z}^+$ that there are finitely many fractional ideals containing $\mathcal{O}_K$ with index r. If $\mathcal{O}_K \subset \mathfrak{a}$ and $[\mathfrak{a} : \mathcal{O}_K] = r$, then $r\mathfrak{a} \subset \mathcal{O}_K$, so $\mathcal{O}_K \subset \mathfrak{a} \subset \frac{1}{r}\mathcal{O}_K$. Since $[\frac{1}{r}\mathcal{O}_K : \mathcal{O}_K] = r^n$, $\frac{1}{r}\mathcal{O}_K/\mathcal{O}_K$ is finite, so there are finitely many such $\mathfrak{a}$. We're done. ■

The proof of Theorem 5.3 tells us the ideal classes in $\mathrm{Cl}(K)$ are represented by fractional ideals $\mathfrak{a}$ such that $\mathcal{O}_K \subset \mathfrak{a}$ and $[\mathfrak{a} : \mathcal{O}_K] \leqslant C$, where

$$C = \prod_{\sigma \colon K \to \mathbf{C}} \sum_{i=1}^{n} |\sigma(e_i)|$$

for some choice of $\mathbf{Z}$-basis $\{e_1, \ldots, e_n\}$ of $\mathcal{O}_K$. That doesn't mean C is a bound on the number of ideal classes; it is a bound on the index with which some fractional ideal in each ideal class contains $\mathcal{O}_K$. There could be several fractional ideals containing $\mathcal{O}_K$ with the same index, but there are only a finite number of them.

We will call C the *Kronecker bound* since it essentially occurs in Kronecker's thesis [31, p. 15] in the special case of $\mathbf{Q}(\zeta_p)$ and Kronecker pointed out later [32, pp. 64–65] that the argument using this bound applies to any number field.

Counting fractional ideals that contain $\mathcal{O}_K$ with a given index might feel a bit strange compared to counting ideals inside $\mathcal{O}_K$ with a given index. Using inversion, we will pass to the second point of view.

Theorem 5.4. *The ideal classes of $\mathcal{O}_K$ are*

- *represented by ideals in $\mathcal{O}_K$ with norm at most C,*
- *generated as a group by prime ideals $\mathfrak{p}$ with $\mathrm{N}(\mathfrak{p}) \leqslant C$.*

Proof. We already know the ideal classes are represented by fractional ideals $\mathfrak{a}$ where $\mathcal{O}_K \subset \mathfrak{a}$ and $[\mathfrak{a} : \mathcal{O}_K] \leqslant C$. Write the condition $\mathcal{O}_K \subset \mathfrak{a}$ as $\mathfrak{a}^{-1} \subset \mathcal{O}_K$. We will show $[\mathcal{O}_K : \mathfrak{a}^{-1}] = [\mathfrak{a} : \mathcal{O}_K]$, so inversion exchanges the fractional ideals containing $\mathcal{O}_K$ with index at most C and the ideals contained in $\mathcal{O}_K$ with index (norm) at most C.

Write $\mathfrak{a}^{-1} = \mathfrak{b}$, which is an ideal in $\mathcal{O}_K$, and write $\mathfrak{a} = \frac{1}{a}\mathfrak{c}$ for $a \in \mathcal{O}_K$ and $\mathfrak{c} \subset \mathcal{O}_K$. Then $(a) = \mathfrak{b}\mathfrak{c}$, so $\mathrm{N}((a)) = \mathrm{N}(\mathfrak{b})\,\mathrm{N}(\mathfrak{c})$ and $[\mathfrak{a} : \mathcal{O}_K] = [\frac{1}{a}\mathfrak{c} : \mathcal{O}_K] = [\mathfrak{c} : a\mathcal{O}_K] = \mathrm{N}((a))/\,\mathrm{N}(\mathfrak{c}) = \mathrm{N}(\mathfrak{b}) = [\mathcal{O}_K : \mathfrak{b}] = [\mathcal{O}_K : \mathfrak{a}^{-1}]$.

Ideals in $\mathcal{O}_K$ with norm at most C are products of prime ideals with norm at most C, so the ideal classes of such primes generate $\mathrm{Cl}(K)$. ∎

Just like $\mathcal{O}_K$, the ideal classes of any Dedekind domain A form a group since all fractional A-ideals are invertible. The group of ideal classes of fractional A-ideals is called the ideal class group of A and is written as $\mathrm{Cl}(A)$, so what we wrote before as $\mathrm{Cl}(K)$ for number fields K is really $\mathrm{Cl}(\mathcal{O}_K)$. While the ring of integers of a number field has a finite ideal class group, other Dedekind domains can have an infinite ideal class group. For example, the ideal class group of $\mathbf{C}[X, \sqrt{X^3 - X}]$ (which is integrally closed by Exercise 2.10) turns out to be isomorphic to the torus $\mathbf{C}/(\mathbf{Z} + \mathbf{Z}i)$. In a sense, the "reason" ideal class groups of number fields are finite is that $\mathbf{Z}/m\mathbf{Z}$ is finite for $m \neq 0$; we did use that finiteness in the proof of Theorem 5.3. To justify this idea, the integral closure of $\mathbf{F}[x]$ in a finite extension of $\mathbf{F}(x)$ is finite when $\mathbf{F}$ is a finite field and the proof of that uses finiteness of $\mathbf{F}[[x]/(f(x))$ for nonzero $f(x)$ (Exercise 5.13).

The group $\mathrm{Cl}(A)$ is trivial if and only if A is a PID, which is equivalent to A being a UFD (Theorem 4.87), so $\mathrm{Cl}(A)$ is a measure of how far A is from having unique factorization of elements. The ideal class group is abelian, and a theorem of Claborn [9, pp. 219–222] says every abelian group is the ideal class group of some Dedekind domain. It is believed that every finite abelian group is the class group of some number field, but this is still unsolved.

Since $x\mathfrak{a} = xA \cdot \mathfrak{a}$ and the principal fractional A-ideals form a group under multiplication ($xA \cdot yA = xyA$ and $(xA)^{-1} = \frac{1}{x}A$), we can think about ideal classes as cosets for the subgroup of principal fractional A-ideals. Therefore when A is Dedekind, $\mathrm{Cl}(A)$ can be regarded as a quotient group

$$\mathrm{Cl}(A) = \{\text{fractional } A\text{-ideals}\} \,/\, \{\text{principal fractional } A\text{-ideals}\}\,. \tag{5.3}$$

In any Dedekind domain, all the fractional ideals are invertible. It turns out that the converse is true as well. This will be a consequence of the next two lemmas.

Let A be a domain with fraction field F. For any two A-modules M and N in F (these modules are not assumed to be finitely generated), we define their product to be the A-module

$$MN := \left\{ \sum_{i=1}^{r} x_i y_i : r \geqslant 1,\ x_i \in M,\ y_i \in N \right\}.$$

The identity for this multiplication is A. Invertibility for this multiplication has a built-in finiteness:

Lemma 5.5. *If $MN = A$ then M and N are finitely generated.*

Proof. Some *finite* sum of products is equal to 1: $x_1x_1' + \cdots + x_kx_k' = 1$ where $x_i \in M$ and $x_i' \in N$. For any $x \in M$,

$$x = 1 \cdot x = x_1(x_1'x) + \cdots + x_k(x_k'x),$$

and $x_i'x \in NM = A$, so $M \subset \sum_{i=1}^{k} Ax_i \subset M$. Thus $M = \sum_{i=1}^{k} Ax_i$. Similarly, $N = \sum_{i=1}^{k} Ax_i'$. ∎

A finitely generated A-module in F certainly has a common denominator, so Lemma 5.5 tells us that invertible A-modules in F are automatically fractional A-ideals.

Lemma 5.6. *If a domain has cancellation of ideals, i.e., always $\mathfrak{ac} = \mathfrak{bc}$ implies $\mathfrak{a} = \mathfrak{b}$ when $\mathfrak{c} \neq (0)$, then the domain is integrally closed.*

Proof. Let A be a domain with cancellation of ideals. Suppose an element x in the fraction field of A is integral over A. We want to show x is in A. Write $x = a/b$ where a and b are in A with $b \neq 0$. Since x is integral over A,

$$x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0 = 0$$

with $n \geqslant 1$ and $c_i \in A$. Let $R = A[x] = A + Ax + \cdots + Ax^{n-1}$. This is a ring and a nonzero A-module in the fraction field of A. Since x has denominator b, by the definition of R we have $b^{n-1}R \subset A$, so R has common denominator b^{n-1}. Therefore

$$\mathfrak{a} := b^{n-1}R = Ab^{n-1} + Ab^{n-2}a + \cdots + Aa^{n-1}$$

is a nonzero A-module in A, *i.e.*, $\mathfrak{a}$ is a nonzero ideal in A. Since R is a ring, $R^2 = R$, so $\mathfrak{a}^2 = b^{2(n-1)}R^2 = b^{n-1}b^{n-1}R = (b)^{n-1}\mathfrak{a}$. Therefore by cancellation of nonzero ideals in A, $\mathfrak{a} = (b)^{n-1} = b^{n-1}A$, so $b^{n-1}R = b^{n-1}A$. This implies $R = A$, so $x \in R = A$. ■

Theorem 5.7. *If all ideal classes for a domain are invertible, then the domain is a Dedekind domain.*

Proof. Let A be the domain. The hypothesis is equivalent to saying all nonzero ideals in A are invertible as fractional A-ideals. By Lemma 5.5, all ideals in A are finitely generated, so A is Noetherian. Invertible ideals can be cancelled, so Lemma 5.6 tells us that A is integrally closed. It remains to show that any nonzero prime ideal $\mathfrak{p}$ is inveritble.

Suppose $\mathfrak{p} \subset \mathfrak{a} \subset A$. Then $\mathfrak{p}\mathfrak{a}^{-1} \subset A$, so $\mathfrak{p}\mathfrak{a}^{-1}$ is an ideal and $\mathfrak{p} = \mathfrak{p}\mathfrak{a}^{-1} \cdot \mathfrak{a}$. By Theorem 4.3, $\mathfrak{p} \supset \mathfrak{p}\mathfrak{a}^{-1}$ or $\mathfrak{p} \supset \mathfrak{a}$. The first condition implies $\mathfrak{p}\mathfrak{a} = \mathfrak{p}$ and the second condition implies $\mathfrak{p} = \mathfrak{a}$. Therefore $\mathfrak{a}$ is A or $\mathfrak{p}$, so $\mathfrak{p}$ is maximal. ■

For domains like $\mathbf{Z}[\sqrt{5}]$ which are not Dedekind domains, the set of all their ideal classes under multiplication is just a monoid ("group without inverses"). We get a group by focusing on the invertible classes. For a domain A, its ideal class group $\mathrm{Cl}(A)$ is, by definition, the group of its invertible ideal classes where the group law is multiplication of ideal classes and $[(1)]$ is the identity. Equivalently,

$$\mathrm{Cl}(A) = \{\text{invertible fractional } A\text{-ideals}\} \,/\, \{\text{principal fractional } A\text{-ideals}\}\,.$$

Compare this with (5.3), where A is Dedekind.

5.2 Ideal Classes for $\mathbf{Q}(\sqrt{-5})$

As an example of a class group computation, we will show $\mathbf{Q}(\sqrt{-5})$ has two ideal classes and then look at several applications of that.

Theorem 5.8. *The ideal class group of* $\mathbf{Q}(\sqrt{-5})$ *has order two.*

Proof. Use $\{e_1, e_2\} = \left\{1, \sqrt{-5}\right\}$. Since

$$C = (|e_1| + |e_2|)(|\overline{e_1}| + |\overline{e_2}|) = (1 + \sqrt{5})^2 \approx 10.4,$$

Theorem 5.4 says the group $\mathrm{Cl}(\mathbf{Q}(\sqrt{-5}))$ is

- represented by $\mathfrak{a} \subset \mathbf{Z}[\sqrt{-5}]$ such that $\mathrm{N}(\mathfrak{a}) \leqslant 10$,
- generated by primes $\mathfrak{p}$ with $\mathrm{N}(\mathfrak{p}) \leqslant 10$.

Let's find all such $\mathfrak{p}$.

If $\mathrm{N}(\mathfrak{p}) \leqslant 10$, then $\mathfrak{p}$ divides (2), (3), (5), or (7). We already know from Section 4.4 that

$$(2) = \mathfrak{p}_2^2,\ (3) = \mathfrak{p}_3\mathfrak{p}_3',\ (5) = (\sqrt{-5})^2,\ (7) = \mathfrak{p}_7\mathfrak{p}_7'.$$

In particular, $\mathfrak{p}_2$ is the unique prime factor of (2).

In $\mathrm{Cl}(\mathbf{Q}(\sqrt{-5}))$ principal ideals become trivial, so

$$[\mathfrak{p}_2]^2 = 1,\ [\mathfrak{p}_3][\mathfrak{p}_3'] = 1,\ [\mathfrak{p}_7][\mathfrak{p}_7'] = 1.$$

Thus $\mathrm{Cl}(\mathbf{Q}(\sqrt{-5}))$ is generated by $\mathfrak{p}_2$, either prime ideal of norm 3, and either prime ideal of norm 7. Since $(1 + \sqrt{-5}) = \mathfrak{p}_2\mathfrak{p}_3$ and $(3 + \sqrt{-5}) = \mathfrak{p}_2\mathfrak{p}_7$, $[\mathfrak{p}_3]$ and $[\mathfrak{p}_7]$ both equal $[\mathfrak{p}_2]^{-1}$. Thus $\mathrm{Cl}(\mathbf{Q}(\sqrt{-5})) = \langle[\mathfrak{p}_2]\rangle$. Since $\mathfrak{p}_2$ is not principal and its square is principal, $\mathrm{Cl}(\mathbf{Q}(\sqrt{-5})) \cong \mathbf{Z}/2\mathbf{Z}$. ■

Two consequences of this, for any nonzero ideal $\mathfrak{a}$ in $\mathbf{Z}[\sqrt{-5}]$, are

- $\mathfrak{a}^2$ is principal since $[\mathfrak{a}]^2 = 1$.
- either $\mathfrak{a}$ is principal or $[\mathfrak{a}] = [\mathfrak{p}_2]$, in which case $[\mathfrak{a}\mathfrak{p}_2] = [\mathfrak{a}][\mathfrak{p}_2] = [\mathfrak{p}_2]^2 = 1$, so $\mathfrak{a}\mathfrak{p}_2$ is principal.

We saw this in Section 4.4 for prime ideals of small norm in $\mathbf{Z}[\sqrt{-5}]$: their squares are all principal ideals and each nonprincipal prime of small norm has a principal product with $\mathfrak{p}_2$ (*e.g.*, $\mathfrak{p}_2\mathfrak{p}_3 = (1+\sqrt{-5})$).

Theorem 5.9. *For a prime number p, $-5 \equiv \Box \bmod p \Longleftrightarrow p$ or $2p$ has the form x^2+5y^2 for some integers x and y, and we can't have both p and $2p$ of that form.*

Proof. First we will show

$$-5 \equiv \Box \bmod p \Longleftrightarrow (p) = \mathfrak{p}\mathfrak{p}', \tag{5.4}$$

where $\mathfrak{p}$ and $\mathfrak{p}'$ are (possibly equal) prime ideals in $\mathbf{Z}[\sqrt{-5}]$.

Having $-5 \equiv \Box \bmod p$ is the same thing as $T^2+5 \bmod p$ having a nontrivial factorization, and by Kummer's factorization theorem that is the same as (p) having a nontrivial factorization in $\mathbf{Z}[\sqrt{-5}]$, which must be $\mathfrak{p}\mathfrak{p}'$ since $\mathrm{N}((p)) = p^2$. That settles (5.4).

Next we show

$$p = x^2+5y^2 \text{ for some } x, y \in \mathbf{Z} \Longleftrightarrow (p) = \mathfrak{p}\mathfrak{p}' \text{ with principal } \mathfrak{p}, \mathfrak{p}'. \tag{5.5}$$

The key point is that the prime ideals $\mathfrak{p}$ and $\mathfrak{p}'$ in (5.5) are principal.

($\Rightarrow$) If $p = x^2+5y^2$ for some x and y in $\mathbf{Z}$, then $(p) = (x+y\sqrt{-5})(x-y\sqrt{-5})$. The principal ideals on the right both have norm $x^2+5y^2 = p$, so they are prime ideals.

($\Leftarrow$) Suppose $(p) = (\alpha)(\beta)$ where (α) and (β) are principal prime ideals. Taking norms of both sides shows (α) and (β) have norm p. Writing $\alpha = x+y\sqrt{-5}$, we get $p = \mathrm{N}((\alpha)) = |x^2+5y^2| = x^2+5y^2$.

Finally, we show

$$2p = x^2+5y^2 \text{ for some } x, y \in \mathbf{Z} \Longleftrightarrow (p) = \mathfrak{p}\mathfrak{p}' \text{ with nonprincipal } \mathfrak{p}, \mathfrak{p}'. \tag{5.6}$$

($\Rightarrow$) If $2p = x^2+5y^2$ for some integers x and y, then $(x+y\sqrt{-5}) = \mathfrak{p}_2\mathfrak{p}$, where $\mathfrak{p}$ has norm p. The ideal $\mathfrak{p}$ can't be principal, because if it were then $\mathfrak{p}_2$ would be a principal fractional ideal and thus a principal ideal, but we know $\mathfrak{p}_2$ is not a principal ideal. Similarly, $(x-y\sqrt{-5})$ has ideal norm $2p$ so $(x-y\sqrt{-5}) = \mathfrak{p}_2\mathfrak{p}'$, where $\mathfrak{p}'$ has norm p and $\mathfrak{p}'$ is not principal. Multiplying these factorizations of

$(x+y\sqrt{-5})$ and $(x-y\sqrt{-5})$, we get

$$(2p) = (x+y\sqrt{-5})(x-y\sqrt{-5}) = \mathfrak{p}_2\mathfrak{p}\mathfrak{p}_2\mathfrak{p}' = \mathfrak{p}_2^2\mathfrak{p}\mathfrak{p}' = (2)\mathfrak{p}\mathfrak{p}',$$

so $(p) = \mathfrak{p}\mathfrak{p}'$ with nonprincipal $\mathfrak{p}$ and $\mathfrak{p}'$.

($\Leftarrow$) If $(p) = \mathfrak{p}\mathfrak{p}'$ with nonprincipal prime factors, then $\mathfrak{p}$ and $\mathfrak{p}'$ have norm p. Because there are only two ideal classes, the product of two nonprincipal ideals is principal, so $\mathfrak{p}_2\mathfrak{p} = (x+y\sqrt{-5})$ for some x and y in $\mathbf{Z}$. Taking the norm of both sides, $2p = |x^2+5y^2| = x^2+5y^2$.

Since the right sides of (5.5) and (5.6) are not compatible, by unique factorization of ideals, (5.4) tells us that $-5 \equiv \square \bmod p$ is equivalent to p or $2p$ being a value of x^2+5y^2 but both can't happen. ■

The condition $2p = x^2+5y^2$ can be recast in terms of a representation theorem for p itself: $p = 2m^2+2mn+3n^2$ for some $m, n \in \mathbf{Z}$. If $2p = x^2+5y^2$ then reducing mod 2 gives $0 \equiv x^2+y^2 \equiv x+y \bmod 2$, so $x \equiv y \bmod 2$. Write $x = y+2m$, so

$$2p = (y+2m)^2+5y^2 = 4m^2+4my+6y^2 \Longrightarrow p = 2m^2+2my+3y^2.$$

Conversely, if $p = 2m^2+2mn+3n^2$, then $2p = 4m^2+4mn+6n^2 = (2m+n)^2+5n^2$. Therefore Theorem 5.9 can be recast as saying

$$-5 \equiv \square \bmod p \Longleftrightarrow p \text{ is } x^2+5y^2 \text{ or } 2x^2+2xy+3y^2 \text{ for some } x, y \in \mathbf{Z}, \quad (5.7)$$

and only one of these possibilities occurs.

Equation (5.7) is the way Gauss and Lagrange would have said $\mathbf{Q}(\sqrt{-5})$ has 2 ideal classes.

5.3 More Examples

Here is a procedure for finding $\mathrm{Cl}(K)$:

- Pick a $\mathbf{Z}$-basis for $\mathcal{O}_K$, say $\{e_1, \ldots, e_n\}$.

- Set the Kronecker bound to be

$$C = \prod_{\sigma\colon K\to\mathbf{C}} \Big(\sum_{i=1}^{n} |\sigma(e_i)|\Big).$$

The group $\mathrm{Cl}(K)$ is generated by primes $\mathfrak{p}$ where $\mathrm{N}(\mathfrak{p}) \leqslant C$.

- Find all primes $\mathfrak{p}$ such that $\mathrm{N}(\mathfrak{p}) \leqslant C$.
- Figure out relations among $[\mathfrak{p}]$ where $\mathrm{N}(\mathfrak{p}) \leqslant C$.

Writing $\mathrm{N}(\mathfrak{p}) = p^f$, if $\mathrm{N}(\mathfrak{p}) \leqslant C$ then $p \leqslant C$, so we factor all (p) where $p \leqslant C$ and look at its prime ideal factors with norm at most C to get generators for $\mathrm{Cl}(K)$. The last step above is the hardest. You may happen to find enough elements in $\mathcal{O}_K$ to show all the prime ideals in your list are principal (*e.g.*, if $\alpha \in \mathfrak{p}$ and $\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| = \mathrm{N}(\mathfrak{p})$, then $\mathfrak{p} = (\alpha)$), in which case $h = 1$, but if you are left with some ideals that you suspect are not principal and want to prove they aren't (so $h > 1$), how do you do that? One way to show $\mathfrak{a}$ is not principal is to compute $\mathrm{N}(\mathfrak{a})$ and, after crossing your fingers, hope you can show there is no element α such that $\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| = \mathrm{N}(\mathfrak{a})$.

Example 5.10. Let $K = \mathbf{Q}(\sqrt{5})$ and $\mathcal{O}_K = \mathbf{Z}[\frac{1+\sqrt{5}}{2}]$. Using $\{1, \frac{1+\sqrt{5}}{2}\}$ as a $\mathbf{Z}$-basis for $\mathcal{O}_K$,

$$C = \left(1 + \left|\frac{1+\sqrt{5}}{2}\right|\right)\left(1 + \left|\frac{\sqrt{5}-1}{2}\right|\right) = 2 + \sqrt{5} \approx 4.2.$$

We need to factor all (p) where $p \leqslant 4$. That means we will factor $T^2 - T - 1 \bmod p$ for $p = 2$ and $p = 3$. It is irreducible both times, so (2) and (3) are both prime. Since $\mathrm{N}((2)) = 4$ and $\mathrm{N}((3)) = 9$, the only prime ideal with norm at most 4 is (2), which is principal, so $\mathrm{Cl}(\mathcal{O}_K)$ is trivial. Thus $\mathbf{Z}[\frac{1+\sqrt{5}}{2}]$ is a PID, and we did not show this by checking if the ring is Euclidean. (It is Euclidean, but we don't discuss how to show that.)

Example 5.11. Let $K = \mathbf{Q}(\sqrt{-23})$ and as a $\mathbf{Z}$-basis of $\mathcal{O}_K$ choose $\{1, \frac{1+\sqrt{-23}}{2}\}$, which implies

$$C = \left(1 + \left|\frac{1+\sqrt{-23}}{2}\right|\right)\left(1 + \left|\frac{1-\sqrt{-23}}{2}\right|\right) \approx 11.8.$$

Table 5.1 lists the factorization of $T^2 - T + 6 \bmod p$ for $p \leqslant C$.

The primes ideals with norm at most 11 are $\mathfrak{p}_2, \mathfrak{p}_2', \mathfrak{p}_3$, and $\mathfrak{p}_3'$. Since $\mathfrak{p}_2\mathfrak{p}_2' = (2)$ and $\mathfrak{p}_3\mathfrak{p}_3' = (3)$, in $\mathrm{Cl}(K)$ we have the relations $[\mathfrak{p}_2'] = [\mathfrak{p}_2]^{-1}$ and $[\mathfrak{p}_3'] = [\mathfrak{p}_3]^{-1}$. Since $\mathrm{N}((\frac{1+\sqrt{-23}}{2})) = 6$, we can set $(\frac{1+\sqrt{-23}}{2}) = \mathfrak{p}_2\mathfrak{p}_3$. (This equation distinguishes $\mathfrak{p}_2$ from $\mathfrak{p}_2'$ and $\mathfrak{p}_3$ from $\mathfrak{p}_3'$, which up to this point have appeared

p	$T^2 - T + 6 \bmod p$	(p)
2	$T(T-1)$	$\mathfrak{p}_2\mathfrak{p}_2'$
3	$T(T-1)$	$\mathfrak{p}_3\mathfrak{p}_3'$
5	irreducible	(5)
7	irreducible	(7)
11	irreducible	(11)

Table 5.1: Factoring prime numbers in $\mathbf{Z}[\frac{1+\sqrt{-23}}{2}]$.

in symmetric roles.) So $[\mathfrak{p}_3] = [\mathfrak{p}_2]^{-1}$. Thus $\mathrm{Cl}(K) = \langle[\mathfrak{p}_2]\rangle$. Is $\mathfrak{p}_2$ principal? This ideal has norm 2, and for $m, n \in \mathbf{Z}$,

$$\mathrm{N}\left(m + n\frac{1+\sqrt{-23}}{2}\right) = \left(m + \frac{n}{2}\right)^2 + 23\left(\frac{n}{2}\right)^2,$$

which is never 2. (For nonzero n the norm is at least $23/4 > 2$ and for $n = 0$ the norm is a perfect square.) Since no $\alpha \in \mathcal{O}_K$ has norm 2, $[\mathfrak{p}_2] \neq 1$. Also

$$\mathrm{N}\left(1 + \frac{1+\sqrt{-23}}{2}\right) = \mathrm{N}\left(\frac{3}{2} + \frac{1}{2}\sqrt{-23}\right) = 8.$$

Since

$$\frac{1+\sqrt{-23}}{2} \equiv 0 \bmod \mathfrak{p}_2 \Longrightarrow 1 + \frac{1+\sqrt{-23}}{2} \equiv 1 \not\equiv 0 \bmod \mathfrak{p}_2,$$

we must have $(1 + \frac{1+\sqrt{-23}}{2}) = \mathfrak{p}_2'^3$, so $[\mathfrak{p}_2']^3 = 1$ and therefore $[\mathfrak{p}_2]^3 = 1$. This shows $[\mathfrak{p}_2]$ has order 3, so $\mathrm{Cl}(K) = \left\{[(1)], [\mathfrak{p}_2], [\mathfrak{p}_2^2]\right\}$. For any nonzero ideal $\mathfrak{a}$ in $\mathbf{Z}[\frac{1+\sqrt{-23}}{2}]$,

- $\mathfrak{a}^3$ is principal since $[\mathfrak{a}]^3 = 1$,
- either $\mathfrak{a}$ is principal or $\mathfrak{a}\mathfrak{p}_2$ is principal (if $[\mathfrak{a}] = [\mathfrak{p}_2^2]$) or $\mathfrak{a}\mathfrak{p}_2^2$ is principal (if $[\mathfrak{a}] = [\mathfrak{p}_2]$) and only one of these can happen.

The next two examples show the Kronecker bound C can get big for fields of small degree.

Example 5.12. Let $K = \mathbf{Q}(\alpha)$ where $\alpha^3 - \alpha - 1 = 0$, so $\mathcal{O}_K = \mathbf{Z}[\alpha]$ (Example 3.45). Using the $\mathbf{Z}$-basis $\left\{1, \alpha, \alpha^2\right\}$, $C \approx 28.08$.

Example 5.13. Let $K = \mathbf{Q}(\beta)$ where $\beta^5 - \beta - 1 = 0$, so $\mathcal{O}_K = \mathbf{Z}[\beta]$ (Example 3.27). Using the $\mathbf{Z}$-basis $\left\{1, \beta, \beta^2, \beta^3, \beta^4\right\}$, $C \approx 3454.4$.

By changing the $\mathbf{Z}$-basis we can get some savings in C.

Example 5.14. In $\mathbf{Z}[\sqrt{103}]$ with $\mathbf{Z}$-basis $\{1, \sqrt{103}\}$, $C \approx 124.29$. Replacing $\sqrt{103} \approx 10.14$ with the smaller number $\sqrt{103} - 10$ gives us a $\mathbf{Z}$-basis $\{1, \sqrt{103} - 10\}$ for which $C \approx 24.29$.

In Table 5.2 we list the first squarefree positive and negative d for which the quadratic field $\mathbf{Q}(\sqrt{d})$ has each possible class group structure from sizes 2 to 9. For example, the first imaginary quadratic field $\mathbf{Q}(\sqrt{d})$, ordered by $|d|$, whose class group is a product of two groups of order 2 occurs when $d = -21$.

Group	$\mathbf{Z}/2\mathbf{Z}$	$\mathbf{Z}/3\mathbf{Z}$	$\mathbf{Z}/4\mathbf{Z}$	$(\mathbf{Z}/2\mathbf{Z})^2$	$\mathbf{Z}/5\mathbf{Z}$	$\mathbf{Z}/6\mathbf{Z}$
$d > 0$	10	79	82	130	401	235
$d < 0$	-5	-23	-14	-21	-47	-26

Group	$\mathbf{Z}/7\mathbf{Z}$	$\mathbf{Z}/8\mathbf{Z}$	$\mathbf{Z}/4\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$	$(\mathbf{Z}/2\mathbf{Z})^3$	$\mathbf{Z}/9\mathbf{Z}$	$(\mathbf{Z}/3\mathbf{Z})^2$
$d > 0$	577	226	399	1155	1129	32009
$d < 0$	-71	-41	-65	-105	-199	-4027

Table 5.2: Quadratic fields $\mathbf{Q}(\sqrt{d})$ with particular class groups.

5.4 The Class Number

The number of ideal classes in a number field K (really, the number of ideal classes in $\mathcal{O}_K$) is called the *class number* of K and is written $h(K)$.[2] Saying $h(K) = 1$ is another way of saying $\mathcal{O}_K$ is a PID. We know that $h(\mathbf{Q}) = 1$, $h(\mathbf{Q}(i)) = 1$, $h(\mathbf{Q}(\sqrt{-5})) = 2$, and $h(\mathbf{Q}(\sqrt{-23})) = 3$.

The importance of class numbers in Diophantine equations is illustrated by Kummer's work on Fermat's last theorem. This was Fermat's notorious claim that he could show, by a marvelous proof that didn't fit in the margin, that the equation $x^n + y^n = z^n$ has no solution in positive integers x, y, and z when $n \geqslant 3$. If this is true for an exponent n then it is true for any multiple of n. Every number $n \geqslant 3$ is divisible by an odd prime or by 4, so it suffices to focus on these exponents. Fermat himself had settled the case $n = 4$, so suppose $x^p + y^p = z^p$ where p is an odd prime and x, y, and z are positive integers. Any common factor of x or y is a factor of z and its p-th power can be cancelled from all the terms, so we may assume $(x, y) = 1$. The sum of pth powers on the

[2] The use of h as the notation for the number of ideal classes goes back to Dirichlet (1838).

left side can be factored using pth roots of unity:

$$\prod_{i=0}^{p-1}(x+\zeta_p^i y) = z^p \tag{5.8}$$

where ζ_p is a nontrivial pth root of unity. This equation is in $\mathbf{Z}[\zeta_p]$, which we will see in Section 6.6 is the ring of integers of $\mathbf{Q}(\zeta_p)$, although Kummer did not know this; for him $\mathbf{Z}[\zeta_p]$ was just the natural ring to work in for this problem.

If $\mathbf{Z}[\zeta_p]$ has unique factorization and the factors on the left side of (5.8) are pairwise relatively prime, then each factor is a pth power up to unit multiple: $x+\zeta_p^i y = u_i w_i^p$. For $x+\zeta_p^i y$ to be nearly a pth power for all i from 0 to $p-1$ seems like a very strong condition to impose on two integers x and y, so one can anticipate that there should be a contradiction from this. Kummer devised a method to make this intuition precise, but he knew that it was not automatic for the factors $x+\zeta_p^i y$ to be relatively prime and he also discovered that the assumption of unique factorization is wrong when $p = 23$. He found a class number hypothesis which allowed him to get around these problems.

Theorem 5.15 (Kummer, 1847). *If p is an odd prime and $p \nmid h(\mathbf{Q}(\zeta_p))$ then $x^p + y^p = z^p$ has no solution in positive integers x, y, z.*

The importance of p not dividing $h(\mathbf{Q}(\zeta_p))$ for Kummer was similar to the importance of 3 not dividing $h(\mathbf{Q}(\sqrt{-5}))$ in the proof of Theorem **??**: if $p \nmid h(\mathbf{Q}(\zeta_p))$ then an ideal in $\mathbf{Z}[\zeta_p]$ whose pth power is principal has to be principal, which is important if we want to convert (5.8) into an equation with ideals and later come back to recover information about numbers. A proof of Theorem 5.15 is in [6, pp. 223–224, 378–381]. It is not easy and requires subtle properties of units in $\mathbf{Z}[\zeta_p]$. For comparison, $\mathbf{Z}[\sqrt{-5}]$ has units ± 1 so there are no unit problems in Theorem **??**.

It turns out that $h(\mathbf{Q}(\zeta_p)) = 1$ for $p \leqslant 19$ and Table 5.3 lists $h(\mathbf{Q}(\zeta_p))$ for all the remaining primes p below 50. We see 37 is the only prime in this range which does not fit the hypothesis in Kummer's theorem, so Kummer had proved Fermat's last theorem for every prime exponent below 50 other than $p = 37$, which was a striking achievement compared to other work on Fermat's last theorem at the time. Before Kummer, the only settled cases were $p = 3, 5$, and 7. (Kummer did *not* actually compute all the class numbers in Table 5.3. He found a method to decide if $p \nmid h(\mathbf{Q}(\zeta_p))$ which is simpler to carry out by hand than computing $h(\mathbf{Q}(\zeta_p))$.)

p	23	29	31	37	41	43	47
$h(\mathbf{Q}(\zeta_p))$	3	2^3	3^2	37	11^2	211	$5 \cdot 139$

Table 5.3: Class number of $\mathbf{Q}(\zeta_p)$.

The class numbers in Table 5.3 are growing, and Kummer conjectured that $h(\mathbf{Q}(\zeta_p)) = 1$ only for $p \leqslant 19$. This was proved independently by Montgomery and Uchida in 1971. Ultimately Fermat's last theorem was settled completely by Wiles and Taylor [59], [62] using techniques that make no use whatsoever of factorizations like (5.8).

There are many open questions about ideal class groups of number fields. Here are a few of them, which are all believed to have the answer "yes."

1. Are there infinitely many number fields with class number 1? It has been suggested that the number fields $\mathbf{Q}(\zeta_{2^n} + \zeta_{2^n}^{-1})$ might all have class number 1. Weber showed 2 is not a factor of any class number in this tower. Fukuda and Komatsu [20] showed a prime factor of any class number in this tower is greater than 10^8.

2. Are there infinitely many real quadratic fields with class number 1? This goes back to Gauss. The data suggest that about 76% of $\mathbf{Q}(\sqrt{p})$ with prime p have class number 1. In contrast to real quadratic fields, it is known by work of Baker [3], Heegner [24], and Stark [56] that there are only 9 imaginary quadratic fields with class number 1: $\mathbf{Q}(\sqrt{d})$ for $d = -1, -2, -3, -7, -11, -19, -43, -67, -163$. The strikingly different behavior in class numbers of real and imaginary quadratic fields is related to the unit group being infinite in the real quadratic case and finite in the imaginary quadratic case. It's usually hard to separate the study of ideal class groups and unit groups (*e.g.*, to prove an ideal is nonprincipal we usually need to know about the units, as in the proof of Theorem 4.51), but in the imaginary quadratic case the unit group is finite and explicitly known.

3. Are there infinitely many quadratic fields where the odd part of the class group (the subgroup of elements with odd order) is cyclic? (The 2-Sylow subgroup of the class group of a quadratic field is generally not cyclic; for instance, see Exercise 5.12.) In examples, the odd part of the class group shows a definite bias for being cyclic. Notice, for instance, how much larger

$|d|$ is in Table 5.2 when $\mathbf{Q}(\sqrt{d})$ first has a noncyclic class group of size 9 compared to the first cyclic class group of size 9. The Cohen–Lenstra heuristics [12] give precise conjectures about the frequency with which the odd part of the class group of a quadratic field has specific structural properties (*e.g.*, being cyclic, having order divisible by a particular prime, having a particular p-Sylow subgroup). Their heuristics were extended to class groups of higher-degree number fields by Cohen and Martinet [13]. Numerical data once cast some doubt on the higher-degree heuristics, but a special case of the Cohen–Martinet heuristics was proved by Bhargava [4].

4. Does every finite abelian group arise (up to isomorphism) as the class group of some number field? Table 5.2 might suggest that every finite abelian group could be the class group of a real and imaginary quadratic field, but this is not true: Chowla [40, p. 447] showed $(\mathbf{Z}/2\mathbf{Z})^n$ is not the class group of any imaginary quadratic field for all large n (probably $n \geqslant 5$ is enough; all $n < 5$ occur) and Shanks [53] showed no imaginary quadratic field has class group $(\mathbf{Z}/5\mathbf{Z})^2$, $(\mathbf{Z}/7\mathbf{Z})^2$, or $(\mathbf{Z}/11\mathbf{Z})^2$. The real quadratic case is different, *e.g.*, the Cohen–Lenstra heuristics predict that any finite abelian group of odd order is the odd part of the class group of infinitely many real quadratic fields.

5. Are there infinitely regular primes? A prime p is called *regular* if $p \nmid h(\mathbf{Q}(\zeta_p))$. These are the primes to which Kummer's work on Fermat's last theorem can be applied. Below 100 all primes are regular except for 37, 59, and 67. (The fields $\mathbf{Q}(\zeta_{59})$, and $\mathbf{Q}(\zeta_{67})$ have class numbers $3 \cdot 59 \cdot 233$ and $67 \cdot 12739$.) It is known that there are infinitely many irregular primes [6, pp. 381–382], and all the numerical data suggest regular primes appear more often than irregular primes, but there is no proof that there actually are infinitely many regular primes.

6. For every prime p, does the maximal real subfield $\mathbf{Q}(\zeta_p + \zeta_p^{-1})$ of $\mathbf{Q}(\zeta_p)$ have class number not divisible by p? This is Vandiver's conjecture. It has been checked into the millions, although probabilistic heuristics suggest counterexamples would occur only rarely, so the lack of counterexamples so far is not yet compelling.

The behavior of class numbers in towers of number fields is not straightforward. For cyclotomic fields, there is divisibility in towers: if $\mathbf{Q}(\zeta_m) \subset \mathbf{Q}(\zeta_n)$

then $h(\mathbf{Q}(\zeta_m)) \mid h(\mathbf{Q}(\zeta_n))$. But in general if $K \subset L$ there is no reason to expect $h(K) \mid h(L)$, or even $h(K) \leqslant h(L)$. For instance, $h(\mathbf{Q}(\sqrt{-5})) = 2$ but $h(\mathbf{Q}(i, \sqrt{-5})) = 1$ (Exercise 7.26).

There was once a famous open question, going back to Fürtwangler in the 1920s, asking: is every number field K a subfield of a number field with class number 1? The belief for many years was that the answer is yes, but in 1964 Golod and Shafarevich [21] showed the answer is no and they gave explicit counterexamples among imaginary quadratic fields K. Brumer [7], Kuzmin [34], and Roquette and Zassenhaus [46] extended this work and it turns out that there are infinitely many counterexamples K in every degree greater than 1.

5.5 Exercises

1. a) For a prime $p \neq 5$ such that $-5 \equiv \Box \bmod p$, show $p = x^2 + 5y^2 \Leftrightarrow p \equiv 1, 4 \bmod 5$, while $2p = x^2 + 5y^2 \Leftrightarrow p \equiv 2, 3 \bmod 5$.

   b) Here is a complete list of $p < 100$ for which $-5 \equiv \Box \bmod p$:

   $$3,\ 7,\ 23,\ 29,\ 41,\ 43,\ 47,\ 61,\ 67,\ 83,\ 89.$$

   Write either p or $2p$ in the form $x^2 + 5y^2$, and for primes of the second type write p in the form $2x^2 + 2xy + 3y^2$.

2. Use the Kronecker bound to show $\mathbf{Q}(i)$ and $\mathbf{Q}(\sqrt{2})$ have class number 1.

3. a) Show $\mathbf{Q}(\sqrt{-6})$ has class number 2.

   b) Show $-6 \equiv \Box \bmod p$ if and only if $p = x^2 + 6y^2$ or $2p = x^2 + 6y^2$ for some x and y in $\mathbf{Z}$, but not both.

   c) Show $2p = x^2 + 6y^2$ for some integers x and y if and only if $p = 2x'^2 + 3y'^2$ for some integers x' and y'.

   d) For any prime $p \neq 2$, show $p = x^2 + 6y^2 \Leftrightarrow -6 \equiv \Box \bmod p$ and $p \equiv 1, 7 \bmod 8$, while $p = 2x^2 + 3y^2 \Leftrightarrow -6 \equiv \Box \bmod p$ and $p \equiv 3, 5 \bmod 8$.

4. Let $K = \mathbf{Q}(\sqrt{229})$, so $\mathcal{O}_K = \mathbf{Z}[\frac{1+\sqrt{229}}{2}]$. Factor $(2), (3), (5)$, and (7) into prime ideals in $\mathcal{O}_K$. For $1 \leqslant a \leqslant 10$, factor $\mathrm{N}(a + \sqrt{229}) = |a^2 - 229|$ and use some of the data you find to show the subgroup of the ideal class group generated by primes dividing 2, 3, 5, 7, and 11 is generated by a prime (either prime) dividing 3.

5. Show $\mathbf{Q}(\sqrt{14})$ has class number 1.

6. Show $\mathbf{Q}(\sqrt{-15})$ has class number 2. (Note $-15 \equiv 1 \bmod 4$.)

7. Show $\mathbf{Q}(\sqrt{-11})$ has class number 1 and use this to show the integral solutions to $y^2 = x^3 - 11$ are $(3, \pm 4)$ and $(15, \pm 58)$.

8. Show $\mathbf{Q}(\sqrt{10})$ has class number 2.

9. Show $\mathbf{Q}(\sqrt{-13})$ has class number 2 and use this to show the integral solutions to $y^2 = x^3 - 13$ are $(17, \pm 70)$.

10. Show $\mathbf{Q}(\sqrt{-14})$ has a class group that is cyclic of order 4, generated by the ideal class of a prime ideal dividing 3.

11. Let K be a quadratic field and $\mathrm{Gal}(K/\mathbf{Q}) = \{1, \sigma\}$. For any nonzero ideal $\mathfrak{a} \subset \mathcal{O}_K$, write $\sigma(\mathfrak{a}) = \{\sigma(\alpha) : \alpha \in \mathfrak{a}\}$. This is called the conjugate ideal.

    a) Show $\sigma(\mathfrak{a})$ is an ideal, for principal ideals $\sigma((\alpha)) = (\sigma(\alpha))$, and $\sigma(\mathfrak{a}\mathfrak{b}) = \sigma(\mathfrak{a})\sigma(\mathfrak{b})$.

    b) In $\mathbf{Z}[i]$ and $\mathbf{Z}[\sqrt{-5}]$, find examples of ideals where $\sigma(\mathfrak{a}) = \mathfrak{a}$ but $\mathfrak{a}$ cannot be generated by an ordinary integer.

    c) Show $\mathfrak{a}\sigma(\mathfrak{a}) = \mathrm{N}(\mathfrak{a})\mathcal{O}_K$, so $[\mathfrak{a}]^{-1} = [\sigma(\mathfrak{a})]$ in $\mathrm{Cl}(K)$. (Hint: Both sides are multiplicative in $\mathfrak{a}$. Compare the hth power of both sides, where $h = h(K)$.)

12. Tables of class numbers of imaginary quadratic fields are filled mostly with even numbers. (See [6, p. 425].) This exercise explains why: an imaginary quadratic field with odd class number must be $\mathbf{Q}(i)$ or $\mathbf{Q}(\sqrt{-p})$ for a prime $p \equiv 3 \bmod 4$.

    Let $n = p_1 \cdots p_t$ be a product of t distinct primes and $K = \mathbf{Q}(\sqrt{-n})$. Every imaginary quadratic field except $\mathbf{Q}(i)$ looks like this. Set $\mathfrak{p}_i = (p_i, \sqrt{-n})$ in $\mathcal{O}_K$.

    a) Show $\mathfrak{p}_i^2 = (p_i)$ for all i, so $[\mathfrak{p}_i]^2 = 1$ in $\mathrm{Cl}(K)$.

    b) If $t \geqslant 2$, show no element of $\mathcal{O}_K$ has norm p_i, so each ideal class $[\mathfrak{p}_i]$ has order 2 and K has even class number. (If you take separate cases according as $-n \equiv 1 \bmod 4$ or $-n \not\equiv 1 \bmod 4$, keep Remark 1.22 in mind.)

    c) If $t = 1$, so $K = \mathbf{Q}(\sqrt{-p})$ for a prime number p, the method of part b doesn't produce an ideal class of order 2 since $(p, \sqrt{-p}) = (\sqrt{-p})$ is principal. If $p \equiv 1 \bmod 4$ show the ideal $\mathfrak{p} = (2, 1+\sqrt{-p})$ satisfies $\mathfrak{p}^2 = (2)$ and $\mathfrak{p}$ is not principal. (Therefore $[\mathfrak{p}]$ has order 2 and K has even class number when $p \equiv 1 \bmod 4$.)

d) Show the product of the ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_t$ is principal, but no nonempty proper subset has a principal product, so the subgroup generated by the ideal classes $[\mathfrak{p}_i]$ in $\mathrm{Cl}(\mathbf{Q}(\sqrt{-n}))$ is isomorphic to $(\mathbf{Z}/2\mathbf{Z})^{t-1}$. If $n \equiv 1 \bmod 4$, show the group generated by the ideal classes $[\mathfrak{p}_1], \dots, [\mathfrak{p}_t], [(2, 1+\sqrt{-n})]$ is isomorphic to $(\mathbf{Z}/2\mathbf{Z})^t$. (This turns out to be the 2-torsion subgroup of the class group, but don't show that.)

Note. This exercise shows that when $N > 1$ is not a prime power, so it has at least two different prime factors, the class group of $\mathbf{Q}(\sqrt{-N})$ has elements of order 2. There is a relationship between elements of order 2 in this class group and nontrivial factorizations of the squarefree part of N. Shanks [15, Sect. 5.6.4] devised an algorithm to factor N based on this.

13. Let $\mathbf{F}$ be a finite field. We want to show the integral closure of $\mathbf{F}[X]$ in any finite extension of $\mathbf{F}(X)$ has a finite ideal class group. (It is a Dedekind domain by Corollary 4.61.) The proof that number fields have finite ideal class groups doesn't appear to work in the function field case, at first glance, since function fields over finite fields don't have embeddings into $\mathbf{C}$. Fill in the details of the following alternate approach to the number field case which carries over to the function field case.

a) Let K be a number field and $\{e_1, \dots, e_n\}$ be a $\mathbf{Z}$-basis of $\mathcal{O}_K$. Show the norm map from K to $\mathbf{Q}$ is a homogeneous polynomial function in the coordinates of this basis: for $c_1, \dots, c_n \in \mathbf{Q}$, $\mathrm{N}_{K/\mathbf{Q}}(c_1e_1 + \cdots + c_ne_n) = P(c_1, \dots, c_n)$ where $P(X_1, \dots, X_n)$ is homogeneous of degree n with integral coefficients. This is not a surprise: $\mathrm{N}_{\mathbf{Q}(\sqrt{d})/\mathbf{Q}}(x + y\sqrt{d}) = x^2 - dy^2$ is homogeneous of degree 2 and (2.3) gives a homogeneous cubic polynomial formula for the norm on $\mathbf{Q}(\sqrt[3]{d})/\mathbf{Q}$. (Hint: $\mathrm{N}_{K/\mathbf{Q}}(c_1e_1 + \cdots + c_ne_n) = \det(\sum_{i=1}^n c_i[m_{e_i}])$.)

b) Show $|P(c_1, \dots, c_n)| \leqslant (\max|c_i|)^n ||P||$, where $||P||$ is the sum of the absolute values of the coefficients of P.

c) For the $\mathbf{Z}$-basis $\{1, \sqrt{10}\}$ of $\mathbf{Z}[\sqrt{10}]$, show $||P|| = 11$, while for the $\mathbf{Z}$-basis $\{1, \sqrt{10} - 3\}$ show $||P|| = 8$. Check these are smaller than the Kronecker bounds for these two $\mathbf{Z}$-bases.

d) Show the proof of the finiteness of $\mathrm{Cl}(K)$ (Theorem 5.3) goes through using $||P||$ in place of the Kronecker bound.

e) Now let K denote a finite extension of $\mathbf{F}(X)$, rather than a number field, and let R be the integral closure of $\mathbf{F}[X]$ in K. Show $\mathrm{Cl}(R)$ is finite. (Hint: For a nonzero ideal $\mathfrak{a}$ in R, place $[R : \mathfrak{a}]$ between two consecutive powers of q^n, where $q = \#\mathbf{F}$ and $n = [K : \mathbf{F}(X)]$. The norm properties in Exercise 4.31 are useful. Note also that $\deg(g + h) \leqslant \max(\deg g, \deg h)$ in $\mathbf{F}[X]$, which is stronger than $\deg(g + h) \leqslant \deg g + \deg h$, so the $\mathbf{F}[X]$-analogue of $||P||$ in part b should be a maximum and not a sum. Don't confuse $\deg g$ and $q^{\deg g}$.)

f) The ring $\mathbf{F}_2[x, y]$, where $y^2 - y = x^3$, is integrally closed by Exercise 2.11. Show its ideal class group has order 3 and is generated by the ideal class of either prime lying over x. (First show the class group is generated by prime ideals lying over irreducibles in $\mathbf{F}_2[x]$ of degree at most 3. Use the $\mathbf{F}_2[x]$-valued norm from Exercise 4.31 to determine how these irreducibles, as well as y, $x + y$, and $x + 1 + y$, decompose into prime ideals in $\mathbf{F}_2[x, y]$ in order to get relations in the class group.)

g) The ring $\mathbf{F}_3[x, y]$, where $y^2 = x^3 - x$, is integrally closed by Exercise 2.10. Show its ideal class group is a product of two groups of order 2 and is generated by the ideal classes of a prime lying over x and a prime lying over $x + 1$ (or $x - 1$).

14. In a UFD, if a product of two nonzero elements without a common factor (besides units) is equal to an nth power, for $n > 1$, the two elements are themselves nth powers up to unit multiple.

    Show that in any $\mathcal{O}_K$ which is not a UFD, there will always be a counterexample: there is some integer $n > 1$ and nonzero α and β in $\mathcal{O}_K$ without a common factor such that $\alpha\beta = \gamma^n$ for some γ in $\mathcal{O}_K$ but α and β are not nth powers up to unit multiple in $\mathcal{O}_K$. (Hint: Translate this into a problem about ideals and the ideal class group. Let $n = p$ be a prime dividing the size of the ideal class group and use ideals of order p in the ideal class group.)

15. If A is a UFD, show a fractional A-ideal $\mathfrak{a}$ is invertible if and only if it is principal. In particular, this gives another proof that a Dedekind domain which is a UFD is a PID. (Hint: By scaling we may assume $\mathfrak{a} \subset A$. If $\mathfrak{a}$ is invertible, by Lemma 5.5 $\mathfrak{a}$ is finitely generated and we can write

$\mathfrak{a} = Ax_1 + \cdots + Ax_k$ and $\mathfrak{a}^{-1} = Ay_1 + \cdots + Ay_k$. Write $y_i = a_i/b_i$ in lowest terms and show $\mathfrak{a} = bA$ where b is the least common multiple of $b_1, \ldots, b_k$.)

16. When $\mathrm{SL}_2(\mathbf{Z}[\sqrt{-6}])$ acts on $\mathbf{P}^1(\mathbf{Q}(\sqrt{-6}))$, determine if $[2+\sqrt{-6}, 5]$ and $[2, \sqrt{-6}]$ are in the same orbit, and if so find a matrix taking one point to the other.

17. Let $L = \mathbf{Q}(\sqrt{2}, \sqrt{-5})$.

    a) Show $\alpha = \frac{1+\sqrt{-5}}{\sqrt{2}} \in \mathcal{O}_L$ and $\{1, \alpha\}$ is a $\mathbf{Z}[\sqrt{2}]$-basis of $\mathcal{O}_L$.

    b) Show $\mathcal{O}_L = \mathbf{Z}[\sqrt{-5}] + \frac{1}{\sqrt{2}}\mathfrak{p}_2$, where $\mathfrak{p}_2 = (2, 1+\sqrt{-5})$, so the Steinitz class of $\mathcal{O}_L$ as a $\mathbf{Z}[\sqrt{-5}]$-module is $[\mathfrak{p}_2]$.

18. Let $d \geqslant 2$ be a squarefree positive integer and p be a prime not dividing d such that $p \equiv 3 \bmod 4$. Let $F = \mathbf{Q}(\sqrt{-dp})$ and $E = F(\sqrt{-p}) = \mathbf{Q}(\sqrt{-dp}, \sqrt{-p})$. Show
$$\mathcal{O}_E = \mathcal{O}_F e_1 \oplus \mathfrak{p} e_2,$$
where $e_1 = \frac{1+\sqrt{-p}}{2}$, $e_2 = \frac{1}{\sqrt{-p}}$, and $\mathfrak{p} = (p, \sqrt{-dp}) = p\mathcal{O}_F + \sqrt{-dp}\mathcal{O}_F$. The ideal $\mathfrak{p}$ is the unique prime over p in $\mathcal{O}_F$.

19. If F is a number field with class number 1 and E/F is a finite extension, show $\mathcal{O}_E$ has an $\mathcal{O}_F$-basis which contains the number 1.

20. Show Theorem 4.89 is equivalent to the following statement: for any nonzero ideal $\mathfrak{c}$ in a Dedekind domain A, each ideal class in $\mathrm{Cl}(A)$ contains an ideal relatively prime to $\mathfrak{c}$.

21. a) Suppose $\mathfrak{a}$ is a nonzero ideal in a Dedekind domain A such that $\mathfrak{a}^2$ is principal. Write $\mathfrak{a}^2 = (t_1)$ and $\mathfrak{a} = (t_1, t_2)$ with nonzero t_1 and t_2 (it's possible by Theorem 4.88). Show there are t_1' and t_2' in $\mathfrak{a}$ such that $t_1t_1' + t_2t_2' = t_1$ (the obvious choice $t_1' = 1$ and $t_2' = 0$ won't work if $\mathfrak{a} \neq (1)$ since we're requiring t_1' and t_2' come from $\mathfrak{a}$) and the matrix $\left(\begin{smallmatrix} t_1'/t_1 & t_2'/t_1 \\ -t_2/t_1 & 1 \end{smallmatrix}\right)$ provides an isomorphism from $\mathfrak{a} \oplus \mathfrak{a}$ to $A \oplus A$.

    b) In $\mathbf{Z}[\sqrt{-6}]$, let $\mathfrak{p} = (3, \sqrt{-6})$. Write down an explicit A-module isomorphism $\mathfrak{p} \oplus \mathfrak{p} \cong A \oplus A$.

22. Let A be a Dedekind domain with fraction field F, V be an F-vector space, and M be an A-module in V which is isomorphic to an A-fractional ideal $\mathfrak{a}$, say by $\varphi \colon M \to \mathfrak{a}$.

a) Pick any $v \neq 0$ in M, so $Av \subset M$. Show $\mathfrak{a}\mathfrak{b} = A\varphi(v)$ for some ideal $\mathfrak{b}$ in A.

b) Show $M = (\frac{1}{\varphi(v)}\mathfrak{a})v = \mathfrak{a}\left(\frac{1}{\varphi(v)}v\right)$. (The first formula shows M is a set of fractional multiples of a v, but the fractional ideal may be larger than A. The second formula shows M is a set of $\mathfrak{a}$-multiples of a vector, but the vector may not lie in M.)

23. a) Let A be a commutative ring. If P and P' are projective A-modules, show $P \oplus P'$ is a projective A-module. Is an arbitrary direct sum $\bigoplus_{i \in I} P_i$ of projective modules projective?

b) If A is a Dedekind domain and M is a finitely generated A-module, show $M \cong M_{\text{tor}} \oplus M/M_{\text{tor}}$, so the torsion submodule of M can be split off as a direct summand. (Hint: The A-module M/M_{tor} is finitely generated and torsion-free. Consider the natural surjective map $M \twoheadrightarrow M/M_{\text{tor}}$.)

c) The unit circle S^1, as an abelian group ($\mathbf{Z}$-module) has torsion subgroup T equal to the roots of unity. Show T is not a direct summand of S^1: $S^1 \ncong T \oplus G$ for any abelian group G.

24. (Staying relatively prime to an ideal)

Fix a nonzero ideal $\mathfrak{c}$ in a Dedekind domain A. We say a fractional A-ideal is relatively prime to $\mathfrak{c}$ if it has the form $\mathfrak{a}\mathfrak{b}^{-1}$ where $\mathfrak{a}$ and $\mathfrak{b}$ are ideals in A that are relatively prime to $\mathfrak{c}$ in the usual sense for ideals in A. (This is analogous to saying $9/7$ is relatively prime to 40, but not to 35.) Let $I(\mathfrak{c})$ be all the fractional A-ideals relatively prime to $\mathfrak{c}$ and $P(\mathfrak{c})$ be the principal fractional A-ideals relatively prime to $\mathfrak{c}$.

a) Show $I(\mathfrak{c})$ and $P(\mathfrak{c})$ are groups under multiplication.

b) Show the function $I(\mathfrak{c}) \to \text{Cl}(A)$ where $\mathfrak{a}\mathfrak{b}^{-1} \mapsto [\mathfrak{a}][\mathfrak{b}]^{-1}$ for $\mathfrak{a}$ and $\mathfrak{b}$ relatively prime to $\mathfrak{c}$ is a well-defined group homomorphism that is surjective with kernel $P(\mathfrak{c})$, so $I(\mathfrak{c})/P(\mathfrak{c}) \cong \text{Cl}(A)$. This shows the ideal class group of A can be constructed using only ideals relatively prime to a chosen ideal. (Look at the previous exercise.)

c) If $xA \in P(\mathfrak{c})$ show $x = a/b$ where a and b are in A and the ideals aA and bA are each relatively prime to $\mathfrak{c}$. (Hint: $\mathfrak{a}\mathfrak{b}^{-1} = (\mathfrak{a}\mathfrak{n})(\mathfrak{b}\mathfrak{n})^{-1}$ for any $\mathfrak{n}$.)

CHAPTER 6

RAMIFICATION

In analysis, the critical values of a smooth function $f(x)$ (those numbers $f(x_0)$ where $f'(x_0) = 0$) are of obvious importance. The analogue of this concept in a number field K is the primes p such that the ideal $p\mathcal{O}_K$ has a repeated prime ideal factor. Such p are called the ramified primes in K. They turn out to be the prime factors of the discriminant of K and we will see there is a close connection between ramification and Eisenstein polynomials. To illustrate the use of ramification as a special structure in number fields, we will use it to determine the ring of integers in cyclotomic fields and certain radical extensions of $\mathbf{Q}$.

6.1 Critical Values of Polynomials

If $f(x) \in \mathbf{R}[x]$, the solutions to $f'(x_0) = 0$ are the critical points of f, and the corresponding critical values $f(x_0)$ include all the maximum and minimum values of f. The terms critical point and critical value make sense for polynomials $f(z) \in \mathbf{C}[z]$, but what is their significance? There is no maximum and minimum for complex-valued functions.

Critical values of $f(z)$ are precisely those $c \in \mathbf{C}$ where the equation $f(z) = c$ has fewer solutions than usual. Suppose $f(z)$ has degree $n \geqslant 1$. The function $f\colon \mathbf{C} \to \mathbf{C}$ is surjective by the fundamental theorem of algebra: the equation

$f(z) = c$ is solvable for any $c \in \mathbf{C}$ since the polynomial $f(z) - c$ has a root. If we pick any $c \in \mathbf{C}$ and ask how often $f(z) = c$, the answer is at most n since the polynomial $f(z) - c$ has degree n.

Theorem 6.1. *When $n = \deg f \geqslant 1$ and $c \in \mathbf{C}$, the equation $f(z) = c$ has fewer than n solutions if and only if there is an $r \in \mathbf{C}$ such that $f(r) = c$ and $f'(r) = 0$. Such c exist when $n \geqslant 2$ and there are finitely many of them.*

Proof. The equation $f(z) = c$ has fewer than n solutions if and only if $f(z) - c$ has a multiple root. Suppose r is any root and write $f(z) - c = (z - r)g(z)$. Then r is a multiple root of $f(z) - c$ if and only if $g(r) = 0$. Differentiating the equation $f(z) - c = (z - r)g(z)$ and setting $z = r$, we get $f'(r) = g(r)$, so r is a multiple root of $f(z) - c$ if and only if $f(r) = c$ and $f'(r) = 0$. In particular, multiple roots of $f(z) - c$ are roots of $f'(z)$.

When $n \geqslant 2$, $f'(z)$ is a nonconstant polynomial, so it has a root in $\mathbf{C}$. When $f'(r) = 0$, r is a multiple root of $f(z) - f(r)$, so the equation $f(z) = f(r)$ has fewer than n solutions. Thus there exist $c \in \mathbf{C}$ such that $\#\{z : f(z) = c\} < n$, and there are finitely many such c since $f'(z)$ has finitely many roots. ■

Example 6.2. Let $f(z) = z^3 - 3z^2 + 1$, so $f'(z) = 3z^2 - 6z = 3z(z - 2)$ has roots 0 and 2. Since $f(0) = 1$ and $f(2) = -3$, $\#\{z : f(z) = c\} = 3$ when $c \neq 1$ or -3. When $c = 1$ or -3, there turn out to be only two solutions each. To find them, we factor

$$f(z) - 1 = z^3 - 3z^2 = z^2(z - 3)$$

and

$$f(z) - (-3) = z^3 - 3z^2 + 4 = (z - 2)^2(z + 1).$$

So $\{z : f(z) = 1\} = \{0, 3\}$ and $\{z : f(z) = -3\} = \{2, -1\}$.

What does this have to do with number theory? The factorization

$$f(z) - c = (z - r_1)^{e_1} \cdots (z - r_k)^{e_k} \tag{6.1}$$

as the number c varies (the roots r_i and the multiplicities e_i change with c) is analogous to the prime ideal factorization

$$p\mathcal{O}_K = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_g^{e_g} \tag{6.2}$$

as the prime number p varies. Table 6.1 uses the notations of (6.1) and (6.2) to describe analogous concepts in the polynomial and number field settings.

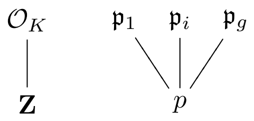
Polynomial	Number Field
$\deg f$	$[K:\mathbf{Q}]$
$c \in \mathbf{C}$	prime p
r_i such that $f(r_i) = c$	$\mathfrak{p}_i$ such that $\mathfrak{p}_i \cap \mathbf{Z} = p\mathbf{Z}$

Table 6.1: Comparing Polynomials and Number Fields.

The special numbers c where the equation $f(z) = c$ has fewer than n solutions, which are the c such that $f(z) - c$ has a multiple root (some e_i in (6.1) is greater than 1), have as a number-theoretic counterpart the primes p where $p\mathcal{O}_K$ has a multiple prime ideal factor (some e_i in (6.2) is greater than 1). Since in analysis it is important to understand the critical values of a function, you should believe it is important to know the primes p such that $p\mathcal{O}_K$ has a multiple prime ideal factor, even if you don't see yet what it could be good for.

6.2 Ramified and Unramified Primes

Let K be a number field and p be a prime number. The ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_g$ in (6.2) are the prime ideal factors of $(p) = p\mathcal{O}_K$. We say the $\mathfrak{p}_i$'s *lie over* p. The terminology is suggested by the following diagram.



Figures 6.1 and 6.2 are prime ideal diagrams for $\mathbf{Z}[\sqrt{-5}]$ and $\mathbf{Z}[\sqrt[3]{2}]$.

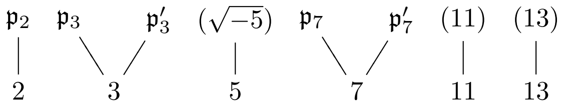


Figure 6.1: Primes lying over primes in $\mathbf{Z}[\sqrt{-5}]$.

Definition 6.3. In the notation of (6.2), if each $e_i = 1$ then we say p is *unramified* in K. If $e_i > 1$ for some i then we say p is *ramified* in K.

Saying p is unramified in K just means (p) has no multiple prime ideal factors. And p is ramified in K when (p) has a multiple prime ideal factor. If

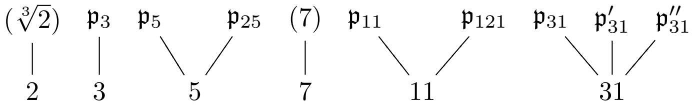


Figure 6.2: Primes lying over primes in $\mathbf{Z}[\sqrt[3]{2}]$.

we can link the factorization of (p) with the factorization of a polynomial $f(T)$ mod p, as in the Dedekind–Kummer factorization theorems, then p ramifies in K when $f(T)$ mod p has a multiple irreducible factor.

Example 6.4. Ramified primes in $\mathbf{Q}(i)$ are those p where T^2+1 mod p has a multiple factor. That only happens at 2, where $T^2+1 \equiv (T+1)^2$ mod 2, and 2 is ramified in $\mathbf{Q}(i)$: $(2) = (1+i)^2$. The primes 3 and 5 are both unramified in $\mathbf{Q}(i)$, with (3) being prime and $(5) = (1+2i)(1-2i)$.

Example 6.5. Ramified primes in $\mathbf{Q}(\sqrt{-5})$ are those p where T^2+5 mod p has a multiple factor. The only such p are 2 and 5: $T^2+5 \equiv (T+1)^2$ mod 2 and $T^2+5 \equiv T^2$ mod 5. So 2 and 5 are the only ramified primes in $\mathbf{Q}(\sqrt{-5})$, with $(2) = (2, 1+\sqrt{-5})^2$ and $(5) = (\sqrt{-5})^2$.

Example 6.6. Let $K = \mathbf{Q}(\alpha)$, where $\alpha^3 - \alpha - 1 = 0$. We saw in Example 3.45 that $\mathcal{O}_K = \mathbf{Z}[\alpha]$, so (p) factors in $\mathcal{O}_K$ the way $T^3 - T - 1$ factors mod p. See Table 6.2 for the factorizations of 5, 23, and 31. The primes 5 and 31 are unramified in K while 23 is ramified. One of the primes over 23 has an exponent larger than 1.

p	$T^3 - T - 1 \bmod p$	(p)
5	$(T-2)(T^2+2T+3)$	$\mathfrak{p}_5\mathfrak{p}_{25}$
23	$(T-3)(T-10)^2$	$\mathfrak{p}_{23}\mathfrak{p}'^2_{23}$
31	$(T-4)(T-7)(T-20)$	$\mathfrak{p}_{31}\mathfrak{p}'_{31}\mathfrak{p}''_{31}$

Table 6.2: Some primes factored in $\mathbf{Z}[\alpha]$, $\alpha^3 - \alpha - 1 = 0$.

To each prime ideal $\mathfrak{p}$ such that $\mathfrak{p} \mid (p)$, we associate two integers:

- the largest exponent $e \geqslant 1$ such that $\mathfrak{p}^e \mid (p)$. (We write $\mathfrak{p}^e || (p)$ to mean e is as large as possible, *e.g.*, $5^2 || 75$.) The exponent $e = e(\mathfrak{p}|p)$ is called the *ramification index* of $\mathfrak{p}$ over p.

