Chapter 11

Factorization

Rien n'est beau que le vrai.

Hermann Minkowski

1. FACTORIZATION OF INTEGERS AND POLYNOMIALS

This chapter is a study of division in rings. Because it is modeled on properties of the ring of integers, we will begin by reviewing these properties. Some have been used without comment in earlier chapters of the book, and some have already been proved.

The property from which all others follow is division with remainder: If a, b are integers and $a \neq 0$, there exist integers q, r so that

$$b = aq + r, \tag{1.1}$$

and $0 \leq r < |a|$. This property is often stated only for positive integers, but we allow a and b to take on negative values too. That is why we use the absolute value $|a|$ to bound the remainder. The proof of the existence of (1.1) is a simple induction argument.

We've already seen some of the most important consequences of division with remainder, but let us recall them. In Chapter 10, we saw that every subgroup of $\mathbb{Z}^+$ is an ideal and that every ideal of $\mathbb{Z}$ is principal, that is, it has the form $d\mathbb{Z}$ for some integer $d \geq 0$. As was proved in Chapter 2 (2.6), this implies that a greatest common divisor of a pair of integers a, b exists and that it is an integer linear combination of a and b. If a and b have no factor in common other than ± 1, then 1 is a linear combination of a and b with integer coefficients:

$$ra + sb = 1, \tag{1.2}$$

for some $r, s \in \mathbb{Z}$. This implies the fundamental property of prime integers, which was proved in Chapter 3 (2.8). We restate it here:

(1.3) **Proposition.** Let p be a prime integer, and let a, b be integers. If p divides the product ab, then p divides a or b. □

(1.4) **Theorem.** *Fundamental Theorem of Arithmetic:* Every integer $a \neq 0$ can be written as a product

$$a = cp_1 \cdots p_k,$$

where $c = \pm 1$, the p_i are positive prime integers, and $k \geq 0$. This expression is unique except for the ordering of the prime factors.

Proof. First, a prime factorization exists. To prove this, it is enough to consider the case that a is greater than 1. By induction on a, we may assume the existence proved for all positive integers $b < a$. Either a is prime, in which case the product has one factor, or there is a proper divisor $b \neq a$. Then $a = bb'$ and also $b' \neq a$. Both b and b' are smaller than a, and by induction they can be factored into primes. Setting their factorizations side by side gives a factorization of a.

Second, the factorization is unique. Suppose that

$$\pm p_1 \cdots p_n = a = \pm q_1 \cdots q_m.$$

The signs certainly agree. We apply (1.3), with $p = p_1$. Since p_1 divides the product $q_1 \cdots q_m$, it divides some q_i, say q_1. Since q_1 is prime, $p_1 = q_1$. Cancel p_1 and proceed by induction. □

The structure of the ring of integers is closely analogous to that of a polynomial ring $F[x]$ in one variable over a field. Whenever a property of one of these rings is derived, we should try to find an analogous property of the other. We have already discussed division with remainder for polynomials in Chapter 10, and we have seen that every ideal of the polynomial ring $F[x]$ is principal [Chapter 10 (3.21)].

A polynomial $p(x)$ with coefficients in a field F is called *irreducible* if it is not constant and if its only divisors of lower degree in $F[x]$ are constants. This means that the only way that p can be written as a product of two polynomials is $p = cp_1$, where c is a constant and p_1 is a constant multiple of p. The irreducible polynomials are analogous to prime integers. It is customary to normalize them by factoring out their leading coefficients, so that they become monic.

The proof of the following theorem is similar to the proof of the analogous statements for the ring of integers:

(1.5) **Theorem.** Let F be a field, and let $F[x]$ denote the polynomial ring in one variable over F.

(a) If two polynomials f, g have no common nonconstant factor, then there are polynomials $r, s \in F[x]$ such that $rf + sg = 1$.

(b) If an irreducible polynomial $p \in F[x]$ divides a product fg, then p divides one of the factors f or g.

(c) Every nonzero polynomial $f \in F[x]$ can be written as a product

$$f = cp_1 \cdots p_k,$$

where c is a nonzero constant, the p_i are monic irreducible polynomials in $F[x]$, and $k \geq 0$. This factorization is unique, except for the ordering of the terms. □

The constant factor c which appears in the third part of this theorem is analogous to the factor ± 1 in (1.4). These are the units in their respective rings. The unit factors are there because we normalized primes to be positive, and irreducible polynomials to be monic. We can allow negative primes or nonmonic irreducible polynomials if we wish. The unit factor can then be absorbed, if $k > 0$. But this complicates the statement of uniqueness slightly.

(1.6) **Examples.** Over the complex numbers, every polynomial of positive degree has a root α and therefore has a divisor of the form $x - \alpha$. So the irreducible polynomials are linear, and the irreducible factorization of a polynomial has the form

$$f(x) = c(x - \alpha_1) \cdots (x - \alpha_n), \tag{1.7}$$

where α_i are the roots of $f(x)$, repeated as necessary. The uniqueness of this factorization is not surprising.

When $F = \mathbb{R}$, there are two classes of irreducible polynomials: linear polynomials and irreducible quadratic polynomials. A real quadratic polynomial $x^2 + bx + c$ is irreducible if and only if its discriminant $b^2 - 4c$ is negative, in which case it has a pair of complex conjugate roots. The fact that every irreducible polynomial over the complex numbers is linear implies that no higher-degree polynomial is irreducible over the reals. Suppose that a polynomial $f(x)$ has real coefficients a_i and that α is a complex, nonreal root of $f(x)$. Then the complex conjugate $\overline{\alpha}$ is different from α and is also a root. For, since f is a real polynomial, its coefficients a_i satisfy the relation $a_i = \overline{a}_i$. Then

$$f(\overline{\alpha}) = a_n\overline{\alpha}^n + \cdots + a_1\overline{\alpha} + a_0 = \overline{a}_n\overline{\alpha}^n + \cdots + \overline{a}_1\overline{\alpha} + \overline{a}_0 = \overline{f(\alpha)} = \overline{0} = 0.$$

The quadratic polynomial $g(x) = (x - \alpha)(x - \overline{\alpha}) = x^2 - (\alpha + \overline{\alpha})x + \alpha\overline{\alpha}$ has real coefficients $-(\alpha + \overline{\alpha})$ and $\alpha\overline{\alpha}$, and both of its linear factors appear on the right side of the complex factorization (1.7) of $f(x)$. Thus $g(x)$ divides $f(x)$. So the factorization of $f(x)$ into irreducible real polynomials is obtained by grouping conjugate pairs in the complex factorization. □

Factorization of polynomials is more complicated for polynomials with rational coefficients than for real or complex polynomials, because there exist irreducible polynomials in $\mathbb{Q}[x]$ of arbitrary degree. For example, $x^5 - 3x^4 + 3$ is irreducible in $\mathbb{Q}[x]$. We will see more examples in Section 4. Neither the form of the irreducible factorization nor its uniqueness is intuitively clear for rational polynomials.

For future reference, we note the following elementary fact:

(1.8) **Proposition.** Let F be a field, and let $f(x)$ be a polynomial of degree n with coefficients in F. Then f has at most n roots in F.

Proof. An element $\alpha \in F$ is a root of f if and only if $x - \alpha$ divides f [Chapter 10 (3.20)]. If so, then we can write $f(x) = (x - \alpha)q(x)$, where $q(x)$ is a polynomial of degree $n - 1$. If β is another root of f, then $f(\beta) = (\beta - \alpha)q(\beta) = 0$. Since F is a field, the product of nonzero elements of F is not zero. So one of the two elements $\beta - \alpha$, $q(\beta)$ is zero. In the first case $\beta = \alpha$, and in the second case β is one of the roots of $q(x)$. By induction on n, we may assume that $q(x)$ has at most $n - 1$ roots in F. Then there are at most n possibilities for β. □

The fact that F is a field is crucial to Theorem (1.5) and to Proposition (1.8), as the following example shows. Let R be the ring $\mathbb{Z}/8\mathbb{Z}$. Then in the polynomial ring $R[x]$, we have

$$x^2 - 1 = (x + 1)(x - 1) = (x + 3)(x - 3).$$

The polynomial $x^2 - 1$ has four roots modulo 8, and its factorization into irreducible polynomials is not unique.

2. *UNIQUE FACTORIZATION DOMAINS, PRINCIPAL IDEAL DOMAINS, AND EUCLIDEAN DOMAINS*

Having seen that factorization of polynomials is analogous to factorization of integers, it is natural to ask whether other rings can have such properties. Relatively few such rings exist, but the ring of Gauss integers is one interesting example. This section explores ways in which various parts of the theory can be extended.

We begin by introducing the terminology used in studying factorization. It is natural to assume that the given ring R is an integral domain, so that the Cancellation Law is available, and we will make this assumption throughout. We say that an element a *divides* another element b (abbreviated $a \mid b$) if $b = aq$ for some $q \in R$. The element a is a *proper divisor* of b if $b = aq$ for some $q \in R$ and if neither a nor q is a unit. A nonzero element a of R is called *irreducible* if it is not a unit and if it has no proper divisor. Two elements a, a' are called *associates* if each divides the other. It is easily seen that a, a' are associates if and only if they differ by a unit factor, that is, if $a' = ua$ for some unit u.

The concepts of divisor, unit, and associate can be interpreted in terms of the principal ideals generated by the elements. Remember that an ideal I is called *principal* if it is generated by a single element:

$$I = (a). \tag{2.1}$$

Keep in mind the fact that (a) consists of all elements which are multiples of a, that is, which are divisible by a. Then

(2.2)
$$u \text{ is a unit} \Leftrightarrow (u) = (1)$$
$$a \text{ and } a' \text{ are associates} \Leftrightarrow (a) = (a')$$
$$a \text{ divides } b \Leftrightarrow (a) \supset (b)$$
$$a \text{ is a proper divisor of } b \Leftrightarrow (1) > (a) > (b).$$

The proof of these equivalences is straightforward, and we omit it.

Now suppose that we hope for a theorem analogous to the Fundamental Theorem of Arithmetic in an integral domain R. We may divide the statement of the theorem into two parts. First, a given element a must be a product of irreducible elements, and second, this product must be essentially unique.

Consider the first part. We assume that our element a is not zero and not a unit; otherwise we have no hope of writing it as a product of irreducible elements. Then we attempt to factor a, proceeding as follows: If a is irreducible itself, we are done. If not, then a has a proper factor, so it decomposes in some way as a product, $a = a_1 b_1$, where neither a_1 nor b_1 is a unit. We continue factoring a_1 and b_1 if possible, and we hope that this procedure terminates; in other words, we hope that after a finite number of steps all the factors are irreducible. The condition that this procedure always terminates has a neat description in terms of principal ideals:

(2.3) **Proposition.** Let R be an integral domain. The following conditions are equivalent:

(a) For every nonzero element a of R which is not a unit, the process of factoring a terminates after finitely many steps and results in a factorization $a = b_1 \cdots b_k$ of a into irreducible elements of R.

(b) R does not contain an infinite increasing chain of principal ideals
$$(a_1) < (a_2) < (a_3) < \dots .$$

Proof. Suppose that R contains an infinite increasing sequence $(a_1) < (a_2) < \dots$. Then $(a_n) < (1)$ for every n, because $(a_n) < (a_{n+1}) \subset (1)$. Since $(a_{n-1}) < (a_n)$, a_n is a proper divisor of a_{n-1}, say $a_{n-1} = a_n b_n$ where a_n, b_n are not units. This provides a nonterminating sequence of factorizations of a_1: $a_1 = a_2 b_2 = a_3 b_3 b_2 = a_4 b_4 b_3 b_2 \dots$. Conversely, such a sequence of factorizations gives us an increasing chain of ideals. □

The second condition of this proposition is often called the *ascending chain condition* for principal ideals. However, to emphasize the factorization property, we will say that *existence of factorizations* holds in R if the equivalent conditions of the proposition are true.

It is easy to describe domains in which existence of factorizations fails. One example is obtained by adjoining all 2^k-th roots of x_1 to the polynomial ring $F[x_1]$:

(2.4)
$$R = F[x_1, x_2, x_3, \dots],$$

with the relations $x_2^2 = x_1$, $x_3^2 = x_2$, $x_4^2 = x_3$, and so on. We can factor the element x_1 indefinitely in this ring, and correspondingly there is an infinite chain $(x_1) < (x_2) < \dots$ of principal ideals.

It turns out that we need infinitely many generators for a ring to make an example such as the one just given, so we will rarely encounter such rings. In practice, the second part of the Fundamental Theorem is the one which gives the most trouble. Factorization into irreducible elements will usually be possible, but it will not be unique.

Units in a ring complicate the statement of uniqueness. It is clear that unit factors should be disregarded, since there is no end to the possibility of adding unit factors in pairs uu^{-1}. For the same reason, *associate* factors should be considered equivalent. The units in the ring of integers are ± 1, and in this ring it was natural to normalize irreducible elements (primes) to be positive; similarly, we may normalize irreducible polynomials by insisting that they be monic. We don't have a reasonable way to normalize elements of an arbitrary integral domain, so we will allow some ambiguity. It is actually neater to work with *principal ideals* than with elements: Associates generate the same principal ideal. However, it isn't too cumbersome to use elements here, and we will stay with them. The importance of ideals will become clear in the later sections of this chapter.

We will call an integral domain R a *unique factorization domain* if it has the following properties:

(2.5)

(i) Existence of factorizations is true for R. In other words, the process of factoring a nonzero element a which is not a unit terminates after finitely many steps and yields a factorization $a = p_1 \cdots p_m$, where each p_i is irreducible.

(ii) The irreducible factorization of an element is unique in the following sense: If a is factored in two ways into irreducible elements, say $a = p_1 \cdots p_m = q_1 \cdots q_n$, then $m = n$, and with suitable ordering of the factors, p_i is an associate of q_i for each i.

So in the statement of uniqueness, associate factorizations are considered equivalent.

Here is an example in which uniqueness of factorization is not true. The ring is the integral domain

$$R = \mathbb{Z}[\sqrt{-5}]. \tag{2.6}$$

It consists of all complex numbers of the form $a + b\sqrt{-5}$, where $a, b \in \mathbb{Z}$. The units in this ring are ± 1, and the integer 6 has two essentially different factorizations in R:

$$6 = 2 \cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5}). \tag{2.7}$$

It is not hard to show that all four terms $2, 3, 1+\sqrt{-5}, 1-\sqrt{-5}$ are irreducible elements of R. Since the units are ± 1, the associates of 2 are 2 and -2. So 2 is not an associate of $1 \pm \sqrt{-5}$, which shows that the two factorizations are essentially different and hence that R is not a unique factorization domain.

The crucial property of prime integers is that if a prime divides a product, it divides one of the factors. We will call an element p of an integral domain R *prime* if it has these properties: p is not zero and not a unit, and if p divides a product of elements of R, it divides one of the factors. These are the properties from which uniqueness of the factorization is derived.

(2.8) **Proposition.** Let R be an integral domain. Suppose that existence of factorizations holds in R. Then R is a unique factorization domain if and only if every irreducible element is prime.

The proof is a simple extension of the arguments used in (1.3) and (1.4); we leave it as an exercise. □

It is important to distinguish between the two concepts of irreducible element and prime element. They are equivalent in unique factorization domains, but most rings contain irreducible elements which are not prime. For instance, in the ring $R = \mathbb{Z}[\sqrt{-5}]$ considered above, the element 2 has no proper factor, so it is irreducible. It is not prime because, though it divides the product $6 = (1+\sqrt{-5})(1-\sqrt{-5})$, it does not divide either factor.

Since irreducible elements in a unique factorization domain are prime, the phrases *irreducible factorization* and *prime factorization* are synonymous. We can use them interchangeably when we are working in a unique factorization domain, but not otherwise.

There is a simple way of deciding whether an element a divides another element b in a unique factorization domain, in terms of their irreducible (or prime) factorizations.

(2.9) **Proposition.** Let R be a unique factorization domain, and let $a = p_1 \cdots p_r$, $b = q_1 \cdots q_s$ be given prime factorizations of two elements of R. Then a divides b in R if and only if $s \geq r$, and with a suitable ordering of the factors q_i of b, p_i is an associate of q_i for $i = 1, \ldots, r$. □

(2.10) **Corollary.** Let R be a unique factorization domain, and let a, b be elements of R which are not both zero. There exists a *greatest common divisor* d of a, b, with the following properties:

(i) d divides a and b;

(ii) if an element e of R divides a and b, then e divides d. □

It follows immediately from the second condition that any two greatest common divisors of a, b are associates. However, *the greatest common divisor need not have the form* $ra + sb$. For example, we will show in the next section that the integer polynomial ring $\mathbb{Z}[x]$ is a unique factorization domain [see (3.8)]. In this ring, the elements 2 and x have greatest common divisor 1, but 1 is not a linear combination of these elements with integer polynomial coefficients.

Another important property of the ring of integers is that every ideal of $\mathbb{Z}$ is principal. An integral domain in which every ideal is principal is called a *principal ideal domain*.

(2.11) **Proposition.**

(a) In an integral domain, a prime element is irreducible.

(b) In a principal ideal domain, an irreducible element is prime.

We leave the proofs of (2.9–2.11) as exercises. □

(2.12) **Theorem.** A principal ideal domain is a unique factorization domain.

Proof. Suppose that R is a principal ideal domain. Then every irreducible element of R is prime. So according to Proposition (2.8), we need only prove the existence of factorizations for R. By Proposition (2.3), this is equivalent to showing that R contains no infinite increasing chain of principal ideals. We argue by contradiction. Suppose that $(a_1) < (a_2) < (a_3) < \dots$ is such a chain.

(2.13) **Lemma.** Let R be any ring. The union of an increasing chain of ideals $I_1 \subset I_2 \subset I_3 \subset \dots$ is an ideal.

Proof. Let I denote the union of the chain. If u, v are in I, then they are in I_n for some n. Then $u + v$ and ru are also in I_n; hence they are in I. □

We apply this lemma to the union I of our chain of principal ideals and use the hypothesis that R is a principal ideal domain to conclude that I is principal, say $I = (b)$. Now since b is in the union of the ideals (a_n), it is in one of these ideals. But if $b \in (a_n)$, then $(b) \subset (a_n)$, and on the other hand $(a_n) \subset (a_{n+1}) \subset (b)$. Therefore $(a_n) = (a_{n+1}) = (b)$. This contradicts the assumption that $(a_n) < (a_{n+1})$, and this contradiction completes the proof. □

The converse of Theorem (2.12) is not true. The ring $\mathbb{Z}[x]$ of integer polynomials is a unique factorization domain [see (3.8)], but it is not a principal ideal domain.

(2.14) **Proposition.**

(a) Let p be a nonzero element of a principal ideal domain R. Then $R/(p)$ is a field if and only if p is irreducible.

(b) The maximal ideals are the principal ideals generated by irreducible elements.

Proof. Since an ideal M is maximal if and only if R/M is a field, the two parts are equivalent. We will prove the second part. A principal ideal (a) contains another principal ideal (b) if and only if a divides b. The only divisors of an irreducible element p are the units and the associates of p. Therefore the only principal ideals which contain (p) are (p) and (1). Since every ideal of R is principal, this shows that an irreducible element generates a maximal ideal. Conversely, let b be a polynomial

having a proper factorization $b = aq$, where neither a nor q is a unit. Then $(b) < (a) < (1)$, and this shows that (b) is not maximal. □

Let us now abstract the procedure of division with remainder. To do so, we need a notion of *size* of an element of a ring. Appropriate measures are

(2.15) *absolute value*, if $R = \mathbb{Z}$,

degree of a polynomial, if $R = F[x]$,

(absolute value)^2, if $R = \mathbb{Z}[i]$.

In general, a *size function* on an integral domain R will be any function

(2.16) $$\sigma: R - \{0\} \longrightarrow \{0, 1, 2, \ldots\}$$

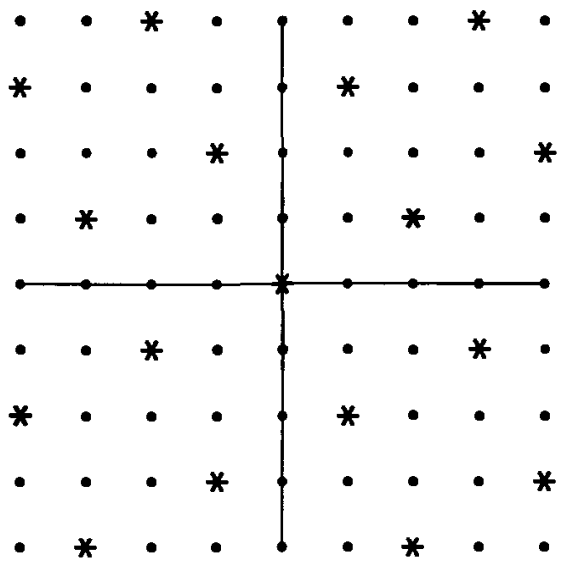
from the set of nonzero elements of R to the nonnegative integers. An integral domain R is a *Euclidean domain* if there is a size function σ on R such that the division algorithm holds:

(2.17) *Let $a, b \in R$ and suppose that $a \neq 0$. There are elements $q, r \in R$ such that $b = aq + r$, and either $r = 0$ or $\sigma(r) < \sigma(a)$.*

We do not require the elements q, r to be uniquely determined by a and b.

(2.18) **Proposition.** The rings $\mathbb{Z}$, $F[x]$, and $\mathbb{Z}[i]$ are Euclidean domains. □

The ring of integers and the polynomial ring have already been discussed. Let us show that the ring of Gauss integers is a Euclidean domain, with size function the function $\sigma = |\ \ |^2$. The elements of $\mathbb{Z}[i]$ form a square lattice in the complex plane, and the multiples of a given element a form a *similar lattice*, the ideal $(a) = Ra$. If we write $a = re^{i\theta}$, then (a) is obtained by rotating through the angle θ followed by stretching by the factor $r = |a|$:



(2.19) **Figure.** $*$ = ideal (a), R = $\mathbb{Z}[i]$

It is clear that for every complex number b, there is at least one point of the lattice (a) whose square distance from b is $\leq \frac{1}{2}|a|^2$. Let that point be aq, and set $r = b - aq$. Then $|r|^2 \leq \frac{1}{2}|a|^2 < |a|^2$, as required. Note that since there may be more than one choice for the element aq, this division with remainder is not unique.