- the exponent $f \geqslant 1$ such that $\mathrm{N}(\mathfrak{p}) = p^f$. We call $f = f(\mathfrak{p}|p)$ the *residue field degree* of $\mathfrak{p}$ over p since $f = \dim_{\mathbf{Z}/p\mathbf{Z}}(\mathcal{O}_K/\mathfrak{p})$.

In the notation $e(\mathfrak{p}|p)$ and $f(\mathfrak{p}|p)$, the vertical bar is just a separating device and doesn't stand for the divisibility relation, although it is true that $\mathfrak{p} \mid (p)$.

Example 6.7. In $\mathbf{Z}[\sqrt{-5}]$, $e(\mathfrak{p}_2|2) = 2$, $e(\mathfrak{p}_3|3) = 1$, and $f((11)|11) = 2$.

Example 6.8. In $\mathbf{Z}[\sqrt[3]{2}]$, $e(\mathfrak{p}_2|2) = 3$ and $f(\mathfrak{p}_5|5) = 1$.

For each prime p we have a set of numbers $e_1, \ldots, e_g$ and $f_1, \ldots, f_g$ associated to the prime ideals lying over p. There is a constraint linking them together. Taking ideal norms of both sides of (6.2), and setting $n = [K : \mathbf{Q}]$,

$$p^n = p^{e_1 f_1} \cdots p^{e_g f_g} = p^{e_1 f_1 + \cdots + e_g f_g} \Longrightarrow \boxed{n = e_1 f_1 + \cdots + e_g f_g}.$$

Let's express this fundamental identity without subscripts: for all primes p,

$$[K : \mathbf{Q}] = \sum_{\mathfrak{p}|(p)} e(\mathfrak{p}|p) f(\mathfrak{p}|p). \tag{6.3}$$

In particular, each ramification index e_i, residue field degree f_i, and the number of prime ideal factors g are all at most $[K : \mathbf{Q}]$.

Definition 6.9. In the notation of (6.2), we say

1. p is *inert* in K if (p) is prime $(g = 1, e_1 = 1, f_1 = n)$,

2. p is *split completely* in K if $(p) = \mathfrak{p}_1 \cdots \mathfrak{p}_n$ $(g = n$, all $e_i = 1$, all $f_i = 1)$,

3. p is *totally ramified* in K if $(p) = \mathfrak{p}^n$ $(g = 1,\ e_1 = n,\ f_1 = 1)$.

Example 6.10. In $\mathbf{Q}(i)$, 2 is totally ramified since $(2) = (1+i)^2$, while any prime $p \equiv 1 \bmod 4$ is split completely in $\mathbf{Q}(i)$ since $T^2 + 1 \bmod p$ has two different roots. A prime $p \equiv 3 \bmod 4$ is inert in $\mathbf{Q}(i)$ since $T^2 + 1 \bmod p$ is irreducible.

Example 6.11. In $\mathbf{Q}(\sqrt[3]{2})$, the primes 2 and 3 are totally ramified by Table 4.2.

Example 6.12. In $\mathbf{Q}(\alpha)$, where $\alpha^3 - \alpha - 1 = 0$, Table 6.2 says 23 is ramified but not totally ramified.

p	$T^4+2T^2+3T+1 \bmod p$	(p)	e_i's	f_i's
2	irreducible	(2)	1	4
3	$(T^2+1)^2$	$\mathfrak{p}_9^2$	2	2
5	irreducible	(5)	1	4
7	$(T-1)(T-4)(T^2+5T+2)$	$\mathfrak{p}_7\mathfrak{p}_7'\mathfrak{p}_{49}$	$1,1,1$	$1,1,2$
11	irreducible	(11)	1	4
13	$(T-10)^2(T^2+7T+3)$	$\mathfrak{p}_{13}^2\mathfrak{p}_{169}$	$2,1$	$1,2$
43	$(T-4)(T-9)(T-13)(T-17)$	$\mathfrak{p}_{43}\mathfrak{p}_{43}'\mathfrak{p}_{43}''\mathfrak{p}_{43}'''$	$1,1,1,1$	$1,1,1,1$

Table 6.3: Some e_i and f_i data in $\mathbf{Z}[\alpha]$, $\alpha^4+2\alpha^2+3\alpha+1=0$.

Example 6.13. Let $K=\mathbf{Q}(\alpha)$, where α is a root of T^4+2T^2+3T+1. Then $\mathcal{O}_K=\mathbf{Z}[\alpha]$ (Example 3.49) and we factor small primes in Table 6.3. The primes 3 and 13 are ramified in K, but not totally ramified. The other primes are unramified. The first prime which splits completely in K is 43.

Example 6.14. In Table 6.4, we use Table 4.7 to describe which primes are inert, split, or ramified in a quadratic field $\mathbf{Q}(\sqrt{d})$ for d a squarefree integer. (Inert and split primes are both unramified.) In a quadratic field any prime that is ramified is totally ramified, unlike 23 in the cubic field of Example 6.6 or 3 and 13 in the quartic field of Example 6.13.

(p)	Condition on d, $p\neq 2$	Condition on d, $p=2$
Inert: (p)	$d\not\equiv\square \bmod p$	$d\equiv 5 \bmod 8$
Split: $\mathfrak{p}\mathfrak{p}'$	$d\equiv\square \bmod p$	$d\equiv 1 \bmod 8$
Ramified: $\mathfrak{p}^2$	$p\mid d$	$d\equiv 2,3 \bmod 4$

Table 6.4: Ramification in $\mathbf{Q}(\sqrt{d})$, d squarefree.

6.3 Most Primes are Unramified

The ramified primes in a number field are analogous to the critical values of a polynomial in $\mathbf{C}[z]$. A polynomial in $\mathbf{C}[z]$ has finitely many critical values and we will now show a number field has only finitely many ramified primes. Equivalently, all but finitely many prime numbers are unramified in a number field. The argument depends on an interplay between discriminants in characteristic 0 and in characteristic p.

Theorem 6.15. *For monic $f(T) \in \mathbf{Z}[T]$ and a prime number p,*

$$\operatorname{disc}(f(T)) \bmod p = \operatorname{disc}(f(T) \bmod p).$$

Note Theorem 6.15 allows all monic polynomials, not just irreducible ones. Writing the discriminant of a polynomial in terms of its roots, the discriminant on the left side involves roots of a polynomial in characteristic 0 and the discriminant on the right side involves roots of a polynomial in characteristic p.

Proof. The number $\operatorname{disc}(f(T))$ is in $\mathbf{Z}$ and the number $\operatorname{disc}(f(T) \bmod p)$ is in $\mathbf{F}_p$. To check the reduction mod p of the first number is the second number, we will work in a number field where $f(T)$ has all of its roots. The roots are algebraic integers since $f(T)$ is monic.

Let K be a number field in which $f(T)$ splits completely:

$$f(T) = (T - \alpha_1) \cdots (T - \alpha_n) \quad \text{in } \mathcal{O}_K[T].$$

Let $\mathfrak{p}$ be a prime ideal in $\mathcal{O}_K$ lying over p, so if we reduce both sides mod $\mathfrak{p}$ then we have a factorization of $f(T) \bmod p$:

$$\overline{f}(T) = (T - \overline{\alpha}_1) \cdots (T - \overline{\alpha}_n) \quad \text{in } (\mathcal{O}_K/\mathfrak{p})[T].$$

Then

$$\begin{aligned} \operatorname{disc}(f(T)) \bmod \mathfrak{p} &= \prod_{i<j} (\alpha_j - \alpha_i)^2 \bmod \mathfrak{p} \\ &= \prod_{i<j} (\overline{\alpha}_j - \overline{\alpha}_i)^2 \\ &= \operatorname{disc}(f(T) \bmod p). \end{aligned}$$

Since $\mathfrak{p} \mid (p)$, the field $\mathcal{O}_K/\mathfrak{p}$ has characteristic p, so it contains $\mathbf{Z}/p\mathbf{Z}$. Since $\operatorname{disc}(f(T)) \in \mathbf{Z}$, its reduction mod $\mathfrak{p}$ lies in $\mathbf{Z}/p\mathbf{Z}$, so $\operatorname{disc}(f(T)) \bmod p = \operatorname{disc}(f(T) \bmod p)$. ∎

Theorem 6.16. *Let $K = \mathbf{Q}(\alpha)$, where α is an algebraic integer with minimal polynomial $f(T) \in \mathbf{Z}[T]$. Then for all but finitely many primes p, $f(T) \bmod p$ is separable, and these p are unramified in K. More precisely, if*

$$f(T) \equiv \pi_1(T) \cdots \pi_g(T) \bmod p,$$

where the $\pi_i(T)$'s are distinct monic irreducibles in $\mathbf{F}_p[T]$, then $(p) = \mathfrak{p}_1 \cdots \mathfrak{p}_g$ with $\mathrm{N}(\mathfrak{p}_i) = p^{\deg \pi_i}$.

Proof. The reduction $f(T) \bmod p$ is separable if and only if $\operatorname{disc}(f(T) \bmod p) \neq 0$, which is equivalent to $p \nmid \operatorname{disc}(f(T))$ by Theorem 6.15. The integer $\operatorname{disc}(f(T))$ is nonzero, so $p \nmid \operatorname{disc}(f(T))$ for all but finitely many p.

Since $\operatorname{disc}(f(T)) = \operatorname{disc}(\mathbf{Z}[\alpha]) = [\mathcal{O}_K : \mathbf{Z}[\alpha]]^2 \operatorname{disc}(K)$, if $p \nmid \operatorname{disc}(f(T))$ then p does not divide $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$, so Dedekind's factorization theorem lets us factor p in terms of the way $f(T) \bmod p$ factors. The lack of multiple irreducible factors of $f(T) \bmod p$ directly translates into the lack of multiple prime ideal factors of (p), so p is unramified in K. ■

Corollary 6.17. *Let $g(T) \in \mathbf{Z}[T]$ be monic and α be a root. If $g(T) \bmod p$ is separable then p is unramified in $\mathbf{Q}(\alpha)$. In particular, all but finitely many primes are unramified in $\mathbf{Q}(\alpha)$.*

Proof. We are not assuming $g(T)$ is irreducible. Let $f(T) \in \mathbf{Z}[T]$ be the minimal polynomial of α, so $f(T) \mid g(T)$ in $\mathbf{Z}[T]$. Write $g(T) = f(T)h(T)$. When $g(T) \bmod p$ is separable, also its factor $f(T) \bmod p$ is separable, so p is unramified in $\mathbf{Q}(\alpha)$ (and $p \nmid [\mathcal{O}_{\mathbf{Q}(\alpha)} : \mathbf{Z}[\alpha]]$) by Theorem 6.16. ■

Corollary 6.17 provides the easiest sufficient condition for a prime p to be unramified in a number field. The idea to take away from this is that separable polynomials mod p (which are polynomials with no multiple roots) are related to p being unramified (which means $p\mathcal{O}_K$ has no multiple prime ideal factor). We can check a polynomial is separable by checking it is relatively prime to its derivative.

Example 6.18. Let $K = \mathbf{Q}(\sqrt[n]{a})$ for some nonzero integer a. The notation $\sqrt[n]{a}$ denotes any fixed solution to $\alpha^n = a$. We are *not* assuming $T^n - a$ is irreducible. (We could use $a = 1$, for instance, with α an nth root of unity.)

Since $T^n - a$ has derivative nT^{n-1}, $T^n - a \bmod p$ is separable provided $p \nmid na$. Therefore any $p \nmid na$ is unramified in $\mathbf{Q}(\sqrt[n]{a})$. Whether p ramifies in $\mathbf{Q}(\sqrt[n]{a})$ if $p \mid na$ depends on the situation, *e.g.*, 2 is unramified in $\mathbf{Q}(\sqrt{5})$ but is ramified in $\mathbf{Q}(\sqrt{3})$ (see Table 6.4).

In the notation of Theorem 6.16, the way (p) factors in the integers of $\mathbf{Q}(\alpha)$ matches the way $f(T) \bmod p$ factors if $p \nmid \operatorname{disc}(f(T))$, but not generally if $p \mid \operatorname{disc}(f(T))$.

Example 6.19. Let $K = \mathbf{Q}(\sqrt{5})$. The polynomial $T^2 - 5$ has discriminant 20. Although $T^2 - 5 \equiv (T-1)^2 \bmod 2$, it is not true that $(2) = \mathfrak{p}^2$. In fact, (2) is prime (see Table 4.6).

Example 6.20. Let $K = \mathbf{Q}(\alpha)$ where α is a root of $T^3 - 12T + 2$, whose discriminant is 6804. Although $T^3 - 12T + 2 \equiv T^3 \bmod 2$, we are not justified in saying $(2) = \mathfrak{p}^3$ simply from the mod 2 factorization. But in Example 4.47 we showed $2 \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$, which shows Dedekind's theorem applies, so the factorization $(2) = \mathfrak{p}^3$ is in fact correct.

Example 6.21. Let $K = \mathbf{Q}(\sqrt[3]{10})$. The polynomial $T^3 - 10$ has discriminant 2700. Although $T^3 - 10 \equiv (T-1)^3 \bmod 3$, it is not true that $(3) = \mathfrak{p}^3$. We saw in Example 4.48 that $(3) = \mathfrak{p}_3^2\mathfrak{p}_3'$.

Example 6.22. The number $\alpha = \sqrt{10} + \sqrt{13}$ has minimal polynomial $f(T) = T^4 - 46T^2 + 9$ over $\mathbf{Q}$, whose discriminant is $2^{14} \cdot 3^2 \cdot 5^2 \cdot 13^2$, so any prime other than 2, 3, 5, or 13 is unramified in K. Let $K = \mathbf{Q}(\alpha)$. In Table 6.5 we factor $f(T) \bmod p$ for $p \leqslant 13$. It is separable modulo 7 and 11 but not modulo 2, 3, 5, or 13, so there is no justification for reading off how 2, 3, 5, or 13 factor in K based merely on the data in the table.

p	$T^4 - 46T^2 + 9 \bmod p$	(p)
2	$(T+1)^4$	?
3	$T^2(T-1)(T-2)$	?
5	$(T^2+2)^2$	?
7	$(T^2+1)(T^2+2)$	$\mathfrak{p}_{49}\mathfrak{p}_{49}'$
11	$(T^2+4)(T^2+5)$	$\mathfrak{p}_{121}\mathfrak{p}_{121}'$
13	$(T+6)^2(T+7)^2$	?

Table 6.5: Factoring primes in $\mathbf{Q}(\sqrt{10} + \sqrt{13})$.

The factorizations of 2, 3, 5, and 13 in K turn out to be

$$(2) = \mathfrak{p}_4^2, \quad (3) = \mathfrak{p}_3\mathfrak{p}_3'\mathfrak{p}_3''\mathfrak{p}_3''', \quad (5) = \mathfrak{p}_{25}^2, \quad (13) = \mathfrak{p}_{13}^2\mathfrak{p}_{13}'^2,$$

which happens to match the shape of $f(T) \bmod p$ for $p = 5$ and 13 but not for $p = 2$ and 3. In particular, $f(T) \bmod 3$ has a multiple irreducible factor, but 3 is unramified (even split completely) in K.

One way of proving two number fields are not isomorphic is to find a prime number whose ideal factorizations in the integers of the two fields have different shapes. Theorem 6.16 provides a simple method of finding such primes: writing the two fields as $\mathbf{Q}(\alpha)$ and $\mathbf{Q}(\beta)$ where α and β are algebraic integers with minimal polynomials $f(T)$ and $g(T)$ in $\mathbf{Z}[T]$, find a prime p where $f(T)$ and $g(T)$ are separable mod p with factorizations of different shapes.

Example 6.23. The polynomials $f(T) = T^3 + 6T - 6$ and $g(T) = T^3 - 6T - 10$ are both irreducible since they are Eisenstein at 2. Let $K = \mathbf{Q}(\alpha)$ and $L = \mathbf{Q}(\beta)$, where $f(\alpha) = 0$ and $g(\beta) = 0$. Theorem 6.16 and Table 6.6 tell us the prime 5 is unramified in K and L with different factorizations, so $K \not\cong L$. Although 7 is unramified in K and L, it factors in the same way so this doesn't distinguish K from L. We can't use Theorem 6.16 to read off the factorizations of (2) and (3) in K and L since the polynomial factorizations mod 2 and 3 have repeated factors.

p	$T^3 + 6T - 6 \bmod p$	$p\mathcal{O}_K$	$T^3 - 6T - 10 \bmod p$	$p\mathcal{O}_L$
2	T^3	?	T^3	?
3	T^3	?	$(T-1)^3$	?
5	$T(T+1)(T-1)$	$\mathfrak{p}_5\mathfrak{p}_5'\mathfrak{p}_5''$	irreducible	(5)
7	$(T+5)(T^2+2T+5)$	$\mathfrak{p}_7\mathfrak{p}_{49}$	$(T+5)(T^2+2T+3)$	$\mathfrak{q}_7\mathfrak{q}_{49}$

Table 6.6: Factoring primes in two cubic fields.

In practice the method of Example 6.23 works well to tell fields apart, but it does not always work: there are nonisomorphic number fields in which each prime number factors with the same shape. See Example 6.44.

6.4 Dedekind's Discriminant Theorem

Only finitely many primes ramify in a number field. Now we will determine precisely which primes these are.

Theorem 6.24 (Dedekind, 1882). *A prime p ramifies in K if and only if p divides* $\operatorname{disc}(K)$*.*

This theorem provides the principal theoretical significance of the discriminant. We will prove Theorem 6.24 in two cases: first if $\mathcal{O}_K = \mathbf{Z}[\alpha]$ and then the general case. Logically the first case is not necessary in the proof of the general case, but the first case is included because the general case is not easy and the first case provides a simpler insight into why ramification of primes is related to discriminants. As in the proofs of the Dedekind–Kummer factorization theorems, our proof of Theorem 6.24 will use the ring $\mathcal{O}_K/(p)$.

Proof. (Assuming $\mathcal{O}_K = \mathbf{Z}[\alpha]$) Let α have minimal polynomial $f(T) \in \mathbf{Z}[T]$, so $\mathbf{Z}[T]/(f(T)) \cong \mathcal{O}_K$ by sending T to α. Factor $f(T)$ modulo p, say $\overline{f}(T) = \pi_1^{e_1} \cdots \pi_g^{e_g}$. From Kummer's factorization theorem we know the exponents $e_1, \ldots, e_g$ appearing here are the same as in the prime ideal factorization of (p) in $\mathcal{O}_K$, so p ramifies in K if and only if $\overline{f}(T)$ has a multiple irreducible factor. The $\pi_i(T)$'s are all separable (every irreducible polynomial in $\mathbf{F}_p[T]$ is separable), so saying $\overline{f}(T)$ has a multiple irreducible factor is equivalent to saying $\overline{f}(T)$ has a multiple root (in a splitting field).[1] Equivalently, (p) has a multiple prime factor in $\mathcal{O}_K$ if and only if $\overline{f} = f(T) \bmod p$ has a multiple root. A polynomial has a multiple root if and only if its discriminant is 0. Therefore

$$p \text{ ramifies in } K \iff \operatorname{disc} \overline{f} = 0 \text{ in } \mathbf{F}_p.$$

By Theorem 6.15,

$$\operatorname{disc}(\overline{f}) = 0 \text{ in } \mathbf{F}_p \iff p \mid \operatorname{disc} f \text{ in } \mathbf{Z}.$$

By Theorem 3.25, since $\mathcal{O}_K = \mathbf{Z}[\alpha]$

$$\operatorname{disc} f = \operatorname{disc}(\mathbf{Z}[\alpha]) = \operatorname{disc}(K).$$

Putting everything together,

$$p \text{ ramifies in } K \iff p \mid \operatorname{disc}(K). \quad \blacksquare$$

In brief, the reason discriminants are related to ramification is that ramification in number fields occurs at prime numbers which have a *multiple prime ideal factor*, a polynomial has discriminant 0 exactly when it has a *multiple root*, and a number mod p is 0 exactly when the number is *divisible* by p.

The way we used the condition $\mathcal{O}_K = \mathbf{Z}[\alpha]$ in the proof of Theorem 6.24 is through the isomorphism $\mathbf{Z}[T]/(f(T)) \cong \mathcal{O}_K$ by $T \mapsto \alpha$. But what we really used is the mod p isomorphism $\mathbf{F}_p[T]/(\overline{f}) \cong \mathcal{O}_K/(p)$ by $T \mapsto \overline{\alpha}$, and that isomorphism holds under the condition that $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$, which is weaker than

[1] Compare with $T^p - u$ in $\mathbf{F}_p(u)[T]$, which is irreducible but has a multiple root in a splitting field: if $r^p - u = 0$ then $T^p - u = T^p - r^p = (T - r)^p$, so its only root is r.

$\mathcal{O}_K = \mathbf{Z}[\alpha]$. Moreover, $\operatorname{disc}(f(T)) = [\mathcal{O}_K : \mathbf{Z}[\alpha]]^2 \operatorname{disc}(K)$, so $p \mid \operatorname{disc}(f(T))$ if and only if $p \mid \operatorname{disc}(K)$ provided that $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$. Therefore our proof of Theorem 6.24 for a prime p goes through with the hypothesis $\mathcal{O}_K = \mathbf{Z}[\alpha]$ weakened to $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$. As long as to each prime p we can find an $\alpha_p \in \mathcal{O}_K$ such that $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha_p]]$, the above proof of Theorem 6.24 applies to p (using α_p in the role of α). Dedekind tried to prove there is an α_p for each p in any number field, and it must have been a shock to find the example of "Dedekind's field" where it isn't true (Theorem 4.49): $2 \mid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$ for all α. The possibility that $p \mid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$ for all α means a completely general proof of Theorem 6.24 has to abandon the special rings $\mathbf{Z}[\alpha]$ and polynomial discriminants. We will prove Theorem 6.24 in general using the reduction mod p of ring discriminants. (Ring discriminants over PIDs are defined in Section 3.7.)

By Theorem 3.66, the characteristic polynomial, trace, and norm of an algebraic integer in a number field K can be computed relative to any basis of the field extension $K/\mathbf{Q}$ or relative to any basis of the ring extension $\mathcal{O}_K/\mathbf{Z}$. The answers both ways are the same, so for any $\mathbf{Z}$-basis $\{x_1, \ldots, x_n\}$ of $\mathcal{O}_K$ we have

$$\operatorname{disc}(K) = \operatorname{disc}_{\mathbf{Z}}(\mathcal{O}_K) = \det(\operatorname{Tr}_{K/\mathbf{Q}}(x_i x_j)) = \det(\operatorname{Tr}_{\mathcal{O}_K/\mathbf{Z}}(x_i x_j)).$$

We want to justify saying

$$\operatorname{disc}_{\mathbf{Z}}(\mathcal{O}_K) \bmod p = \operatorname{disc}_{\mathbf{Z}/p\mathbf{Z}}(\mathcal{O}_K/(p)), \tag{6.4}$$

which is an analogue of Theorem 6.15 for ring discriminants.

Writing $\mathcal{O}_K = \mathbf{Z}x_1 \oplus \cdots \oplus \mathbf{Z}x_n$ we have $p\mathcal{O}_K = \mathbf{Z}px_1 \oplus \cdots \oplus \mathbf{Z}px_n$, so

$$\mathcal{O}_K/(p) \cong \bigoplus_{i=1}^{n} (\mathbf{Z}/p\mathbf{Z})\overline{x}_i,$$

which has dimension n over $\mathbf{Z}/p\mathbf{Z}$. This vector space decomposition doesn't tell us anything about the multiplicative structure of $\mathcal{O}_K/(p)$. Since (p) may not be prime in $\mathcal{O}_K$, $\mathcal{O}_K/(p)$ need not be a domain. For instance, $\mathbf{Z}[i]/(2)$ and $\mathbf{Z}[i]/(5)$ are not domains.

Theorem 6.25. *For $\alpha \in \mathcal{O}_K$ and p any prime number,*

$$\begin{aligned}\chi_{\mathcal{O}_K/\mathbf{Z},\alpha}(T) \bmod p &= \chi_{(\mathcal{O}_K/(p))/(\mathbf{Z}/p\mathbf{Z}),\overline{\alpha}}(T),\\ \mathrm{Tr}_{\mathcal{O}_K/\mathbf{Z}}(\alpha) \bmod p &= \mathrm{Tr}_{(\mathcal{O}_K/(p))/(\mathbf{Z}/p\mathbf{Z})}(\overline{\alpha}),\\ \mathrm{N}_{\mathcal{O}_K/\mathbf{Z}}(\alpha) \bmod p &= \mathrm{N}_{(\mathcal{O}_K/(p))/(\mathbf{Z}/p\mathbf{Z})}(\overline{\alpha}).\end{aligned}$$

Proof. Pick a **Z**-basis of $\mathcal{O}_K$, say $\{x_1, \ldots, x_n\}$. Write $\alpha x_j = c_{1j}x_1 + \cdots + c_{nj}x_n$ ($c_{ij} \in \mathbf{Z}$). Then in $\mathcal{O}_K/(p)$, $\overline{\alpha}\,\overline{x}_j = \overline{c}_{1j}\overline{x}_1 + \cdots + \overline{c}_{nj}\overline{x}_n$, so $[m_\alpha] = (c_{ij})$ and $[m_{\overline{\alpha}}] = (\overline{c}_{ij})$. That is really the whole point, since now we see the mod p reduction of $\chi_{\mathcal{O}_K/\mathbf{Z},\alpha}(T)$ is

$$\det(TI_n - (c_{ij})) \bmod p = \det(TI_n - (\overline{c}_{ij})),$$

which implies $\chi_{\mathcal{O}_K/\mathbf{Z},\alpha}(T) \bmod p = \chi_{(\mathcal{O}_K/(p))/(\mathbf{Z}/p\mathbf{Z}),\overline{\alpha}}(T)$. Equating suitable coefficients on both sides gives the formulas for the trace and norm. ■

Theorem 6.25 can be described as saying the characteristic polynomial, trace, and norm all commute with reduction mod p. More precisely, the diagrams

$$\begin{tikzcd}
\mathcal{O}_K \arrow[r, "\text{reduction}"] \arrow[d, "\chi"'] & \mathcal{O}_K/(p) \arrow[d, "\chi"] \\
\mathbf{Z}[T] \arrow[r, "\text{reduction}"'] & (\mathbf{Z}/p\mathbf{Z})[T]
\end{tikzcd}$$

$$\begin{tikzcd}
\mathcal{O}_K \arrow[r, "\text{reduction}"] \arrow[d, "\mathrm{Tr}"'] & \mathcal{O}_K/(p) \arrow[d, "\mathrm{Tr}"] \\
\mathbf{Z} \arrow[r, "\text{reduction}"'] & \mathbf{Z}/p\mathbf{Z}
\end{tikzcd}$$

and

$$\begin{tikzcd}
\mathcal{O}_K \arrow[r, "\text{reduction}"] \arrow[d, "\mathrm{N}"'] & \mathcal{O}_K/(p) \arrow[d, "\mathrm{N}"] \\
\mathbf{Z} \arrow[r, "\text{reduction}"'] & \mathbf{Z}/p\mathbf{Z}
\end{tikzcd}$$

all commute.

Theorem 6.26. *For any number field K and prime p, (6.4) is true:*

$$\mathrm{disc}_{\mathbf{Z}}(\mathcal{O}_K) \bmod p = \mathrm{disc}_{\mathbf{Z}/p\mathbf{Z}}(\mathcal{O}_K/(p)).$$

The right side is, by definition, the discriminant of any $\mathbf{Z}/p\mathbf{Z}$-basis of $\mathcal{O}_K/(p)$, so it is only well-defined up to multiplication by a nonzero square factor. The left side of the equation is a single number mod p since all $\mathbf{Z}$-bases of $\mathcal{O}_K$ have the same discriminant. So the meaning of this theorem is that the square class (nonzero square or nonsquare or 0) of $\operatorname{disc}_{\mathbf{Z}}(\mathcal{O}_K)$ mod p is the common square class of discriminants of all $\mathbf{Z}/p\mathbf{Z}$-bases of $\mathcal{O}_K/(p)$.

Proof. For a $\mathbf{Z}$-basis $\{x_i\}$ of $\mathcal{O}_K$, $\{x_i \bmod p\mathcal{O}_K\}$ is a $\mathbf{Z}/p\mathbf{Z}$-basis of $\mathcal{O}_K/(p)$, so

$$\begin{aligned}\operatorname{disc}_{\mathbf{Z}}(\mathcal{O}_K) \bmod p &= \det(\operatorname{Tr}_{\mathcal{O}_K/\mathbf{Z}}(x_ix_j)) \bmod p\\ &= \det(\operatorname{Tr}_{\mathcal{O}_K/\mathbf{Z}}(x_ix_j) \bmod p)\\ &= \det(\operatorname{Tr}_{(\mathcal{O}_K/(p))/(\mathbf{Z}/p\mathbf{Z})}(\overline{x}_i\overline{x}_j)) \quad \text{by Theorem 6.25}\\ &= \operatorname{disc}_{\mathbf{Z}/p\mathbf{Z}}(\mathcal{O}_K/(p)).\end{aligned}$$

■

Lemma 6.27. *Let A be a commutative ring and B_1 and B_2 be commutative ring extensions of A which are each finite free A-modules. Then,*

$$\operatorname{disc}_A(B_1 \times B_2) = \operatorname{disc}_A(B_1)\operatorname{disc}_A(B_2).$$

This equality is only up to unit (square) multiple in A, or it is an exact equality in A if we choose A-module bases of B_1, B_2, and $B_1 \times B_2$ compatibly, as we'll see in the proof.

Proof. Pick A-module bases for B_1 and B_2:

$$B_1 = \bigoplus_{i=1}^{m} Ae_i, \quad B_2 = \bigoplus_{j=1}^{n} Af_j.$$

As an A-module basis for $B_1 \times B_2$ we will use $\{e_1, \dots, e_m, f_1, \dots, f_n\}$. Since $e_if_j = 0$ in $B_1 \times B_2$, the matrix whose determinant is $\operatorname{disc}_A(B_1 \times B_2)$ is a block diagonal matrix

$$\begin{pmatrix} (\operatorname{Tr}_{(B_1\times B_2)/A}(e_ie_k)) & O \\ O & (\operatorname{Tr}_{(B_1\times B_2)/A}(f_jf_\ell)) \end{pmatrix}.$$

For any $x \in B_1$, multiplication by x on $B_1 \times B_2$ kills the B_2 component and acts on the B_1-component in the way x multiplies on B_1, so a matrix for multiplication by x on $B_1 \times B_2$ is a matrix whose upper left block is a matrix for multiplication by x on B_1 and other blocks are 0. Thus

$$\mathrm{Tr}_{(B_1\times B_2)/A}(x) = \mathrm{Tr}_{B_1/A}(x) \quad \text{for } x \in B_1.$$

Similarly, $\mathrm{Tr}_{(B_1\times B_2)/A}(x) = \mathrm{Tr}_{B_2/A}(x)$ for $x \in B_2$. Thus

$$\begin{aligned}\mathrm{disc}_A(B_1\times B_2) &= \det\begin{pmatrix} (\mathrm{Tr}_{B_1/A}(e_ie_k)) & O \\ O & (\mathrm{Tr}_{B_2/A}(f_jf_\ell)) \end{pmatrix} \\ &= \det(\mathrm{Tr}_{B_1/A}(e_ie_k))\det(\mathrm{Tr}_{B_2/A}(f_jf_\ell)) \\ &= \mathrm{disc}_A(B_1)\,\mathrm{disc}_A(B_2). \quad \blacksquare\end{aligned}$$

Now we are ready for the general proof of Dedekind's discriminant theorem.

Proof. Writing $(p) = \mathfrak{p}_1^{e_1}\cdots\mathfrak{p}_g^{e_g}$, the Chinese remainder theorem tells us

$$\mathcal{O}_K/(p) \cong \mathcal{O}_K/\mathfrak{p}_1^{e_1}\times\cdots\times\mathcal{O}_K/\mathfrak{p}_g^{e_g}. \tag{6.5}$$

Since $\mathfrak{p}_i^{e_i} \mid (p)$, $p = 0$ in each $\mathcal{O}_K/\mathfrak{p}_i^{e_i}$, so all the rings in (6.5) contain $\mathbf{F}_p$.

Since $\mathrm{disc}(K) = \mathrm{disc}_{\mathbf{Z}}(\mathcal{O}_K)$, Theorem 6.26 tells us $p \mid \mathrm{disc}(K)$ if and only if $\mathrm{disc}_{\mathbf{F}_p}(\mathcal{O}_K/(p)) = 0$. Using (6.5) and Lemma 6.27,

$$\mathrm{disc}_{\mathbf{F}_p}(\mathcal{O}_K/(p)) = \prod_{i=1}^{g}\mathrm{disc}_{\mathbf{F}_p}(\mathcal{O}_K/\mathfrak{p}_i^{e_i}). \tag{6.6}$$

For any prime-power ideal $\mathfrak{p}^e$ such that $\mathfrak{p}^e \mid (p)$, we will show $\mathrm{disc}_{\mathbf{F}_p}(\mathcal{O}_K/\mathfrak{p}^e) = 0$ if and only if $e > 1$. Therefore $\mathrm{disc}_{\mathbf{F}_p}(\mathcal{O}_K/(p)) = 0$ if and only if some $e_i > 1$, which means p ramifies.

Suppose $e > 1$. Any $x \in \mathfrak{p} - \mathfrak{p}^e$ reduces to a nonzero *nilpotent* element in $\mathcal{O}_K/\mathfrak{p}^e$: $\overline{x} \neq 0$ but $\overline{x}^e = 0$. We can extend $\overline{x}$ to an $\mathbf{F}_p$-basis of $\mathcal{O}_K/\mathfrak{p}^e$, say $\{\overline{x}_1, \ldots, \overline{x}_n\}$ with $\overline{x} = \overline{x}_1$. Writing the trace map $\mathrm{Tr}_{(\mathcal{O}_K/\mathfrak{p}^e)/\mathbf{F}_p}$ as Tr for short, the first column of the matrix $(\mathrm{Tr}(\overline{x}_i\overline{x}_j))$ contains the numbers $\mathrm{Tr}(\overline{x}_i\overline{x})$. These traces are all 0: $\overline{x}_i\overline{x}$ is nilpotent, so the linear transformation $m_{\overline{x}_i\overline{x}}$ on $\mathcal{O}_K/\mathfrak{p}^e$ is nilpotent and thus its eigenvalues all equal zero, so their sum (the trace) is 0. Since one column of the trace pairing matrix $(\mathrm{Tr}(\overline{x}_i\overline{x}_j))$ is all 0, $\mathrm{disc}_{\mathbf{F}_p}(\mathcal{O}_K/\mathfrak{p}^e) = 0$.

Now suppose $e = 1$. Then $\mathcal{O}_K/\mathfrak{p}^e = \mathcal{O}_K/\mathfrak{p}$ is a finite field of characteristic p. Extensions of finite fields are separable, so $\mathrm{disc}_{\mathbf{F}_p}(\mathcal{O}_K/\mathfrak{p}) \neq 0$ (Theorem 3.22). $\blacksquare$

Nilpotent elements played a role in this proof, since nilpotent linear maps are the simplest ones that are guaranteed to have trace 0. Nilpotency also turns out to be another way to think about what it means for p to to be ramified.

Theorem 6.28. *A prime number p is ramified in a number field K if and only if the ring $\mathcal{O}_K/(p)$ has a nonzero nilpotent element. Moreover, the following conditions on a prime number p are equivalent:*

(a) p is unramified in K,

(b) the ring $\mathcal{O}_K/(p)$ is isomorphic to a product of fields,

(c) the only nilpotent element in the ring $\mathcal{O}_K/(p)$ is 0.

Proof. Negating (a) and (c) gives us the equivalence of p being ramified with $\mathcal{O}_K/(p)$ having a nonzero nilpotent element.

When p is unramified in K, all $e_i = 1$ in (6.5), so $\mathcal{O}_K/(p)$ is isomorphic to a product of (finite) fields. In a product of fields the only nilpotent element is 0. So (a) $\Rightarrow$ (b) $\Rightarrow$ (c).

When p is ramified, some $e_i > 1$, so $\mathcal{O}_K/(p)$ is isomorphic to a product of rings including some $\mathcal{O}_K/\mathfrak{p}_i^{e_i}$ where $e_i > 1$. The ring $\mathcal{O}_K/\mathfrak{p}_i^{e_i}$ has a nonzero nilpotent element, namely the reduction mod $\mathfrak{p}_i^{e_i}$ of any $x \in \mathfrak{p}_i - \mathfrak{p}_i^{e_i}$. Then $\mathcal{O}_K/(p)$ has the nonzero nilpotent element $(0, \ldots, 0, x, 0, \ldots, 0)$. Thus (c) $\Rightarrow$ (a) (we proved the contrapositive). ∎

6.5 Total Ramification and Eisenstein Polynomials

Recall that p is called totally ramified in K if $p\mathcal{O}_K = \mathfrak{p}^n$, where $n = [K : \mathbf{Q}]$. This is the largest possible exponent that could appear, by (6.3). Being totally ramified is stronger than $p\mathcal{O}_K$ being a prime ideal power: it can happen that $p\mathcal{O}_K = \mathfrak{p}^e$ with $e < n$ (see $p = 3$ in Example 6.13), and such p are not totally ramified.

We will see that p being totally ramified is closely related to Eisenstein polynomials with respect to p, which are (monic) polynomials in $\mathbf{Z}[T]$

$$f(T) = T^n + c_{n-1}T^{n-1} + \cdots + c_1T + c_0$$

such that each coefficient c_i is divisible by p and the constant term c_0 is not divisible by p^2. Such polynomials are irreducible in $\mathbf{Q}[T]$, and this Eisenstein criterion for irreducibility is the way everyone first meets Eisenstein polynomials. (It is also how the criterion was discovered by Eisenstein; he developed it as an irreducibility test for $\mathbf{Q}(i)[T]$ [14, Chap. 15].) Before getting to ramification, we show the Eisenstein condition is a simple way to guarantee that $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$.

Lemma 6.29. *Let $K/\mathbf{Q}$ be a number field with degree n. Assume $K = \mathbf{Q}(\alpha)$, where $\alpha \in \mathcal{O}_K$ and its minimal polynomial over $\mathbf{Q}$ is Eisenstein at p. For $a_0, a_1, \ldots, a_{n-1} \in \mathbf{Z}$, if*

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} \equiv 0 \bmod p\mathcal{O}_K \tag{6.7}$$

then $a_i \equiv 0 \bmod p\mathbf{Z}$ for all i.

Proof. We will argue by induction from a_0 up to a_{n-1}.

Multiply through the congruence (6.7) by α^{n-1}, making every term a multiple of α^n except for the first term, which becomes $a_0\alpha^{n-1}$. Since α is the root of an Eisenstein polynomial at p, $\alpha^n \equiv 0 \bmod p\mathcal{O}_K$, so

$$a_0\alpha^{n-1} \equiv 0 \bmod p\mathcal{O}_K.$$

Take norms down to $\mathbf{Z}$: since $a_0\alpha^{n-1} = p\beta$ for some β in $\mathcal{O}_K$, $a_0^n \operatorname{N}_{K/\mathbf{Q}}(\alpha)^{n-1} = p^n \operatorname{N}_{K/\mathbf{Q}}(\beta)$, so

$$a_0^n \operatorname{N}_{K/\mathbf{Q}}(\alpha)^{n-1} \equiv 0 \bmod p^n\mathbf{Z}. \tag{6.8}$$

The norm of α is, up to sign, the constant term of its characteristic polynomial for $K/\mathbf{Q}$. Since α generates $K/\mathbf{Q}$, its characteristic polynomial is its minimal polynomial, which is Eisenstein by hypothesis. Therefore $\operatorname{N}_{K/\mathbf{Q}}(\alpha)$ is divisible by p exactly once, so in (6.8) $\operatorname{N}_{K/\mathbf{Q}}(\alpha)^{n-1}$ is divisible by p exactly $n-1$ times. The modulus in (6.8) is p^n, so $p \mid a_0^n$, so $p \mid a_0$. Now the congruence (6.7) becomes

$$a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} \equiv 0 \bmod p\mathcal{O}_K$$

Multiply this by α^{n-2} to get $a_1\alpha^{n-1} \equiv 0 \bmod p\mathcal{O}_K$ and take norms again. The conclusion now will be $p \mid a_1$. We can now take out the a_1-term from the original congruence and iterate this idea all the way to the last term, so each a_i is divisible by p. ∎

Nonexample 6.30. When $K = \mathbf{Q}(\sqrt{5})$, $1 + \sqrt{5} \equiv 0 \bmod 2\mathcal{O}_K$; the coefficients of 1 and $\sqrt{5}$ are not even, which is no contradiction since the minimal polynomial of $\sqrt{5}$ over $\mathbf{Q}$ is not Eisenstein at 2.

Theorem 6.31. *Let $K/\mathbf{Q}$ be a number field with degree n. Assume $K = \mathbf{Q}(\alpha)$, where α is an algebraic integer whose minimal polynomial over $\mathbf{Q}$ is Eisenstein at p. If*

$$r_0 + r_1\alpha + \cdots + r_{n-1}\alpha^{n-1} \in \mathcal{O}_K$$

with $r_i \in \mathbf{Q}$, then each r_i has no p in its denominator.

Proof. Assume some r_i has a p in its denominator. Let d be the least common denominator, so $p \mid d$, $dr_i \in \mathbf{Z}$ for all i, and some dr_i is not a multiple of p. Then

$$dr_0 + dr_1\alpha + \cdots + dr_{n-1}\alpha^{n-1} \in p\mathcal{O}_K,$$

so Lemma 6.29 tells us $dr_i \in p\mathbf{Z}$ for every i. This is a contradiction. ■

Theorem 6.32. *Let $K = \mathbf{Q}(\alpha)$ where α is the root of an Eisenstein polynomial at p. Then $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$.*

Proof. We argue by contradiction. Suppose $p \mid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$. Then by Cauchy's theorem, the quotient group $\mathcal{O}_K/\mathbf{Z}[\alpha]$ has an element of order p: there is some $\gamma \in \mathcal{O}_K$ such that $\gamma \notin \mathbf{Z}[\alpha]$ but $p\gamma \in \mathbf{Z}[\alpha]$. Write

$$\gamma = r_0 + r_1\alpha + \cdots + r_{n-1}\alpha^{n-1}$$

with $n = [K : \mathbf{Q}]$ and $r_i \in \mathbf{Q}$. Since $\gamma \notin \mathbf{Z}[\alpha]$, some r_i is not in $\mathbf{Z}$. Since $p\gamma \in \mathbf{Z}[\alpha]$ we have $pr_i \in \mathbf{Z}$. Hence r_i has a p in its denominator, which contradicts Theorem 6.31. ■

Example 6.33. We show the ring of algebraic integers of $K = \mathbf{Q}(\sqrt[3]{2})$ is $\mathbf{Z}[\sqrt[3]{2}]$ in a different way than in Section 3.1. We have

$$\operatorname{disc}(\mathbf{Z}[\sqrt[3]{2}]) = [\mathcal{O}_K : \mathbf{Z}[\sqrt[3]{2}]]^2 \operatorname{disc}(K).$$

Since $\operatorname{disc}(\mathbf{Z}[\sqrt[3]{2}]) = \operatorname{disc}(T^3 - 2) = -108 = -2^2 3^3$, 2 and 3 are the only primes which could divide $[\mathcal{O}_K : \mathbf{Z}[\sqrt[3]{2}]]$. Since $\sqrt[3]{2}$ is the root of $T^3 - 2$, which is Eisenstein at 2, 2 does not divide $[\mathcal{O}_K : \mathbf{Z}[\sqrt[3]{2}]]$ by Theorem 6.32. The number $\sqrt[3]{2} + 1$ is a root of $(T-1)^3 - 2 = T^3 - 3T^2 + 3T - 3$, which is Eisenstein at 3, so 3 does not divide $[\mathcal{O}_K : \mathbf{Z}[\sqrt[3]{2} + 1]] = [\mathcal{O}_K : \mathbf{Z}[\sqrt[3]{2}]]$. Hence $[\mathcal{O}_K : \mathbf{Z}[\sqrt[3]{2}]]$ must be 1, so $\mathcal{O}_K = \mathbf{Z}[\sqrt[3]{2}]$.

Example 6.34. We show the ring of algebraic integers of $K = \mathbf{Q}(\sqrt[4]{2})$ is $\mathbf{Z}[\sqrt[4]{2}]$. Since

$$\operatorname{disc}(\mathbf{Z}[\sqrt[4]{2}]) = [\mathcal{O}_K : \mathbf{Z}[\sqrt[4]{2}]]^2 \operatorname{disc}(K)$$

and $\operatorname{disc}(\mathbf{Z}[\sqrt[4]{2}]) = \operatorname{disc}(T^4 - 2) = -2^{11}$, $[\mathcal{O}_K : \mathbf{Z}[\sqrt[4]{2}]]$ is a power of 2. Because $\sqrt[4]{2}$ is a root of $T^4 - 2$ which is Eisenstein at 2, 2 does not divide $[\mathcal{O}_K : \mathbf{Z}[\sqrt[4]{2}]]$. Therefore the index is 1.

Example 6.35. We show the ring of algebraic integers of $K = \mathbf{Q}(\sqrt[5]{2})$ is $\mathbf{Z}[\sqrt[5]{2}]$. The discriminant of $T^5 - 2$ is $2^4 5^5$, so the only prime factors of $[\mathcal{O}_K : \mathbf{Z}[\sqrt[5]{2}]]$ could be 2 and 5. Since $\sqrt[5]{2}$ is a root of $T^5 - 2$, which is Eisenstein at 2, and $\sqrt[5]{2} - 2$ is a root of

$$(T + 2)^5 - 2 = T^5 + 10T^4 + 40T^3 + 80T^2 + 80T + 30,$$

which is Eisenstein at 5, neither 2 nor 5 divides the index since $\mathbf{Z}[\sqrt[5]{2} - 2] = \mathbf{Z}[\sqrt[5]{2}]$.

Perhaps at this point you're intrigued: does $\mathbf{Q}(\sqrt[n]{2})$ have ring of integers $\mathbf{Z}[\sqrt[n]{2}]$ for all n? This can be answered using ramification, but it will not be done here.

Example 6.36. We return to Example 6.23, where $K = \mathbf{Q}(\alpha)$ and $L = \mathbf{Q}(\beta)$ for roots α and β of the polynomials $f(T) = T^3 + 6T - 6$ and $g(T) = T^3 - 6T - 10$. Each polynomial has discriminant $-1836 = -2^2 \cdot 3^3 \cdot 17$, so the discriminants of K and L are factors of 1836. Their complementary factors in 1836 are $[\mathcal{O}_K : \mathbf{Z}[\alpha]]^2$ and $[\mathcal{O}_L : \mathbf{Z}[\beta]]^2$. These square factors must divide $2^2 \cdot 3^2$.

Since $f(T)$ and $g(T)$ are both Eisenstein at 2, $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ and $[\mathcal{O}_L : \mathbf{Z}[\beta]]$ are not divisible by 2. Since $f(T)$ and $g(T - 2)$ are Eisenstein at 3, $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ and $[\mathcal{O}_L : \mathbf{Z}[\beta + 2]] = [\mathcal{O}_L : \mathbf{Z}[\beta]]$ are not divisible by 3. Therefore $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ and $[\mathcal{O}_L : \mathbf{Z}[\beta]]$ both equal 1, so $\operatorname{disc}(K) = \operatorname{disc}(L) = -1836$, $\mathcal{O}_K = \mathbf{Z}[\alpha]$, and $\mathcal{O}_L = \mathbf{Z}[\beta]$.

The link between Eisenstein polynomials and totally ramified primes is described in the following two theorems, which are converses of each other.

Theorem 6.37. *Let $K = \mathbf{Q}(\alpha)$, where α is the root of a polynomial of degree n which is Eisenstein at a prime p. Then p is totally ramified in K and $p\mathcal{O}_K = \mathfrak{p}^n$ where $\mathfrak{p} = (p, \alpha)$.*

Proof. By Theorem 6.32, $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$, so we can factor p in $\mathcal{O}_K$ by factoring $f(T) \bmod p$, where $f(T)$ is the minimal polynomial for α. An Eisenstein polynomial at p reduces mod p to T^n, so $(p) = (p, \alpha)^n$. ■

Theorem 6.38. *Let K be a number field of degree n, and suppose a prime p is totally ramified in K. Then $K = \mathbf{Q}(\alpha)$ for some α which is the root of an Eisenstein polynomial at p. More precisely, if we write $p\mathcal{O}_K = \mathfrak{p}^n$ then we can take for α any element of $\mathfrak{p} - \mathfrak{p}^2$.*

Proof. Taking the ideal norm of both sides of the equation $p\mathcal{O}_K = \mathfrak{p}^n$, $p^n = \mathrm{N}(\mathfrak{p})^n$, so $\mathrm{N}(\mathfrak{p}) = p$: the prime ideal lying over a totally ramified prime has prime norm.

For $\alpha \in \mathfrak{p} - \mathfrak{p}^2$,

$$(\alpha) = \mathfrak{p}\mathfrak{a}, \tag{6.9}$$

where $\mathfrak{p}$ does not divide $\mathfrak{a}$. Taking ideal norms in (6.9),

$$|\,\mathrm{N}_{K/\mathbf{Q}}(\alpha)| = p\,\mathrm{N}(\mathfrak{a}).$$

Since $\mathfrak{p}$ is the only prime dividing (p), $\mathrm{N}(\mathfrak{a})$ is not divisible by p (Corollary 4.28), so p divides $\mathrm{N}_{K/\mathbf{Q}}(\alpha)$ exactly once. Write the characteristic polynomial $f(T) = \chi_{K/\mathbf{Q},\alpha}(T)$ as

$$f(T) = T^n + c_{n-1}T^{n-1} + \cdots + c_1T + c_0 \in \mathbf{Z}[T].$$

We will show this polynomial is Eisenstein at p, and therefore it is irreducible of degree n, so $K = \mathbf{Q}(\alpha)$. The constant term c_0 of $f(T)$ is $\pm\mathrm{N}_{K/\mathbf{Q}}(\alpha)$, which is divisible by p exactly once. To show each c_i is a multiple of p, we will use induction. The method will be similar to Lemma 6.29, but we can't use that result since the hypotheses of Lemma 6.29 are exactly what we are trying to prove here.

Reduce the equation $f(\alpha) = 0$ modulo $\mathfrak{p}^n$:

$$\alpha^n + c_{n-1}\alpha^{n-1} + \cdots + c_1\alpha + c_0 \equiv 0 \bmod \mathfrak{p}^n.$$

Since $\alpha \in \mathfrak{p}$, we can drop the term α^n. Suppose by induction that $c_0, \ldots, c_{i-1} \equiv 0 \bmod \mathfrak{p}^n$ and $i \leqslant n-1$. Then

$$c_{n-1}\alpha^{n-1} + \cdots + c_i\alpha^i \equiv 0 \bmod \mathfrak{p}^n.$$

Reduce both sides of this mod $\mathfrak{p}^{i+1}$ (which makes sense since $\mathfrak{p}^{i+1} \mid \mathfrak{p}^n$). Since α^{i+1} is in $\mathfrak{p}^n$, we are left with

$$c_i\alpha^i \equiv 0 \bmod \mathfrak{p}^{i+1}.$$

Therefore $\mathfrak{p}^{i+1} \mid (c_i)(\alpha)^i$. Since (α) is divisible by $\mathfrak{p}$ just once, $\mathfrak{p} \mid (c_i)$, so $c_i \in \mathfrak{p} \cap \mathbf{Z} = p\mathbf{Z}$. ■

Example 6.39. Let $K = \mathbf{Q}(\sqrt[3]{2})$. Since $\sqrt[3]{2}$ is a root of $T^3 - 2$, which is Eisenstein at 2, the prime number 2 is totally ramified in K. Indeed, $(2) = (\sqrt[3]{2})^3$. Similarly, since $K = \mathbf{Q}(\sqrt[3]{2}+1)$ and $\sqrt[3]{2}+1$ is a root of

$$(T-1)^3 - 2 = T^3 - 3T^2 + 3T - 3,$$

which is Eisenstein at 3, we must have $(3) = \mathfrak{p}^3$. In fact,

$$(\sqrt[3]{2}+1)^3 = 3(1+\sqrt[3]{2}+\sqrt[3]{4}),$$

and the second factor is a unit in $\mathbf{Z}[\sqrt[3]{2}]$ (its inverse is $\sqrt[3]{2}-1$), so $(3) = (\sqrt[3]{2}+1)^3$.

Example 6.40. Let $K = \mathbf{Q}(\sqrt{-5})$. Since $1+\sqrt{-5}$ is a field generator and is a root of $T^2 - 2T + 6$, which is Eisenstein at 2, we have $(2) = \mathfrak{p}^2$ for some prime ideal $\mathfrak{p}$. The ideal $\mathfrak{p}$ is $(2, 1+\sqrt{-5})$.

Example 6.41. Let $K = \mathbf{Q}(\alpha)$ where $\alpha^3 - 12\alpha + 2 = 0$. In Example 3.48 we showed $2 \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$ and in Example 4.47 we showed $(2) = \mathfrak{p}_2^3$. Now we see both conditions follow immediately from $T^3 - 12T + 2$ being Eisenstein at 2. We saw in Example 4.47 that $(3) = \mathfrak{p}_3^3$: $\beta = \frac{1}{3}(1+\alpha+\alpha^2)$ is a root of $T^3 - 9T^2 + 21T - 7$, which factors mod 3 as $(T-1)^3$, and we saw that $3 \nmid [\mathcal{O}_K : \mathbf{Z}[\beta]]$ by tedious computations. The minimal polynomial of $\beta - 1$ is $T^3 - 6T^2 + 6T + 6$, which is Eisenstein at 3. That proves the index $[\mathcal{O}_K : \mathbf{Z}[\beta - 1]] = [\mathcal{O}_K : \mathbf{Z}[\beta]]$ is not divisible by 3 and $(3) = \mathfrak{p}_3^3$ without having to directly compute the index or the ideal factorization (but we did have to find β first).

Example 6.42. Let $K = \mathbf{Q}(\sqrt[3]{10})$. Since $T^3 - 10$ is Eisenstein at 2 and 5, neither of these primes divides $[\mathcal{O}_K : \mathbf{Z}[\sqrt[3]{10}]]$ and both primes are totally ramified in K. This conclusion was reached earlier with a lot of (implicit) computation in Examples 3.50 and 4.48.

Example 6.43. No discussion of Eisenstein polynomials would be complete without a discussion of pth roots of unity. For any prime p, let ζ_p be a root of unity of order p. Since $\zeta_p^p = 1$ but $\zeta_p \neq 1$, ζ_p is is a root of

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

which is not Eisenstein but the translate $\Phi_p(T+1)$ is:

$$\Phi_p(T+1) = \frac{(T+1)^p - 1}{T} = \sum_{k=1}^{p} \binom{p}{k} T^{k-1} = T^{p-1} + pT^{p-2} + \cdots + p.$$

The binomial coefficients $\binom{p}{k}$ are multiples of p when $1 \leqslant k \leqslant p-1$, so the non-leading coefficients are divisible by p and the constant term is divisible by p once. A root of $\Phi_p(T+1)$ is $\zeta_p - 1$, so p is totally ramified in $\mathbf{Q}(\zeta_p - 1) = \mathbf{Q}(\zeta_p)$ and in fact $(p) = (p, \zeta_p - 1)^{p-1}$ by Theorem 6.37.

For any prime power $p^r > 1$, let ζ_{p^r} be a root of unity of order p^r. Then $\zeta_{p^r}^{p^r} = 1$ but $\zeta_{p^{r-1}} \neq 1$, so ζ_{p^r} is a root of

$$\Phi_{p^r}(T) := \frac{T^{p^r} - 1}{T^{p^{r-1}-1}} = \Phi_p(T^{p^{r-1}}) = 1 + T^{p^{r-1}} + T^{2p^{r-1}} + \cdots + T^{(p-1)p^{r-1}},$$

The translate $\Phi_{p^r}(T+1)$ has constant term $\Phi_{p^r}(1) = p$ and its non-leading coefficients are all multiples of p. Rather than show that from scratch, we can work mod p:

$$\begin{aligned} T^{p^r} - 1 \equiv \Phi_{p^r}(T)(T^{p^{r-1}} - 1) \bmod p \quad &\Longrightarrow \quad (T-1)^{p^r} \equiv \Phi_{p^r}(T)(T-1)^{p^{r-1}} \bmod p \\ &\Longrightarrow \quad \Phi_{p^r}(T) \equiv (T-1)^{(p-1)p^{r-1}} \bmod p, \end{aligned}$$

so $\Phi_{p^r}(T+1) \equiv T^{(p-1)p^{r-1}} \bmod p$, which means all non-leading coefficients are $0 \bmod p$. (For $r = 1$ this is an alternate proof that $\Phi_p(T+1)$ is Eisenstein at p.) So, as in the case $r = 1$, p totally ramifies in $\mathbf{Q}(\zeta_{p^r} - 1) = \mathbf{Q}(\zeta_{p^r})$ for any r and its factorization is $(p) = (p, \zeta_{p^r} - 1)^{\varphi(p^r)}$.

Example 6.44. Let $K = \mathbf{Q}(\sqrt[8]{3})$ and $L = \mathbf{Q}(\sqrt[8]{48})$. Both K and L have degree 8 over $\mathbf{Q}$, since $T^8 - 3$ and $T^8 - 48 = T^8 - 16 \cdot 3$ are Eisenstein at 3. These two fields are nonisomorphic because they both lie in $\mathbf{Q}(\sqrt[8]{3}, \zeta_8)$ and they correspond to nonconjugate subgroups of $\mathrm{Gal}(\mathbf{Q}(\sqrt[8]{3}, \zeta_8)/\mathbf{Q})$ (Exercise 6.32c). We will show, however, that each prime p factors in the same way in K and L. So nonisomorphic number fields can have prime factorizations of the same shape for all p.

Because T^8-3 and T^8-48 are both Eisenstein at 3, 3 is totally ramified in both K and L. The polynomial $(T+1)^8-3$ is Eisenstein at 2, so 2 is totally ramified in K. Is 2 totally ramified in L? With the aid of PARI, $\frac{1}{8}\sqrt[8]{48}^5+\frac{1}{2}\sqrt[8]{48}-1$ is a root of

$$T^8+8T^7+28T^6+56T^5+46T^4-40T^3-116T^2-88T-26, \tag{6.10}$$

which is Eisenstein at 2. Thus 2 is totally ramified in K and L.

For any prime p other than 2 or 3, T^8-3 and T^8-48 are both separable in $\mathbf{F}_p[T]$, so the way p factors in K and L matches the way T^8-3 and T^8-48 factor in $\mathbf{F}_p[T]$. Since $48=3\cdot 16$, to show these two polynomials factor in the same way mod p it suffices to show 16 mod p is an 8th power for all $p>3$, since if $16\equiv c_p^8 \bmod p$ then $T^8-48\equiv T^8-3c_p^8\equiv c_p^8((T/c_p)^8-3) \bmod p$, so T^8-48 becomes T^8-3 in $\mathbf{F}_p[T]$ after some scaling.

We can write 16 as a 4th power in several ways:

$$16=2^4=(2i)^4=(-2)^4. \tag{6.11}$$

For $p\neq 2$, at least one of 2, -1, or -2 is a square in $\mathbf{F}_p^\times$: the kernel of $x\mapsto x^2$ on $\mathbf{F}_p^\times$ has a kernel of size 2, so it has an image of index 2, which means $\mathbf{F}_p^\times/(\mathbf{F}_p^\times)^2$ has order 2, and therefore two numbers in $\mathbf{F}_p^\times$ which are not squares have a product that is a square. Depending on which of 2, -2, or -1 is a square in $\mathbf{F}_p^\times$, one of the formulas for 16 in (6.11) shows 16 is an 8th power in $\mathbf{F}_p$ (*e.g.*, if -1 is a square then the formula $2i=(1+i)^2$ show $(2i)^4$ is an 8th power of $1+\sqrt{-1}$).

In practice, when the Eisenstein criterion is used to prove a polynomial $f(T)$ in $\mathbf{Z}[T]$ is irreducible, the criterion is applied not to $f(T)$ but to some translate $f(T+c)$ (*e.g.*, to prove the pth cyclotomic polynomial $\Phi_p(T)$ is irreducible for prime p, show $\Phi_p(T+1)$ is Eisenstein at p). In terms of a root α of $f(T)$ the number field $\mathbf{Q}(\alpha)=\mathbf{Q}(\alpha-c)$ would have to be totally ramified at p in order for $f(T+c)$ to be Eisenstein at p. For a ramified prime to be totally ramified is rather special, except for quadratic fields where ramification is forced to be total by degree considerations. Therefore Theorem 6.37 is telling us that we can't expect to prove a random irreducible $f(T)\in\mathbf{Z}[T]$ is irreducible by finding an Eisenstein translate $f(T+c)$.

The following divisibility criterion makes this point even stronger, in terms of the multiplicity of prime factors of the discriminant.

Theorem 6.45. *Let $K = \mathbf{Q}(\alpha)$ where α is the root of an Eisenstein polynomial at p with degree n.*

(a) *If $p \nmid n$, then $p^{n-1} || \operatorname{disc}(K)$. That is, $p^{n-1} \mid \operatorname{disc}(K)$ and $p^n \nmid \operatorname{disc}(K)$.*

(b) *If $p \mid n$, then $p^n \mid \operatorname{disc}(K)$. More precisely, the multiplicity of p in $\operatorname{disc}(K)$ is at least n and at most $nd + n - 1$, where $p^d || n$.*

This is telling us the exact power of p in $\operatorname{disc}(K)$ if $p \nmid n$, and upper and lower bounds on that power of p when $p \mid n$. For instance, Theorem 6.45 says the power of 2 in $\operatorname{disc}(\mathbf{Q}(\sqrt[4]{2}))$ is at least 2^4 and at most $2^{4\cdot 2+4-1} = 2^{11}$. The discriminant in fact is -2^{11} (Example 6.34).

Proof. Since $\operatorname{disc}(\mathbf{Z}[\alpha]) = [\mathcal{O}_K : \mathbf{Z}[\alpha]]^2 \operatorname{disc}(K)$ and $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$ by Theorem 6.32, the highest powers of p in $\operatorname{disc}(K)$ and $\operatorname{disc}(\mathbf{Z}[\alpha])$ are the same. From the formula

$$\operatorname{disc}(\mathbf{Z}[\alpha]) = \operatorname{disc}(f(T)) = \pm \mathrm{N}_{K/\mathbf{Q}}(f'(\alpha))$$

we now look for the highest power of p that is a factor of $\mathrm{N}_{K/\mathbf{Q}})(f'(\alpha))$.

Write $(p) = \mathfrak{p}^n$. Let the minimal polynomial of α over $\mathbf{Q}$ be $f(T) = T^n + c_{n-1}T^{n-1} + \cdots + c_1T + c_0$, so $\mathrm{N}((\alpha)) = |\,\mathrm{N}_{K/\mathbf{Q}}(\alpha)| = |c_0|$ is divisible by p exactly once. The only prime ideal over p in K is $\mathfrak{p}$, so (α) is divisible by $\mathfrak{p}$ exactly once: $(\alpha) = \mathfrak{p}\mathfrak{a}$ with $\mathfrak{p} \nmid \mathfrak{a}$.

a) We have

$$f'(\alpha) = n\alpha^{n-1} + (n-1)c_{n-1}\alpha^{n-2} + \cdots + 2c_2\alpha + c_1, \tag{6.12}$$

and each c_i is divisible by p, and thus $c_i \equiv 0 \bmod \mathfrak{p}^n$, so all terms on the right side (6.12) except the first term are $0 \bmod \mathfrak{p}^n$. Thus

$$f'(\alpha) \equiv n\alpha^{n-1} \bmod \mathfrak{p}^n. \tag{6.13}$$

We know $(\alpha)^{n-1} = \mathfrak{p}^{n-1}\mathfrak{a}^{n-1}$ is divisible by $\mathfrak{p}^{n-1}$ and not by $\mathfrak{p}^n$. Therefore if $n \not\equiv 0 \bmod \mathfrak{p}$, which is the same as $n \not\equiv 0 \bmod p$, (6.13) says $(f'(\alpha))$ is divisible by $\mathfrak{p}$ exactly $n-1$ times. So $(f'(\alpha)) = \mathfrak{p}^{n-1}\mathfrak{b}$ where $\mathfrak{p} \nmid \mathfrak{b}$, and taking ideal norms gives us $\pm \mathrm{N}_{K/\mathbf{Q}}(f'(\alpha)) = p^{n-1}b$ where $p \nmid b$.

b) If $n \equiv 0 \bmod p$ then (6.13) tells us $\mathfrak{p}^n \mid (f'(\alpha))$, so $p^n \mid \mathrm{N}_{K/\mathbf{Q}}(f'(\alpha))$.

To get an upper bound on the highest power of p in $|\,\mathrm{N}_{K/\mathbf{Q}}(f'(\alpha))| = \mathrm{N}((\alpha))$ when $p \mid n$, it will suffice to get an upper bound on the highest power of $\mathfrak{p}$ in

$(f'(\alpha))$, since $\mathfrak{p}$ is the only prime factor of (p). We return to (6.12). Each of the terms on the right side, when not 0, is divisible[2] by a different power of $\mathfrak{p}$. The first term $n\alpha^{n-1}$ is divisible by $p^d\mathfrak{p}^{n-1} = \mathfrak{p}^{nd+n-1}$. For $1 \leqslant i \leqslant n-1$, if $c_i \neq 0$ then $ic_i\alpha^{i-1}$ is divisible by $p^{d_i}\mathfrak{p}^{i-1} = \mathfrak{p}^{nd_i+i-1}$, where $d_i \geqslant 1$ is the highest power of p in ic_i. The $\mathfrak{p}$-power in $n\alpha^{n-1}$ has this form too if we set $c_n = 1$ and $d_n = d$. The $\mathfrak{p}$-exponents $nd_i + i - 1$ for $1 \leqslant i \leqslant n$ are all incongruent mod n, so they are unequal in $\mathbf{Z}$. Therefore the highest $\mathfrak{p}$-power dividing $(f'(\alpha))$ is the smallest $\mathfrak{p}$-power on the list:

$$\min_{1\leqslant i\leqslant n, c_i\neq 0}(nd_i + i - 1).$$

Since $c_n = 1 \neq 0$, this minimum is at most $nd_n + n - 1 = nd + n - 1$. ■

Corollary 6.46. *If $f(T) \in \mathbf{Z}[T]$ is monic of degree n and $f(T+c)$ is Eisenstein at a prime p for some $c \in \mathbf{Z}$, then $p^{n-1} \mid \operatorname{disc}(f(T))$.*

Proof. Since $f(T+c)$ and $f(T)$ have the same discriminant, which is a multiple of $\operatorname{disc}(K)$, if $f(T+c)$ is Eisenstein at p then $p^{n-1} \mid \operatorname{disc}(f(T))$. When $p \mid n$ we even have $p^n \mid \operatorname{disc}(f(T))$. ■

Example 6.47. For primes p and q not equal to 3 or to each other, let $K = \mathbf{Q}(\sqrt[3]{pq^2})$. We will find conditions on p and q that imply $\mathcal{O}_K$ is not monogenic (does not have a power basis).

View K as a subfield of $\mathbf{R}$, so there is no ambiguity about the meaning of cube roots. Set $\alpha = \sqrt[3]{pq^2}$. The field K contains $\beta = \sqrt[3]{p^2q} = \frac{1}{q}\alpha^2$, and $\alpha = \frac{1}{p}\beta^2$. Neither $\mathbf{Z}[\alpha]$ nor $\mathbf{Z}[\beta]$ can be $\mathcal{O}_K$, since $\beta \notin \mathbf{Z}[\alpha]$ and $\alpha \notin \mathbf{Z}[\beta]$. The $\mathbf{Z}$-lattice $\mathbf{Z} + \mathbf{Z}\alpha + \mathbf{Z}\beta$ properly contains them both:

$$\mathbf{Z}[\alpha], \mathbf{Z}[\beta] \subsetneqq \mathbf{Z} + \mathbf{Z}\alpha + \mathbf{Z}\beta \subset \mathcal{O}_K.$$

Since $\mathbf{Z}[\alpha] = \mathbf{Z} + \mathbf{Z}\alpha + \mathbf{Z}q\beta$ and $\mathbf{Z}[\beta] = \mathbf{Z} + \mathbf{Z}\beta + \mathbf{Z}p\alpha$,

$$[\mathbf{Z} + \mathbf{Z}\alpha + \mathbf{Z}\beta : \mathbf{Z}[\alpha]] = q \text{ and } [\mathbf{Z} + \mathbf{Z}\alpha + \mathbf{Z}\beta : \mathbf{Z}[\beta]] = p. \tag{6.14}$$

The lattice $\mathbf{Z} + \mathbf{Z}\alpha + \mathbf{Z}\beta$ is a ring since $\alpha^2 = q\beta$, $\beta^2 = p\alpha$, and $\alpha\beta = pq$.

Since $\operatorname{disc}(T^3 - pq^2) = -27p^2q^4$ and $\operatorname{disc}(T^3 - p^2q) = -27p^4q^2$,

$$-27p^2q^4 = [\mathcal{O}_K : \mathbf{Z}[\alpha]]^2 \operatorname{disc}(K) \text{ and } -27p^4q^2 = [\mathcal{O}_K : \mathbf{Z}[\beta]]^2 \operatorname{disc}(K). \tag{6.15}$$

[2] We say an element $x \in \mathcal{O}_K$ is divisible by an ideal $\mathfrak{b}$ when $\mathfrak{b} \mid (x)$, or equivalently $x \equiv 0 \bmod \mathfrak{b}$.

The polynomials $T^3 - pq^2$ and $T^3 - p^2q$ are Eisenstein at p and q respectively, so (6.15) tells us $\mathrm{disc}(K)$ is divisible by p^2 and q^2. The index $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ is divisible by q since β has order q in $\mathcal{O}_K/\mathbf{Z}[\alpha]$. Similarly, $[\mathcal{O}_K : \mathbf{Z}[\beta]]$ is divisible by p. Feeding this information into (6.15), we have two possibilities:

- $\mathrm{disc}(K) = -3p^2q^2$, $[\mathcal{O}_K : \mathbf{Z}[\alpha]] = 3q$, and $[\mathcal{O}_K : \mathbf{Z}[\beta]] = 3p$.
- $\mathrm{disc}(K) = -27p^2q^2$, $[\mathcal{O}_K : \mathbf{Z}[\alpha]] = q$, and $[\mathcal{O}_K : \mathbf{Z}[\beta]] = p$.

Either way, 3 divides $\mathrm{disc}(K)$, so the primes that ramify in K are 3, p, and q.

If the second possibility happens then (6.14) tells us

$$\mathcal{O}_K = \mathbf{Z} + \mathbf{Z}\alpha + \mathbf{Z}\beta.$$

Also, $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ and $[\mathcal{O}_K : \mathbf{Z}[\beta]]$ are relatively prime, so we can use Dedekind's factorization theorem to factor any prime in $\mathcal{O}_K$.

A way to force the second possibility is to make 3 totally ramified, since in that case we are assured that $3^3 \mid \mathrm{disc}(K)$ by Theorem 6.45 with $n = 3$. We force 3 to be totally ramified by having K generated by the root of an Eisenstein polynomial at 3. While $f(T) = T^3 - pq^2$ is not Eisenstein at 3, consider the polynomial

$$f(T \pm 1) = T^3 \pm 3T^2 + 3T \pm 1 - pq^2.$$

Since $pq^2 \not\equiv 0 \bmod 3$, for some choice of sign the constant term will be a multiple of 3. We want it to be divisible by 3 exactly once. This means $pq^2 \equiv 2, 4, 5, 7 \bmod 9$. There are many such choices. One is $p = 2$ and $q = 5$. For such a choice, $\sqrt[3]{pq^2} \pm 1$ will be the root of an Eisenstein polynomial at 3.

For $\mathcal{O}_K$ not to be monogenic, we want to force the index $[\mathcal{O}_K : \mathbf{Z}[\gamma]]$ to be larger than 1 for every $\gamma \in \mathcal{O}_K - \mathbf{Z}$. Write $\gamma = a + b\alpha + c\beta$ for some integers a, b, and c, where b and c are not both 0. Since $\mathbf{Z}[a + b\alpha + c\beta] = \mathbf{Z}[b\alpha + c\beta]$, we can assume $a = 0$. The index $[\mathcal{O}_K : \mathbf{Z}[b\alpha + c\beta]]$ is the absolute value of the determinant of a matrix expressing a basis of $\mathbf{Z}[b\alpha + c\beta]$ in terms of a basis of $\mathcal{O}_K$ (Theorem 3.10). Using the basis $1, \alpha, \beta$ for $\mathcal{O}_K$, we find

$$\begin{pmatrix} 1 \\ b\alpha + c\beta \\ (b\alpha + c\beta)^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & b & c \\ 2pqbc & pc^2 & qb^2 \end{pmatrix} \begin{pmatrix} 1 \\ \alpha \\ \beta \end{pmatrix}.$$

The determinant of the matrix is $qb^3 - pc^3$, so $\mathcal{O}_K = \mathbf{Z}[b\alpha + c\beta]$ if and only if $qb^3 - pc^3 = \pm 1$. This implies $qb^3 \equiv \pm 1 \bmod p$ and $pc^3 \equiv \pm 1 \bmod q$, so $p \bmod q$

is a cube and $q \bmod p$ is a cube. As long as either $p \bmod q$ or $q \bmod p$ is not a cube, the index can't be 1 for any b or c. Let's summarize.

Theorem 6.48. *For distinct primes p and q not equal to 3 such that $pq^2 \equiv 2, 4, 5, 7 \bmod 9$, the ring of integers of $\mathbf{Q}(\sqrt[3]{pq^2})$ has $\mathbf{Z}$-basis $\{1, \sqrt[3]{pq^2}, \sqrt[3]{p^2q}\}$ and has discriminant $-27p^2q^2$. If $p \bmod q$ is not a cube or $q \bmod p$ is not a cube then this ring is not monogenic.*

The smallest p and q where $pq^2 \equiv 2, 4, 5, 7 \bmod 9$ and one of $p \bmod q$ and $q \bmod p$ is not a cube is $p = 7$ and $q = 5$, so $pq^2 = 175$ and $p^2q = 245$. Therefore the field $K = \mathbf{Q}(\sqrt[3]{175})$ does not have a power basis for its ring of integers, which is $\mathbf{Z} + \mathbf{Z}\sqrt[3]{175} + \mathbf{Z}\sqrt[3]{245}$. Moreover, in the course of the proof we saw $[\mathcal{O}_K : \mathbf{Z}[\sqrt[3]{175}]] = 5$ and $[\mathcal{O}_K : \mathbf{Z}[\sqrt[3]{245}]] = 7$, so we can't prove $\mathcal{O}_K$ isn't monogenic by finding a prime factor common to all indices $[\mathcal{O}_K : \mathbf{Z}[\gamma]]$, which was the method used for Dedekind's field (Theorem 4.49).

6.6 Cyclotomic Fields

The cyclotomic fields $\mathbf{Q}(\zeta_m)$, where ζ_m is a primitive mth root of unity, are the most important family of number fields after quadratic fields. Here are a few illustrations (hardly exhaustive) of their role.

- Kummer discovered ideal factorization (without the language of ideals) in his study of these fields, principally the fields $\mathbf{Q}(\zeta_p)$ for prime p.
- The Kronecker–Weber theorem says every finite abelian extension of $\mathbf{Q}$ is inside a cyclotomic field. The search for a similar theorem describing finite abelian extensions of other number fields[3] led to class field theory.
- Iwasawa's study of the p-part of the class groups of the fields $\mathbf{Q}(\zeta_{p^n})$ for a fixed prime p led to Iwasawa theory [60, Chap. 13].
- In most number fields it is not easy to write down units of infinite order explicitly (*e.g.*, there's no formula for a nontrivial solution of Pell's equation), but inside cyclotomic fields are cyclotomic units [60, Chap. 8], which generate "most" of the unit group and also lead to the most concrete examples of Euler systems [11, Chap. 5].

[3] If $K \neq \mathbf{Q}$, some finite abelian extension of K is not in any $K(\zeta_m)$, so the most naive generalization of the Kronecker–Weber theorem to number fields beyond $\mathbf{Q}$ is always wrong.

- Mihailescu's proof [49] of Catalan's conjecture, which says that the only consecutive perfect powers in $\mathbf{Z}^+$ are 8 and 9, uses arithmetic in $\mathbf{Q}(\zeta_p)$.

The degree formula $[\mathbf{Q}(\zeta_m) : \mathbf{Q}] = \varphi(m)$ is often proved in courses on Galois theory. As an illustration of techniques in algebraic number theory, we will use ramification to prove that degree formula in another way and also to show $\mathbf{Z}[\zeta_m]$ is the ring of integers of $\mathbf{Q}(\zeta_m)$. Then we will see how every prime number factors into prime ideals in $\mathbf{Z}[\zeta_m]$.

The two basic features we need about ramification are the following:

- For $r \geqslant 1$, p is totally ramified in $\mathbf{Q}(\zeta_{p^r})$ with its unique prime ideal factor being $(p, \zeta_{p^r} - 1)$ (Example 6.43).

- If m is not divisible by p then p is unramified in $\mathbf{Q}(\zeta_m)$ (Example 6.18).

These should be compared with the way $T^m - 1$ factors mod p:

- For $r \geqslant 1$, $T^{p^r} - 1 \equiv (T-1)^{p^r} \bmod p$.

- If m is not divisible by p then $T^m - 1 \bmod p$ is separable (distinct irreducible factors).

(Although $T^m - 1$ is *not* the minimal polynomial of ζ_m over $\mathbf{Q}$ when $m > 1$, the ramification information about p in $\mathbf{Q}(\zeta_m)$ can be remembered in terms of how $T^m - 1 \bmod p$ factors.)

Writing $m = p^r m'$ where $p \nmid m'$, the field $\mathbf{Q}(\zeta_m)$ is the composite of $\mathbf{Q}(\zeta_{p^r})$ and $\mathbf{Q}(\zeta_{m'})$. The crucial point to keep in mind with the diagram below is that p is totally ramified in $\mathbf{Q}(\zeta_{p^r})$ and is unramified in $\mathbf{Q}(\zeta_{m'})$.

$$\begin{tikzcd}
& \mathbf{Q}(\zeta_m) \arrow[dl, dash] \arrow[dr, dash] & \\
\mathbf{Q}(\zeta_{p^r}) \arrow[dr, dash, "p \text{ tot. ram.}"'] & & \mathbf{Q}(\zeta_{m'}) \arrow[dl, dash, "p \text{ unram.}"] \\
& \mathbf{Q} &
\end{tikzcd}$$

We will prove some general theorems about composites of two number fields such that a prime is totally ramified in one field and unramified in the other field. Theorems about cyclotomic fields will fall out as special cases using the above field diagram.

If E/F is an extension of number fields and $\mathfrak{a}$ is an ideal in $\mathcal{O}_F$, we can extend it to an ideal in $\mathcal{O}_E$:

$$\mathfrak{a}\mathcal{O}_E := \left\{ \sum_{i=1}^{r} a_i x_i : r \geqslant 1, a_i \in \mathfrak{a}, x_i \in \mathcal{O}_E \right\}. \tag{6.16}$$

This is the smallest ideal in $\mathcal{O}_E$ containing $\mathfrak{a}$. When $\mathfrak{p}$ is a (nonzero) prime ideal in $\mathcal{O}_F$, the extended ideal $\mathfrak{p}\mathcal{O}_E$ need not be prime (*e.g.*, the extension of $5\mathbf{Z}$ to $\mathbf{Z}[i]$ is $5\mathbf{Z}[i]$, which factors since $5 = (1+2i)(1-2i)$). We want to show $\mathfrak{p}\mathcal{O}_E$ is not the unit ideal, so it has prime ideal factors.

Lemma 6.49. *If E/F is a finite extension of number fields and $\mathfrak{p}$ is a nonzero prime ideal in $\mathcal{O}_F$ then $\mathfrak{p}\mathcal{O}_E \neq \mathcal{O}_E$.*

Proof. If $\mathfrak{p}\mathcal{O}_E = \mathcal{O}_E$ then 1 belongs to $\mathfrak{p}\mathcal{O}_E$:

$$1 = \sum_{i=1}^{r} a_i x_i,$$

where $a_i \in \mathfrak{p}$ and $x_i \in \mathcal{O}_E$. For any $b \in \mathfrak{p}^{-1}$, multiplying by b gives us

$$b = \sum_{i=1}^{r} (ba_i) x_i.$$

On the right side, $ba_i \in \mathfrak{p}^{-1}\mathfrak{p} = \mathcal{O}_F$ and $x_i \in \mathcal{O}_E$. Therefore $b \in \mathcal{O}_E \cap F = \mathcal{O}_F$, so $\mathfrak{p}^{-1} \subset \mathcal{O}_F$. However, $\mathfrak{p}^{-1} \not\subset \mathcal{O}_F$, so we have a contradiction. ■

Remark 6.50. Since $\mathfrak{p}\mathcal{O}_E \neq \mathcal{O}_E$, there is a prime ideal $\mathfrak{P}$ in $\mathcal{O}_E$ such that $\mathfrak{P} \mid \mathfrak{p}\mathcal{O}_E$. Typically there may be more than one such $\mathfrak{P}$. Associating to $\mathfrak{p}$ the prime ideal factors of $\mathfrak{p}\mathcal{O}_E$ is usually a multi-valued (one-to-many) map, but in the other direction we we can always recover $\mathfrak{p}$ from any $\mathfrak{P}$ dividing $\mathfrak{p}\mathcal{O}_E$ by taking intersections: $\mathfrak{P} \cap \mathcal{O}_F = \mathfrak{p}$. (This is analogous to $\mathfrak{p} \cap \mathbf{Z} = p\mathbf{Z}$ for any $\mathfrak{p}$ dividing $p\mathcal{O}_F$.) Indeed, $\mathfrak{p} \subset \mathfrak{P} \cap \mathcal{O}_F \subset \mathcal{O}_F$ and $\mathfrak{p}$ is maximal while $1 \notin \mathfrak{P} \cap \mathcal{O}_F$.

Lemma 6.51. *If a prime p is unramified in a number field K then it is unramified in every subfield of K. If p is totally ramified in K then it is totally ramified in every subfield of K.*

Proof. Let $\mathbf{Q} \subset F \subset K$, so the natural ring homomorphism $\mathcal{O}_F/p\mathcal{O}_F \to \mathcal{O}_K/p\mathcal{O}_K$ is injective. If p is unramified in K then $\mathcal{O}_K/p\mathcal{O}_K$ has no nonzero nilpotent elements by Theorem 6.28, so $\mathcal{O}_F/p\mathcal{O}_F$ has no nonzero nilpotent elements, so p is unramified in F.

If p is totally ramified in K, write $p\mathcal{O}_K = \mathfrak{p}^n$. Since $f(\mathfrak{p}|p) = 1$, the natural embedding $\mathbf{F}_p \to \mathcal{O}_K/\mathfrak{p}$ is an isomorphism. Since $\mathfrak{p}$ is the only prime ideal factor of $p\mathcal{O}_K$, the only prime ideal factor of $p\mathcal{O}_F$ is $\mathfrak{p} \cap \mathcal{O}_F$ (Remark 6.50). Set $\mathfrak{q} = \mathfrak{p} \cap \mathcal{O}_F$ and write $p\mathcal{O}_F = \mathfrak{q}^e$ for some $e \geqslant 1$. Then $[F : \mathbf{Q}] = e(\mathfrak{q}|p)f(\mathfrak{q}|p) = ef(\mathfrak{q}|p)$. Since $\mathcal{O}_F/\mathfrak{q} \hookrightarrow \mathcal{O}_K/\mathfrak{p}$ and $[\mathcal{O}_K/\mathfrak{p} : \mathbf{F}_p] = 1$, we get $[\mathcal{O}_F/\mathfrak{q} : \mathbf{F}_p] = 1$. Therefore $[F : \mathbf{Q}] = e$, so p is totally ramified in F. ■

Theorem 6.52. *Let K and L be number fields with $m = [K : \mathbf{Q}]$, $n = [L : \mathbf{Q}]$, and assume there is a prime p that is totally ramified in K and unramified in L. Then*

(a) $K \cap L = \mathbf{Q}$,

(b) $[KL : \mathbf{Q}] = mn$.

$$\begin{tikzcd}
 & KL \arrow[dl, no head] \arrow[dr, no head] & \\
K & & L \\
 & \mathbf{Q} \arrow[ul, no head, "p \text{ tot. ram.}"] \arrow[ur, no head, "p \text{ unram.}"'] &
\end{tikzcd}$$

Proof. a) Let $F = K \cap L$. Then p is unramified and totally ramified in F by Lemma 6.51. Total ramification in F implies $p\mathcal{O}_F = \mathfrak{p}^{[F:\mathbf{Q}]}$. Being unramified in F implies all prime factors of $p\mathcal{O}_F$ have multiplicity 1, so $[F : \mathbf{Q}] = 1$.

b) Since p is totally ramified in K, $K = \mathbf{Q}(\alpha)$ where α is an algebraic integer that is the root of an Eisenstein polynomial at p, say

$$f(T) = T^m + c_{m-1}T^{m-1} + \cdots + c_1T + c_0 \in \mathbf{Z}[T].$$

The composite field KL is $L(\alpha)$, so $[KL : \mathbf{Q}] = [L(\alpha) : L][L : \mathbf{Q}] = [L(\alpha) : L]n$. The polynomial $f(T)$, viewed in $\mathcal{O}_L[T]$, is Eisenstein at any $\mathfrak{p}_i$: $c_j \equiv 0 \bmod \mathfrak{p}_i$ since $p \in \mathfrak{p}_i$, and since c_0 is divisible by p just once and p is divisible by $\mathfrak{p}_i$ just once (unramified), c_0 is divisible by $\mathfrak{p}_i$ just once. Eisenstein polynomials are irreducible (Exercise 4.24), so $[L(\alpha) : L] = \deg f = m$. ■

If K and L are finite extensions of some field F, the simplest way to conclude that $K \cap L = F$ and $[KL : F] = [K : F][L : F]$ is to assume $[K : F]$ and $[L : F]$ are relatively prime. That is a consequence of field degree considerations only, taking into account no special structure in the fields. For number fields, Theorem 6.52 tells us we can use the special structure of ramification to show $[KL:\mathbf{Q}] = [K:\mathbf{Q}][L:\mathbf{Q}]$ without hypotheses on $[K:\mathbf{Q}]$ and $[L:\mathbf{Q}]$.

Example 6.53. The field $\mathbf{Q}(\sqrt[3]{2})$ is totally ramified at 2 and $\mathbf{Q}(\alpha)$ where $\alpha^3 - \alpha - 1 = 0$ is unramified at 2 (its discriminant is -23), so $[\mathbf{Q}(\sqrt[3]{2}, \alpha) : \mathbf{Q}] = 9$.

Theorem 6.54. *For every $m \geqslant 1$, $[\mathbf{Q}(\zeta_m) : \mathbf{Q}] = \varphi(m)$.*

Proof. We may take $m > 1$. When $m = p^r$ is a power of a prime p then $\zeta_{p^r} - 1$ is a root of the Eisenstein polynomial $\Phi_{p^r}(T+1)$, which is irreducible by Eisenstein's criterion, so the degree of the field $\mathbf{Q}(\zeta_{p^r} - 1) = \mathbf{Q}(\zeta_{p^r})$ over $\mathbf{Q}$ is $\deg(\Phi_{p^r}(T+1)) = p^{r-1}(p-1) = \varphi(p^r)$.

We now induct on the number of prime factors of m. When m is not a prime power, let p be a prime factor of m and write $m = p^r m'$ with $r \geqslant 1$ and $p \nmid m'$. Then $\mathbf{Q}(\zeta_m) = \mathbf{Q}(\zeta_{p^r}, \zeta_{m'})$.

- Since p^r is a power of p, p is totally ramified in $\mathbf{Q}(\zeta_{p^r})$.
- Since $p \nmid m'$, p is unramified in $\mathbf{Q}(\zeta_{m'})$.

Apply Theorem 6.52 to the pair of fields $\mathbf{Q}(\zeta_{p^r})$ and $\mathbf{Q}(\zeta_{m'})$ and conclude $[\mathbf{Q}(\zeta_{p^r}, \zeta_{m'}) : \mathbf{Q}] = [\mathbf{Q}(\zeta_{p^r}) : \mathbf{Q}][\mathbf{Q}(\zeta_{m'}) : \mathbf{Q}]$. From the prime-power case, $[\mathbf{Q}(\zeta_{p^r}) : \mathbf{Q}] = \varphi(p^r)$. Since m' has fewer prime factors than m, $[\mathbf{Q}(\zeta_{m'}) : \mathbf{Q}] = \varphi(m')$ by induction, so $[\mathbf{Q}(\zeta_{p^r}, \zeta_{m'}) : \mathbf{Q}] = \varphi(p^r)\varphi(m') = \varphi(m)$. ■

Quadratic fields are all described as $\mathbf{Q}(\sqrt{d})$ with nonsquare $d \in \mathbf{Z}$, but there are coincidences: $\mathbf{Q}(\sqrt{d}) = \mathbf{Q}(\sqrt{d'})$ if and only if d'/d is a perfect square. The squarefree $d \neq 1$ describe each quadratic field exactly once. Cyclotomic fields $\mathbf{Q}(\zeta_m)$ have a similar feature. They are parametrized by positive integers m, but there are coincidences: $\mathbf{Q}(\zeta_3) = \mathbf{Q}(\zeta_6)$ and $\mathbf{Q}(\zeta_5) = \mathbf{Q}(\zeta_{10})$ since $-\zeta_3$ has order 6 and $-\zeta_5$ has order 10. The extent to which different positive integers parametrize the same cyclotomic field is more restricted than for quadratic fields.

Corollary 6.55. *For $m < n$, $\mathbf{Q}(\zeta_m) = \mathbf{Q}(\zeta_n)$ if and only if m is odd and $n = 2m$.*

Proof. When m is odd, $-\zeta_m$ has order $2m$, so $\zeta_{2m} \in \mathbf{Q}(\zeta_m)$. Therefore $\mathbf{Q}(\zeta_{2m}) \subset \mathbf{Q}(\zeta_m)$. The reverse inclusion is clear, so $\mathbf{Q}(\zeta_m) = \mathbf{Q}(\zeta_{2m})$.

Conversely, assume $\mathbf{Q}(\zeta_m) = \mathbf{Q}(\zeta_n)$ with $m < n$. Computing the degree of this field over $\mathbf{Q}$, $\varphi(m) = \varphi(n)$. Since ζ_m and ζ_n are in $\mathbf{Q}(\zeta_m)$, so is $\zeta_{[m,n]}$.

(It is a basic theorem from group theory that in a finite abelian group, if there are elements with orders m and n then there is an element of order $[m,n]$. Apply this to the group $\langle \zeta_m, \zeta_n \rangle$.) Therefore $\mathbf{Q}(\zeta_{[m,n]}) \subset \mathbf{Q}(\zeta_m)$. The reverse inclusion is obvious, so $\mathbf{Q}(\zeta_{[m,n]}) = \mathbf{Q}(\zeta_m)$. Taking degrees over $\mathbf{Q}$ tells us $\varphi([m,n]) = \varphi(m)$. Write $[m,n] = md$. Then (Exercise 6.9)

$$\varphi([m,n]) = \varphi(md) = \varphi(m)\varphi(d)\frac{(m,d)}{\varphi((m,d))} \geqslant \varphi(m)\varphi(d).$$

Therefore $\varphi(m) \geqslant \varphi(m)\varphi(d)$, so $\varphi(d) = 1$, so $d = 1$ or 2. Thus $[m,n] = m$ or $[m,n] = 2m$. Since n is a factor of $[m,n]$ and $n > m$, we must have $n = 2m$. If m is even then $\varphi(n) = \varphi(2m) = 2\varphi(m) > \varphi(m)$, a contradiction. Thus m is odd. ■

By Corollary 6.55, we can describe all cyclotomic fields as $\mathbf{Q}(\zeta_m)$ uniquely by using only m that are not twice an odd number, which means using $m \not\equiv 2 \bmod 4$. A close look at the proof shows that the discrepancy with 2 comes from the fact that $\varphi(d) = 1$ (that is, $(\mathbf{Z}/d\mathbf{Z})^\times$ is trivial) only for $d = 1$ and $d = 2$.

We turn now to the computation of the integers of $\mathbf{Q}(\zeta_m)$. It will be based on the following general theorem about rings of integers in a composite of two number fields.

Theorem 6.56. *Let F be a field and K/F and L/F be finite extensions in a common larger field. Set $m = [K : F]$ and $n = [L : F]$. Assume $[KL : F] = mn$. Let $\{e_1, \ldots, e_m\}$ be an F-basis of K and $\{f_1, \ldots, f_n\}$ be an F-basis of L.*

(a) *The set of products $\{e_i f_j\}$ is an F-basis of KL.*

(b) *For $\alpha \in K$, $\chi_{KL/L,\alpha}(T) = \chi_{K/F,\alpha}(T)$. In particular, for $\alpha \in K$,*

$$\mathrm{Tr}_{KL/L}(\alpha) = \mathrm{Tr}_{K/F}(\alpha).$$

(c) *Let K and L be number fields of respective degrees m and n over $\mathbf{Q}$ and assume $[KL : \mathbf{Q}] = mn$. Set*

$$d = (\mathrm{disc}(K), \mathrm{disc}(L)), \quad \mathcal{O}_K\mathcal{O}_L = \left\{ \sum_{k=1}^{r} x_k y_k : r \geqslant 1, x_k \in \mathcal{O}_K, y_k \in \mathcal{O}_L \right\}.$$

Then $\mathcal{O}_K\mathcal{O}_L \subset \mathcal{O}_{KL} \subset \frac{1}{d}\mathcal{O}_K\mathcal{O}_L$, so $\mathcal{O}_{KL} = \mathcal{O}_K\mathcal{O}_L$ if K and L have relatively prime discriminants.

The first two parts of this theorem are field theory. Only the last part involves number fields.

Proof. (a) We will show $\{e_if_j\}$ spans KL as an F-vector space. Since this spanning set has size mn and $mn = [KL : F]$ by hypothesis, this spanning set is a basis.

Since K and L are *finite* extensions of F, $KL = \{\sum_{k=1}^{r} x_ky_k : r \geqslant 1, x_k \in K, y_k \in L\}$. For $x \in K$ and $y \in L$, write $x = \sum_{i=1}^{m} a_ie_i$ and $y = \sum_{j=1}^{n} b_jf_j$, where a_i and b_j are in F. Then $xy = \sum_{i,j} a_ib_je_if_j \in \sum_{i,j} Fe_if_j$. Taking finite sums of these, we get $KL \subset \sum_{i,j} Fe_jf_j$, so $KL = \sum_{i,j} Fe_if_j$.

(b) For $\alpha \in K$, let $(c_{k\ell})$ be the matrix for $m_\alpha \colon K \to K$ as an F-linear map with respect to the F-basis $\{e_1, \ldots, e_m\}$ of K:

$$\alpha e_\ell = \sum_{k=1}^{m} c_{k\ell}e_k. \tag{6.17}$$

The set $\{e_1, \ldots, e_m\}$ is also an L-basis of KL: $\{e_if_j\}$ spans KL over F, so it also spans over L, and e_if_j is a scalar multiple of e_i over L. Thus $\{e_1, \ldots, e_m\}$ spans KL over L, and $[KL : L] = [KL : F]/[L : F] = mn/n = m$, so this spanning set is a basis. By (6.17), $(c_{k\ell})$ is the matrix for multiplication by α on KL over L with respect to $\{e_1, \ldots, e_m\}$, so $\chi_{KL/L,\alpha}(T) = \det(TI_n - (c_{k\ell})) = \chi_{K/F,\alpha}(T)$.

(c) Let $\mathcal{O}_K$ have $\mathbf{Z}$-basis $\{e_1, \ldots, e_m\}$ and $\mathcal{O}_L$ have $\mathbf{Z}$-basis $\{f_1, \ldots, f_n\}$. Then $\mathcal{O}_K\mathcal{O}_L = \sum_{i,j} \mathbf{Z}e_if_j$. For $z \in \mathcal{O}_{KL}$, write

$$z = \sum_{i,j} c_{ij}e_if_j$$

with $c_{ij} \in \mathbf{Q}$. We want to show each c_{ij} has denominator d.

Since $d = (\mathrm{disc}(K), \mathrm{disc}(L))$, to show each c_{ij} has denominator d we first show the c_{ij}'s have denominator dividing $\mathrm{disc}(K)$. We will adapt the method used to show $\mathcal{O}_K \subset \frac{1}{\mathrm{disc}(M)}M$ when M is a $\mathbf{Z}$-lattice in K (see (3.6)). Treating KL as an L-vector space with basis $\{e_1, \ldots, e_m\}$,

$$z = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} c_{ij}f_j \right) e_i,$$

with the inner sums $\sum_j c_{ij}f_j$ lying in L. Set $\alpha_i := \sum_{j=1}^{n} c_{ij}f_j \in L$, so $z =$

$\sum_{i=1}^m \alpha_i e_i$. For $i' = 1, \ldots, m$,

$$ze_{i'} = \sum_{i=1}^m \alpha_i e_i e_{i'}.$$

Apply $\mathrm{Tr}_{KL/L}$ to both sides:

$$\mathrm{Tr}_{KL/L}(ze_{i'}) = \sum_{i=1}^m \alpha_i \mathrm{Tr}_{KL/L}(e_i e_{i'}) = \sum_{i=1}^m \alpha_i \mathrm{Tr}_{K/\mathbf{Q}}(e_i e_{i'}).$$

Thus

$$\begin{pmatrix} \mathrm{Tr}_{KL/L}(ze_1) \\ \vdots \\ \mathrm{Tr}_{KL/L}(ze_m) \end{pmatrix} = \begin{pmatrix} \mathrm{Tr}_{K/\mathbf{Q}}(e_1e_1) & \cdots & \mathrm{Tr}_{K/\mathbf{Q}}(e_1e_m) \\ \vdots & \ddots & \vdots \\ \mathrm{Tr}_{K/\mathbf{Q}}(e_1e_m) & \cdots & \mathrm{Tr}_{K/\mathbf{Q}}(e_me_m) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix}.$$

The traces on the left are in $\mathcal{O}_L$, since $ze_{i'}$ is an algebraic integer, and the traces on the right are in $\mathbf{Z}$. From this equation and Cramer's formula, each α_i can be written as a ratio with numerator in $\mathcal{O}_L$ and denominator $\det(\mathrm{Tr}_{K/\mathbf{Q}}(e_ie_{i'})) = \mathrm{disc}(K)$. Since $\mathcal{O}_L = \bigoplus_{j=1}^n \mathbf{Z}f_j$ and $\alpha_i \in \frac{1}{\mathrm{disc}(K)}\mathcal{O}_L$, each c_{ij} is a fraction with denominator $\mathrm{disc}(K)$.

By a similar argument with the roles of K and L reversed, each c_{ij} can be written as a fraction with denominator $\mathrm{disc}(L)$. Therefore the reduced form denominator of each c_{ij} divides both $\mathrm{disc}(K)$ and $\mathrm{disc}(L)$, so it divides their gcd d. This puts z in $\frac{1}{d}\sum_{i,j}\mathbf{Z}e_if_j = \frac{1}{d}\mathcal{O}_K\mathcal{O}_L$, so $\mathcal{O}_{KL} \subset \frac{1}{d}\mathcal{O}_K\mathcal{O}_L$. ■

Theorem 6.57. *For any $m \geqslant 1$, the ring of integers in $\mathbf{Q}(\zeta_m)$ is $\mathbf{Z}[\zeta_m]$.*

Proof. First we assume $m = p^r$ is a prime power.

Let $K = \mathbf{Q}(\zeta_{p^r})$. For any prime $q \neq p$, $T^{p^r} - 1 \bmod q$ is separable, so its factor $\Phi_{p^r}(T) \bmod q$ is separable. Thus $q \nmid \mathrm{disc}(\Phi_{p^r}(T))$. Therefore the only prime factor of $\mathrm{disc}(\Phi_{p^r}(T))$ can be p, so it is a power of p (up to sign). Since

$$\mathrm{disc}(\Phi_{p^r}(T)) = \mathrm{disc}(\mathbf{Z}[\zeta_{p^r}]) = [\mathcal{O}_K : \mathbf{Z}[\zeta_{p^r}]]^2 \,\mathrm{disc}(K),$$

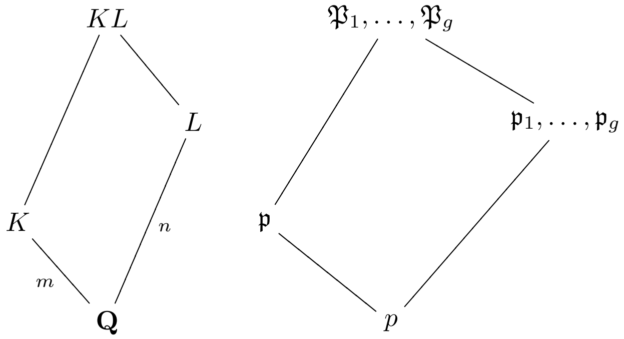
the index $[\mathcal{O}_K : \mathbf{Z}[\zeta_{p^r}]]$ is a power of p. Since $\Phi_{p^r}(T+1)$ is Eisenstein at p with root $\zeta_{p^r} - 1$, the index $[\mathcal{O}_K : \mathbf{Z}[\zeta_{p^r} - 1]]$ is not divisible by p by Theorem 6.32. The ring $\mathbf{Z}[\zeta_{p^r} - 1]$ is $\mathbf{Z}[\zeta_{p^r}]$, so $[\mathcal{O}_K : \mathbf{Z}[\zeta_{p^r}]]$ is a both a power of p and not divisible by p, so this index is 1, which means $\mathcal{O}_K = \mathbf{Z}[\zeta_{p^r}]$.

Now suppose m is not a prime power, so it has more than one prime factor. Write $m = p^r m'$ where p doesn't divide m. By induction on the number of prime factors we may suppose $\mathbf{Q}(\zeta_{m'})$ has ring of integers $\mathbf{Z}[\zeta_{m'}]$. Since p does not divide m', it is unramified in $\mathbf{Q}(\zeta_{m'})$, so $p \nmid \operatorname{disc}(\mathbf{Q}(\zeta_{m'}))$. The only prime that ramifies in $\mathbf{Q}(\zeta_{p^r})$ is p, so $\mathbf{Q}(\zeta_{p^r})$ and $\mathbf{Q}(\zeta_{m'})$ have relatively prime discriminants. Theorem 6.56 tells us the ring of integers of $\mathbf{Q}(\zeta_{p^r}, \zeta_{m'})$ is $\mathbf{Z}[\zeta_{p^r}]\mathbf{Z}[\zeta_{m'}] = \mathbf{Z}[\zeta_{p^r}, \zeta_{m'}] = \mathbf{Z}[\zeta_m]$. ■

Remark 6.58. Kummer studied arithmetic in the ring $\mathbf{Z}[\zeta_m]$ because it is the "natural" subring of $\mathbf{Q}(\zeta_m)$. He didn't know $\mathbf{Z}[\zeta_m]$ is integrally closed. He was lucky that it is, since otherwise there wouldn't have been a unique factorization of ideals to discover in that ring. (For any number field K, a proper subring of $\mathcal{O}_K$ whose fraction field is K is not integrally closed.)

To factor prime numbers into prime ideals in $\mathbf{Q}(\zeta_m)$, we will use the following theorem about ramification in a composite of number fields such that a prime number is totally ramified in one field and unramified in the other field.

Theorem 6.59. *Let K and L be number fields with $m = [K : \mathbf{Q}]$, $n = [L : \mathbf{Q}]$, and assume there is a prime p that is totally ramified in K and unramified in L. Write $p\mathcal{O}_K = \mathfrak{p}^m$ and $p\mathcal{O}_L = \mathfrak{p}_1 \cdots \mathfrak{p}_g$. Then $p\mathcal{O}_{KL} = \mathfrak{P}_1^m \cdots \mathfrak{P}_g^m$, where $\mathfrak{p}_i\mathcal{O}_{KL} = \mathfrak{P}_i^m$ and $f(\mathfrak{P}_i|p) = f(\mathfrak{p}_i|p)$.*


KL
L
K
n
m
$\mathbf{Q}$
$\mathfrak{P}_1, \ldots, \mathfrak{P}_g$
$\mathfrak{p}_1, \ldots, \mathfrak{p}_g$
$\mathfrak{p}$
p

The important idea conveyed by this theorem is that when we compose two number fields such that a prime p is totally ramified in one and is unramified in the other, each field controls a different aspect of the factorization of (p) in the composite field: the ramification indices in the composite field match the one p has in the totally ramified extension and the residue field degrees and number of prime ideal factors in the composite field match those parameters for p in the unramified extension. Loosely, the field in which the relevant parameter (e_i, f_i, g) is nontrivial contributes that value to the composite field.

Proof. The extension of ideals from one ring of integers $\mathcal{O}_F$ to a larger ring of integers $\mathcal{O}_E$ is multiplicative: $\mathfrak{a}\mathcal{O}_E \cdot \mathfrak{a}'\mathcal{O}_E = \mathfrak{a}\mathfrak{a}'\mathcal{O}_E$ (Exercise 6.3). In particular, when we extend the ideal equations $p\mathcal{O}_K = \mathfrak{p}^m$ and $p\mathcal{O}_L = \mathfrak{p}_1 \cdots \mathfrak{p}_g$ to $\mathcal{O}_{KL}$ we get two formulas for $p\mathcal{O}_{KL}$:

$$p\mathcal{O}_{KL} = (\mathfrak{p}\mathcal{O}_{KL})^m, \quad p\mathcal{O}_{KL} = \mathfrak{p}_1\mathcal{O}_{KL} \cdots \mathfrak{p}_g\mathcal{O}_{KL}.$$

By Lemma 6.49, for each prime $\mathfrak{p}_i$ there is at least one prime in $\mathcal{O}_{KL}$, say $\mathfrak{P}_i$, which divides $\mathfrak{p}_i\mathcal{O}_{KL}$. There is a natural embedding of the residue field $\mathcal{O}_K/\mathfrak{p}_i$ into the residue field $\mathcal{O}_{LK}/\mathfrak{P}_i$ since the kernel of the natural map $\mathcal{O}_K \to \mathcal{O}_{KL}/\mathfrak{P}_i$ is $\mathfrak{P}_i \cap \mathcal{O}_K = \mathfrak{p}_i$ (Remark 6.50). From the tower of naturally embedded residue fields

$$\mathbf{Z}/p\mathbf{Z} \subset \mathcal{O}_K/\mathfrak{p}_i \subset \mathcal{O}_{KL}/\mathfrak{P}_i$$

we obtain $f(\mathfrak{P}_i|p) \geqslant f(\mathfrak{p}_i|p)$.

To estimate $e(\mathfrak{P}_i|p)$, we use the equation $p\mathcal{O}_{KL} = (\mathfrak{p}\mathcal{O}_{KL})^m$, which tells us that any prime ideal factor of $p\mathcal{O}_{KL}$ has multiplicity divisible by m, so $e(\mathfrak{P}_i|p) \geqslant m$. Since the ideals $\mathfrak{P}_i$ are primes over p in KL, but we don't yet know if they are all such primes, we have $\sum_{i=1}^{g} e(\mathfrak{P}_i|p)f(\mathfrak{P}_i|p) \leqslant [KL : \mathbf{Q}]$. If we are missing some primes over p in KL then this inequality is strict. Since

$$mn = [KL : \mathbf{Q}] \geqslant \sum_{i=1}^{g} e(\mathfrak{P}_i|p)f(\mathfrak{P}_i|p) \geqslant \sum_{i=1}^{g} mf(\mathfrak{p}_i|p) = m\sum_{i=1}^{g} f(\mathfrak{p}_i|p) = mn,$$

there is equality throughout, so the $\mathfrak{P}_i$'s are all the primes lying over p in KL, $e(\mathfrak{P}_i|p) = m$, and $f(\mathfrak{P}_i|p) = f(\mathfrak{p}_i|p)$. Since $p\mathcal{O}_{KL} = \mathfrak{P}_1^m \cdots \mathfrak{P}_g^m$, $p\mathcal{O}_{KL} = \mathfrak{p}_1\mathcal{O}_{KL} \cdots \mathfrak{p}_g\mathcal{O}_{KL}$, and $\mathfrak{P}_i \mid \mathfrak{p}_i\mathcal{O}_{KL}$, we have $\mathfrak{p}_i\mathcal{O}_{KL} = \mathfrak{P}_i^m$ for all i. ■

Corollary 6.60. *If $p \nmid m'$ then the residue field degree of each prime over p in $\mathbf{Q}(\zeta_{m'})$ is the order of p in $(\mathbf{Z}/m'\mathbf{Z})^\times$.*

Proof. Let $\mathfrak{p}$ be a prime over p in $\mathbf{Q}(\zeta_{m'})$. The residue field $\mathbf{Z}[\zeta_{m'}]/\mathfrak{p}$ is generated as a ring over $\mathbf{F}_p$ by $\zeta_{m'}$ mod $\mathfrak{p}$: $\mathbf{Z}[\zeta_{m'}]/\mathfrak{p} = \mathbf{F}_p(\zeta_{m'} \bmod \mathfrak{p})$. What is the degree of this finite field over $\mathbf{F}_p$?

Since $T^{m'}-1$ is separable in characteristic p, its factorization

$$T^{m'}-1=\prod_{i=0}^{m'-1}(T-\zeta_{m'}^i)$$

in characteristic 0 must reduce mod $\mathfrak{p}$ to distinct linear factors, so the powers $\zeta_{m'}^i$ mod $\mathfrak{p}$ are distinct. That is, $\zeta_{m'}$ continues to have order m' when we reduce it mod $\mathfrak{p}$. (Compare this with reducing i in $\mathbf{Z}[i]/(1+i)=\mathbf{F}_2$: $i\equiv 1 \bmod 1+i$, so the order of i drops from 4 down to 1. Here $p=2$ divides $m'=4$.)

The theory of finite fields tells us $[\mathbf{F}_p(\zeta_{m'} \bmod \mathfrak{p}):\mathbf{F}_p]$ is the smallest $f\geqslant 1$ such that $\zeta_{m'}^{p^f}\equiv\zeta_{m'} \bmod \mathfrak{p}$. Since $\zeta_{m'}$ mod $\mathfrak{p}$ has order m', $\zeta_{m'}^{p^f}\equiv\zeta_{m'} \bmod \mathfrak{p}$ if and only if $p^f\equiv 1 \bmod m'$. The least such f is the order of p mod m'. ■

Since the order of p in $(\mathbf{Z}/m'\mathbf{Z})^\times$ is determined by p mod m', the shape of the factorization of p in $\mathbf{Q}(\zeta_{m'})$ depends only on p mod m': primes in the same congruence class of $(\mathbf{Z}/m'\mathbf{Z})^\times$ factor in the same way in $\mathbf{Q}(\zeta_{m'})$.

Example 6.61. In $\mathbf{Q}(\zeta_7)$, 7 is totally ramified. Any prime p other than 7 does not divide 7, so p is unramified in $\mathbf{Q}(\zeta_7)$: $e(\mathfrak{p}|p)=1$ for all $\mathfrak{p}$ lying over p in $\mathbf{Q}(\zeta_7)$. The shape of the factorization of p in $\mathbf{Q}(\zeta_7)$ is determined by p mod 7 and is given in Table 6.7. The second column is the common residue field degree of all primes over p in $\mathbf{Q}(\zeta_7)$. For example, if $p\equiv 2 \bmod 7$ then each prime over p in $\mathbf{Q}(\zeta_7)$ has a residue field of size p^3.

p mod 7	order of p mod 7	(p)
1	1	splits comp.
6	2	$\mathfrak{p}\mathfrak{p}'\mathfrak{p}''$
$2,4$	3	$\mathfrak{p}\mathfrak{p}'$
$3,5$	6	prime

Table 6.7: Factoring primes in $\mathbf{Q}(\zeta_7)$.

Example 6.62. In $\mathbf{Q}(\zeta_8)$, 2 is totally ramified. Any prime p other than 2 is unramified in $\mathbf{Q}(\zeta_8)$ and the shape of the factorization of p in $\mathbf{Q}(\zeta_8)$ depends on p mod 8 as described in Table 6.8. Notice in particular that no prime stays prime in $\mathbf{Q}(\zeta_8)$, since the group $(\mathbf{Z}/8\mathbf{Z})^\times$ has no element of order 4 (no generator).

Example 6.63. In $\mathbf{Q}(\zeta_9)$, 3 is totally ramified and the shape of the factorization of any other prime is in Table 6.9.

p mod 8	order of p mod 8	(p)
1	1	splits comp.
$3, 5, 7$	2	$\mathfrak{p}\mathfrak{p}'$

Table 6.8: Factoring primes in $\mathbf{Q}(\zeta_8)$.

p mod 9	order of p mod 9	(p)
1	1	splits comp.
8	2	$\mathfrak{p}\mathfrak{p}'\mathfrak{p}''$
$4, 7$	3	$\mathfrak{p}\mathfrak{p}'$
$2, 5$	6	prime

Table 6.9: Factoring primes in $\mathbf{Q}(\zeta_9)$.

Now we factor any prime in any cyclotomic field.

Theorem 6.64. *Write $m = p^r m'$ where $p \nmid m'$. In $\mathbf{Z}[\zeta_m]$,*

$$(p) = (\mathfrak{p}_1 \cdots \mathfrak{p}_g)^{\varphi(p^r)},$$

where the $\mathfrak{p}_i$'s are distinct, for all i the residue field degree $f(\mathfrak{p}_i|p)$ is the order f of $p \bmod m'$, and $g = \varphi(m')/f$. In particular, if $p \nmid m$ then

$$(p) = \mathfrak{p}_1 \cdots \mathfrak{p}_g,$$

where the $\mathfrak{p}_i$'s are distinct and $f(\mathfrak{p}_i|p)$ is the order of $p \bmod m$ for all i.

Proof. To analyze the behavior of p in $\mathbf{Q}(\zeta_m)$ we focus separate attention on the subfields $\mathbf{Q}(\zeta_{p^r})$ and $\mathbf{Q}(\zeta_{m'})$ in the diagram below. In $\mathbf{Q}(\zeta_{p^r})$, p is totally ramified with prime factor $(p, \zeta_{p^r} - 1)$ and p is unramified in $\mathbf{Q}(\zeta_{m'})$. Let

$\mathfrak{q}_1, \ldots, \mathfrak{q}_g$ be the prime factors of p in $\mathbf{Q}(\zeta_{m'})$. (We don't yet know what g is.)

$$\begin{tikzcd}
 & \mathbf{Q}(\zeta_m) \arrow[ddl, dash] \arrow[dr, dash] & \\
 & & \mathbf{Q}(\zeta_{m'}) \arrow[ddl, dash, "\varphi(m')"] \\
\mathbf{Q}(\zeta_{p^r}) \arrow[dr, dash, "\varphi(p^r)"'] & & \\
 & \mathbf{Q} &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & \mathfrak{p}_1, \ldots, \mathfrak{p}_g \arrow[ddl, dash] \arrow[dr, dash] & \\
 & & \mathfrak{q}_1, \ldots, \mathfrak{q}_g \arrow[ddl, dash] \\
(p, \zeta_{p^r} - 1) \arrow[dr, dash] & & \\
 & p &
\end{tikzcd}$$

The factorization of p in $\mathbf{Q}(\zeta_m)$ now follows from Theorem 6.59 with $K = \mathbf{Q}(\zeta_{p^r})$ and $L = \mathbf{Q}(\zeta_{m'})$, allowing that $\mathfrak{p}_i$ and $\mathfrak{P}_i$ there have become $\mathfrak{q}_i$ and $\mathfrak{p}_i$ here: $\mathfrak{q}_i\mathbf{Z}[\zeta_m] = \mathfrak{p}_i^{\varphi(p^r)}$ for some prime ideal $\mathfrak{p}_i$, $e(\mathfrak{p}_i|p) = \varphi(p^r)$, and $f(\mathfrak{p}_i|p) = f(\mathfrak{q}_i|p)$. In Corollary 6.60 we saw all primes over p in $\mathbf{Q}(\zeta_{m'})$ have residue field degree equal to the order f of p mod m'. Since $[\mathbf{Q}(\zeta_{m'}) : \mathbf{Q}] = \sum_{i=1}^{g} e(\mathfrak{q}_i|p)f(\mathfrak{q}_i|p) = \sum_{i=1}^{g} f = fg$, $g = \varphi(m')/f$. ■

The shape of factorizations of prime numbers in cyclotomic fields fits into broader themes:

- For any prime p, the primes in $\mathbf{Q}(\zeta_m)$ lying over p have the same ramification index and the same residue field degree. This is a general feature of the factorization of any p in a *Galois* extension. (We've seen many non-Galois cubic fields where a ramified p factors as $\mathfrak{p}\mathfrak{q}^2$ or an unramified p factors as $(p) = \mathfrak{p}\mathfrak{q}$ with $\mathrm{N}(\mathfrak{p}) = p$ and $\mathrm{N}(\mathfrak{q}) = p^2$.)

- The shape of the factorization of a prime $p \nmid m$ in $\mathbf{Q}(\zeta_m)$ is determined by p mod m, so factorizations of unramified primes in $\mathbf{Q}(\zeta_m)$ are governed by congruence conditions.

  One of the main results of class field theory is that the shape of the factorization of unramified primes in a number field can be described by congruence conditions on the primes to some modulus if and only if the extension is an *abelian* Galois extension of $\mathbf{Q}$. (The primes which split completely in $\mathbf{Q}(\sqrt[3]{2}, \zeta_3)$, a nonabelian Galois extension of $\mathbf{Q}$, are $31, 43, 109, 127, 157, 223, \ldots$, and are not described by a congruence condition.)

Corollary 6.65. *An odd prime p is ramified in $\mathbf{Q}(\zeta_m)$ if and only if $p \mid m$, and 2 is ramified in $\mathbf{Q}(\zeta_m)$ if and only if $4 \mid m$.*

An odd prime p is inert in $\mathbf{Q}(\zeta_m)$ if and only if $p \bmod m$ generates $(\mathbf{Z}/m\mathbf{Z})^\times$ and p splits completely in $\mathbf{Q}(\zeta_m)$ if and only if $p \equiv 1 \bmod m$. The same conditions apply to 2 provided m is odd.

Proof. By Theorem 6.64, the ramification index of each prime over p in $\mathbf{Q}(\zeta_m)$ is $\varphi(p^r)$, where p^r is the largest power of p dividing m. We have $\varphi(p^r) = 1$ if and only if $p^r = 1$ or 2. So p is ramified in $\mathbf{Q}(\zeta_m)$ if and only if $\varphi(p^r) > 1$, which is equivalent to $r \geqslant 1$ (*i.e.*, $p \mid m$) when $p \neq 2$ and to $r \geqslant 2$ (*i.e.*, $4 \mid m$) when $p = 2$.

A prime p that is inert or splits completely in $\mathbf{Q}(\zeta_m)$ is unramified, in which case whether it is inert or splits completely is determined by whether its residue field degree is $\varphi(m)$ or 1. The residue field degree is the order of $p \bmod m$ when $p \nmid m$. ■

With the convention of labeling cyclotomic fields as $\mathbf{Q}(\zeta_m)$ with $m \not\equiv 2 \bmod 4$ (see Corollary 6.55), Corollary 6.65 assumes a simpler form which has no exception for the prime 2: p ramifies in $\mathbf{Q}(\zeta_m)$ if and only if $p \mid m$, p is inert if and only if $p \bmod m$ generates $(\mathbf{Z}/m\mathbf{Z})^\times$, and p splits completely if and only if $p \equiv 1 \bmod m$.[4]

Example 6.66. In $\mathbf{Q}(\zeta_{15})$ and $\mathbf{Q}(\zeta_{21})$, no prime stays prime since $(\mathbf{Z}/15\mathbf{Z})^\times$ and $(\mathbf{Z}/21\mathbf{Z})^\times$ are not cyclic.

6.7 Exercises

1. Let $f(T) \in \mathbf{Z}[T]$ be Eisenstein at a prime p. If p is unramified in a number field K, show $f(T)$ is Eisenstein at every prime ideal $\mathfrak{p}$ lying over p in K (see Exercise 4.24c).

2. Let $f(T) \in \mathbf{Z}[T]$ be monic of degree n and $K = \mathbf{Q}(\alpha)$ where α is a root of $f(T)$.

   a) If $(p) = (x)^n$ for some $x \in \mathcal{O}_K$, show $n = [K : \mathbf{Q}]$ and (x) is a prime ideal, so $f(T)$ is irreducible over $\mathbf{Q}$ and p is totally ramified in K.

[4] All odd primes are 1 mod 2, so actually p splits completely in $\mathbf{Q}(\zeta_m)$ if and only if $\{p \equiv 1 \bmod m\}$ for all $m \geqslant 1$ except $m = 2$

b) Show $p = \prod_{i\in(\mathbf{Z}/p^r\mathbf{Z})^\times}(1-\zeta_{p^r}^i)$. (Hint: Look at the factorization of $\Phi_{p^r}(T)$.)

c) When $p \nmid i$, show $1-\zeta_{p^r}^i$ and $1-\zeta_{p^r}$ divide each other. (Hint: show ζ_{p^r} and $\zeta_{p^r}^i$ are each powers of each other.)

b) In $\mathbf{Q}(\zeta_{p^r})$, show $(p) = (\zeta_{p^r}-1)^{\varphi(p^r)}$ and conclude that the cyclotomic polynomial $\Phi_{p^r}(T)$ is irreducible in $\mathbf{Q}[T]$. (This is an alternate proof of irreducibility, not relying on the Eisenstein criterion.)

3. Let F be a number field and E/F be a finite extension. For any ideal $\mathfrak{a}$ in $\mathcal{O}_F$, the extended ideal $\mathfrak{a}\mathcal{O}_E$ is defined in (6.16).

   a) For $\alpha \in \mathcal{O}_F$, show $(\alpha\mathcal{O}_F)\mathcal{O}_E = \alpha\mathcal{O}_E$.

   b) For ideals $\mathfrak{a}$ and $\mathfrak{a}'$ in $\mathcal{O}_F$, show $(\mathfrak{a}\mathcal{O}_E)(\mathfrak{a}'\mathcal{O}_E) = (\mathfrak{a}\mathfrak{a}')\mathcal{O}_E$, so extending ideals in $\mathcal{O}_F$ to ideals in $\mathcal{O}_E$ commutes with ideal multiplication.

4. In the quadratic fields $\mathbf{Q}(\sqrt{d})$ with squarefree $d \neq 1$ and $|d| \leqslant 10$, determine how 2 factors (is it inert, split completely, or ramified).

5. In $\mathbf{Z}[\sqrt{14}]$, $(2) = \mathfrak{p}_2^2$ and $(7) = \mathfrak{p}_7^2$ for some prime ideals of norm 2 and 7. By Exercise 5.5, $\mathfrak{p}_2$ and $\mathfrak{p}_7$ are principal. Find generators of these two ideals.

6. Show the ring of integers of $\mathbf{Q}(\sqrt{3}, \sqrt{5})$ is $\mathbf{Z}[\sqrt{3}, \frac{1+\sqrt{5}}{2}]$.

7. Show the ring of integers of $\mathbf{Q}(\sqrt[4]{3})$ is $\mathbf{Z}[\sqrt[4]{3}]$.

8. Determine the splitting laws for all primes in $\mathbf{Q}(\zeta_5)$ and $\mathbf{Q}(\zeta_{12})$. That is, describe all ways prime numbers factor and which prime numbers have which shape for their factorization.

9. For positive integers a and b, show $\varphi(ab) = \varphi(a)\varphi(b)(a,b)/\varphi((a,b))$. In particular, $\varphi(ab) \geqslant \varphi(a)\varphi(b)$ with equality if and only if $(a,b) = 1$.

10. Define $K = \mathbf{Q}(\alpha)$ and $L = \mathbf{Q}(\beta)$, where α and β are roots of the following polynomials:

$$f(T) = T^3 - 21T - 28, \quad g(T) = T^3 - 21T - 35.$$

   a) Show K and L are cubic fields with the same discriminant $3969 = 3^4 \cdot 7^2$.

   b) Factor the primes up to 11 in both fields and conclude that K and L are not isomorphic.

c) Repeat parts a and b with $f(T) = T^3 - T^2 - 30T + 64$ and $g(T) = T^3 + T^2 - 30T + 27$ (this time the common discriminant is $8281 = 7^2 \cdot 13^2$).

11. (Continuation of Exercise 4.13) For $i = 1, 2, 3$, define three cubic fields $K_i = \mathbf{Q}(\alpha_i)$ where α_i is the root of $f_i(T)$:

$$f_1(T) = T^3 - 18T - 6, \quad f_2(T) = T^3 - 36T - 78, \quad f_3(T) = T^3 - 54T - 150.$$

These polynomials are all Eisenstein at 2 and 3, so they are irreducible over $\mathbf{Q}$.

a) Check each polynomial has 3 real roots and discriminant $22356 = 2^2 \cdot 3^5 \cdot 23$.

b) Show K_i has ring of integers $\mathbf{Z}[\alpha_i]$ for $i = 1, 2, 3$. Therefore each K_i has discriminant 22356.

c) Prove the K_i's are nonisomorphic by finding primes with different splitting behavior in each pair of fields.

12. Let α be a root of $f(T) = T^4 + 2T^2 - 2$ and β be a root of $g(T) = T^4 - 2T^2 - 2$.

a) In $\mathbf{Q}(\alpha)$, show 2 is totally ramified and 11 splits completely.

b) In $\mathbf{Q}(\beta)$, show 2 is totally ramified and 11 does not split completely.

c) Show 3 ramifies in both fields, with $(3) = \mathfrak{p}_9^2$ in $\mathbf{Q}(\alpha)$ and $(3) = \mathfrak{p}_3^2\mathfrak{p}_3'^2$ in $\mathbf{Q}(\beta)$. (Hint: Look at ramification in a quadratic subfield.)

d) Show the two fields both have discriminant $-4608 = -2^9 \cdot 3^2$.

13. For a *squarefree* integer $a \neq \pm 1$, verify the following is a $\mathbf{Z}$-basis of $\mathbf{Q}(\sqrt[3]{a})$ and is the splitting of 3:

a) $\{1, \sqrt[3]{a}, \sqrt[3]{a^2}\}$ and $(3) = \mathfrak{p}^3$ if $a \not\equiv \pm 1 \bmod 9$.

b) $\{1, \sqrt[3]{a}, \frac{1}{3}(1 \pm \sqrt[3]{a} + \sqrt[3]{a^2})\}$ and $(3) = \mathfrak{p}^2\mathfrak{p}'$ if $a \equiv \pm 1 \bmod 9$, where the sign on $\sqrt[3]{a}$ in the third member of the basis agrees with $a \bmod 9$.

14. Not all pure cubic fields $\mathbf{Q}(\sqrt[3]{a})$ can have a squarefree, so the previous exercise does not cover the most general case. Verify the following $\mathbf{Z}$-basis and splitting of (3) in the indicated pure cubic fields.

a) $\mathbf{Q}(\sqrt[3]{20})$: $\{1, \sqrt[3]{2^2 \cdot 5}, \sqrt[3]{2 \cdot 5^2}\}$, $(3) = \mathfrak{p}^3$,

b) $\mathbf{Q}(\sqrt[3]{28})$: $\{1, \sqrt[3]{2^2 \cdot 7}, \frac{1}{3}(2 - \sqrt[3]{2^2 \cdot 7} + \sqrt[3]{2 \cdot 7^2})\}$, $(3) = \mathfrak{p}^2\mathfrak{p}'$.

A $\mathbf{Z}$-basis for the integers in a general pure cubic field $\mathbf{Q}(\sqrt[3]{m})$ and in a general biquadratic field $\mathbf{Q}(\sqrt{m}, \sqrt{n})$ can be found in [37, pp. 49–52].

15. The first n such that $\mathbf{Q}(\sqrt[n]{2})$ does not have ring of integers $\mathbf{Z}[\sqrt[n]{2}]$ is 1093. What are the next three such n?

16. Give an alternate proof that if $K = \mathbf{Q}(\alpha)$ where α is the root of an Eisenstein polynomial at p then p is totally ramified: letting $\mathfrak{p}$ be a prime in K lying over p, use the Eisenstein polynomial to show $(\alpha)^n = p(\beta)$ where $\beta \not\equiv 0 \bmod \mathfrak{p}$ and conclude that $e(\mathfrak{p}|p) = n$.

17. a) If $f(T) \in \mathbf{Z}[T]$ and some $f(T+c)$ is Eisenstein at some prime p, show there is such an Eisenstein translate where $0 \leqslant c \leqslant p^2 - 1$.

b) If $f(T)$ is monic in $\mathbf{Z}[T]$ of degree n and $f(T+c)$ is Eisenstein at p, show in the field $K = \mathbf{Q}(\alpha)$, where $f(\alpha) = 0$, that $(p) = (p, \alpha - c)^n$.

c) The table below gives some examples of polynomials with an Eisenstein translate.

$f(T)$	p	$T \mapsto T + c$	$f(T+c)$
$T^3 - T^2 - 9T + 8$	7	$T \mapsto T - 2$	$T^3 - 7T^2 + 7T + 14$
$T^3 - T^2 - 23T - 13$	2	$T \mapsto T - 1$	$T^3 - 4T^2 - 18T + 8$
$T^3 - T^2 - 48T - 63$	5	$T \mapsto T - 3$	$T^3 - 10T^2 - 15T + 45$
$T^3 - T^2 - 33T - 53$	7	$T \mapsto T - 2$	$T^3 - 7T^2 - 21T + 154$

The whole trick about finding Eisenstein translates is figuring out what c to translate by. The next table factors the discriminant of each polynomial and factors $f(T) \bmod p$ from the previous table. Compare the tables to make a prediction about how to choose c so $f(T+c)$ might be Eisenstein.

$f(T)$	$\operatorname{disc} f$	p	$f(T) \bmod p$
$T^3 - T^2 - 9T + 8$	$7^2 \cdot 53$	7	$(T+2)^3$
$T^3 - T^2 - 23T - 13$	$2^5 \cdot 5^2 \cdot 7^2$	2	$(T+1)^3$
$T^3 - T^2 - 48T - 63$	$3^3 \cdot 5^2 \cdot 419$	5	$(T+3)^3$
$T^3 - T^2 - 33T - 53$	$7^2 \cdot 757$	7	$(T+2)^3$

Put your idea to work by determining for each polynomial in the next table every prime for which the polynomial has an Eisenstein translate (there may be none).

$f(T)$	$\operatorname{disc} f$
$T^3 - 21T + 26$	$2^3 \cdot 3^4 \cdot 29$
$T^3 - T^2 - 10T + 1$	$3^2 \cdot 11 \cdot 43$
$T^3 - 2T^2 - 9T + 2$	$2^2 \cdot 31^2$
$T^3 - T^2 - 7T + 3$	$2^5 \cdot 7^2$
$T^3 - T^2 - 23T - 13$	$2^5 \cdot 5^2 \cdot 7^2$
$T^3 - T^2 - 43T + 116$	$13^2 \cdot 277$
$T^3 - T^2 - 51T + 81$	$2^2 \cdot 3^2 \cdot 5 \cdot 2393$

18. (Bhargava) Let R be any commutative ring which additively has a finite $\mathbf{Z}$-basis. Examples include any $\mathcal{O}_K$, $\mathbf{Z}^n$ as a product ring, $\mathbf{Z}[x]/(x^3 - x)$, and $\mathbf{Z}[x,y]/(x^2, xy, y^2)$. (Nonexamples are $\mathbf{Z}[1/2]$, which has no $\mathbf{Z}$-basis, and $\mathbf{Z}[x,y]/(x^2, xy)$, which has an infinite $\mathbf{Z}$-basis.) A product of any two such rings is again such a ring.

    Define the discriminant of R to be $\operatorname{disc}(R) = \det(\operatorname{Tr}_{R/\mathbf{Z}}(e_ie_j))$ for any $\mathbf{Z}$-basis $\{e_1, \ldots, e_n\}$ of R. This is independent of the basis or its ordering since the only unit square in $\mathbf{Z}$ is 1. Prove in the following steps that Stickelberger's congruence holds: $\operatorname{disc}(R) \equiv 0, 1 \bmod 4$.

    a) Letting n be the rank of R as a $\mathbf{Z}$-module, construct an n-dimensional $\mathbf{Q}$-algebra A which contains R. (Either make sense of formal products of rational numbers with R or use tensor products.)

    b) Show A has only finitely many maximal ideals. (Hint: If $\mathfrak{m}_1, \ldots, \mathfrak{m}_k$ are maximal ideals, the diagonal map $A \to \prod_{i=1}^k A/\mathfrak{m}_i$ is onto by the Chinese remainder theorem. Compare dimensions over $\mathbf{Q}$.)

    c) If $\mathfrak{m}_1, \ldots, \mathfrak{m}_t$ are *all* the maximal ideals of A, show every element of $\bigcap_{i=1}^t \mathfrak{m}_i$ is nilpotent. (Hint: Study the minimal polynomial over $\mathbf{Q}$ of any element of $\bigcap_{i=1}^t \mathfrak{m}_i$ to show it is nilpotent.)

    d) If A has a nonzero nilpotent element, show $\operatorname{disc}(R) = 0$.

    e) If the only nilpotent element of A is 0, show $A \cong \prod_{i=1}^t A/\mathfrak{m}_i$, which is a finite product of number fields.

    f) Use part e and Lemma 6.27 to compare $\operatorname{disc}(R)$ with the discriminant of a product of rings of integers of number fields, thereby deriving Stickelberger's congruence for $\operatorname{disc}(R)$ from the special case of $\operatorname{disc}(\mathcal{O}_K)$.

    g) Can you find a proof of Stickelberger's congruence for $\operatorname{disc}(R)$ which does not involve reduction to the special case of rings of integers?

19. Let E/F be a finite separable extension of fields and $\{e_1, \ldots, e_n\}$ be an F-basis of E with dual basis $\{e_1^\vee, \ldots, e_n^\vee\}$ relative to the trace pairing on E. That is, $\mathrm{Tr}_{E/F}(e_i e_j^\vee) = \delta_{ij}$. Show the trace pairing matrices $(\mathrm{Tr}_{E/F}(e_i e_j))$ and $(\mathrm{Tr}_{E/F}(e_i^\vee e_j^\vee))$ are inverses of each other. In particular,

$$\mathrm{disc}_{E/F}(e_1, \ldots, e_n)\,\mathrm{disc}_{E/F}(e_1^\vee, \ldots, e_n^\vee) = 1.$$

20. Let $K = \mathbf{Q}(\alpha)$, where $\alpha^3 - 9\alpha - 6 = 0$. The polynomial $T^3 - 9T - 6$ is Eisenstein at 3, so it is irreducible in $\mathbf{Q}[T]$.

a) Compute $\mathrm{Tr}_{K/\mathbf{Q}}(a + b\alpha + c\alpha^2)$ with $a, b, c \in \mathbf{Q}$. Use linearity of the trace to ease the computations.

b) Find the dual basis to $\{1, \alpha, \alpha^2\}$ relative to the trace pairing $K \times K \to \mathbf{Q}$.

c) Compute the ring of integers $\mathcal{O}_K$.