We could also proceed algebraically. We divide the complex number b by a: $b = aw$, where $w = x + yi$ is a complex number, not necessarily a Gauss integer. Then we choose the nearest Gauss integer point (m, n) to (x, y), writing $x = m + x_0$, $y = n + y_0$, where m, n are integers and x_0, y_0 are real numbers such that $-\frac{1}{2} \leq x_0, y_0 < \frac{1}{2}$. Then $(m + ni)a$ is the required point of Ra. For, $|x_0 + y_0 i|^2 < \frac{1}{2}$ and $|b - (m + ni)a|^2 = |a(x_0 + y_0 i)|^2 < \frac{1}{2}|a|^2$.

One can copy the discussion of factorization of integers with minor changes to prove this proposition:

(2.20) **Proposition.** A Euclidean domain is a principal ideal domain, and hence it is a unique factorization domain. □

(2.21) **Corollary.** The rings $\mathbb{Z}$, $\mathbb{Z}[i]$, and $F[x]$ (F a field) are principal ideal domains and unique factorization domains. □

In the ring $\mathbb{Z}[i]$ of Gauss integers, the element 3 is irreducible, hence prime, but 2 and 5 are not irreducible because

$$2 = (1 + i)(1 - i) \quad \text{and} \quad 5 = (2 + i)((2 - i). \tag{2.22}$$

These are the prime factorizations of 2 and 5 in $\mathbb{Z}[i]$.

There are four units in the ring $\mathbb{Z}[i]$, namely $\{\pm 1, \pm i\}$. So every nonzero element α of this ring has four associates, namely the elements $\pm\alpha$, $\pm i\alpha$. The associates of $2 + i$, for example are

$$2 + i, \quad -2 - i, \quad -1 + 2i, \quad 1 - 2i.$$

There is no really natural way to normalize primes in $\mathbb{Z}[i]$, though if pressed we would choose the unique associate lying in the first quadrant and not on the imaginary axis. It is better to accept the ambiguity of (2.5) here or else work with principal ideals.

3. GAUSS'S LEMMA

Theorem (1.5) applies to the ring $\mathbb{Q}[x]$ of polynomials with rational coefficients: Every polynomial $f(x) \in \mathbb{Q}[x]$ can be expressed uniquely in the form $cp_1 \cdots p_k$, where $c \in \mathbb{Q}$ and p_i are monic polynomials which are irreducible over $\mathbb{Q}$. Now suppose that a polynomial $f(x)$ has integer coefficients, $f(x) \in \mathbb{Z}[x]$, and that it factors in $\mathbb{Q}[x]$. Can it be factored without leaving $\mathbb{Z}[x]$? We are going to prove that it can, and that $\mathbb{Z}[x]$ is a unique factorization domain.

Here is an example of a prime factorization in $\mathbb{Z}[x]$:

$$6x^3 + 9x^2 + 9x + 3 = 3(2x + 1)(x^2 + x + 1).$$

As we see from this example, irreducible factorizations are slightly more complicated in $\mathbb{Z}[x]$ than in $\mathbb{Q}[x]$. First, the prime integers are irreducible elements of $\mathbb{Z}[x]$, so they may appear in the prime factorization of a polynomial. Second, the factor $2x + 1$ isn't monic. If we want to stay with integer coefficients, we can't ask for monic factors.

The integer factors of a polynomial $f(x) = a_n x^n + \cdots + a_0$ in $\mathbb{Z}[x]$ are common divisors of its coefficients $a_0, \ldots, a_n$. A polynomial $f(x)$ is called *primitive* if its coefficients $a_0, \ldots, a_n$ have no common integer factor except for the units ± 1 and if its highest coefficient a_n is positive.

(3.1) **Lemma.** Every nonzero polynomial $f(x) \in \mathbb{Q}[x]$ can be written as a product

$$f(x) = cf_0(x),$$

where c is a rational number and $f_0(x)$ is a primitive polynomial in $\mathbb{Z}[x]$. Moreover, this expression for f is unique. The polynomial f has integer coefficients if and only if c is an integer. If so, then $|c|$ is the greatest common divisor of the coefficients of f, and the sign of c is the sign of the leading coefficient of f.

The rational number c which appears in this lemma is called the *content* of $f(x)$. If f has integer coefficients, then the content divides f in $\mathbb{Z}[x]$. Also, f is primitive if and only if its content is 1.

Proof of the Lemma. To find f_0, we first multiply f by an integer to clear the denominators in its coefficients. This will give us a polynomial f_1 with integer coefficients. Then we factor out the greatest common divisor of the coefficients of f_1 and adjust the sign of the leading coefficient. The resulting polynomial f_0 is primitive, and $f = cf_0$ for some rational number c. This proves existence.

To prove uniqueness, suppose that $cf_0(x) = dg_0(x)$, where $c, d \in \mathbb{Q}$ and f_0, g_0 are primitive polynomials. We will show that $c = d$ and $f_0 = g_0$. Clearing denominators reduces us to the case that c and d are integers. Let $\{a_i\}, \{b_i\}$ denote the coefficients of f_0, g_0 respectively. Then $ca_i = db_i$ for all i. Since the greatest common divisor of $\{a_0, \ldots, a_n\}$ is 1, c is the greatest common divisor of $\{ca_0, \ldots, ca_n\}$. Similarly, d is the greatest common divisor of $\{db_0, \ldots, db_n\} = \{ca_0, \ldots, ca_n\}$. Hence $c = \pm d$ and $f_0 = \pm g_0$. Since f_0 and g_0 have positive leading coefficients, $f_0 = g_0$ and $c = d$. If f has integer coefficients, clearing of the denominator is not necessary; hence c is an integer, and up to sign it is the greatest common divisor of the coefficients, as stated. □

As we have already observed, the Substitution Principle gives us a homomorphism

$$\mathbb{Z}[x] \longrightarrow \mathbb{F}_p[x], \tag{3.2}$$

where $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ is the field with p elements. This homomorphism sends a polynomial $f(x) = a_m x^m + \cdots + a_0$ to its residue $\bar{f}(x) = \bar{a}_m x^m + \cdots + \bar{a}_0$ modulo p. We will now use it to prove Gauss's Lemma.

(3.3) **Theorem.** *Gauss's Lemma:* A product of primitive polynomials in $\mathbb{Z}[x]$ is primitive.

Proof. Let the polynomials be f and g, and let $h = fg$ be their product. Since the leading coefficients of f and g are positive, the leading coefficient of h is, too. To show that h is primitive, we will show that no prime integer p divides all the coefficients of $h(x)$. This will show that the content of h is 1. Consider the homomorphism $\mathbb{Z}[x] \longrightarrow \mathbb{F}_p[x]$ defined above. We have to show that $\bar{h} \neq 0$. Since f is primitive, its coefficients are not all divisible by p. So $\bar{f} \neq 0$. Similarly, $\bar{g} \neq 0$. Since the polynomial ring $\mathbb{F}_p[x]$ is an integral domain, $\bar{h} = \bar{f}\bar{g} \neq 0$, as required. □

(3.4) **Proposition.**

(a) Let f, g be polynomials in $\mathbb{Q}[x]$, and let f_0, g_0 be the associated primitive polynomials in $\mathbb{Z}[x]$. If f divides g in $\mathbb{Q}[x]$, then f_0 divides g_0 in $\mathbb{Z}[x]$.

(b) Let f be a primitive polynomial in $\mathbb{Z}[x]$, and let g be any polynomial with integer coefficients. Suppose that f divides g in $\mathbb{Q}[x]$, say $g = fq$, with $q \in \mathbb{Q}[x]$. Then $q \in \mathbb{Z}[x]$, and hence f divides g in $\mathbb{Z}[x]$.

(c) Let f, g be polynomials in $\mathbb{Z}[x]$. If they have a common nonconstant factor in $\mathbb{Q}[x]$, then they have a common nonconstant factor in $\mathbb{Z}[x]$ too.

Proof. To prove (a), we may clear denominators so that f and g become primitive. Then (a) is a consequence of (b). To prove (b), we apply (3.1) in order to write the quotient in the form $q = cq_0$, where q_0 is primitive and $c \in \mathbb{Q}$. By Gauss's Lemma, fq_0 is primitive, and the equation $g = cfq_0$ shows that it is the primitive polynomial g_0 associated to g. Therefore $g = cg_0$ is the expression for g referred to in Lemma (3.1), and c is the content of g. Since $g \in \mathbb{Z}[x]$, it follows that $c \in \mathbb{Z}$, hence that $q \in \mathbb{Z}[x]$. Finally, to prove (c), suppose that f, g have a common factor h in $\mathbb{Q}[x]$. We may assume that h is primitive, and then by (b) h divides both f and g in $\mathbb{Z}[x]$. □

(3.5) **Corollary.** If a nonconstant polynomial f is irreducible in $\mathbb{Z}[x]$, then it is irreducible in $\mathbb{Q}[x]$. □

(3.6) **Proposition.** Let f be an integer polynomial with positive leading coefficient. Then f is irreducible in $\mathbb{Z}[x]$ if and only if either

(i) f is a prime integer, or

(ii) f is a primitive polynomial which is irreducible in $\mathbb{Q}[x]$.

Proof. Suppose that f is irreducible. As in Lemma (3.1), we may write $f = cf_0$, where f_0 is primitive. Since f is irreducible, this can not be a proper factorization. So either c or f_0 is 1. If $f_0 = 1$, then f is constant, and to be irreducible, a constant polynomial must be a prime integer. If $c = 1$, then f is primitive, and is irreducible in $\mathbb{Q}[x]$ by the previous corollary. The converse, that integer primes and primitive irreducible polynomials are irreducible elements of $\mathbb{Z}[x]$, is clear. □

(3.7) **Proposition.** Every irreducible element of $\mathbb{Z}[x]$ is a prime element.

Proof. Let f be irreducible, and suppose f divides gh, where $g, h \in \mathbb{Z}[x]$.

Case 1: $f = p$ is a prime integer. Write $g = cg_0$ and $h = dh_0$ as in (3.1). Then g_0h_0 is primitive, and hence some coefficient a of g_0h_0 is not divisible by p. But since p divides gh, the corresponding coefficient, which is cda, is divisible by p. Hence p divides c or d, so p divides g or h.

Case 2: f is a primitive polynomial which is irreducible in $\mathbb{Q}[x]$. By (2.11b), f is a prime element of $\mathbb{Q}[x]$. Hence f divides g or h in $\mathbb{Q}[x]$. By (3.4), f divides g or h in $\mathbb{Z}[x]$. □

(3.8) **Theorem.** The polynomial ring $\mathbb{Z}[x]$ is a unique factorization domain. Every nonzero polynomial $f(x) \in \mathbb{Z}[x]$ which is not ± 1 can be written as a product

$$f(x) = \pm p_1 \cdots p_m \, q_1(x) \cdots q_n(x),$$

where the p_i are prime integers and the $q_i(x)$ are irreducible primitive polynomials. This expression is unique up to arrangement of the factors.

Existence of factorizations is easy to prove for $\mathbb{Z}[x]$, so this theorem follows from Propositions (3.7) and (2.8). □

Now let R be any unique factorization domain, and let F be its field of fractions [Chapter 10 (6.5)]. Then $R[x]$ is a subring of $F[x]$, and the results of this section can be copied, replacing $\mathbb{Z}$ by R and $\mathbb{Q}$ by F throughout. The only change to be made is that instead of normalizing primitive polynomials it is better to allow ambiguity caused by unit factors, as in the previous section. The main results are these:

(3.9) **Theorem.** Let R be a unique factorization domain with field of fractions F.

(a) Let f, g be polynomials in $F[x]$, and let f_0, g_0 be the associated primitive polynomials in $R[x]$. If f divides g in $F[x]$, then f_0 divides g_0 in $R[x]$.

(b) Let f be a primitive polynomial in $R[x]$, and let g be any polynomial in $R[x]$. Suppose that f divides g in $F[x]$, say $g = fq$, with $q \in F[x]$. Then $q \in R[x]$, and hence f divides g in $R[x]$.

(c) Let f, g be polynomials in $R[x]$. If they have a common nonconstant factor in $F[x]$, then they have a common nonconstant factor in $R[x]$ too.

(d) If a nonconstant polynomial f is irreducible in $R[x]$, then it is irreducible in $F[x]$.

(e) $R[x]$ is a unique factorization domain.

The proof of Theorem (3.9) follows the pattern established for the ring $\mathbb{Z}[x]$, and we omit it. □

Since $R[x_1, \ldots, x_n] \approx R[x_1, \ldots, x_{n-1}][x_n]$, we obtain this corollary:

(3.10) **Corollary.** The polynomial rings $\mathbb{Z}[x_1, \ldots, x_n]$ and $F[x_1, \ldots, x_n]$, where F is a field, are unique factorization domains. □

So the ring $\mathbb{C}[x, y]$ of complex polynomials in two variables is a unique factorization domain. In contrast to the case of one variable, however, where every complex polynomial is a product of linear ones, complex polynomials in two variables are often irreducible, and hence prime.

The irreducibility of a polynomial $f(x, y)$ can sometimes be proved by studying the locus $W = \{f(x, y) = 0\}$ in $\mathbb{C}^2$. Suppose that f factors, say

$$f(x, y) = g(x, y)h(x, y),$$

where g, h are nonconstant polynomials. Then $f(x, y) = 0$ if and only if one of the two equations $g(x, y) = 0$ or $h(x, y) = 0$ holds. So if we let $U = \{g(x, y) = 0\}$, $V = \{h(x, y) = 0\}$ denote these two varieties in $\mathbb{C}^2$, then

$$W = U \cup V.$$

It may be possible to see geometrically that W has no such decomposition.

For example, we can use this method to show that the polynomial

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

is irreducible. Since the total degree of f is 2, any proper factor of f has to be linear, of the form $g(x, y) = ax + by + c$. And the solutions to a linear equation lie on a line, whereas $\{f = 0\}$ is a circle. Of course when we speak of lines and circles, we are actually talking about the real loci in $\mathbb{R}^2$. So this reasoning shows that f is irreducible in $\mathbb{R}[x, y]$. But in fact, the real locus of a circle has enough points to show irreducibility in $\mathbb{C}[x, y]$ too. Suppose that $f = gh$ in $\mathbb{C}[x, y]$, where g and h are linear as before. Then every point of the real circle $x^2 + y^2 - 1 = 0$ lies on one of the complex loci U, V. So at least one of these loci contains two real points. There is exactly one complex line (a *line* being the locus of solutions of a linear equation $ax + by + c = 0$) which passes through two given points, and if these points are real, the linear equation defining the line is also real, up to a constant factor. This is proved by writing down the equation of a line through two points explicitly. So if f has a linear factor, then it has a real one. But the circle does not contain a line.

One can also prove that $x^2 + y^2 - 1$ is irreducible algebraically, using the method of undetermined coefficients (see Section 4, exercise 17).

4. EXPLICIT FACTORIZATION OF POLYNOMIALS

We now pose the problem of determining the factors of a given integer polynomial

$$f(x) = a_n x^n + \cdots + a_j x + a_0. \tag{4.1}$$

What we want are the irreducible factors in $\mathbb{Q}[x]$, and by (3.5) this amounts to determining the irreducible factors in $\mathbb{Z}[x]$. Linear factors can be found fairly easily. If $b_1 x + b_0$ divides $f(x)$, then b_1 divides a_n and b_0 divides a_0. There are finitely many integers which divide a_n and a_0, so we can try all possibilities. In each case, we carry out the division and determine whether the remainder is zero. Or we may substitute the rational number $r = -b_0/b_1$ into $f(x)$ to see if it is a root.

Though things are not so clear for factors of higher degree, Kronecker showed that the factors can be determined with a finite number of computations. His method is based on the Lagrange interpolation formula. Unfortunately this method requires too many steps to be practical except for factors of low degree, and a lot of work has been done on the problem of efficient computation. One of the most useful methods is computation modulo p, using the homomorphism $\mathbb{Z}[x] \longrightarrow \mathbb{F}_p[x]$. If our polynomial $f(x)$ factors in $\mathbb{Z}[x]$: $f = gh$, then its residue $\bar{f}(x)$ modulo p also factors: $\bar{f} = \bar{g}\bar{h}$. And since there are only finitely many polynomials of each degree in $\mathbb{F}_p[x]$, all factorizations there can be carried out in finitely many steps.

(4.2) **Proposition.** Let $f(x) = a_n x^n + \cdots + a_0 \in \mathbb{Z}[x]$ be an integer polynomial, and let p be a prime integer which does not divide a_n. If the residue $\bar{f}$ of f modulo p is irreducible, then f is irreducible in $\mathbb{Q}[x]$.

Proof. This follows from an inspection of the homomorphism. We need the assumption that p does not divide a_n in order to rule out the possibility that a factor g of f could reduce to a constant in $\mathbb{F}_p[x]$. This assumption is preserved if we replace f by the associated primitive polynomial. So we may assume that f is primitive. Since p does not divide a_n, the degrees of f and $\bar{f}$ are equal. If f factors in $\mathbb{Q}[x]$, then it also factors in $\mathbb{Z}[x]$, by Corollary (3.5). Let $f = gh$ be a proper factorization in $\mathbb{Z}[x]$. Since f is primitive, g and h have positive degree. Since $\deg f = \deg \bar{f}$ and $\bar{f} = \bar{g}\bar{h}$, it follows that $\deg g = \deg \bar{g}$ and $\deg h = \deg \bar{h}$, hence that $\bar{f} = \bar{g}\bar{h}$ is a proper factorization, which shows that $\bar{f}$ is reducible. □

Suppose we suspect that a given polynomial $f(x) \in \mathbb{Z}[x]$ is irreducible. Then we can try reduction modulo p for a few low primes, $p = 2$ or 3 for instance, and hope that $\bar{f}$ turns out to be of the same degree and irreducible. If so, we will have proved that f is irreducible too. Note also that since $\mathbb{F}_p$ is a field, the results of Theorem (1.5) hold for the ring $\mathbb{F}_p[x]$.

Unfortunately, there exist integer polynomials which are irreducible, though they can be factored modulo p for every prime p. The polynomial $x^4 - 10x^2 + 1$ is an example. So the method of reduction modulo p will not always work. But it does work quite often.

The irreducible polynomials in $\mathbb{F}_p[x]$ can be found by the "sieve" method. The *sieve of Eratosthenes* is the name given to the following method of determining the primes less than a given number n. We list the integers from 2 to n. The first one, 2, is prime because any proper factor of 2 must be smaller than 2, and there is no smaller integer on the list. We make a note of the fact that 2 is prime, and then we cross out the multiples of 2 from our list. Except for 2 itself, they are not prime. The first integer which is left, 3, is a prime because it isn't divisible by any smaller prime. We note that 3 is a prime and then cross out the multiples of 3 from our list. Again, the smallest remaining integer, 5, is a prime, and so on.

2 3 ~~4~~ 5 ~~6~~ 7 ~~8~~ ~~9~~ ~~10~~ 11 ~~12~~ 13 ~~14~~ ~~15~~ ~~16~~ 17 ~~18~~ 19 ....

This method will also determine the irreducible polynomials in $\mathbb{F}_p[x]$. We list all polynomials, degree by degree, and then cross out products. For example, the

linear polynomials in $\mathbb{F}_2[x]$ are x and $x + 1$. They are irreducible. The polynomials of degree 2 are x^2, $x^2 + x$, $x^2 + 1$, and $x^2 + x + 1$. The first three are divisible by x or by $x + 1$, so the last one is the only irreducible polynomial of degree 2 over $\mathbb{F}_2$.

(4.3) *The irreducible polynomials of degree* ≤ 4 *over* $\mathbb{F}_2$:

$$x, \quad x + 1; \quad x^2 + x + 1; \quad x^3 + x^2 + 1, \quad x^3 + x + 1;$$
$$x^4 + x^3 + 1, \quad x^4 + x + 1, \quad x^4 + x^3 + x^2 + x + 1.$$

By trying the polynomials on this list, we can factor all polynomials of degree 9 or less in $\mathbb{F}_2[x]$.

As a sample application of 4.2, the polynomial $x^4 - 6x^3 + 12x^2 - 3x + 9$ is irreducible in $\mathbb{Q}[x]$, because its residue in $\mathbb{F}_2[x]$ is $x^4 + x + 1$.

(4.4) *The monic irreducible polynomials of degree* 2 *over* $\mathbb{F}_3$:

$$x^2 + 1, \quad x^2 + x - 1, \quad x^2 - x - 1.$$

Reduction modulo p may help describe the factorization of a polynomial even though the residue is reducible. Consider the polynomial $f(x) = x^3 + 6x + 3$ for instance. Reducing modulo 3, we obtain x^3. This doesn't look like a promising tool. However, suppose that $f(x)$ were reducible, say $(ax + b)(cx^2 + dx + e) = x^3 + 6x + 3$. Then the residue of $ax + b$ would have to divide x^3 in $\mathbb{F}_3[x]$, which would imply $b \equiv 0$ (modulo 3). Similarly, we could conclude $e \equiv 0$ (modulo 3). It is impossible to satisfy both of these conditions, because $be = 3$. Therefore no such factorization exists, and $f(x)$ is irreducible.

The principle at work in this example is called the Eisenstein Criterion.

(4.5) **Proposition.** *Eisenstein Criterion:* Let $f(x) = a_n x^n + \cdots + a_0 \in \mathbb{Z}[x]$ be an integer polynomial, and let p be a prime integer. Suppose that the coefficients of f satisfy the following conditions:

(i) p does not divide a_n;

(ii) p divides the other coefficients $a_{n-1}, \ldots, a_0$;

(iii) p^2 does not divide a_0.

Then f is irreducible in $\mathbb{Q}[x]$. If f is primitive, it is irreducible in $\mathbb{Z}[x]$.

For example, $x^4 + 50x^2 + 30x + 20$ is irreducible in $\mathbb{Q}[x]$ and in $\mathbb{Z}[x]$.

Proof of the Eisenstein Criterion. Assume that the conditions are met for f. Let $\bar{f}$ denote the residue modulo p. The hypotheses (i) and (ii) imply that $\bar{f} = \bar{a}_n x^n$ and that $\bar{a}_n \neq 0$. If f is reducible in $\mathbb{Q}[x]$, then it will factor in $\mathbb{Z}[x]$ into factors of positive degree, say $f = gh$. Then $\bar{g}$ and $\bar{h}$ divide $\bar{a}_n x^n$, and hence each of these polynomials is a monomial. Therefore all coefficients of g and of h, except the highest ones, are divisible by p. Let the constant coefficients of g, h be b_0, c_0. Then the constant coefficient of f is $a_0 = b_0 c_0$. Since p divides b_0 and c_0, it follows that p^2 divides a_0, which contradicts (iii). This shows that f is irreducible. The last assertion follows from (3.6). □

One of the most important applications of the Eisenstein Criterion is to prove the irreducibility of the *cyclotomic polynomial* $x^{p-1} + x^{p-2} + \cdots + x + 1$, whose roots are the pth roots of unity, the powers of $\zeta = e^{2\pi i/p}$:

(4.6) **Corollary.** Let p be a prime. The polynomial $f(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$ is irreducible in $\mathbb{Q}[x]$.

Proof. We note that $(x - 1)f(x) = x^p - 1$. Next, we make the substitution $x = y + 1$ into this product, obtaining

$$yf(y + 1) = (y + 1)^p - 1 = y^p + \binom{p}{1}y^{p-1} + \binom{p}{2}y^{p-2} + \cdots + \binom{p}{p-1}y.$$

We have $\binom{p}{i} = p(p - 1)\cdots(p - i - 1)/i!$. If $i < p$, then the prime p isn't a factor of $i!$, so $i!$ divides the product $(p - 1)\cdots(p - i + 1)$ of the remaining terms in the numerator of the integer $\binom{p}{i}$. This implies that $\binom{p}{i}$ is divisible by p. Dividing the expansion of $yf(y + 1)$ by y shows that $f(y + 1)$ satisfies the conditions of the Eisenstein Criterion, hence that it is an irreducible polynomial. It follows that $f(x)$ is irreducible too. □

It is instructive to examine the statement analogous to the Eisenstein Criterion when the ring of integers is replaced by the polynomial ring $\mathbb{C}[t]$. Then $\mathbb{Z}[x]$ gets replaced by $\mathbb{C}[t][x] \approx \mathbb{C}[t, x]$, the polynomial ring in two variables.

(4.7) **Proposition.** Let $f(t, x)$ be an element of $\mathbb{C}[t, x]$, written as a polynomial in x whose coefficients are polynomials in t: $f(t, x) = a_n(t)x^n + \cdots + a_1(t)x + a_0(t)$. Suppose that

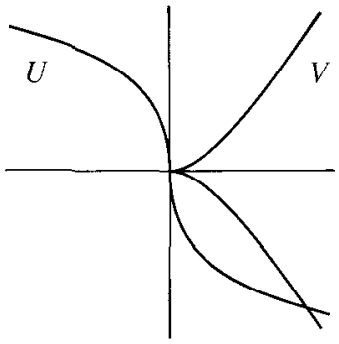
(i) t does not divide $a_n(t)$,
(ii) t divides $a_{n-1}(t), \ldots, a_0(t)$,
(iii) t^2 does not divide $a_0(t)$.

Then $f(t, x)$ is irreducible in the ring $\mathbb{C}(t)[x]$. If f is primitive, meaning that it has no factor which is a polynomial in t alone, then f is irreducible in $\mathbb{C}[t, x]$.

This can be proved exactly as we proved (4.5), replacing $\mathbb{F}_p[x]$ by $\mathbb{C}[x] = \mathbb{C}[t, x]/(t)$. But let us examine the geometry of this situation by considering the locus $W = \{f(t, x) = 0\}$ in complex 2-space. Conditions (i) and (ii) of (4.7) imply that $f(0, x) = cx^n$, where $c = a_n(0) \neq 0$. Consequently the only solution of $f(t, x) = 0$ with $t = 0$ is $t = x = 0$, so the variety W meets the x-axis $\{t = 0\}$ only at the origin.

Suppose that $f(t, x)$ is reducible: $f(t, x) = g(t, x)h(t, x)$. Then W is the union of the two varieties $U = \{g = 0\}$ and $V = \{h = 0\}$. Also, $cx^n = f(0, x) = g(0, x)h(0, x)$. Hence $g(0, x)$ is a constant times x^r, and $h(0, x)$ is a constant times x^{n-r}, where r is the degree of g in the variable x. Therefore g and h both vanish at

the origin. It follows that the origin is a *singular point* of W, meaning that the partial derivatives $\partial f/\partial x$ and $\partial f/\partial t$ both vanish at $(0, 0)$. This is checked by differentiating the product gh. On the other hand, $\partial f/\partial t\,(0, 0) = da_0/dt\,(0)$, and this is the linear coefficient of $a_0(t)$. If it vanishes, t^2 divides $a_0(t)$, contrary to (4.7iii). □



5. PRIMES IN THE RING OF GAUSS INTEGERS

We have seen that the ring of Gauss integers is a Euclidean domain. Its units are $\{\pm 1, \pm i\}$, and every element which is not zero and not a unit is a product of prime elements. In this section we will study these prime elements, called *Gauss primes*, and their relation to prime integers. We looked at some examples in Section 2, where we saw that the prime integer 5 factors in $\mathbb{Z}[i]$: $5 = (2 + i)(2 - i)$, while 3 does not factor; 3 is a Gauss prime. Remember that since there are four units, there are four associate factorizations of the integer 5 which we consider equivalent:

$$(2 + i)(2 - i) = (-2 - i)(-2 + i) = (1 - 2i)(1 + 2i) = (-1 + 2i)(-1 - 2i).$$

We will now show that the examples 3 and 5 exhibit the two ways that prime integers can factor in the ring $\mathbb{Z}[i]$. The story is summed up in this theorem:

(5.1) **Theorem.**

(a) Let p be a prime integer. Then either p is a Gauss prime, or else it is the product of two complex conjugate Gauss primes: $p = \pi\overline{\pi}$.

(b) Let π be a Gauss prime. Then either $\pi\overline{\pi}$ is a prime integer, or else it is the square of a prime integer.

(c) The prime integers which are Gauss primes are those congruent to 3 modulo 4; that is, $p = 3, 7, 11, 19, \ldots$.

(d) Let p be a prime integer. The following are equivalent:

(i) p is a product of two complex conjugate Gauss primes.
(ii) p is the sum of two integer squares: $p = a^2 + b^2$, with $a, b \in \mathbb{Z}$.
(iii) The congruence $x^2 \equiv -1$ (modulo p) has an integer solution.
(iv) $p \equiv 1$ (modulo 4), or $p = 2$; that is, $p = 2, 5, 13, 17, \ldots$.

It will take some time to prove all parts of this theorem.

The following lemma follows directly from the definition of a Gauss integer:

(5.2) **Lemma.** A Gauss integer which is a real number is an ordinary integer. An ordinary integer d divides another integer a in $\mathbb{Z}[i]$ if and only if d divides a in $\mathbb{Z}$. Moreover, d divides a Gauss integer $a + bi$ if and only if d divides both a and b.

Now to prove part (a) of the theorem, let p be an integer prime. Then p is not a unit in the ring $\mathbb{Z}[i]$. Hence it has a Gauss prime divisor, say $\pi = a + bi$, where $a, b \in \mathbb{Z}$. The complex conjugate $\overline{\pi} = a - bi$ also divides p because $p = \overline{p}$, so $\pi\overline{\pi} = a^2 + b^2$ divides p^2 in the ring of Gauss integers. Being an integer, $\pi\overline{\pi}$ is an integer divisor of p^2. There are two possibilities: π may be an associate of p. In this case, p is a Gauss prime. Otherwise π is a proper divisor of p in the ring of Gauss integers, and then $\pi\overline{\pi}$ is a proper divisor of p^2 in the ring $\mathbb{Z}$. Since $\pi\overline{\pi}$ is a positive integer, $\pi\overline{\pi} = p$ in this case.

We can turn this argument around to prove (b). Let π be a Gauss prime. Then $\pi\overline{\pi}$ is a positive integer, say $\pi\overline{\pi} = n$. We factor n into primes in the ring of integers. This factorization will also be a factorization in the Gauss integers, though not necessarily a prime factorization. Since π is a Gauss prime which divides n in $\mathbb{Z}[i]$, it divides one of the integer prime factors of n. Thus π divides an integer prime p. Then $\pi\overline{\pi}$ is an integer divisor of p^2, hence $\pi\overline{\pi} = p$ or p^2.

Note that part (c) of Theorem (5.1) is a formal consequence of (a) and of the equivalence of conditions (d)(i) and (d)(iv). So we need not consider part (c) further, and we now turn to the proof of part (d). It is easy to see that (i) and (ii) of part (d) are equivalent: Suppose that $p = \pi\overline{\pi}$ for some Gauss prime $\pi = a + bi$. Then $p = \pi\overline{\pi} = (a + bi)(a - bi) = a^2 + b^2$, so p is a sum of two integer squares. Conversely, if $p = a^2 + b^2$, then $p = (a + bi)(a - bi)$ provides a factorization of p in the ring of Gauss integers, which is a prime factorization because of (a).

The equivalence of (d)(i) and (d)(iii) of Theorem (5.1) is harder to prove. To do so, we go back to the formal construction of the Gauss integers. The ring $\mathbb{Z}[i]$ is obtained from the ring $\mathbb{Z}$ by adjoining an element i with the relation $i^2 + 1 = 0$. So there is an isomorphism

$$\mathbb{Z}[x]/(x^2 + 1) \xrightarrow{\sim} \mathbb{Z}[i]. \tag{5.3}$$

Let (p) denote the principal ideal generated by a prime integer p in the ring of Gauss integers. Its elements are the Gauss integers $a + bi$ such that a and b are both divisible by p. Denote by R' the quotient ring $\mathbb{Z}[i]/(p)$. Then R' can also be thought of as the ring obtained by introducing the two relations

$$x^2 + 1 = 0 \quad \text{and} \quad p = 0 \tag{5.4}$$

into the polynomial ring $\mathbb{Z}[x]$. So we have an isomorphism

$$\mathbb{Z}[x]/(x^2 + 1, p) \xrightarrow{\sim} \mathbb{Z}[i]/(p) = R', \tag{5.5}$$

where $(x^2 + 1, p)$ denotes the ideal of $\mathbb{Z}[x]$ generated by the two elements.

(5.6) **Lemma.** Let p be a prime integer. The following statements are equivalent:

(i) p is a Gauss prime;

(ii) the ring $R' = \mathbb{Z}[i]/(p)$ is a field;
(iii) $x^2 + 1$ is an irreducible polynomial in the ring $\mathbb{F}_p[x]$.

Proof. The equivalence of the first two statements follows from Proposition (2.14). What we are really after is the equivalence of (i) and (iii), and at first glance, these two statements do not seem to be related at all. It was in order to obtain this equivalence that we introduced the auxiliary ring R'. The proof is based on the following elementary but remarkably useful observation, which follows from the Third Isomorphism Theorem [Chapter 10 (4.3b)]:

(5.7) *To construct the ring R', it does not matter which of the two relations (5.4) is introduced into the ring $\mathbb{Z}[x]$ first.*

So let us reverse the order and begin by killing the element p. The Substitution Principle tells us what we will get. The kernel of the homomorphism $\mathbb{Z}[x] \longrightarrow \mathbb{F}_p[x]$ is precisely the ideal $p\mathbb{Z}[x]$. Since this map is surjective, it induces an isomorphism

$$\mathbb{Z}[x]/p\mathbb{Z}[x] \xrightarrow{\sim} \mathbb{F}_p[x].$$

We now introduce our other relation $x^2 + 1 = 0$ into this ring, interpreting the coefficients of this polynomial as elements of $\mathbb{F}_p$. The result is an isomorphism

$$\mathbb{F}_p[x]/(x^2 + 1) \xrightarrow{\sim} R'. \tag{5.8}$$

Proposition (2.14), applied to the ring $\mathbb{F}_p[x]$, shows that R' is a field if and only if $x^2 + 1$ is irreducible in $\mathbb{F}_p[x]$. □

We can now prove the equivalence of conditions (d)(i) and (d)(iii) of (5.1). We know by Lemma (5.6) that p is a Gauss prime if and only if $x^2 + 1$ is an irreducible polynomial in the ring $\mathbb{F}_p[x]$. Since it is a quadratic polynomial, $x^2 + 1$ is reducible if it has a root in $\mathbb{F}_p$ and irreducible if it has no root. Also, the residue of an integer a (modulo p) is a root of $x^2 + 1$ if and only if $a^2 \equiv -1$ (modulo p). Thus the congruence $x^2 \equiv -1$ (modulo p) has a solution if and only if $x^2 + 1$ is reducible modulo p, which happens if and only if p is not a Gauss prime. The equivalence of (i) and (iii) follows.

It remains to prove the equivalence of condition (iv) of part (d) with one of the other conditions. We will show its equivalence with condition (iii). The congruence $x^2 \equiv -1$ (modulo 2) does have the solution $x = 1$, so it is sufficient to look at the other primes, that is, at the odd primes. The following lemma does the job:

(5.9) **Lemma.** Let p be an odd prime, and let $\overline{a}$ denote the residue of an integer a modulo p.

(a) The integer a solves the congruence $x^2 \equiv -1$ (modulo p) if and only if its residue $\overline{a}$ is an element of order 4 in the multiplicative group of the field $\mathbb{F}_p$.
(b) The multiplicative group $\mathbb{F}_p^\times$ contains an element of order 4 if and only if $p \equiv 1$ (modulo 4).

Proof. There is exactly one element of order 2 in $\mathbb{F}_p^\times$, namely the residue of -1. This is because an element of order 2 is a root of the polynomial $x^2 - 1$, and we know the roots of this polynomial: They are ± 1 in any field [see (1.7)]. If a residue $\overline{a}$ has order 4 in $\mathbb{F}_p^\times$, then $\overline{a}^2$ has order 2; hence $\overline{a}^2 = -1$, which means $a^2 \equiv -1$ (modulo p). Conversely, if $a^2 \equiv -1$ (modulo p), then $\overline{a}$ has order 4 in $\mathbb{F}_p^\times$. This proves part (a) of the lemma.

Now the order of the group $\mathbb{F}_p^\times$ is $p - 1$. So if this group contains an element of order 4, then $p - 1$ is divisible by 4, or equivalently $p \equiv 1$ (modulo 4). Conversely, suppose that $p - 1$ is divisible by 4, and let H be the Sylow-2 subgroup of $\mathbb{F}_p^\times$, whose order is the largest power 2^r of 2 which divides $p - 1$. Since 4 divides $p - 1$, the order of H is at least 4, so there is an element $\overline{a}$ in H different from ± 1. This element does not have order 2, nor does it have order 1. But since H is a 2-group, the order of $\overline{a}$ is a power of 2. So some power of $\overline{a}$ has order exactly 4.

This completes the proof of Theorem (5.1). □

6. ALGEBRAIC INTEGERS

In the next sections we are going to study factorization of algebraic numbers in a simple but important case, that of quadratic imaginary integers. The ring of Gauss integers is our model here. It was in order to extend the properties of factorization of ordinary integers to algebraic numbers that ideals were first introduced, and the extension is very beautiful.

In contrast to most of the topics we have studied, the arithmetic of quadratic number fields is not of universal importance. It has many applications to arithmetic, but not so many in other areas of mathematics. Our reason for including this topic, aside from its elegance, is its historical importance. Many of our algebraic tools were first developed in order to extend arithmetic properties of the integers to algebraic numbers.

A typical application of algebraic numbers to arithmetic is to the problem of determining integer points on an ellipse such as

$$x^2 + 5y^2 = p, \tag{6.1}$$

where for simplicity we assume that p is a prime. To determine integer points on the circle $x^2 + y^2 = p$, we may begin by factoring the left side, obtaining $(x + iy)(x - iy) = p$, and then use arithmetic in the Gauss integers to analyze the factorization. We did this in our proof of Theorem (5.1). The analogous procedure for equation (6.1) leads to

$$(x + \sqrt{-5}y)(x - \sqrt{-5}y) = p,$$

so we may attempt an analysis in the ring $\mathbb{Z}[\sqrt{-5}]$. However, as we have seen, factorization is not unique in this ring. We will have some trouble.

Another example is the famous Fermat Equation

$$x^3 + y^3 = z^3. \tag{6.2}$$

It was proved by Euler that this equation has no integer solutions, except for the trivial solutions in which one of the variables is zero. To analyze it, we may bring y^3 to the other side and factor, obtaining

$$x^3 = (z - y)(z - \zeta y)(z - \bar{\zeta} y), \tag{6.3}$$

where

$$\zeta = \tfrac{1}{2}(-1 + \sqrt{-3}) = e^{2\pi i/3} \tag{6.4}$$

is a complex cube root of 1. One can then analyze this equation using arithmetic in the ring $\mathbb{Z}[\zeta]$. This ring happens to be a Euclidean domain, so unique factorization is available. Unfortunately, the proof that (6.2) has no nontrivial solution is fairly complicated, so we will not give it.

Problems of this type, which ask for integer solutions of polynomial equations, are called *Diophantine problems*. We will analyze a few of them in Section 12, when the necessary tools have been assembled.

A complex number α is called algebraic if it is the root of a nonzero polynomial $f(x)$ with rational coefficients (Chapter 10, Section 1). We can, of course, clear denominators in the coefficients of the polynomial $f(x)$. So if α is an algebraic number, then it is also the root of a polynomial with integer coefficients. The number α is called an *algebraic integer* if it is the root of a *monic* polynomial with integer coefficients, a polynomial of the form

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0, \quad \text{with } a_i \in \mathbb{Z}. \tag{6.5}$$

Thus the cube root of unity ζ, being a root of the polynomial $x^3 - 1$, is an algebraic integer.

Let α be an algebraic number. The set of all polynomials in $\mathbb{Q}[x]$ which have α as a root is the kernel of the substitution homomorphism

$$\mathbb{Q}[x] \longrightarrow \mathbb{C}, \quad \text{defined by } f(x) \rightsquigarrow f(\alpha).$$

So it is a principal ideal, generated by an irreducible element $f(x)$ of the polynomial ring which is called the *irreducible polynomial for α over $\mathbb{Q}$*. (Why is f irreducible?) It is the polynomial of lowest degree having α as a root and is unique up to a constant factor. The degree of the irreducible polynomial for α is also called the *degree* of α over $\mathbb{Q}$.

We may choose this irreducible polynomial $f(x)$ for α to be a *primitive* polynomial in $\mathbb{Z}[x]$. Then $f(x)$ also generates the ideal of $\mathbb{Z}[x]$ of all integer polynomials having α as a root.

(6.6) **Proposition.** The kernel of the map $\mathbb{Z}[x] \longrightarrow \mathbb{C}$ sending $x \rightsquigarrow \alpha$ is the principal ideal of $\mathbb{Z}[x]$ generated by the primitive irreducible polynomial for α.

Proof. Let $f(x)$ be the primitive irreducible polynomial for α. If $g \in \mathbb{Z}[x]$ has α as a root, then f divides g in $\mathbb{Q}[x]$, and hence f divides g in $\mathbb{Z}[x]$ too, by (3.4). So g is in the principal ideal of $\mathbb{Z}[x]$ generated by f. □

Note that the leading coefficient of a polynomial $f(x)$ divides the leading coefficient of any multiple in $\mathbb{Z}[x]$. So it follows from Proposition (6.6) that if the primitive irreducible polynomial $f(x)$ for α is *not* monic, then α is not the root of any monic integer polynomial.

(6.7) **Proposition.** An algebraic number α is an algebraic integer if and only if the primitive irreducible polynomial for α is monic. Equivalently, α is an algebraic integer if and only if the monic irreducible polynomial for α in $\mathbb{Q}[x]$ has integer coefficients. □

The primitive irreducible polynomial for the cube root of unity ζ is $x^2 + x + 1$.

(6.8) **Corollary.** A rational number r is an algebraic integer if and only if it is an ordinary integer.

For, the monic irreducible polynomial over $\mathbb{Q}$ of a rational number r is $x - r$. □

Proposition (6.7) can be used to decide whether or not an algebraic number is an algebraic integer, provided that we can compute its irreducible polynomial. For example, $\alpha = \frac{1}{2}(1 + \sqrt{2})$ is a root of $4x^2 - 4x - 1$. This is the primitive irreducible polynomial for α. Hence α is not an algebraic integer.

The concept of algebraic integer was one of the most important discoveries of number theory. It is not easy to explain quickly why it is the right definition to use, but roughly speaking, we can think of the leading coefficient of the primitive irreducible polynomials $f(x)$ for α as a "denominator." If α is the root of an integer polynomial $f(x) = dx^n + a_{n-1}x^{n-1} + \cdots + a_0$, then $d\alpha$ is an algebraic integer, because it is a root of the monic integer polynomial

$$x^n + a_{n-1}x^{n-1} + da_{n-2}x^{n-2} + \cdots + d^{n-2}a_1x + d^{n-1}a_0. \tag{6.9}$$

Thus we can "clear the denominator" in any algebraic number α by multiplying it with a suitable integer to get an algebraic integer. The leading coefficient is, however, not a precise denominator. Thus if $\alpha = \frac{1}{2}(1 + \sqrt{2})$, then 2α is an algebraic integer, while the leading coefficient of its primitive irreducible polynomial is 4.

In another direction, the example of the algebraic integer $\zeta = \frac{1}{2}(-1 + \sqrt{-3})$ shows that we must not jump to conclusions just because some expression for an algebraic number has denominators.

Explicit computation with algebraic integers is not very easy. It is a fact that they form a subring of $\mathbb{C}$, that is, that sums and products of algebraic integers are algebraic integers, but this isn't obvious. Rather than develop a general theory, we will work out the case of quadratic extensions explicitly.

A *quadratic number field* $F = \mathbb{Q}[\sqrt{d}]$ consists of all complex numbers

$$a + b\sqrt{d}, \quad \text{with} \quad a, b \in \mathbb{Q}, \tag{6.10}$$

where d is a fixed integer, positive or negative, which is not a rational square. The notation $\sqrt{d}$ will stand for the positive square root if $d > 0$ and for the positive

imaginary square root if $d < 0$. If d has a square integer factor, we can pull it out of the radical and put it into b without changing the field. Therefore it is customary to assume that d is *square free,* meaning that $d = \pm p_1 \cdots p_r$ where the p_i are distinct primes, or that $d = -1$. So the values we take are

$$d = -1, \pm 2, \pm 3, \pm 5, \pm 6, \pm 7, \pm 10, \ldots .$$

The field F is called a *real quadratic number field* if $d > 0$, or an *imaginary quadratic number field* if $d < 0$.

We will now compute the algebraic integers in F. The computation for a special value of d is no simpler than the general case. Nevertheless, you may wish to substitute a value such as $d = 5$ when going over this computation. We set

$$\delta = \sqrt{d}. \tag{6.11}$$

When d is negative, δ is purely imaginary. Let

$$\alpha = a + b\delta$$

be any element of F which is not in $\mathbb{Q}$, that is, such that $b \neq 0$. Then $\alpha' = a - b\delta$ is also in F. If d is negative, α' is the complex conjugate of α. Note that α is a root of the polynomial

$$(x - \alpha)(x - \alpha') = x^2 - (\alpha+\alpha')x + \alpha\alpha' = x^2 - 2ax + (a^2 - b^2 d). \tag{6.12}$$

This polynomial has the rational coefficients $-2a$ and $a^2 - b^2 d$. Since α is not a rational number, it is not the root of a linear polynomial. So (6.12) is irreducible and is therefore the monic irreducible polynomial for α over $\mathbb{Q}$. According to (6.7), α is an algebraic integer if and only if (6.12) has integer coefficients. Thus we have the following corollary:

(6.13) **Corollary.** $\alpha = a + b\delta$ is an algebraic integer if and only if $2a$ and $a^2 - b^2 d$ are integers. □

This corollary also holds when $b = 0$, because if a^2 is an integer, then so is a. If we like, we can use the conditions of the corollary as a definition of the integers in F.