21. For a prime p and integer $d \geqslant 1$, let $n = p^d$ and $K = \mathbf{Q}(\sqrt[n]{p})$. Show $\mathrm{disc}(K) = \pm p^{nd+n-1}$ and pin down the correct sign depending on n.

22. For primes p and q which are not equal to each other or equal to 3, let $K = \mathbf{Q}(\sqrt[3]{p^2 q})$. Set $\alpha = \sqrt[3]{p^2 q}$ and $\beta = \sqrt[3]{pq^2}$. The lattice $A = \mathbf{Z} + \mathbf{Z}\alpha + \mathbf{Z}\beta$ is a ring.

a) Show the dual lattice of A relative to the trace pairing on K is

$$A^\vee = \mathbf{Z}\frac{1}{3} + \mathbf{Z}\frac{\beta}{3pq} + \mathbf{Z}\frac{\alpha}{3pq} = \frac{1}{3pq}\left(\mathbf{Z}pq + \mathbf{Z}\alpha + \mathbf{Z}\beta\right).$$

b) As a fractional A-ideal, show $A^\vee = \frac{1}{3pq}(\alpha, \beta)$.

23. Let K be a number field and $\mathfrak{p}$ be a prime ideal in $\mathcal{O}_K$ lying over a prime number p. We ask: when is the unique ring homomorphism $\mathbf{Z} \to \mathcal{O}_K/\mathfrak{p}^r$ surjective for all r? (Having this be true for all r is the same as having this be true for all sufficiently large r, because surjectivity for modulus $\mathfrak{p}^r$ forces surjectivity for smaller powers of $\mathfrak{p}$ since a congruence modulo $\mathfrak{p}^r$ is valid modulo any smaller power of $\mathfrak{p}$.)

a) Taking $K = \mathbf{Q}(i)$ and $\mathfrak{p} = (1 + i)$, show $\mathbf{Z} \to \mathbf{Z}[i]/\mathfrak{p}^r$ is surjective for $r = 1$ but not for any $r \geqslant 2$.

b) Back in the general case, let $e = e(\mathfrak{p}|p)$ and $f = f(\mathfrak{p}|p)$. Show $\mathfrak{p}^e \cap \mathbf{Z} = p\mathbf{Z}$ and then argue from this that if $\mathbf{Z} \to \mathcal{O}_K/\mathfrak{p}^e$ is onto then $e = 1$ and $f = 1$.

c) Conversely, if $e(\mathfrak{p}|p) = 1$ and $f(\mathfrak{p}|p) = 1$ show $\mathbf{Z} \to \mathcal{O}_K/\mathfrak{p}^r$ is surjective for all $r \geqslant 1$. (Hint: Show $\mathfrak{p}^r \cap \mathbf{Z} = p^r\mathbf{Z}$ for all r.)

d) Take $K = \mathbf{Q}(\sqrt{-5})$, and $\mathfrak{p} = (3, 1 + \sqrt{-5})$. Since $e(\mathfrak{p}|3) = 1$ and $f(\mathfrak{p}|3) = 1$, part c says there is an $a \in \mathbf{Z}$ such that $1 + 2\sqrt{-5} \equiv a \bmod \mathfrak{p}^4$. Find one. (Hint: The prime $\mathfrak{p}$ has norm 3. Think about solving $x^2 \equiv -5 \bmod 3^4$. There will be two solutions and you have to be careful to see which solution really works for you. The solution you don't use would be useful for congruences modulo powers of the conjugate ideal $\overline{\mathfrak{p}}$.)

24. Let $K = \mathbf{Q}(\alpha)$, where α is an algebraic integer whose minimal polynomial

$$T^n + c_{n-1}T^{n-1} + \cdots + c_1T + c_0$$

in $\mathbf{Z}[T]$ is Eisenstein with respect to some prime p.

a) Show $\mathbf{Z}[\alpha] \cap p^r\mathcal{O}_K = p^r\mathbf{Z}[\alpha]$ for all $r \geqslant 1$.

b) For $F = \mathbf{Q}(\sqrt{5})$, show $\mathbf{Z}[\sqrt{5}] \cap 2\mathcal{O}_F$ properly contains $2\mathbf{Z}[\sqrt{5}]$ and then compute $[\mathbf{Z}[\sqrt{5}] \cap 2\mathcal{O}_F : 2\mathbf{Z}[\sqrt{5}]]$. (This doesn't contradict part a since $\sqrt{5}$ is not a root of an Eisenstein polynomial with respect to 2.)

25. Let $K = \mathbf{Q}(\alpha)$ with α an algebraic integer and p be a prime number.

a) Show the following conditions are equivalent.

- the natural ring homomorphism $\mathbf{Z}[\alpha]/p\mathbf{Z}[\alpha] \to \mathcal{O}_K/p\mathcal{O}_K$ is an isomorphism,
- $p\mathcal{O}_K \cap \mathbf{Z}[\alpha] = p\mathbf{Z}[\alpha]$,
- $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$.

This basically means the converse of (4.2) is true. (Hint: Think about the use of Cauchy's theorem in the proof of Theorem 6.32.)

b) If $\mathbf{Z}[\alpha]/p\mathbf{Z}[\alpha]$ is a field, show $p\mathcal{O}_K$ is prime and $p \nmid [\mathcal{O}_K : \mathbf{Z}[\alpha]]$.

c) If $\mathbf{Z}[\alpha \cap p\mathcal{O}_K = p\mathbf{Z}[\alpha]$, show $\mathbf{Z}[\alpha] \cap p^r\mathcal{O}_K = p^r\mathbf{Z}[\alpha]$ for all $r \geqslant 1$. (It is not crucial in this last part that p is actually prime.)

26. Let $K = \mathbf{Q}(\sqrt{a}, \sqrt{b})$ where a and b are distinct squarefree integers other than 1. Show the quadratic subfields of K are $\mathbf{Q}(\sqrt{a})$, $\mathbf{Q}(\sqrt{b})$, and $\mathbf{Q}(\sqrt{c})$, where $c = ab/(a,b)^2$ is squarefree and not 1.

27. Let K be a cubic field where $d = |\mathrm{disc}(K)| > 1$ is squarefree. (The easiest way to find such K is with monic irreducible cubic $f(T) \in \mathbf{Z}[T]$ such that $|\mathrm{disc}\, f| = d$. Set $K = \mathbf{Q}(\alpha)$, where $f(\alpha) = 0$.)

    a) Show d is odd and $\mathrm{disc}(K) = d^*(= (-1)^{(d-1)/2}d)$.

    b) Show any prime p dividing d is ramified but not totally ramified in K, so $p\mathcal{O}_K = \mathfrak{p}^2\mathfrak{p}'$ for distinct primes $\mathfrak{p}$ and $\mathfrak{p}'$ of norm p.

    c) Show $K(\sqrt{d^*})$ is a Galois extension of $\mathbf{Q}$ in which every prime number not dividing d is unramified.

28. For an integer a and $n \geqslant 2$, assume $T^n - a$ is irreducible. Write $\sqrt[n]{a}$ for any root of it. If $\mathbf{Q}(\sqrt[n]{a})$ has ring of integers $\mathbf{Z}[\sqrt[n]{a}]$, show

    a) a is squarefree (Hint: if $\alpha = \sqrt[3]{p^2q}$ then $(\alpha^2/p)^3 = pq^2$),

    b) for every positive divisor d of n, $T^d - a$ is irreducible and $\mathbf{Q}(\sqrt[d]{a})$ has ring of integers $\mathbf{Z}[\sqrt[d]{a}]$.

29. Let p be a prime and a be a nonzero integer other than ± 1 which is not divisible by any pth power.

    a) Show $T^p - a$ is irreducible in $\mathbf{Q}[T]$.

    b) Show any prime dividing a is totally ramified in $\mathbf{Q}(\sqrt[p]{a})$.

    c) If $(p, a) = 1$, show $\mathrm{disc}(\mathbf{Q}(\sqrt[p]{a}))$ is divisible by p, so p is ramified.

    d) If $(p, a) = 1$ and $a^{p-1} \equiv 1 \bmod p^2$, show $(p) = \mathfrak{p}^{p-1}\mathfrak{p}'$ where $\mathrm{N}(\mathfrak{p}) = \mathrm{N}(\mathfrak{p}') = p$.

30. Let $K = \mathbf{Q}(\sqrt[4]{17})$. The discriminant of $T^4 - 17$ is $-2^8 \cdot 17^3$.

    a) Use Theorem 3.28, Stickelberger's theorem, and Theorem 6.45 to show $\mathrm{disc}(K) = -2^a \cdot 17^3$, where $2 \leqslant a \leqslant 8$. (In particular, 2 must be ramified.)

    b) Show $\{1, \gamma, \gamma', \gamma''\}$ is a $\mathbf{Z}$-basis of $\mathcal{O}_K$, where $\gamma = \sqrt[4]{17}$, $\gamma' = \frac{1}{2}(1+\gamma^2) = \frac{1}{2}(1+\sqrt{17})$, and $\gamma'' = \frac{1}{4}(1+\gamma)(1+\gamma^2)$.

    c) Show $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ is even for all α, so $\mathcal{O}_K$ does not have a power basis. (Hint: Use the method of Theorem 4.49.)

d) In $\mathcal{O}_K$, show $(2) = \mathfrak{p}_2^2\mathfrak{p}_2'\mathfrak{p}_2''$. (Hint: Since 2 splits completely in $\mathbf{Q}(\sqrt{17})$, there are at least two primes over 2 in K. And γ'' has minimal polynomial $T^4 - T^3 - 6T^2 - 16T - 16$. Look at the minimal polynomial of $\gamma'' - 1$.)

31. Let $K = \mathbf{Q}(\alpha)$ where $\alpha^4 + 8\alpha + 12 = 0$.

a) Two factorizations of $T^4 + 8T + 12 \bmod p$ are

$$\begin{aligned} T^4 + 8T + 12 &\equiv (T-4)(T^3 + 4T + T + 2) \bmod 5, \\ T^4 + 8T + 12 &\equiv (T^2 + 4T + 7)(T^2 + 13T + 9) \bmod 17. \end{aligned}$$

Explain from this why $T^4 + 8T + 12$ is irreducible in $\mathbf{Q}[T]$.

b) Factor the ideal $(\alpha + c)$ into primes for $c = 0, \pm 1, \pm 2$.

c) Use the ideal equations $(12) = (\alpha^4 + 8\alpha)$ and $(\alpha^4) = (8\alpha + 12)$ to show $(3) = \mathfrak{p}_3\mathfrak{p}_3'^3$.

d) Show $\alpha^2/2 \in \mathcal{O}_K$ and $\mathrm{N}_{K/\mathbf{Q}}(\alpha^2/2 - 1) = 8$.

e) The ratio $2/(\alpha^2/2 - 1)$ has norm 2. Show it is an algebraic integer and $(2) = \mathfrak{p}_2^4$.

32. Let $K = \mathbf{Q}(\sqrt[8]{3})$ and $L = \mathbf{Q}(\sqrt[8]{48})$.

a) Factor the primes 5, 7, 11, and 13 in K and L.

b) Show $\mathrm{disc}(K) = \mathrm{disc}(L)$. (Hint: Use a computer to find the discriminant of the polynomial in (6.10).)

c) Show K and L both lie in $\mathbf{Q}(\sqrt[8]{3}, \zeta_8)$ and they correspond to nonconjugate subgroups of $\mathrm{Gal}(\mathbf{Q}(\sqrt[8]{3}, \zeta_8)/\mathbf{Q})$, so K and L are nonisomorphic fields.

33. Set $\alpha = \zeta_7^2 + \zeta_7^4 + \zeta_7^5 + \zeta_7^6$, where $\zeta_7 = e^{2\pi i/7}$.

a) Show $|\alpha|^2 = \alpha\overline{\alpha} = 2$, so $|\alpha| = \sqrt{2}$.

b) Show $\alpha \neq \sqrt{2}\zeta$ for any root of unity ζ in $\mathbf{C}$. (Hint: 2 does not ramify in $\mathbf{Q}(\zeta_7)$.)

34. Set the mth *cyclotomic polynomial* to be

$$\Phi_m(T) := \prod_{i \in (\mathbf{Z}/m\mathbf{Z})^\times} (T - \zeta_m^i).$$

Its degree is $\varphi(m)$, hence the notation Φ for the polynomial. Since $[\mathbf{Q}(\zeta_m):\mathbf{Q}] = \varphi(m)$, $\Phi_m(T)$ is the minimal polynomial of ζ_m over $\mathbf{Q}$, so $\Phi_m(T) \in \mathbf{Z}[T]$. Some examples are

$$\Phi_1(T) = T-1,\ \Phi_2(T) = T+1,\ \Phi_3(T) = T^2+T+1,\ \Phi_4(T) = T^2+1,$$

$$\Phi_5(T) = T^4+T^3+T^2+T+1,\ \Phi_6(T) = T^2-T+1,\ \Phi_{12}(T) = T^4-T^2+1.$$

a) Show $\Phi_{p_1^{e_1}\cdots p_k^{e_k}}(T) = \Phi_{p_1\cdots p_k}(T^{p_1^{e_1-1}\cdots p_k^{e_k-1}})$ for distinct primes p_i with $e_i \geqslant 1$.

b) Show $\Phi_{2m}(T) = \Phi_m(-T)$ for odd m.

c) When $m = p^r m'$, show $\Phi_m(T) \equiv \Phi_{m'}(T)^{\varphi(p^r)} \bmod p$.

d) If $p \nmid m$, let f be the (multiplicative) order of $p \bmod m'$. Show

$$\Phi_m(T) \equiv \pi_1(T)\pi_2(T)\cdots\pi_g(T) \bmod p,$$

where the π_i's are distinct monic irreducibles in $\mathbf{F}_p[T]$, $\deg \pi_i = f$ for all i, and $g = \varphi(m)/f$.

e) If $(\mathbf{Z}/m\mathbf{Z})^\times$ is not cyclic, show $\Phi_m(T) \bmod p$ is reducible for *all* p. The smallest example is $m = 8$: $\Phi_8(T) = T^4+1$.

35. a) For a prime power p^r, show $\mathrm{N}_{\mathbf{Q}(\zeta_{p^r})/\mathbf{Q}}(\zeta_{p^r}-1) = p$.

    b) As a strong contrast to part a, if $m > 1$ is not a prime power show $\mathrm{N}_{\mathbf{Q}(\zeta_m)/\mathbf{Q}}(\zeta_m - 1) = 1$, so $\zeta_m - 1$ is a unit in $\mathbf{Z}[\zeta_m]$ when m is not a prime power.

36. Let $\mathbf{F}$ be a finite field and $K/\mathbf{F}(X)$ be a finite extension of degree n with R the integral closure of $\mathbf{F}[X]$ in K. By Corollary 4.61, R is a Dedekind domain (whether or not $K/\mathbf{F}(X)$ is separable).

    For any irreducible π in $\mathbf{F}[X]$, factor it into prime ideals in R as

    $$\pi R = \mathfrak{p}_1^{e_1}\cdots\mathfrak{p}_g^{e_g}$$

    and set $\#(R/\mathfrak{p}_i) = q_\pi^{f_i}$, where $q_\pi = \#(\mathbf{F}[X]/\pi)$.

    Call π ramified in R if some $e_i > 1$ and totally ramified in R if $\pi R = \mathfrak{p}^n$.

    a) Show $\sum_{i=1}^g e_i f_i = n$. (Hint: Use one of the norm functions on ideals in R from Exercise 4.31.)

b) Show π is ramified in R if and only if $\pi \mid \operatorname{disc}_{\mathbf{F}[X]}(R)$.

c) If $K = \mathbf{F}(X, \alpha)$ where $\alpha \in R$ has a minimal polynomial over $\mathbf{F}[X]$ that is Eisenstein at π, show $\pi \nmid [R : \mathbf{F}[X][\alpha]]_{\mathbf{F}[X]}$ and π is totally ramified in R. (Here $[R : \mathbf{F}[X][\alpha]]_{\mathbf{F}[X]} = \operatorname{card}_{\mathbf{F}[X]}(R/\mathbf{F}[X][\alpha])$.) The analogue of Cauchy's theorem in Exercise 3.19a will be useful here.

d) If π is totally ramified in R, show $K = \mathbf{F}(X, \alpha)$ for some $\alpha \in R$ whose minimal polynomial over $\mathbf{F}[X]$ is Eisenstein at π. (Hint: Use one of the norm functions on ideals in R from Exercise 4.31.)

e) (Continuation of Exercise 3.16) Let $\mathbf{F}$ be a finite field of characteristic 2, and for an odd positive integer d let $\mathcal{O}_d$ be the integral closure of $\mathbf{F}[X]$ in $\mathbf{F}(X, \alpha_d)$ where $\alpha_d^2 - \alpha_d - 1/X^d = 0$. Show X is totally ramified in $\mathcal{O}_d$ and conclude that $\mathcal{O}_d = \mathbf{F}[X] + \mathbf{F}[X]\beta_d$, where $\beta_d := X^{(d+1)/2}\alpha_d$.

37. Let k be a field of characteristic p and $L = k(X^{1/p})$. Assume k is not perfect, meaning that k has elements which are not pth powers in k. (An example is $k = \mathbf{F}_p(u)$.)

a) Show the integral closure of $k[X]$ in L is $k[X^{1/p}]$. (This is a PID, being isomorphic to $k[Y]$.)

b) Compute $\operatorname{disc}_{k[X]}(k[X^{1/p}])$.

c) For $c \in k$, show $X - c$ is totally ramified in $k[X^{1/p}]$ if c is a pth power in k, but if c is not a pth power in k then $X - c$ stays prime (so is unramified?) in $k[X^{1/p}]$.

d) What do parts b and c suggest about the prospect of applying Dedekind's discriminant theorem to an integral closure of a general PID?

CHAPTER 7

NUMBER FIELDS IN EUCLIDEAN SPACE

7.1 The Euclidean Embedding

When we proved class groups are finite, we used field embeddings of K into $\mathbf{R}$ and $\mathbf{C}$. We will now make a more sustained use of these embeddings, in combination with a basic theorem about lattices and volumes in Euclidean space. The benefits of this development will include:

1. a better constant for computing class groups than the Kronecker bound C which we found in Section 5.1,
2. Minkowski's discriminant theorem: when $[K:\mathbf{Q}] > 1$, $\operatorname{disc}(K) \neq \pm 1$,
3. Dirichlet's unit theorem: $\mathcal{O}_K^\times$ is finitely generated (with a formula for the number of generators).

Our arguments will make *no* use of prime ideal factorization.

The standard pictures of $\mathbf{Z}$ and $\mathbf{Z}[i]$ are as discrete subsets of $\mathbf{R}$ and $\mathbf{C}$. See Figures 7.1 and 7.2. In contrast, $\mathbf{Z}[\sqrt{2}]$ is dense in $\mathbf{R}$: since $0 < \sqrt{2} - 1 < 1$, $(\sqrt{2}-1)^n \to 0$ as $n \to \infty$ so the integral multiples of all the powers $(\sqrt{2}-1)^n$ are a dense subset of $\mathbf{R}$. See Figure 7.3.

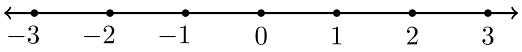


Figure 7.1: $\mathbf{Z}$ lying discretely in $\mathbf{R}$.

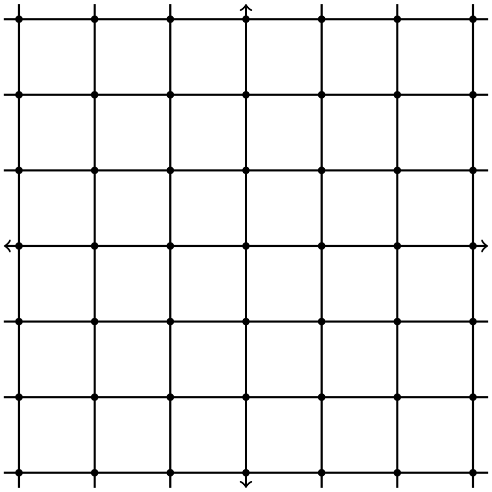


Figure 7.2: $\mathbf{Z}[i]$ lying discretely in $\mathbf{C}$.

We can make $\mathbf{Z}[\sqrt{2}]$ look discrete by using *both* real embeddings of the abstract field $\mathbf{Q}(\gamma)$, where $\gamma^2 = 2$. It has two embeddings into $\mathbf{R}$: $\sigma_1(a + b\gamma) = a + b\sqrt{2}$ and $\sigma_2(a + b\gamma) = a - b\sqrt{2}$. Define $\theta \colon \mathbf{Q}(\gamma) \to \mathbf{R}^2$ to have component functions σ_1 and σ_2:

$$\begin{aligned}\theta(a + b\gamma) &= (\sigma_1(a + b\gamma), \sigma_2(a + b\gamma))\\ &= (a + b\sqrt{2}, a - b\sqrt{2})\\ &= a(1, 1) + b(\sqrt{2}, -\sqrt{2}).\end{aligned}$$

This function θ is injective, since it already is in each component. The vectors $(1, 1)$ and $(\sqrt{2}, -\sqrt{2})$ are linearly independent over $\mathbf{R}$, so they are a basis of $\mathbf{R}^2$. Then $\theta(\mathbf{Z}[\gamma])$, which corresponds to letting a and b run over $\mathbf{Z}$, is a discrete subset of $\mathbf{R}^2$. See Figure 7.4. This provides a picture of $\mathbf{Z}[\sqrt{2}]$ as a discrete set in the plane. (Projection of $\theta(\mathbf{Z}[\gamma])$ to the x-axis has image the previous dense copy of $\mathbf{Z}[\sqrt{2}]$ in $\mathbf{R}$. This is a good example where a continuous function

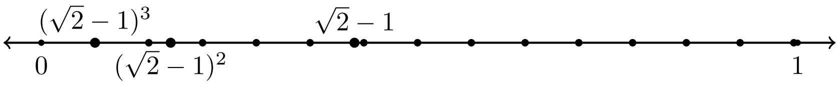


Figure 7.3: $\sqrt{2}-1, (\sqrt{2}-1)^2, (\sqrt{2}-1)^3$, and integral multiples of $(\sqrt{2}-1)^3$.

$\mathbf{R}^2 \to \mathbf{R}$ sends a discrete subset to a dense subset.)

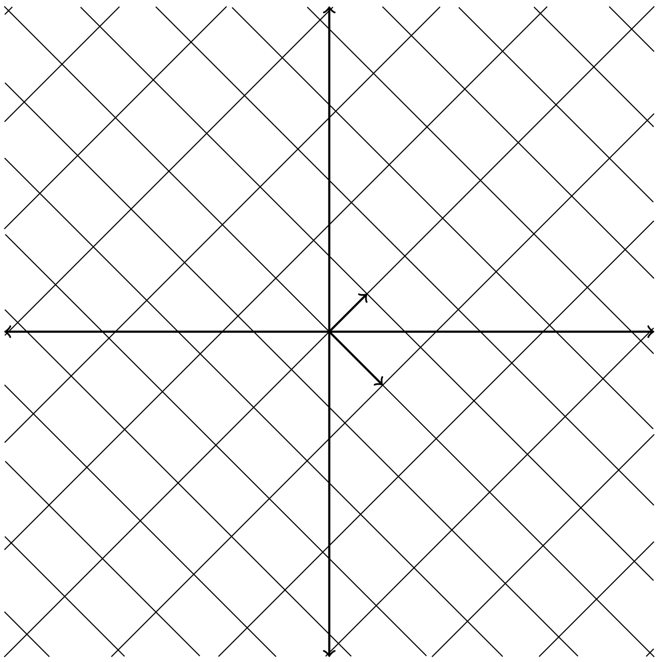


Figure 7.4: The lattice $\theta(\mathbf{Z}[\gamma]) = \mathbf{Z}(1,1) + \mathbf{Z}(\sqrt{2}, -\sqrt{2})$ in $\mathbf{R}^2$.

Turning to the general setting, consider a number field K of degree n. Write $K = \mathbf{Q}(\gamma)$, so any embedding of K into $\mathbf{C}$ is determined by where γ goes. The minimal polynomial of γ over $\mathbf{Q}$ is irreducible of degree n and therefore has n distinct roots in $\mathbf{C}$ (irreducibles in characteristic 0 are separable), so there are n embeddings $\sigma\colon K \to \mathbf{C}$.

We call $\sigma\colon K \to \mathbf{C}$ a *real embedding* if $\sigma(K) \subset \mathbf{R}$. Otherwise, we call σ a *complex embedding* and $\overline{\sigma}(x) = \overline{\sigma(x)}$ is another complex embedding since $\sigma(K)$ has some non-real values. The complex embeddings all come in conjugate pairs

$\{\sigma, \overline{\sigma}\}$. Set r_1 to be the number of real embeddings of K and $2r_2$ to be the number of complex embeddings of K. That means r_2 is the number of *pairs* of complex conjugate embeddings. If we write $K = \mathbf{Q}(\gamma)$, r_1 is the number of real roots of the minimal polynomial of γ over $\mathbf{Q}$ and $2r_2$ is the number of non-real complex roots of that polynomial. In Table 7.1 are examples of r_1 and r_2. Check these values, especially r_2 in the cubic examples.

K	n	r_1	r_2
$\mathbf{Q}$	1	1	0
$\mathbf{Q}(i)$	2	0	1
$\mathbf{Q}(\sqrt{2})$	2	2	0
$\mathbf{Q}(\sqrt[3]{2})$	3	1	1
$\mathbf{Q}(\alpha), \alpha^3 - 9\alpha - 9 = 0$	3	3	0
$\mathbf{Q}(\zeta_m),\ m > 2$	$\varphi(m)$	0	$\varphi(m)/2$

Table 7.1: Examples of r_1 and r_2.

Counting up all the embeddings, we get

$$r_1 + 2r_2 = n.$$

This is analogous to the identity $e_1f_1 + \cdots + e_gf_g = n$ associated to each prime number. We call K *totally real* if $r_2 = 0$ and *totally complex* if $r_1 = 0$. For example, $\mathbf{Q}$ and $\mathbf{Q}(\sqrt{2})$ are totally real and $\mathbf{Q}(\zeta_m)$ for $m > 2$ is totally complex (when $m = 4$ this field is $\mathbf{Q}(i)$).

Definition 7.1. Label the real and complex embeddings of K as

$$\sigma_1, \ldots, \sigma_{r_1}, \sigma_{r_1+1}, \overline{\sigma}_{r_1+1}, \ldots, \sigma_{r_1+r_2}, \overline{\sigma}_{r_1+r_2}.$$

The *Euclidean embedding* of K is $\theta_K \colon K \to \mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$, where

$$\theta_K(\alpha) = (\sigma_1(\alpha), \ldots, \sigma_{r_1}(\alpha), \sigma_{r_1+1}(\alpha), \ldots, \sigma_{r_1+r_2}(\alpha)).$$

Since each coordinate of θ_K is injective, θ_K is injective. The first r_1 coordinates of $\theta_K(\alpha)$ are real numbers and the last r_2 coordinates are complex numbers. The definition of θ_K uses only half the complex embeddings, one from each pair of complex conjugate embeddings. Since θ_K depends on the choice of one from each pair of complex conjugate embeddings and on the indexing of all embeddings, it is a non-canonical function, although θ_K is called the "canonical embedding" in [48, p. 56].

Let's look at formulas for the Euclidean embeddings of two number fields.

Example 7.2. Let $K = \mathbf{Q}(\sqrt{2})$ with $\sigma_1(\sqrt{2}) = \sqrt{2}$ and $\sigma_2(\sqrt{2}) = -\sqrt{2}$. Then $\theta_K \colon K \to \mathbf{R}^2$ by

$$\theta_K(a + b\sqrt{2}) = (a + b\sqrt{2}, a - b\sqrt{2}) = a(1,1) + b(\sqrt{2}, \sqrt{2}).$$

This is the mapping illustrated in Figure 7.4, which embeds $\mathbf{Q}(\sqrt{2})$ into $\mathbf{R}^2$ and $\mathbf{Z}[\sqrt{2}]$ as a discrete subset of $\mathbf{R}^2$.

Example 7.3. Let $K = \mathbf{Q}(\sqrt[3]{2})$ with $\sigma_1(\sqrt[3]{2}) = \sqrt[3]{2}$, $\sigma_2(\sqrt[3]{2}) = \omega\sqrt[3]{2}$, and $\overline{\sigma}_2(\sqrt[3]{2}) = \overline{\omega}\sqrt[3]{2} = \omega^2\sqrt[3]{2}$, where $\omega = e^{2\pi i/3}$. Then $\theta_K \colon K \to \mathbf{R} \times \mathbf{C}$ by

$$\begin{aligned}\theta_K(a + b\sqrt[3]{2} + c\sqrt[3]{4}) &= (a + b\sqrt[3]{2} + c\sqrt[3]{4}, a + b\omega\sqrt[3]{2} + c\omega^2\sqrt[3]{4}) \\ &= a(1,1) + b(\sqrt[3]{2}, \omega\sqrt[3]{2}) + c(\sqrt[3]{4}, \omega^2\sqrt[3]{4}).\end{aligned}$$

The first component of $\theta_K(\alpha) = (\sigma_1(\alpha), \sigma_2(\alpha))$ is a real number and the second component is a complex number. By replacing σ_2 with $\overline{\sigma}_2$, we get a second Euclidean embedding of K: $\alpha \mapsto (\sigma_1(\alpha), \overline{\sigma}_2(\alpha))$.

We want to show the Euclidean image of $\mathcal{O}_K$ is a lattice in the Euclidean space $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$, as we saw with $\mathbf{Z}[\sqrt{2}]$ inside $\mathbf{R}^2$. For this purpose, we use the formula

$$\mathrm{Tr}_{K/\mathbf{Q}}(x) = \sum_{\sigma} \sigma(x),$$

where the sum runs over every real and complex embedding σ of K (Theorem 8.18). The analogous norm formula is

$$\mathrm{N}_{K/\mathbf{Q}}(x) = \prod_{\sigma} \sigma(x)$$

and we used that already in (5.1).

Theorem 7.4. *If $\alpha_1, \ldots, \alpha_n$ is a basis for $K/\mathbf{Q}$, then $\theta_K(\alpha_1), \ldots, \theta_K(\alpha_n)$ is a basis of $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ as an $\mathbf{R}$-vector space.*

Example 7.5. We saw this before for the Euclidean image of the $\mathbf{Q}$-basis $\{1, \sqrt{2}\}$ of $\mathbf{Q}(\sqrt{2})$.

Proof. It suffices to prove this for just one $\mathbf{Q}$-basis of K, because if some $\mathbf{Q}$-basis of K has a Euclidean image that is linearly dependent over $\mathbf{R}$ then $\theta_K(K)$ lies in a hyperplane of $\mathbf{R}^n$, so the Euclidean image of any $\mathbf{Q}$-basis of K is not a basis of $\mathbf{R}^n$.

We will use a power basis $\{1, \alpha, \ldots, \alpha^{n-1}\}$. Since $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ has dimension n, being a basis is the same as being linearly independent, so we just check linear independence.

Putting the vectors $\theta_K(1)$, $\theta_K(\alpha), \ldots, \theta_K(\alpha^{n-1})$ into the columns of a matrix,

$$\begin{pmatrix} | & | & & | \\ \theta_K(1) & \theta_K(\alpha) & \cdots & \theta_K(\alpha^{n-1}) \\ | & | & & | \end{pmatrix} = \begin{pmatrix} 1 & \sigma_1(\alpha) & \cdots & \sigma_1(\alpha)^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \sigma_{r_1+r_2}(\alpha) & \cdots & \sigma_{r_1+r_2}(\alpha)^{n-1} \end{pmatrix}.$$

The right side is a Vandermonde matrix, with determinant $\prod_{i<j}(\sigma_j(\alpha)-\sigma_i(\alpha))$. Since α generates K over $\mathbf{Q}$, $\sigma_j(\alpha) \neq \sigma_i(\alpha)$ when $\sigma_j \neq \sigma_i$. The determinant is not 0, so the columns are linearly independent over $\mathbf{R}$. ■

Letting $\alpha_1, \ldots, \alpha_n$ be a $\mathbf{Z}$-basis of $\mathcal{O}_K$, $\theta_K(\mathcal{O}_K)$ is a lattice since it is the $\mathbf{Z}$-span of $\theta_K(\alpha_1), \ldots, \theta_K(\alpha_n)$, which is a basis of $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$. Games based on the lattice representation of $\mathbf{Z}[\zeta_5]$ in $\mathbf{C}^2$ and $\mathbf{Z}[\sqrt{2}]$ in $\mathbf{R}^2$ are at

http://www.math.brown.edu/~res/Java/App12x/test1.html

and

http://www.math.brown.edu/~res/Java/App32/test1.html.

For an article about these games, see [50].

The component functions of $\theta_K(\alpha) = (\ldots, \sigma_i(\alpha), \ldots)$ are field homomorphisms, so θ_K is an injective ring homomorphism, where we treat $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ as a ring with componentwise operations. Note $\mathbf{C} \cong \mathbf{R}^2$ as $\mathbf{R}$-vector spaces but not as rings: the ring operations in $\mathbf{C}$ are not componentwise (and the ring $\mathbf{R}^2$ is not a field). So $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2} \cong \mathbf{R}^n$ as $\mathbf{R}$-vector spaces but not as rings, unless K is totally real ($r_2 = 0$).

The Euclidean embedding $K \to \mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ provides us with a new proof of the formula for the sign of $\operatorname{disc}(K)$ in Theorem 3.28. This sign is $(-1)^{r_2}$. (It was written in Theorem 3.28 as $(-1)^P$, where $P = r_2$.) The ring $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ is finite-dimensional over $\mathbf{R}$, so we can talk about the $\mathbf{R}$-discriminant of a basis in the sense of (3.18). The Euclidean embedding of K into $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ takes a $\mathbf{Q}$-basis to an $\mathbf{R}$-basis, and using such a basis shows us that $[m_{\theta_K(\alpha)}] = [m_\alpha]$ for any $\alpha \in K$, so $\mathrm{Tr}_{K/\mathbf{Q}}(\alpha) = \mathrm{Tr}_{(\mathbf{R}^{r_1}\times\mathbf{C}^{r_2})/\mathbf{R}}(\theta_K(\alpha))$. Therefore

$$\mathrm{disc}_{K/\mathbf{Q}}(\alpha_1, \ldots, \alpha_n) = \mathrm{disc}_{(\mathbf{R}^{r_1}\times\mathbf{C}^{r_2})/\mathbf{R}}(\theta_K(\alpha_1), \ldots, \theta_K(\alpha_n)).$$

Changing a basis changes the discriminant by a square (in $\mathbf{Q}$ on the left and in $\mathbf{R}$ on the right), so it does not change the sign. Therefore we can compute the sign of $\mathrm{disc}(K)$ by computing the sign of the discriminant of any $\mathbf{R}$-basis of $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$. Using the basis with 1 in each $\mathbf{R}$-factor and $\{1, i\}$ in each $\mathbf{C}$-factor, by Lemma 6.27, its discriminant is

$$\mathrm{disc}_{\mathbf{R}/\mathbf{R}}(1)^{r_1}\,\mathrm{disc}_{\mathbf{C}/\mathbf{R}}(1, i)^{r_2} = (-4)^{r_2},$$

whose sign is $(-1)^{r_2}$.

7.2 Minkowski's Convex Body Theorem

7.2.1 Statement and Proof

Let $\Lambda \subset \mathbf{R}^n$ be a lattice. For any $\mathbf{Z}$-basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of Λ, we call

$$\left\{\sum_{i=1}^{n} c_i\mathbf{v}_i : 0 \leqslant c_i \leqslant 1\right\}$$

a *fundamental box* for Λ. It depends on the choice of basis.

Example 7.6. Let $\Lambda = \mathbf{Z}^2$. The basis $\{(1,0),(0,1)\}$ has as a fundamental box the shaded square in Figure 7.5. The basis $\{(1,2),(0,1)\}$ has as a fundamental box the shaded parallelogram in Figure 7.6. The lattice $\mathbf{Z}^2$ is the same in both figures, even though the lines drawn are different.

For any fundamental box B of a lattice Λ in $\mathbf{R}^n$,

$$\mathbf{R}^n = \bigcup_{\mathbf{v}\in\Lambda}(B + \mathbf{v}).$$

This is illustrated in Figure 7.7, where we have shaded in a typical parallelogram from Figure 7.4 and see it is the translate of a fundamental box for that lattice

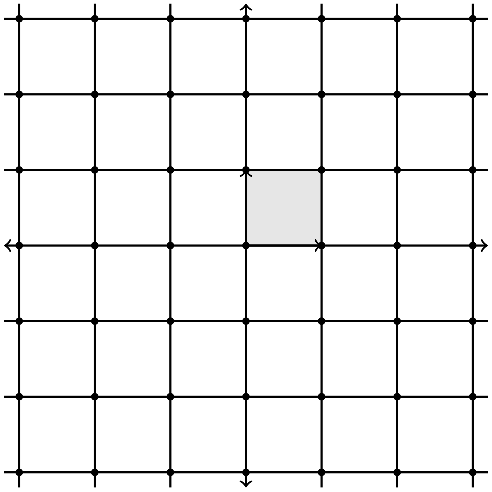


Figure 7.5: $\mathbf{Z}^2$ with basis $\{(1,0),(0,1)\}$.

by a vector in the lattice.

While a fundamental box is not intrinsic to a lattice in $\mathbf{R}^n$, its volume is intrinsic.

Theorem 7.7. *All fundamental boxes of a lattice $\Lambda \subset \mathbf{R}^n$ have the same volume.*

Of course for $n = 1$ and $n = 2$, volume means length and area. In Figures 7.5 and 7.6, the fundamental boxes for $\mathbf{Z}^2$ both have area 1.

Proof. Pick two $\mathbf{Z}$-bases of Λ, say $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$. Their associated boxes are

$$B = \left\{ \sum_{i=1}^{n} c_i \mathbf{v}_i : 0 \leqslant c_i \leqslant 1 \right\} \qquad \text{and} \qquad B' = \left\{ \sum_{i=1}^{n} c_i \mathbf{v}'_i : 0 \leqslant c_i \leqslant 1 \right\}.$$

We can write

$$\mathbf{v}'_j = \sum_{i=1}^{n} a_{ij} \mathbf{v}_i, \qquad a_{ij} \in \mathbf{Z} \tag{7.1}$$

since the $\mathbf{v}_i$'s are a $\mathbf{Z}$-basis of Λ. Since the $\mathbf{v}'_j$'s are also a $\mathbf{Z}$-basis of Λ, (a_{ij}) is

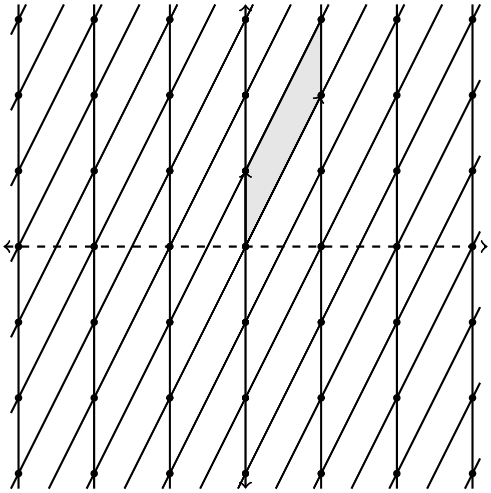


Figure 7.6: $\mathbf{Z}^2$ with basis $\{(1,2),(0,1)\}$.

in $\mathrm{GL}_n(\mathbf{Z})$, so $\det(a_{ij}) = \pm 1$. The volume of B is

$$\left|\det\begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix}\right|$$

and there is a similar formula for the volume of B'. Since

$$\begin{pmatrix} | & & | \\ \mathbf{v}'_1 & \cdots & \mathbf{v}'_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix}(a_{ij})$$

(check the indexing in (7.1) makes this correct!) taking determinants of both sides and then absolute values gives us $\mathrm{vol}(B') = \mathrm{vol}(B)$. ■

We call the common volume of all fundamental boxes for a lattice Λ the *volume of* Λ. (It would be better to call this the *covolume* of Λ since it is really the natural[1] volume of $\mathbf{R}^n/\Lambda$.)

[1] There is a unique Haar measure $\mathrm{d}\overline{\mathbf{x}}$ on $\mathbf{R}^n/\Lambda$ making the integration formula

$$\int_{\mathbf{R}^n} f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{\mathbf{R}^n/\Lambda}\Big(\sum_{\mathbf{v}\in\Lambda} f(\mathbf{x}+\mathbf{v})\Big)\,\mathrm{d}\overline{\mathbf{x}}$$

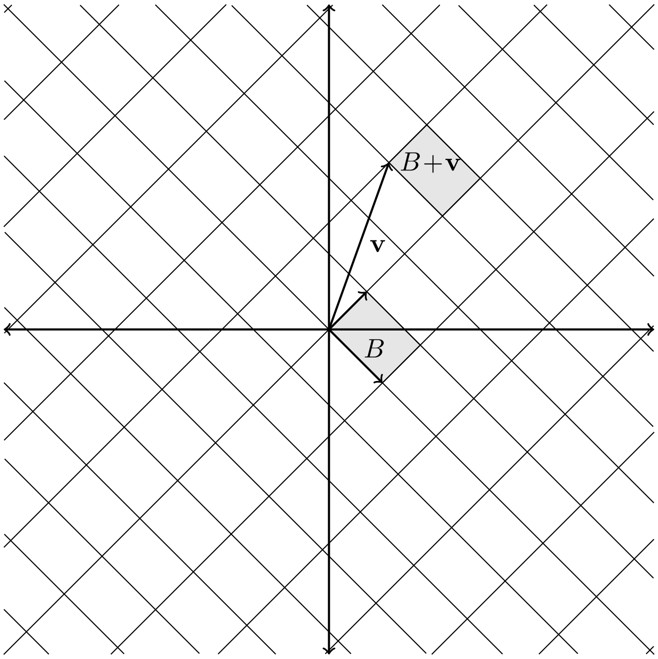


Figure 7.7: A fundamental box for $\mathbf{Z}(1,1)+\mathbf{Z}(\sqrt{2},-\sqrt{2})$ and one translate of it.

Example 7.8. In $\mathbf{R}^2$, $\mathrm{vol}(\mathbf{Z}^2)=1$ and the image of $\mathbf{Z}[\sqrt{2}]$ under the Euclidean embedding has volume

$$\mathrm{vol}(\mathbf{Z}(1,1)+\mathbf{Z}(\sqrt{2},-\sqrt{2}))=\left|\det\begin{pmatrix}1 & \sqrt{2}\\ 1 & -\sqrt{2}\end{pmatrix}\right|=2\sqrt{2}.$$

So the boxes in Figure 7.4 all have area $2\sqrt{2}$.

The connection we will exploit between lattices and volumes is through the following theorem.

Theorem 7.9 (Minkowski). *Let $\Lambda\subset\mathbf{R}^n$ be a lattice and $X\subset\mathbf{R}^n$ be a bounded subset which is convex*[2] *and centrally symmetric.*[3] *If $\mathrm{vol}(X)>2^n\,\mathrm{vol}(\Lambda)$, then X meets Λ in a nonzero vector. If X is compact and $\mathrm{vol}(X)\geqslant 2^n\,\mathrm{vol}(\Lambda)$,*

valid for all continuous $f\colon\mathbf{R}^n\to\mathbf{R}$ with compact support. Letting f be the characteristic function of a fundamental box for Λ, we get that the volume of the box equals $\int_{\mathbf{R}^n/\Lambda}\mathrm{d}\overline{\mathbf{x}}$.

[2] *convex*: if $\mathbf{v}$ and $\mathbf{w}$ are in X then so is $t\mathbf{v}+(1-t)\mathbf{w}$ for any $t\in[0,1]$.

[3] *centrally symmetric*: if $\mathbf{v}\in X$, then $-\mathbf{v}\in X$.

then X meets Λ in a nonzero vector.

Before proving Minkowski's theorem, we illustrate it in Figure 7.8. The lattice pictured there is $\Lambda = \mathbf{Z}(1,-1) + \mathbf{Z}(1,2)$, whose shaded fundamental box has area 3. (Points with integral coordinates on the axes are marked for ease of reference, but they don't necessarily belong to Λ.) The solid curve is the ellipse $(x/3)^2 + (y/1.3)^2 = 1$, whose area is $\pi \cdot 3 \cdot 1.3 \approx 12.25$. Taking for X the *interior* of the ellipse, since $12.25 > 4 \cdot 3$ there has to be a nonzero point from Λ inside X, and the points $P = (1,-1)$ and $-P = (-1,1)$ work. (The lattice point $Q = (2,1) = (1,-1) + (1,2)$ is just barely outside the ellipse and $(3,0) = 2(1,-1) + (1,2)$ is on the boundary.) The dotted curves in Figure 7.8 are rotations of the solid ellipse around the origin by 35 degrees and 65 degrees. Minkowski's theorem says there is a nonzero point from Λ inside each of them, and we can check this explicitly: the 35-degree rotation contains Q (it barely misses containing P and $R = (1,2)$) and the 65-degree rotation contains R (but not P or Q).

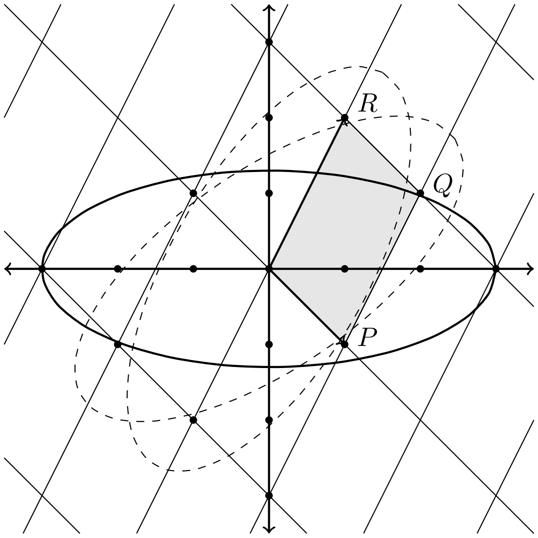


Figure 7.8: Some convex, centrally symmetric regions in $\mathbf{R}^2$ and a lattice.

Remark 7.10. A fussy reader may ask what we mean by volume in $\mathbf{R}^n$. All the convex regions we will work with will be described by rather simple inequalities and their volumes may safely be computed by multiple integrals. For those who have studied measure theory and are sensitive to the issue of nonmeasurable subsets, Minkowski proved that any bounded convex subset of $\mathbf{R}^n$ is measurable with finite volume.

The factor 2^n in Minkowski's theorem is crucial. With any smaller constant there, the theorem breaks down.

Nonexample 7.11. Let $X = (-1, 1)^2$ be the open square in $\mathbf{R}^2$ with vertices $(\pm 1, \pm 1)$, as in Figure 7.9. Then $\mathrm{vol}(X) = 4$ and $X \cap \mathbf{Z}^2 = \{(0, 0)\}$. This construction generalizes to any $\mathbf{R}^n$, using for X the open n-dimensional box $(-1, 1)^n$ with vertices at the 2^n vectors $(\pm 1, \ldots, \pm 1)$, which meets $\mathbf{Z}^n$ only in $(0, \ldots, 0)$. The closure of X has volume 2^n and meets $\mathbf{Z}^n$ in many nonzero vectors (such as its vertices). Minkowski's theorem is valid with $>$ replaced by $\geqslant$ if we add the condition that X is *compact*. See [48, p. 55].

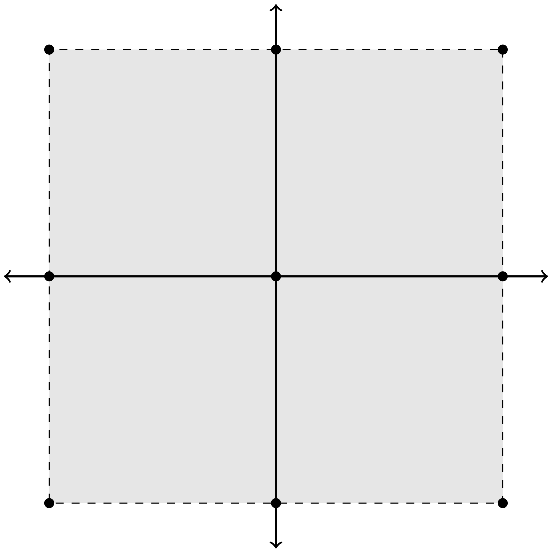


Figure 7.9: Open square with area 4 meeting $\mathbf{Z}^2$ in $(0, 0)$.

Proof of the convex body theorem. Let $\Lambda' = 2\Lambda$, so $\mathrm{vol}(\Lambda') = 2^n \,\mathrm{vol}(\Lambda)$. Let B be a fundamental box for Λ, so $2B$ is a fundamental box for 2Λ. Since

$$\mathbf{R}^n = \bigcup_{\mathbf{v} \in \Lambda'} (2B + \mathbf{v}),$$

we can translate any vector in $\mathbf{R}^n$ by a member of Λ' to a point in $2B$.

Assume $\operatorname{vol}(X) > 2^n \operatorname{vol}(\Lambda) = \operatorname{vol}(\Lambda') = \operatorname{vol}(2B)$. When we translate the points in X by Λ' to lie in $2B$, there must be two distinct points $\mathbf{x}_1$ and $\mathbf{x}_2$ in X that translate by Λ' to the same point in $2B$:

$$\mathbf{x}_1 - 2\mathbf{v}_1 = \mathbf{x}_2 - 2\mathbf{v}_2$$

for some $\mathbf{v}_1, \mathbf{v}_2 \in \Lambda$. (If this did not happen, so different points in X translate by Λ' to different points in $2B$, then the volume of X is at most the volume of $2B$, which is a contradiction.) Since $\mathbf{x}_1 \neq \mathbf{x}_2$, $\mathbf{v}_1 \neq \mathbf{v}_2$. Then

$$\frac{\mathbf{x}_1 - \mathbf{x}_2}{2} = \mathbf{v}_1 - \mathbf{v}_2 \neq \mathbf{0}.$$

The difference $\mathbf{v}_1 - \mathbf{v}_2$ is a nonzero vector in Λ. Since X is centrally symmetric, $-\mathbf{x}_2 \in X$. Since X is convex, $\frac{\mathbf{x}_1 - \mathbf{x}_2}{2} = \frac{1}{2}\mathbf{x}_1 + \frac{1}{2}(-\mathbf{x}_2) \in X$. Thus X meets Λ in a nonzero vector $\mathbf{v}_1 - \mathbf{v}_2$.

Now assume X is compact and $\operatorname{vol}(X) \geqslant 2^n \operatorname{vol}(\Lambda)$. For $N \geqslant 1$ consider the region $X_N = (1+1/N)X$. Its volume is greater $2^n \operatorname{vol}(\Lambda)$, so by what we already proved there is a nonzero $\mathbf{v}_N \in X_N \cap \Lambda \subset 2X \cap \Lambda$. Since $2X$ is compact and Λ is discrete, the intersection $2X \cap \Lambda$ is finite. Letting N range over the positive integers, some nonzero $\mathbf{v} \in \Lambda$ is in X_N for infinitely many N, say $\mathbf{v} \in X_{N_i}$. Therefore $\mathbf{v}$ is in $\bigcap_{N_i}(1 + 1/N_i)X \subset \overline{X} = X$. ■

Minkowski's theorem has a simple proof, so it may not seem significant. It is profound through applications with clever choices of lattice and convex body.

Example 7.12. Minkowski's theorem can be used to reprove Fermat's two-square theorem: if p is a prime number and $p \equiv 1 \bmod 4$, then $p = a^2 + b^2$ for some integers a and b. From a solution $c \in \mathbf{Z}$ to $c^2 \equiv -1 \bmod p$ (c exists because $p \equiv 1 \bmod 4$) we write down the lattice[4] $\Lambda_c = \mathbf{Z}(c, 1) + \mathbf{Z}(p, 0) \subset \mathbf{Z}^2$, whose volume is $|\det(\begin{smallmatrix} c & p \\ 1 & 0 \end{smallmatrix})| = p$. The open disc $X = \{(a, b) \in \mathbf{R}^2 : a^2 + b^2 < 2p\}$ has area $2p\pi$ and $4 \operatorname{vol}(\Lambda_c) = 4p$. Since $2\pi p > 4p$, X meets Λ_c in a nonzero vector (a, b). Then $0 < a^2 + b^2 < 2p$. Every element of Λ_c has squared length a multiple of p (here we use $c^2 \equiv -1 \bmod p$), so $a^2 + b^2$ must equal p.

Using a lattice in $\mathbf{R}^4$, this argument can be extended to prove the four-square theorem (every nonnegative integer is a sum of 4 squares). See [57, Sect. 7.3].

[4] The congruence $a^2 + b^2 \equiv 0 \bmod p$ is the same as $a \equiv \pm cb \bmod p$. Writing $a = \pm cb + pk$, $(a, b) = b(\pm c, 1) + k(p, 0)$, so the solutions to $a^2 + b^2 \equiv 0 \bmod p$ lie on one of the two lattices $\Lambda_{\pm c}$. This is a natural explanation of where Λ_c is coming from.

To apply Minkowski's theorem to number fields, we will translate operations on K to operations on $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$. An element of $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ will be written as

$$\mathbf{v} = (x_1, \ldots, x_{r_1}, z_1, \ldots, z_{r_2}).$$

Table 7.2 provides analogous operations in both settings, keeping in mind that θ_K includes only half the complex embeddings of K while the trace and norm are a sum and product over all embeddings.

K	$\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$
addition	addition
multiplication	multiplication
$\mathbf{Q}$-basis	$\mathbf{R}$-basis
$\mathrm{Tr}_{K/\mathbf{Q}}(\gamma) = \sum_\sigma \sigma(\gamma)$	$T(\mathbf{v}) = \sum_{i=1}^{r_1} x_i + \sum_{j=1}^{r_2} 2\,\mathrm{Re}(z_j)$
$\mathrm{N}_{K/\mathbf{Q}}(\gamma) = \prod_\sigma \sigma(\gamma)$	$N(\mathbf{v}) = \prod_{i=1}^{r_1} x_i \cdot \prod_{j=1}^{r_2} (z_j \overline{z}_j)$
$\langle \gamma, \gamma' \rangle := \mathrm{Tr}_{K/\mathbf{Q}}(\gamma\gamma')$	$\langle \mathbf{v}, \mathbf{v}' \rangle := T(\mathbf{v}\mathbf{v}')$

Table 7.2: Operations on K translated to $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$.

Compatibility in Table 7.2, row by row, means: θ_K is additive and multiplicative, it sends a $\mathbf{Q}$-basis of K to an $\mathbf{R}$-basis of $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$, $T(\theta_K(\gamma)) = \mathrm{Tr}_{K/\mathbf{Q}}(\gamma)$, $N(\theta_K(\gamma)) = \mathrm{N}_{K/\mathbf{Q}}(\gamma)$, and $T(\theta_K(\gamma)\theta_K(\gamma')) = \mathrm{Tr}_{K/\mathbf{Q}}(\gamma\gamma')$.

Identifying $\mathbf{C} = \mathbf{R} + \mathbf{R}i$ with $\mathbf{R}^2$ lets us treat the vector space $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ as $\mathbf{R}^n$. While $\langle \mathbf{v}, \mathbf{v}' \rangle$ is a symmetric bilinear form on $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$, it is not generally the same as the dot product on $\mathbf{R}^n$. We need the dot product because of its connection to volumes, while $\langle \mathbf{v}, \mathbf{v}' \rangle$ is related to the trace pairing on K. To express $\langle \mathbf{v}, \mathbf{v}' \rangle$ in terms of the dot product, expand $\mathbf{v}$ to a vector in $\mathbf{R}^n$:

$$\mathbf{v} = (x_1, \ldots, x_{r_1}, a_1, b_1, \ldots, a_{r_2}, b_{r_2}) \in \mathbf{R}^n,$$

where $z_j = a_j + b_j i$. A formula for $\langle \mathbf{v}, \mathbf{v}' \rangle = T(\mathbf{v}\mathbf{v}')$ is

$$\begin{aligned}\langle \mathbf{v}, \mathbf{v}' \rangle &= T(x_1x_1', \ldots, x_{r_1}x_{r_1}', z_1z_1', \ldots, z_{r_2}z_{r_2}') \\ &= \sum_{i=1}^{r_1} x_i x_i' + \sum_{j=1}^{r_2} 2\,\mathrm{Re}(z_j z_j').\end{aligned}$$

Since $\mathrm{Re}(z_j z_j') = a_j a_j' - b_j b_j'$,

$$\langle \mathbf{v}, \mathbf{v}' \rangle = \sum_{i=1}^{r_1} x_i x_i' + \sum_{j=1}^{r_2} 2(a_j a_j' - b_j b_j'). \tag{7.2}$$

Comparing this to

$$\begin{aligned}\mathbf{v} \cdot \mathbf{v}' &= (x_1, \ldots, x_{r_1}, a_1, b_1, \ldots, a_{r_2}, b_{r_2}) \cdot (x_1', \ldots, x_{r_1}', a_1', b_1', \ldots, a_{r_2}', b_{r_2}') \\ &= \sum_{i=1}^{r_1} x_i x_i' + \sum_{j=1}^{r_2} (a_j a_j' + b_j b_j'),\end{aligned}$$

we are inspired to write $2(a_j a_j' - b_j b_j')$ in (7.2) as $a_j(2a_j') + b_j(-2b_j')$ to make $\langle \mathbf{v}, \mathbf{v}' \rangle$ have the form of a dot product in $\mathbf{R}^n$: it equals

$$(x_1, \ldots, x_{r_1}, a_1, b_1, \ldots, a_{r_2}, b_{r_2}) \cdot (x_1', \ldots, x_{r_1}', 2a_1', -2b_1', \ldots, 2a_{r_2}', -2b_{r_2}').$$

Therefore $\langle \mathbf{v}, \mathbf{v}' \rangle = \mathbf{v} \cdot J(\mathbf{v}')$, where

$$\begin{aligned} J(\mathbf{v}') &= (x_1', \ldots, x_{r_1}', 2a_1', -2b_1', \ldots, 2a_{r_2}', -2b_{r_2}') \\ &= \begin{pmatrix} I_{r_1} & O & \cdots & O \\ O & \left(\begin{smallmatrix} 2 & 0 \\ 0 & -2 \end{smallmatrix}\right) & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & \left(\begin{smallmatrix} 2 & 0 \\ 0 & -2 \end{smallmatrix}\right) \end{pmatrix} \begin{pmatrix} x_1' \\ \vdots \\ x_{r_1}' \\ a_1' \\ b_1' \\ \vdots \\ a_{r_2}' \\ b_{r_2}' \end{pmatrix}. \end{aligned}$$

Note J is given by a diagonal matrix with $\det(J) = (-4)^{r_2}$. When K is totally real, $J = I_n$ so $\langle \mathbf{v}, \mathbf{v}' \rangle$ coincides with the dot product $\mathbf{v} \cdot \mathbf{v}'$. Otherwise we have some extra 2's floating around if we pass between $\langle \mathbf{v}, \mathbf{v}' \rangle$ and $\mathbf{v} \cdot \mathbf{v}'$.

Theorem 7.13. *For a* $\mathbf{Q}$*-basis* $\{\alpha_1, \ldots, \alpha_n\}$ *of* K*, the lattice in* $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ *spanned by* $\{\theta_K(\alpha_1), \ldots, \theta_K(\alpha_n)\}$ *has volume*

$$\frac{1}{2^{r_2}} \sqrt{|\det(\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i \alpha_j))|} = \frac{1}{2^{r_2}} \sqrt{|\mathrm{disc}_{K/\mathbf{Q}}(\alpha_1, \ldots, \alpha_n)|}.$$

Proof. For $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in $\mathbf{R}^n$, the box they span has volume $\sqrt{|\det(\mathbf{v}_i \cdot \mathbf{v}_j)|}$.

(Theorem 3.23.) So the box in $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ spanned by $\{\theta_K(\alpha_i)\}$ has volume

$$\sqrt{|\det(\theta_K(\alpha_i) \cdot \theta_K(\alpha_j))|}.$$

Since

$$\langle \mathbf{v}, \mathbf{v}' \rangle = \mathbf{v} \cdot J(\mathbf{v}')$$

for all $\mathbf{v}$ and $\mathbf{v}'$ in $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$, we have

$$\mathbf{v} \cdot \mathbf{v}' = \left\langle \mathbf{v}, J^{-1}(\mathbf{v}') \right\rangle,$$

so the box spanned by the $\mathbf{v}_i$'s has volume

$$\sqrt{\left|\det\left(\langle \theta_K(\alpha_i), J^{-1}(\theta_K(\alpha_j))\rangle\right)\right|}. \tag{7.3}$$

There is an identity which can be used to express $\det(\left\langle \theta_K(\alpha_i), J^{-1}(\theta_K(\alpha_j))\right\rangle)$ in terms of $\det(\langle \theta_K(\alpha_i), \theta_K(\alpha_j)\rangle)$: for any linear map $A\colon \mathbf{R}^n \to \mathbf{R}^n$ and basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ of $\mathbf{R}^n$ and symmetric bilinear form $\langle \cdot, \cdot \rangle\colon \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$,

$$\det(\langle \mathbf{e}_i, A\mathbf{e}_j \rangle) = (\det A)\det(\langle \mathbf{e}_i, \mathbf{e}_j \rangle). \tag{7.4}$$

To prove this, the left side for fixed $\mathbf{e}_1, \dots, \mathbf{e}_n$ and fixed $\langle \cdot, \cdot \rangle$ is a function of A, sending $\mathrm{M}_n(\mathbf{R})$ to $\mathbf{R}$. As a function of the columns of A, $\det(\langle \mathbf{e}_i, A\mathbf{e}_j \rangle)$ is alternating and multilinear. A standard characterization of the determinant is that it is the unique alternating multilinear function $\mathrm{M}_n(\mathbf{R}) \to \mathbf{R}$ up to scaling. So there is a $c \in \mathbf{R}$ such that

$$\det(\langle \mathbf{e}_i, A\mathbf{e}_j \rangle) = c \cdot \det A$$

for all $A \in \mathrm{M}_n(\mathbf{R})$. At $A = I_n$ we get $c = \det(\langle \mathbf{e}_i, \mathbf{e}_j \rangle)$, which gives us (7.4).

Applying (7.4) to (7.3),

$$\sqrt{\left|\det\left(\langle \theta_K(\alpha_i), J^{-1}(\theta_K(\alpha_j))\rangle\right)\right|} = \sqrt{|\det(J^{-1})\det(\langle \theta_K(\alpha_i), \theta_K(\alpha_j)\rangle)|}.$$

Since $\det J = (-4)^{r_2}$, the box spanned by the $\theta_K(\alpha_i)$'s has volume

$$\begin{aligned}\sqrt{\left|\det\left(\langle\theta_K(\alpha_i), J^{-1}(\theta_K(\alpha_j))\rangle\right)\right|} &= \frac{1}{2^{r_2}}\sqrt{|\det(\langle\theta_K(\alpha_i),\theta_K(\alpha_j)\rangle)|}\\ &= \frac{1}{2^{r_2}}\sqrt{\left|\det(\mathrm{Tr}_{K/\mathbf{Q}}(\alpha_i\alpha_j))\right|}\\ &= \frac{1}{2^{r_2}}\sqrt{\left|\mathrm{disc}_{K/\mathbf{Q}}(\alpha_1,\ldots,\alpha_n)\right|}. \qquad \blacksquare\end{aligned}$$

Example 7.14. For any nonzero ideal $\mathfrak{a}$ in $\mathcal{O}_K$, a $\mathbf{Z}$-basis of $\mathfrak{a}$ is a $\mathbf{Q}$-basis of K, so $\theta_K(\mathfrak{a})$ is a lattice in $\mathbf{R}^{r_1}\times\mathbf{C}^{r_2}$. The volume of this lattice is

$$\frac{1}{2^{r_2}}\sqrt{|\mathrm{disc}(\mathfrak{a})|} = \frac{1}{2^{r_2}}\sqrt{[\mathcal{O}_K:\mathfrak{a}]^2\,|\mathrm{disc}(K)|} = \frac{1}{2^{r_2}}\sqrt{|\mathrm{disc}(K)|}\,\mathrm{N}(\mathfrak{a}).$$

If K is totally real, Theorem 7.13 tells us the Euclidean image of $\mathcal{O}_K$ is a lattice with volume $\sqrt{|\mathrm{disc}(K)|}$, which reminds us that discriminants are supposed to be something like squared volume.[5] In the general case we need an additional power of 2 in the volume formula, but that's just a technicality. All the main ideas in the proof are present in the totally real case.

7.2.2 Better Bound for Class Number

In Theorem 5.3, we showed that $\mathrm{Cl}(K)$ is finite by finding a constant $C>0$ such that every nonzero ideal $\mathfrak{a}\subset\mathcal{O}_K$ contains a nonzero α such that $\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| \leqslant C[\mathcal{O}_K:\mathfrak{a}]$. We will use Minkowski's theorem to prove a similar result with C replaced by another, often smaller, constant.

Theorem 7.15. *Every nonzero ideal $\mathfrak{a}\subset\mathcal{O}_K$ contains a nonzero α such that*

$$\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| \leqslant M[\mathcal{O}_K:\mathfrak{a}],$$

where $M = \frac{n!}{n^n}\left(\frac{4}{\pi}\right)^{r_2}\sqrt{|\mathrm{disc}(K)|}$.

Proof. We will apply Minkowski's theorem to a certain convex body in $\mathbf{R}^{r_1}\times\mathbf{C}^{r_2}$ and the lattice $\theta_K(\mathfrak{a})$. Recall $\mathrm{N}_{K/\mathbf{Q}}(\alpha) = N(\theta_K(\alpha))$. Since we want to find a nonzero element of $\mathfrak{a}$ with a particular bound on the absolute value of its norm, it is natural to consider the regions

$$\{\mathbf{v}\in\mathbf{R}^{r_1}\times\mathbf{C}^{r_2} : |N(\mathbf{v})|\leqslant t\} \tag{7.5}$$

[5] Recall we saw before that the lattice $\mathbf{Z}[\sqrt{2}]$ in $\mathbf{R}^2$ has volume $2\sqrt{2}$ and $\sqrt{|\,\mathrm{disc}(\mathbf{Z}[\sqrt{2}])|} = \sqrt{8} = 2\sqrt{2}$.

but these are usually *not* convex (or bounded). See the shaded part of Figure 7.10 for $K = \mathbf{Q}(\sqrt{2})$.

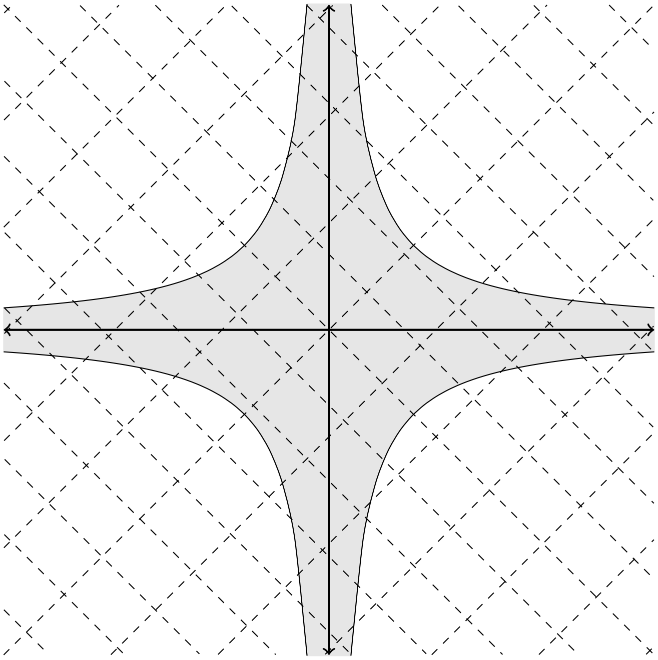


Figure 7.10: The region $\{(x_1, x_2) \in \mathbf{R}^2 : |x_1 x_2| \leqslant 5\}$ and $\theta(\mathbf{Z}[\sqrt{2}])$.