The possibilities for a and b depend on the congruence class of d modulo 4. Note that since d is assumed to be square free, the case $d \equiv 0$ (modulo 4) has been ruled out, so $d \equiv 1, 2$, or 3 (modulo 4).

(6.14) **Proposition.** The algebraic integers in the quadratic field $F = \mathbb{Q}[\sqrt{d}]$ have the form $\alpha = a + b\delta$, where:

(a) If $d \equiv 2$ or 3 (modulo 4), then a and b are integers.
(b) If $d \equiv 1$ (modulo 4), then either $a, b \in \mathbb{Z}$ or $a, b \in \mathbb{Z} + \frac{1}{2}$.

The cube root of unity $\zeta = \frac{1}{2}(-1 + \sqrt{-3})$ is an example of an algebraic integer of the second type. On the other hand, since $-1 \equiv 3$ (modulo 4), the integers in the field $\mathbb{Q}[i]$ are just the Gauss integers.

Proof of the Proposition. Since the coefficients of the irreducible polynomial (6.12) for α are $2a$ and $a^2 - b^2d$, α is certainly an algebraic integer if a and b are integers. Assume that $d \equiv 1$ (modulo 4) and that $a, b \in \mathbb{Z} + \frac{1}{2}$. (We say that they are *half integers.*) Then $2a \in \mathbb{Z}$. To show that $a^2 - b^2d \in \mathbb{Z}$, we write $a = \frac{1}{2}m$, $b = \frac{1}{2}n$, where m, n are odd integers. Computing modulo 4, we find

$$m^2 - n^2d \equiv (\pm 1)^2 - (\pm 1)^2 \cdot 1 \equiv 0 \text{ (modulo 4)}.$$

Hence $a^2 - b^2d = \frac{1}{4}(m^2 - n^2d) \in \mathbb{Z}$, as required.

Conversely, suppose that α is an algebraic integer. Then $2a \in \mathbb{Z}$ by Corollary (6.13). There are two cases: either $a \in \mathbb{Z}$ or $a \in \mathbb{Z} + \frac{1}{2}$.

Case 1: $a \in \mathbb{Z}$. It follows that $b^2d \in \mathbb{Z}$ too. Now if we write $b = m/n$, where m, n are relatively prime integers and $n > 0$, then $b^2d = m^2d/n^2$. Since d is square free, it can't cancel a square in the denominator. So $n = 1$. If a is an integer, b must be an integer too.

Case 2: $a \in \mathbb{Z} + \frac{1}{2}$ is a half integer, say $a = \frac{1}{2}m$ as before. Then $4a^2 \in \mathbb{Z}$, and the condition $a^2 - b^2d \in \mathbb{Z}$ implies that $4b^2d \in \mathbb{Z}$ but $b^2d \notin \mathbb{Z}$. Therefore b is also a half integer, say $b = \frac{1}{2}n$, where n is odd. In order for this pair of values for a, b to satisfy $a^2 - b^2d \in \mathbb{Z}$, we must have $m^2 - n^2d \equiv 0$ (modulo 4). Computing modulo 4, we find that $d \equiv 1$ (modulo 4). □

A convenient way to write all the integers in the case $d \equiv 1$ (modulo 4) is to introduce the algebraic integer

$$\eta = \tfrac{1}{2}(1 + \delta), \tag{6.15}$$

which is a root of the monic integer polynomial

$$x^2 - x + \tfrac{1}{4}(1 - d). \tag{6.16}$$

(6.17) **Proposition.** Assume that $d \equiv 1$ (modulo 4). Then the algebraic integers in $F = \mathbb{Q}[\sqrt{d}]$ are $a + b\eta$, where $a, b \in \mathbb{Z}$. □

It is easy to show by explicit calculation that the integers in F form a ring R in each case, called the *ring of integers in* F. Computation in R can be carried out by high school algebra.

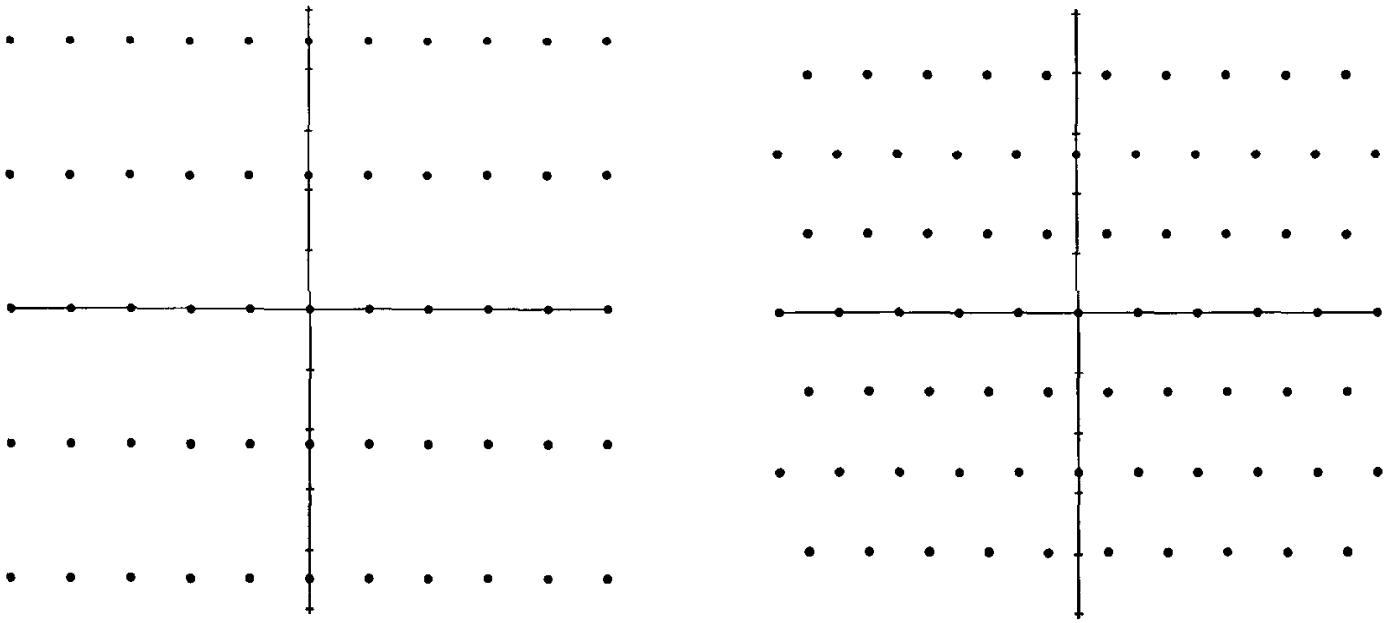
The *discriminant* of F is defined to be the discriminant of the polynomial $x^2 - d$ in the case $R = \mathbb{Z}[\delta]$ and the discriminant of the polynomial $x^2 - x + \frac{1}{4}(1 - d)$ if $R = \mathbb{Z}[\eta]$. This discriminant will be denoted by D. Thus

$$D = \begin{cases} 4d & \text{if } d \equiv 2, 3 \\ d & \text{if } d \equiv 1 \end{cases} \text{ (modulo 4)}. \tag{6.18}$$

Since D can be computed in terms of d, it isn't very important to introduce a separate notation for it. However, some formulas become independent of the congruence class when they are expressed in terms of D rather than d.

The imaginary quadratic case $d < 0$ is slightly easier to treat than the real one, so we will concentrate on it in the next sections. In the imaginary case, the ring R

forms a lattice in the complex plane which is rectangular if $d \equiv 2, 3$ (modulo 4), and "isosceles triangular" if $d \equiv 1$ (modulo 4). When $d = -1, R$ is the ring of Gauss integers, and the lattice is square. When $d = -3$, the lattice is equilateral triangular. Two other examples are depicted below.



$d = -5$ $d = -7$

(6.19) **Figure.** Integers in some imaginary quadratic fields.

The property of being a lattice is very special to rings such as those we are considering here, and we will use geometry to analyze them. Thinking of R as a lattice is also useful for intuition.

It will be helpful to carry along a specific example as we go. We will use the case $d = -5$ for this purpose. Since $-5 \equiv 3$ (modulo 4), the ring of integers forms a rectangular lattice, and $R = \mathbb{Z}[\delta]$, where $\delta = \sqrt{-5}$.

7. FACTORIZATION IN IMAGINARY QUADRATIC FIELDS

Let R be the ring of integers of an imaginary quadratic number field $F = \mathbb{Q}[\delta]$. If $\alpha = a + b\delta$ is in R, so is its complex conjugate $\overline{\alpha} = a - b\delta$. We call the *norm* of α the integer

$$N(\alpha) = \alpha\overline{\alpha}. \tag{7.1}$$

It is also equal to $a^2 - b^2 d$ and to $|\alpha|^2$, and it is the constant term of the irreducible polynomial for α over $\mathbb{Q}$. Thus $N(\alpha)$ is a positive integer unless $\alpha = 0$. Note that

$$N(\beta\gamma) = N(\beta)N(\gamma). \tag{7.2}$$

This formula gives us some control of possible factors of an element α of R. Say that $\alpha = \beta\gamma$. Then both terms on the right side of (7.2) are positive integers. So to check for factors of α, it is enough to look at elements β whose norm divides $N(\alpha)$; this is not too big a job if a and b are reasonably small.

In particular, let us ask for *units* of R:

(7.3) **Proposition.**

(a) An element α of R is a unit if and only if $N(\alpha) = 1$.

(b) The units of R are $\{\pm 1\}$ unless $d = -1$ or -3. If $d = -1$, so that R is the ring of Gauss integers, the units are $\{\pm 1, \pm i\}$, and if $d = -3$ they are the powers of the 6th root of unity $\frac{1}{2}(1 + \sqrt{-3})$.

Proof. If α is a unit, then $N(\alpha)N(\alpha^{-1}) = N(1) = 1$. Since $N(\alpha)$ and $N(\alpha^{-1})$ are positive integers, they are both equal to 1. Conversely, if $N(\alpha) = \alpha\overline{\alpha} = 1$, then $\overline{\alpha} = \alpha^{-1}$. So $\alpha^{-1} \in R$, and α is a unit. Thus α is a unit if and only if it lies on the unit circle in the complex plane. The second assertion follows from the configuration of the lattice R [see Figure (6.19)]. □

Next we investigate factorization of an element $\alpha \in R$ into irreducible factors.

(7.4) **Proposition.** Existence of factorizations is true in R.

Proof. If $\alpha = \beta\gamma$ is a proper factorization in R, then β, γ aren't units. So by Proposition (7.3), $N(\alpha) = N(\beta)N(\gamma)$ is a proper factorization in the ring of integers. The existence of factorizations in R now follows from the existence of factorizations in $\mathbb{Z}$. □

However, factorization into irreducible elements will not be unique in most cases. We gave a simple example with $d = -5$ in Section 2:

$$6 = 2 \cdot 3 = (1 + \delta)(1 - \delta), \tag{7.5}$$

where $\delta = \sqrt{-5}$. For example, to show that $1 + \delta$ is irreducible, we note that its norm is $(1 + \delta)(1 - \delta) = 6$. A proper factor must have norm 2 or 3, that is, absolute value $\sqrt{2}$ or $\sqrt{3}$. There are no such points in the lattice R.

The same method provides examples for other values of d:

(7.6) **Proposition.** The only ring R with $d \equiv 3$ (modulo 4) which is a unique factorization domain is the ring of Gauss integers.

Proof. Assume that $d \equiv 3$ (modulo 4), but that $d \neq -1$. Then

$$1 - d = 2\left(\frac{1 - d}{2}\right) \quad \text{and} \quad 1 - d = (1 + \delta)(1 - \delta).$$

There are two factorizations of $1 - d$ in R. The element 2 is irreducible because $N(2) = 4$ is the smallest value >1 taken on by $N(\alpha)$. [The only points of R inside

the circle of radius 2 about the origin are $0, 1, -1$, when $d = -5, -13, -17, \ldots$. See Figure (6.19).] So if there were a common refinement of the above factorizations, 2 would divide either $1 + \delta$ or $1 - \delta$ in R, which it does not: $\frac{1}{2} \pm \frac{1}{2}\delta$ is not in R when $d \equiv 3$ (modulo 4). □

Notice that this reasoning breaks down if $d \equiv 1$ (modulo 4). In that case, 2 does divide $1 + \delta$, because $\frac{1}{2} + \frac{1}{2}\delta \in R$. In fact, there are more cases of unique factorization when $d \equiv 1$ (modulo 4). The following theorem is very deep, and we will not prove it:

(7.7) **Theorem.** Let R be the ring of integers in the imaginary quadratic field $\mathbb{Q}(\sqrt{d})$. Then R is a unique factorization domain if and only if d is one of the integers $-1, -2, -3, -7, -11, -19, -43, -67, -163$.

Gauss proved for these values of d that R is a unique factorization domain. We will learn how to do this. He also conjectured that there were no others. This much more difficult part of the theorem was finally proved by Baker and Stark in 1966, after the problem had been worked on for more than 150 years.

Ideals were introduced to rescue the uniqueness of factorization. As we know (2.12), R must contain some nonprincipal ideals unless it is a unique factorization domain. We will see in the next section how these nonprincipal ideals serve as substitutes for elements.

Note that every nonzero ideal A is a *sublattice* of R: It is a subgroup under addition, and it is discrete because R is discrete. Moreover, if α is a nonzero element of A, then $\alpha\delta$ is in A too, and $\alpha, \alpha\delta$ are linearly independent over $\mathbb{R}$. However, not every sublattice is an ideal.

(7.8) **Proposition.** If $d \equiv 2$ or 3 (modulo 4), the nonzero ideals of R are the sublattices which are closed under multiplication by δ. If $d \equiv 1$ (modulo 4), they are the sublattices which are closed under multiplication by $\eta = \frac{1}{2}(1 + \delta)$.

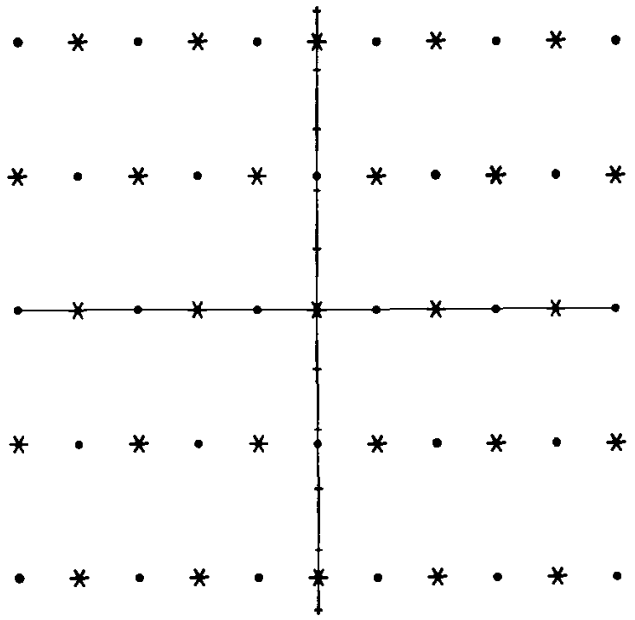
Proof. To be an ideal, a subset A must be closed under addition and under multiplication by elements of R. Any lattice is closed under addition and under multiplication by integers. So if it is also closed under multiplication by δ, then it is also closed under multiplication by an element of the form $a + b\delta$, with $a, b \in \mathbb{Z}$. This includes all elements of R if $d \equiv 2, 3$ (modulo 4). The proof in the case that $d \equiv 1$ (modulo 4) is similar. □

In order to get a feeling for the possibilities, we will describe the ideals of the ring $R = \mathbb{Z}[\sqrt{-5}]$ before going on. The most interesting ideals are those which are not principal.

(7.9) **Theorem.** Let $R = Z[\delta]$, where $\delta = \sqrt{-5}$, and let A be a nonzero ideal of R. Let α be a nonzero element of A of minimal absolute value $|\alpha|$. There are two cases:

Case 1: A is the principal ideal (α), which has the lattice basis $(\alpha, \alpha\delta)$.
Case 2: A has the lattice basis $(\alpha, \frac{1}{2}(\alpha + \alpha\delta))$, and is not a principal ideal.

The second case can occur only if $\frac{1}{2}(\alpha + \alpha\delta)$ is an element of R. The ideal $A = (2, 1 + \delta)$, which is depicted below, is an example.



(7.10) **Figure.** The ideal $(2, 1 + \delta)$ in the ring $\mathbb{Z}[\delta]$, $\delta = \sqrt{-5}$.

The statement of Proposition (7.9) has a geometric interpretation. Notice that the lattice basis $(\alpha, \alpha\delta)$ of the principal ideal (α) is obtained from the lattice basis $(1, \delta)$ of R by multiplication by α. If we write $\alpha = re^{i\theta}$, then the effect of multiplication by α is to rotate the complex plane through the angle θ and then stretch by the factor r. So (α) and R are similar geometric figures, as we noted in Section 2. Similarly, the basis $(\alpha, \frac{1}{2}(\alpha + \alpha\delta))$ is obtained by multiplication by $\frac{1}{2}\alpha$ from the basis $(2, 1 + \delta)$. So the ideals listed in Case 2 are geometric figures similar to the one depicted in Figure (7.10). The similarity classes of ideals are called the *ideal classes*, and their number is called the *class number* of R. Thus Proposition (7.9) implies that the class number of $\mathbb{Z}[\sqrt{-5}]$ is 2. We will discuss ideal classes for other quadratic imaginary fields in Section 10.

The proof of Theorem (7.9) is based on the following lemma about lattices in the complex plane:

(7.11) **Lemma.** Let r be the minimum absolute value among nonzero elements of a lattice A, and let γ be an element of A. Let D be the disc of radius $\frac{1}{n}r$ about the point $\frac{1}{n}\gamma$. There is no point of A in the interior of D other than its center $\frac{1}{n}\gamma$.

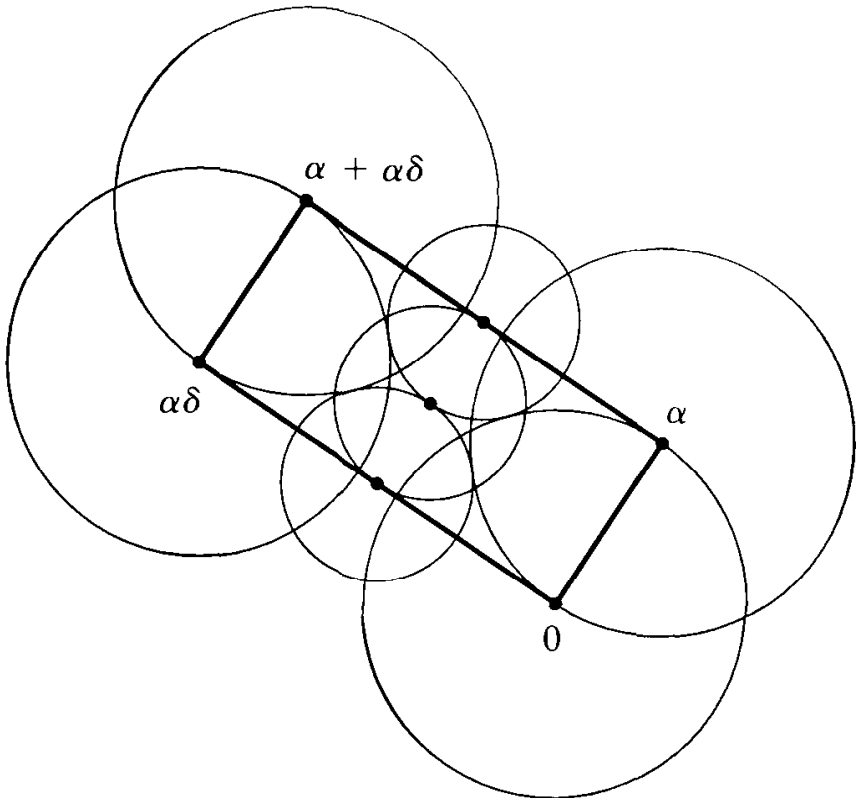
The point $\frac{1}{n}\gamma$ may lie in A or not. This depends on A and on γ.

Proof. Let β be a point in the interior of D. Then by definition of the disc, $|\beta - \frac{1}{n}\gamma| < \frac{1}{n}r$, or equivalently, $|n\beta - \gamma| < r$. If $\beta \in A$, then $n\beta - \gamma \in A$ too.

In this case, $n\beta - \gamma$ is an element of A of absolute value less than r, which implies that $n\beta - \gamma = 0$, hence that $\beta = \frac{1}{n}\gamma$. □

Proof of Theorem (7.9). Let α be the chosen element of A of minimal absolute value r. The principal ideal $(\alpha) = R\alpha$ consists of the complex numbers $(a + b\delta)\alpha$, with $a, b \in \mathbb{Z}$. So it has the lattice basis $(\alpha, \alpha\delta)$ as is asserted in the proposition. Since A contains α, it contains the principal ideal (α) too, and if $A = (\alpha)$ we are in Case 1.

Suppose that $A > (\alpha)$, and let β be an element of A which is not in (α). We may choose β to lie in the rectangle whose four vertices are $0, \alpha, \alpha\delta, a + a\delta$ [see Chapter 5 (4.14)]. Figure (7.13) shows a disc of radius r about the four vertices of this rectangle, and a disc of radius $\frac{1}{2}r$ about the three half lattice points $\frac{1}{2}\alpha\delta, \frac{1}{2}(\alpha + \alpha\delta)$, and $\alpha + \frac{1}{2}\alpha\delta$. Notice that the interiors of these discs cover the rectangle. According to Lemma (7.11), the only points of the interiors which can lie in A are the centers of the discs. Since β is not in (α), it is not a vertex of the rectangle. So β must be one of the half lattice points $\frac{1}{2}\alpha\delta, \frac{1}{2}(\alpha + \alpha\delta)$, or $\alpha + \frac{1}{2}\alpha\delta$.



(7.13) **Figure.**

This exhausts the information which we can get from the fact that A is a lattice. We now use the fact that A is an ideal to rule out the two points $\frac{1}{2}\alpha\delta$ and $\alpha + \frac{1}{2}\alpha\delta$. Suppose that $\frac{1}{2}\alpha\delta \in A$. Multiplying by δ, we find that $\frac{1}{2}\alpha\delta^2 = -\frac{5}{2}\alpha \in A$ too and since $\alpha \in A$ that $\frac{1}{2}\alpha \in A$. This contradicts our choice of α. Next, we note that if $\alpha + \frac{1}{2}\alpha\delta$ were in A, then $\frac{1}{2}\alpha\delta$ would be in A too, which has been ruled out. The remaining possibility is that $\beta = \frac{1}{2}(\alpha + \alpha\delta)$. If so, we are in Case 2. □

8. IDEAL FACTORIZATION

Let R be the ring of integers in an imaginary quadratic field. In order to avoid confusion, we will denote ordinary integers by latin letters $a, b, \ldots$, elements of R by greek letters $\alpha, \beta, \ldots$, and ideals by capital letters $A, B, \ldots$. We will consider only *nonzero* ideals of R.

The notation $A = (\alpha, \beta, \ldots, \gamma)$ stands for the ideal generated by the elements $\alpha, \beta, \ldots, \gamma$. Since an ideal is a plane lattice, it has a lattice basis consisting of two elements. Any lattice basis generates the ideal, but we must distinguish between the notions of a lattice basis and a generating set. We also need to remember the dictionary (2.2) which relates elements to the principal ideals they generate.

Dedekind extended the notion of divisibility to ideals using the following definition of ideal multiplication: Let A and B be ideals in a ring R. We would like to define the product ideal AB to be the set of all products $\alpha\beta$, where $\alpha \in A$ and $\beta \in B$. Unfortunately, this set of products is usually not an ideal: It will not be closed under sums. To get an ideal, we must put into AB all *finite sums of products*

$$\sum_i \alpha_i \beta_i, \quad \text{where } \alpha_i \in A \text{ and } \beta_i \in B. \tag{8.1}$$

The set of such sums is the smallest ideal of R which contains all products $\alpha\beta$, and we denote this *product ideal* by AB. (This use of the product notation is different from its use in group theory [Chapter 2 (8.5)].) The definition of multiplication of ideals is not as simple as we might hope, but it works reasonably well.

Notice that multiplication of ideals is commutative and associative, and that R is a unit element. This is why $R = (1)$ is often called the unit ideal:

$$AR = RA = A, \quad AB = BA, \quad A(BC) = (AB)C. \tag{8.2}$$

(8.3) **Proposition.**

(a) The product of principal ideals is principal: If $A = (\alpha)$ and $B = (\beta)$, then $AB = (\alpha\beta)$.

(b) Assume that $A = (\alpha)$ is principal, but let B be arbitrary. Then

$$AB = \alpha B = \{\alpha\beta \mid \beta \in B\}.$$

(c) Let $\alpha_1, \ldots, \alpha_m$ and $\beta_1, \ldots, \beta_n$ be generators for the ideals A and B respectively. Then AB is generated as an ideal by the mn products $\alpha_i\beta_j$.

We leave this proof as an exercise. □

In analogy with divisibility of elements of a ring, we say that an ideal A *divides* another ideal B if there is an ideal C such that $B = AC$.

To see how multiplication of ideals can be used, let us go back to the example $d = -5$, in which $2 \cdot 3 = (1 + \delta)(1 - \delta)$. For uniqueness of factorization to hold in the ring $R = \mathbb{Z}[\delta]$, there would have to be an element $\rho \in R$ dividing both 2 and

$1 + \delta$. This is the same as saying that 2 and $1 + \delta$ should be in the principal ideal (p). There is no such element. However, there is an *ideal*, not a principal ideal, which contains 2 and $1 + \delta$, namely the ideal generated by these two elements. This ideal $A = (2, 1 + \delta)$ is depicted in Figure (7.10). We can make three other ideals using the factors of 6:

$$\bar{A} = (2, 1 - \delta), \quad B = (3, 1 + \delta), \quad \bar{B} = (3, 1 - \delta).$$

The first of these ideals is denoted by $\bar{A}$ because it is the complex conjugate of the ideal A:

$$\bar{A} = \{\bar{\alpha} \mid \alpha \in A\}. \tag{8.4}$$

As a lattice, $\bar{A}$ is obtained by reflecting the lattice A about the real axis. That the complex conjugate of any ideal is an ideal is easily seen. Actually, it happens that our ideal A is equal to its complex conjugate $\bar{A}$, because $1 - \delta = 2 - (1 + \delta) \in A$. This is an accidental symmetry of the lattice A: The ideals B and $\bar{B}$ are not the same.

Now let us compute the products of these ideals. According to Proposition (8.3c), the ideal $A\bar{A}$ is generated by the four products of the generators $(2, 1 + \delta)$ and $(2, 1 - \delta)$ of A and $\bar{A}$:

$$A\bar{A} = (4, 2 + 2\delta, 2 - 2\delta, 6).$$

Each of these four generators is divisible by 2, so $A\bar{A} \subset (2)$. On the other hand, $2 = 6 - 4$ is in $A\bar{A}$. Therefore $(2) \subset A\bar{A}$, so

$$A\bar{A} = (2)\,!$$

[The notation (2) is ambiguous, because it can denote both $2R$ and $2\mathbb{Z}$. It stands for $2R$ here.] Next, the product AB is generated by the four products:

$$AB = (6, 2 + 2\delta, 3 + 3\delta, -4 + 2\delta).$$

Each of these four elements is divisible by $1 + \delta$. Since $1 + \delta$ is in AB, we find that $AB = (1 + \delta)$. Similarly, $\bar{A}\bar{B} = (1 - \delta)$ and $B\bar{B} = (3)$.

It follows that the principal ideal (6) is the product of the four ideals:

$$(6) = (2)(3) = (A\bar{A})(B\bar{B}) = (AB)(\bar{A}\bar{B}) = (1 + \delta)(1 - \delta). \tag{8.5}$$

Isn't this beautiful? The ideal factorization $(6) = A\bar{A}B\bar{B}$ has provided a common refinement of the two factorizations (2.7).

The rest of this section is devoted to proving unique factorization of ideals in the rings of integers of an imaginary quadratic number field. We will follow the discussion of factorization of elements as closely as possible.

The first thing to do is to find an analogue for ideals of the notion of a prime element.

(8.6) **Proposition.** Let P be an ideal of a ring R which is not the unit ideal. The following conditions are equivalent:

(i) If α, β are elements of R such that $\alpha\beta \in P$, then $\alpha \in P$ or $\beta \in P$.

(ii) If A, B are ideals of R such that $AB \subset P$, then $A \subset P$ or $B \subset P$.

(iii) The quotient ring R/P is an integral domain.

An ideal which satisfies one of these conditions is called a *prime ideal.*

For example, every maximal ideal is prime, because if M is maximal, then R/M is a field, and a field is an integral domain. The zero ideal of a ring R is prime if and only if R is an integral domain.

Proof of the Proposition: The conditions for $\bar{R} = R/P$ to be an integral domain are that $\bar{R} \neq 0$ and that $\overline{\alpha\beta} = 0$ implies $\bar{\alpha} = 0$ or $\bar{\beta} = 0$. These conditions translate back to $P \neq R$ and if $\alpha\beta \in P$ then $\alpha \in P$ or $\beta \in P$. Thus (i) and (iii) are equivalent. The fact that (ii) implies (i) is seen by taking $A = (\alpha)$ and $B = (\beta)$. The only surprising implication is that (i) implies (ii). Assume that (i) holds, and let A, B be ideals such that $AB \subset P$. If A is not contained in P, there is some element $\alpha \in A$ which is not in P. If β is an element of B, then $\alpha\beta \in AB$; hence $\alpha\beta \in P$. By part (i), $\beta \in P$. Since this is true for all of its elements, $B \subset P$ as required. □

We now go back to imaginary quadratic number fields.

(8.7) **Lemma.** Let $A \subset B$ be lattices in $\mathbb{R}^2$. There are only finitely many lattices L between A and B, that is, such that $A \subset L \subset B$.

Proof. Let (α_1, α_2) be a lattice basis for A, and let P be the parallelogram with vertices $0, \alpha_1, \alpha_2, \alpha_1 + \alpha_2$. There are finitely many elements of B contained in P [Chapter 5 (4.12)], so if L is a lattice between A and B, there are finitely many possibilities for the set $L \cap P$. Call this set S. The proof will be completed by showing that S and A determine the lattice L. To show this, let γ be an element of L. Then there is an element of $\alpha \in A$ such that $\gamma - \alpha$ is in P, hence in S. [See the proof of (4.14) in Chapter 5]. Symbolically, we have $L = S + A$. This describes L in terms of S and A, as required. □

(8.8) **Proposition.** Let R be the ring of integers in an imaginary quadratic number field.

(a) Let B be a nonzero ideal of R. There are finitely many ideals between B and R.

(b) Every proper ideal of R is contained in a maximal ideal.

(c) The nonzero prime ideals of R are the maximal ideals.

Proof.

(a) This follows from lemma (8.7).

(b) Let B be a proper ideal. Then B is contained in only finitely many ideals. We can search through them to find a maximal ideal.

(c) We have already remarked that maximal ideals are prime. Conversely, let P be a nonzero prime ideal. Then P has finite index in R. So R/P is a finite integral do-

main, and hence it is a field [Chapter 10 (6.4)]. This shows that P is a maximal ideal. □

(8.9) **Theorem.** Let R be the ring of integers in an imaginary quadratic field F. Every nonzero ideal of R which is not the whole ring is a product of prime ideals. This factorization is unique, up to order of the factors.

This remarkable theorem can be extended to other rings of algebraic integers, but it is a very special property of such rings. Most rings do not admit unique factorization of ideals. Several things may fail, and we want to take particular note of one of them. We know that a principal ideal (α) contains another principal ideal (β) if and only if α divides β in the ring. So the definition of a prime element π can be restated as follows: If $(\pi) \supset (\alpha\beta)$, then $(\pi) \supset (\alpha)$ or $(\pi) \supset (\beta)$. The second of the equivalent definitions (8.6) of a prime ideal is the analogous statement for ideals: If $P \supset AB$, then $P \supset A$ or $P \supset B$. So if inclusion of ideals were equivalent with divisibility, the proof of uniqueness of factorizations would carry over to ideals. Unfortunately the cumbersome definition of product ideal causes trouble. In most rings, the inclusion $A \supset B$ does not imply that A divides B. This weakens the analogy between prime ideal and prime element. It will be important to establish the equivalence of inclusion and divisibility in the particular rings we are studying. This is done below, in Proposition (8.11).

We now proceed with the proof of Theorem (8.9). For the rest of this section, R will denote the ring of integers in an imaginary quadratic number field. The proof is based on the following lemma:

(8.10) **Main Lemma.** Let R be the ring of integers in an imaginary quadratic number field. The product of a nonzero ideal and its conjugate is a principal ideal of R generated by an ordinary integer:

$$A\bar{A} = (n), \quad \text{for some } n \in \mathbb{Z}.$$

The most important point here is that for every ideal A there is some ideal B such that AB is principal. That $\bar{A}$ does the job and that the product ideal is generated by an ordinary integer are less important points.

We will prove the lemma at the end of the section. Let us assume it for now and derive some consequences for multiplication of ideals. Because these consequences depend on the Main Lemma, they are not true for general rings.

(8.11) **Proposition.** Let R be the ring of integers in an imaginary quadratic number field.

(a) *Cancellation Law:* Let A, B, C be nonzero ideals of R. If $AB \supset AC$ then $B \supset C$. If $AB = AC$, then $B = C$.

(b) If A and B are nonzero ideals of R, then $A \supset B$ if and only if A divides B, that is, if and only if $B = AC$ for some ideal C.

(c) Let P be a nonzero prime ideal of R. If P divides a product AB of ideals, then P divides one of the factors A or B.

Proof. (a) Assume that $AB \supset AC$. If $A = (\alpha)$ is principal, then $AB = \alpha B$ and $AC = \alpha C$ (8.3). Viewing these sets as subsets of the complex numbers, we multiply the relation $\alpha B \supset \alpha C$ on the left by α^{-1} to conclude that $B \supset C$. So the assertion is true when A is principal. In general, if $AB \supset AC$, then multiply both sides by $\bar{A}$ and apply the Main Lemma: $nB = \bar{A}AB \supset \bar{A}AC = nC$, and apply what has been shown. The case that $AB = AC$ is the same.

(b) The implication which is not clear is that if A contains B then A divides B. We will first check this when $A = (\alpha)$ is principal. In this case, to say that $(\alpha) \supset B$ means that α divides every element β of B. Let $C = \alpha^{-1}B$ be the set of quotients, that is, the set of elements $\alpha^{-1}\beta$, with $\beta \in B$. You can check that C is an ideal and that $\alpha C = B$. Hence $B = AC$ in this case. Now let A be arbitrary, and assume that $A \supset B$. Then $(n) = \bar{A}A \supset \bar{A}B$. By what has already been shown, there is an ideal C such that $nC = \bar{A}B$, or $\bar{A}AC = \bar{A}B$. By the Cancellation Law, $AC = B$.

(c) To prove part (c) of the proposition, we apply part (b) to translate divisibility into inclusion. Then (c) follows from the definition of prime ideal. □

Proof of Theorem (8.9). There are two things to prove. First we must show that every proper, nonzero ideal A is a product of prime ideals. If A is not itself prime, then it is not maximal, so we can find a proper ideal A_1 strictly larger than A. Then A_1 divides A (8.11b), so we can write $A = A_1B_1$. It follows that $A \subset B_1$. Moreover, if we had $A = B_1$, the Cancellation Law would imply $R = A_1$, contradicting the fact that A_1 is a proper ideal. Thus $A < B_1$. Similarly, $A < A_1$. Since there are only finitely many ideals between A and R, this process of factoring an ideal terminates. When it does, all factors will be maximal, and hence prime. So every proper ideal A can be factored into primes.

Now to prove uniqueness, we apply the property (8.11c) of prime ideals: If $P_1 \cdots P_r = Q_1 \cdots Q_s$, with P_i, Q_j prime, then P_1 divides $Q_1 \cdots Q_s$, and hence it divides one of the factors, say Q_1. Since Q_1 is maximal, $P_1 = Q_1$. Cancel by (8.11a) and use induction on r. □

(8.12) **Theorem.** The ring of integers R is a unique factorization domain if and only if it is a principal ideal domain. If so, then the factorizations of elements and of ideals correspond naturally.

Proof. We already know that a principal ideal domain has unique factorization (2.12). Conversely, suppose that R is a unique factorization domain, and let P be any nonzero prime ideal of R. Then P contains an irreducible element, say π. For, any nonzero element α of P is a product of irreducible elements, and, by definition of prime ideal, P contains one of its irreducible factors. By (2.8), an irreducible element π is prime, that is, (π) is a prime ideal. By (8.6), (π) is maximal. Since

$(\pi) \subset P$, it follows that $(\pi) = P$, hence that P is principal. By Theorem (8.9), every nonzero ideal A is a product of primes; hence it is principal (8.3a). Thus R is a principal ideal domain. The last assertion of the theorem is clear from (2.2). □

Proof of the Main Lemma (8.10). We can generate A as a lattice by two elements, say α, β. Then A is certainly generated as an ideal by these same elements, and moreover $\overline{\alpha}, \overline{\beta}$ generate $\overline{A}$. Hence the four products $\alpha\overline{\alpha}$, $\alpha\overline{\beta}$, $\overline{\alpha}\beta$, $\beta\overline{\beta}$ generate the ideal $A\overline{A}$. Consider the three elements $\alpha\overline{\alpha}$, $\beta\overline{\beta}$, and $\alpha\overline{\beta} + \overline{\alpha}\beta$ of $A\overline{A}$. They are all equal to their conjugates and hence are rational numbers. Since they are algebraic integers, they are ordinary integers. Let n be their greatest common divisor in $\mathbb{Z}$. Then n is a linear combination of $\alpha\overline{\alpha}$, $\beta\overline{\beta}$, $\alpha\overline{\beta} + \overline{\alpha}\beta$ with integer coefficients. Hence n is in the product ideal $A\overline{A}$. Therefore $A\overline{A} \supset (n)$. If we show that n divides each of the four generators of the ideal $A\overline{A}$ in R, then it will follow that $(n) \supset A\overline{A}$, hence that $(n) = A\overline{A}$, as was to be shown.

Now by construction, n divides $\alpha\overline{\alpha}$ and $\beta\overline{\beta}$ in $\mathbb{Z}$, hence in R. So we have to show that n divides $\alpha\overline{\beta}$ and $\overline{\alpha}\beta$ in R. The elements $(\alpha\overline{\beta})/n$ and $(\overline{\alpha}\beta)/n$ are roots of the polynomial $x^2 - rx + s$, where

$$r = \frac{\alpha\overline{\beta} + \overline{\alpha}\beta}{n} \quad \text{and} \quad s = \frac{\alpha\overline{\alpha}}{n}\frac{\beta\overline{\beta}}{n}.$$

By definition of n, these elements r, s are integers, so this is a monic equation in $\mathbb{Z}[x]$. Hence $(\alpha\overline{\beta})/n$ and $(\overline{\alpha}\beta)/n$ are algebraic integers, as required. □

Note. This is the only place where the definition of algebraic integer is used directly. The lemma would be false if we took a smaller ring than R, for example, if we didn't take the elements with half integer coefficients when $d \equiv 1$ (modulo 4).

9. THE RELATION BETWEEN PRIME IDEALS OF R AND PRIME INTEGERS

We saw in Section 5 how the primes in the ring of Gauss integers are related to integer primes. A similar analysis can be made for the ring R of integers in a quadratic number field. The main difference is that R is usually not a principal ideal domain, and therefore we should speak of prime ideals rather than of prime elements. This complicates the analogues of parts (c) and (d) of Theorem (5.1), and we will not consider them here. [However, see (12.10).]

(9.1) **Proposition.** Let P be a nonzero prime ideal of R. There is an integer prime p so that either $P = (p)$ or $P\overline{P} = (p)$. Conversely, let p be a prime integer. There is a prime ideal P of R so that either $P = (p)$ or $P\overline{P} = (p)$.

The proof follows that of parts (a) and (b) of Theorem (5.1) closely. □

The second case of (9.1) is often subdivided into two cases, according to whether or not P and $\bar{P}$ are equal. The following terminology is customary: If (p) is a prime ideal, then we say that p *remains prime* in R. If $P\bar{P} = (p)$, then we say that p *splits* in R, unless $P = \bar{P}$, in which case we say that P *ramifies* in R.

Let us analyze the behavior of primes further. Assume that $d \equiv 2$ or 3 (modulo 4). In this case, $R = \mathbb{Z}[\delta]$ is isomorphic to $\mathbb{Z}[x]/(x^2 - d)$. To ask for prime ideals containing the ideal (p) is equivalent to asking for prime ideals of the ring $R/(p)$ [Chapter 10 (4.3)]. Note that

$$R/(p) \approx \mathbb{Z}[x]/(x^2 - d,p). \tag{9.2}$$

Interchanging the order of the two relations $x^2 - d = 0$ and $p = 0$ as in the proof of Theorem (5.1), we find the first part of the proposition below. The second part is obtained in the same way, using the polynomial (6.16).

(9.3) **Proposition.**

(a) Assume that $d \equiv 2$ or 3 (modulo 4). An integer prime p remains prime in R if and only if the polynomial $x^2 - d$ is irreducible over $\mathbb{F}_p$.

(b) Assume that $d \equiv 1$ (modulo 4). Then p remains prime if and only if the polynomial $x^2 - x + \frac{1}{4}(1 - d)$ is irreducible over $\mathbb{F}_p$. □

10. IDEAL CLASSES IN IMAGINARY QUADRATIC FIELDS

As before, R denotes the ring of integers in an imaginary quadratic number field. In order to analyze the extent to which uniqueness of factorization of elements fails in R, we introduce an equivalence relation on ideals which is compatible with ideal multiplication and such that the principal ideals form one equivalence class. It is reasonably clear which relation to use: We call two ideals A, B *similar* ($A \sim B$) if there are nonzero elements $\sigma, \tau \in R$ so that

$$\sigma B = \tau A. \tag{10.1}$$

This is an equivalence relation. The equivalence classes for this relation are called *ideal classes*, and the ideal class of A will be denoted by $\langle A \rangle$.

We could also take the element $\lambda = \sigma^{-1}\tau$ of the quadratic number field $F = \mathbb{Q}[\delta]$ and say that A and B are similar if

$$B = \lambda A, \quad \text{for some } \lambda \in \mathbb{Q}[\delta]. \tag{10.2}$$

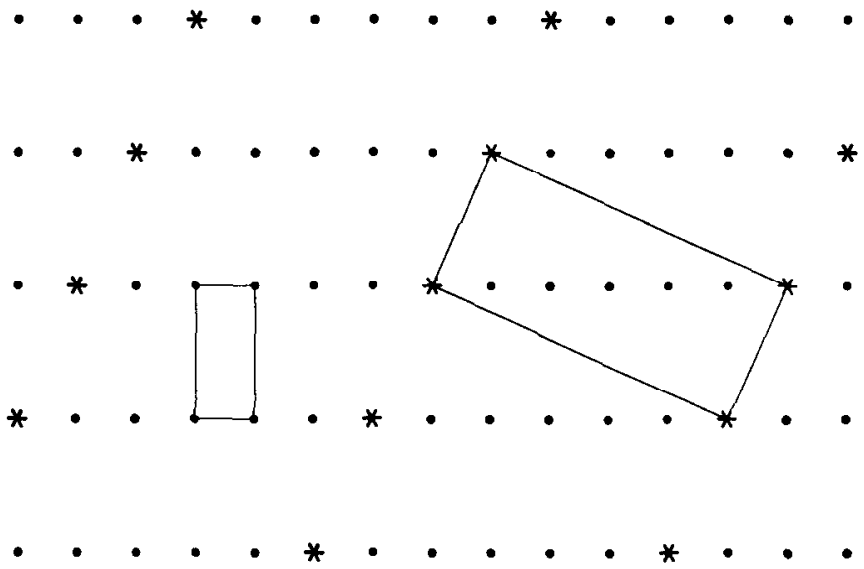
Similarity has a nice geometric interpretation. Two ideals A and B are similar if the lattices in the complex plane which represent them are similar geometric figures, by a similarity which is *orientation-preserving*. To see this, note that a lattice looks the same at all points. So a similarity can be assumed to relate 0 in A to 0 in B. Then it will be described as a rotation followed by a stretching or shrinking,

that is, as multiplication by a complex number λ. Since multiplication by λ carries a nonzero element $\alpha \in A$ to an element $\lambda\alpha = \beta \in B$, $\lambda = \beta\alpha^{-1}$ is automatically in the field F.