Instead of defining regions by directly bounding $|N(\mathbf{v})|$, which is a product, we will instead bound a sum using the arithmetic-geometric mean inequality, which says for $a_1, \ldots, a_n \geqslant 0$ that

$$\frac{a_1 + \cdots + a_n}{n} \geqslant \sqrt[n]{a_1 \cdots a_n}.$$

Applying this to $\{a_1, \ldots, a_n\} = \{|x_1|, \ldots, |x_{r_1}|, |z_1|, |\overline{z}_1|, \ldots, |z_{r_2}|, |\overline{z}_{r_2}|\}$, we get

$$\frac{\sum_{i=1}^{r_1} |x_i| + 2\sum_{j=1}^{r_2} |z_j|}{n} \geqslant \sqrt[n]{|N(\mathbf{v})|}.$$

This suggests we consider bounded regions of the form

$$X_t = \left\{\mathbf{v} \in \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} : \sum_{i=1}^{r_1} |x_i| + 2\sum_{j=1}^{r_2} |z_j| \leqslant t\right\}$$

for $t > 0$. These are convex, centrally symmetric, and compact. See the shaded part of Figure 7.11 for $K = \mathbf{Q}(\sqrt{2})$. We will show in Lemma 7.16 that

$$\operatorname{vol}(X_t) = 2^{r_1} \left(\frac{\pi}{2}\right)^{r_2} \frac{1}{n!} t^n.$$

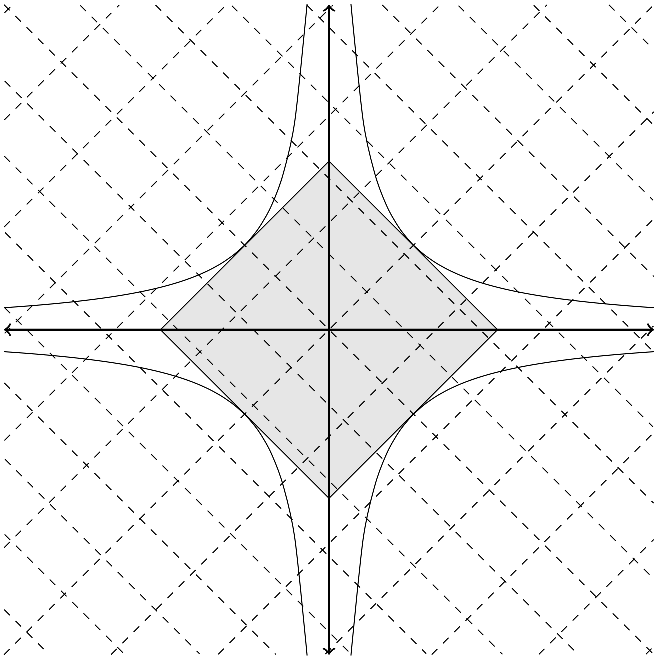


Figure 7.11: The region $\{(x_1, x_2) \in \mathbf{R}^2 : |x_1| + |x_2| \leqslant 2\sqrt{5}\}$ and $\theta(\mathbf{Z}[\sqrt{2}])$.

To apply Minkowski's convex body theorem to X_t and $\theta_K(\mathfrak{a})$, we want t to satisfy

$$\operatorname{vol}(X_t) \geqslant 2^n \operatorname{vol}(\theta_K(\mathfrak{a})) = 2^n \frac{1}{2^{r_2}} \sqrt{|\operatorname{disc}(K)|}\, \mathrm{N}(\mathfrak{a}) = 2^{r_1+r_2} \sqrt{|\operatorname{disc}(K)|}\, \mathrm{N}(\mathfrak{a}),$$

where the volume of $\theta_K(\mathfrak{a})$ is computed from Example 7.14. So we want

$$\begin{aligned} 2^{r_1}\left(\frac{\pi}{2}\right)^{r_2}\frac{1}{n!}t^n &\geqslant 2^{r_1+r_2}\sqrt{|\mathrm{disc}(K)|}\,\mathrm{N}(\mathfrak{a}) \\ \Longrightarrow t^n &\geqslant \left(\frac{2}{\pi}\right)^{r_2} n!2^{r_2}\sqrt{|\mathrm{disc}(K)|}\,\mathrm{N}(\mathfrak{a}) \\ &= \left(\frac{4}{\pi}\right)^{r_2} n!\sqrt{|\mathrm{disc}(K)|}\,\mathrm{N}(\mathfrak{a}). \end{aligned}$$

There are many choices of t which fit this inequality. For any of them, $\theta_K(\mathfrak{a})$ meets X_t in a nonzero vector. Which choice of t should we make? Let N be the greatest integer $\leqslant (4/\pi)^{r_2}n!\sqrt{|\mathrm{disc}(K)|}\,\mathrm{N}(\mathfrak{a})$. By continuity we can choose t^n close enough to $(4/\pi)^{r_2}n!\sqrt{|\mathrm{disc}(K)|}\,\mathrm{N}(\mathfrak{a})$ from above so that N is also the greatest integer $\leqslant t^n$.

Having chosen t, let $\theta_K(\alpha)$ be a nonzero element of $\theta_K(\mathfrak{a})\cap X_t$, so α is a nonzero element of $\mathfrak{a}$ and

$$\sum_\sigma |\sigma(\alpha)| \leqslant t,$$

where the sum runs over all real and complex embeddings σ of K. We have

$$\sqrt[n]{\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right|} = \sqrt[n]{\prod_\sigma |\sigma(\alpha)|} \leqslant \frac{1}{n}\sum_\sigma |\sigma(\alpha)| \leqslant \frac{t}{n}.$$

Thus

$$\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| \leqslant \left(\frac{t}{n}\right)^n = \frac{t^n}{n^n} \Longrightarrow n^n\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| \leqslant t^n.$$

Since $n^n\,\mathrm{N}_{K/\mathbf{Q}}(\alpha)$ is an integer less than t^n, we can now bring in the number N:

$$n^n\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| \leqslant N \leqslant \left(\frac{4}{\pi}\right)^{r_2} n!\sqrt{|\mathrm{disc}(K)|}\,\mathrm{N}(\mathfrak{a}).$$

Divide the left and right sides by n^n and we're done. ■

The constant

$$M = \frac{n!}{n^n}\left(\frac{4}{\pi}\right)^{r_2}\sqrt{|\mathrm{disc}(K)|}$$

in Theorem 7.15 is called the *Minkowski bound* for K. (While the Kronecker bound for K depends on a $\mathbf{Z}$-basis of $\mathcal{O}_K$, the Minkowski bound doesn't.) Using the Minkowski bound in place of the Kronecker bound, the proof of Theorem 5.4 shows $\mathrm{Cl}(K)$ is

- *represented* by ideals $\mathfrak{a} \subset \mathcal{O}_K$ such that

$$\mathrm{N}(\mathfrak{a}) \leqslant \frac{n!}{n^n}\left(\frac{4}{\pi}\right)^{r_2}\sqrt{|\mathrm{disc}(K)|},$$

- *generated* by classes of primes $\mathfrak{p}$ with $\mathrm{N}(\mathfrak{p}) \leqslant \frac{n!}{n^n}(\frac{4}{\pi})^{r_2}\sqrt{|\mathrm{disc}(K)|}$.

Table 7.3 provides some examples where we previously (Sections 5.2 and 5.3) computed the Kronecker bound and then found the class numbers of $\mathbf{Q}(\sqrt{-5})$ and $\mathbf{Q}(\sqrt{5})$ by hand, but not of $\mathbf{Q}(\alpha)$ and $\mathbf{Q}(\beta)$ because the Kronecker bounds for them are larger. We can compute the class numbers for all these number fields using the Minkowski bound. In $\mathbf{Q}(\sqrt{-5})$, the only ideals of norm less than 2.85 are (1) and $\mathfrak{p}_2$, and $\mathfrak{p}_2$ is not principal, so $h(\mathbf{Q}(\sqrt{-5})) = 2$. The Minkowski bounds for the second and third fields in Table 7.3 are less than 2, so $h = 1$ for them. In $\mathbf{Q}(\beta)$, no prime ideal has norm 2 or 3 since $T^5 - T - 1$ has no root in $\mathbf{F}_2$ or $\mathbf{F}_3$, so $h(\mathbf{Q}(\beta)) = 1$.

K	Kronecker bound	Minkowski bound
$\mathbf{Q}(\sqrt{-5})$	10.4	$\frac{2!}{2^2}\left(\frac{4}{\pi}\right)\sqrt{20} \approx 2.85$
$\mathbf{Q}(\sqrt{5})$	4.2	$\frac{2!}{2^2}\sqrt{5} \approx 1.12$
$\mathbf{Q}(\alpha)$, $\alpha^3 - \alpha - 1 = 0$	28.08	$\frac{3!}{3^3}\left(\frac{4}{\pi}\right)\sqrt{23} \approx 1.35$
$\mathbf{Q}(\beta)$, $\beta^5 - \beta - 1 = 0$	3454.4	$\frac{5!}{5^5}\left(\frac{4}{\pi}\right)^2\sqrt{2869} \approx 3.3$

Table 7.3: Comparing the Kronecker bound and the Minkowski bound.

If we want to determine $h(K)$ using the Minkowski bound, the hardest term to compute in this constant is $\mathrm{disc}(K)$. If we don't know $\mathcal{O}_K$ then we can replace $\mathrm{disc}(K)$ in the Minkowski bound by $\mathrm{disc}(\mathbf{Z}[\alpha])$ for any algebraic integer α such that $K = \mathbf{Q}(\alpha)$. After all, $\mathrm{disc}(\mathbf{Z}[\alpha]) = [\mathcal{O}_K : \mathbf{Z}[\alpha]]^2\,\mathrm{disc}(K)$, so $|\mathrm{disc}(\mathbf{Z}[\alpha])|$ is at least as large as $|\mathrm{disc}(K)|$, and using it would have the effect of replacing the true Minkowski bound by a larger bound, so that will work too.

Let's return to the volume computation left out of the proof of Theorem 7.15.

Lemma 7.16. *For $t > 0$, the set $\{\mathbf{v} \in \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} : \sum_{i=1}^{r_1} |x_i| + 2\sum_{j=1}^{r_2} |z_j| \leqslant t\}$ has volume $2^{r_1}(\frac{\pi}{2})^{r_2}\frac{t^{r_1+2r_2}}{(r_1+2r_2)!}$.*

Proof. For $t > 0$, let

$$V_{r_1,r_2}(t) = \mathrm{vol}\left(\left\{\mathbf{v} \in \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} : \sum_{i=1}^{r_1} |x_i| + 2\sum_{j=1}^{r_2} |z_j| \leqslant t\right\}\right).$$

Scaling an interval by t scales its length by t, while scaling a disc by t scales its area by t^2. Therefore $V_{r_1,r_2}(t) = V_{r_1,r_2}(1)t^{r_1+2r_2}$.

We will prove the formula for $V_{r_1,r_2}(t)$ by two inductive arguments. First we assume $r_2 = 0$ and show $V_{r_1,0}(t) = 2^{r_1}t^{r_1}/r_1!$. Since $V_{r_1,0}(t) = V_{r_1,0}(1)t^{r_1}$, it remains to compute

$$V_{r_1,0}(1) = \mathrm{vol}\left(\left\{\mathbf{v} \in \mathbf{R}^{r_1} : \sum_{i=1}^{r_1} |x_i| \leqslant 1\right\}\right)$$

and show it is $2^{r_1}/r_1!$. When $r_1 = 1$ this is the length of $(-1,1)$, which is 2. For $r_1 > 1$, we compute the volume by slices, in the same way the volume of a nice solid in $\mathbf{R}^3$ can be computed in freshman calculus by integrating the cross-sectional area across the solid:

$$\begin{aligned} V_{r_1,0}(1) &= \int_{-1}^{1} \mathrm{vol}\left(\left\{\mathbf{v} \in \mathbf{R}^{r_1-1} : \sum_{i=1}^{r_1-1} |x_i| < 1-|x|\right\}\right) \mathrm{d}x \\ &= \int_{-1}^{1} V_{r_1-1,0}(1-|x|)\,\mathrm{d}x \\ &= \int_{-1}^{1} V_{r_1-1,0}(1)(1-|x|)^{r_1-1}\,\mathrm{d}x \\ &= 2V_{r_1-1,0}(1)\int_{0}^{1}(1-x)^{r_1-1}\,\mathrm{d}x \\ &= \frac{2}{r_1}V_{r_1-1,0}(1). \end{aligned}$$

By induction, $V_{r_1,0}(1) = 2^{r_1}/r_1!$.

Now assume $r_2 > 0$. We will get a recursive formula for $V_{r_1,r_2}(1)$ in terms of $V_{r_1,r_2-1}(1)$ by integrating with respect to z_{r_2} first. Write z_{r_2} in polar coordinates and note that $|z_{r_2}| < 1/2$ in the region whose volume we're computing:

$$\begin{aligned} V_{r_1,r_2}(1) &= \int_0^{2\pi}\int_0^{1/2} \mathrm{vol}\left(\left\{\mathbf{v} \in \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} : \sum_{i=1}^{r_1}|x_i| + 2\sum_{j=1}^{r_2-1}|z_j| \leqslant 1-2r\right\}\right) r\mathrm{d}r\mathrm{d}\theta \\ &= \int_0^{2\pi}\int_0^{1/2} V_{r_1,r_2-1}(1-2r)r\mathrm{d}r\mathrm{d}\theta \\ &= \int_0^{2\pi}\int_0^{1/2} V_{r_1,r_2-1}(1)(1-2r)^{r_1+2(r_2-1)}r\mathrm{d}r\mathrm{d}\theta \\ &= 2\pi V_{r_1,r_2-1}(1)\int_0^{1/2}(1-2r)^{r_1+2r_2-2}r\mathrm{d}r. \end{aligned}$$

In the integral, make the change of variables $r = u/2$ so $u \in (0, 1)$:

$$\begin{aligned} V_{r_1,r_2}(1) &= \frac{\pi}{2} V_{r_1,r_2-1}(1) \int_0^1 (1-u)^{r_1+2r_2-2} u \, du \\ &= \frac{\pi}{2} V_{r_1,r_2-1}(1) \int_0^1 u^{r_1+2r_2-2}(1-u) \, du \\ &= \frac{\pi}{2} V_{r_1,r_2-1}(1) \frac{1}{(r_1+2r_2)(r_2+2r_2-1)}. \end{aligned}$$

By induction, $V_{r_1,r_2}(1) = 2^{r_1}(\frac{\pi}{2})^{r_2} \frac{1}{(r_1+2r_2)!}$. ■

7.2.3 Class Group Calculations

We will use the Minkowski bound to compute class groups of various quadratic fields. Let $K = \mathbf{Q}(\sqrt{d})$, with d a squarefree integer. The Minkowski bounds for real and imaginary K are in Table 7.4.

d mod 4	Real	Imaginary
1	$\frac{1}{2}\sqrt{d}$	$\frac{2}{\pi}\sqrt{\lvert d\rvert}$
2, 3	$\sqrt{d}$	$\frac{4}{2\pi}\sqrt{\lvert d\rvert}$

Table 7.4: Minkowski bound for quadratic fields

Example 7.17. If the Minkowski bound is less than 2 then $\mathrm{Cl}(K)$ is represented by the unit ideal so $h(K) = 1$. This tells us the following quadratic fields have class number 1:

$$\mathbf{Q}(\sqrt{2}),\ \mathbf{Q}(\sqrt{3}),\ \mathbf{Q}(\sqrt{5}),\ \mathbf{Q}(\sqrt{13}),\ \mathbf{Q}(i),\ \mathbf{Q}(\sqrt{-2}),\ \mathbf{Q}(\sqrt{-3}),\ \text{and } \mathbf{Q}(\sqrt{-7}).$$

There are other real and imaginary quadratic fields with class number 1, but the Minkowski bound in the other cases is not less than 2, so we need extra work to show the class number is 1. For example, $\mathbf{Q}(\sqrt{-11})$ and $\mathbf{Q}(\sqrt{-19})$ have Minkowski bound between 2 and 3. In each case 2 is inert, so $h = 1$ for these two fields.

Example 7.18. Let $K = \mathbf{Q}(\sqrt{82})$. We will show the class group is cyclic of order 4.

Here $n = 2$, $r_2 = 0$, and $\mathrm{disc}(K) = 4 \cdot 82$, so the Minkowski bound is ≈ 9.055. We look at the primes lying over 2, 3, 5, and 7.

The following table describes how (p) factors from the way $T^2 - 82$ factors modulo p.

p	$T^2 - 82 \bmod p$	(p)
2	T^2	$\mathfrak{p}_2^2$
3	$(T+1)(T-1)$	$\mathfrak{p}_3\mathfrak{p}_3'$
5	irreducible	(5)
7	irreducible	(7)

Thus, the class group of $\mathbf{Q}(\sqrt{82})$ is generated by $[\mathfrak{p}_2]$ and $[\mathfrak{p}_3]$, with $\mathfrak{p}_2^{-1} \sim \mathfrak{p}_2$ and $\mathfrak{p}_3' \sim \mathfrak{p}_3^{-1}$.

Since $\mathrm{N}_{K/\mathbf{Q}}(10+\sqrt{82}) = 18 = 2 \cdot 3^2$, and $10+\sqrt{82}$ is not divisible by 3, $(10+\sqrt{82})$ is divisible by just one of $\mathfrak{p}_3$ and $\mathfrak{p}_3'$. Let $\mathfrak{p}_3$ be that prime (explicitly, $\mathfrak{p}_3 = (3, 10+\sqrt{82}) = (3, 1+\sqrt{82})$), so $(10+\sqrt{82}) = \mathfrak{p}_2\mathfrak{p}_3^2$. Thus $\mathfrak{p}_2 \sim \mathfrak{p}_3^{-2}$, so the class group of K is generated by $[\mathfrak{p}_3]$ and

$$[\mathfrak{p}_2]^2 = 1, \quad [\mathfrak{p}_3]^2 = [\mathfrak{p}_2].$$

Therefore $[\mathfrak{p}_3]$ has order dividing 4.

We will show $\mathfrak{p}_2$ is nonprincipal, so $[\mathfrak{p}_3]$ has order 4, and thus K has a class group $\langle[\mathfrak{p}_3]\rangle \cong \mathbf{Z}/4\mathbf{Z}$.

If $\mathfrak{p}_2 = (x+y\sqrt{82})$, then $x^2 - 82y^2 = \pm 2$. We want to show this has no integral solution. Since $1+\sqrt{82}$ has norm -1, we only need to focus on $x^2 - 82y^2 = 2$. The obvious first thing to do is look for a contradiction with congruences. Since $82 = 2 \cdot 41$, we reduce mod 41 and get $x^2 \equiv 2 \bmod 41$. This is no contradiction, since $17^2 \equiv 2 \bmod 41$. In fact, searching for a contradiction with congruences is hopeless, because $x^2 - 82y^2 \equiv 2 \bmod m$ is solvable for every m (Exercise 7.6). We need a different idea. One possibility is to use the Diophantine approximation method of Section 1.4 for $x^2 - dy^2 = n$. We will instead use the ideal factorization method of Section 4.6.

Since $\mathfrak{p}_2^2 = (2)$, if $\mathfrak{p}_2 = (x+y\sqrt{82})$ then

$$2 = (x+y\sqrt{82})^2 u \tag{7.6}$$

for some unit u in $\mathbf{Z}[\sqrt{82}]$. Taking norms of both sides shows u must have positive norm, so u has norm 1.

The unit group of $\mathbf{Z}[\sqrt{82}]$ is $\pm(9+\sqrt{82})^{\mathbf{Z}}$, with $9+\sqrt{82}$ having norm -1. Therefore the positive units of norm 1 are the integral powers of $(9+\sqrt{82})^2$, which are all squares. A unit u that is a square in (7.6) can be absorbed into $(x+y\sqrt{82})^2$, so we have to be able to solve $2=(x+y\sqrt{82})^2$ in integers x and y. Equating coefficients of a basis on both sides quickly leads to a contradiction, so $\mathfrak{p}_2$ is not principal.

Example 7.19. Let $K=\mathbf{Q}(\sqrt{-14})$. We will show the class group is cyclic of order 4. (This was Exercise 5.10, using the Kronecker bound.)

Here $n=2, r_2=1$, and $\mathrm{disc}(K)=-56$. The Minkowski bound is ≈ 4.764, so the class group is generated by primes dividing (2) and (3). The following table shows how (2) and (3) factor in $\mathcal{O}_K$ based on how T^2+14 factors modulo 2 and modulo 3.

p	$T^2+14 \bmod p$	(p)
2	T^2	$\mathfrak{p}_2^2$
3	$(T-1)(T+1)$	$\mathfrak{p}_3\mathfrak{p}_3'$

Since $\mathfrak{p}_2^2\sim 1$, $\mathfrak{p}_2\sim\mathfrak{p}_2^{-1}$. Since $\mathfrak{p}_3\mathfrak{p}_3'\sim 1$, $\mathfrak{p}_3'\sim\mathfrak{p}_3^{-1}$. Therefore the class group of K is generated by $[\mathfrak{p}_2]$ and $[\mathfrak{p}_3]$.

Both $\mathfrak{p}_2$ and $\mathfrak{p}_3$ are nonprincipal, since the equations $a^2+14b^2=2$ and $a^2+14b^2=3$ have no integral solutions.

To find relations between $\mathfrak{p}_2$ and $\mathfrak{p}_3$, we use $\mathrm{N}_{K/\mathbf{Q}}(2+\sqrt{-14})=18=2\cdot 3^2$. The ideal $(2+\sqrt{-14})$ is divisible by only one of $\mathfrak{p}_3$ and $\mathfrak{p}_3'$, since $2+\sqrt{-14}$ is not a multiple of 3. Without loss of generality, we may let $\mathfrak{p}_3$ be the prime of norm 3 dividing $(2+\sqrt{-14})$. Then $\mathfrak{p}_2\mathfrak{p}_3^2\sim 1$, so

$$\mathfrak{p}_3^2\sim\mathfrak{p}_2^{-1}\sim\mathfrak{p}_2,$$

so the class group of K is generated by $[\mathfrak{p}_3]$. Since $\mathfrak{p}_2$ is nonprincipal and $\mathfrak{p}_2^2\sim 1$, $[\mathfrak{p}_3]$ has order 4. Thus, the class group of K is cyclic of order 4.

Example 7.20. Let $K=\mathbf{Q}(\sqrt{-30})$. We will show the class group is a product of two cyclic groups of order 2.

Here $n=2, r_2=1$, and $\mathrm{disc}(K)=-120$. The Minkowski bound is ≈ 6.97, so the class group is generated by primes dividing 2, 3, and 5.

The following table shows how these primes factor into prime ideals.

p	$T^2+30 \bmod p$	(p)
2	T^2	$\mathfrak{p}_2^2$
3	T^2	$\mathfrak{p}_3^2$
5	T^2	$\mathfrak{p}_5^2$

For $a, b \in \mathbf{Z}$, $\mathrm{N}_{K/\mathbf{Q}}(a + b\sqrt{-30}) = a^2 + 30b^2$ is never 2, 3, or 5. Therefore $\mathfrak{p}_2$, $\mathfrak{p}_3$, and $\mathfrak{p}_5$ are nonprincipal and their ideal classes have order 2 in the class group of K. Moreover, since $\mathrm{N}_{K/\mathbf{Q}}(\sqrt{-30}) = 30 = 2 \cdot 3 \cdot 5$, $(\sqrt{-30}) = \mathfrak{p}_2\mathfrak{p}_3\mathfrak{p}_5$. Thus, in the class group, $\mathfrak{p}_2\mathfrak{p}_3\mathfrak{p}_5 \sim 1$, so $[\mathfrak{p}_2]$ and $[\mathfrak{p}_3]$ generate the class group.

The relation $\mathfrak{p}_2\mathfrak{p}_3\mathfrak{p}_5 \sim 1$ in the class group can be rewritten as

$$[\mathfrak{p}_2][\mathfrak{p}_3] = [\mathfrak{p}_5]^{-1} = [\mathfrak{p}_5].$$

Since $\mathfrak{p}_5$ is nonprincipal and $[\mathfrak{p}_2]$ and $[\mathfrak{p}_3]$ have order 2 in the class group, $[\mathfrak{p}_2] \neq [\mathfrak{p}_3]$. Therefore the class group of K is a product of two cyclic groups of order 2.

Example 7.21. Let $K = \mathbf{Q}(\sqrt{79})$. We will show the class group is cyclic of order 3. (This is the first real quadratic field $\mathbf{Q}(\sqrt{d})$, ordered by squarefree d, with a class number greater than 2.)

Here $n = 2, r_2 = 0$, and $\mathrm{disc}(K) = 4 \cdot 79$. The Minkowski bound is ≈ 8.88, so the class group is generated by primes dividing 2, 3, 5, and 7. The following table shows how these primes factor in $\mathcal{O}_K$.

p	$T^2 - 79 \bmod p$	(p)
2	$(T-1)^2$	$\mathfrak{p}_2^2$
3	$(T+1)(T-1)$	$\mathfrak{p}_3\mathfrak{p}_3'$
5	$(T+2)(T-2)$	$\mathfrak{p}_5\mathfrak{p}_5'$
7	$(T+3)(T-3)$	$\mathfrak{p}_7\mathfrak{p}_7'$

Therefore the class group is generated by $[\mathfrak{p}_2]$, $[\mathfrak{p}_3]$, $[\mathfrak{p}_5]$, and $[\mathfrak{p}_7]$. At this point $\mathfrak{p}_3$, $\mathfrak{p}_5$, and $\mathfrak{p}_7$ can be either prime lying over 3, 5, and 7.

Here is a table which factors $|\,\mathrm{N}_{K/\mathbf{Q}}(a + \sqrt{79})|$ for a running from 1 to 10.

a	$\lvert\,\mathrm{N}_{K/\mathbf{Q}}(a+\sqrt{79})\rvert$
1	$2 \cdot 3 \cdot 13$
2	$3 \cdot 5^2$
3	$2 \cdot 5 \cdot 7$
4	$3^2 \cdot 7$
5	$2 \cdot 3^3$
6	43
7	$2 \cdot 3 \cdot 5$
8	$3 \cdot 5$
9	2
10	$3 \cdot 7$

From $a = 9$, we see $\mathfrak{p}_2 = (9+\sqrt{79}) \sim 1$. From $a = 3$, $(3+\sqrt{79}) = \mathfrak{p}_2\mathfrak{p}_5\mathfrak{p}_7$ (which pins down a choice of $\mathfrak{p}_5$ and $\mathfrak{p}_7$). From $a = 7$ we can write $(7+\sqrt{79}) = \mathfrak{p}_2\mathfrak{p}_3\mathfrak{p}_5'$ (We can't have $\mathfrak{p}_5$ dividing $(7+\sqrt{79})$ because then $7+\sqrt{79} \equiv 3+\sqrt{79} \bmod \mathfrak{p}_5$, so $4 \equiv 0 \bmod \mathfrak{p}_5$, which is false.) Passing to ideal classes, $1 = [\mathfrak{p}_5][\mathfrak{p}_7]$ and $1 = [\mathfrak{p}_3][\mathfrak{p}_5'] = [\mathfrak{p}_3][\mathfrak{p}_5]^{-1}$, so the class group of K is generated by $[\mathfrak{p}_5] = [\mathfrak{p}_3]$.

Consider now $a = 5$. Since $5+\sqrt{79}$ has absolute norm $2 \cdot 27$ and is not divisible by 3, $(5+\sqrt{79})$ is only divisible by one of $\mathfrak{p}_3$ or $\mathfrak{p}_3'$. Having already chosen $\mathfrak{p}_3$ as a factor of $(7+\sqrt{79})$, it is not also a factor of $(5+\sqrt{79})$ since the generators of those ideals differ by 2. So $(5+\sqrt{79}) = \mathfrak{p}_2\mathfrak{p}_3'^3 \sim \mathfrak{p}_3'^3 \sim \mathfrak{p}_3^{-3}$. Thus, the class group is either trivial or cyclic of order 3.

We now show $\mathfrak{p}_3$ is not principal, so the class group is cyclic of order 3. Our method will be similar to the work with $\mathbf{Q}(\sqrt{82})$ in Example 7.18. In particular, we need knowledge of the unit group of $\mathbf{Z}[\sqrt{79}]$.

If $\mathfrak{p}_3$ is principal then so is $\mathfrak{p}_3'$. Writing $\mathfrak{p}_3' = (\alpha)$, the factorization of $(5+\sqrt{79})$ implies

$$\begin{aligned}(\alpha^3) &= \mathfrak{p}_3'^3\\ &= (5+\sqrt{79})\mathfrak{p}_2^{-1}\\ &= (5+\sqrt{79})(9+\sqrt{79})^{-1}\\ &= (-17+2\sqrt{79}).\end{aligned}$$

Thus

$$\alpha^3 = (-17+2\sqrt{79})u, \tag{7.7}$$

where u is a unit in $\mathbf{Z}[\sqrt{79}]$. The fundamental unit of $\mathbf{Z}[\sqrt{79}]$ is

$$\varepsilon = 80 + 9\sqrt{79},$$

so $\mathbf{Z}[\sqrt{79}]^\times = \pm\varepsilon^{\mathbf{Z}}$. In (7.7), u only matters as a unit modulo cubes of units (a unit cube can be absorbed into α^3). Therefore we may assume $u = 1, \varepsilon$, or ε^2, By a direct calculation,

$$(-17+2\sqrt{79})\varepsilon = 64+7\sqrt{79}, \quad (-17+2\sqrt{79})\varepsilon^2 = 9937+1118\sqrt{79}.$$

Writing $\alpha = x + y\sqrt{79}$ for unknown integers x and y, we have

$$\alpha^3 = x(x^2+3\cdot 79y^2) + y(3x^2+79y^2)\sqrt{79}.$$

Taking ideal norms in the hypothetical equation $(x+y\sqrt{79}) = \mathfrak{p}_3$, $|x^2-79y^2| = 3$, so both x and y are nonzero. Therefore the coefficient $y(3x^2+79y^2)$ of $\sqrt{79}$ in α^3 is, in absolute value, at least $3+79 = 82$. The coefficients of $\sqrt{79}$ in $-17+2\sqrt{79}$ and $(-17+2\sqrt{79})\varepsilon$, are 2 and 7, so neither of these is α^3.

Could $\alpha^3 = 9937 + 1118\sqrt{79}$? If so, then

$$y(3x^2+79y^2) = 1118 = 2\cdot 13\cdot 43.$$

Thus y (which must be positive by this equation) has 8 possibilities. For each choice of y, we try to solve for x as an integer. One possibility works: $y = 2$ and $x = 9$. So $\alpha = 9+2\sqrt{79}$. But this number has norm -235, not ± 3. We have a contradiction, so $\mathfrak{p}_3$ is not principal.

Example 7.22. Let $K = \mathbf{Q}(\sqrt{-65})$. We will show its class group is isomorphic to $\mathbf{Z}/2\mathbf{Z}\times\mathbf{Z}/4\mathbf{Z}$.

The Minkowski bound is $(4/\pi)\sqrt{65} \approx 10.26$, so we should factor 2, 3, 5, and 7 in $\mathcal{O}_K = \mathbf{Z}[\sqrt{-65}]$. From the following table, the class group is generated by $[\mathfrak{p}_2]$, either prime lying over 3, and $[\mathfrak{p}_5]$.

p	$T^2+65 \bmod p$	(p)
2	$(T+1)^2$	$\mathfrak{p}_2^2$
3	$(T+1)(T+2)$	$\mathfrak{p}_3\mathfrak{p}_3'$
5	T^2	$\mathfrak{p}_5^2$
7	T^2+65	(7)

If we factor $\mathrm{N}(a+\sqrt{-65}) = a^2+65$ for small a, looking for only factors of 2, 3, and 5, then we get examples at $a = 4$ and $a = 5$.

a	a^2+65
1	$3\cdot 11$
2	$3\cdot 23$
3	$2\cdot 37$
4	3^4
5	$2\cdot 3^2\cdot 5$

Since $(4+\sqrt{-65})$ is not divisible by (3), the ideal $(4+\sqrt{-65})$ is divisible by only one of the prime factors of (3). Choose $\mathfrak{p}_3$ as that prime, so

$$(4+\sqrt{-65}) = \mathfrak{p}_3^4.$$

Then

$$(5+\sqrt{-65}) = \mathfrak{p}_2\mathfrak{p}_3'^2\mathfrak{p}_5,$$

so the class group is generated by $[\mathfrak{p}_2]$ and $[\mathfrak{p}_3]$.

Since $\mathfrak{p}_2^2 = (2)$ and $\mathfrak{p}_3^4 = (4+\sqrt{-65})$, $[\mathfrak{p}_2]^2 = [1]$ and $[\mathfrak{p}_3]^4 = [1]$. The ideal $\mathfrak{p}_2$ is nonprincipal, since there is no integral solution to the equation $2 = x^2+65y^2$. The only integral solution to $9 = x^2+65y^2$ is $x = \pm 3$ and $y = 0$, so if $\mathfrak{p}_3^2$ were principal then $\mathfrak{p}_3^2 = (3) = \mathfrak{p}_3\mathfrak{p}_3'$, and that is false ($\mathfrak{p}_3 \neq \mathfrak{p}_3'$). Therefore $[\mathfrak{p}_2]$ has order 2 and $[\mathfrak{p}_3]$ has order 4. Can $[\mathfrak{p}_3]^2 = [\mathfrak{p}_2]$? If so, then $[\mathfrak{p}_2\mathfrak{p}_3^2] = [\mathfrak{p}_2]^2 = [1]$, so $\mathfrak{p}_2\mathfrak{p}_3^2$ is principal. But $18 = x^2+65y^2$ has no integral solution. Therefore the subgroups $\langle[\mathfrak{p}_2]\rangle$ and $\langle[\mathfrak{p}_3]\rangle$ of $\mathrm{Cl}(K)$ intersect trivially, so the class group is

$$\langle[\mathfrak{p}_2],[\mathfrak{p}_3]\rangle \cong \langle[\mathfrak{p}_2]\rangle \times \langle[\mathfrak{p}_3]\rangle \cong \mathbf{Z}/2\mathbf{Z}\times\mathbf{Z}/4\mathbf{Z}.$$

We end this section with one example of a class group computation for a cubic field.

Example 7.23. The Minkowski bound for $\mathbf{Q}(\sqrt[3]{2})$ is $(6/27)(4/\pi)\sqrt{108} \approx 2.94$, so we only need to factor (2). Since $(2) = (\sqrt[3]{2})^3$, $h = 1$: the ring $\mathbf{Z}[\sqrt[3]{2}]$ is a PID.

The lesson to draw from these examples is that computing a class group where $h > 1$ requires a concrete understanding of the unit group. The class group and unit group are intimately bound up with each other, and this is why our knowledge of class groups of imaginary quadratic fields (where the unit group is finite, and usually just ±1) is more extensive than that of real quadratic fields.

7.2.4 Application: Lower Bounds on Discriminants

Since the primes which ramify in a number field K are the prime factors of $\mathrm{disc}(K)$, to say no prime ramifies in K is the same as saying $\mathrm{disc}(K) = \pm 1$. Kronecker had conjectured that when $K \neq \mathbf{Q}$ there is always some prime which ramifies in K, and this was an open problem for a decade until Minkowski used his convex body theorem to show $|\mathrm{disc}(K)| > 1$.

Theorem 7.24 (Minkowski). *For any number field $K \neq \mathbf{Q}$, $|\mathrm{disc}(K)| > 1$. In particular, there is a prime number which ramifies in K.*

Proof. Set $n = [K : \mathbf{Q}] \geqslant 2$. Use $\mathfrak{a} = \mathcal{O}_K$ in Theorem 7.15 to conclude there is a nonzero $\alpha \in \mathcal{O}_K$ such that $\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| \leqslant \frac{n!}{n^n}\left(\frac{4}{\pi}\right)^{r_2}\sqrt{|\mathrm{disc}(K)|}$. Since $\mathrm{N}_{K/\mathbf{Q}}(\alpha)$

is a nonzero integer,

$$1 \leqslant \frac{n!}{n^n}\left(\frac{4}{\pi}\right)^{r_2}\sqrt{|\mathrm{disc}(K)|}.$$

So we have

$$|\mathrm{disc}(K)| \geqslant \left(\frac{\pi}{4}\right)^{2r_2}\frac{n^{2n}}{n!^2} \geqslant \left(\frac{\pi}{4}\right)^{n}\frac{n^{2n}}{n!^2} \tag{7.8}$$

since $0 < \frac{\pi}{4} < 1$. Check that

$$a_n = \left(\frac{\pi}{4}\right)^{n}\frac{n^{2n}}{n!^2}$$

is strictly increasing for $n \geqslant 2$ ($\frac{a_{n+1}}{a_n} > 1$) and

$$a_2 = \left(\frac{\pi}{4}\right)^{2}\frac{16}{4} = \frac{\pi^2}{4} > 1.$$ ■

Example 7.25. For any number field K of degree n,

$$|\mathrm{disc}(K)| \geqslant \left(\frac{\pi}{4}\right)^{n}\frac{n^{2n}}{n!^2} \sim \frac{1}{2\pi n}\left(\frac{\pi e^2}{4}\right)^{n} \tag{7.9}$$

by Stirling's formula. Since $\pi e^2/4 \approx 5.8$, $|\mathrm{disc}(K)|$ grows very quickly with n.

n	$\left(\frac{\pi}{4}\right)^n \frac{n^{2n}}{n!^2}$
2	2.47
3	9.81
4	43.29

Table 7.5: Lower bound on $|\mathrm{disc}(K)|$ in terms of $n = [K : \mathbf{Q}]$.

By Table 7.5, no $\mathcal{O}_K$ has $\mathrm{disc}(K) = 4$ or 9: such K must be a quadratic field, but no quadratic field has discriminant a perfect square. Recall Stickelberger's theorem: $\mathrm{disc}(K) \equiv 0, 1 \bmod 4$. Now we see some integers that are 0 or 1 mod 4 are not the discriminant of any number field.

Corollary 7.26. *If K and L are number fields with relatively prime discriminants then $K \cap L = \mathbf{Q}$.*

Proof. If $K \cap L$ is not $\mathbf{Q}$ then some prime p ramifies in $K \cap L$, by Theorem 7.24. A prime which ramifies in a number field also ramifies in any larger number field, so p ramifies in both K and L. But that means $\mathrm{disc}(K)$ and $\mathrm{disc}(L)$ have a common prime factor, which is a contradiction. ■

Corollary 7.27. *For monic separable $f(T) \in \mathbf{Z}[T]$ with $\deg f > 1$ other than $(T+a)(T+a+1)$, $|\operatorname{disc} f| > 1$. In particular, $|\operatorname{disc} f| > 1$ if $f(T)$ is irreducible with degree greater than* 1*.*

Notice the wholly elementary nature of this corollary: it is a statement about polynomials,[6] not number fields.

Proof. First suppose $f(T)$ splits over $\mathbf{Q}$ with roots $r_1, \ldots, r_n$. Then the r_i's are distinct integers and $\operatorname{disc} f = \prod_{i<j}(r_j - r_i)^2$. If $|r_j - r_i| > 1$ at least once then $\operatorname{disc} f$ has a prime factor. So $|\operatorname{disc} f| > 1$ unless $\deg f = 2$ and the roots of $f(T)$ differ by 1, which is the exceptional case $(T+a)(T+a+1)$ for some a.

We may now assume $f(T)$ has an irreducible factor with degree at least 2. If $g(T) \mid f(T)$ in $\mathbf{Z}[T]$ then $\operatorname{disc} g \mid \operatorname{disc} f$ in $\mathbf{Z}$ (Exercise 3.17), so we may assume $f(T)$ is irreducible with degree at least 2.

Let $f(\alpha) = 0$ and set $K = \mathbf{Q}(\alpha)$. Then $[K : \mathbf{Q}] = \deg f \geqslant 2$, so $|\operatorname{disc}(K)| > 1$. The connection between polynomial discriminants (product of squared differences of roots) and ring discriminants (determinant of trace pairing matrix) is that $\operatorname{disc} f(T) = \operatorname{disc}(\mathbf{Z}[\alpha]) = [\mathcal{O}_K : \mathbf{Z}[\alpha]]^2 \operatorname{disc}(K)$, so $|\operatorname{disc} f(T)| > 1$. ■

Remark 7.28. Why would Kronecker conjecture that every number field other than $\mathbf{Q}$ ramifies at some prime? This is a natural proposal from the analogy between number fields and function fields over $\mathbf{C}$, interpreting the latter as function fields of compact Riemann surfaces. The bottom field $\mathbf{Q}$ in algebraic number theory is analogous to the "bottom" compact Riemann surface $\mathbf{P}^1(\mathbf{C})$ (the only one of genus 0). If X is a compact Riemann surface that is not isomorphic to $\mathbf{P}^1(\mathbf{C})$, any nonconstant holomorphic map $X \to \mathbf{P}^1(\mathbf{C})$ is ramified at some point (by the Riemann–Hurwitz formula), and this has the flavor of Theorem 7.24.

7.2.5 Application: Hermite's Theorem

The degree is a crude measure of the size of a number field, since there are infinitely many number fields in $\overline{\mathbf{Q}}$ (distinct fields or fields up to isomorphism) with each degree. The next theorem shows the discriminant is a better parameter for counting how big a number field is.

[6] It's important the coefficients are *integers*: $T^3 - T + \frac{1}{3}$ has discriminant 1. Curiously, the only rational a and b such that $\operatorname{disc}(T^3 + aT + b) = 1$ are $(a, b) = (-1, \pm\frac{1}{3})$, and that is equivalent to Fermat's last theorem for exponent 3 [45, pp. 32–33].

Theorem 7.29 (Hermite). *For $B > 0$, $\#\{K \subset \overline{\mathbf{Q}} : |\mathrm{disc}(K)| \leqslant B\}$ is finite. In other words, there are finitely many number fields $K \subset \overline{\mathbf{Q}}$ with a given bound on $|\mathrm{disc}(K)|$.*

Proof. Since $|\mathrm{disc}(K)|$ grows with $[K : \mathbf{Q}]$ by the asymptotic lower bound in (7.9), bounding $|\mathrm{disc}(K)|$ from above also bounds $[K : \mathbf{Q}]$ from above. Moreover, when we specify $[K : \mathbf{Q}]$ there are finitely many choices for r_1 and r_2. Therefore it suffices to show there are finitely many K with given values for $\mathrm{disc}(K)$, $[K : \mathbf{Q}]$, r_1, and r_2. The plan of the proof is to use this data to find an $\alpha \in \mathcal{O}_K$ such that $K = \mathbf{Q}(\alpha)$ and $|\sigma(\alpha)|$ is bounded above in terms of $|\,\mathrm{disc}(K)|$ as σ runs over all real and complex embeddings of K.

First we consider the case that $r_1 > 0$ (K has a real embedding). For $t > 0$, set

$$\begin{aligned} X_t &= \{\mathbf{v} \in \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} : |x_1| < t, \text{ all other coordinates are less than } 1\} \\ &= (-t,t) \times (-1,1)^{r_1-1} \times \{z : |z| < 1\}^{r_2}. \end{aligned}$$

The region X_t is a product of intervals and discs. It is convex and centrally symmetric, with volume

$$\mathrm{vol}(X_t) = 2t \cdot 2^{r_1-1}\pi^{r_2} = 2^{r_1}\pi^{r_2}t.$$

For which t does X_t meet $\theta_K(\mathcal{O}_K)$? It does if we can apply Minkowski's convex body theorem:

$$\begin{aligned} 2^{r_1}\pi^{r_2}t &\overset{?}{>} 2^n \,\mathrm{vol}(\theta_K(\mathcal{O}_K)) \\ &= 2^n \frac{1}{2^{r_2}}\sqrt{|\mathrm{disc}(K)|} \\ &= 2^{r_1+r_2}\sqrt{|\mathrm{disc}(K)|}, \end{aligned}$$

so

$$2^{r_1}\pi^{r_2}t > 2^n \,\mathrm{vol}(\theta_K(\mathcal{O}_K)) \iff t > \left(\frac{2}{\pi}\right)^{r_2}\sqrt{|\mathrm{disc}(K)|}.$$

We could use $t = \sqrt{|\mathrm{disc}(K)|} + 1$. For such t, there is a nonzero α in $\mathcal{O}_K$ such that $\theta_K(\alpha) \in X_t$, so $|\sigma_1(\alpha)| \leqslant t$ and $|\sigma(\alpha)| \leqslant 1$ for all $\sigma \neq \sigma_1$. Then

$$1 \leqslant \left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right| = \prod_{\sigma} |\sigma(\alpha)|.$$

This implies $|\sigma_1(\alpha)| \geqslant 1$, since otherwise the right side is less than 1. So $\sigma_1(\alpha) \neq \sigma(\alpha)$ for all $\sigma \neq \sigma_1$. Thus $K = \mathbf{Q}(\alpha)$. (This is basic field theory: an element of K has its $\mathbf{Q}$-conjugates $\sigma(\alpha)$ in $\mathbf{C}$ repeated $[K : \mathbf{Q}(\alpha)]$ times as σ runs over all embeddings of K into $\mathbf{C}$.) The minimal polynomial of α over $\mathbf{Q}$ is $\prod_\sigma (T - \sigma(\alpha))$ and the upper bounds on all $|\sigma(\alpha)|$ provide upper bounds on the size of the coefficients of the polynomial, which are in $\mathbf{Z}$. So there are finitely many such polynomials, hence finitely many α (since α is a root of such a polynomial), and thus there are finitely many such fields $K = \mathbf{Q}(\alpha)$.

Now suppose $r_1 = 0$, *i.e.*, K is totally complex. For $t > 0$, set

$$X_t = \{\mathbf{v} \in \mathbf{C}^{r_2} : |\mathrm{Re}(z_1)| < 1,\ |\mathrm{Im}(z_1)| < t,\ |z_j| < 1,\ j \neq 1\}.$$

This is a product of a rectangle with width 2 and height $2t$ and $r_2 - 1$ open unit discs, so it is convex and centrally symmetric with

$$\mathrm{vol}(X_t) = 4\pi^{r_2-1}t.$$

It's easy to find t making $\mathrm{vol}(X_t) > 2^n \,\mathrm{vol}(\theta_K(\mathcal{O}_K))$. In fact, as in the previous case, we can use $t = \sqrt{|\mathrm{disc}(K)|} + 1$. For this t, there is a nonzero α in $\mathcal{O}_K$ such that

$$|\mathrm{Re}(\sigma_1(\alpha))| < 1,\ |\mathrm{Im}(\sigma_1(\alpha))| < t,\ |\sigma(\alpha)| < 1 \text{ for } \sigma \neq \sigma_1, \overline{\sigma}_1.$$

Since $|\sigma_1(\alpha)| = |\overline{\sigma}_1(\alpha)|$, from $1 \leqslant \left|\mathrm{N}_{K/\mathbf{Q}}(\alpha)\right|$, we must have $|\sigma_1(\alpha)| \geqslant 1$, which implies $\sigma_1(\alpha) \neq \sigma(\alpha)$ for $\sigma \neq \sigma_1, \overline{\sigma}_1$. Could $\sigma_1(\alpha) = \overline{\sigma}_1(\alpha)$? If so, then $\sigma_1(\alpha) \in \mathbf{R}$, so $|\sigma_1(\alpha)| = |\mathrm{Re}(\sigma_1(\alpha))| < 1$, which is a contradiction. Thus $\sigma_1(\alpha) \neq \sigma(\alpha)$ for all $\sigma \neq \sigma_1$, so $K = \mathbf{Q}(\alpha)$ and we can end with the same argument as in the first case. ■

The number field other than $\mathbf{Q}$ with least absolute discriminant is $\mathbf{Q}(\sqrt{-3})$. The next few number fields ordered by absolute discriminant are also quadratic: $\mathbf{Q}(i)$, $\mathbf{Q}(\sqrt{5})$, and $\mathbf{Q}(\sqrt{-7})$. The cubic field with least absolute discriminant is $\mathbf{Q}(\alpha)$, where $\alpha^3 - \alpha - 1 = 0$ and the discriminant is -23. The minimality of this cubic field can be proved by working through the proof of Hermite's theorem for cubic fields. See [23, pp. 620–625]. Do you think this is related to $\mathbf{Q}(\sqrt{-23})$ being the quadratic field with smallest discriminant in absolute value whose class number is 3? (That class number was computed in Section 5.3.)

The next cubic field ordered by $|\,\mathrm{disc}(K)|$ is $\mathbf{Q}(\beta)$ where $\beta^3 + \beta + 1 = 0$, with $\mathrm{disc}(K) = -31$, and the one after that is $\mathbf{Q}(\gamma)$ where $\gamma^3 - 2\gamma^2 + 2 = 0$, with

$\mathrm{disc}(K) = -44$. Tables of the first thirty cubic fields having $(r_1, r_2) = (1,1)$ and $(3,0)$, ordered by absolute discriminant, can be found in [1, pp. 376–377]. Note that when a cubic field is not Galois over $\mathbf{Q}$, there are three copies of the field in $\overline{\mathbf{Q}}$, so in $\overline{\mathbf{Q}}$ there are really three cubic fields with least absolute discriminant, but there is only one up to isomorphism.

Hermite's theorem suggests a way to make tables of number fields: order them by $|\mathrm{disc}(K)|$. At John Jones's web page http://hobbes.la.asu.edu/NFDB you can find tables of number fields with searchable constraints on the discriminant, degree, ramified primes, and other parameters. (Within reasonable bounds the tables there are complete, but there are no promises.)

A folklore conjecture is that there is a $c > 0$ such that

$$\#\{K \subset \overline{\mathbf{Q}} : |\,\mathrm{disc}(K)| \leqslant x\} \sim cx$$

as $x \to \infty$. This problem is completely wide open, but there are precise conjectures and theorems if we fix the degree: for each $n \geqslant 2$, is there is a $c_n > 0$ such that

$$\#\{K \subset \overline{\mathbf{Q}} : [K : \mathbf{Q}] = n, |\,\mathrm{disc}(K)| \leqslant x\} \sim c_n x \tag{7.10}$$

as $x \to \infty$? (In both counts, a distinction should be made between counting number fields as subfields of $\overline{\mathbf{Q}}$ and counting number fields in $\overline{\mathbf{Q}}$ up to isomorphism. Both formulations are reasonable.) The case $n = 2$ in (7.10) is classical, with $c_2 = 6/\pi^2$. Davenport and Heilbronn proved (7.10) when $n = 3$ in 1971 and about 30 years later Bhargava proved (7.10) when $n = 4$ and $n = 5$. See Bhargava's survey paper [5] for an account of work on (7.10) and a conjectured formula for c_n for all n.

We know bounding $|\mathrm{disc}(K)|$ bounds $[K : \mathbf{Q}]$, but bounding $[K : \mathbf{Q}]$ does *not* bound $|\,\mathrm{disc}(K)|$ (try quadratic fields). However, bounding the degree *and* specifying the ramified primes (the prime factors of the discriminant does bound the discriminant, because there is an upper bound on the power of a prime p as a factor of $|\,\mathrm{disc}(K)|$ which depends solely on p and $n = [K : \mathbf{Q}]$: letting $e_1, \ldots, e_g, f_1, \ldots, f_g$ be the ramification indices and residue field degrees for primes over p in K, and letting $p^{k_i} \| e_i$, the multiplicity of p in $\mathrm{disc}(K)$ is at most

$$\sum_{i=1}^{g} f_i(e_i - 1 + e_i k_i) \leqslant n(1 + \log_p n).$$

The bound on the right comes from removing -1 on the left and using the inequalities $k_i \leqslant \log_p(e_i) \leqslant \log_p(n)$. The bound on the left was conjectured by Dedekind, and its proof by Hensel (1894) was one of the first applications of p-adic numbers. (A proof can be found in [51, p. 58].)

Corollary 7.30. *For each positive integer n and finite set of primes $S = \{p_1, \ldots, p_r\}$, there are finitely many number fields in $\overline{\mathbf{Q}}$ with degree n that are ramified at the primes in S.*

Proof. If K is a number field of degree n whose ramified primes are a subset of S, then each p_i divides $\operatorname{disc}(K)$ with multiplicity at most $n(1 + \log_{p_i} n)$, so $|\operatorname{disc}(K)| \leqslant \prod_{i=1}^{r} p_i^{n(1+\log_{p_i} n)}$. Since $|\operatorname{disc}(K)|$ is bounded, there are only finitely many such K by Hermite's theorem. ■

Rather than focusing on fields where there is ramification at a particular finite set of primes S, it is more common in practice to ask that ramification not occur outside S, or equivalently that it can only occur in S but without insisting it happens everywhere in S. We are *restricting* the ramification, not *insisting* on it. The usual phrase is "unramified outside S." For example, if $S = \{2, 3\}$, the quadratic fields unramified outside S are

$$\mathbf{Q}(i),\quad \mathbf{Q}(\sqrt{2}),\quad \mathbf{Q}(\sqrt{-2}),\quad \mathbf{Q}(\sqrt{3}),\quad \mathbf{Q}(\sqrt{-3}),\quad \mathbf{Q}(\sqrt{6}),\quad \mathbf{Q}(\sqrt{-6}),$$

while the quadratic fields which are ramified over S (meaning there is ramification at each prime in S) are

$$\mathbf{Q}(\sqrt{3}),\quad \mathbf{Q}(\sqrt{6}),\quad \mathbf{Q}(\sqrt{-6}).$$

To appreciate why anyone would care about something like Corollary 7.30, here are three other theorems in the same spirit, *i.e.*, after restricting a degree or a dimension as well as a set of places where something bad happens, only finitely many objects of that type can occur. In all cases, "finitely many" means "finitely many up to isomorphism".

- For each positive integer n and finite set of points S in $\mathbf{P}^1(\mathbf{C})$, there are finitely many compact Riemann surfaces X and nonconstant holomorphic maps $f\colon X \to \mathbf{P}^1(\mathbf{C})$ with degree n that are unramified outside S. (A choice of a map $X \to \mathbf{P}^1(\mathbf{C})$ is like viewing a number field K as an extension of $\mathbf{Q}$.)

- For each positive integer n, number field K, and finite set of primes S in K, there are finitely many elliptic curves over K that have good reduction outside S. This was proved by Shafarevich, who conjectured the next theorem at the 1962 International Congress.
- For each positive integer n, number field K, and finite set of primes S in K, there are finitely many abelian varieties over K of dimension n that have good reduction outside S. This was proved by Faltings [19] in his work on the Mordell conjecture. In the case $n = 1$ it is Shafarevich's theorem stated previously.

Corollary 7.30 is often more useful than Hermite's theorem itself, so much so that some people refer to Corollary 7.30 as Hermite's theorem. There are several basic theorems in arithmetic geometry whose proofs depend on Corollary 7.30 or its variants (Exercise 7.32). For example, the proof of the Mordell–Weil theorem [25, Part C] uses Corollary 7.30 and the two other main finiteness theorems of algebraic number theory: finiteness of the class group of any number field and Dirichlet's unit theorem, to which we turn next.

7.3 Dirichlet's Unit Theorem

7.3.1 Statement and Proof

What is the structure of $\mathcal{O}_K^\times$?

Example 7.31. The unit group of $\mathbf{Z}[\sqrt{2}]$ is $\pm(1+\sqrt{2})^{\mathbf{Z}}$. This was proved in Theorem 1.29.

Example 7.32. In $\mathbf{Z}[\sqrt[3]{2}]$, $\sqrt[3]{2}-1$ is a root of $(T+1)^3-2 = T^3+3T^2+3T-1$, so $\sqrt[3]{2}-1 \in \mathbf{Z}[\sqrt[3]{2}]^\times$. Explicitly,

$$(\sqrt[3]{2}-1)(1+\sqrt[3]{2}+\sqrt[3]{4}) = 1.$$

Since $\sqrt[3]{2}-1$ is real and not ± 1, its integral powers gives us infinitely many units in $\mathbf{Z}[\sqrt[3]{2}]$. Is every unit in $\mathbf{Z}[\sqrt[3]{2}]$ a power of $\sqrt[3]{2}-1$, up to sign? We will find out in Section 7.3.2.

Theorem 7.33 (Dirichlet, 1846). *The unit group $\mathcal{O}_K^\times$ is finitely generated with r_1+r_2-1 independent[7] units of infinite order: there are $u_1, \ldots, u_{r_1+r_2-1}$ in*

[7] We say numbers $x_1, \ldots, x_r$ in a multiplicative abelian group are *independent* if there are

$\mathcal{O}_K^\times$ such that every unit can be written uniquely in the form $\zeta u_1^{m_1}\cdots u_{r_1+r_2-1}^{m_{r_1+r_2-1}}$ where ζ is a root of unity in K and the m_i's are integers. As a group,

$$\mathcal{O}_K^\times = \mu(K)u_1^{\mathbf{Z}}\cdots u_{r_1+r_2-1}^{\mathbf{Z}} \cong \mathbf{Z}/w\mathbf{Z}\oplus\mathbf{Z}^{r_1+r_2-1},$$

where $\mu(K)$ is the group of roots of unity in K and $w = \#\mu(K)$.

Theorem 7.33 is called Dirichlet's unit theorem, or just the unit theorem. We call r_1+r_2-1 the *rank* of $\mathcal{O}_K^\times$. It is the size of a basis of $\mathcal{O}_K^\times/\mu(K)$, which is a finitely generated torsion-free abelian group. Before proving Dirichlet's unit theorem, let's look at some cases with small unit rank to see what the theorem means.

If $r_1+r_2-1=0$, the unit theorem says $\mathcal{O}_K^\times$ is finite (and cyclic). This happens only for $K=\mathbf{Q}$ ($r_1=1, r_2=0$) and K an imaginary quadratic field ($r_1=0, r_2=1$).

When is $r_1+r_2-1=1$? This happens when $(r_1,r_2)=(2,0),(1,1)$, or $(0,2)$, which is precisely when K is

- real quadratic (*e.g.*, $\mathbf{Q}(\sqrt{2})$),
- cubic with one real embedding (*e.g.*, $\mathbf{Q}(\sqrt[3]{2})$),
- totally complex quartic (*e.g.*, $\mathbf{Q}(\zeta_5)$).

In all three cases, $\mathcal{O}_K^\times=\pm u^{\mathbf{Z}}$ for some unit u. In the quadratic and cubic cases, if we view K inside $\mathbf{R}$ there is a unique choice of u satisfying $u>1$.

Explicit examples of unit groups of rank 1 and 2 are listed in Table 7.6.

K	r_1	r_2	r_1+r_2-1	$\mu(K)$	$\mathcal{O}_K^\times$
$\mathbf{Q}(\sqrt{2})$	2	0	1	±1	$\pm(1+\sqrt{2})^{\mathbf{Z}}$
$\mathbf{Q}(\sqrt{5})$	2	0	1	±1	$\pm(\frac{1+\sqrt{5}}{2})^{\mathbf{Z}}$
$\mathbf{Q}(\zeta_5)$	0	2	1	$\langle-\zeta_5\rangle=\mu_{10}$	$\mu_{10}(\frac{1+\sqrt{5}}{2})^{\mathbf{Z}}$
$\mathbf{Q}(\sqrt[4]{2})$	2	1	2	±1	$\pm(1+\sqrt[4]{2})^{\mathbf{Z}}(1+\sqrt{2})^{\mathbf{Z}}$

Table 7.6: Examples of unit groups.

To prove the unit theorem, we want to view the units as a lattice in a Euclidean space. Let $n=[K:\mathbf{Q}]$, $r=r_1+r_2-1$, and $V=\mathbf{R}^{r_1}\times\mathbf{C}^{r_2}$ (a

no multiplicative relations $x_1^{m_1}\cdots x_r^{m_r}=1$ with $m_i\in\mathbf{Z}$ except the one where each exponent m_i is 0; this means linear independence over $\mathbf{Z}$ if we view the group as a $\mathbf{Z}$-module.

commutative ring). We'll identify K with its image $\theta_K(K)$ in V. For example, $N(\alpha) = \mathrm{N}_{K/\mathbf{Q}}(\alpha)$ for any $\alpha \in K$, where $N\colon V \to \mathbf{R}$ is defined in Table 7.2. Set

$$U = \theta_K(\mathcal{O}_K^\times) \subset V^\times = (\mathbf{R}^\times)^{r_1} \times (\mathbf{C}^\times)^{r_2}.$$

For $u \in U$, $N(u) = \pm 1$, so $|N(u)| = 1$. Thus $\mathcal{O}_K^\times$, in its incarnation as U, lies on the hypersurface

$$G = \{\mathbf{v} \in V : |N(\mathbf{v})| = 1\}, \tag{7.11}$$

which is not a lattice.

Example 7.34. Let $K = \mathbf{Q}(\sqrt{2})$, $\mathcal{O}_K = \mathbf{Z}[\sqrt{2}]$, and $V = \mathbf{R}^2$. Then $N(x_1, x_2) = x_1 x_2$ and $G = \{(x_1, x_2) : |x_1 x_2| = 1\}$, which is a union of two hyperbolas. See Figure 7.12, where we overlay the two hyperbolas on the Euclidean image of $\mathbf{Z}[\sqrt{2}]$ from Figure 7.4. (To keep the notation simple, we write θ for θ_K in the picture.) Points in the Euclidean image of $\mathbf{Z}[\sqrt{2}]^\times$ are marked. A picture only of the hyperbola and the units on it is in Figure 7.13. This is the picture to have in mind: U will turn out to be a discrete subset of G.

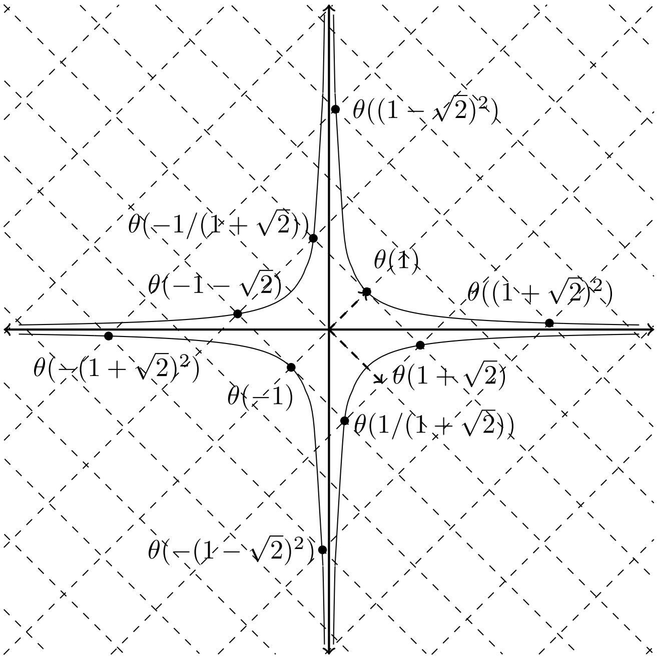


Figure 7.12: The graph of $G = \{\mathbf{v} \in \mathbf{R}^2 : |N(\mathbf{v})| = 1\}$ for $K = \mathbf{Q}(\sqrt{2})$, I.

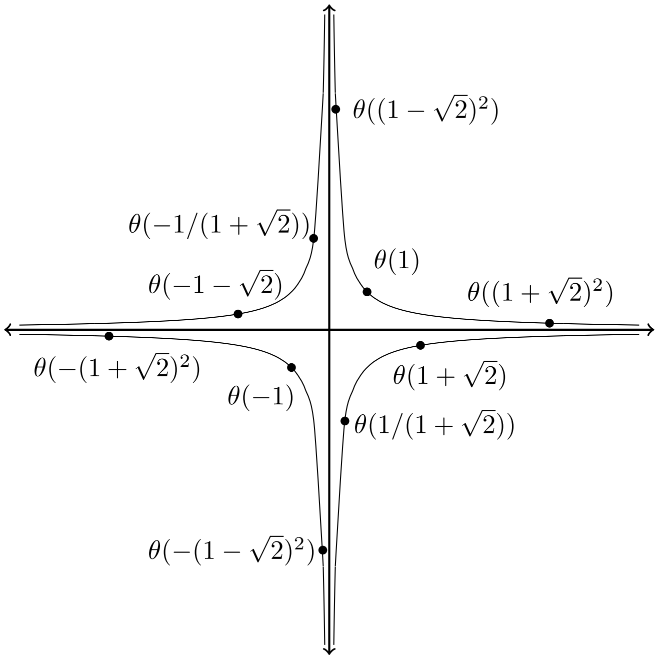


Figure 7.13: The graph of $G = \{\mathbf{v} \in \mathbf{R}^2 : |N(\mathbf{v})| = 1\}$ for $K = \mathbf{Q}(\sqrt{2})$, II.

Lattice methods can be applied to the study of units by using logarithms. Define the *log mapping* $L\colon V^\times \to \mathbf{R}^{r_1+r_2}$ by

$$L(\dots, x_i, \dots, z_j, \dots) = (\dots, \log|x_i|, \dots, 2\log|z_j|, \dots). \tag{7.12}$$

Note the absolute values inside the logarithms in (7.12): these are all logarithms of positive real numbers. The log mapping L is a continuous group homomorphism which is surjective with kernel $\{\pm 1\}^{r_1} \times (S^1)^{r_2}$, where $S^1 = \{z \in \mathbf{C} : |z| = 1\}$.

The proof of the unit theorem will be broken into three parts.

- Identify the kernel of the log mapping on the unit group U.
- Show the image $L(U)$ is a discrete[8] subgroup of $\mathbf{R}^{r_1+r_2}$ with rank at most $r_1 + r_2 - 1$.

[8] To be clear, this means $L(U)$ intersects each bounded subset of $\mathbf{R}^{r_1+r_2}$ in finitely many points. It is equivalent to saying the topology which $L(U)$ inherits from $\mathbf{R}^{r_1+r_2}$ is the discrete topology.

- Show $L(U)$ has rank $r_1 + r_2 - 1$.

The third step is the hardest one; before that step we won't even know if there are any units in K other than roots of unity.

What is the kernel of the log mapping on U? If u is a root of unity in U then $L(u) = \mathbf{0}$. The converse is also true.

Theorem 7.35 (Kronecker, 1857). *If $u \in U$ and $L(u) = \mathbf{0}$ then u is a root of unity. That is, if $u \in \mathcal{O}_K^\times$ and $|\sigma(u)| = 1$ for all real and complex embeddings σ of K, then $u^m = 1$ for some $m \geqslant 1$.*

Proof. The set of units u fitting the hypotheses of the theorem is a subgroup of $\mathcal{O}_K^\times$. We will show there are only finitely many minimal polynomials for u in $\mathbf{Z}[T]$, so this subgroup is finite and therefore u has finite order, which is another way of saying it's a root of unity in K.