An ideal B is similar to the unit ideal R if and only if $B = \lambda R$ for some λ in the field. Then λ is an element of B, hence of R. In this case, B is the principal ideal (λ). So we have the following:

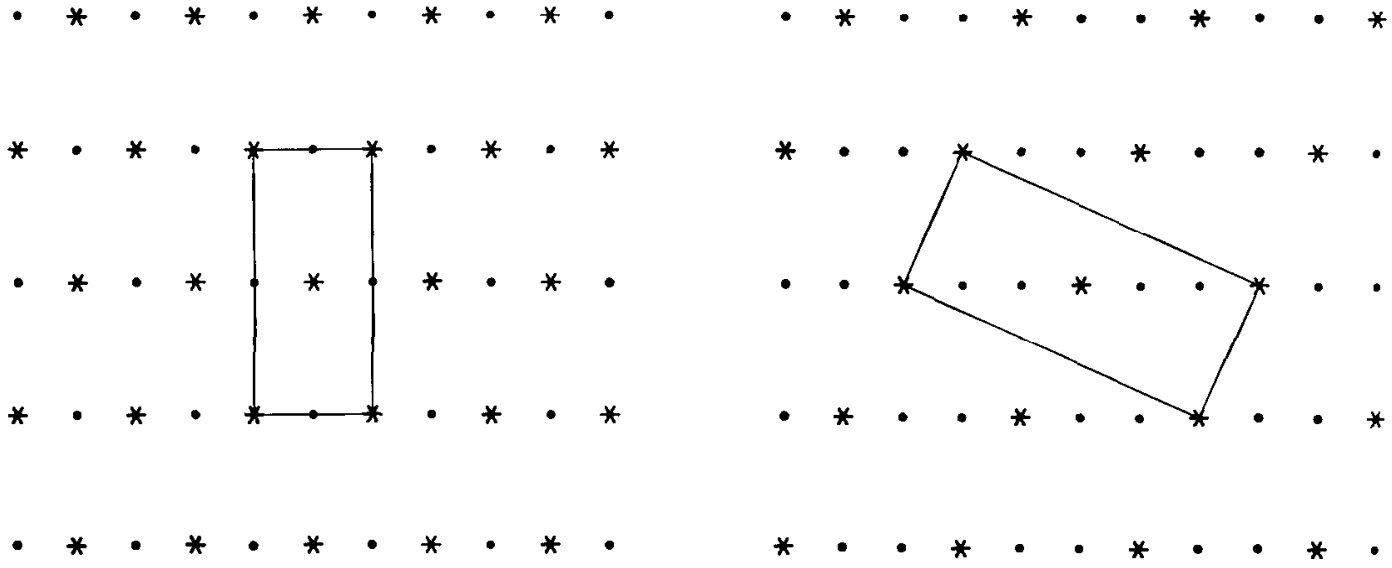
(10.3) **Proposition.** The ideal class $\langle R \rangle$ consists of the principal ideals. □

Figure (10.4) shows the principal ideal $(1 + \delta)$ in the ring $\mathbb{Z}[\delta]$, where $\delta^2 = -5$.



(10.4) **Figure.** The principal ideal $1 + \delta$.

We saw in (7.9) that there are two ideal classes. Each of the ideals $A = (2, 1 + \delta)$ and $B = (3, 1 + \delta)$, for example, represents the class of nonprincipal ideals. In this case $2B = (1 + \delta)A$. These ideals are depicted in Figure (10.5).



(10.5) **Figure.** The ideals $(2, 1 + \delta)$ and $(3, 1 + \delta)$.

(10.6) **Proposition.** The ideal classes form an abelian group $\mathscr{C}$, with law of composition induced by multiplication of ideals:

$$\langle A\rangle\langle B\rangle = \text{class of } AB = \langle AB\rangle;$$

the class of the principal ideals is the identity: $\langle R\rangle = \langle 1\rangle$.

Proof. If $A \sim A'$ and $B \sim B'$, then $A' = \lambda A$ and $B' = \mu B$ for some $\lambda, \mu \in F = \mathbb{Q}[\delta]$; hence $A'B' = \lambda\mu AB$. This shows that $\langle AB\rangle = \langle A'B'\rangle$, hence that this law of composition is well-defined. Next, the law is commutative and associative because multiplication of ideals is, and the class of R is an identity (8.2). Finally, $A\bar{A} = (n)$ is principal by the Main Lemma (8.10). Since the class of the principal ideal (n) is the identity in $\mathscr{C}$, we have $\langle A\rangle\langle\bar{A}\rangle = \langle R\rangle$, so $\langle\bar{A}\rangle = \langle A\rangle^{-1}$. □

(10.7) **Corollary.** Let R be the ring of integers in an imaginary quadratic number field. The following assertions are equivalent:

(i) R is a principal ideal domain;

(ii) R is a unique factorization domain;

(iii) the ideal class group $\mathscr{C}$ of R is the trivial group.

For to say that $\mathscr{C}$ is trivial is the same as saying that every ideal is similar to the unit ideal, which by Proposition (10.3) means that every ideal is principal. By Theorem (8.12), this occurs if and only if R is a unique factorization domain. □

Because of Corollary (10.7), it is natural to count the ideal classes and to consider this count, called the *class number*, a measure of nonuniqueness of factorization of elements in R. More precise information is given by the structure of $\mathscr{C}$ as a group. As we have seen (7.9), there are two ideal classes in the ring $\mathbb{Z}[\sqrt{-5}]$, so its ideal class group is a cyclic group of order 2 and its class number is 2.

We will now show that the ideal class group $\mathscr{C}$ is always a finite group. The proof is based on a famous lemma of Minkowski about lattice points in convex regions. A bounded subset S of the plane $\mathbb{R}^2$ is called *convex* and *centrally symmetric* if it has these properties:

(10.8) (a) *Convexity:* If $p, q \in S$, then the line segment joining p to q is in S.

(b) *Central symmetry:* If $p \in S$, then $-p \in S$.

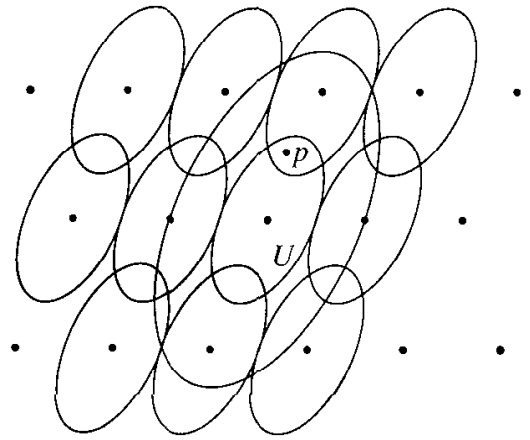
Notice that these conditions imply that $0 \in S$, unless S is empty.

(10.9) **Minkowski's Lemma.** Let L be a lattice in $\mathbb{R}^2$, and let S be a convex, centrally symmetric subset of $\mathbb{R}^2$. Let $\Delta(L)$ denote the area of the parallelogram spanned by a lattice basis for L. If

$$\text{Area}(S) > 4\Delta(L),$$

then S contains a lattice point other than 0.

Proof. Define U to be the convex set similar to S, but with half the linear dimension. In other words, we put $p \in U$ if $2p \in S$. Then U is also convex and centrally symmetric, and $\text{Area}(U) = \frac{1}{4}\text{Area}(S)$. So the above inequality can be restated as $\text{Area}(U) > \Delta(L)$.



(10.10) **Figure.**

(10.11) **Lemma.** There is an element $\alpha \in L$ such that $U \cap (U + \alpha)$ is not empty.

Proof. Let P be the parallelogram spanned by a lattice basis for L. The translates $P + \alpha$ with $\alpha \in L$ cover the plane without overlapping except along their edges. The heuristic reason that the lemma is true is this: There is one translate $U + \alpha$ for each translate $P + \alpha$, and the area of U is larger than the area of P. So the translates $U + \alpha$ must overlap. To make this precise, we note that since U is a bounded set, it meets finitely many of the translates $P + \alpha$, say it meets $P + \alpha_1, \ldots, P + \alpha_k$. Denote by U_i the set $(P + \alpha_i) \cap U$. Then U is cut into the pieces $U_1, \ldots, U_k$, and $\text{Area}(U) = \Sigma\ \text{Area}(U_i)$. We translate U_i back to P by subtracting α_i, setting $V_i = U_i - \alpha_i$, and we note that $V_i = P \cap (U - \alpha_i)$. So V_i is a subset of P, and $\text{Area}(V_i) = \text{Area}(U_i)$. Then $\Sigma\ \text{Area}(V_i) = \text{Area}(U) > \Delta(L) = \text{Area}(P)$. This implies that two of the sets V_i must overlap, that is, that for some $i \neq j$, $(U - \alpha_i) \cap (U - \alpha_j)$ is nonempty. Adding α_i and setting $\alpha = \alpha_i - \alpha_j$, we find that $U \cap (U + \alpha)$ is nonempty too.

Returning to the proof of Minkowski's Lemma, choose α as in Lemma (10.11), and let p be a point of $U \cap (U + \alpha)$. From $p \in U + \alpha$, it follows that $p - \alpha \in U$. By central symmetry, $q = \alpha - p \in U$ too. The midpoint between p and q is $\frac{1}{2}\alpha$, which is also in U, because U is convex. Therefore $\alpha \in S$, as required. □

(10.12) **Corollary.** Any lattice L in $\mathbb{R}^2$ contains a nonzero vector α such that

$$|\alpha|^2 \leq 4\Delta(L)/\pi.$$

Proof. We apply Minkowski's Lemma, taking for S a circle of radius r about the origin. The lemma guarantees the existence of a nonzero lattice point in S, provided that $\pi r^2 > 4\Delta(L)$, or that $r^2 > 4\Delta(L)/\pi$. So for any positive number ϵ, there is a lattice point α with $|\alpha|^2 < 4\Delta(L)/\pi + \epsilon$. Since there are only finitely many lattice points in a bounded region and since ϵ can be arbitrarily small, there is a lattice point satisfying the desired inequality. □

We now return to ideals in the ring R of integers in an imaginary quadratic field. There are two measures for the size of an ideal, which turn out to be the same. The first is the index in R. Since an ideal A is a sublattice of R, it has finite index:

$$[R : A] = \textit{number of additive cosets of } A \textit{ in } R.$$

This index can be expressed in terms of the area of the parallelogram spanned by basis vectors:

(10.13) **Lemma.** Let (a_1, a_2) and (b_1, b_2) be lattice bases for lattices $B \supset A$ in $\mathbb{R}^2$, and let $\Delta(A)$ and $\Delta(B)$ be the areas of the parallelograms spanned by these bases. Then $[B : A] = \Delta(A)/\Delta(B)$.

We leave the proof as an exercise. □

(10.14) **Corollary.**

(a) Let A be a plane lattice. The area $\Delta(A)$ is independent of the lattice basis for A.
(b) If $C \supset B \supset A$ are lattices, then $[C:A] = [C:B][B:A]$. □

It is easy to compute the area $\Delta(R)$ using the description (6.14) of the ring:

$$\Delta(R) = \tfrac{1}{2}\sqrt{|D|} = \begin{cases} \sqrt{|d|} & \text{if } d \equiv 2, 3 \pmod 4 \\ \tfrac{1}{2}\sqrt{|d|} & \text{if } d \equiv 1 \pmod 4 \end{cases}, \tag{10.15}$$

where D is the discriminant (6.18).

The other measure of the size of an ideal can be obtained from the Main Lemma (8.10): We write $A\bar{A} = (n)$ and take the integer n (chosen > 0, of course). This is analogous to the norm of an element (7.1) and is therefore called the *norm* of the ideal:

$$N(A) = n, \quad \text{if } A\bar{A} = (n). \tag{10.16}$$

It has the multiplicative property

$$N(AB) = N(A)N(B), \tag{10.17}$$

because $AB\overline{AB} = A\bar{A}B\bar{B} = (nm)$ if $N(B) = m$. Note also that if A is the principal ideal (α), then its norm is the norm of α:

$$N((\alpha)) = \alpha\bar{\alpha} = N(\alpha), \tag{10.18}$$

because $(\alpha)(\bar{\alpha}) = (\alpha\bar{\alpha})$.

(10.19) **Lemma.** For any nonzero ideal A of R,

$$[R : A] = N(A).$$

(10.20) **Corollary.** *Multiplicative property of the index:* Let A and B be nonzero ideals of R. Then

$$[R : AB] = [R : A][R : B]. \square$$

Let us defer the proof of Lemma (10.19) and derive the finiteness of the class number from it.

(10.21) **Theorem.** Let $\mu = 2\sqrt{|D|}/\pi$. Every ideal class contains an ideal A such that $N(A) \leq \mu$.

Proof. Let A be an ideal. We have to find another ideal A' in the class of A whose norm is not greater than μ. We apply Corollary (10.12): There is an element $\alpha \in A$ with

$$N(\alpha) = |\alpha|^2 \leq 4\Delta(A)/\pi.$$

Then $A \supset (\alpha)$. This implies that A divides (α), that is, that $AC = (\alpha)$ for some ideal C. By the multiplicative property of norms (10.17) and by (10.18), $N(A)N(C) = N(\alpha) \leq 4\Delta(A)/\pi$. Using (10.13), (10.14), and (10.19), we write $\Delta(A) = [R:A]\Delta(R) = \frac{1}{2}N(A)\sqrt{|D|}$. Substituting for $\Delta(A)$ and cancelling $N(A)$, we find $N(C) \leq \mu$.

Now since CA is a principal ideal, the class $\langle C \rangle$ is the inverse of $\langle A \rangle$, i.e., $\langle C \rangle = \langle \bar{A} \rangle$. So we have shown that $\langle \bar{A} \rangle$ contains an ideal whose norm satisfies the required inequality. Interchanging the roles of A and $\bar{A}$ completes the proof. $\square$

The finiteness of the class number follows easily:

(10.22) **Theorem.** The ideal class group $\mathcal{C}$ is finite.

Proof. Because of (10.19) and (10.21), it is enough to show that there are finitely many ideals with index $[R:A] \leq \mu$, so it is enough to show that there are only finitely many sublattices $L \subset R$ with $[R : L] \leq \mu$. Choose an integer $n \leq \mu$, and let L be a sublattice such that $[R : L] = n$. Then R/L is an abelian group of order n, so multiplication by n is the zero map on this group. The translation of this fact to R is the statement $nR \subset L$: Sublattices of index n contain nR. Lemma (8.7) implies that there are finitely many such lattices L. Since there are also finitely many possibilities for n, we are done. $\square$

The ideal class group can be computed explicitly by checking which of the sublattices $L \subset R$ of index $\leq \mu$ are ideals. However, this is not efficient. It is better to look directly for prime ideals. Let $[\mu]$ denote the largest integer less than μ.

(10.23) **Proposition.** The ideal class group $\mathcal{C}$ is generated by the classes of the prime ideals P which divide integer primes $p \leq [\mu]$.

Proof. We know that every class contains an ideal A of norm $N(A) \leq \mu$, and since $N(A)$ is an integer, $N(A) \leq [\mu]$. Suppose that an ideal A with norm $\leq \mu$ is factored into prime ideals: $A = P_1 \cdots P_r$. Then $N(A) = N(P_1) \cdots N(P_k)$, by (10.17). Hence $N(P_i) \leq [\mu]$ for each i. So the classes of prime ideals P of norm $\leq [\mu]$ form a set of generators of $\mathcal{C}$, as claimed. □

To apply this proposition, we examine each prime integer $p \leq [\mu]$. If p remains prime in R, then the prime ideal (p) is principal, so its class is trivial. We throw out these primes. If p does not remain prime in R, then we include the class of one of its two prime ideal factors P in our set of generators. The class of the other prime factor is its inverse. It may still happen that P is a principal ideal, in which case we discard it. The remaining primes generate $\mathcal{C}$.

Table (10.24) gives a few values which illustrate different groups.

TABLE 10.24 SOME IDEAL CLASS GROUPS

d	D	$[\mu]$	*Ideal class group*
−2	−8	1	trivial
−5	−20	2	order 2
−13	−52	4	order 2
−14	−56	4	order 4, cyclic
−21	−84	5	Klein four group
−23	−23	3	order 3
−26	−104	6	order 6
−47	−47	4	order 5
−71	−71	5	order 7

(10.25) **Examples.** To apply Proposition (10.23), we factor (p) into prime ideals for all prime integers $p \leq \mu$.

(a) $d = -7$. In this case $[\mu] = 1$. Proposition (10.23) tells us that the class group $\mathcal{C}$ is generated by the empty set of prime ideals. So $\mathcal{C}$ is trivial, and R is a unique factorization domain.

(b) $d = -67$. Here $R = \mathbb{Z}[\eta]$, where $\eta = \frac{1}{2}(1 + \delta)$, and $[\mu] = 5$. The ideal class group is generated by the prime ideals dividing $2, 3, 5$. According to Proposition (9.3), a prime integer p remains prime in R if and only if the polynomial $x^2 - x + 17$ is irreducible modulo p. This is true for each of the primes $2, 3, 5$. So the primes in question are principal, and the ideal class group is trivial.

(c) $d = -14$. Here $[\mu] = 4$, so $\mathcal{C}$ is generated by prime ideals dividing (2) and (3). The polynomial $x^2 + 14$ is reducible, both modulo 2 and modulo 3, so by (9.3) neither of these integers remains prime in R. Say that $(2) = P\bar{P}$ and $(3) = Q\bar{Q}$. As in the discussion of $\mathbb{Z}[\sqrt{-5}]$, we find that $P = (2, \delta) = \bar{P}$. The ideal class $\langle P \rangle$ has order 2 in $\mathcal{C}$.

To compute the order of the class $\langle Q \rangle$, we may compute the powers of the ideal explicitly and find the first power whose lattice is similar to R. This is not efficient. It is better to compute the norms of a few small elements of R, hoping to deduce a relation among the generators. The most obvious elements to try are δ and $1 + \delta$. But $N(\delta) = 14$ and $N(1 + \delta) = 15$. These are not as good as we may hope for, because they involve the primes 7 and 5, whose factors are not among our generators. We'd rather not bring in these extra primes. The element $2 + \delta$ is better: $N(2 + \delta) = (2 + \delta)(2 - \delta) = 2 \cdot 3 \cdot 3$. This gives us the ideal relation

$$(2 + \delta)(2 - \delta) = P\bar{P}Q\bar{Q}Q\bar{Q} = P^2Q^2\bar{Q}^2.$$

Since $2 + \delta$ and $2 - \delta$ are not associates, they do not generate the same ideal. On the other hand, they generate conjugate ideals. Taking these facts into account, the only possible prime factorizations of $(2 + \delta)$ are PQ^2 and $P\bar{Q}^2$. Which case we have depends on which factor of (3) we label as Q. So we may suppose that $(2 + \delta) = PQ^2$. Then since $(2 + \delta)$ is a principal ideal, $\langle P \rangle \langle Q \rangle^2 = 1$ in $\mathcal{C}$. Hence $\langle Q \rangle^2 = \langle P \rangle^{-1} = \langle P \rangle$. This shows that $\mathcal{C}$ is the cyclic group of order 4 generated by $\langle Q \rangle$.

(d) $d = -23$, and hence $R = \mathbb{Z}[\eta]$ where $\eta = \frac{1}{2}(1 + \delta)$. Then $[\mu] = 3$, so $\mathcal{C}$ is generated by the classes of the prime ideals dividing (2) and (3). Both of these primes split in R, because the polynomial $x^2 - x + 6$ is reducible modulo 2 and modulo 3 (9.3). In fact, $(2) = P\bar{P}$, where P has the lattice base $(2,\eta)$ [see (7.8)]. This is not a principal ideal.

Say that $(3) = Q\bar{Q}$. To determine the structure of the ideal class group, we note that $N(\eta) = 2 \cdot 3$ and $N(1+\eta) = 2 \cdot 2 \cdot 2$. Therefore

$$(\eta)(\overline{\eta}) = P\bar{P}Q\bar{Q} \quad \text{and} \quad (1+\eta)(\overline{1+\eta}) = (8) = (2)^3 = P^3\bar{P}^3.$$

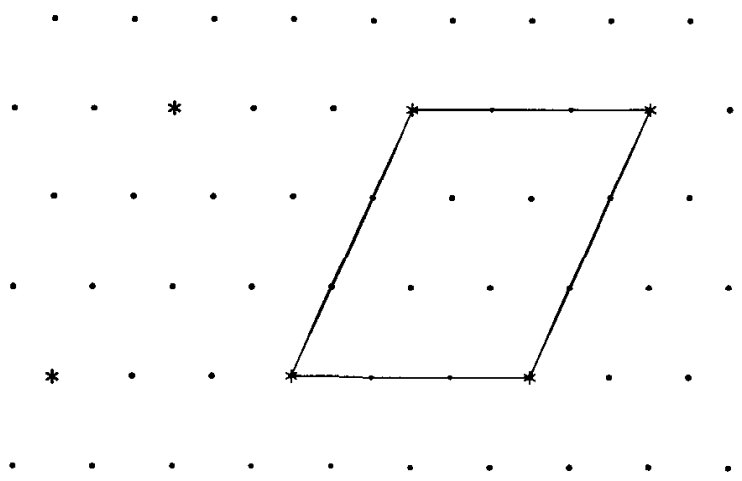
Interchanging the roles of $P, \bar{P}$ and of $Q, \bar{Q}$ as necessary, we obtain $(\eta) = PQ$ and $(1 + \eta) = P^3$ or $\bar{P}^3$. Therefore $\langle P \rangle^3 = \langle 1 \rangle$ and $\langle Q \rangle = \langle P \rangle^{-1}$ in $\mathcal{C}$. The ideal class group is a cyclic group of order 3. □

Proof of Lemma (10.19). This lemma is true for the unit ideal R. We will prove that $[R : P] = N(P)$ if P is a prime ideal, and we will show that if P is prime and if A is an arbitrary nonzero ideal, then $[R : AP] = [R : A][R : P]$. It will follow that if $[R:A] = N(A)$, then $[R:AP] = N(AP)$. Induction on the length of the prime factorization of an ideal will complete the proof.

(10.26) **Lemma.** Let n be an ordinary integer, and let A be an ideal. Then

$$[R : nA] = n^2[R : A].$$

Proof. We know that $R \supset A \supset nA$, and therefore (10.14b) $[R : nA] = [R : A][A : nA]$. Thus we must show that $[A : nA] = n^2$. Now A is a lattice, and nA is the sublattice obtained by stretching by the factor n:



(10.27) **Figure.** $3A = \{*\}$.

Clearly, $[A : nA] = n^2$, as required. □

We return to the proof of Lemma (10.19). There are two cases to consider for the ideal P. According to (9.1), there is an integer prime p so that either $P = (p)$ or $P\bar{P} = (p)$.

In the first case, $N(P) = p^2$, and $AP = pA$. We can use Lemma (10.26) twice to conclude that $[R : AP] = p^2[R : A]$ and $[R : P] = p^2[R : R] = p^2$. Thus $[R : AP] = [R : A][R : P]$ and $[R : P] = N(P)$, as required.

In the second case, $N(P) = p$. We consider the chain of ideals $A > AP > AP\bar{P}$. It follows from the Cancellation Law (8.11a) that this is a strictly decreasing chain of ideals, hence that

$$[R : A] < [R : AP] < [R : AP\bar{P}]. \tag{10.28}$$

Also, since $P\bar{P} = (p)$, we have $AP\bar{P} = pA$. Therefore we may apply Lemma (10.26) again, to conclude that $[R : AP\bar{P}] = p^2[R : A]$. Since each index (10.28) is a proper division of the next, the only possibility is that $[R : AP] = p[R : A]$. Applying this to the case $A = R$ shows that $[R : P] = p = N(P)$. So we find $[R : AP] = [R : A][R : P]$ and $[R : P] = N(P)$ again. This completes the proof. □

11. REAL QUADRATIC FIELDS

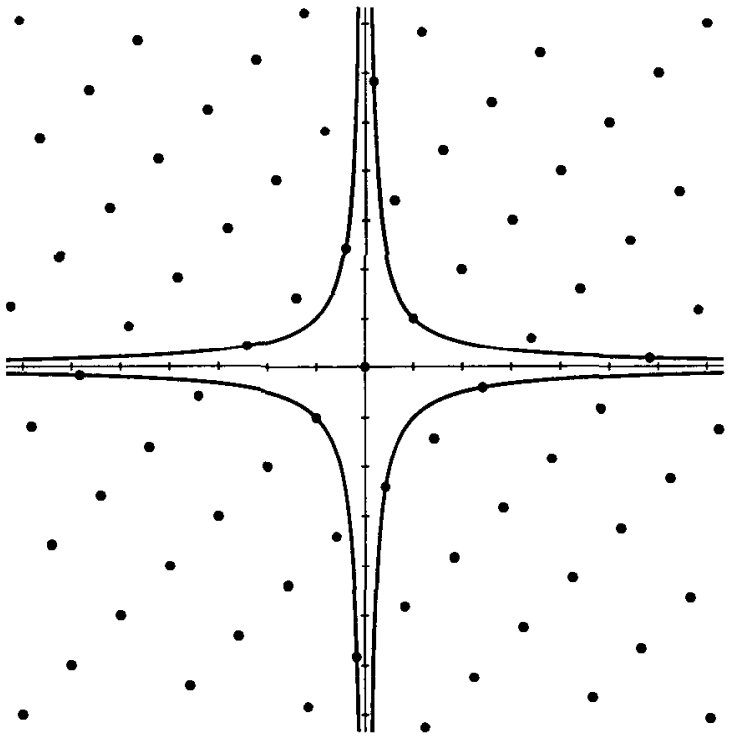
In this section we will take a brief look at real quadratic number fields $\mathbb{Q}[\delta]$, where $\delta^2 = d > 0$. We will use the field $\mathbb{Q}[\sqrt{2}]$ as an example. The ring of integers in this field is

$$R = \mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Z}\}. \tag{11.1}$$

Since $Q[\sqrt{d}]$ is a subfield of the real numbers, the ring of integers is not embedded as a lattice in the complex plane, but we can represent R as a lattice by using the coefficients (a, b) as coordinates. A slightly more convenient representation of R as a lattice is obtained by associating to the algebraic integer $a + b\sqrt{d}$ the point (u, v), where

$$u = a + b\sqrt{d}, \quad v = a - b\sqrt{d}. \tag{11.2}$$

The resulting lattice is depicted below for the case $d = 2$:



(11.3) **Figure.** The lattice $\mathbb{Z}[\sqrt{2}]$.

Since the (u, v)-coordinates are related to the (a, b)-coordinates by the linear transformation (11.2), there is no essential difference between the two ways of depicting R, though since the transformation is not orthogonal, the shape of the lattice is different in the two representations.

Recall that the field $\mathbb{Q}[\sqrt{d}]$ is isomorphic to the abstractly constructed field

$$F \approx \mathbb{Q}[x]/(x^2 - d). \tag{11.4}$$

Let us replace $\mathbb{Q}[\sqrt{d}]$ by F and denote the residue of x in F by δ. So this element δ is an abstract square root of d rather than the positive real square root. Then the coordinates u, v represent the two ways that the abstractly given field F can be embedded into the real numbers; namely u sends $\delta \rightsquigarrow \sqrt{d}$ and v sends $\delta \rightsquigarrow -\sqrt{d}$.

For $\alpha = a + b\delta \in \mathbb{Q}[\delta]$, let us denote by α' the "conjugate" element $a - b\delta$. The *norm* of α is defined to be

$$N(\alpha) = \alpha\alpha' = a^2 - db^2, \tag{11.5}$$

in analogy with the imaginary quadratic case (7.1). If α is an algebraic integer, then $N(\alpha)$ is an integer, not necessarily positive, and

$$N(\alpha\beta) = \alpha\beta\alpha'\beta' = N(\alpha)N(\beta). \tag{11.6}$$

With this definition of norm, the proof of unique factorization of ideals into prime ideals in imaginary quadratic fields carries over.

There are two notable differences between real and imaginary quadratic fields. The first is that, for real quadratic fields, ideals in the same class are not similar geometric figures when embedded as lattices in the (u, v)-plane by (11.2). In particular, principal ideals need not be similar to the lattice R. The reason is simple: Multiplication by an element $\alpha = a + b\delta$ stretches the u-coordinate by the factor $a + b\sqrt{d}$, and it stretches the v-coordinate by the different factor $a - b\sqrt{d}$. This fact complicates the geometry slightly, and it is the reason that we developed the imaginary quadratic case first. It does not change the theory in an essential way: The class number is still finite.

The second difference is more important. It is that there are infinitely many units in the rings of integers in a real quadratic field. Since the norm $N(\alpha)$ of an algebraic integer is an ordinary integer, a unit must have norm ± 1 as before [see (7.3)], and if $N(\alpha) = \alpha\alpha' = \pm 1$, then $\pm\alpha'$ is the inverse of α, so α is a unit. For example,

$$\alpha = 1 + \sqrt{2}, \quad \alpha^2 = 3 + 2\sqrt{2} \tag{11.7}$$

are units in the ring $R = \mathbb{Z}[\sqrt{2}]$. Their norms are -1 and 1 respectively. The element α has infinite order in the group of units of R.

The condition $N(\alpha) = a^2 - 2b^2 = \pm 1$ for units translates in (u, v)-coordinates to

$$uv = \pm 1. \tag{11.8}$$

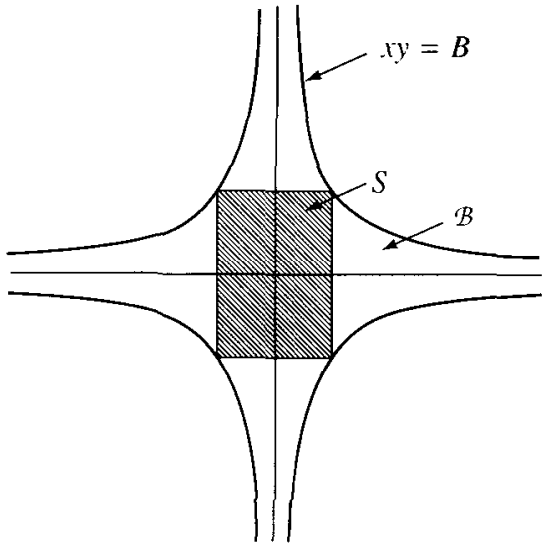
The units are the points of the lattice which lie on one of the two hyperbolas $uv = 1$ and $uv = -1$. These hyperbolas are depicted in Figure (11.3). It is a remarkable fact that real quadratic fields always have infinitely many units or, what amounts to the same thing, that the lattice of integers always contains infinitely many points on the hyperbola $uv = 1$. This fact is not obvious, either algebraically or geometrically.

(11.9) **Theorem.** Let R be the ring of integers in a real quadratic number field. The group of units in R is infinite.

(11.10) **Lemma.** Let Δ denote the area of the parallelogram spanned by a lattice basis of R, in its embedding into the (u, v)-plane. There are infinitely many elements β of R whose norm $N(\beta)$ is bounded, in fact, such that $|N(\beta)| \leq B$, where B is any real number $> \Delta$.

Proof. In the embedding into the (u, v)-plane, the elements of norm r are the lattice points on the hyperbola $xy = r$, and the elements whose norm is bounded in

absolute value by a positive number B are those lying in the region $\mathcal{B}$ bounded by the four branches of the hyperbolas $xy = B, xy = -B$.



(11.11) **Figure.**

Choose an arbitrary positive real number u_0. Then the rectangle S whose vertices are $(\pm u_0, \pm B/u_0)$ lies entirely in the region $\mathcal{B}$, and the area of this rectangle is $4B$. So if $B > \Delta$, then Minkowski's Lemma guarantees the existence of a nonzero lattice point α in S. The norm of this point is bounded by B. This is true for all u_0, and if u_0 is very large, the rectangle S is very narrow. On the other hand, there are no lattice points on the u_0-axis, because there are no nonzero elements in R of norm zero. So no particular lattice point is contained in all the rectangles S. It follows that there are infinitely many lattice points in $\mathcal{B}$. □

Since there are only finitely many integers r in the interval $-B \leq r \leq B$, Lemma (11.10) implies the following corollary:

(11.12) **Corollary.** For some integer r, there are infinitely many elements of R of norm r. □

Let r be an integer. We will call two elements $\beta_i = m_i + n_i\delta$ of R *congruent modulo* r if r divides $\beta_1 - \beta_2$ in R. If $d \equiv 2$ or 3 (modulo 4), this just means that $m_1 \equiv m_2$ and $n_1 \equiv n_2$ (modulo r).

(11.13) **Lemma.** Let β_1, β_2 be elements of R with the same norm r, and which are congruent modulo r. Then β_1/β_2 is a unit of R.

Proof. It suffices to show that β_1/β_2 is in R, because the same argument will show that $\beta_2/\beta_1 \in R$, hence that β_1/β_2 is a unit. Let $\beta_i' = m_i - n_i\delta$ be the conjugate of β_i. Then $\beta_1/\beta_2 = \beta_1\beta_2'/\beta_2\beta_2' = \beta_1\beta_2'/r$. But $\beta_2' \equiv \beta_1'$ (modulo r), so $\beta_1\beta_2' \equiv \beta_1\beta_1' = r$ (modulo r). Therefore r divides $\beta_1\beta_2'$, which shows that $\beta_1/\beta_2 \in R$, as required. □

Proof of Theorem (11.9). We choose r so that there are infinitely many elements $\beta = m + n\delta$ of norm r. We partition the set of these elements according to the congruence classes modulo r. Since there are finitely many congruence classes, some class contains infinitely many elements. The ratio of any two of these elements is a unit. □

12. SOME DIOPHANTINE EQUATIONS

Diophantine equations are polynomial equations with integer coefficients, which are to be solved in the integers. The most famous is the *Fermat Equation*

$$x^n + y^n = z^n. \tag{12.1}$$

Fermat's "Last Theorem" asserts that if $n \geq 3$ this equation has no integer solutions x, y, z, except for the trivial solutions in which one of the variables is zero. Fermat wrote this theorem in the margin of a book, asserting that the margin did not contain enough room for his proof. No proof is known today, though the theorem has been proved for all $n < 10^5$. Also, a theorem proved by Faltings in 1983, which applies to this equation as well as to many others, shows that there are only *finitely many* integer solutions for any given value of n.

This section contains a few examples of Diophantine equations which can be solved using the arithmetic of imaginary quadratic numbers. They are included only as samples. An interested reader should look in a book on number theory for a more organized discussion.

We have two methods at our disposal, namely arithmetic of quadratic number fields and congruences, and we will use both.

(12.2) **Example.** Determination of the integers n such that the equation

$$x^2 + y^2 = n$$

has an integer solution.

Here the problem is to determine the integers n which are sums of two squares or, equivalently, such that there is a point with integer coordinates on the circle $x^2 + y^2 = n$. Theorem (5.1) tells us that when p is a prime, the equation $x^2 + y^2 = p$ has an integer solution if and only if either $p = 2$ or $p \equiv 1$ (modulo 4). It is not difficult to extend this result to arbitrary integers. To do so, we interpret a sum of squares $a^2 + b^2$ as the norm $\alpha\overline{\alpha}$ of the Gauss integer $\alpha = a + bi$. Then the problem is to decide which integers n are the norms of Gauss integers. Now if a Gauss integer α is factored into Gauss primes, say $\alpha = \pi_1 \cdots \pi_k$, then its norm factors too: $N(\alpha) = N(\pi_1) \cdots N(\pi_k)$. So if n is the norm of a Gauss integer, then it is a product of norms of Gauss primes, and conversely. The norms of Gauss primes are the primes $p \equiv 1$ (modulo 4), the squares of primes $p \equiv 3$ (modulo 4), and the prime 2. Thus we have the following theorem:

(12.3) **Theorem.** The equation $x^2 + y^2 = n$ has an integer solution if and only if every prime p which is congruent 3 modulo 4 has an even exponent in the factorization of n. □

(12.4) **Example.** Determination of the integer solutions of the equation

$$y^2 + 13 = x^3.$$

We factor the left side of the equation, obtaining

$$(y + \delta)(y - \delta) = x^3,$$

where $\delta = \sqrt{-13}$. The ring of integers $R = \mathbb{Z}[\delta]$ is not a unique factorization domain, so we will analyze this equation using ideal factorization.

(12.5) **Lemma.** Let a, b be integers, and let R be any ring containing $\mathbb{Z}$ as a subring. If a and b are contained in a common proper ideal A of R, then they have a common prime factor in $\mathbb{Z}$.

Proof. We prove the contrapositive. If a, b have no common prime factor in $\mathbb{Z}$, then we can write $1 = ra + sb$, $r, s \in \mathbb{Z}$. This equation shows that if a, b are in an ideal A of R, then $1 \in A$ too. Hence A is not a proper ideal. □

(12.6) **Lemma.** Let x, y be an integer solution of the equation (12.4). The two elements $y + \delta$ and $y - \delta$ have no common prime ideal factor in R.

Proof. Let P be a prime ideal of R which contains $y + \delta$ and $y - \delta$. Then $2y \in P$ and $2\delta \in P$. Since P is a prime ideal, either $2 \in P$, or else $y \in P$ and $\delta \in P$.

In the first case, 2 and $y^2 + 13$ are not relatively prime integers by Lemma (12.5), and since 2 is prime, it divides $y^2 + 13$ in $\mathbb{Z}$. This implies that 2 divides x and that 8 divides $y^2 + 13 = x^3$. So y must be odd. Then $y^2 \equiv 1$ (modulo 4); hence $y^2 + 13 \equiv 2$ (modulo 4). This contradicts $x^3 \equiv 0$ (modulo 8).

Suppose that $y, \delta \in P$. Then $13 \in P$, and hence 13 and y are not relatively prime in $\mathbb{Z}$, that is, 13 divides y. Therefore 13 divides x, and reading the equation $y^2 + 13 = x^3$ modulo 13^2, we obtain $13 \equiv 0$ (modulo 13^2), which is a contradiction. So we have shown that $y + \delta$ and $y - \delta$ are relatively prime in R. □

We now read the equation $(y + \delta)(y - \delta) = (x)^3$ as an equality of principal ideals of R, and we factor the right side into primes, say

$$(y + \delta)(y - \delta) = (P_1 \cdots P_s)^3.$$

On the right we have a cube, and the two ideals on the left have no common prime factor. It follows that each of these ideals is a cube too, say $(y + \delta) = A^3$ and $(y - \delta) = \bar{A}^3$ for some ideal A. Looking at our table of ideal classes, we find that the ideal class group of R is cyclic of order 2. So the ideal classes of A and A^3 are equal. Since A^3 is a principal ideal, so is A, say $A = (u + v\delta)$, for some integers

u, v. We have been lucky. Since the units in R are ± 1, $(u + v\delta)^3 = \pm(y + \delta)$. Changing sign if necessary, we may assume that $(u + v\delta)^3 = y + \delta$.

We now complete the analysis by studying the equation $y + \delta = (u + v\delta)^3$. We expand the right side, obtaining

$$y + \delta = (u^3 - 39uv^2) + (3u^2v - 13v^3)\delta.$$

So $y = u^3 - 39uv^2$ and $1 = (3u^2 - 13v^2)v$. The second equation implies that $v = \pm 1$ and that $3u^2 - 13 = \pm 1$. The only possibilities are $u = \pm 2$ and $v = -1$. Then $y = \pm 70$ and $x = (u + v\delta)(u - v\delta) = 17$. These values do give solutions, so the integer solutions of the equation $y^2 + 13 = x^3$ are $x = 17$ and $y = \pm 70$. □

(12.7) **Example.** Determination of the prime integers p such that

$$x^2 + 5y^2 = p$$

has an integer solution.

Let $\delta = \sqrt{-5}$, and let $R = \mathbb{Z}[\delta]$. We know (9.3a) that the principal ideal (p) splits in R if and only if the congruence $x^2 \equiv -5$ (modulo p) has an integer solution. If $(p) = P\bar{P}$ and if P is a principal ideal, say $P = (a + b\delta)$, then $(p) = (a + b\delta)(a - b\delta) = (a^2 + 5b^2)$. Since the only units in R are ± 1, $a^2 + 5b^2 = \pm p$, and since $a^2 + 5b^2$ is positive, $a^2 + 5b^2 = p$.

Unfortunately, R is not a principal ideal domain. So it is quite likely that $(p) = P\bar{P}$ but that P is not a principal ideal. To analyze the situation further, we use the fact that there are exactly two ideal classes in R. The principal ideals form one class, and the other class is represented by any nonprincipal ideal. The ideal $A = (2, 1 + \delta)$ is one nonprincipal ideal, and we recall that for this ideal $A^2 = A\bar{A} = (2)$. Now since the ideal class group is cyclic of order 2, the product of any two ideals in the same class is principal. Suppose that $(p) = P\bar{P}$ and that P is not a principal ideal. Then AP is principal, say $AP = (a + b\delta)$. Then $(a + b\delta)(a - b\delta) = AP\overline{AP} = (2p)$. We find that $a^2 + 5b^2 = 2p$.

(12.8) **Lemma.** Let p be an odd prime. The congruence $x^2 \equiv -5$ (modulo p) has a solution if and only if one of the two equations $x^2 + 5y^2 = p$ or $x^2 + 5y^2 = 2p$ has an integer solution.

Proof. If the congruence has a solution, then $(p) = P\bar{P}$, and the two cases are decided as above, according to whether or not P is principal. Conversely, if $a^2 + 5b^2 = p$, then (p) splits in R, and we can apply (9.3a). If $a^2 + 5b^2 = 2p$, then $(a + b\delta)(a - b\delta) = (2p) = A\bar{A}(p)$. It follows from unique factorization of ideals that (p) splits too, so (9.3a) can be applied again. □

This lemma does not solve our original problem, but we have made progress. In most such situations we could not complete our analysis. But here we are lucky again, or rather this example was chosen because it admits a complete solution: The two cases can be distinguished by congruences. If $a^2 + 5b^2 = p$, then one of the

two integers a, b is odd and the other is even. We compute the congruence modulo 4, finding that $a^2 + 5b^2 \equiv 1$ (modulo 4). Hence $p \equiv 1$ (modulo 4) in this case. If $a^2 + 5b^2 = 2p$, we compute the congruences modulo 8. Since $p \equiv 1$ or 3 (modulo 4), we know that $2p \equiv 2$ or 6 (modulo 8). Any square is congruent 0, 1, or 4 (modulo 8). Hence $5b^2 \equiv 0, 5$, or 4 (modulo 8), which shows that $a^2 + 5b^2$ can not be congruent to 2 (modulo 8). Thus $p \equiv 3$ (modulo 4) in this case. We have therefore proved the following lemma:

(12.9) **Lemma.** Let p be an odd prime. Assume that the congruence $x^2 \equiv -5$ (modulo p) has a solution. Then $x^2 + 5y^2 = p$ has an integer solution if $p \equiv 1$ (modulo 4), and $x^2 + 5y^2 = 2p$ has an integer solution if $p \equiv 3$ (modulo 4).

There remains finally the problem of characterizing the odd primes p such that the congruence $x^2 \equiv -5$ has a solution modulo p. This is done by means of the amazing *Quadratic Reciprocity Law*, which asserts that $x^2 \equiv 5$ (modulo p) has a solution if and only if $x^2 \equiv p$ (modulo 5) has one! And the second congruence has a solution if and only if $p \equiv \pm 1$ (modulo 5). Combining this with the previous lemma and with the fact that -1 is a square modulo 5, we find:

(12.10) **Theorem.** Let p be an odd prime. The equation $x^2 + 5y^2 = p$ has an integer solution if and only if $p \equiv 1$ (modulo 4) and $p \equiv \pm 1$ (modulo 5). □

Nullum vero dubium nobis esse videtur,
quin multa eaque egregia in hoc genere adhuc lateant
in quibus alii vires suas exercere possint.

Karl Friedrich Gauss

EXERCISES

1. Factorization of Integers and Polynomials

1. Let a, b be positive integers whose sum is a prime p. Prove that their greatest common divisor is 1.
2. Define the greatest common divisor of a set of n integers, and prove its existence.
3. Prove that if d is the greatest common divisor of $a_1, \ldots, a_n$, then the greatest common divisor of $a_1/d, \ldots, a_n/d$ is 1.
4. **(a)** Prove that if n is a positive integer which is not a square of an integer, then $\sqrt{n}$ is not a rational number.
   (b) Prove the analogous statement for nth roots.
5. **(a)** Let a, b be integers with $a \neq 0$, and write $b = aq + r$, where $0 \leq r < |a|$. Prove that the two greatest common divisors (a, b) and (a, r) are equal.
   (b) Describe an algorithm, based on (a), for computing the greatest common divisor.

(c) Use your algorithm to compute the greatest common divisors of the following: (a) 1456, 235, (b) 123456789, 135792468.