All $\mathbf{Q}$-conjugates of u in $\mathbf{C}$ have the form $\sigma(u)$ where σ is some real or complex embedding of K. The minimal polynomial of u over $\mathbf{Q}$ factors over $\mathbf{C}$ as a product $\prod_{i=1}^{d}(T - u_i)$, where $d = [\mathbf{Q}(u) : \mathbf{Q}] \leqslant [K : \mathbf{Q}]$ and $|u_i| = 1$ for all i. The coefficient of T^j in the minimal polynomial is a sum of $\binom{d}{d-j}$ products of terms which each have absolute value 1, so the coefficient has absolute value at most $\binom{d}{d-j} = \binom{d}{j}$. Each coefficient is in $\mathbf{Z}$ (since u is an algebraic integer), so there are only finitely many choices for the coefficients. The degree is bounded too, so there are finitely many minimal polynomials for u in $\mathbf{Z}[T]$. ■

Remark 7.36. The hypotheses that $u \in \mathcal{O}_K^\times$ and $|\sigma(u)| = 1$ for all σ can be weakened to $u \in \mathcal{O}_K$ and $|\sigma(u)| \leqslant 1$ (with $u \neq 0$) and the conclusion of Theorem 7.35 still holds by the same proof. However, the theorem breaks down if $u \in K$ and $|\sigma(u)| \leqslant 1$: the algebraic number $\frac{3}{5} + \frac{4}{5}i$ has $\mathbf{Q}$-conjugates $\frac{3}{5} \pm \frac{4}{5}i$, which both lie on the unit circle, but $\frac{3}{5} + \frac{4}{5}i$ is not a root of unity (it's not even an algebraic integer). Some algebraic integers in $\mathbf{C}$ which are not on the unit circle can have $\mathbf{Q}$-conjugates that are on the unit circle. See Exercise 7.17.

There are only finitely many roots of unity in K, since a root of unity of order m has degree $\varphi(m)$ and $\varphi(m) \to \infty$ (somewhat irregularly) as $m \to \infty$. Therefore $\mu(K)$ is a finite subgroup of $K^\times$, which implies it is a cyclic group. (Proof #1: any real or complex embedding of K makes $\mu(K)$ a subgroup of S^1, whose finite subgroups are all cyclic. Proof #2: it is a theorem in abstract algebra that any finite subgroup of the nonzero elements of any field is a cyclic group.) Note that if K has a real embedding then $\mu(K) = \{\pm 1\}$, but even totally complex number fields can have only ±1 as the roots of unity inside them, such as any imaginary quadratic field other than $\mathbf{Q}(i)$ or $\mathbf{Q}(\sqrt{-3})$.

Our next step in proving the unit theorem is studying the image of the log mapping on U. We have $U \subset G \subset V^\times$, where the "norm 1" hypersurface G is defined in (7.11), so $L(U) \subset L(G) \subset L(V^\times) = \mathbf{R}^{r_1+r_2}$. The image $L(G)$ is the hyperplane

$$H = \left\{(y_1, \dots, y_{r_1+r_2}) \in \mathbf{R}^{r_1+r_2} : \sum_i y_i = 0\right\} \tag{7.13}$$

of dimension $r_1 + r_2 - 1$.[9] If $r_1 + r_2 = 1$ (this means K is $\mathbf{Q}$ or an imaginary quadratic field) then $H = L(U) = \{\mathbf{0}\}$, so $\mathcal{O}_K^\times = \mu(K)$ and the unit theorem is proved in this case. In fact, the condition $r_1 + r_2 = 1$ only happens when K is $\mathbf{Q}$ or an imaginary quadratic field, in which case we already knew there are finitely many units (Theorem 1.36). *From now on assume* $r_1 + r_2 > 1$.

A picture of $L(U)$ when $K = \mathbf{Q}(\sqrt{2})$ is in Figure 7.14, where the points

$$L(\theta_K(\pm(1+\sqrt{2})^n)) = (n\log(1+\sqrt{2}), n\log(\sqrt{2}-1))$$

are plotted. (Don't confuse $\mathbf{R}^2$ as the target of the log mapping for $\mathbf{Q}(\sqrt{2})$ with $\mathbf{R}^2$ as the target of the Euclidean embedding for $\mathbf{Q}(\sqrt{2})$. The Euclidean embedding of a number field lies in $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ while the log mapping for a number field lies in $\mathbf{R}^{r_1+r_2}$, and these look the same when K is totally real.)

Theorem 7.37. *The group $L(U)$ is discrete in H.*

Proof. We will show $L(U)$ is discrete in the larger space $\mathbf{R}^{r_1+r_2}$. This means $L(U)$ meets any bounded subset of $\mathbf{R}^{r_1+r_2}$ in a finite set. Suppose $b > 0$ and $L(u)$ is in the box $[-b, b]^{r_1+r_2}$. Then $|\sigma(u)| \leqslant e^b$ for real σ and $|\sigma(u)| \leqslant e^{b/2}$ for complex σ, which implies the coefficients of the minimal polynomial of u in $\mathbf{Z}[T]$ are bounded in absolute value depending only on b and $[K : \mathbf{Q}]$ (a mild extension of the argument in the proof of Theorem 7.35). There are finitely many such polynomials and thus finitely many roots. ■

Not every discrete subgroup of $\mathbf{R}^n$ is a lattice: consider $\mathbf{Z} \times \{0\}$ in $\mathbf{R}^2$.

Lemma 7.38. *A subgroup of $\mathbf{R}^n$ which is discrete is the $\mathbf{Z}$-span of an $\mathbf{R}$-linearly independent subset of $\mathbf{R}^n$. In particular, a discrete subgroup of $\mathbf{R}^n$ which contains a basis of $\mathbf{R}^n$ is a lattice.*

[9] This is the first place where it is important we used the coefficient of 2 in the components of L at complex embeddings. If we did not use 2 there then we'd have to use it in the linear relation defining H.

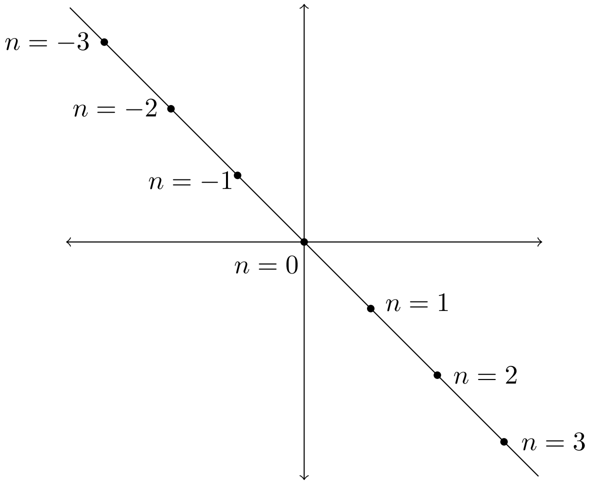


Figure 7.14: The hyperplane $x + y = 0$ and $L(\theta_K(\pm(1+\sqrt{2})^n))$, $K = \mathbf{Q}(\sqrt{2})$.

Proof. Let L be a discrete subgroup of $\mathbf{R}^n$. We may assume L is nonzero. Let $\{e_1, \ldots, e_d\}$ be the largest $\mathbf{R}$-linearly independent subset of L and set $L' = \sum_{i=1}^d \mathbf{Z}e_i$. We will show the index is $[L : L']$ is finite. For any $v \in L$, the set $\{e_1, \ldots, e_d, v\}$ is not linearly independent, so we can write $v = \sum_{i=1}^d c_i e_i$ with $c_i \in \mathbf{R}$. Split up each c_i into its integral and fractional parts, say $c_i = a_i + \varepsilon_i$ where $a_i \in \mathbf{Z}$ and $0 \leqslant \varepsilon_i < 1$. Then $v = v' + w$ where $v' = \sum_{i=1}^d a_i e_i$ and $w = \sum_{i=1}^d \varepsilon_i e_i$. Since $v' \in L'$, $w = v - v' \in L$ and $v \equiv w \bmod L'$. Thus L/L' is represented by combinations of the e_i's with coefficients between 0 and 1. Such combinations are in a bounded part of $\mathbf{R}^n$, which meets L in finitely many vectors by discreteness of L. Thus L/L' has finitely many representatives, so $[L : L'] < \infty$.

Writing $m = [L : L']$, we have $mL \subset L'$, so $L' \subset L \subset (1/m)L'$. This places L between two finite free $\mathbf{Z}$-modules of rank d, so $L \cong \mathbf{Z}^d$ too. We already know that L contains an $\mathbf{R}$-linearly independent subset of size d. All we need to show now is that any $\mathbf{Z}$-basis of L is linearly independent over $\mathbf{R}$.

We have $L = \sum_{i=1}^d \mathbf{Z}e_i$, where $\{e_1, \ldots, e_d\}$ is linearly independent over $\mathbf{R}$. Since we showed $L \cong \mathbf{Z}^d$ as an additive group, any $\mathbf{Z}$-basis of L contains d elements. Let $\{e'_1, \ldots, e'_d\}$ be any $\mathbf{Z}$-basis of L, so $L = \sum_{i=1}^d \mathbf{Z}e'_i$. Since $\{e_1, \ldots, e_d\}$ and $\{e'_1, \ldots, e'_d\}$ are both $\mathbf{Z}$-bases of L, they are linked by an integral matrix with determinant ± 1. It is left as an exercise to use this change-of-basis matrix to show that the e_i's being linearly independent over $\mathbf{R}$ implies the e_i''s are also linearly independent over $\mathbf{R}$.

When L contains a basis of $\mathbf{R}^n$, we can use $d = n$ to see that L is a lattice in $\mathbf{R}^n$. ■

Remark 7.39. Some authors call every discrete subgroup of $\mathbf{R}^n$ a lattice and what we call a lattice would be called a "full lattice." That is the convention in [6], for instance.

By Theorem 7.37 and Lemma 7.38 (applied to H as $\mathbf{R}^{r_1+r_2-1}$), $L(U)$ is the $\mathbf{Z}$-span of an $\mathbf{R}$-linearly independent subset of H, so this spanning set has size at most $r_1 + r_2 - 1$.Write

$$L(U) = \bigoplus_{i=1}^{r'} \mathbf{Z}L(u_i)$$

for some $u_1, \ldots, u_{r'}$ in U with $r' \leqslant r_1 + r_2 - 1$. From our knowledge that the kernel of L on U is the roots of unity in K,

$$\mathcal{O}_K^\times = \mu(K)u_1^{\mathbf{Z}} \cdots u_{r'}^{\mathbf{Z}}$$

and the $\mathbf{Z}$-linear independence of the $L(u_i)$'s implies the multiplicative independence of the u_i's. To complete the proof of the Dirichlet unit theorem we need to show $r' = r_1 + r_2 - 1$. At this point we don't yet even know that $r' > 0$. (The hypothetical option $r' = 0$ means $L(U) = \{\mathbf{0}\}$ and we haven't ruled that out yet if $r_1 + r_2 > 1$.) Since we know $L(U)$ is a discrete subgroup of $H \cong \mathbf{R}^{r_1+r_2-1}$, Lemma 7.38 implies that the following theorem will complete the proof of the unit theorem.

Theorem 7.40. *There is a basis of H in $L(U)$.*

Proof. For a vector $\mathbf{c} = (c_1, c_2, \ldots, c_{r_1+r_2})$ where $c_i > 0$, define

$$X_{\mathbf{c}} = \{(\ldots, x_i, \ldots, z_j, \ldots) \in \mathbf{R}^{r_1+r_2} : |x_i| < c_i, |z_j| < c_{r_1+j}\}.$$

This is a product of intervals and discs, with volume

$$2^{r_1}c_1 \ldots c_{r_1}\pi^{r_2}c_{r_1+1}^2 \cdots c_{r_1+r_2}^2 = 2^{r_1}\pi^{r_2}c_1 \ldots c_{r_1}c_{r_1+1}^2 \cdots c_{r_1+r_2}^2.$$

To apply Minkowski's convex body theorem to $X_{\mathbf{c}}$ and $\theta_K(\mathcal{O}_K)$, we need

$$2^{r_1}\pi^{r_2}c_1\dots c_{r_1}c_{r_1+1}^2\cdots c_{r_1+r_2}^2 > 2^n\operatorname{vol}(\theta_K(\mathcal{O}_K)) = 2^n\frac{1}{2^{r_2}}\sqrt{|\operatorname{disc}(K)|},$$

which is equivalent to

$$c_1\dots c_{r_1}c_{r_1+1}^2\cdots c_{r_1+r_2}^2 > \left(\frac{2}{\pi}\right)^{r_2}\sqrt{|\operatorname{disc}(K)|}.$$

We will use this kind of inequality to create an infinite sequence of numbers $\alpha_1, \alpha_2, \dots$ in $\mathcal{O}_K$ whose norms are uniformly bounded above in absolute value, which means the ideals $(\alpha_1), (\alpha_2), \dots$ have repetitions since there are finitely many ideals with norm below any given bound. If $(\alpha_s) = (\alpha_t)$ where $s < t$ then $\alpha_s = \alpha_t u$ for some unit $u \in \mathcal{O}_K^\times$. A careful method of selecting the α_s's will assure us that u is a unit of infinite order and varying the method of selection will produce $r_1 + r_2 - 1$ independent units.

Let $\sigma_1, \dots, \sigma_{r_1+r_2}$ be embeddings of K used to define the Euclidean embedding θ_K: the first r_1 are real and the remaining r_2 are complex. (There are r_2 additional complex embeddings $\overline{\sigma}_{r_1+j}$.)

Fix an embedding σ_k. There are $r_1 + r_2 - 1$ additional embeddings in the definition of θ_K. For any nonzero $\alpha \in \mathcal{O}_K$, set $c_i = |\sigma_i(\alpha)|$ for all $i \neq k$. (This is a nonempty choice since $r_1 + r_2 - 1 > 0$.) Choose c_k so that

$$\left(\frac{2}{\pi}\right)^{r_2}\sqrt{|\operatorname{disc}(K)|} < c_1\dots c_{r_1}c_{r_1+1}^2\cdots c_{r_1+r_2}^2 < \left(\frac{2}{\pi}\right)^{r_2}\sqrt{|\operatorname{disc}(K)|+1}.$$

By the first inequality there is some nonzero $\alpha' \in \mathcal{O}_K$ such that $\theta_K(\alpha') \in X_{\mathbf{c}}$, so $|\sigma_i(\alpha')| < c_i$ for all i. Therefore by the second inequality,

$$\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha')\right| < \prod_{i=1}^{r_1} c_i \prod_{j=1}^{r_2} c_{j+r_2}^2 < \left(\frac{2}{\pi}\right)^{r_2}\sqrt{|\operatorname{disc}(K)|+1}.$$

From the definition of c_i, $|\sigma_i(\alpha')| < |\sigma_i(\alpha)|$ when $i \neq k$.

This argument can be repeated with α' playing the role of α, leading to a sequence $\{\alpha_s\}_{s\geqslant 1}$ in $\mathcal{O}_K - \{0\}$ such that, for all s,

$$\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha_s)\right| < \left(\frac{2}{\pi}\right)^{r_2}\sqrt{|\operatorname{disc}(K)|+1} \quad\text{and}\quad |\sigma_i(\alpha_{s+1})| < |\sigma_i(\alpha_s)| \quad\text{when}\quad i \neq k.$$

Since the ideals (α_s) have uniformly bounded norm there is a repetition: for some $s < t$, $\alpha_s = u\alpha_t$ where $u \in \mathcal{O}_K^\times$. For $i \neq k$, $|\sigma_i(\alpha_s)| = |\sigma_i(u)|\,|\sigma_i(\alpha_t)| < |\sigma_i(u)|\,|\sigma_i(\alpha_s)|$, so $|\sigma_i(u)| > 1$. This holds for all $i \neq k$, so the vector $L(u)$ has all positive coordinates except for its kth coordinate, which is negative since all the coordinates of $L(u)$ add up to 0. In particular, u is not a root of unity so we have (finally) proved the existence of a unit of infinite order in $\mathcal{O}_K$ when $r_1 + r_2 > 1$.

This argument produced a unit after fixing an embedding σ_k of K. That unit, which we now write more carefully as u_k instead of u, satisfies $\log|\sigma_i(u_k)| > 0$ for all $i \neq k$ and $\log|\sigma_k(u_k)| < 0$. To complete the proof we will show the vectors $L(u_1), \ldots, L(u_{r_1+r_2-1})$ in H are linearly independent over $\mathbf{R}$. Consider the matrix with jth column $L(u_j)$ for $1 \leqslant j \leqslant r_1 + r_2 - 1$. Its sign pattern is

$$\begin{pmatrix} - & + & \cdots & + \\ + & - & \cdots & + \\ \vdots & \vdots & \ddots & \vdots \\ + & + & \cdots & - \\ + & + & \cdots & + \end{pmatrix}.$$

We will show the columns are linearly independent by an argument which shows that any matrix with the above sign pattern and column sums equal to 0 has linearly independent columns.

Suppose $\sum_{j=1}^{r_1+r_2-1} c_j L(u_j) = \mathbf{0}$ for some c_j's in $\mathbf{R}$ which are not all 0. Dividing through by the c_j with largest absolute value, we may assume some c_k is 1 and $|c_j| \leqslant 1$ for $j \neq k$. Looking at the kth coordinate in this linear dependence relation, $\sum_{j=1}^{r_1+r_2-1} c_j \log|\sigma_k(u_j)| = 0$. When $j \neq k$, $\log|\sigma_k(u_j)| > 0$, so $c_j \log|\sigma_k(u_j)| \leqslant \log|\sigma_k(u_j)|$. This inequality is also true when $j = k$ since $c_k = 1$. Therefore

$$0 = \sum_{j=1}^{r_1+r_2-1} c_j \log|\sigma_k(u_j)| \leqslant \sum_{j=1}^{r_1+r_2-1} \log|\sigma_k(u_j)| = -\log|\sigma_k(u_{r_1+r_2})| < 0,$$

a contradiction. ■

Here is a second proof of Theorem 7.40, which is based on [28]. It replaces the construction of a large linearly independent subset with a compactness argument.

Proof. For $G = \{\mathbf{v} \in V^\times : |N(\mathbf{v})| = 1\}$ and $U = \mathcal{O}_K^\times \subset G$, we will show G/U is compact in the quotient topology. (The topology we give G is the one it inherits as a subset of V.) As an example, in Figure 7.13 G/U has representatives on the closed arc from $\theta(1)$ to $\theta((1+\sqrt{2})^2)$, which is compact. We will actually show G/U is sequentially compact (and thus compact): every sequence has a convergent subsequence.

Let $C \subset V$ be any compact, convex, centrally symmetric region with

$$\operatorname{vol}(C) \geqslant 2^n \operatorname{vol}(\theta_K(\mathcal{O}_K)). \tag{7.14}$$

(Explicitly, we know $\operatorname{vol}(\theta_K(\mathcal{O}_K)) = \frac{1}{2^{r_2}}\sqrt{|\operatorname{disc}(K)|}$.) For example, we can take C to be a large closed ball around the origin in V.

In V we can multiply (V is a ring), and multiplication by any $\mathbf{v} \in V^\times$ is an invertible linear transformation on V with determinant $N(\mathbf{v})$. The set $\mathbf{v}C$ is convex (an invertible linear transformation on V preserves convexity) and centrally symmetric with volume $|N(\mathbf{v})|\operatorname{vol}(C)$ since a linear transformation scales volume by the absolute value of its determinant. Pick a sequence $\{g_jU\} \subset G/U$. Since $|N(g_j)| = 1$ (definition of G), g_jC has volume $\operatorname{vol}(C)$. By (7.14) and Minkowski's convex body theorem, g_jC meets $\mathcal{O}_K - \{0\}$. Write

$$g_jc_j = a_j$$

for some $c_j \in C$ and nonzero $a_j \in \mathcal{O}_K$. The numbers $|N(a_j)| = |N(c_j)|$ are bounded since N is continuous on the compact set C. Since $|N(a_j)| = |\mathrm{N}_{K/\mathbf{Q}}(a_j)| = [\mathcal{O}_K : a_j\mathcal{O}_K]$ and only finitely many ideals have index below a given bound, the list of ideals $a_j\mathcal{O}_K$ must have infinitely many repetitions, so by passing to a subsequence we may reindex and assume they are all equal: $a_1\mathcal{O}_K = a_j\mathcal{O}_K$ for all j. Thus $a_1u_j = a_j$ for some $u_j \in \mathcal{O}_K^\times$. (We are discovering units here, nonconstructively.) The c_j's lie in a compact set C, so by passing to a further subsequence we can assume the c_j's converge, say to $\ell \in V$. Since $|N(c_j)| = |N(a_j)| = |N(a_1)| > 0$, by continuity $|N(\ell)| = |N(a_1)| > 0$, which implies $\ell \in V^\times$. Since $c_j \to \ell$, we have $c_j^{-1} \to \ell^{-1}$ in $V^\times$. Passing to cosets modulo U in $V^\times$,

$$g_jU = c_j^{-1}a_jU = c_j^{-1}a_1u_jU = c_j^{-1}a_1U \to \ell^{-1}a_1U$$

and $\ell^{-1}a_1 \in G$. So G/U is (sequentially) compact.

Since $L(G) = H$, the natural map $G/U \to H/L(U)$ induced by L is onto. Since G/U is compact and L is continuous, $H/L(U)$ is compact. We know from

Theorem 7.37 that $L(U)$ is the $\mathbf{Z}$-span of some $\mathbf{R}$-linearly independent subset of H, say of size r'. Extend this set to a basis of H: $H = \bigoplus_{j=1}^{r_1+r_2-1} \mathbf{R}e_j$ and $L(U) = \bigoplus_{j=1}^{r'} \mathbf{Z}e_j$. Then $H/L(U) \cong (\mathbf{R}/\mathbf{Z})^{r'} \times \mathbf{R}^{r_1+r_2-1-r'}$. Compactness of $H/L(U)$ implies $r' = r_1 + r_2 - 1$. ■

Remark 7.41. The argument at the end of this proof leads to a topological description of lattices that does not refer directly to bases: a subgroup Λ in $\mathbf{R}^n$ is a lattice if and only if Λ is discrete and $\mathbf{R}^n/\Lambda$ is compact (in the quotient topology). Compactness of $\mathbf{R}^n/\Lambda$ is related to Λ reaching out in all directions in $\mathbf{R}^n$. The subgroup $\mathbf{Z} \times \{0\}$ in $\mathbf{R}^2$ is discrete but has no vertical aspect, so $\mathbf{R}^2/(\mathbf{Z} \times \{0\}) \cong (\mathbf{R}/\mathbf{Z}) \times \mathbf{R}$ is not compact.

Remark 7.42. Dirichlet proved his unit theorem not for $\mathcal{O}_K$, but for rings of the form $\mathbf{Z}[\alpha]$[10] and he did not have Minkowski's convex body theorem available (it was 45 years in the future). In its place he used the pigeonhole principle, whose significance in mathematics he was the first to recognize. You can see the idea of the pigeonhole principle at work, with volumes, in the proof of the convex body theorem. For a proof of the unit theorem along the lines of Dirichlet's original argument, see [30, Sect. 2.9, 2.10].

Units $u_1, \ldots, u_{r_1+r_2-1}$ as in the unit theorem, which generate the unit group modulo roots of unity and are multiplicatively independent, are called a set of *fundamental units* for $\mathcal{O}_K^\times$ (or, by abuse of language, for K).

Example 7.43. Let K be a real quadratic field, so it has unit rank 1: the unit group has the form $\pm u^{\mathbf{Z}}$ for some u. The choices for a fundamental unit of K are u, $-u$, $1/u$, and $-1/u$. These are the units which turn into a generator of $L(\mathcal{O}_K^\times) \cong \mathbf{Z}$ under the log mapping. If we view K inside $\mathbf{R}$ by a specific real embedding, there is exactly one fundamental unit which is greater than 1 and it is called *the* fundamental unit of K (it depends on the choice of real embedding). In $\mathbf{Q}(\sqrt{2})$, for instance, the fundamental unit is $1 + \sqrt{2}$.

7.3.2 Units in Cubic Fields with Rank 1

In a cubic field with one real embedding, the unit group has rank 1. As in a real quadratic field, if u is a fundamental unit then the other fundamental units are $-u$, $1/u$, and $-1/u$. If we view the field inside $\mathbf{R}$, then there is a least unit

[10] In Dirichlet's time, the distinction between rings $\mathbf{Z}[\alpha]$ and rings $\mathcal{O}_K$ was not at all clear cut; everyone worked with $\mathbf{Z}[\alpha]$.

greater than 1 and that unit is called the fundamental unit. To prove a unit is fundamental, we will use the following inequality.

Theorem 7.44 (Artin). *If K is a cubic number field with one real embedding and $v > 1$ is any unit in $\mathcal{O}_K$, then $|\mathrm{disc}(K)| < 4v^3 + 24$.*

Proof. This will be a very tedious calculation. The reader may want to read the corollary and its applications first and then return to this proof.

Since v is a unit and is not ± 1, $v \notin \mathbf{Q}$. Thus $\mathbf{Q}(v) = K$, so $\mathbf{Z}[v]$ is a $\mathbf{Z}$-lattice in K. From $\mathbf{Z}[v] \subset \mathcal{O}_K$, $|\,\mathrm{disc}(K)| \leqslant |\,\mathrm{disc}(\mathbf{Z}[v])|$. We will show $|\,\mathrm{disc}(\mathbf{Z}[v])| < 4v^3 + 24$.

Let $\sigma\colon K \to \mathbf{C}$ be one of the non-real embeddings of K. Then $\mathrm{N}_{K/\mathbf{Q}}(v) = v\sigma(v)\overline{\sigma}(v) = v|\sigma(v)|^2 > 0$, so v has norm 1. Let $x = \sqrt{v}$ (as a positive real number), so $1 = x^2|\sigma(v)|^2$. Therefore $|\sigma(v)| = 1/x$, so in polar form $\sigma(v) = x^{-1}e^{it}$ for some real number t. Then

$$\begin{aligned}
\mathrm{disc}(\mathbf{Z}[v]) &= ((\sigma(v) - v)(\overline{\sigma}(v) - v)(\sigma(v) - \overline{\sigma}(v)))^2 \\
&= ((x^{-1}e^{it} - x^2)(x^{-1}e^{-it} - x^2)(x^{-1}e^{it} - x^{-1}e^{-it}))^2 \\
&= ((x^{-2} + x^4 - 2x\cos t)(-2ix^{-1}\sin t))^2 \\
&= -4(\sin^2 t)(x^3 + x^{-3} - 2\cos t)^2,
\end{aligned}$$

so

$$\begin{aligned}
\frac{1}{4}|\,\mathrm{disc}(\mathbf{Z}[v])| &= (\sin^2 t)(x^3 + x^{-3} - 2\cos t)^2 \\
&= (1 - \cos^2 t)(x^3 + x^{-3} - 2\cos t)^2.
\end{aligned}$$

Set $a = x^3 + x^{-3}$, so $a > 2$ since $x > 0$, and set $c = \cos t$, so $c \in [-1, 1]$. Then $(1/4)\,\mathrm{disc}(\mathbf{Z}[u]) = (1 - c^2)(a - 2c)^2$. Set $f(y) = (1 - y^2)(a - 2y)^2$. What is its maximal value on $[-1, 1]$? On this interval f takes nonnegative values and vanishes at the endpoints, so we check where f' vanishes in $[-1, 1]$. By calculus, $f'(y) = 2(a - 2y)(4y^2 - ay - 2)$, where the linear factor vanishes at $a/2 > 1$ and the quadratic factor has roots $(a \pm \sqrt{a^2 + 32})/8$; since $a > 2$, the root with the $+$ sign is greater than 1 and the other root is in $(-1, 0)$. Call this root y_0. It is the only root of f' in $[-1, 1]$, so $f(y_0)$ is the maximum value of f on $[-1, 1]$. Thus

$$\frac{1}{4}|\mathrm{disc}(\mathbf{Z}[v])| = f(c) \leqslant f(y_0) = (1 - y_0^2)(a - 2y_0)^2.$$

Expanding the square and using the equation $ay_0 = 4y_0^2 - 2$ (since y_0 is a root of the quadratic factor of f'), we get

$$(1-y_0^2)(a-2y_0)^2 = a^2 - 4y_0^4 - 4y_0^2 + 4.$$

Substituting $a = x^3 + x^{-3}$,

$$(1-y_0^2)(a-2y_0)^2 = x^6 + 6 + (x^{-6} - 4y_0^4 - 4y_0^2).$$

We will show $x^{-6} < 4y_0^2$, so the right side is less than $x^6 + 6 = v^3 + 6$. Then $|\operatorname{disc}(\mathbf{Z}[v])| < 4v^3 + 24$, as desired.

Since $y_0 \in (-1, 0)$ while $x^{-1} \in (0,1)$, the inequality $x^{-6} < 4y_0^2$ is the same as $1 < 4|y_0|x^3$. To prove this inequality, let's write down formulas for y_0 and x in terms of a. Since y_0 is the smaller root of $4y^2 - ay - 2$, $y_0 = (a - \sqrt{a^2+32})/8$. Since $x^3 + x^{-3} = a$, multiplying by x^3 and using the quadratic formula shows $x^3 = (a + \sqrt{a^2-4})/2$. Therefore we want to show

$$1 < \frac{(\sqrt{a^2+32} - a)(a + \sqrt{a^2-4})}{4}.$$

Check by calculus that the expression on the right is an increasing function of a for $a \geqslant 2$. At 2 the right side is 2, so since $a = x^3 + x^{-3} > 2$ we are done. ■

Corollary 7.45. *Let K be a cubic field with $r_1 = 1$. Viewing K in $\mathbf{R}$, if $u > 1$ is a unit of $\mathcal{O}_K$ and $4u^{3/2} + 24 < |\operatorname{disc}(K)|$ then u is the fundamental unit of $\mathcal{O}_K$.*

Proof. Let ε be the fundamental unit, so $u = \varepsilon^n$ with $n \geqslant 1$. We want to show $n = 1$. If $n \geqslant 2$ then Artin's inequality with $v = \varepsilon$ says

$$|\operatorname{disc}(K)| < 4\varepsilon^3 + 24 = 4u^{3/n} + 24 \leqslant 4u^{3/2} + 24,$$

which contradicts the inequality in the statement of the corollary. ■

Example 7.46. Let $K = \mathbf{Q}(\sqrt[3]{2})$, so $\operatorname{disc}(K) = -108$. Since

$$1 = \sqrt[3]{2}^3 - 1 = (\sqrt[3]{2} - 1)(\sqrt[3]{4} + \sqrt[3]{2} + 1),$$

we have a unit $u = 1 + \sqrt[3]{2} + \sqrt[3]{4} \approx 3.847$. Since $4u^{3/2} + 24 \approx 54.185 < 108$, u is the fundamental unit of $\mathcal{O}_K$ so its reciprocal $\sqrt[3]{2} - 1$ is also a fundamental unit.

Example 7.47. Let $K = \mathbf{Q}(\alpha)$, where $\alpha^3 + 2\alpha + 1 = 0$. The polynomial $T^3 + 2T + 1$ is irreducible modulo 3, so $K/\mathbf{Q}$ is cubic. It has one real root, approximately $-.45$. Since $\operatorname{disc}(T^3 + 2T + 1) = -59$, $\mathcal{O}_K = \mathbf{Z}[\alpha]$. Clearly α is a unit. We view K in $\mathbf{R}$. Since $\alpha \approx -.45$, we get a unit greater than 1 using

$$u = -\frac{1}{\alpha} \approx 2.205.$$

Since $4u^{3/2} + 24 \approx 37.10 < 59$, u is the fundamental unit of $\mathcal{O}_K$.

Now we look at some cubic fields with unit rank 1 where Artin's inequality does not suffice to prove a unit is the fundamental unit.

Nonexample 7.48. Let $K = \mathbf{Q}(\alpha)$ where $\alpha^3 - \alpha - 1 = 0$. This cubic field has discriminant -23 and α is a unit in $\mathcal{O}_K = \mathbf{Z}[\alpha]$. The unique real root of $T^3 - T - 1$ is approximately 1.324, so it is natural to guess that α is the fundamental unit. Since $|\operatorname{disc}(K)| < 24$, we can't use Artin's inequality to test if a unit is the fundamental unit: $|\operatorname{disc}(K)| < 4v^3 + 24$ for all units $v > 1$ in $\mathbf{Z}[\alpha]$.

Since we know the unit group (modulo ± 1) is infinite cyclic, to show α is the fundamental unit we show α is the smallest unit greater than 1: no unit $u \in \mathbf{Z}[\alpha]^\times$ satisfies $1 < u < \alpha$. Let $\sigma\colon K \to \mathbf{C}$ be one of the complex embeddings of K, so $\mathrm{N}_{K/\mathbf{Q}}(u) = u\sigma(u)\overline{\sigma}(u) = u|\sigma(u)|^2 > 0$. Therefore $\mathrm{N}_{K/\mathbf{Q}}(u) = 1$. Since $u \notin \mathbf{Q}$, the minimal polynomial of u over $\mathbf{Q}$ is $T^3 + aT^2 + bT - 1$ for some integers a and b. The roots are $u, \sigma(u)$, and $\overline{\sigma}(u)$, so

$$a = -(u + \sigma(u) + \overline{\sigma}(u)), \quad b = u\sigma(u) + u\overline{\sigma}(u) + \sigma(u)\overline{\sigma}(u).$$

Then

$$|a| \leqslant u + 2|\sigma(u)|, \quad |b| \leqslant 2u|\sigma(u)| + |\sigma(u)|^2.$$

Since $1 = u|\sigma(u)|^2$, the bound $1 < u$ implies $|\sigma(u)| < 1$, so from $1 < u < \alpha$ we get

$$|a| < \alpha + 2 \approx 3.3, \quad |b| \leqslant 2\alpha + 1 \approx 3.6.$$

Thus a and b both lie in $\{0, \pm 1, \pm 2, \pm 3\}$. Among all irreducible $T^3 + aT^2 + bT - 1$ with a and b in this set, any such polynomial with a root in K must have discriminant equal to a nonzero square multiple of $\operatorname{disc}(K) = -23$ (because, with r being a root, $\operatorname{disc}(\mathbf{Z}[r]) = [\mathcal{O}_K : \mathbf{Z}[r]]^2 \operatorname{disc}(K)$). Several polynomials have this feature, including $T^3 - T - 1$ itself, but aside from $T^3 - T - 1$, the real root of such a polynomial is always larger than α. Therefore α is the fundamental unit of $\mathcal{O}_K$.

Nonexample 7.49. We will determine the fundamental unit of $K = \mathbf{Q}(\sqrt[3]{175})$. From Example 6.47, $\mathcal{O}_K = \mathbf{Z} + \mathbf{Z}\alpha + \mathbf{Z}\beta$, where $\alpha = \sqrt[3]{175} = \sqrt[3]{5^2 \cdot 7}$ and $\beta = \sqrt[3]{245} = \sqrt[3]{5 \cdot 7^2}$. The PARI command `bnfinit(x^3 - 175).fu` returns the answer `[Mod(2/5*x^2 + 48*x - 281, x^3-175)]`, which is telling us that a generator of the unit group of K, up to sign, is $v := \frac{2}{5}\alpha^2 + 48\alpha - 281 = 48\alpha + 2\beta - 281$. When we set α and β to be real numbers, v is approximately $-.000004409$, so the fundamental unit of K greater than 1 is

$$u = -\frac{1}{v} = 13516\alpha + 12082\beta + 75601 \approx 226802.996,$$

which is a root of $T^3 - 226803T^2 + 843T - 1$.

The computer told us what the fundamental unit is supposed to be, and now let's prove it is not misleading us: u is the fundamental unit of K. Let $\varepsilon > 1$ be the unknown fundamental unit, so $u = \varepsilon^n$ for some $n \geqslant 1$. Artin's inequality says $|\operatorname{disc}(K)| < 4\varepsilon^3 + 24$. By Example 6.47, $\operatorname{disc}(K) = -27 \cdot 5^2 \cdot 7^2 = 33075$, so $\varepsilon^3 > 33051/4$. Raising both sides to the nth power,

$$(13516\alpha + 12082\beta + 75601)^3 > \left(\frac{33051}{4}\right)^n,$$

so

$$n < \frac{\log((13516\alpha + 12082\beta + 75601)^3)}{\log(33051/4)} \approx 4.1.$$

The only options for n other than 1 are 2, 3, and 4. We will prove u is not a square or a cube, and that forces $n = 1$: $u = \varepsilon$.

A standard way to prove an algebraic integer in some number field is not an nth power is to show it is not an nth power mod $\mathfrak{p}$ for some prime ideal $\mathfrak{p}$.[11] We used this idea before in the proof of Theorem 4.51, and now we use it again. It is easiest to work modulo a prime with residue field degree 1 (so the residue field is $\mathbf{F}_p$). While there are some primes in $\mathcal{O}_K$ of small norm with residue field degree 1, it turns out that $u \bmod \mathfrak{p}$ is a cube for $\mathfrak{p}$ with small norm. Let's look at the prime numbers which split completely in K, whose factors provide us with three degree 1 prime ideals. For p to split completely in K means $T^3 - 175 \bmod p$ has

[11] Although rare, it is possible that an algebraic integer is not an nth power but is congruent to an nth power mod $\mathfrak{p}$ for all $\mathfrak{p}$. This phenomenon is explained by the Grunwald–Wang theorem in class field theory.

three different roots, and the first such prime is 37:

$$T^3 - 175 \equiv (T-3)(T-4)(T-30) \bmod 37.$$

Let $\mathfrak{p}$ be the prime ideal corresponding to the first factor: $\mathfrak{p} = (37, \alpha - 3)$ and $\mathcal{O}_K/\mathfrak{p} \cong \mathbf{F}_{37}$. Since $\alpha \equiv 3 \bmod \mathfrak{p}$ and $\alpha\beta = 35$, $\beta \equiv 35/3 \equiv 24 \bmod \mathfrak{p}$. In $\mathcal{O}_K/\mathfrak{p}$,

$$u = 13516\alpha + 12082\beta + 75601 \equiv 11\alpha + 20\beta + 10 \equiv 11 \cdot 3 + 20 \cdot 24 + 10 \equiv 5 \bmod \mathfrak{p}.$$

In $\mathbf{F}_{37}$, 5 is neither a square nor a cube, so we're done.

7.3.3 Units in Real Multiquadratic Fields

A real quadratic field has unit rank 1 and finding a fundamental unit in it is essentially the same as solving Pell's equation. More precisely, units in $\mathbf{Q}(\sqrt{d})$ for squarefree $d > 1$ come from integral solutions of $x^2 - dy^2 = \pm 1$ if $d \equiv 2, 3 \bmod 4$ and $x^2 - dy^2 = \pm 4$ if $d \equiv 1 \bmod 4$ (Section 1.3).

A real biquadratic field looks like $\mathbf{Q}(\sqrt{a}, \sqrt{b})$, where a and b are distinct squarefree positive integers, and it has unit rank $4 - 1 = 3$. Where can we find 3 units? There are 3 quadratic subfields ($\mathbf{Q}(\sqrt{a})$, $\mathbf{Q}(\sqrt{b})$, and $\mathbf{Q}(\sqrt{ab})$) and each has a fundamental unit. Could a fundamental unit from each quadratic subfield give us a set of fundamental units for the biquadratic field? Not always.

Example 7.50. In the field $\mathbf{Q}(\sqrt{2}, \sqrt{3})$, PARI says a system of fundamental units is $1 + \sqrt{2}$, $\sqrt{2} + \sqrt{3}$, and $\frac{\sqrt{2} - \sqrt{6}}{2}$. (The field is $\mathbf{Q}(\sqrt{2} + \sqrt{3})$ and $\sqrt{2} + \sqrt{3}$ has minimal polynomial $T^4 - 10T^2 + 1$ over $\mathbf{Q}$; in PARI, the command `bnfinit(x^4 - 10*x^2 + 1).fu` returns the list of units given above, as polynomials in $\sqrt{2} + \sqrt{3}$.)

$$\begin{tikzcd}
& \mathbf{Q}(\sqrt{2}, \sqrt{3}) \arrow[dl, no head] \arrow[d, no head] \arrow[dr, no head] & \\
\mathbf{Q}(\sqrt{2}) \arrow[dr, no head] & \mathbf{Q}(\sqrt{3}) \arrow[d, no head] & \mathbf{Q}(\sqrt{6}) \arrow[dl, no head] \\
& \mathbf{Q} &
\end{tikzcd}$$

Fundamental units for the three quadratic subfields are $1 + \sqrt{2}$, $2 + \sqrt{3}$, and $5 + 2\sqrt{6}$. The units $2 + \sqrt{3}$ and $5 + 2\sqrt{6}$ can't be part of a system of fundamental units for $\mathbf{Q}(\sqrt{2}, \sqrt{3})$ since $2 + \sqrt{3} = (\frac{\sqrt{2} + \sqrt{6}}{2})^2$ and $5 + 2\sqrt{6} = (\sqrt{2} + \sqrt{3})^2$. A unit that is a square can't belong to a set of fundamental units for the same reason that a vector in $\mathbf{Z}^n$ whose coordinates are all even can't be part of a $\mathbf{Z}$-basis. (In fact, using the log mapping on units, that is exactly why squared units are not part of a set of fundamental units.)

Example 7.51. The field $\mathbf{Q}(\sqrt{5}, \sqrt{34})$ has quadratic subfields $\mathbf{Q}(\sqrt{5})$, $\mathbf{Q}(\sqrt{34})$, and $\mathbf{Q}(\sqrt{170})$. Fundamental units for these quadratic subfields are $\frac{1+\sqrt{5}}{2} \approx 1.618$, $35 + 6\sqrt{34} \approx 69.985$, and $13 + \sqrt{170} \approx 26.038$. PARI says a system of fundamental units for $\mathbf{Q}(\sqrt{5}, \sqrt{34})$ is $.618\ldots$, $26.038\ldots$, and $-69.985\ldots$, with the second and third units being immediately recognized from our previous data. The number $.618\ldots$ is the inverse of $\frac{1+\sqrt{5}}{2}$, as anyone familiar with the Golden ratio will see right away.

More generally, consider a real multiquadratic field

$$K = \mathbf{Q}(\sqrt{d_1}, \ldots, \sqrt{d_k}),$$

where the d_i's are nonsquare positive integers which are multiplicatively independent modulo squares (that is, they are independent in $\mathbf{Q}^\times/(\mathbf{Q}^\times)^2$). By Galois theory, $\mathrm{Gal}(K/\mathbf{Q}) \cong \{\pm 1\}^k$ with automorphisms determined by the sign changes by which they affect each $\sqrt{d_i}$. Since K is totally real, $r_1 = [K : \mathbf{Q}] = 2^k$ and $r_2 = 0$. The unit rank of K is $r_1 + r_2 - 1 = 2^k - 1$, so $\mathcal{O}_K^\times \cong \{\pm 1\} \times \mathbf{Z}^{2^k - 1}$. Let's show that, as in the biquadratic case, we can find $2^k - 1$ nontrivial units from quadratic subfields.

For any nonempty subset $I = \{i_1, \ldots, i_m\}$ of $\{1, 2, \ldots, k\}$, set $d_I = d_{i_1} \cdots d_{i_m}$. Then d_I is not a perfect square, so $\mathbf{Q}(\sqrt{d_I})$ is a quadratic subfield of K. If $I \neq J$, $d_I d_J$ is not a perfect square, so we have written down $2^k - 1$ quadratic subfields. (In fact, by Galois theory these are all the quadratic subfields. Duality theory of finite abelian groups says the number of subgroups of $\{\pm 1\}^k$ with index 2 equals the number of subgroups with order 2, which is the number of elements of order 2, and every nontrivial element of $\{\pm 1\}^k$ has order 2, so there are $2^k - 1$ such subgroups.) Each $\mathbf{Q}(\sqrt{d_I})$ is real and therefore has unit rank 1. Choosing a unit other than ± 1 from each of these fields gives us $2^k - 1$ units of infinite order in K. (This choosing need not be abstract; a method of finding a nontrivial unit in a real quadratic field was described back in Section 1.3.)

Theorem 7.52. *With notation as above, let u_I be any unit in $\mathbf{Q}(\sqrt{d_I})$ other than ± 1. The $2^k - 1$ units u_I are multiplicatively independent: if $\prod_I u_I^{a_I} = 1$, where the exponents a_I are in $\mathbf{Z}$, then each a_I is 0.*

Proof. Our argument is taken from a paper of Luca and Shparlinski [36, Lemma 2]. The special feature of a unit in a real quadratic field is that its $\mathbf{Q}$-conjugate is, up to sign, its inverse. This fact will interact well with multiplicative relations.

A $\mathbf{Q}$-basis of K is all the square roots $\sqrt{d_I}$ together with 1 (we could avoid the special role for 1 by setting $d_\varnothing = 1$ and $1 = \sqrt{d_\varnothing}$). For any nonempty subset J in $\{1, 2, \ldots, k\}$, there is a $\sigma \in \mathrm{Gal}(K/\mathbf{Q})$ such that $\sigma(\sqrt{d_J}) = -\sqrt{d_J}$ and $\sigma(\sqrt{d_I}) = \sqrt{d_I}$ for all $I \neq J$. Since σ is the identity on $\mathbf{Q}(\sqrt{d_I})$ for $I \neq J$ and is nontrivial on $\mathbf{Q}(\sqrt{d_J})$, $\sigma(u_I) = u_I$ while $\sigma(u_J) = \pm u_J^{-1}$.

Applying σ to $\prod_I u_I^{a_I} = 1$ turns it into $\prod_{I \neq J} u_I^{a_I} \cdot (\pm u_J^{-1})^{a_J} = 1$. Dividing one multiplicative relation by the other, $(\pm u_J^2)^{a_J} = 1$. Since u_J has infinite order, $a_J = 0$. ■

Corollary 7.53. *The units u_I generate a subgroup of $\mathcal{O}_K^\times$ with finite index.*

Proof. By their multiplicative independence, the u_I's generate a group of rank $2^k - 1$, which is the rank of $\mathcal{O}_K^\times/\{\pm 1\}$. ■

7.3.4 Units in Cyclotomic Fields

The unit rank of a cyclotomic field $\mathbf{Q}(\zeta_m)$, for $m > 2$, is $\varphi(m)/2 - 1$. Is it possible, as in multiquadratic fields, to write down explicitly $\varphi(m)/2 - 1$ independent units?

First suppose $m = p$ is an odd prime, so $\varphi(p)/2 - 1 = (p-3)/2$. We seek $(p-3)/2$ independent units in $\mathbf{Z}[\zeta_p]$. The numbers $1 - \zeta_p^k$ for $k = 1, 2, \ldots, p-1$ all generate the same ideal $(1 - \zeta_p)$ in $\mathbf{Z}[\zeta_p]$ (it is the unique prime lying over p), so their ratios are units in $\mathbf{Z}[\zeta_p]$. Set

$$u_k = \frac{1 - \zeta_p^k}{1 - \zeta_p} = 1 + \zeta_p + \zeta_p^2 + \cdots + \zeta_p^{k-1}$$

for $k = 1, 2, \ldots, p-1$. Of course $u_1 = 1$, so it's not interesting. Moreover, $u_{p-k} = -\zeta_p^{-k} u_k$, so if we want independent units we focus on $2 \leqslant k \leqslant (p-1)/2$. The number of these units is $(p-1)/2 - 1 = (p-3)/2$. Working in $\mathbf{C}$ with $\zeta_p = e^{2\pi i/p}$, $\left|1 - \zeta_p^k\right| > |1 - \zeta_p|$ for such k (see Figure 7.15 for $p = 11$), so $|u_k| > 1$ and that implies u_k has infinite order.

Generalizing beyond primes, for any prime power $p^r > 2$ the ratios

$$\frac{1 - \zeta_{p^r}^k}{1 - \zeta_{p^r}}$$

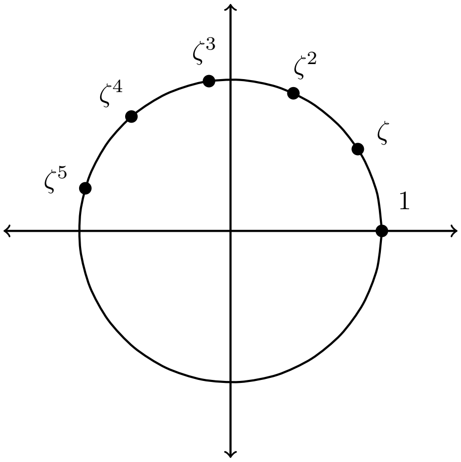


Figure 7.15: The 11th roots of unity in the upper half-plane.

are units of infinite order when $(k, p) = 1$. Taking $2 \leqslant k < \frac{1}{2}p^r$ and $(k, p) = 1$ gives us $\varphi(p^r)/2 - 1$ units. They are independent, but a proof is *not* easy, even when $r = 1$ (this special case goes back to Kummer). See [60, Theorem 8.2]. The group these units generate along with $\pm\zeta_{p^r}$ is called the group of *cyclotomic units* in $\mathbf{Q}(\zeta_{p^r})$. This group has finite index in the unit group of $\mathbf{Q}(\zeta_{p^r})$ which is closely related to the class number of $\mathbf{Q}(\zeta_{p^r} + \zeta_{p^r}^{-1})$ and is often greater than 1. In particular, the term "cyclotomic unit" means something more specific than "unit in a cyclotomic field".

If m is composite and greater than 2, it is natural to guess that we can get a set of $\varphi(m)/2 - 1$ independent units with the ratios

$$\frac{1 - \zeta_m^k}{1 - \zeta_m}$$

for $2 \leqslant k < \frac{1}{2}m$ and $(k, m) = 1$. Alas, this is generally false: such units can have nontrivial relations. For example, the unit rank of $\mathbf{Q}(\zeta_{39})$ is $\varphi(39)/2 - 1 = 11$ and there is a multiplicative relation among the above 11 units where $2 \leqslant k \leqslant 19$ with $(k, 39) = 1$:

$$v_2 v_5 v_7 v_8 v_{11} v_{19} = v_4 v_{10} v_{14} v_{16} v_{17} \zeta_{39}^{-5}, \tag{7.15}$$

where $v_k = (1 - \zeta_{39}^k)/(1 - \zeta_{39})$. (We can raise both sides of (7.15) to the 39th power to eliminate the root of unity factor on the right, but it looks better to display the relation in this simpler form with an additional root of unity floating around.) There is a simple reason why a relation like (7.15) occurs: when ζ is a root of unity whose order is *not* a prime power and is not 2 mod 4, $1-\zeta$ is a unit (Exercise 7.38) so in fact (7.15) is an equation secretly involving 12 units

$$1-\zeta_{39}^{k} \text{ for } k = 1,2,4,5,7,8,10,11,14,16,17,19. \tag{7.16}$$

The unit rank of $\mathbf{Q}(\zeta_{39})$ is 11, not 12, so it's not a surprise that something like (7.15) occurs. If we scrap the ratios and work directly with the 12 units in (7.16), can we get a group of rank 11 with them? No. These units admit two different relations:

$$\begin{aligned}(1-\zeta)(1-\zeta^4)(1-\zeta^{10})(1-\zeta^{14})(1-\zeta^{16})(1-\zeta^{17}) &= -\zeta^{-8},\\ (1-\zeta^2)(1-\zeta^5)(1-\zeta^7)(1-\zeta^8)(1-\zeta^{11})(1-\zeta^{19}) &= -\zeta^{-13},\end{aligned}$$

where $\zeta = \zeta_{39}$. (If you divide these two relations by $(1-\zeta)^6$ and fiddle with roots of unity you will recover (7.15).) The units in (7.16) therefore generate a group with rank at most 10. Numerically, PARI says the rank is 10.

An explicit finite-index subgroup of the unit group of $\mathbf{Q}(\zeta_m)$ for composite m is due to Ramachandra. See [60, Theorem 8.3].

7.3.5 The Regulator

In a number field K, let $r = r_1 + r_2 - 1$ and assume $r > 0$. If we find r units $u_1, \ldots, u_r$ in $\mathcal{O}_K^\times$, how can we check that they are multiplicatively independent?

The log mapping $L\colon V^\times \to \mathbf{R}^{r_1+r_2}$ is a homomorphism whose kernel is roots of unity, so $u_1, \ldots, u_r$ are multiplicatively independent if and only if the vectors $L(\theta_K(u_1)), \ldots L(\theta_K(u_r))$ are $\mathbf{Z}$-linearly independent. Since $L(U)$ is discrete, $\mathbf{Z}$-linear independence is the same as $\mathbf{R}$-linear independence (see the proof of Lemma 7.38). Putting these log vectors into a matrix as columns

$$\begin{pmatrix} \log|\sigma_1(u_1)| & \log|\sigma_1(u_2)| & \cdots & \log|\sigma_1(u_r)| \\ \log|\sigma_2(u_1)| & \log|\sigma_2(u_2)| & \cdots & \log|\sigma_2(u_r)| \\ \vdots & \vdots & \ddots & \vdots \\ 2\log|\sigma_{r_1+r_2}(u_1)| & 2\log|\sigma_{r_1+r_2}(u_2)| & \cdots & 2\log|\sigma_{r_1+r_2}(u_r)| \end{pmatrix}, \tag{7.17}$$

how can we decide when the columns are linearly independent over $\mathbf{R}$? (The rows corresponding to complex embeddings have 2's in them from the definition of L.)

This nonsquare matrix has one more row than columns. Because the column sums are 0, any linear dependence relation among the columns of the matrix after one row is removed is also a linear dependence relation for the original columns. Linear independence of the columns in (7.17) is therefore equivalent to linear independence of the columns in the square matrix we get after removing any one row. That means the determinants of the $r \times r$ minors of (7.17) are all zero or all nonzero. In fact these determinants are all equal up to sign. We can see this by considering the $r \times r$ matrix obtained by augmenting the above matrix with an extra column of 1's:

$$\begin{pmatrix} \log|\sigma_1(u_1)| & \log|\sigma_1(u_2)| & \cdots & \log|\sigma_1(u_r)| & 1 \\ \log|\sigma_2(u_1)| & \log|\sigma_2(u_2)| & \cdots & \log|\sigma_2(u_r)| & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 2\log|\sigma_{r_1+r_2}(u_1)| & 2\log|\sigma_{r_1+r_2}(u_2)| & \cdots & 2\log|\sigma_{r_1+r_2}(u_r)| & 1 \end{pmatrix}.$$

The determinant of this new matrix remains the same if we add all rows to one row. Since the entries in each column except the last one add up to 0, this addition makes all the entries of the chosen row equal to 0 except the last entry, which becomes $r_1 + r_2$. Computing the determinant by expansion on that row shows the determinant of this square matrix is $\pm(r_1+r_2)$ times the determinant of an $r \times r$ minor. This proves all the $r \times r$ minors in the matrix (7.17) have the same determinant up to sign.

Definition 7.54. For a set of r units $u_1, \ldots, u_r$ in K, its *regulator* $R(u_1, \ldots, u_r)$ is the absolute value of the determinant of any $r \times r$ minor of (7.17).

In abbreviated notation,

$$R(u_1, \ldots, u_r) = |\det(\delta_i \log|\sigma_i(u_j)|)_{1 \leqslant i,j \leqslant r}|,$$

where σ_i runs through all the real embeddings of K and one from each pair of complex conjugate embeddings of K, δ_i is 1 when σ_i is real and $\delta_i = 2$ when σ_i is complex, and *any one σ_i is removed* so we have as many σ_i's as we do u_j's. The regulator is unchanged by permuting the columns, so it is a function of the u_i's as an unordered set. It is nonzero if and only if the u_i's are multiplicatively independent, which means it is easy to check numerically that a set of units is independent. (By comparison, proving an independent set of units is fundamental is hard and this remains an important task in computational number theory.) If the regulator of a set of units vanishes then any integral vector killed by (7.17) provides us with the exponents for a multiplicative dependence relation (up to root of unity factors). This is precisely how the multiplicative relations among units in $\mathbf{Q}(\zeta_{39})$, such as (7.15), were found numerically.

Two sets of r independent units which generate the same group up to roots of unity have the same regulator (Exercise 7.44). Therefore we can define the *regulator of* K to be the regulator of any system of fundamental units of K when $r > 0$ and write this as $R(K)$. If $r = 0$ we define $R(K) = 1$.

Example 7.55. Let K be a real quadratic field with fundamental unit u. We have

$$R(K) = |\log|\sigma(u)||,$$

where $\sigma\colon K \to \mathbf{R}$ is either real embedding. If we identify K with its image in $\mathbf{R}$ and choose for u the fundamental unit that is greater than 1, then the absolute value signs can be dropped: $R(K) = \log u$. For instance, $R(\mathbf{Q}(\sqrt{2})) = \log(1+\sqrt{2}) \approx .881$.

Example 7.56. The regulator of $\mathbf{Q}(\sqrt[3]{2})$ is $\log(1+\sqrt[3]{2}+\sqrt[3]{4}) \approx 1.347$.

Example 7.57. A system of fundamental units for $\mathbf{Q}(\sqrt[4]{2})$ is $1+\sqrt[4]{2}$ and $1+\sqrt{2}$. There are two real embeddings and one pair of complex conjugate embeddings. Using the two real embeddings, the regulator of $\mathbf{Q}(\sqrt[4]{2})$ is

$$\left|\det\begin{pmatrix} \log|1+\sqrt[4]{2}| & \log|1+\sqrt{2}| \\ \log|1-\sqrt[4]{2}| & \log|1+\sqrt{2}| \end{pmatrix}\right|$$

while using one real embedding and a complex embedding the regulator is

$$\left|\det\begin{pmatrix} \log|1+\sqrt[4]{2}| & \log|1+\sqrt{2}| \\ 2\log|1+i\sqrt[4]{2}| & 2\log|1-\sqrt{2}| \end{pmatrix}\right|.$$

(For the purpose of embeddings into $\mathbf{R}$ or $\mathbf{C}$, treat $\sqrt{2}$ as $\sqrt[4]{2}^2$ when making computations.) These two numbers are equal and are approximately 2.158.

The regulator should be thought of as a volume for the units, much as the discriminant of K is a (squared) volume for the ring of integers. Strictly, $R(K)$ is the volume of a fundamental region for the log mapping on the units after it is projected to one of the coordinate hyperplanes in $\mathbf{R}^{r_1+r_2}$.

Example 7.58. The regulator of $\mathbf{Q}(\sqrt{2})$ is $\log(1+\sqrt{2})$. In Figure 7.16 is the image of the log mapping for $\mathbf{Q}(\sqrt{2})$, where the hyperplane H is the line

$x+y=0$. A fundamental region for the unit group of $\mathbf{Z}[\sqrt{2}]$ in this hyperplane is the segment connecting the origin to either $L(\theta_K(1+\sqrt{2}))$ or its negative. These segments have length $\sqrt{2}\log(1+\sqrt{2})$. When we project a fundamental region to either coordinate axis, the image is a segment with length $\log(1+\sqrt{2})$.

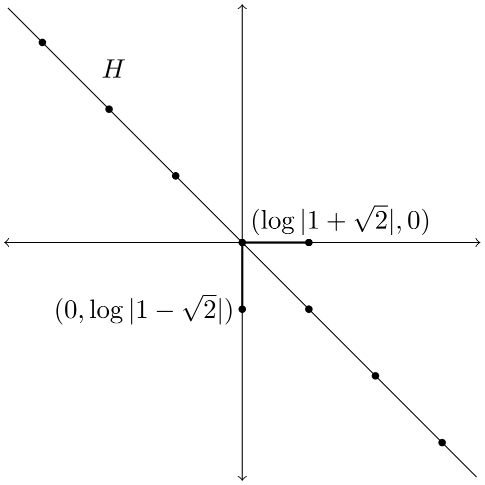


Figure 7.16: The regulator for $K = \mathbf{Q}(\sqrt{2})$ as a length.

Example 7.59. Let's write a unit in $\mathbf{Q}(\sqrt[4]{2})$ in terms of a fundamental set of units. The number

$$u = 5 - \sqrt[4]{2} - 2\sqrt{2} + 2\sqrt[4]{2}^3 \approx 4.345$$

is a unit in $\mathbf{Q}(\sqrt[4]{2})$. How can we write $u = \pm(1+\sqrt[4]{2})^a(1+\sqrt{2})^b$ for some integers a and b? Since $u > 0$, the sign on the right is positive. Applying to this equation the two real embeddings where $\sigma_1(\sqrt[4]{2}) = \sqrt[4]{2}$ and $\sigma_2(\sqrt[4]{2}) = -\sqrt[4]{2}$, and then taking absolute values and logarithms, we get two linear equations that can be combined into the matrix equation

$$\begin{pmatrix} \log|\sigma_1(u)| \\ \log|\sigma_2(u)| \end{pmatrix} = \begin{pmatrix} \log\left|1+\sqrt[4]{2}\right| & \log\left|1+\sqrt{2}\right| \\ \log\left|1-\sqrt[4]{2}\right| & \log\left|1+\sqrt{2}\right| \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}.$$

We can solve for $\binom{a}{b}$ by inverting the matrix, whose determinant is the regulator of $\mathbf{Q}(\sqrt[4]{2})$ (up to sign). Numerically, $a \approx 2.999$ and $b \approx -1.004$, so of course we expect $a = 3$ and $b = -1$, which can be verified by a direct computation.

7.4 The Zeta-Function of a Number Field

The *Riemann zeta-function* is defined for complex s with $\mathrm{Re}(s) > 1$ by the series

$$\zeta(s) = \sum_{n \geqslant 1} \frac{1}{n^s},$$

which is absolutely convergent by the integral test. This section describes (mostly without proofs) some basic properties of the Riemann zeta-function and its generalization by Dedekind to the zeta-function of a number field.

The series defining $\zeta(s)$ is called its Dirichlet series representation, since Dirichlet first introduced series of the form $\sum c_n/n^s$. The link between $\zeta(s)$ and number theory comes from a product representation over the primes:

$$\zeta(s) = \prod_p \frac{1}{1 - 1/p^s}.$$

This is called the Euler product for $\zeta(s)$. The equality of the Dirichlet series and Euler product follows from unique factorization in $\mathbf{Z}$: expand each $1/(1 - 1/p^s)$ into a geometric series and multiply them all together to get the sum of each $1/n^s$ exactly once, which is $\zeta(s)$. (Of course justification is needed for this analytic computation, as each geometric series expansion is actually valid for $\mathrm{Re}(s) > 0$ but the product of all of them is only convergent for $\mathrm{Re}(s) > 1$.)

The function $\zeta(s)$ had been used by Euler and Dirichlet only for real s. Riemann (1859) was the first person to consider $\zeta(s)$ as a function of a complex variable. It is analytic on the half-plane $\mathrm{Re}(s) > 1$ and Riemann extended it to a meromorphic function on $\mathbf{C}$ that has a simple pole at $s = 1$. His goal was to use $\zeta(s)$ to prove the prime number theorem:

$$\#\{\text{primes } p \leqslant x\} \sim \frac{x}{\log x} \text{ as } x \to \infty.$$

Riemann presented an outline for how a proof should proceed and a proof along these lines was completed in 1896 independently by Hadamard and de la Vallée-Poussin. Wiener (1932) showed that the prime number theorem is equivalent to the function $\zeta(s)$ being nonzero on the line $\mathrm{Re}(s) = 1$, which is the edge of the region of convergence for its Dirichlet series formula $\sum_{n \geqslant 1} 1/n^s$.

Riemann discovered a symmetry in $\zeta(s)$ called the functional equation. It involves the Gamma function $\Gamma(s)$, which is initially defined as an analytic function on $\mathrm{Re}(s) > 0$ by $\Gamma(s) = \int_0^\infty x^s e^{-x}\, \mathrm{d}x/x$. Using integration by parts,

$\Gamma(s+1) = s\Gamma(s)$, and that equation can be used to extend $\Gamma(s)$ to $\mathbf{C}$ as a meromorphic function with simple poles at the integers $0, -1, -2, \dots$ (and no zeros). Setting

$$Z(s) = \pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s), \tag{7.18}$$

Riemann's functional equation for the zeta-function is

$$Z(1-s) = Z(s).$$

This equation can be expressed directly in terms of $\zeta(s)$ as

$$\zeta(1-s) = 2\cos\left(\frac{\pi s}{2}\right)(2\pi)^{-s}\zeta(s),$$

but the functional equation with $Z(s)$ is conceptually a better way to think about it. Because $\zeta(s) \neq 0$ for $\mathrm{Re}(s) > 1$ (essentially by the Euler product representation) and $\pi^{-s/2}\Gamma(s/2) \neq 0$ for $\mathrm{Re}(s) > 1$, $Z(s) \neq 0$ for $\mathrm{Re}(s) > 1$ and the functional equation for $Z(s)$ implies $Z(1-s) \neq 0$ for $\mathrm{Re}(s) > 1$, so $Z(s) \neq 0$ for $\mathrm{Re}(s) < 0$. Since $\Gamma(s/2)$ has simple poles at $s = 0, -2, -4, \dots$ and no zeros, the nonvanishing of $Z(s)$ on $\mathrm{Re}(s) < 0$ implies by (7.18) that $\zeta(s)$ has simple zeros when s is a negative even integer. These are called the trivial zeros of $\zeta(s)$. The nontrivial zeros satisfy $0 < \mathrm{Re}(s) < 1$. The *Riemann hypothesis* says that all nontrivial zeros of $\zeta(s)$ line on the line $\mathrm{Re}(s) = 1/2$. For example, the first zero on $\zeta(s)$ in the upper half-plane is approximately $\frac{1}{2} + 14.1347i$. The Riemann hypothesis has been checked numerically to a very large height.

Dedekind (1871) introduced a generalization of $\zeta(s)$ for each number field K. The *Dedekind zeta-function* of K is defined on the half-plane $\mathrm{Re}(s) > 1$ by the Dirichlet series

$$\zeta_K(s) = \sum_{\mathfrak{a} \neq (0)} \frac{1}{\mathrm{N}(\mathfrak{a})^s}, \tag{7.19}$$

where $\mathfrak{a}$ runs over the nonzero ideals in $\mathcal{O}_K$. Note $\zeta_{\mathbf{Q}}(s) = \zeta(s)$. There is an Euler product for $\zeta_K(s)$ over the nonzero prime ideals in $\mathcal{O}_K$:

$$\zeta_K(s) = \prod_{\mathfrak{p}} \frac{1}{1 - 1/\mathrm{N}(\mathfrak{p})^s}, \qquad \mathrm{Re}(s) > 1. \tag{7.20}$$

This product formula is essentially a consequence of unique factorization of ideals: if we expand each $1/(1 - 1/\mathrm{N}(\mathfrak{p})^s)$ into a geometric series and multiply the terms together then we get each $1/\mathrm{N}(\mathfrak{a})^s$ exactly once.

One difference between $\zeta_K(s)$ for $K \neq \mathbf{Q}$ and $\zeta(s)$ is that there can be (and in fact really are) multiple terms in the Dirichlet series for $\zeta_K(s)$ which are equal, since several ideals can have the same norm. If we collect equal terms together then we get a Dirichlet series for $\zeta_K(s)$ over the positive integers:

$$\zeta_K(s) = \sum_{n \geqslant 1} \frac{a_n}{n^s}, \text{ where } a_n = \#\{\mathfrak{a} : \mathrm{N}(\mathfrak{a}) = n\}. \tag{7.21}$$

For example,

$$\zeta_{\mathbf{Q}(i)}(s) = 1 + \frac{1}{2^s} + \frac{1}{4^s} + \frac{2}{5^s} + \frac{1}{8^s} + \frac{1}{9^s} + \frac{2}{10^s} + \cdots + \frac{3}{25^2} + \cdots.$$

When a series $\sum_{n \geqslant 1} c_n/n^s$ converges in some right half-plane $\{a + bi : a > a_0\}$, it defines an analytic function $f(s)$ there and the coefficients c_n are uniquely determined. (For example, the limit of $f(s)$ as $s \to \infty$ along the real line is c_1 and the same limit for $2^s(f(s) - c_1)$ is c_2.) Therefore $\zeta_K(s)$, as an analytic function on $\mathrm{Re}(s) > 1$, determines the numbers $\#\{\mathfrak{a} : \mathrm{N}(\mathfrak{a}) = n\}$ for $n \geqslant 1$. Note that a Dirichlet series written as a sum over ideals of $\mathcal{O}_K$, as in the initial definition (7.19) of $\zeta_K(s)$ uses more ingredients than complex analysis alone can recover: a series $\sum_{n \geqslant 1} c_n/n^s$ determines it coefficients c_n but we can have $\sum_{\mathfrak{a}} c_{\mathfrak{a}}/\mathrm{N}(\mathfrak{a})^s = \sum_{\mathfrak{a}} c'_{\mathfrak{a}}/\mathrm{N}(\mathfrak{a})^s$ while $c_{\mathfrak{a}} \neq c'_{\mathfrak{a}}$ (we just need $\sum_{\mathrm{N}(\mathfrak{a})=n} c_{\mathfrak{a}} = \sum_{\mathrm{N}(\mathfrak{a})=n} c'_{\mathfrak{a}}$ for all $n \geqslant 1$).

the function $\zeta_K(s)$ only determines coefficients of its Dirichlet series (7.20) over the positive integers, not over ideals.

Landau (1903) extended $\zeta_K(s)$ from $\mathrm{Re}(s) > 1$ to $\mathrm{Re}(s) > 1 - 1/[K : \mathbf{Q}]$ as a meromorphic function and used analytic properties of $\zeta_K(s)$ near the line $\mathrm{Re}(s) = 1$ to prove the prime ideal theorem:

$$\#\{\mathfrak{p} : \mathrm{N}(\mathfrak{p}) \leqslant x\} \sim \frac{x}{\log x} \text{ as } x \to \infty. \tag{7.22}$$

Riemann's analytic continuation and functional equation for $\zeta(s)$ were generalized to $\zeta_K(s)$ by Hecke (1917). To express his results we use two Gamma-type functions:

$$\Gamma_{\mathbf{R}}(s) = \pi^{-s/2}\Gamma\left(\frac{s}{2}\right) \quad \text{and} \quad \Gamma_{\mathbf{C}}(s) = 2(2\pi)^{1-s}\Gamma(s).$$

One reason for the extra factor of 2 at the front of $\Gamma_{\mathbf{C}}(s)$ is that there is an analytic identity $\Gamma_{\mathbf{R}}(s)\Gamma_{\mathbf{R}}(s+1) = \Gamma_{\mathbf{C}}(s)$ with no extra constants floating around.

Theorem 7.60 (Hecke). *For any number field K, $\zeta_K(s)$ can be extended meromorphically to* $\mathbf{C}$*, with a simple pole at $s = 1$ and no other poles. There is a functional equation $Z_K(1-s) = Z_K(s)$, where*

$$Z_K(s) = |\operatorname{disc}(K)|^{s/2}\Gamma_{\mathbf{R}}(s)^{r_1}\Gamma_{\mathbf{C}}(s)^{r_2}\zeta_K(s).$$

(An elementary explanation for the subscripts $\mathbf{R}$ and $\mathbf{C}$ in $\Gamma_{\mathbf{R}}(s)$ and $\Gamma_{\mathbf{C}}(s)$ is that these functions are factors in any $Z_K(s)$ with multiplicity equal to the number of real embeddings and pairs of complex conjugate embeddings of K.)

Let's use the functional equation to determine some zeros of $\zeta_K(s)$. From the Euler product (7.20), $\zeta_K(s) \neq 0$ for $\operatorname{Re}(s) > 1$. The functions $\Gamma_{\mathbf{R}}(s)$ and $\Gamma_{\mathbf{C}}(s)$ are also nonvanishing for $\operatorname{Re}(s) > 1$, so $Z_K(s) \neq 0$ for $\operatorname{Re}(s) > 1$, and thus also $Z_K(s) \neq 0$ for $\operatorname{Re}(s) < 0$ by the functional equation. Therefore the zeros of $\zeta_K(s)$ in the half-plane $\operatorname{Re}(s) < 0$ are the points where $\Gamma_{\mathbf{R}}(s)^{r_1}\Gamma_{\mathbf{C}}(s)^{r_2}$ has poles. Since $\Gamma_{\mathbf{R}}(s)$ involves $\Gamma(s/2)$ while $\Gamma_{\mathbf{C}}(s)$ involves $\Gamma(s)$, $\Gamma_{\mathbf{R}}(s)$ has poles at $s = 0, -2, -4, \dots$ and $\Gamma_{\mathbf{C}}(s)$ has poles at $s = 0, -1, -2, \dots$, the only place in the half-plane $\operatorname{Re}(s) < 0$ where $\zeta_K(s)$ can vanish is at the negative integers.

Corollary 7.61. *The order of vanishing of $\zeta_K(s)$ at negative integers is as follows: for $k \geqslant 1$,*

$$\operatorname{ord}_{s=-2k}\zeta_K(s) = r_1 + r_2, \quad \operatorname{ord}_{s=1-2k}\zeta_K(s) = r_2.$$

Moreover, $\operatorname{ord}_{s=0}\zeta_K(s) = r_1 + r_2 - 1$.

Proof. At negative integers $Z_K(s) \neq 0$, so the order of vanishing of $\zeta_K(s)$ at a negative integer is the order of the pole of $\Gamma_{\mathbf{R}}(s)^{r_1}\Gamma_{\mathbf{C}}(s)^{r_2}$ at that number. The functions $\Gamma_{\mathbf{R}}(s)$ and $\Gamma_{\mathbf{C}}(s)$ have simple poles, which makes it straightforward to compute the order of the pole for $\Gamma_{\mathbf{R}}(s)^{r_1}\Gamma_{\mathbf{C}}(s)^{r_2}$ at each negative integer.

Since r_1 or r_2 is positive, $r_1 + r_2 > 0$, so $\zeta_K(s)$ vanishes at negative even integers. It vanishes at negative odd integers unless K is totally real ($r_2 = 0$).

To find the order of vanishing of $\zeta_K(s)$ at $s = 0$, rewrite the functional equation $Z_K(s) = Z_K(1-s)$ as

$$\zeta_K(s) = |\operatorname{disc}(K)|^{1/2-s}\frac{\Gamma_R(1-s)^{r_1}\Gamma_{\mathbf{C}}(1-s)^{r_2}}{\Gamma_{\mathbf{R}}(s)^{r_1}\Gamma_{\mathbf{C}}(s)^{r_2}}\zeta_K(1-s). \tag{7.23}$$

At $s = 0$, $\Gamma(s)$ has a simple pole so $\Gamma_{\mathbf{R}}(s)$ and $\Gamma_{\mathbf{C}}(s)$ do as well. Since $\Gamma_{\mathbf{R}}(1) =$

$\pi^{-1/2}\Gamma(1/2) = 1$ (it is a classical analytic fact that $\Gamma(1/2) = \sqrt{\pi}$) and $\Gamma_{\mathbf{C}}(1) = 2$, the Γ-ratio on the right side of (7.23) has order of vanishing $r_1 + r_2$ at $s = 0$. The function $\zeta_K(1-s)$ has a simple pole at $s = 0$ from the pole of $\zeta_K(s)$ at $s = 1$. Thus $\zeta_K(s)$ has order of vanishing $r_1 + r_2 - 1$ at $s = 0$. ■

Let's take stock of where $\zeta_K(s)$ vanishes at integers $\leqslant 0$. We have $\zeta_K(0) = 0$ unless K is $\mathbf{Q}$ or an imaginary quadratic field. If $r_2 > 0$, *i.e.*, K is not totally real, then $\zeta_K(s)$ vanishes at all negative integers. If K is totally real then $\zeta_K(s)$ vanishes at the negative even integers but is nonzero at the negative odd integers. (By a theorem of Klingen and Siegel [29], [55], for totally real K the values of $\zeta_K(s)$ at negative odd integers are rational numbers, hence of natural arithmetic interest. As a special case, taking $K = \mathbf{Q}$, the values of the Riemann zeta-function at negative odd integers are essentially Bernoulli numbers.)

The integers $\leqslant 0$ at which we have shown $\zeta_K(s)$ vanishes are called trivial zeros of $\zeta_K(s)$. All nontrivial zeros of $\zeta_K(s)$ satisfy $0 < \mathrm{Re}(s) < 1$ and the *generalized Riemann hypothesis* says all such zeros lie on the line $\mathrm{Re}(s) = 1/2$. For example, the first nontrivial zero of $\zeta_{\mathbf{Q}(i)}(s)$ in the upper half-plane is approximately $\frac{1}{2} + 6.0209i$. There have been numerical checks of the generalized Riemann hypothesis for several number fields $K \neq \mathbf{Q}$, but they are not nearly as extensive as for $\zeta(s)$ itself.

Many theorems in number theory begin "Assuming the generalized Riemann hypothesis,...". A striking example is Weinberger's theorem [61] about rings of algebraic integers which are Euclidean domains (that is, a domain admitting a division algorithm). Any Euclidean domain is a PID and it is natural to ask if there are any PIDs which are not Euclidean. Motzkin (1949) showed $\mathbf{Z}[\frac{1+\sqrt{-19}}{2}]$ is not[12] Euclidean, while this ring is a PID since it is the ring of integers of $\mathbf{Q}(\sqrt{-19})$ and $h(\mathbf{Q}(\sqrt{-19})) = 1$ by Example 7.17. There are three other imaginary quadratic fields whose ring of integers is a PID but is not Euclidean: $\mathbf{Q}(\sqrt{-43})$, $\mathbf{Q}(\sqrt{-67})$, and $\mathbf{Q}(\sqrt{-163})$. Incredibly, these are probably the only such examples among all rings of algebraic integers.

Theorem 7.62 (Weinberger, 1973). *Assume the generalized Riemann hypothesis is true. If K is not $\mathbf{Q}$ or an imaginary quadratic field and $\mathcal{O}_K$ is a* PID *then $\mathcal{O}_K$ is Euclidean.*

[12] Many classical examples of rings $\mathcal{O}_K$ which are Euclidean have $|\mathrm{N}_{K/\mathbf{Q}}(\alpha)|$ as the Euclidean function. Such $\mathcal{O}_K$ are called norm-Euclidean. To say $\mathcal{O}_K$ is not Euclidean is stronger than saying it is not norm-Euclidean. For example, the integers of $\mathbf{Q}(\sqrt{69})$ is Euclidean but not norm Euclidean [10].

What is special about K not being $\mathbf{Q}$ or an imaginary quadratic field? Such K have units of infinite order by Dirichlet's unit theorem, and this is used in Weinberger's proof.

Whenever a theorem is proved with the generalized Riemann hypothesis, it is natural to seek techniques that will prove the theorem unconditionally. In the case of Weinberger's theorem, Harper and Murty [22] proved unconditionally that if $\mathcal{O}_K$ is a PID then $\mathcal{O}_K$ is Euclidean when K is a Galois extension of $\mathbf{Q}$ with unit rank at least 4 (this includes all Galois extensions with degree at least 10). Their argument was modified by Narkiewicz [41] to show Weinberger's theorem is true unconditionally for real quadratic fields with at most two exceptions (of course there are probably no exceptions).

The residue of $\zeta_K(s)$ at $s = 1$ is given by a famous formula of Dedekind:[13]

$$\mathrm{Res}_{s=1}\,\zeta_K(s) = \frac{2^{r_1}(2\pi)^{r_2}h(K)R(K)}{w_K\sqrt{|\mathrm{disc}(K)|}}, \tag{7.24}$$

where $h(K)$ is the class number of K, $R(K)$ is the regulator of K, and w_K is the number of roots of unity in K. The formula (7.24) was found by Dirichlet in the mid-19th century for quadratic fields, so it is called Dirichlet's class number formula or the analytic class number formula.

This formula provides an important numerical method to estimate the product $h(K)R(K)$. Comparing the two formulas

$$\lim_{s\to 1^+}(s-1)\zeta_K(s) = \frac{2^{r_1}(2\pi)^{r_2}h(K)R(K)}{w_K\sqrt{|\mathrm{disc}(K)|}}, \quad \lim_{s\to 1^+}(s-1)\zeta(s) = 1,$$

we take the ratio and get

$$\lim_{s\to 1^+}\frac{\zeta_K(s)}{\zeta(s)} = \frac{2^{r_1}(2\pi)^{r_2}h(K)R(K)}{w_K\sqrt{|\mathrm{disc}(K)|}}.$$

On the left side, when $s > 1$ we can express the zeta ratio using Euler products:

$$\frac{\zeta_K(s)}{\zeta(s)} = \prod_p \frac{1-1/p^s}{\prod_{\mathfrak{p}|p}(1-1/\mathrm{N}(\mathfrak{p})^s)}.$$

[13] When a meromorphic function $f(s)$ has a simple pole at 1 then its residue at 1 is $\lim_{s\to 1}(s-1)f(s)$. Dedekind did not know $\zeta_K(s)$ has a meaning beyond $\mathrm{Re}(s) > 1$, so for him the residue formula for $\zeta_K(s)$ was $\lim_{s\to 1^+}(s-1)\zeta_K(s)$.

Formally setting $s = 1$, we would like to say

$$\lim_{s\to 1^+} \frac{\zeta_K(s)}{\zeta(s)} = \prod_p \frac{1 - 1/p}{\prod_{\mathfrak{p}|p}(1 - 1/\operatorname{N}(\mathfrak{p}))},$$

but it is not clear if the right side is meaningful since the Euler products for $\zeta(s)$ and $\zeta_K(s)$ separately diverge at $s = 1$. Nevertheless, the ratio of these Euler products does make sense at $s = 1$. A classical theorem of Mertens says that

$$\prod_{p\leqslant x} \frac{1}{1 - 1/p} \sim e^\gamma \log x \tag{7.25}$$

as $x \to \infty$, where $\gamma = .577\ldots$ is Euler's constant. (Here $\sim$ means the ratio of the two sides tends to 1.) Rosen [47] established an analogue for number fields:

$$\prod_{p\leqslant x} \prod_{\mathfrak{p}|p} \frac{1}{1 - 1/\operatorname{N}(\mathfrak{p})} \sim \rho_K e^\gamma \log x \tag{7.26}$$

as $x \to \infty$, where $\rho_K = \operatorname{Res}_{s=1} \zeta_K(s)$. Taking the ratio of (7.26) divided by (7.25), we obtain

$$\lim_{x\to\infty} \prod_{p\leqslant x} \frac{1 - 1/p}{\prod_{\mathfrak{p}|p}(1 - 1/\operatorname{N}(\mathfrak{p}))} = \rho_K = \frac{2^{r_1}(2\pi)^{r_2} h(K)R(K)}{w_K\sqrt{|\operatorname{disc}(K)|}}.$$

Therefore

$$h(K)R(K) = \lim_{x\to\infty} \frac{w_K\sqrt{|\operatorname{disc}(K)|}}{2^{r_1}(2\pi)^{r_2}} \prod_{p\leqslant x} \frac{1 - 1/p}{\prod_{\mathfrak{p}|p}(1 - 1/\operatorname{N}(\mathfrak{p}))}. \tag{7.27}$$

Here is how this can be used. First we try to compute the class group and unit group separately. Find generators for the class group and a set of relations on them. Let h' denote the size of the group described by the generators and relations. If the (abelian) group described by these generators and relations is not yet the class group (*e.g.*, the class group has a generator $[\mathfrak{p}]$ with $[\mathfrak{p}]^8 = [1]$, but perhaps the true order of $[\mathfrak{p}]$ is 1, 2, or 4), then the class group is a further quotient group, which means $h(K)$ is a factor of h', say $h' = ah(K)$ for a positive integer a. Next find $r_1 + r_2 - 1$ independent units and compute their regulator R'. If the group which these units generate, along with the roots of unity in K, has index b in the full unit group then $R' = bR(K)$ (see Exercise 7.44), so $h'R' = abh(K)R(K)$. On the left side $h'R'$ has been computed by separate computations with ideal classes and units. On the right side $h(K)R(K)$ can be computed by (7.27) with moderately large x (we ignore error estimates here). The ratio $h'R'/h(K)R(K)$ is a positive integer ab telling us how far off we are in the computation of the class group and unit group.

Example 7.63. Let $K = \mathbf{Q}(\sqrt{3}, \sqrt{5})$. Its ring of integers is $\mathbf{Z}[\sqrt{3}, \frac{1+\sqrt{5}}{2}]$ (Exercise 6.6), from which one computes $\operatorname{disc}(K) = 3600$. Obviously $r_1 = 4$, $r_2 = 0$, and $w_K = 2$.

The Minkowski bound for K is 5.625. How do 2, 3, and 5 factor in $\mathcal{O}_K$? The primes 2 and 3 are both ramified in $\mathbf{Q}(\sqrt{3})$ and inert in $\mathbf{Q}(\sqrt{5})$ while 5 is inert in $\mathbf{Q}(\sqrt{3})$ and ramified in $\mathbf{Q}(\sqrt{5})$, so each of the primes has $e = 2$ and $f = 2$ in $\mathbf{Q}(\sqrt{3}, \sqrt{5})$: $(2) = \mathfrak{P}_4^2$, $(3) = \mathfrak{P}_9^2$, and $(5) = \mathfrak{P}_{25}^2$. Trivially $\mathfrak{P}_9 = (\sqrt{3})$ and $\mathfrak{P}_{25} = (\sqrt{5})$. In $\mathbf{Z}[\sqrt{3}]$ we have $(2) = (1+\sqrt{3})^2$, so in $\mathcal{O}_K$ we have $\mathfrak{P}_4 = (1+\sqrt{3})$. Thus $h(K) = 1$.

The unit rank of K is 3, and three obvious units in K are the fundamental units in the three quadratic subfields: $2 + \sqrt{3}$, $\frac{1+\sqrt{5}}{2}$, and $4 + \sqrt{15}$. Their regulator is $5.230695\ldots$. Call this R'.

Now we compute $h(K)R(K) = R(K)$ using partial Euler products, so we have to factor prime numbers in $\mathcal{O}_K$. We already know how (2), (3), and (5) factor in $\mathcal{O}_K$. Since $K = \mathbf{Q}(\sqrt{3} + \sqrt{5})$ and $\sqrt{3} + \sqrt{5}$ has minimal polynomial $T^4 - 16T^2 + 4$, whose discriminant is $2^{14} \cdot 3^2 \cdot 5^2$, we can factor any prime $p > 5$ in $\mathcal{O}_K$ by factoring $T^4 - 16T^2 + 4 \bmod p$. By (7.27)

$$R(K) = \lim_{x \to \infty} 7.5 \prod_{p \leqslant x} \frac{1 - 1/p}{\prod_{\mathfrak{p}|p}(1 - 1/\mathrm{N}(\mathfrak{p}))}. \tag{7.28}$$

When $x = 100$, 1000, and 10000, the right side of (7.28) is approximately 2.71047, 2.61479, and 2.62870. The corresponding approximations to $R'/R(K)$ are 1.9298, 2.0004, and 1.9898. This ratio should be an integer, so it ought to be 2. That suggests the group generated by the units $2 + \sqrt{3}$, $\frac{1+\sqrt{5}}{2}$, and $4 + \sqrt{15}$, along with ± 1, has index 2 in $\mathcal{O}_K^\times$. We search for a new unit. Fortunately, $\sqrt{3} + \frac{1+\sqrt{5}}{2}$ turns out to be a unit and the regulator of $2 + \sqrt{3}$, $\frac{1+\sqrt{5}}{2}$, and $\sqrt{3} + \frac{1+\sqrt{5}}{2}$ is $R'' = 2.615347\ldots$. This is close to the numerical approximations we already found for $R(K)$ using (7.28), which leads us to believe (but we have not proved) that we have found a system of fundamental units for K.

Another application of $\zeta_K(s)$ is lower bounds on $|\operatorname{disc}(K)|$. Minkowski gave a lower bound on $|\operatorname{disc}(K)|$ in (7.8) using his convex body theorem. Because the completed zeta-function $Z_K(s) = |\operatorname{disc}(K)|^{s/2}\Gamma_{\mathbf{R}}(s)^{r_1}\Gamma_{\mathbf{C}}(s)^{r_2}\zeta_K(s)$ involves a power of $|\operatorname{disc}(K)|$, Stark introduced analytic techniques to find lower bounds on $\log|\operatorname{disc}(K)|$ in terms of the behavior of $Z_K'(s)/Z_K(s)$. (The factor $|\operatorname{disc}(K)|^{s/2}$, after logarithmic differentiation, becomes $\frac{1}{2}\log|\operatorname{disc}(K)|$.) Odlyzko improved Stark's bounds in a series of papers, but before describing one of Odlyzko's bounds let's first get a lower bound on $\log|\operatorname{disc}(K)|$ using Minkowski's inequality (7.8) so we can make a comparison.

Theorem 7.64. *For any number field K with degree $n \geqslant 2$, $\log|\operatorname{disc}(K)| > 1.39n - 1.93$.*

Proof. Taking the logarithm of both sides of (7.8),

$$\log|\operatorname{disc}(K)| \geqslant \left(\log\frac{\pi}{4}\right)n + 2(n\log n - \log(n!)).$$

From estimates related to Stirling's formula, $\log(n!) < n\log n - n + \frac{1}{2}\log(2\pi n) + \frac{1}{12n}$, so

$$\log|\operatorname{disc}(K)| > \left(2 + \log\frac{\pi}{4}\right)n - \left(\log 2\pi + \log n + \frac{1}{6n}\right).$$

The function $(\log x)/x$ has maximum value $1/e$, so $\log n < \frac{1}{e}n$. Therefore

$$\log|\operatorname{disc}(K)| > \left(2 + \log\frac{\pi}{4} - \frac{1}{e}\right)n - \left(\log 2\pi + \frac{1}{12}\right) > 1.39n - 1.93.$$

■

Odlyzko's lower bound on $\log|\operatorname{disc}(K)|$ uses an auxiliary parameter $\sigma > 1$, which can be adjusted to make the lower bound as large as possible. We will be content to give just a special case.

Theorem 7.65 (Odlyzko). *If K is a totally complex number field with degree n, $\log|\operatorname{disc}(K)| > 2.17n - 4.95$.*

Proof. See [42] or [60, Sect. 11.4] using $\sigma = 2.8$. ■

Example 7.66. Let $F = \mathbf{Q}(\sqrt[3]{2}, \zeta_3)$, so $[F : \mathbf{Q}] = 6$. We will show every proper extension of F is ramified at some prime in F. If E is an unramified extension of F and $n = [E : \mathbf{Q}]$, then $|\operatorname{disc}(E)|^{1/n} = |\operatorname{disc}(F)|^{1/6}$ (see the proof of Theorem **??**), so $\log|\operatorname{disc}(E)| = \frac{n}{6}\log|\operatorname{disc}(F)|$. The discriminant of F turns out to be -34992 and $\log(34992) \approx 10.4628$.

Since E is totally complex (F has no real embedding, so the larger field E has none either), $\log|\operatorname{disc}(E)| > 2.17n - 4.95$. Therefore

$$\frac{n}{6} \cdot 10.463 > 2.17n - 4.95 \Longrightarrow 11.62 > n.$$

Our lower bound on discriminants has become an upper bound on the degree over $\mathbf{Q}$ of an unramified extension of F. Since E contains F, n is a multiple of 6. The upper bound implies $n = 6$, so $E = F$.

If we tried to use Theorem 7.64 instead of Theorem 7.65 then we get $\frac{n}{6} \cdot 10.463 > 1/39 - 1/93$, so $.35n > -1.93$, which doesn't imply any constraint at all.

The zeta-function of K is one of the basic tools used in analytic number theory to prove theorems about K, so it is natural to ask how much $\zeta_K(s)$ knows about K.

Theorem 7.67. *For a number field K, its zeta-function $\zeta_K(s)$ determines the following information about K:*

- *r_1 and r_2,*
- *$[K : \mathbf{Q}]$,*
- *$\operatorname{disc}(K)$,*
- *$h(K)R(K)/w_K$,*
- *in each factorization $(p) = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_g^{e_g}$ the numbers g and the residue field degrees $f_i = f(\mathfrak{p}_i|p)$,*
- *which primes split completely in K.*

Proof. By Theorem 7.61, the order of vanishing of $\zeta_K(s)$ at $s = -1$ and $s = -2$ tells us r_1 and $r_1 + r_2$, so r_1, r_2, and $[K : \mathbf{Q}] = r_1 + 2r_2$ are determined by $\zeta_K(s)$.

Rewriting the functional equation (7.23) as

$$|\operatorname{disc}(K)|^{s-1/2} = \frac{\Gamma_R(1-s)^{r_1}\Gamma_{\mathbf{C}}(1-s)^{r_2}\zeta_K(1-s)}{\Gamma_{\mathbf{R}}(s)^{r_1}\Gamma_{\mathbf{C}}(s)^{r_2}\zeta_K(s)}$$

show the function $\zeta_K(s)$, which determines r_1 and r_2, also determines $|\operatorname{disc}(K)|$. The sign of $\operatorname{disc}(K)$ is $(-1)^{r_2}$, and $\zeta_K(s)$ knows r_2, so $\zeta_K(s)$ determines $\operatorname{disc}(K)$. From the analytic class number formula, $\zeta_K(s)$ determines $h(K)R(K)/w_K$.

Write $\zeta_K(s) = \sum_{n \geqslant 1} a_n/n^s$, with $a_n = \#\{\mathfrak{a} : \mathrm{N}(\mathfrak{a}) = n\}$. Since the norm of a prime ideal is a power of the prime number it lies over, expanding the Euler product for $\zeta_K(s)$ into a Dirichlet series shows the series $\sum_{k \geqslant 0} a_{p^k}/p^{ks}$ for each prime p equals the product of the Euler factors at prime ideals over p:

$$\sum_{k \geqslant 0} \frac{a_{p^k}}{p^{ks}} = \prod_{\mathfrak{p}|p} \frac{1}{1 - 1/\mathrm{N}(\mathfrak{p})^s}.$$

Let $p\mathcal{O}_K = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_g^{e_g}$. Then

$$\sum_{k \geqslant 0} \frac{a_{p^k}}{p^{ks}} = \prod_{i=1}^{g} \frac{1}{1 - 1/p^{f_i s}}. \tag{7.29}$$

Since any Dirichlet series $\sum_{n \geqslant 1} c_n/n^s$ that converges in a right half-plane uniquely determines its coefficients c_n, $\zeta_K(s)$ knows the subseries (7.29) for each p. Setting $z = 1/p^s$ in (7.29) shows $\zeta_K(s)$ analytically determines $\prod_{i=1}^g (1 - z^{f_i})$ for $|z| < 1/p$ (corresponding to $\mathrm{Re}(s) > 1$), which is enough to determine the rational function $\prod_{i=1}^g (1 - z^{f_i})$ for all complex z (a rational function is determined by its values at infinitely many points). Since $\prod_{i=1}^g (1 - z^{f_i})$ has 1 as a root with multiplicity g, we see that the number of prime ideal factors of $p\mathcal{O}_K$ is determined by $\zeta_K(s)$. In particular, p splits completely in K if and only if $g = [K : \mathbf{Q}]$, so $\zeta_K(s)$ determines the prime numbers which split completely in K. (Alternate explanation: p splits completely if and only if $a_p = [K : \mathbf{Q}]$.)

It is left to the reader (Exercise 7.45) to show $f_1, \ldots, f_g$ are determined by the rational function $\prod_{i=1}^g (1 - z^{f_i})$, and thus by $\zeta_K(s)$. ■

7.5 Exercises

1. The ring $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ is a finite-dimensional $\mathbf{R}$-vector space, so it has trace and norm maps to $\mathbf{R}$ as defined at the start of Section 3.7.

   On $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$, show

$$\mathrm{Tr}_{(\mathbf{R}^{r_1} \times \mathbf{C}^{r_2})/\mathbf{R}}((x_1, \ldots, x_{r_1}, z_1, \ldots, z_{r_2})) = \sum_{i=1}^{r_1} x_i + \sum_{j=1}^{r_2} 2\,\mathrm{Re}(z_j) \tag{7.30}$$

and

$$\mathrm{N}_{(\mathbf{R}^{r_1}\times\mathbf{C}^{r_2})/\mathbf{R}}((x_1,\dots,x_{r_1},z_1,\dots,z_{r_2})) = \prod_{i=1}^{r_1} x_i \prod_{j=1}^{r_2} |z_j|^2 \tag{7.31}$$

and the trace pairing is given by

$$\mathrm{Tr}_{(\mathbf{R}^{r_1}\times\mathbf{C}^{r_2})/\mathbf{R}}(vv') = v\cdot Jv', \tag{7.32}$$

where $v = (x_1,\dots,x_{r_1},z_1,\dots,z_{r_2})$, $v' = (x'_1,\dots,x'_{r_1},z'_1,\dots,z'_{r_2})$, and on the right side we view v and v' in $\mathbf{R}^n$ rather than $\mathbf{R}^{r_1}\times\mathbf{C}^{r_2}$ by identifying $\mathbf{C} = \mathbf{R}+\mathbf{R}i$ with $\mathbf{R}^2$ using basis $\{1,i\}$, and J is the $n\times n$ real block matrix

$$J = \begin{pmatrix} I_{r_1} & O & \cdots & O \\ O & \left(\begin{smallmatrix}2&0\\0&-2\end{smallmatrix}\right) & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & \left(\begin{smallmatrix}2&0\\0&-2\end{smallmatrix}\right) \end{pmatrix}$$

where there are r_2 copies of $\left(\begin{smallmatrix}2&0\\0&-2\end{smallmatrix}\right)$ along the diagonal after one $r_1\times r_1$ identity matrix.

The right sides of (7.30), (7.31), and (7.32) were used in Table 7.2 in order to express the trace, norm, and trace pairing on a number field in terms of functions on its Euclidean image in $\mathbf{R}^{r_1}\times\mathbf{C}^{r_2}$, without investigating whether these formulas on $\mathbf{R}^{r_1}\times\mathbf{C}^{r_2}$ have an intrinsic meaning. We now see they do.

2. Let K be a number field of degree n, with real embeddings $\sigma_1,\dots,\sigma_{r_1}$ and pairs of complex conjugate embeddings $\sigma_{r_1+1},\overline{\sigma}_{r_1+1},\dots,\sigma_{r_1+r_2},\overline{\sigma}_{r_1+r_2}$. The Euclidean embedding $\theta_K\colon K\to\mathbf{R}^{r_1}\times\mathbf{C}^{r_2}$ is given by

$$\theta_K(\alpha) = (\sigma_1(\alpha),\dots,\sigma_{r_1}(\alpha),\sigma_{r_1+1}(\alpha),\dots,\sigma_{r_1+r_2}(\alpha)).$$

a) Construct an $\mathbf{R}$-algebra isomorphism $\mathbf{R}\otimes_{\mathbf{Q}}K\cong\mathbf{R}^{r_1}\times\mathbf{C}^{r_2}$ that restricts to the Euclidean embedding on K, in the sense that $1\otimes\alpha\mapsto\theta_K(\alpha)$ for $\alpha\in K$. (Use Theorem 7.4 to verify your map is surjective.) This is how the classical method of studying number fields through real and complex embeddings can be made "coordinate-free", by embedding K in a *canonical* Euclidean space $\mathbf{R}\otimes_{\mathbf{Q}}K$ (using $\alpha\mapsto 1\otimes\alpha$) in place of the noncanonical Euclidean embedding $\theta_K\colon K\hookrightarrow\mathbf{R}^{r_1}\times\mathbf{C}^{r_2}$.

b) Without using the isomorphism in part a, show that on $\mathbf{R} \otimes_{\mathbf{Q}} K$,

$$\operatorname{Tr}_{(\mathbf{R}\otimes_{\mathbf{Q}}K)/\mathbf{R}}(x\otimes\alpha) = x\operatorname{Tr}_{K/\mathbf{Q}}(\alpha) \text{ and } \operatorname{N}_{(\mathbf{R}\otimes_{\mathbf{Q}}K)/\mathbf{R}}(x\otimes\alpha) = x^n\operatorname{N}_{K/\mathbf{Q}}(\alpha)$$

on elementary tensors and the trace pairing on $\mathbf{R} \otimes_{\mathbf{Q}} K$ is given by

$$\operatorname{Tr}_{(\mathbf{R}\otimes_{\mathbf{Q}}K)/\mathbf{R}}((x\otimes\alpha)(y\otimes\beta)) = xy\operatorname{Tr}_{K/\mathbf{Q}}(\alpha\beta)$$

on products of elementary tensors.

3. (Continuation of Exercise **??**) Describe the maximal ideals of the ring $\mathbf{R} \otimes_{\mathbf{Q}} K$ in terms of real and complex embeddings of K.

4. If we convert the trace pairing $K \times K \to \mathbf{Q}$ into a $\mathbf{Q}$-linear map $K \otimes_{\mathbf{Q}} K \to \mathbf{Q}$ and base extend that to an $\mathbf{R}$-linear map

$$\mathbf{R} \otimes_{\mathbf{Q}} (K \otimes_{\mathbf{Q}} K) \to \mathbf{R} \otimes_{\mathbf{Q}} \mathbf{Q} \cong \mathbf{R}$$

and compose this with the natural $\mathbf{R}$-linear isomorphism $\mathbf{R}\otimes_{\mathbf{Q}}(K\otimes_{\mathbf{Q}}K) \cong (\mathbf{R} \otimes_{\mathbf{Q}} K) \otimes_{\mathbf{R}} (\mathbf{R} \otimes_{\mathbf{Q}} K)$ to get an $\mathbf{R}$-linear map

$$(\mathbf{R} \otimes_{\mathbf{Q}} K) \otimes_{\mathbf{R}} (\mathbf{R} \otimes_{\mathbf{Q}} K) \to \mathbf{R}$$

and pull that back to an $\mathbf{R}$-bilinear map

$$(\mathbf{R} \otimes_{\mathbf{Q}} K) \times (\mathbf{R} \otimes_{\mathbf{Q}} K) \to \mathbf{R},$$

show this last map is the trace pairing on $\mathbf{R} \otimes_{\mathbf{Q}} K$. So the base extension by $\mathbf{R}$ of the trace pairing on K is the trace pairing on $\mathbf{R} \otimes_{\mathbf{Q}} K$.

5. a) Compare the Minkowski bound for $\mathbf{Q}(\sqrt{103})$ with the bounds found in Example 5.14.

   b) Show $\mathbf{Q}(\sqrt{103})$, $\mathbf{Q}(\sqrt{-163})$, and $\mathbf{Q}(\zeta_5)$ have class number 1. (Be careful: for prime p, the degree of $\mathbf{Q}(\zeta_p)$ over $\mathbf{Q}$ is $p-1$, not p.)

6. a) In Example 7.18 we used ideal factorizations to show $x^2 - 82y^2 = 2$ has no integral solution. Prove this in a second way by the method of Section 1.4.

   b) The equation $x^2 - 82y^2 = 2$ has rational solutions $(10/3, 1/3)$ and $(18/11, 1/11)$. Use this to show the congruence $x^2 - 82y^2 \equiv 2 \bmod m$ is solvable for every m.

7. a) Show $\mathbf{Q}(\sqrt{-51})$ has class number 2.

   b) Use part a to show the equation $y^2 = x^3 - 51$ has no integral solutions.

   c) For an odd prime p, show by induction that if $(a, p) = 1$ and $a \equiv \Box \bmod p$ then $a \equiv \Box \bmod p^r$ for all $r \geqslant 1$. If $(a, 2) = 1$ and $a \equiv 1 \bmod 8$, show $a \equiv \Box \bmod 2^r$ for all $r \geqslant 1$.

   d) The equation $y^2 = x^3 - 51$ has a rational solution: $(1375/9, 50986/27)$. Use this and part c to show $y^2 \equiv x^3 - 51 \bmod m$ is solvable for all $m \geqslant 2$.

8. Show $\mathbf{Q}(\sqrt{-31})$ has class number 3 and a prime ideal dividing (2) generates the ideal class group.

9. Show $\mathbf{Q}(\sqrt{-39})$ has a class group that is cyclic of order 4.

10. Show $\mathbf{Q}(\sqrt{-38})$ has class number 6.

11. Show $\mathbf{Q}(\sqrt{-62})$ has a class group that is cyclic of order 8.

12. Show $\mathbf{Q}(\sqrt{-210})$ has class group isomorphic to $(\mathbf{Z}/(2))^3$ which is generated by any three of $[\mathfrak{p}_2], [\mathfrak{p}_3], [\mathfrak{p}_5], [\mathfrak{p}_7]$.

13. In $\mathbf{Z}[\sqrt{79}]$, factor $(a + \sqrt{79})$ into prime ideals for $a = 2, 4, 8$, and 10 using the data and notation from Example 7.21.

14. Let $K = \mathbf{Q}(\alpha)$, where $\alpha^3 - \alpha - 10 = 0$. The polynomial $T^3 - T - 10$ has a unique real root ≈ 2.308.

    a) Show $[K : \mathbf{Q}] = 3$, $\mathcal{O}_K = \mathbf{Z}[\alpha]$, and $h(K)$ divides 4.

    b) If $h(K) = 4$, show $\mathrm{Cl}(K)$ is cyclic.

    c) Use knowledge of how 2 factors into prime ideals in K to show $(\alpha + 2) = (\alpha - 2)^2$ as ideals (this is natural to guess if you made a table in part b of the values $|\,\mathrm{N}_{K/\mathbf{Q}}(\alpha + m)|$ where $|m|$ is small) and then write the ratio $\frac{\alpha+2}{(\alpha-2)^2}$ explicitly in the form $a\alpha^2 + b\alpha + c$ with integers a, b, and c.

    d) Show the unit found in part c is not a square and use it to show $h(K) = 4$. (It is convenient to use reduction mod $\mathfrak{p}$ for a well-chosen prime $\mathfrak{p}$, like at the end of Nonexample 7.49.)

    e) Determine $\mathcal{O}_K^\times$.

15. Draw a picture of (7.5) for all possibilities when $r_1+2r_2=3$. Is the region convex or bounded?

16. Redo the proof of the Minkowski bound using the convex and centrally-symmetric region

$$X_t = \{(x_1,\dots,z_1,\dots) \in \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} : |x_i| < t, |z_j| < t\}.$$

This is a simpler type of region than the one used in the main text: if an algebraic integer has Euclidean image in X_t, its norm has absolute value at most t^n, where $n = [K:\mathbf{Q}]$, a bound which does not need the trickery of the arithmetic-geometric mean inequality.

Work through the proof of Theorem 7.15 using X_t for suitable t and find a number $b > 0$ (depending on K) such that each ideal class in $\mathcal{O}_K$ contains an integral ideal with norm at most b. (Be sure you compute the volume of X_t correctly. It is generally not $(2t)^n$.) Show b is *always* weaker than the Minkowski bound when $n \geqslant 3$ (what if $n = 2$?) Does this alternate bound allow you to prove every number field $K \neq \mathbf{Q}$ has $|\operatorname{disc}(K)| > 1$?

17. Show each of the following polynomials is irreducible in $\mathbf{Q}[T]$, has two real roots which are inverses of each other, and the remaining complex roots have absolute value 1.

a) $T^4 - T^3 - T^2 - T + 1$

b) $T^6 - T^5 - T^4 - T^3 - T^2 - T + 1$.

18. In $\mathbf{Z}[\sqrt[3]{2}]$, find a single generator for each prime ideal factor of (3) and (5) in Table 4.3. (Hint: Compare the known norm of an ideal in that table with the norm of an element in the ideal.)

19. Let $K = \mathbf{Q}(\alpha)$, where $\alpha^3 - 9\alpha - 9 = 0$. The polynomial $T^3 - 9T - 9$ is irreducible in $\mathbf{Q}[T]$ since it's irreducible mod 2. It has three real roots, which are approximately

$$-2.226681, \quad -1.184792, \quad 3.411474.$$

a) Compute $\mathcal{O}_K$. (Hint: Theorem 6.32.)

b) Determine how all the prime numbers below 20 factor in $\mathcal{O}_K$.

c) Use the Minkowski bound to compute $h(K)$.

d) Show $\alpha+1$ and $\alpha+2$ are multiplicatively independent units and they generate a subgroup of $\mathcal{O}_K^\times$ with finite index. (Hint: Use the log mapping to turn this into a linear independence problem.)

20. Let $K = \mathbf{Q}(\alpha)$ where $\alpha^3 - 9\alpha - 2 = 0$.

a) Show $[K : \mathbf{Q}] = 3$ and K is totally real.

b) Show $\mathcal{O}_K = \mathbf{Z}[\alpha]$.

c) Make a table of $\left|\mathrm{N}_{K/\mathbf{Q}}(\alpha+m)\right|$ for small $|m|$ and use it show $h(K) = 1$.

d) Find two independent units using principal ideals with different generators.

21. a) Show $\mathbf{Q}(\sqrt[3]{3})$ has ring of integers $\mathbf{Z}[\sqrt[3]{3}]$ and class number 1.

b) Show the fundamental unit of $\mathbf{Q}(\sqrt[3]{3})$ is $4 + 3\sqrt[3]{3} + 2\sqrt[3]{9}$.

22. a) Show $\mathbf{Q}(\sqrt[3]{6})$ has ring of integers $\mathbf{Z}[\sqrt[3]{6}]$ and class number 1.

b) Check $(T-2)^3 - 6$ is Eisenstein at 2 and show $(2-\sqrt[3]{6})$ is the unique prime over 2.

c) Since $(2) = (2-\sqrt[3]{6})^3$ as ideals, the ratio $u := \frac{2}{(2-\sqrt[3]{6})^3} \approx 326.9908$ is a unit in $\mathbf{Z}[\sqrt[3]{6}]$. Write it explicitly in the form $u = a + b\sqrt[3]{6} + c\sqrt[3]{36}$ with integers a, b, and c, and prove u is the fundamental unit.

23. Show $\mathbf{Q}(\alpha)$, where α is the real root of $T^3 + T + 1$ (discriminant is -31), has fundamental unit $-1/\alpha \approx 1.46$.

24. Let K be a cubic field which is totally real: $r_1 = 3$ and $r_2 = 0$.

a) The polynomial $T^3 - 3T - 1$ has three real roots. Let α be a root and $E = \mathbf{Q}(\alpha)$. Show $[E : \mathbf{Q}] = 3$, 3 is totally ramified in E, and $\mathcal{O}_E = \mathbf{Z}[\alpha]$.

b) Show $h(E) = 1$.

c) The unit group of E has rank 2 by Dirichlet's unit theorem. One unit is α. Verify $\alpha + 1$ is a unit, and check multiplicative independence of the pair $\{\alpha, \alpha+1\}$ numerically.

d) Show $\alpha - 2$, $2\alpha + 3$, and $3\alpha + 1$ are all units.

e) Use the log mapping to find an expression for each of $\alpha - 2$, $2\alpha + 3$, and $3\alpha + 1$ in the form $\pm\alpha^a(\alpha+1)^b$ with integral exponents a and b.

25. Let $E = \mathbf{Q}(\alpha)$ and $F = \mathbf{Q}(\beta)$, where $\alpha^3 - 8\alpha - 15 = 0$ and $\beta^3 - \beta^2 - 7\beta - 12 = 0$. The polynomials $T^3 - 8T - 15$ and $T^3 - T^2 - 7T - 12$ are both irreducible mod 13, so E and F are cubic fields.

    a) Verify that $\mathcal{O}_E$ has $\mathbf{Z}$-basis $\{1, \alpha, \alpha^2\}$, $\mathcal{O}_F$ has $\mathbf{Z}$-basis $\{1, \beta, \beta^2\}$, and $\operatorname{disc}(E) = \operatorname{disc}(F) = -4027$. (The number 4027 is prime.)

    b) Determine the shape of the factorization of $2, 3, 5, \dots$ in E and F until you reach a prime where the shape is not the same, thus showing $E \not\cong F$.

    c) Determine the class numbers of E and F using the Minkowski bound.

26. Let $K = \mathbf{Q}(i, \sqrt{-5})$. By Exercise 3.11, $\mathcal{O}_K = \mathbf{Z}[(i + \sqrt{-5})/2]$ and $\operatorname{disc}(K) = 400$. Show K has class number 1. (Hint: Letting $\alpha = \frac{i+\sqrt{-5}}{2}$, it is convenient to factor the ideals $(\alpha^2 + \alpha + c)$ for $c = 1$, 2, and 3.)

27. (Continuation of Exercise 6.32) Find the Minkowski bounds for $\mathbf{Q}(\sqrt[8]{3})$ and $\mathbf{Q}(\sqrt[8]{48})$.

28. For a number field K, its zeta-function has a simple pole at $s = 1$ and a zero of order $r_1 + r_2 - 1$ at $s = 0$. Use the functional equation for $\zeta_K(s)$ and the analytic class number formula (the residue at $s = 1$) to show the power series for $\zeta_K(s)$ at $s = 0$ starts off as

$$\zeta_K(s) = -\frac{h(K)R(K)}{w_K} s^{r_1+r_2-1} + \text{higher powers of } s. \tag{7.33}$$

29. a) Use the inequality in Theorem **??** to show a cubic field F with $|\operatorname{disc}(F)| < 31.39$ has no proper unramified extension and find an example of such a field.[14]

    b) Show quartic fields with $|\operatorname{disc}(F)| < 158.32$ have no proper unramified extension and find an example of such a field.

    c) Use Odlyzko's lower bound on discriminants to broaden the inequalities in parts a and b, for totally complex fields, to $|\operatorname{disc}(F)| < 58.06$ and $|\operatorname{disc}(F)| < 512.6$, respectively.

30. Show the number field $\mathbf{Q}(\sqrt[3]{7})$ has ring of integers $\mathbf{Z}[\sqrt[3]{7}]$, fundamental unit $2 - \sqrt[3]{7} \approx .087$, and class number 3 with its class group generated by a prime of norm 3.

[14] CHECK: Complete this, check numerics.

31. (Continuation of Exercise 7.30) Here is a method to find examples of number fields having unramified extensions of a chosen degree. For $n \geqslant 2$, the monic polynomial $f(T) = T^n + A_{n-1}T^{n-1} + \cdots + A_1T + A_0$ with indeterminate coefficients has a discriminant that is a polynomial function of $A_0, \ldots, A_{n-1}$ with integral coefficients, *e.g.*, $\operatorname{disc}(T^3 + AT + B) = -4A^3 - 27B^2$. It might not be possible to solve $\operatorname{disc}(T^n + a_{n-1}T^{n-1} + \cdots + a_1T + a_0) = 1$ in integers, but we can certainly solve it often with algebraic integers.

    a) Let $a_0, \ldots, a_{n-1}$ be algebraic integers satisfying $\operatorname{disc}(T^n + a_{n-1}T^{n-1} + \cdots + a_1T + a_0) = 1$. Set $K = \mathbf{Q}(a_0, \ldots, a_{n-1})$ and $f(T) = T^n + a_{n-1}T^{n-1} + \cdots + a_1T + a_0 \in \mathcal{O}_K[T]$. If $f(T)$ is irreducible over K and α is a root of $f(T)$, show the field $K(\alpha)$ has ring of integers $\mathcal{O}_K[\alpha]$ and it is unramified at all primes in K.

    b) Why won't this construction ever work when $n = 2$? (Try to make an example over $\mathbf{Z}$ and it should become obvious.)

    c) The discriminant of $T^3 - \sqrt[3]{7}T + 1$ is equal to 1. Show this polynomial is irreducible over $\mathbf{Q}(\sqrt[3]{7})$, so a root of it generates a cubic unramified extension of $\mathbf{Q}(\sqrt[3]{7})$. (Hint: use reduction mod $\mathfrak{p}$ for a prime ideal of small norm, as in Exercise 4.24a.)

32. Generalize Corollary 7.30: if F is any number field, $x > 0$, and S is a finite set of nonzero prime ideals in $\mathcal{O}_F$, the set of finite extensions E inside an algebraic closure $\overline{F}$ such that $[E : F] \leqslant x$ and E/F is unramified outside S is finite. (Hint: $F/\mathbf{Q}$ is unramified outside a finite set of prime numbers.)

33. This exercise shows the simplest attempt to state an analogue of Hermite's theorem with $\mathbf{F}_p(X)$ in place of $\mathbf{Q}$ has counterexamples: the algebraic closure of $\mathbf{F}_p(X)$ contains infinitely many extensions of degree p with $\mathbf{F}_p[X]$-discriminant (1).

    a) For a positive integer d such that $d \not\equiv 0 \bmod p$, show $T^p - T - X^p$ is irreducible in $\mathbf{F}_p(X)[T]$.

    b) With d as in part a, let α_d be a root of $T^p - T - X^d$ and set $K_d = \mathbf{F}_p(X, \alpha_d)$, so $[K_d : \mathbf{F}_p(X)] = p$. Show the integral closure of $\mathbf{F}_p[X]$ in K_d is $\mathbf{F}_p[X][\alpha_d]$ and all primes in $\mathbf{F}_p[X]$ are unramified in K_d.

    c) Using Artin–Schreier theory, show $K_d \not\cong K_{d'}$ as extensions of $\mathbf{F}_p(X)$ when d and d' are distinct positive integers that are not multiples of p.

34. The field $\mathbf{Q}(\sqrt{2}, \sqrt{3}, \sqrt{-5})$ has no real embeddings. Show the only roots of unity in it are ±1.

35. For any number field K, show $\mathcal{O}_K^\times/(\mathcal{O}_K^\times)^2$, treated as a vector space over $\mathbf{Z}/2\mathbf{Z}$ (acting by exponents), has dimension $r_1 + r_2$.

36. For any positive integer m, show $\mathcal{O}_K^\times/(\mathcal{O}_K^\times)^m$ is finite while the group $\mathcal{O}_{K,\mathfrak{p}}^\times/(\mathcal{O}_{K,\mathfrak{p}}^\times)^m$ is infinite for any nonzero prime $\mathfrak{p}$ in $\mathcal{O}_K$. This is an important sense in which the localizations of $\mathcal{O}_K$ at its prime ideals are not simpler than $\mathcal{O}_K$ itself.

37. For $d \geqslant 2$, show there is an explicit constant B_d such that for every number field K with degree d, $\#\mu(K) \leqslant B_d$.

38. If m is composite and $m \not\equiv 2 \bmod 4$, show $1-\zeta_m$ is a unit in $\mathbf{Q}(\zeta_m)$. (Hint: For any prime factor p of m, show $(1-\zeta_m) \mid (1-\zeta_p)$.)

39. a) Suppose $[E : F] = 2$, F is totally real ($r_2(F) = 0$), and E is totally complex ($r_1(E) = 0$). Use the Dirichlet unit theorem to prove $\mathcal{O}_F^\times$ and $\mathcal{O}_E^\times$ have the same rank.

    b) Prove the converse of part a: if $F \subsetneqq E$ and $\mathcal{O}_F^\times$ and $\mathcal{O}_E^\times$ have the same rank, then show $[E : F] = 2$, F is totally real, and E is totally complex. (Hint: In addition to using the unit rank formula from Dirichlet's unit theorem, use the field degree formula $n = r_1 + 2r_2$.)

    Remark. It can be shown that when we increase $\mathcal{O}_F^\times$ by the roots of unity in E we nearly get all the units in E: $[\mathcal{O}_E^\times : \mu(E)\mathcal{O}_F^\times]$ is 1 or 2. See [60, Theorem 4.12].

40. A number field K is called a *CM field* if it is a totally complex quadratic extension of a totally real number field. (Such fields arose in the study of abelian varieties with "complex multiplication," hence the name.) Examples include imaginary quadratic fields and $\mathbf{Q}(\zeta_m)$ for $m \geqslant 3$. The previous exercise shows a special feature of these fields in terms of their unit group and that of a subfield.

    a) If K is a CM field, show it has an intrinsic complex conjugation, in the sense that for every field embedding $f\colon K \to \mathbf{C}$ the automorphism $c(\alpha) := f^{-1}(\overline{f(\alpha)})$ on K is independent of f.

    b) Show the only totally real subfield of a CM field K over which K is a quadratic extension is the fixed field of the intrinsic complex conjugation on K from part a. So the totally real subfield in the definition of a CM field is unique. It is usually denoted K^+.

c) Let K be a number field such that for all field embeddings $f\colon K \to \mathbf{C}$ the automorphism $c(\alpha) := f^{-1}(\overline{f(\alpha)})$ is independent of f and c has order 2. Show K is a CM field.

d) If $K/\mathbf{Q}$ is an *abelian* Galois extension, show it is either totally real or a CM field.

e) Show any S_3-extension of $\mathbf{Q}$ (a Galois extension with Galois group isomorphic to S_3) is neither totally real nor a CM field.

f) Give an example of a CM field which is not Galois over $\mathbf{Q}$.

41. A unit in a number field can be reduced mod $\mathfrak{p}$ for any prime $\mathfrak{p}$. The primes of norm up to 25 in $\mathbf{Z}[\sqrt{2}]$ are $(\sqrt{2})$, $(3+\sqrt{2})$, $(3-\sqrt{2})$, (3), $(5+2\sqrt{2})$, $(5-2\sqrt{2})$, $(5+\sqrt{2})$, $(5-\sqrt{2})$, and (5). For each of these primes $\mathfrak{p}$, determine if the unit $1+\sqrt{2}$ is a generator of $(\mathbf{Z}[\sqrt{2}]/\mathfrak{p})^\times$.

42. Generalize Theorem 7.52 to arbitrary multiquadratic fields: find a set of units from quadratic subfields which generates a subgroup of finite index in the unit group of the whole field.

43. Let $K = \mathbf{Q}(\sqrt[3]{175})$. Set $\alpha = \sqrt[3]{175} = \sqrt[3]{5^2 \cdot 7}$ and $\beta = \sqrt[3]{245} = \sqrt[3]{5 \cdot 7^2}$. Let $v = 13516\alpha + 12082\beta + 75601$. In the text, to prove v is the fundamental unit of K we wanted to show v mod $\mathfrak{p}$ is not a square or a cube for some prime $\mathfrak{p}$ and we were able to do both of these when $\mathfrak{p}$ is one of the primes over 37, which is the first prime splitting completely in K. It is natural to ask if we can do this with any prime ideal lying over a smaller prime number.

Since $[\mathcal{O}_K : \mathbf{Z}[\alpha]] = 7$ and $[\mathcal{O}_K : \mathbf{Z}[\beta]] = 5$, we can factor every prime $p \neq 7$ by factoring $T^3 - 175 \bmod p$ and every prime $p \neq 5$ by factoring $T^3 - 245 \bmod p$. In the table below we factor every prime below 37, except for 7.

p	$T^3 - 175 \bmod p$	(p)
2	$(T-1)(T^2+T+1)$	$\mathfrak{p}_2\mathfrak{p}_4$
3	$(T-1)^3$	$\mathfrak{p}_3^3$
5	T^3	$\mathfrak{p}_5^3$
11	$(T+1)(T^2+10T+1)$	$\mathfrak{p}_{11}\mathfrak{p}_{121}$
13	irreducible	(13)
17	$(T+6)(T^2+11T+2)$	$\mathfrak{p}_{17}\mathfrak{p}_{289}$
19	irreducible	(19)
23	$(T+6)(T^2+17T+3)$	$\mathfrak{p}_{23}\mathfrak{p}_{23^2}$
29	$(T-1)(T^2+T+1)$	$\mathfrak{p}_{29}\mathfrak{p}_{29^2}$
31	irreducible	(31)

a) Show $(7) = (7, \beta)^3$.

b) Check $v \equiv 11 \bmod \mathfrak{p}_{17}$ and this is not a square in $\mathcal{O}_K/\mathfrak{p}_7$.

c) Check $v \not\equiv \square \bmod \mathfrak{p}$ where $\mathfrak{p}$ is any prime lying over 17, 23, and 29. (To handle primes with residue field degree 2 it is convenient to use a computer algebra system.)

d) Show $v \equiv \square \bmod \mathfrak{p}$ where $\mathfrak{p}$ is any prime lying over a prime less than 37 other than 17, 23, and 29.

e) For every prime $\mathfrak{p}$ lying over a prime number less than 37, show $v \bmod \mathfrak{p}$ is a cube.

44. In a number field K, let $r = r_1 + r_2 - 1$ and assume $r > 0$.

a) If $u_1, \ldots, u_r$ are a multiplicatively independent sets of r units in K and $v_1, \ldots, v_r$ are a multiplicatively independent subset of the group $\langle u_1, \ldots, u_r,$ show

$$R(v_1, \ldots, v_r) = [\langle u_1, \ldots, u_r\rangle : \langle v_1, \ldots, v_r\rangle] R(u_1, \ldots, u_r).$$

In particular, the index can be computed as a ratio of regulators.

b) If $v_1, \ldots, v_r$ are a multiplicatively independent sets of r units in K, show

$$R(v_1, \ldots, v_r) = [\mathcal{O}_K^\times : \mu(K)\langle v_1, \ldots, v_r\rangle] R(K).$$

c) Show two sets of r independent units in K which generate the same group up to roots of unity have the same regulator.

45. For positive integers $g, g', f_1, \dots, f_g, f'_1, \dots, f'_{g'}$, suppose $\prod_{i=1}^{g}(1 - z^{f_i}) = \prod_{j=1}^{g'}(1 - z^{f'_j})$ for all $z \in \mathbf{C}$ (infinitely many z is enough). Labeling the exponents in increasing order, so $f_1 \leqslant f_2 \leqslant \cdots \leqslant f_g$ and $f'_1 \leqslant f'_2 \leqslant \cdots \leqslant f'_{g'}$, show $g = g'$ and $f_i = f'_i$ for all i. (This is used in the proof of Theorem 7.67.)

46. Let S be a finite set of prime ideals in a number field K. The S-zeta-function of K is the Euler product with the factors associated to primes in S taken out:

$$\zeta_{K,S}(s) = \prod_{\mathfrak{p} \notin S} \frac{1}{1 - 1/\operatorname{N}(\mathfrak{p})^s} = \sum_{(\mathfrak{a}, S) = 1} \frac{1}{\operatorname{N}(\mathfrak{a})^s},$$

where the notation $(\mathfrak{a}, S) = 1$ means $\mathfrak{a}$ is not divisible by any prime ideal in S. (Admittedly the use of S and s in $\zeta_{K,S}(s)$ is unfortunate, but if you want to write this clearly by hand just use visible serifs in the S.)

Since $\zeta_{K,S}(s) = \zeta_K(s) \prod_{\mathfrak{p} \in S}(1 - \operatorname{N}(\mathfrak{p})^{-s})$, the right side provides a meromorphic continuation of $\zeta_{K,S}(s)$ to $\mathbf{C}$. We want to show $\zeta_{K,S}(s)$, as a meromorphic function, determines $\zeta_K(s)$. That is, removing a finite number of Euler factors from $\zeta_K(s)$ does not lose memory of the full zeta-function of K.

a) Use the equation $\zeta_{K,S}(s) = \zeta_K(s) \prod_{\mathfrak{p} \in S}(1 - \operatorname{N}(\mathfrak{p})^{-s})$ to compute the order of vanishing of $\zeta_{K,S}(s)$ at $s = 0$, -1, and -2 to show the meromorphic function $\zeta_{K,S}(s)$ knows $\#S$.

b) It is a theorem, essentially equivalent to the prime ideal theorem, that $\zeta_K(s)$ has no zeros on the imaginary axis except perhaps at $s = 0$. If S is nonempty, look at zeros of $\zeta_{K,S}(s)$ on the positive imaginary axis to show the meromorphic function $\zeta_{K,S}(s)$ determines $\zeta_K(s)$. (Hint: Use the smallest zero of $\zeta_{K,S}(s)$ on the positive imaginary axis to show $\zeta_{K,S}(s)$ determines $\zeta_{K,S-\{\mathfrak{p}\}}(s)$ for some prime $\mathfrak{p}$ in S.)

c) If we know the shape of the factorizations of all but finitely many prime numbers into prime ideals in K, use part b and the previous exercise to show that the shape of the factorizations of all unramified primes in K is determined.

CHAPTER 8

APPENDIX

8.1 Linear Algebra in Field Extensions

Let E/F be a finite extension of fields with degree n. For $\alpha \in E$, let $m_\alpha \colon E \to E$ be multiplication by α: $m_\alpha(x) = \alpha x$. Picking an F-basis $\{e_1, \ldots, e_n\}$ of E lets us view E as F^n and any F-linear map $L \colon E \to E$ as an $n \times n$ matrix $[L] = (a_{ij})$ in $\mathrm{M}_n(F)$, where $L(e_j) = \sum_{i=1}^{n} a_{ij} e_i$. That is, the j-th column of $[L]$ is the set of coordinates of $L(e_j)$ in the basis, so we fill in $[L]$ by columns, not rows. In particular, since m_α is F-linear any F-basis of E provides us with a matrix representation $[m_\alpha]$ for α.

Example 8.1. In $\mathbf{C}$, with $\alpha = a+bi$ and $\mathbf{R}$-basis $\{1, i\}$, we compute $\alpha \cdot 1 = a+bi$ and $\alpha \cdot i = -b + ai$. Therefore

$$[m_\alpha] = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

These 2×2 real matrices add and multiply in the same way as complex numbers.

Example 8.2. Let $E = \mathbf{Q}(\sqrt{2})$ and $\alpha = 7 + 3\sqrt{2}$. Using the $\mathbf{Q}$-basis $\{1, \sqrt{2}\}$, we compute $\alpha \cdot 1 = 7 + 3\sqrt{2}$ and $\alpha \cdot \sqrt{2} = 6 + 7\sqrt{2}$, so

$$[m_\alpha] = \begin{pmatrix} 7 & 6 \\ 3 & 7 \end{pmatrix}.$$

For α, β, and x in E,

$$m_{\alpha+\beta}(x) = (\alpha+\beta)(x) = \alpha x + \beta x = m_\alpha(x) + m_\beta(x) = (m_\alpha + m_\beta)(x)$$

and

$$(m_\alpha \circ m_\beta)(x) = m_\alpha(\beta x) = \alpha(\beta x) = (\alpha\beta)x = m_{\alpha\beta}(x),$$

so $m_{\alpha+\beta} = m_\alpha + m_\beta$ and $m_{\alpha\beta} = m_\alpha \circ m_\beta$. We can recover α from m_α by evaluating at 1: $m_\alpha(1) = \alpha \cdot 1 = \alpha$.

Definition 8.3. The *characteristic polynomial* of α is

$$\chi_{E/F,\alpha}(T) = \det(TI_n - [m_\alpha]) \in F[T].$$

The *trace* $\mathrm{Tr}_{E/F} \colon E \to F$ is

$$\mathrm{Tr}_{E/F}(\alpha) = \mathrm{Tr}([m_\alpha]) \in F,$$

and the *norm* $\mathrm{N}_{E/F} \colon E \to F$ is

$$\mathrm{N}_{E/F}(\alpha) = \det([m_\alpha]) \in F.$$

The trace and norm of α show up as coefficients of the characteristic polynomial of α:

$$\chi_{E/F,\alpha}(T) = \det(TI_n - [m_\alpha]) = T^n - \mathrm{Tr}_{E/F}(\alpha)T^{n-1} + \cdots + (-1)^n \mathrm{N}_{E/F}(\alpha).$$

Example 8.4. For $\alpha = a + bi$ in $\mathbf{C}$,

$$\chi_{\mathbf{C}/\mathbf{R},\alpha}(T) = \det(TI_2 - [m_\alpha]) = T^2 - 2aT + (a^2 + b^2) \in \mathbf{R}[T],$$

so

$$\mathrm{Tr}_{\mathbf{C}/\mathbf{R}}(a+bi) = 2a \qquad \text{and} \qquad \mathrm{N}_{\mathbf{C}/\mathbf{R}}(a+bi) = a^2 + b^2 = \alpha\overline{\alpha}.$$

Example 8.5. Let $E = \mathbf{Q}(\sqrt{r})$ for r a nonsquare rational number, $F = \mathbf{Q}$, and use basis $\{1, \sqrt{r}\}$. For $\alpha = a + b\sqrt{r}$, $[m_\alpha]$ equals

$$\begin{pmatrix} a & rb \\ b & a \end{pmatrix}.$$

Thus $\chi_{\mathbf{Q}(\sqrt{r})/\mathbf{Q},\alpha}(T) = T^2 - 2aT + (a^2 - rb^2)$,

$$\mathrm{Tr}_{\mathbf{Q}(\sqrt{r})/\mathbf{Q}}(a + b\sqrt{r}) = 2a, \quad \mathrm{N}_{\mathbf{Q}(\sqrt{r})/\mathbf{Q}}(a + b\sqrt{r}) = a^2 - rb^2.$$

Example 8.6. Let $E = \mathbf{Q}(\gamma)$ for γ a root of $T^3 - T - 1$, $F = \mathbf{Q}$, and use basis $\{1, \gamma, \gamma^2\}$. For $\alpha = a + b\gamma + c\gamma^2$, $[m_\alpha]$ equals

$$\begin{pmatrix} a & c & b \\ b & a+c & b+c \\ c & b & a+c \end{pmatrix}.$$

We have $\mathrm{Tr}_{\mathbf{Q}(\gamma)/\mathbf{Q}}(a + b\gamma + c\gamma^2) = 3a + 2c$ and

$$\mathrm{N}_{\mathbf{Q}(\gamma)/\mathbf{Q}}(a + b\gamma + c\gamma^2) = a^3 + b^3 + c^3 - ab^2 + ac^2 - bc^2 + 2a^2c - 3abc.$$

Example 8.7. For $c \in F$, we have

$$[m_\alpha] = cI_n = \begin{pmatrix} c & 0 & \cdots & 0 \\ 0 & c & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c \end{pmatrix},$$

so $\chi_{E/F,c}(T) = (T - c)^n$, $\mathrm{Tr}_{E/F}(c) = nc$, and $\mathrm{N}_{E/F}(c) = c^n$.

Remark 8.8. In the literature you might see S or Sp used for the trace since Spur is the German word for trace.

Corollary 8.9. *The trace* $\mathrm{Tr}_{E/F} \colon E \to F$ *is* F*-linear and the norm* $\mathrm{N}_{E/F} \colon E \to F$ *is multiplicative. Moreover,* $\mathrm{N}_{E/F}(E^\times) \subset F^\times$.

Proof. We have equations of linear maps $m_{\alpha+\beta} = m_\alpha + m_\beta$ and $m_{c\alpha} = cm_\alpha$. Taking the trace of both sides, $\mathrm{Tr}_{E/F}(\alpha + \beta) = \mathrm{Tr}_{E/F}(\alpha) + \mathrm{Tr}_{E/F}(\beta)$ and $\mathrm{Tr}_{E/F}(c\alpha) = c\,\mathrm{Tr}_{E/F}(\alpha)$. So the trace is F-linear. Taking determinants of both sides of the equation $m_{\alpha\beta} = m_\alpha \circ m_\beta$, we get $\mathrm{N}_{E/F}(\alpha\beta) = \mathrm{N}_{E/F}(\alpha)\,\mathrm{N}_{E/F}(\beta)$.

Finally, since $\mathrm{N}_{E/F}(1) = 1$, for nonzero α in E we take norms of both sides of $\alpha \cdot (1/\alpha) = 1$ to get $\mathrm{N}_{E/F}(\alpha)\,\mathrm{N}_{E/F}(1/\alpha) = 1$, so $\mathrm{N}_{E/F}(\alpha) \neq 0$. ∎

Here is the crucial property of the characteristic polynomial of α: it provides us with a polynomial in $F[T]$ that has α as a root.

Theorem 8.10. *For all* $\alpha \in E$, $\chi_{E/F,\alpha}(\alpha) = 0$.

Proof. Write

$$\chi_{E/F,\alpha}(T) = T^n + c_{n-1}T^{n-1} + \cdots + c_1T + c_0 \in F[T].$$

The Cayley-Hamilton theorem says $\chi_{E/F,\alpha}([m_\alpha]) = O$. Note

$$\begin{aligned}\chi_{E/F,\alpha}([m_\alpha]) &= [m_\alpha]^n + c_{n-1}[m_\alpha]^{n-1} + \cdots + c_1[m_\alpha] + c_0I_n \\ &= [m_{\alpha^n}] + c_{n-1}[m_{\alpha^{n-1}}] + \cdots + c_1[m_\alpha] + c_0I_n \\ &= [m_{\alpha^n+c_{n-1}\alpha^{n-1}+\cdots+c_1\alpha+c_0}].\end{aligned}$$

Since $[m_0] = O$ too, $\alpha^n + c_{n-1}\alpha^{n-1} + \cdots + c_1\alpha + c_0 = 0$ because m_t determines t. ■

Example 8.11. The complex number $a + bi$ is a root of the real polynomial $\chi_{\mathbf{C}/\mathbf{R},a+bi}(T) = T^2 - 2aT + a^2 + b^2$.

Although $\chi_{E/F,\alpha}(T)$ is monic in $F[T]$ with α as a root, it need not be the minimal polynomial of α over F because its degree is always $n = [E : F]$, whereas the minimal polynomial of α over F has degree varying with α. Call the minimal polynomial $f_\alpha(T)$. (We suppress the reference to the field F in this notation.) The polynomial $f_\alpha(T)$ is a factor of $\chi_{E/F,\alpha}(T)$, and its degree divides n since $\deg f_\alpha(T) = [F(\alpha) : F]$ is a factor of $n = [E : F]$. Replacing E with $F(\alpha)$, we have $\chi_{F(\alpha)/F,\alpha}(T) = f_\alpha(T)$ since $f_\alpha(T)$ is the only monic polynomial of degree $[F(\alpha) : F]$ in $F[T]$ with α as a root.

Theorem 8.12. *For all $\alpha \in E$, $\chi_{E/F,\alpha}(T) = f_\alpha(T)^{n/d}$ where $d = [F(\alpha) : F]$.*

In other words, $\chi_{E/F,\alpha}(T)$ is the power of the minimal polynomial of α having degree $n = [E : F]$. As a simple example, for $c \in F$ its minimal polynomial in $F[T]$ is $T - c$ while its characteristic polynomial for E/F is $(T - c)^n$.

Proof. We create a basis of E/F using α in two steps: first for $F(\alpha)/F$ and then for $E/F(\alpha)$. We use $\{1, \alpha, \ldots, \alpha^{d-1}\}$ as a basis of $F(\alpha)/F$ and let $e_1, \ldots, e_{n/d}$ be any basis of $E/F(\alpha)$. Then a basis of E/F is $\{e_i\alpha^j\}$ and

$$E = \bigoplus_{i=1}^{n/d} F(\alpha)e_i = \bigoplus_{i=1}^{n/d}\bigoplus_{j=0}^{d-1} F\alpha^j e_i.$$

Multiplication by α preserves each F-subspace $F(\alpha)e_i$ of E, and the matrix for multiplication by α on $F(\alpha)$ with respect to the basis $\{1, \alpha, \ldots, \alpha^{d-1}\}$ is also the matrix for multiplication by α on $F(\alpha)e_i$ with respect to the basis $\{e_i, \alpha e_i, \ldots, \alpha^{d-1}e_i\}$, so we have the block matrix decomposition

$$[m_\alpha]_{E/F} = \begin{pmatrix} [m_\alpha]_{F(\alpha)/F} & O & \cdots & O \\ O & [m_\alpha]_{F(\alpha)/F} & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & [m_\alpha]_{F(\alpha)/F} \end{pmatrix}.$$

Thus $\chi_{E/F,\alpha}(T) = \chi_{F(\alpha)/F,\alpha}(T)^{n/d} = f_\alpha(T)^{n/d}$. ■

Example 8.13. Let $K = \mathbf{Q}(\sqrt{2}, \sqrt{3})$. Since $[K : \mathbf{Q}] = 4$ and $\sqrt{2}$ has degree 2 over $\mathbf{Q}$, $\chi_{K/\mathbf{Q},\sqrt{2}}(T) = (T^2 - 2)^2$.

Example 8.14. Let $K = \mathbf{Q}(\sqrt[3]{2})$. Since $[K : \mathbf{Q}] = 3$, the characteristic polynomials of $\sqrt[3]{2}$ and $\sqrt[3]{4}$ for the extension $K/\mathbf{Q}$ must be $T^3 - 2$ and $T^3 - 4$ since these are the minimal polynomials of the two numbers in $\mathbf{Q}[T]$ and their degrees are already 3.

Corollary 8.15. *Over a large enough field, let the minimal polynomial $f_\alpha(T)$ factor as $(T - \alpha_1)\cdots(T - \alpha_d)$. Then*

$$\mathrm{Tr}_{E/F}(\alpha) = \frac{n}{d}(\alpha_1 + \cdots + \alpha_d), \quad \mathrm{N}_{E/F}(\alpha) = (\alpha_1 \cdots \alpha_d)^{n/d}.$$

Proof. The trace of α is the negative of the coefficient of T^{n-1} in $\chi_{E/F,\alpha}(T)$ and the norm is the constant term of $\chi_{E/F,\alpha}(T)$ multiplied by $(-1)^n$. Therefore the formulas for $\mathrm{Tr}_{E/F}(\alpha)$ and $\mathrm{N}_{E/F}(\alpha)$ are immediate from computing these coefficients in $f_\alpha(T)^{n/d}$. ■

This is *not* saying the trace and norm of α are the sum and product of the roots of the minimal polynomial of α over F. Those roots have to be repeated n/d times, where $d = [F(\alpha) : F]$, making a total of n terms in the sum and product.

Corollary 8.16. *Suppose in a large enough field extension the characteristic polynomial of α relative to E/F splits completely as*

$$\chi_{E/F,\alpha}(T) = (T - r_1)\cdots(T - r_n).$$

Then for any $g(T) \in F[T]$,

$$\chi_{E/F,g(\alpha)}(T) = (T - g(r_1))\cdots(T - g(r_n)),$$

so

$$\mathrm{Tr}_{E/F}(g(\alpha)) = \sum_{i=1}^{n} g(r_i), \quad \mathrm{N}_{E/F}(g(\alpha)) = \prod_{i=1}^{n} g(r_i).$$

In particular, $\chi_{E/F,\alpha^m}(T) = (T - r_1^m)\cdots(T - r_n^m)$*, so* $\mathrm{Tr}_{E/F}(\alpha^m) = \sum_{i=1}^n r_i^m$.

Proof. By Theorem 8.12, $\chi_{E/F,\alpha}(T)$ is a power of the minimal polynomial of α in $F[T]$, so every r_i has the same minimal polynomial over F as α.

Set $f(T) = (T - g(r_1))\cdots(T - g(r_n))$. We want to show this is the characteristic polynomial of $g(\alpha)$. The coefficients of $f(T)$ are symmetric polynomials in $r_1, \ldots, r_n$ with coefficients in F, so by the symmetric function theorem $f(T) \in F[T]$. Let $M(T)$ be the minimal polynomial of $g(\alpha)$ over F, so $M(T)$ is irreducible in $F[T]$. Since α and each r_i have the same minimal polynomial over F, the fields $F(\alpha)$ and $F(r_i)$ are isomorphic over F. Applying such an isomorphism to the equation $M(g(\alpha)) = 0$ turns it into $M(g(r_i)) = 0$ (because $M(T)$ and $g(T)$ have coefficients in F), so $M(T)$ is the minimal polynomial for $g(r_i)$ over F since $M(T)$ is monic irreducible in $F[T]$.

We have shown all roots of $f(T)$ have minimal polynomial $M(T)$ in $F[T]$, and $f(T)$ is monic, so $f(T)$ is a power of $M(T)$. By Theorem 8.12, $\chi_{E/F,g(\alpha)}(T) \in F[T]$ is a power of $M(T)$ with degree $[E : F] = n = \deg(f)$, so $f(T) = \chi_{E/F,g(\alpha)}(T)$. ∎

The next result is called the transitivity of the trace and norm.

Theorem 8.17. *Let $L/E/F$ be a tower of finite extensions. For $\alpha \in L$,*

$$\mathrm{Tr}_{L/F}(\alpha) = \mathrm{Tr}_{E/F}(\mathrm{Tr}_{L/E}(\alpha)), \qquad \mathrm{N}_{L/F}(\alpha) = \mathrm{N}_{E/F}(\mathrm{N}_{L/E}(\alpha)).$$

Proof. Let $(e_1, \ldots, e_m)$ be an ordered E-basis of L and $(f_1, \ldots, f_n)$ be an ordered F-basis of E. Thus as an ordered F-basis of L we can use

$$(e_1 f_1, \ldots, e_1 f_n; \ldots; e_m f_1, \ldots, e_m f_n).$$

For $\alpha \in L$, let

$$\alpha e_j = \sum_{i=1}^{m} c_{ij} e_i, \quad c_{ij} f_s = \sum_{r=1}^{n} b_{ijrs} f_r,$$

for $c_{ij} \in E$ and $b_{ijrs} \in F$. Thus $\alpha(e_j f_s) = \sum_i \sum_r b_{ijrs} e_i f_r$. So

$$[m_\alpha]_{L/E} = (c_{ij}), \quad [m_{c_{ij}}]_{E/F} = (b_{ijrs}), \quad [m_\alpha]_{L/F} = ([m_{c_{ij}}]_{E/F}).$$

Thus

$$\begin{aligned}\mathrm{Tr}_{E/F}(\mathrm{Tr}_{L/E}(\alpha)) &= \mathrm{Tr}_{E/F}\left(\sum_i c_{ii}\right)\\ &= \sum_i \mathrm{Tr}_{E/F}(c_{ii})\\ &= \sum_i \sum_r b_{iirr}\\ &= \mathrm{Tr}_{L/F}(\alpha).\end{aligned}$$

The transitivity of the norm, while important, is a lot more complicated to prove. We will prove a special case below in Corollary 8.20. ■

From now on, we focus on separable extensions.

Theorem 8.18. *Let E/F be a finite separable extension of degree n and E'/F be a normal extension admitting an F-embedding of E (this means an embedding which fixes F). There are n different F-embeddings $E \to E'$, and if we write them as $\sigma_1, \ldots, \sigma_n$, then $\mathrm{Tr}_{E/F}(\alpha) = \sum_{i=1}^n \sigma_i(\alpha)$ and $\mathrm{N}_{E/F}(\alpha) = \prod_{i=1}^n \sigma_i(\alpha)$.*

In practice, E' could be a Galois closure of E/F or an algebraic closure of F. If $F = \mathbf{Q}$ and E is a number field, E' could be the complex numbers.

Proof. First we show the number of F-embeddings $E \to E'$ is n. Since E/F is separable we can write $E = F(x)$ by the primitive element theorem. The minimal polynomial $f(T)$ of x over F has degree n and is separable. Any F-embedding $E \to E'$ is determined by its value on x and its sends x to a root of $f(T)$ in E'. Since we assume E has some F-embedding into E', $f(T)$ has a root in E', so $f(T)$ has n roots in E' since E'/F is normal and $f(T)$ is separable. Therefore there are n different F-embeddings of E into E'.

For any $\alpha \in E$, let $d = [F(\alpha) : F]$ and let $\alpha_1, \ldots, \alpha_d$ be the roots in E' of the minimal polynomial of α in $F[T]$. By the extension theorem for field homomorphisms into a normal extension, the values of $\sigma_i(\alpha)$ are $\alpha_1, \ldots, \alpha_d$ each repeated n/d times. Therefore

$$\sum_{i=1}^n \sigma_i(\alpha) = \frac{n}{d}(\alpha_1 + \cdots + \alpha_d),$$

which we recognize as $\mathrm{Tr}_{E/F}(\alpha)$ in Corollary 8.15. The argument for the norm formula is similar. ■

If E/F were an inseparable extension then Theorem 8.18 has an analogue: each σ_i has to be repeated in the sum and product with multiplicity equal to the inseparable degree of E/F.

Corollary 8.19. *If E/F is a finite Galois extension and $G = \mathrm{Gal}(E/F)$, $\mathrm{Tr}_{E/F}(\alpha) = \sum_{\sigma\in G}\sigma(\alpha)$ and $\mathrm{N}_{E/F}(\alpha) = \prod_{\sigma\in G}\sigma(\alpha)$.*

Proof. We can use $E' = E$ in Theorem 8.18 and the σ_i's are the Galois group of E/F. ■

Corollary 8.20. *If E/F is a finite separable extension and E' is a Galois extension of F containing E then $\mathrm{N}_{E'/F}(\alpha) = \mathrm{N}_{E/F}(\mathrm{N}_{E'/E}(\alpha))$.*

Proof. Let $G = \mathrm{Gal}(E'/F)$, $H = \mathrm{Gal}(E'/E)$, and $\sigma_1, \ldots, \sigma_r$ be left H-coset representatives in G. For $\alpha \in E'$,

$$\mathrm{N}_{E'/F}(\alpha) = \prod_{i=1}^{r}\prod_{\tau\in H}(\sigma_i\tau)(\alpha) = \prod_{i=1}^{r}\sigma_i\left(\prod_{\tau\in H}\tau(\alpha)\right) = \prod_{i=1}^{r}\sigma_i(\mathrm{N}_{E'/E}(\alpha)).$$

By Theorem 8.18, this last product is $\mathrm{N}_{E/F}(\mathrm{N}_{E'/E}(\alpha))$. ■

Corollary 8.21. *If E/F is separable then $\mathrm{Tr}_{E/F}\colon E \to F$ is onto. If E/F is inseparable then $\mathrm{Tr}_{E/F}\colon E \to F$ is identically 0.*

This is an extremely important distinction between separable and inseparable extensions.

Proof. Since $\mathrm{Tr}_{E/F}$ is F-linear with values in F, either it is identically 0 or it is surjective.

Suppose E/F is separable. Let E'/F be a Galois closure of E/F and let $\sigma_1, \ldots, \sigma_n$ be all the F-embeddings of E into E'. Then $\mathrm{Tr}_{E/F}(\alpha) = \sigma_1(\alpha) + \cdots + \sigma_n(\alpha)$ (Theorem 8.18). Viewing the functions $\sigma_i\colon E \to E'$ as characters $E^\times \to E'$, the linear independence of characters tells us the sum $\sum_{i=1}^n \sigma_i$ can't be the zero function on $E^\times$, so $\mathrm{Tr}_{E/F} \not\equiv 0$.

Now suppose E/F is inseparable, so it has positive characteristic p. Any finite extension E/F can be decomposed into a separable extension F'/F followed by a purely inseparable extension E/F' on top of it. The degree $[E : F']$ must be a power of p. When E/F is inseparable, $[E : F'] > 1$. By transitivity of the trace, to show $\mathrm{Tr}_{E/F}$ is identically 0 when E/F is inseparable it suffices to show $\mathrm{Tr}_{E/F'}$ is identically 0. That is, we can assume E/F is purely inseparable and $[E : F] > 1$. Write $[E : F] = p^D$ with $D > 0$.

Each $\alpha \in E$ has minimal polynomial of the form $T^{p^m} - c$ in $F[T]$. Therefore $\chi_{E/F,\alpha}(T) = (T^{p^m} - c)^{p^{D-m}} = T^{p^D} - c^{p^{D-m}}$, so $\mathrm{Tr}_{E/F}(\alpha) = 0$. ■

Example 8.22. For $F = \mathbf{F}_2(u)$ and $E = F(\sqrt{u})$,

$$\mathrm{Tr}_{E/F}(a + b\sqrt{u}) = 2a = 0$$

since $\mathbf{F}_2$ has characteristic 2.

Example 8.23. For $F = \mathbf{F}_p(u)$ and $E = F(\sqrt[p]{u})$, $\mathrm{Tr}_{E/F} \equiv 0$.

8.2 Modules over a PID

We review here structure theorems for submodules of a finite free module over a PID and for finitely generated torsion modules over a PID. Both kinds of modules appear in the description and computation of the integral closure of a PID in a finite separable extension of its fraction field in Chapter 3.

We start with two lemmas that have nothing to do with PIDs.

Lemma 8.24. *Let A be any commutative ring and M be an A-module. If $f\colon M \to A^n$ is linear and onto, then there is an A-module isomorphism $h\colon M \cong A^n \oplus \ker f$ where $h(m) = (f(m), *)$.*

Proof. Let $A^n = Ae_1 \oplus \cdots \oplus Ae_n$ and pick $m_i \in M$ such that $f(m_i) = e_i$. Let $g\colon A^n \to M$ by

$$g(c_1e_1 + \cdots + c_ne_n) = c_1m_1 + \cdots + c_nm_n$$

(i.e., g is A-linear and $g(e_i) = m_i$). So $f(g(v)) = v$ for all $v \in A^n$ (just check at $v = e_i$ and use linearity). Define the function $h\colon M \to A^n \oplus \ker f$ by $h(m) = (f(m), m - g(f(m)))$. Check that h is an isomorphism as Exercise 8.6. ■

Any linear surjection $f\colon M \twoheadrightarrow N$ of A-modules induces an isomorphism $M/\ker f \cong N$ but there is usually not a direct sum decomposition $M \cong N \oplus \ker f$. That there is for $N = A^n$ is a useful property of finite free modules. (It holds more generally when N is a projective module, and in fact this property is one of the definitions of a projective module.)

Lemma 8.25. *Every finitely generated torsion-free module M over a domain A can be embedded in a finite free A-module. More precisely, if $M \neq 0$ there is an embedding $M \hookrightarrow A^d$ for some $d \geqslant 1$ such that the image of M intersects each standard coordinate axis of A^d.*

Proof. Let F be the fraction field of A and $x_1, \ldots, x_n$ be a generating set for M as an A-module. We will show n is an upper bound on the size of any A-linearly independent subset of M. Let $f\colon A^n \to M$ be the linear map where $f(e_i) = x_i$ for all i. (By $e_1, \ldots, e_n$ we mean the standard basis of A^n.) Let $y_1, \ldots, y_k$ be linearly independent in M, so their A-span is isomorphic to A^k. Write $y_j = \sum_{i=1}^n a_{ij} x_i$ with $a_{ij} \in A$. We pull the y_j's back to A^n by setting $v_j = (a_{1j}, \ldots, a_{nj})$, so $f(v_j) = y_j$. A linear dependence relation on the v_j's is transformed by f into a linear dependence relation on the y_j's, which is a trivial relation by their linear independence. Therefore $v_1, \ldots, v_k$ is A-linearly independent in A^n, hence F-linearly independent in F^n. By linear algebra over fields, $k \leqslant n$.

From the bound $k \leqslant n$, there is a linearly independent subset of M with maximal size, say $t_1, \ldots, t_d$. Then $\sum_{j=1}^d At_j \cong A^d$. We will find a scalar multiple of M inside of this. For any $x \in M$, the set $\{x, t_1, \ldots, t_d\}$ is linearly dependent by maximality of d, so there is a nontrivial linear relation $ax + \sum_{i=1}^d a_i t_i = 0$, necessarily with $a \neq 0$. Thus $ax \in \sum_{j=1}^d At_j$. Letting x run through the spanning set $x_1, \ldots, x_n$ there is an $a \in A - \{0\}$ such that $ax_i \in \sum_{j=1}^d At_j$ for all i, so $aM \subset \sum_{j=1}^d At_j$. Multiplying by a is an isomorphism of M with aM, so we have the sequence of A-linear maps

$$M \to aM \hookrightarrow \sum_{j=1}^d At_j \to A^d,$$

where the last map is an isomorphism. ■

Remark 8.26. In Lemma 8.25, we can show the size of a generating set for M as an A-module bounds the size of any linearly independent subset of M using tensor products: $F \otimes_A M$ is finite-dimensional over F since M is finitely generated, and the natural map $M \to F \otimes_A M$ by $m \mapsto 1 \otimes m$ is injective since M is torsion-free. The dimension of the vector space turns out to be the size of any maximal linearly independent subset of M.

Theorem 8.27. *When A is a* PID, *any submodule of a free A-module of rank n is free of rank $\leqslant n$.*

Proof. We may assume the free A-module is A^n and will induct on n. The case where $n = 1$ is clear since A is a PID. Say $n \geqslant 1$ and the theorem is proved for A^n. Let $M \subset A^{n+1}$ be a submodule. We want to show M is free of rank $\leqslant n+1$. View $M \subset A^{n+1} = A \oplus A^n$. Let $\pi\colon A^{n+1} \twoheadrightarrow A^n$ be projection from this direct sum and $N = \pi(M) \subset A^n$. So N is free of rank $\leqslant n$ by the induction hypothesis. Since π maps M onto $N = \pi(M)$ and N is finite free, by Lemma 8.24

$$M \cong N \oplus \ker \pi|_M.$$

But $N \oplus \ker \pi|_M = N \oplus (M \cap A)$. Since A is a PID, $M \cap A$ is 0 or is free of rank 1. So M is free of rank $\leqslant n+1$. ■

Theorem 8.27 is false if A is not a PID, even for the A-module A itself: any nonprincipal ideal in A and any principal ideal xA where $xy = 0$ with $x \neq 0$ and $y \neq 0$ are not free as A-modules.

The rank of a finite free module over any (nonzero) commutative ring A is well-defined: if $A^m \cong A^n$ as A-modules then $m = n$. The simplest proof uses a maximal ideal $\mathfrak{m}$ in A. Setting $M = A^m$ and $N = A^n$, if $M \cong N$ as A-modules then it restricts to an isomorphism $\mathfrak{m}M \cong \mathfrak{m}N$, so $M/\mathfrak{m}M \cong N/\mathfrak{m}N$. This says $(A/\mathfrak{m})^m \cong (A/\mathfrak{m})^n$ as A-modules, hence also as $A/\mathfrak{m}$-vector spaces, so $m = n$ from the well-definedness of dimension for vector spaces.

Corollary 8.28. *When A is a* PID*, every finitely generated torsion-free A-module is a finite free A-module.*

Proof. By Lemma 8.25, such a module embeds into a finite free A-module, so it is finite free too by Theorem 8.27. ■

Corollary 8.29. *Let A be a* PID*. If we have a tower of A-modules $M \subset M' \subset M''$ with $M \cong A^n$ and $M'' \cong A^n$ then $M' \cong A^n$.*

Proof. Since M'' is free of rank n and M' is a submodule, Theorem 8.27 tells us $M' \cong A^m$ with $m \leqslant n$. Using Theorem 8.27 on M as a submodule of M', $M \cong A^k$ with $k \leqslant m$. By hypothesis $M \cong A^n$, so $k = n$. Thus $m = n$. ■

Corollaries 8.28 and 8.29 are both generally false when A is not a PID.

Nonexample 8.30. Let $A = \mathbf{Z}[\sqrt{-5}]$ and consider the tower of ideals $3\mathbf{Z}[\sqrt{-5}] \subset (3, 1+\sqrt{-5}) \subset \mathbf{Z}[\sqrt{-5}]$. The bottom and top are principal ideals, so they are free A-modules of rank 1. The middle ideal $(3, 1+\sqrt{-5})$ is finitely generated and torsion-free, but is not principal and therefore is not a free A-module. (A nonzero ideal is a free module only when it is principal, since any two elements in an ideal are linearly related.)

We want to present a convenient way to picture submodules sitting in a finite free modules over a PID: bases can be chosen for the module and submodule which are aligned nicely, in the following sense.

Definition 8.31. If A is a PID, M is a finite free A-module, and M' is a submodule of M, a basis $\{v_1, \dots, v_n\}$ of M and a basis $\{a_1 v_1, \dots, a_m v_m\}$ of M' with $a_i \in A - \{0\}$ is called a pair of *aligned* bases.

A picture will explain well what alignment of bases means, before we jump into the main theorem about them.

Example 8.32. Let $A = \mathbf{Z}$, $M = \mathbf{Z}[i]$ and take $N = (1+2i)\mathbf{Z}[i]$. So $M = \mathbf{Z}+\mathbf{Z}i$ and

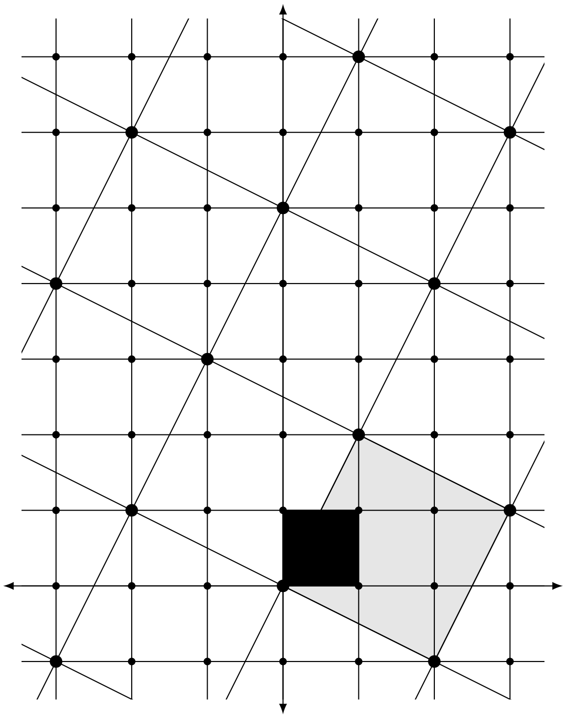
$$N = (1+2i)\mathbf{Z}[i] = (1+2i)\mathbf{Z} + (1+2i)\mathbf{Z}i = \mathbf{Z}(1+2i) + \mathbf{Z}(-2+i).$$

The obvious $\mathbf{Z}$-bases for M and N are $\{1, i\}$ and $\{1+2i, -2+i\}$. In Figure 8.1, we color the fundamental box associated to each basis and translate each box across the plane. The modules M and N are the intersection points of the networks of lines formed by the small and large boxes, respectively. The modules do not know about the lines, which only show us how a choice of basis gives a specific way to picture how the module is generated by the basis.

To see a completely different picture of the same two modules, we use new bases: $\{1+2i, i\}$ for M and $\{1+2i, 5i\}$ for N. These are bases because of the relations

$$\begin{pmatrix} 1+2i \\ i \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad \begin{pmatrix} 1+2i \\ 5i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1+2i \\ -2+i \end{pmatrix},$$

where the two matrices are integral with determinant 1, so the vectors on both sides have the same $\mathbf{Z}$-span. These new bases lead to Figure 8.2, where the fundamental parallelogram for each basis is filled in and looks quite unlike the shaded boxes of Figure 8.1. Translating the parallelograms across the plane produces two new networks of lines (both sharing all the vertical lines) The intersection points are the same as before; make sure you can see the vertices of the large box from Figure 8.1 as intersection points of lines in Figure 8.2. In Figure 8.2 the fundamental parallelograms for M and N actually fit together: five copies of the one for M fill up the one for N. These bases are aligned.



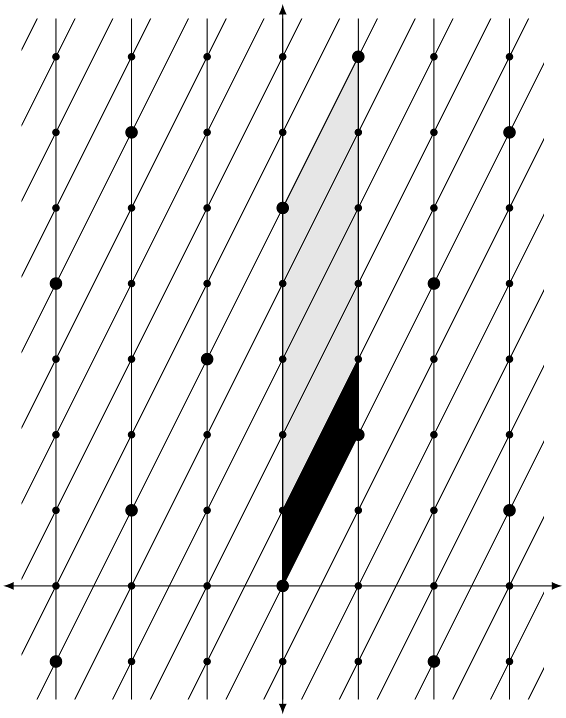
$$\mathbf{Z}[i] = \mathbf{Z}(1+2i) + \mathbf{Z}i, \quad (1+2i) = \mathbf{Z}(1+2i) + \mathbf{Z}(-2+i)$$

Figure 8.1: Nonaligned bases for a lattice and sublattice

Theorem 8.33. *Any finite free A-module M of rank $n \geqslant 1$ and nonzero submodule M' of rank $m \leqslant n$ admit a pair of aligned bases: there is a basis $v_1, \ldots, v_n$ of M and nonzero $a_1, \ldots, a_m \in A$ such that*

$$M = \bigoplus_{i=1}^{n} Av_i \qquad \textit{and} \qquad M' = \bigoplus_{j=1}^{m} Aa_j v_j.$$

Proof. A basis of M gives us coordinate functions for that basis, which are linear maps $M \to A$. Having a set of bases for M and M' as in the theorem means there are a set of compatible coordinate systems on M and M'. To motivate the main idea in the proof, first suppose the theorem is true and let's see what



$$\mathbf{Z}[i] = \mathbf{Z}(1+2i) + \mathbf{Z}i, \quad (1+2i) = \mathbf{Z}(1+2i) + \mathbf{Z}\cdot 5i$$

Figure 8.2: Aligned bases for a lattice and sublattice

it tells us about linear maps $\varphi\colon M \to A$ when they are restricted to M':

$$\varphi(M') = \varphi\Big(\sum_{j=1}^{m} Aa_jv_j\Big) = \sum_{j=1}^{m} Aa_j\varphi(v_j) \in a_1A + \cdots + a_mA.$$

Write $a_1A + \cdots + a_mA = cA$, so $\varphi(M') \subset cA$ for every $\varphi \in M^\vee$. Moreover, writing $c = a_1x_1 + \cdots + a_mx_m$ with $x_j \in A$ and defining $\varphi_c\colon M \to A$ by $\varphi_c(\sum_{i=1}^{n} c_iv_i) = \sum_{j=1}^{m} c_jx_j$, we have $\varphi_c(\sum_{i=1}^{m} a_iv_i) = c$, so

$$\varphi_c(M') = cA.$$

The set of ideals $\varphi(M')$ as φ runs over linear maps $M \to A$ has cA as the unique maximal member for inclusion (this is not saying cA is a maximal ideal!).

Now we start over and define S as the set of ideals $\varphi(M')$ where $\varphi\colon M \to A$

is A-linear. This includes nonzero ideals; for example, let M have A-basis $\{e_1, \dots, e_n\}$, so $M = \bigoplus_{i=1}^n Ae_i$. Some coordinate function for this basis is not identically 0 on M', and the image of that coordinate function on M' is a nonzero ideal in S.

Any nonzero ideal in A is contained in only finitely many ideals, so S contains maximal members with respect to inclusion. Call one of these maximal members a_1A, so $a_1 \neq 0$.[1] We know already that $a_1A = \varphi_1(M')$ for some linear map $\varphi_1 \colon M \to A$. Write

$$a_1 = \varphi_1(v')$$

for $v' \in M'$. Eventually we are going to show φ_1 takes the value 1 on M.

Claim: For any linear map $\varphi \colon M \to A$, $a_1 \mid \varphi(v')$.

To show this, set $\varphi(v') = a_\varphi \in A$. Since A is a PID, $a_1A + a_\varphi A = dA$ for some d, so $a_1A \subset dA$. We have $d = a_1x + a_\varphi y$ for some $x, y \in A$. Then

$$d = x\varphi_1(v') + y\varphi(v') = (x\varphi_1 + y\varphi)(v'),$$

so $dA \subset (x\varphi_1 + y\varphi)(M') \in S$. Hence

$$\varphi_1(M') = a_1A \subset dA \subset (x\varphi_1 + y\varphi)(M').$$

Since $x\varphi_1 + y\varphi$ is a linear map $M \to A$ and $\varphi_1(M')$ is maximal in S,

$$\varphi_1(M') = (x\varphi_1 + y\varphi)(M') = dA = a_1A.$$

Hence $a_1A = a_1A + a_\varphi A$ which implies $a_\varphi \in a_1A$, so $a_1 \mid a_\varphi$.

With the claim proved, we are ready to get our aligned bases in M and M'. Letting $M = \bigoplus_{i=1}^n Ae_i$ for some basis $\{e_1, \dots, e_n\}$, write

$$v' = c_1e_1 + \cdots + c_ne_n$$

for $c_i \in A$. The ith coordinate function for this basis is a linear map $M \to A$ taking the value c_i at v', so c_i is a multiple of a_1 by our claim. Writing $c_i = a_1b_i$,

$$v' = \sum_{i=1}^n c_ie_i = \sum_{i=1}^n a_1b_ie_i = a_1(b_1e_1 + \cdots + b_ne_n) = a_1v_1,$$

[1] We can anticipate a_1A will be the unique maximal member of S, but at the moment it is just some maximal member of S.

say. Then

$$a_1 = \varphi_1(v') = \varphi_1(a_1v_1) = a_1\varphi_1(v_1),$$

so $\varphi_1(v_1) = 1$. There we are: we have found a vector in M at which φ_1 takes the value 1.

We have $M = Av_1 + \ker\varphi_1$ since for any $v \in M$

$$v = \varphi_1(v)v_1 + (v - \varphi_1(v)v_1)$$

and also $Av_1 \cap \ker\varphi_1 = \{0\}$. Thus $M = Av_1 \oplus \ker\varphi_1$. Since M is free of rank n, $\ker\varphi_1$ is free, necessarily of rank $n-1$.

How does M' fit in this decomposition of M?[2] For $w \in M'$, we have

$$w = \varphi_1(w)v_1 + (w - \varphi_1(w)v_1)$$

and the first term is

$$\varphi_1(w)v_1 \in \varphi(M')v_1 = (a_1A)v_1 = Aa_1v_1 = Av' \subset M'.$$

So $w - \varphi_1(w)v_1 \in M'$ also. Therefore

$$M' = \underbrace{(M' \cap Av_1)}_{=Aa_1v_1} \oplus (M' \cap \ker\varphi_1).$$

So $M = Av_1 \oplus \ker\varphi_1$ and $M' = Aa_1v_1 \oplus (M' \cap \ker\varphi_1)$. This last equation tells us $M' \cap \ker\varphi_1$ is free of rank $m-1$ since M' is free of rank m. If $m = 1$ then we're done. If $m > 1$, then we can describe how $M' \cap \ker\varphi_1$ sits in $\ker\varphi_1$ by induction on the rank: we have a basis $v_2, \dots, v_n$ of $\ker\varphi_1$ and $a_2, \dots, a_m \in A - \{0\}$ such that $a_2v_2, \dots, a_mv_m$ is a basis of $M' \cap \ker\varphi_1$. ■

Aligned bases are particularly useful tools when M' has the same rank as M ($m = n$).

Corollary 8.34. *Let A be a* PID*, M be free A-module of rank n, and M' be a submodule of M. Then M' is free of rank n if and only if M/M' is a torsion module.*

[2] Warning: if $M = M_1 \oplus M_2$ and $N \subset M$, usually $N \neq (N \cap M_1) \oplus (N \cap M_2)$. For example, consider $M = \mathbf{Z}^2 = \mathbf{Z}(1,0) \oplus \mathbf{Z}(0,1)$ with $N = \mathbf{Z}(1,1)$.

Proof. The module M' is free of some rank $m \leqslant n$. Write

$$M = \bigoplus_{i=1}^{n} Av_i \qquad \text{and} \qquad M' = \bigoplus_{j=1}^{m} Aa_jv_j.$$

Then $M/M' \cong \bigoplus_{j=1}^{m} A/(a_j) \oplus \bigoplus_{i=m+1}^{n} A$. This is a torsion module if and only if $m = n$. ■

Corollary 8.35. *Let A be a* PID*. Every finitely generated A-module has the form $A^d \oplus T$ where $d \geqslant 0$ and T is a finitely generated torsion module.*

Proof. Let M be a finitely generated A-module, with generators $x_1, \ldots, x_n$. Define $f\colon A^n \twoheadrightarrow M$ by $f(e_i) = x_i$. Then there is a surjective linear map $A^n \twoheadrightarrow M$, so M is isomorphic to a quotient A^n/N. As in the proof of Corollary 8.34, $A^n/N \cong \bigoplus_{j=1}^{m} A/(a_j) \oplus A^{n-m}$ for some $m \leqslant n$. Take $d = n - m$. The direct sum of the $A/(a_j)$'s is a finitely generated torsion module. ■

Corollary 8.36. *Every finitely generated torsion module T over a* PID *A with n generators is a direct sum of cyclic modules: $T \cong A/(a_1) \oplus \cdots \oplus A/(a_n)$, where the a_i's are nonzero.*

Some a_i's might be units, making $A/(a_i) = 0$.

Proof. Take $M = T$ in the proof of Corollary 8.35. ■

8.3 Some PARI commands

The free computer algebra package PARI is designed for computations in number theory. A copy can be downloaded by searching on the internet for "PARI download".

The following list provides some commands in PARI that are useful in algebraic number theory.

Primes and Factoring

`factor(n)` factors the integer n into primes. (This works on rational numbers also and will give prime factorizations with negative exponents.)

`gcd(a,b)` is the greatest common divisor of a and b.

`isprime(n)` returns `1` if n is prime and `0` otherwise.

`prime(n)` returns the nth prime.

`primes(n)` is a vector whose components are the first n primes.

Polynomials

`factor(f(x))` factors $f(x)$ into (monic) irreducibles in $\mathbf{Q}[x]$. (This is the same command as for integers. PARI treats it as a polynomial when there is a variable appearing. If any coefficient has a decimal point then the factorization is done over $\mathbf{C}$.)

`factormod(f(x),p)` factors $f(x) \bmod p$.

`poldisc(f(x))` gives the discriminant of the polynomial $f(x)$.

`polgalois(f(x))` gives the Galois group of the splitting field of $f(x)$ over $\mathbf{Q}$. The output is a vector whose first component is the size of the Galois group and other components describe the group structure.

`polisirreducible(f(x))` returns `1` if $f(x)$ is irreducible in $\mathbf{Q}[x]$ and `0` otherwise.

`polroots(f(x))` is a vector whose components are the roots of $f(x)$ in $\mathbf{C}$.

`subst(F,x,a)` returns the value of F when the variable x in F is replaced by a. Here F can be any algebraic object involving the variable x: a polynomial (in several variables), matrix, vector, and so on.

Linear Algebra

`A = [1,2;4,9]` defines A to be the 2×2 matrix $\left(\begin{smallmatrix}1&2\\4&9\end{smallmatrix}\right)$. Larger matrices can be defined in the same way, using a semicolon to end each row.

`charpoly(A)` is the characteristic polynomial $\det(xI - A)$.

`matdet(A)` is the determinant of the matrix A.

`trace(A)` is the trace of the matrix A.

`v = [1,2,3,6]` defines `v` as a row vector with components 1, 2, 3, and 6.

`v = [1,2,3,6]~` defines `v` as a column vector with components 1, 2, 3, and 6. This is important if you want to multiply a matrix and vector in the usual way, where vectors are written in column form.

`v[n]` is the nth component of the vector `v`. For example, `polroots(f(x))[1]` is the first component of the vector of roots of $f(x)$.

Number Fields

Now $f(x)$ is an irreducible polynomial in $\mathbf{Q}[x]$ and K_f below will denote the number field $\mathbf{Q}(\alpha)$ where α is a root of $f(x)$. Algebraically, this is $\mathbf{Q}[x]/(f(x))$. (If you use a reducible polynomial for $f(x)$, some of the commands will give answers, so make sure your polynomial is irreducible.)

`abs(z)`, `real(z)`, `imag(z)` are the absolute value, real part, and imaginary part of the complex number z.

`algdep(z,n)` is the polynomial in $\mathbf{Z}[x]$ of degree at most n which is most likely to have the complex number z as a root. (If z is not of degree at most n over $\mathbf{Q}$ the answer will be useless, so use of this command requires judgment.)

`bnfclgp(f(x))` gives the ideal class group of K_f. The output is a vector whose first component is the class number and the second component is the cyclic decomposition of the class group.

`bnfinit(f(x))` is a long vector containing information about K_f which is used in unit and class group computations. It is best to assign this a name right away, *e.g.*, `B = bnfinit(f(x));`. The semicolon stops PARI from outputting the data on the screen all at once.

`bnfinit(f(x)).fu` gives the fundamental units of K_f, expressed as polynomials in x mod $f(x)$.

`bnfreg(f(x))` gives the regulator of K_f.

`dirzetak(nfinit(f(x)),N)` gives the coefficients of the first N terms in the Dirichlet series for K_f when it is written as a sum over positive integers. That is, if $\zeta_K(s) = \sum_{n\geqslant 1} a_n/n^s$ then this command returns $[a_1, a_2, \ldots, a_N]$.

`idealfactor(nfinit(f(x)),p)` gives the prime ideal factorization of p in K_f. (If you have already given the vector `nfinit(f(x))` a name, you can use that label in the first component, and you can use `bnfinit(f(x))` there too.) The answer is an array where each row is associated to a different prime ideal. A row has the form `[[p, v, e, f, w] e]`, where e and f are the ramification index and residue field degree for that prime ideal. The vector `v` is related to a second generator γ such that the prime ideal being described is (p, γ) and `w` is related to the inverse of the prime ideal.

`nfbasis(f(x))` gives a $\mathbf{Z}$-basis of K_f.

`nfdisc(f(x))` gives the discriminant of the number field K_f.

`nfinit(f(x))` is a long vector containing information about the number field K_f. It starts off as `[f(x),[r1,r2],d, I,...]` where `r1` and `r2` are r_1 and r_2, `d` is $\mathrm{disc}(K)$, and `I` is the index $[\mathcal{O}_K : \mathbf{Z}[\alpha]]$ for α a root of $f(x)$. As with `bnfinit`, it is best to use this command as a definition, say `v = nfinit(f(x));`.

`nfisincl(f(x),g(x))` is a vector whose components describe the roots of $f(x)$ as polynomials in a root of $g(x)$ if this possible (that is, if the number field defined by a root of $f(x)$ has an embedding into the number field defined by a root of $g(x)$). When $g(x) = f(x)$ and a root of $f(x)$ generates a Galois extension of $\mathbf{Q}$, the output provides formulas for the Galois group acting on a root of $f(x)$.

`nfrootsof1(nfinit(f(x)))` is the number of roots of unity in K_f.

`zetak(zetakinit(f(x)),s)` is $\zeta_{K_f}(s)$, where s is a complex number. (The output may not be accurate if s is unreasonably chosen.)

There are many further commands (*e.g.*, , to add and multiply ideals or test if an ideal is principal), but the above is a basic list to get started.

Note: PARI does arithmetic with fractions exactly. If you want PARI to treat a rational number as a decimal approximation, multiply it by 1.0, *e.g.*, `3/7*1.0`. The numbers π and i are entered in PARI as `Pi` and `I`, and e^z is `exp(z)`. Make sure to include multiplication operations explicitly: $2x$ and $2i$ are entered into PARI as `2*x` and `2*I`, *not* as `2x` and `2I`.

The meaning of any PARI command can be found by typing `?` followed by the command, *e.g.*, `?nfbasis` tells you what `nfbasis` does. Of course this only helps if you know the name of the command. To get a complete list of all PARI commands, type `?`, and a list of the number field commands is `?6`.

For the most part, the PARI commands above receive exact input (*e.g.*, `nfbasis` expects an integral polynomial). The only command where the input is an approximation and the output is expected to be exact, rather than another approximation, is the minimal polynomial command `algdep`. This command produces good answers under reasonable conditions, but when the correct minimal polynomial has very large coefficients there can be errors.

Example 8.37. In PARI, set `f(x) = x^3 + 453603*x^2 + 51438694443*x - 51247953119` and then type `v = polroots(f(x))`. The answer is a vector of length three whose first coordinate is `0.9962831179067027346685176802 + 0.E-28*I`. This is (approximately) the unique real root of $f(x)$. If you now type `algdep(v[1],3)` to find the minimal polynomial over $\mathbf{Q}$ of the first coordinate of v, knowing it should have degree at most 3, the answer is not $f(x)$. Instead it is `287542*x^3 + 101724*x^2 - 365673*x - 21003` (which is very small at that number, roughly 10^{-19}). If you type `algdep(v[1],10)` the answer turns out to be $(x-1)^9(x+1)$, which is wrong in an even worse way.

As practice with these commands, let's run through PARI computations on the quartic field $K = \mathbf{Q}(\sqrt[4]{65})$. (This will be much more meaningful if you download PARI and follow the steps yourself.)

1. What is a $\mathbf{Z}$-basis of $\mathcal{O}_K$?

   Type `nfbasis(x^4-65)` and we get the answer

   ```
   [1, x, 1/2*x^2 + 1/2, 1/4*x^3 + 1/4*x^2 + 1/4*x + 1/4].
   ```

   which means $\{1, \sqrt[4]{65}, \frac{\sqrt{65}+1}{2}, \frac{\sqrt[4]{65}^3+\sqrt{65}+\sqrt[4]{65}+1}{4}\}$ is a $\mathbf{Z}$-basis.

2. What is a polynomial over $\mathbf{Q}$ for the fourth term in the $\mathbf{Z}$-basis?

   Set `r = polroots(x^4-65)[1]`, which is

   ```
   -2.8394115144336774440822629 39 + 0.E-28*I.
   ```

   (This is $-\sqrt[4]{65}$, so setting `r = real(polroots(x^4-65)[1])` would remove the imaginary part.) The command `algdep(r^3/4 + r^2/4 + r/4 + 1/4,4)` returns

   ```
   x^4 - x^3 - 24*x^2 - 256*x - 1024.
   ```

   (It would have been more efficient to set `b = nfbasis(x^4-65);` in the first question and then compute `algdep(subst(b[4],x,r))` to avoiding having to type the polynomial expression in `r` inside `algdep`.)

3. What is the discriminant of K?

   Type `d = nfdisc(x^4 - 65)` and we get

   ```
   -1098500.
   ```

   Its factorization is found with `factor(d)`:

   ```
   [-1 1]

   [2 2]

   [5 3]

   [13 3]
   ```

   which means $\operatorname{disc}(K) = -2^2 \cdot 5^3 \cdot 13^3$. The ramified primes in K are 2, 5, and 13.

4. What is $[\mathcal{O}_K : \mathbf{Z}[\sqrt[4]{65}]]$?

   This can be found in two ways. First, the discriminant of $x^4 - 65$ is found with `poldisc(x^4-65)` and it is

   ```
   -70304000
   ```

whose factorization is $-2^8 \cdot 5^3 \cdot 13^3$. This discriminant divided by disc(K) is $2^6 = 64$, so $[\mathcal{O}_K : \mathbf{Z}[\sqrt[4]{65}]] = 8$. As an alternate solution, the fourth component of `nfinit(x^4-65)` is this index, so we can find it by computing `nfinit(x^4-65)`, which is

```
[x^4 - 65, [2, 1], -10985000, 8, ...]
```

and looking at the fourth component, or directly typing `nfinit(x^4-65)[4]`.

5. What is the shape of the prime ideal factorizations of 2, 3, 5, and 7 in K?

Set `K = nfinit(x^4-65);` and then type `idealfactor(K,2)`. The output is

```
[[2, [-1, 0, 0, 1] , 1, 1, [0, 0, 0, 1] ] 1]

[[2, [0, 1, -1, 0] , 2, 1, [1, 1, 0, 0] ] 2]

[[2, [2, 0, 1, 1] , 1, 1, [1, 0, 1, 1] ] 1]
```

so $2\mathcal{O}_K = \mathfrak{p}_2\mathfrak{p}_2'^2\mathfrak{p}_2''$. Similarly, `idealfactor(K,3)` returns

```
[[3, [0, -1, 2, 0] , 1, 2, [0, 1, -1, 0] ] 1]

[[3, [0, 1, 2, 0] , 1, 2, [0, -1, -1, 0] ] 1]
```

so $3\mathcal{O}_K = \mathfrak{p}_9\mathfrak{p}_9'$. The commands `idealfactor(K,5)` and `idealfactor(K,7)` return

```
[[5, [0, 1, 0, 0] , 4, 1, [2, -1, 2, -1] ] 4]
```

and

```
[[7, [5, 1, 0, 0] , 1, 1, [-2, 3, -1, -3] ] 1]

[[7, [9, 1, 0, 0] , 1, 1, [-1, 3, -2, -3] ] 1]

[[7, [-2, 0, 2, 0] , 1, 2, [-3, 0, 2, 0] ] 1]
```

so $5\mathcal{O}_K = \mathfrak{p}_5^4$ and $7\mathcal{O}_K = \mathfrak{p}_7\mathfrak{p}_7'\mathfrak{p}_{49}$.

For the primes 3 and 7, which don't divide disc$(x^4 - 65)$, we can also obtain the shape of their factorization in K from the degree types of the factorizations of $x^4 - 65$ modulo 3 and 5: `factormod(x^4-65,3)` and `factormod(x^4-65,5)` return

```
[Mod(1, 3)*x^2 + Mod(1, 3)*x + Mod(2, 3) 1]

[Mod(1, 3)*x^2 + Mod(2, 3)*x + Mod(2, 3) 1]
```

and

```
[Mod(1, 7)*x + Mod(2, 7) 1]
```

```
[Mod(1, 7)*x + Mod(5, 7) 1]
[Mod(1, 7)*x^2 + Mod(4, 7) 1].
```

6. What is the class group of K?

   Type `bnfclgp(x^4 - 65)` and we get

   ```
   [4, [2, 2], ...]
   ```

   where only the first two components of the answer are given here. This tells us $h(K) = 4$ and $\mathrm{Cl}(K)$ is a product of two cyclic groups of order 2.

7. What is a system of fundamental units of K?

   Since $r_1 = 2$ and $r_2 = 1$, the unit group has rank 2 by the unit theorem and `bnfinit(x^4 - 65).fu` returns the answer

   ```
   [Mod(x^2 + 8, x^4 - 65),
   Mod(1096*x^3 - 3112*x^2 + 8836*x - 25089, x^4 - 65]
   ```

   which is giving us numbers in $\mathbf{Q}(\sqrt[4]{65})$ as elements of $\mathbf{Q}[x]/(x^4 - 65)$. The unit group (modulo ±1) is generated by $\sqrt{65} + 8$ and $1096\sqrt[4]{65}^3 - 3112\sqrt{65} + 8836\sqrt[4]{65} - 25089$.

8. What is the regulator of K?

   The command `R = bnfreg(x^4-65)` returns

   ```
   63.95045279242670975008629269.
   ```

9. How does $\zeta_K(s)$ begin?

   Recalling that `K = nfinit(x^4-65)`, the command `dirzetak(K,10)` returns

   `[1, 3, 0, 6, 1, 0, 2, 10, 2, 3]` so

   $$\zeta_K(s) = 1 + \frac{3}{2^s} + \frac{6}{4^s} + \frac{1}{5^s} + \frac{2}{7^s} + \frac{10}{8^s} + \frac{2}{9^s} + \frac{3}{10^s} + \cdots .$$

10. What is $\zeta_K(2)$?

    Set `Z = zetakinit(x^4-65);` and then `zetak(Z,2)` has value

    ```
    2.678953090570608822993611623.
    ```

11. Let's numerically check the leading term formulas

    $$\zeta_K(s) = \frac{2^{r_1}(2\pi)^{r_2}hR}{w\sqrt{|\mathrm{disc}(K)|}}\frac{1}{s-1} + \cdots \text{ for } s \text{ near } 1$$

and

$$\zeta_K(s) = -\frac{hR}{w} s^{r_1+r_2-1} + \cdots \text{ for } s \text{ near } 0.$$

Since $r_1 > 0$, $w = 2$ and `2^2*(2*Pi)*4*R/(2*sqrt(abs(d)))` is

```
3.066997687380715729991228380.
```

This is supposed to be the limit of $\zeta_K(s)(s-1)$ as $s \to 1$. Set `g(x) = zetak(Z,x)*(x-1)` and evaluate this for `x` close to 1: `g(1.000001)` is

```
3.066996540925845051253102276
```

which matches the previous computation to 5 digits after the decimal point.

Turning to behavior at $s = 0$, `-4*R/2` has value

```
-127.9009055848534195001725854.
```

This should be the limit of $\zeta_K(s)/s^2$ as $s \to 0$. Set `G(x) = zetak(Z,x)/x^2` and evaluate this at a small value of `x`: `G(1/10^10)` is

```
-127.9009055352882202088191797,
```

which matches the computation of $-hR/w$ to 7 digits after the decimal point.

8.4 A Chronology Until 1927

Below are some key events in the development of algebraic number theory, starting with Fermat. Fermat found no contemporaries with his level of interest in number theory and left almost no record of the proofs of his announced theorems. So Euler, coming a century later, had to start from scratch.

- 1636 (?): Fermat's writes his "last theorem" in margin of Diophantus' *Arithmetica.*

- 1640: Fermat announces his two-square theorem.

- 1654: Fermat describes when $p = x^2 \pm 2y^2$, $p = x^2 + 3y^2$.

- 1657: Fermat challenges English mathematicians to describe all solutions to $x^2 - dy^2 = 1$ and $y^2 = x^3 - 2$ in integers.

- 1749–1759: Euler proves Fermat's theorems on $p = x^2 + y^2$, $p = x^2 + 2y^2$, $p = x^2 + 3y^2$.

- 1768: Lagrange proves Pell's equation has a nontrivial solution.

- 1770: Euler uses $\mathbf{Z}[\sqrt{-2}]$ to solve $y^2 = x^3 - 2$ and $\mathbf{Z}[\sqrt{-3}]$ to prove Fermat's last theorem for exponent 3.

- 1775: Lagrange defines equivalence classes of binary quadratic forms with a fixed discriminant and shows there are finitely many classes.

- 1801: Gauss, in his *Disquisitiones Arithmeticae*, gives the first proof of the quadratic reciprocity law and defines composition on equivalence classes of binary quadratic forms with a fixed discriminant.

- 1831–1832: Gauss develops arithmetic in $\mathbf{Z}[i]$, including unique factorization, to prove biquadratic reciprocity.

- 1837–1847: Kummer studies cyclotomic fields $\mathbf{Q}(\zeta_p)$, discovers nonunique factorization in some $\mathbf{Z}[\zeta_p]$, and introduces "ideal numbers" to restore unique factorization and the ideal class group.

- 1838: Dirichlet uses analysis to obtain formulas for class numbers of quadratic fields.

- 1845: Kronecker's thesis, containing a proof of the unit theorem for cyclotomic fields and an argument that will lead to finiteness of ideal class groups of number fields.

- 1846: Dirichlet proves his unit theorem for rings of the form $\mathbf{Z}[\alpha]$.

- 1857: Dedekind proves quadratic reciprocity in $\mathbf{F}[X]$ for finite fields $\mathbf{F}$ of odd characteristic.

- 1871: Dedekind, in a supplement to the 2nd edition of Dirichlet's *Lectures on Number Theory* (which is refined in later editions), defines ideals in the integers of any number field and proves unique factorization into prime ideals. Equivalence classes of quadratic forms are identified with ideal classes in quadratic fields and Gauss's composition law on those equivalence classes is identified with multiplication of ideal classes. He also defines the zeta-function of a number field.[3]

- 1880: Zolotarev develops semilocal methods of studying number fields.

[3] CHECK: Orders introduced here?

- 1882: Kronecker's memoir on arithmetic in number fields and the polynomials and rational functions over them in several variables; conjectures that if $[K : \mathbf{Q}] > 1$ then $K/\mathbf{Q}$ ramifies at some prime.

- 1882: Dedekind–Weber describe function fields over any algebraically closed field of characteristic 0, from the viewpoint of ideal theory.

- 1891: Minkowski proves Kronecker's discriminant conjecture with convex body theorem.

- 1896: Frobenius introduces Frobenius elements in Galois extensions of $\mathbf{Q}$.

- 1897: Hilbert's *Zahlbericht* appears: quadratic reciprocity law in all number fields, relative extensions, higher ramification groups, points towards class field theory.

- 1897: Hensel introduces p-adic fields to study algebraic numbers using power series expansions, "completing" the ideas of Zolotarev.

- 1897: Weber introduces generalized ideal class groups.

- 1921: Hasse's thesis classifying quadratic forms over $\mathbf{Q}$ using p-adic fields (extended to all number fields in 1924).

- 1922: Chebotarev's thesis on his density theorem for Frobenius elements in Galois groups.

- 1924: Artin's thesis on number theory in quadratic extensions of $\mathbf{F}(X)$: rings of integers, prime ideal factorizations, units, class groups, zeta-functions.

- 1925: F. K. Schmidt's thesis extends Artin's thesis to general finite extensions of $\mathbf{F}(X)$: function fields over a finite field.

- 1927: Noether axiomatizes Dedekind domains.

- 1927: Artin proves his general reciprocity law as an isomorphism of abelian Galois groups of number fields with generalized ideal class groups.

8.5 Exercises

1. Verify Examples 8.5 and 8.6.

2. Let $K = \mathbf{Q}(\alpha)$ where $\alpha^3 + 3\alpha - 6 = 0$.

   a) With respect to the basis $\{1, \alpha, \alpha^2\}$, compute the matrix for multiplication by $x + y\alpha + z\alpha^2$ on K where x, y, and z are rational.

   b) Compute $\mathrm{Tr}_{K/\mathbf{Q}}(x + y\alpha + z\alpha^2)$ for rational x, y, and z. What does this become when $z = 0$?

   c) Compute $\mathrm{N}_{K/\mathbf{Q}}(x + y\alpha + z\alpha^2)$ for rational x, y, and z. What does this become when $z = 0$?

3. Suppose $d \in \mathbf{Z}$ is not a cube. Verify for $x, y, z \in \mathbf{Q}$ that

$$\mathrm{N}_{\mathbf{Q}(\sqrt[3]{d})/\mathbf{Q}}(x + y\sqrt[3]{d} + z\sqrt[3]{d}^2) = x^3 + dy^2 + d^2z^3 - 3dxyz.$$

4. a) If $[K : \mathbf{Q}] = n$ and $\alpha \in K$ has characteristic polynomial $f(T) \in \mathbf{Q}[T]$, for $c \in \mathbf{Q}$ show $\mathrm{N}_{K/\mathbf{Q}}(\alpha - c) = \pm f(c)$. (A common error is to misremember this formula as $\mathrm{N}_{K/\mathbf{Q}}(\alpha + c) = \pm f(c)$.)

   b) In the notation of part a, show more generally for a and b in $\mathbf{Q}$ with $a \neq 0$ that $\mathrm{N}_{K/\mathbf{Q}}(a\alpha - b) = \pm a^n f(b/a)$.

5. Let A be a commutative ring. If every submodule of every finite free A-module is a free A-module, show A is a PID.

6. Let A be a commutative ring, M be an A-module, and $f\colon M \to A^n$ be an A-linear surjection. For the standard basis $e_1, \ldots, e_n$ of A^n, pick $m_i \in M$ such that $f(m_i) = e_i$. Define the linear maps $g\colon A^n \to M$ by $g(e_i) = m_i$ for all i and $h\colon M \to A^n \oplus \ker f$ by

$$h(m) = (f(m), m - g(f(m)))$$

   for all $m \in M$. Show h is an A-module isomorphism. (Hint on surjectivity: for $(x, y) \in A^n \oplus \ker f$ let $m = g(x) + y$. Then $h(m) = (x, y)$.)

7. Suppose A is a PID and π is a prime in A. Inside A^2, set

$$M = A\binom{1}{0} + A\binom{0}{\pi^2} = \left\{\binom{x}{y} : y \equiv 0 \bmod \pi^2\right\}$$

   and

$$N = A\binom{\pi}{0} + A\binom{1}{\pi} = \left\{\binom{x}{y} : y \equiv 0 \bmod \pi, \pi x \equiv y \bmod \pi^2\right\}.$$

a) Find a basis $\{e_1, e_2\}$ of A^2 and a_1 and a_2 in A such that $\{a_1e_1, a_2e_2\}$ is a basis of N. (Such an aligned pairs of bases obviously exists for A^2 and M, by the definition of M.)

b) Show there is no basis $\{e_1, e_2\}$ of A^2 and a_1, a_2, b_1, b_2 in A such that $\{a_1e_1, a_2e_2\}$ is a basis of M and $\{b_1e_1, b_2e_2\}$ is a basis of N. That is, the submodules M and N of A^2 do not admit bases simultaneously aligned with a single basis of A^2.

LIST OF FIGURES

LIST OF TABLES

BIBLIOGRAPHY

[1] S. Alaca and K. S. Williams. *Introductory Algebraic Number Theory.* Cambridge Univ. Press, 2004.

[2] M. F. Atiyah and I. G. MacDonald. *Introduction to Commutative Algebra.* Addison-Wesley, 1969.

[3] A. Baker. Linear forms in the logarithms of algebraic numbers I. *Mathematika*, **13**:204–216, 1966.

[4] M. Bhargava. The density of discriminants of quartic rings and fields. *Ann. of Math.*, **162**:1031–1063, 2005.

[5] M. Bhargava. Higher composition laws and applications. In *Proceedings of the International Congress of Mathematicians, Madrid*, pages 271–294. Eur. Math. Soc., 2006.

[6] Z. I. Borevich and I. R. Shafarevich. *Number Theory.* Academic Press, 1966.

[7] A. Brumer. Ramification and class towers of number fields. *Michigan J. Math.*, **12**:129–131, 1965.

[8] D. Burton. *Elementary Number Theory.* McGraw-Hill, 6th edition, 2007.

[9] L. Claborn. Every abelian group is a class group. *Pacific J. Math.*, **18**:219–222, 1966.

[10] D. A. Clark. A quadratic field which is Euclidean but not norm-Euclidean. *Manuscripta Math.*, **83**:327–330, 1994.

[11] J. Coates and R. Sujatha. *Cyclotomic Fields and Zeta Values.* Springer-Verlag, 2009.

[12] H. Cohen and H. W. Lenstra. Heuristics on class groups of number fields. *Lecture Notes in Mathematics*, **1068**:33–62, 1984.

[13] H. Cohen and J. Martinet. Étude heuristique des groupes de classes des corps de nombres. *J. Reine Angew. Math.*, **404**:39–76, 1990.

[14] D. Cox. *Galois Theory.* Wiley, 2004.

[15] R. Crandall and C. Pomerance. *Prime Numbers: A Computational Perspective.* Springer-Verlag, 2nd edition, 2005.

[16] A. Dubickas and J. Steuding. The Polynomial Pell Equation. *Elem. Math.*, **59**:133–143, 2004.

[17] D. S. Dummit and R. M. Foote. *Abstract Algebra.* John Wiley & Sons, Inc., 3rd edition, 2003.

[18] D. Eisenbud. *Commutative Algebra with a View Towards Algebraic Geometry.* Springer-Verlag, 1995.

[19] G. Faltings. Finiteness theorems for abelian varieties over number fields. In *Arithmetic Geometry*, pages 9–27. Springer-Verlag, 1986.

[20] T. Fukuda and K. Komatsu. Weber's class number problem in the cyclotomic $\mathbf{Z}_2$-extension of $\mathbf{Q}$. *Experimental Mathematics*, **18**:213–222, 2009.

[21] E. S. Golod and I. R. Shafarevich. On the class-field tower. *Izv. Akad. Nauk SSSR, Ser. Mat. (Russian)*, **29**:261–272, 1964.

[22] M. Harper and M. R. Murty. Euclidean rings of algebraic integers. *Canad. J. Math.*, **56**:71–76, 2004.

[23] H. Hasse. *Number Theory.* Springer-Verlag, 1980.

[24] K. Heegner. Diophantintische Analysis und Modulfunktionen. *Math. Z.*, **56**:227–253, 1952.

[25] M. Hindry and J. H. Silverman. *Diophantine Geometry: An Introduction.* Springer-Verlag, 2000.

[26] K. Ireland and M. Rosen. *A Classical Introduction to Modern Number Theory.* Springer-Verlag, 2nd edition, 1990.

[27] N. Jacobson. *Basic Algebra II.* W. H. Freeman and Company, 2nd edition, 1980.

[28] E. Kleinert. Units of Classical Orders: A Survey. *L'Enseignement Math.*, **40**:205–248, 1994.

[29] N. Klingen. *Arithmetical Similarities.* Oxford Univ. Press, 1998.

[30] H. Koch. *Number Theory: Algebraic Numbers and Functions.* Amer. Math. Soc., 2000.

[31] L. Kronecker. De unitatibus complexis. *J. Reine Angew. Math.*, **93**:1–52, 1882.

[32] L. Kronecker. Grundzüge einer arithmetischen Theorie der algebraischen Grössen. *J. Reine Angew. Math.*, **92**:1–122, 1882.

[33] E. Kunz. *Introduction to Commutative Algebra and Algebraic Geometry.* Birkhäuser, 1985.

[34] L. V. Kuzmin. Homologies of profinite groups, the schur multiplier and class field theory. *Izv. Akad. Nauk SSSR, Ser. Mat. (Russian)*, **33**:1220–1254, 1969.

[35] S. Lang. *Algebra.* Springer-Verlag, 3rd edition, 2002.

[36] F. Luca and I. E. Shparlinski. On the Square-free Parts of $\lfloor en! \rfloor$. *Glasgow Math. J.*, **49**:391–403, 2007.

[37] D. Marcus. *Number Fields.* Springer-Verlag, 1977.

[38] H. Matsumura. *Commutative Algebra.* Benjamin-Cummings, 2nd edition, 1980.

[39] H. Matsumura. *Commutative Ring Theory.* Cambridge Univ. Press, 1989.

[40] W. Narkiewicz. *Elementary and Analytic Theory of Algebraic Numbers.* Springer-Verlag, 3rd edition, 2004.

[41] W. Narkiewicz. Euclidean algorithm in small abelian fields. *Funct. Approx. Comment. Math.*, **37**:337–340, 2007.

[42] A. Odlyzko. Some analytic estimates of class numbers and discriminants. *Invent. Math.*, **29**:275–286, 1975.

[43] D. Reed. *Figures of Thought: Mathematics and Mathematical Texts.* Routledge, 1995.

[44] M. Reid. *Undergraduate Algebraic Geometry.* Cambridge Univ. Press, 1989.

[45] P. Ribenboim. *Fermat's Last Theorem for Amateurs.* Springer-Verlag, 2000.

[46] P. Roquette and H. Zassenhaus. A class rank estimate for algebraic number fields. *J. London Math. Soc.*, **44**:31–38, 1969.

[47] M. Rosen. A generalization of Mertens' theorem. *J. Ramanujan Math. Soc.*, **14**:1–19, 1999.

[48] P. Samuel. *Algebraic Theory of Numbers.* Dover, 2008.

[49] R. Schoof. *Catalan's Conjecture.* Springer-Verlag, 2008.

[50] R. E. Schwartz. Lucy and Lily: A Game of Geometry and Number Theory. *Amer. Math. Monthly*, **109**:13–20, 2002.

[51] J-P. Serre. *Local Fields.* Springer-Verlag, 1980.

[52] I. R. Shafarevich. *Basic Algebraic Geometry I.* Springer-Verlag, 2nd edition, 1995.

[53] D. Shanks. On Gauss's class number problems. *Math. Comp.*, **23**:151–163, 1969.

[54] J. E. Shockley. *Introduction to Number Theory.* Holt, Rinehart and Winston, 1967.

[55] C. L. Siegel. *Lectures on the Analytic Theory of Quadratic Forms (1934/35).* Math. Inst. Göttingen, 4th edition, 1995.

[56] H. M. Stark. A complete determination of the complex quadratic fields of class-number one. *Michigan Math. J.*, **14**:1–27, 1967.

[57] I. N. Stewart and D. O. Tall. *Algebraic Number Theory.* Chapman and Hall, 1987.

[58] R. Swan. Factorization of Polynomials over Finite Fields. *Pacific J. Math.*, **12**:1099–1106, 1962.

[59] R. Taylor and A. Wiles. Ring-theoretic properties of certain Hecke algebras. *Ann. of Math.*, **141**:553–572, 1995.

[60] L. Washington. *Introduction to Cyclotomic Fields.* Springer-Verlag, 2nd edition, 1996.

[61] P. J. Weinberger. On Euclidean rings of algebraic integers. In *Proc. of Symposia in Pure Mathematics*, volume **24**, pages 321–332. Amer. Math. Soc., 1973.

[62] A. Wiles. Modular elliptic curves and Fermat's last theorem. *Ann. of Math.*, **141**:443–551, 1995.

INDEX