6. Compute the greatest common divisor of the following polynomials: $x^3 - 6x^2 + x + 4$, $x^5 - 6x + 1$.
7. Prove that if two polynomials f, g with coefficients in a field F factor into linear factors in F, then their greatest common divisor is the product of their common linear factors.
8. Factor the following polynomials into irreducible factors in $\mathbb{F}_p[x]$.
(a) $x^3 + x + 1, p = 2$ (b) $x^2 - 3x - 3, p = 5$ (c) $x^2 + 1, p = 7$
9. Euclid proved that there are infinitely many prime integers in the following way: If $p_1, \ldots, p_k$ are primes, then any prime factor p of $n = (p_1 \cdots p_n) + 1$ must be different from all of the p_i.
(a) Adapt this argument to show that for any field F there are infinitely many monic irreducible polynomials in $F[x]$.
(b) Explain why the argument fails for the formal power series ring $F[[x]]$.
10. *Partial fractions for integers:*
(a) Write the fraction $r = 7/24$ in the form $r = a/8 + b/3$.
(b) Prove that if $n = uv$, where u and v are relatively prime, then every fraction $r = m/n$ can be written in the form $r = a/u + b/v$.
(c) Let $n = n_1 n_2 \cdots n_k$ be the factorization of an integer n into powers of distinct primes: $n_i = p_i^{e_i}$. Prove that every fraction $r = m/n$ can be written in the form $r = m_1/n_1 + \cdots + m_k/n_k$.
11. *Chinese Remainder Theorem:*
(a) Let n, m be relatively prime integers, and let a, b be arbitrary integers. Prove that there is an integer x which solves the simultaneous congruence $x \equiv a$ (modulo m) and $x \equiv b$ (modulo n).
(b) Determine all solutions of these two congruences.
12. Solve the following simultaneous congruences.
(a) $x \equiv 3$ (modulo 15), $x \equiv 5$ (modulo 8), $x \equiv 2$ (modulo 7).
(b) $x \equiv 13$ (modulo 43), $x \equiv 7$ (modulo 71).
13. *Partial fractions for polynomials:*
(a) Prove that every rational function in $\mathbb{C}(x)$ can be written as sum of a polynomial and a linear combination of functions of the form $1/(x - a)^i$.
(b) Find a basis for $\mathbb{C}(x)$ as vector space over $\mathbb{C}$.

*14. Let F be a subfield of $\mathbb{C}$, and let $f \in F[x]$ be an irreducible polynomial. Prove that f has no multiple root in $\mathbb{C}$.

15. Prove that the greatest common divisor of two polynomials f and g in $\mathbb{Q}[x]$ is also their greatest common divisor in $\mathbb{C}[x]$.
16. Let a and b be relatively prime integers. Prove that there are integers m, n such that $a^m + b^n \equiv 1$ (modulo ab).

2. Unique Factorization Domains, Principal Ideal Domains, and Euclidean Domains

1. Prove or disprove the following.
(a) The polynomial ring $\mathbb{R}[x, y]$ in two variables is a Euclidean domain.
(b) The ring $\mathbb{Z}[x]$ is a principal ideal domain.

2. Prove that the following rings are Euclidean domains.
   (a) $\mathbb{Z}[\zeta]$, $\zeta = e^{2\pi i/3}$ (b) $\mathbb{Z}[\sqrt{-2}]$.
3. Give an example showing that division with remainder need not be unique in a Euclidean domain.
4. Let m, n be two integers. Prove that their greatest common divisor in $\mathbb{Z}$ is the same as their greatest common divisor in $\mathbb{Z}[i]$.
5. Prove that every prime element of an integral domain is irreducible.
6. Prove Proposition (2.8), that a domain R which has existence of factorizations is a unique factorization domain if and only if every irreducible element is prime.
7. Prove that in a principal ideal domain R, every pair a, b of elements, not both zero, has a greatest common divisor d, with these properties:
   (i) $d = ar + bs$, for some $r, s \in R$;
   (ii) d divides a and b;
   (iii) if $e \in R$ divides a and b, it also divides d.
   Moreover, d is determined up to unit factor.
8. Find the greatest common divisor of $(11 + 7i, 18 - i)$ in $\mathbb{Z}[i]$.
9. (a) Prove that $2, 3, 1 \pm \sqrt{-5}$ are irreducible elements of the ring $R = \mathbb{Z}[\sqrt{-5}]$ and that the units of this ring are ± 1.
   (b) Prove that existence of factorizations is true for this ring.
10. Prove that the ring $\mathbb{R}[[t]]$ of formal real power series is a unique factorization domain.
11. (a) Prove that if R is an integral domain, then two elements a, b are associates if and only if they differ by a unit factor.
    *(b) Give an example showing that (a) is false when R is not an integral domain.
12. Let R be a principal ideal domain.
    (a) Prove that there is a *least common multiple* $[a, b] = m$ of two elements which are not both zero such that a and b divide m, and that if a, b divide an element $r \in R$, then m divides r. Prove that m is unique up to unit factor.
    (b) Denote the greatest common divisor of a and b by (a, b). Prove that $(a, b)[a, b]$ is an associate of ab.
13. If a, b are integers and if a divides b in the ring of Gauss integers, then a divides b in $\mathbb{Z}$.
14. (a) Prove that the ring R (2.4) obtained by adjoining 2^k-th roots x_k of x to a polynomial ring is the union of the polynomial rings $F[x_k]$.
    (b) Prove that there is no factorization of x_1 into irreducible factors in R.
15. By a *refinement* of a factorization $a = b_1 \cdots b_k$ we mean the expression for a obtained by factoring the terms b_i. Let R be the ring (2.4). Prove that any two factorizations of the same element $a \in R$ have refinements, all of whose factors are associates.
16. Let R be the ring $F[u, v, y, x_1, x_2, x_3, \ldots]/(x_1 y = uv, x_2^2 = x_1, x_3^2 = x_2, \ldots)$. Show that u, v are irreducible elements in R but that the process of factoring uv need not terminate.
17. Prove Proposition (2.9) and Corollary (2.10).
18. Prove Proposition (2.11).
19. Prove that the factorizations (2.22) are prime in $\mathbb{Z}[i]$.
20. The discussion of unique factorization involves only the multiplication law on the ring R, so it ought to be possible to extend the definitions. Let S be a commutative semigroup, meaning a set with a commutative and associative law of composition and with an iden-

tity. Suppose the Cancellation Law holds in S: If $ab = ac$ then $b = c$. Make the appropriate definitions so as to extend Proposition (2.8) to this situation.

*21. Given elements $v_1, \ldots v_n$ in $\mathbb{Z}^2$, we can define a semigroup S as the set of all linear combinations of $(v_1, \ldots, v_n)$ with nonnegative integer coefficients, the law of composition being *addition*. Determine which of these semigroups has unique factorization.

3. Gauss's Lemma

1. Let a, b be elements of a field F, with $a \neq 0$. Prove that a polynomial $f(x) \in F[x]$ is irreducible if and only if $f(ax + b)$ is irreducible.
2. Let $F = \mathbb{C}(x)$, and let $f, g \in \mathbb{C}[x, y]$. Prove that if f and g have a common factor in $F[y]$, then they also have a common factor in $\mathbb{C}[x, y]$.
3. Let f be an irreducible polynomial in $\mathbb{C}[x, y]$, and let g be another polynomial. Prove that if the variety of zeros of g in $\mathbb{C}^2$ contains the variety of zeros of f, then f divides g.
4. Prove that two integer polynomials are relatively prime in $\mathbb{Q}[x]$ if and only if the ideal they generate in $\mathbb{Z}[x]$ contains an integer.
5. Prove Gauss's Lemma without reduction modulo p, in the following way: Let a_i be the coefficient of lowest degree i of f which is not divisible by p. So p divides a_ν if $\nu < i$, but p does not divide a_i. Similarly, let b_j be the coefficient of lowest degree of g which is not divisible by p. Prove that the coefficient of h of degree $i + j$ is not divisible by p.
6. State and prove Gauss's Lemma for Euclidean domains.
7. Prove that an integer polynomial is primitive if and only if it is not contained in any of the kernels of the maps (3.2).
8. Prove that $\det\begin{bmatrix} x & y \\ z & w \end{bmatrix}$ is irreducible in the polynomial ring $\mathbb{C}[x, y, z, w]$.
9. Prove that the kernel of the homomorphism $\mathbb{Z}[x] \longrightarrow \mathbb{R}$ sending $x \rightsquigarrow 1 + \sqrt{2}$ is a principal ideal, and find a generator for this ideal.
10. (a) Consider the map ψ: $\mathbb{C}[x, y] \longrightarrow \mathbb{C}[t]$ defined by $f(x, y) \rightsquigarrow f(t^2, t^3)$. Prove that its kernel is a principal ideal, and that its image is the set of polynomials $p(t)$ such that $p'(0) = 0$.
    (b) Consider the map φ: $\mathbb{C}[x, y] \longrightarrow \mathbb{C}[t]$ defined by $f(x, y) \rightsquigarrow (t^2 - t, t^3 - t^2)$. Prove that $\ker \varphi$ is a principal ideal, and that its image is the set of polynomials $p(t)$ such that $p(0) = p(1)$. Give an intuitive explanation in terms of the geometry of the variety $\{f = 0\}$ in $\mathbb{C}^2$.

4. Explicit Factorization of Polynomials

1. Prove that the following polynomials are irreducible in $\mathbb{Q}[x]$.
   (a) $x^2 + 27x + 213$ (b) $x^3 + 6x + 12$ (c) $8x^3 - 6x + 1$ (d) $x^3 + 6x^2 + 7$
   (e) $x^5 - 3x^4 + 3$
2. Factor $x^5 + 5x + 5$ into irreducible factors in $\mathbb{Q}[x]$ and in $\mathbb{F}_2[x]$.
3. Factor $x^3 + x + 1$ in $\mathbb{F}_p[x]$, when $p = 2, 3, 5$.

4. Factor $x^4 + x^2 + 1$ into irreducible factors in $\mathbb{Q}[x]$.
5. Suppose that a polynomial of the form $x^4 + bx^2 + c$ is a product of two quadratic factors in $\mathbb{Q}[x]$. What can you say about the coefficients of these factors?
6. Prove that the following polynomials are irreducible.
   (a) $x^2 + x + 1$ in the field $\mathbb{F}_2$ **(b)** $x^2 + 1$ in $\mathbb{F}_7$ **(c)** $x^3 - 9$ in $\mathbb{F}_{31}$
7. Factor the following polynomials into irreducible factors in $\mathbb{Q}[x]$.
   (a) $x^3 - 3x - 2$ **(b)** $x^3 - 3x + 2$ **(c)** $x^9 - 6x^6 + 9x^3 - 3$
8. Let p be a prime integer. Prove that the polynomial $x^n - p$ is irreducible in $\mathbb{Q}[x]$.
9. Using reduction modulo 2 as an aid, factor the following polynomials in $\mathbb{Q}[x]$.
   (a) $x^2 + 2345x + 125$ **(b)** $x^3 + 5x^2 + 10x + 5$ **(c)** $x^3 + 2x^2 + 3x + 1$
   (d) $x^4 + 2x^3 + 2x^2 + 2x + 2$ **(e)** $x^4 + 2x^3 + 3x^2 + 2x + 1$
   (f) $x^4 + 2x^3 + x^2 + 2x + 1$ **(g)** $x^5 + x^4 - 4x^3 + 2x^2 + 4x + 1$
10. Let p be a prime integer, and let $f \in \mathbb{Z}[x]$ be a polynomial of degree $2n + 1$, say $f(x) = a_{2n+1}x^{2n+1} + \cdots + a_1x + a_0$. Suppose that $a_{2n+1} \not\equiv 0$ (modulo p), $a_0, a_1, \ldots, a_n \equiv 0$ (modulo p^2), $a_{n+1}, \ldots, a_{2n} \equiv 0$ (modulo p), $a_0 \not\equiv 0$ (modulo p^3). Prove that f is irreducible in $\mathbb{Q}[x]$.
11. Let p be a prime, and let $A \neq I$ be an $n \times n$ integer matrix such that $A^p = I$ but $A \neq I$. Prove that $n \geq p - 1$.
12. Determine the monic irreducible polynomials of degree 3 over $\mathbb{F}_3$.
13. Determine the monic irreducible polynomials of degree 2 over $\mathbb{F}_5$.
14. *Lagrange interpolation formula:*
    (a) Let $x_0, \ldots, x_d$ be distinct complex numbers. Determine a polynomial $p(x)$ of degree n which is zero at $x_1, \ldots, x_n$ and such that $p(x_0) = 1$.
    (b) Let $x_0, \ldots, x_d$; $y_0, \ldots, y_d$ be complex numbers, and suppose that the x_i are all different. There is a unique polynomial $g(x) \in \mathbb{C}[x]$ of degree $\leq d$, such that $g(x_i) = y_i$ for each $i = 0, \ldots, d$. Prove this by determining the polynomial g explicitly in terms of x_i, y_i.

*15. Use the Lagrange interpolation formula to give a method of finding all integer polynomial factors of an integer polynomial in a finite number of steps.

16. Let $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ be a monic polynomial with integer coefficients, and let $r \in \mathbb{Q}$ be a rational root of $f(x)$. Prove that r is an integer.
17. Prove that the polynomial $x^2 + y^2 - 1$ is irreducible by the method of undetermined coefficients, that is, by studying the equation $(ax + by + c)(a'x + b'y + c') = x^2 + y^2 - 1$, where a, b, c, a', b', c' are unknown.

5. Primes in the Ring of Gauss Integers

1. Prove that every Gauss prime divides exactly one integer prime.
2. Factor 30 into primes in $\mathbb{Z}[i]$.
3. Factor the following into Gauss primes.
   (a) $1 - 3i$ **(b)** 10 **(c)** $6 + 9i$
4. Make a neat drawing showing the primes in the ring of Gauss integers in a reasonable size range.
5. Let π be a Gauss prime. Prove that π and $\overline{\pi}$ are associate if and only if either π is associate to an integer prime or $\pi\overline{\pi} = 2$.

6. Let R be the ring $\mathbb{Z}[\sqrt{3}]$. Prove that a prime integer p is a prime element of R if and only if the polynomial $x^2 - 3$ is irreducible in $\mathbb{F}_p[x]$.
7. Describe the residue ring $\mathbb{Z}[i]/(p)$ in each case.
   (a) $p = 2$ **(b)** $p \equiv 1$ (modulo 4) **(c)** $p \equiv 3$ (modulo 4)

*8. Let $R = \mathbb{Z}[\zeta]$, where $\zeta = \frac{1}{2}(-1 + \sqrt{-3})$ is a complex cube root of 1. Let p be an integer prime $\neq 3$. Adapt the proof of Theorem (5.1) to prove the following.
   (a) The polynomial $x^2 + x + 1$ has a root in $\mathbb{F}_p$ if and only if $p \equiv 1$ (modulo 3).
   (b) (p) is a prime ideal of R if and only if $p \equiv -1$ (modulo 3).
   (c) p factors in R if and only if it can be written in the form $p = a^2 + ab + b^2$, for some integers a, b.
   (d) Make a drawing showing the primes of absolute value ≤ 10 in R.

6. Algebraic Integers

1. Is $\frac{1}{2}(1 + \sqrt{3})$ an algebraic integer?
2. Let α be an algebraic integer whose monic irreducible polynomial over $\mathbb{Z}$ is $x^n + a_{n-1}n^{n-1} + \cdots + a_1x + a_0$, and let $R = \mathbb{Z}[\alpha]$. Prove that α is a unit in R if and only if $a_0 = \pm 1$.
3. Let d, d' be distinct square-free integers. Prove that $\mathbb{Q}(\sqrt{d})$ and $\mathbb{Q}(\sqrt{d'})$ are different subfields of $\mathbb{C}$.
4. Prove that existence of factorizations is true in the ring of integers in an imaginary quadratic number field.
5. Let α be the real cube root of 10, and let $\beta = a + b\alpha + c\alpha^2$, with $a, b, c, \in \mathbb{Q}$. Then β is the root of a monic cubic polynomial $f(x) \in \mathbb{Q}[x]$. The irreducible polynomial for α over $\mathbb{Q}$ is $x^3 - 10$, and its three roots are α, $\alpha' = \zeta\alpha$, and $\alpha'' = \zeta^2\alpha$, where $\zeta = e^{2\pi i/3}$. The three roots of f are β, $\beta' = a + b\zeta\alpha + c\zeta^2\alpha^2$, and $\beta'' = a + b\zeta^2\alpha + c\zeta\alpha^2$, so $f(x) = (x - \beta)(x - \beta')(x - \beta'')$.
   (a) Determine f by expanding this product. The terms involving α and α^2 have to cancel out, so they need not be computed.
   (b) Determine which elements β are algebraic integers.
6. Prove Proposition (6.17).
7. Prove that the ring of integers in an imaginary quadratic field is a maximal subring of $\mathbb{C}$ with the property of being a lattice in the complex plane.
8. **(a)** Let $S = \mathbb{Z}[\alpha]$, where α is a complex root of a monic polynomial of degree 2. Prove that S is a lattice in the complex plane.
   (b) Prove the converse: A subring S of $\mathbb{C}$ which is a lattice has the form given in (a).
9. Let R be the ring of integers in the field $\mathbb{Q}[\sqrt{d}]$.
   (a) Determine the elements $\alpha \in R$ such that $R = \mathbb{Z}[\alpha]$.
   (b) Prove that if $R = \mathbb{Z}[\alpha]$ and if α is a root of the polynomial $x^2 + bx + c$ over $\mathbb{Q}$, then the discriminant $b^2 - 4c$ is D (6.18).

7. Factorization in Imaginary Quadratic Fields

1. Prove Proposition (7.3) by arithmetic.
2. Prove that the elements $2, 3, 1 + \sqrt{-5}, 1 - \sqrt{-5}$ are irreducible elements of the ring $\mathbb{Z}[\sqrt{-5}]$.

3. Let $d = -5$. Determine whether or not the lattice of integer linear combinations of the given vectors is an ideal.
   (a) $(5, 1 + \delta)$ **(b)** $(7, 1 + \delta)$ **(c)** $(4 - 2\delta, 2 + 2\delta, 6 + 4\delta)$
4. Let A be an ideal of the ring of integers R in an imaginary quadratic field. Prove that there is a lattice basis for A one of whose elements is a positive integer.
5. Let $R = \mathbb{Z}[\sqrt{-5}]$. Prove that the lattice spanned by $(3, 1 + \sqrt{-5})$ is an ideal in R, determine its nonzero element of minimal absolute value, and verify that this ideal has the form (7.9), Case 2.
6. With the notation of (7.9), show that if α is an element of R such that $\frac{1}{2}(\alpha + \alpha\delta)$ is also in R, then $(\alpha, \frac{1}{2}(\alpha + \alpha\delta))$ is a lattice basis of an ideal.
7. For each ring R listed below, use the method of Proposition (7.9) to describe the ideals in R. Make a drawing showing the possible shapes of the lattices in each case.
   (a) $R = \mathbb{Z}[\sqrt{-3}]$ **(b)** $R = \mathbb{Z}[\frac{1}{2}(1 + \sqrt{-3})]$ **(c)** $R = \mathbb{Z}[\sqrt{-6}]$ **(d)** $R = \mathbb{Z}[\sqrt{-7}]$
   (e) $R = \mathbb{Z}[\frac{1}{2}(1 + \sqrt{-7})]$ **(f)** $R = \mathbb{Z}[\sqrt{-10}]$
8. Prove that R is not a unique factorization domain when $d \equiv 2$ (modulo 4) and $d < -2$.
9. Let $d \leq -3$. Prove that 2 is not a prime element in the ring $\mathbb{Z}[\sqrt{d}]$, but that 2 is irreducible in this ring.

8. Ideal Factorization

1. Let $R = \mathbb{Z}[\sqrt{-6}]$. Factor the ideal (6) into prime ideals explicitly.
2. Let $\delta = \sqrt{-3}$ and $R = \mathbb{Z}[\delta]$. (This is not the ring of integers in the imaginary quadratic number field $\mathbb{Q}[\delta]$.) Let A be the ideal $(2, 1 + \delta)$. Show that $A\bar{A}$ is not a principal ideal, hence that the Main Lemma is not true for this ring.
3. Let $R = \mathbb{Z}[\sqrt{-5}]$. Determine whether or not 11 is an irreducible element of R and whether or not (11) is a prime ideal in R.
4. Let $R = \mathbb{Z}[\sqrt{-6}]$. Find a lattice basis for the product ideal AB, where $A = (2, \delta)$ and $B = (3, \delta)$.
5. Prove that $A \supset A'$ implies that $AB \supset A'B$.
6. Factor the principal ideal (14) into prime ideals explicitly in $R = \mathbb{Z}[\delta]$, where $\delta = \sqrt{-5}$.
7. Let P be a prime ideal of an integral domain R, and assume that existence of factorizations is true in R. Prove that if $a \in P$ then some irreducible factor of a is in P.

9. The Relation Between Prime Ideals of R and Prime Integers

1. Find lattice bases for the prime divisors of 2 and 3 in the ring of integers in (a) $\mathbb{Q}[\sqrt{-14}]$ and (b) $\mathbb{Q}[\sqrt{-23}]$.
2. Let $d = -14$. For each of the following primes p, determine whether or not p splits or ramifies in R, and if so, determine a lattice basis for a prime ideal factor of (p): $2, 3, 5, 7, 11, 13$.
3. **(a)** Suppose that a prime integer p remains prime in R. Prove that $R/(p)$ is then a field with p^2 elements.
   (b) Prove that if p splits in R, then $R/(p)$ is isomorphic to the product ring $\mathbb{F}_p \times \mathbb{F}_p$.

4. Let p be a prime which splits in R, say $(p) = P\bar{P}$, and let $\alpha \in P$ be any element which is not divisible by p. Prove that P is generated as an ideal by (p, α).
5. Prove Proposition (9.3b).
6. If $d \equiv 2$ or 3 (modulo 4), then according to Proposition (9.3a) a prime integer p remains prime in the ring of integers of $\mathbb{Q}[\sqrt{d}]$ if the polynomial $x^2 - d$ is irreducible modulo p.
   (a) Prove the same thing when $d \equiv 1$ (modulo 4) and $p \neq 2$.
   (b) What happens to $p = 2$ in this case?
7. Assume that $d \equiv 2$ or 3 (modulo 4). Prove that a prime integer p ramifies in R if and only if $p = 2$ or p divides d.
8. State and prove an analogue of problem 7 when d is congruent 1 modulo 4.
9. Let p be an integer prime which ramifies in R, and say that $(p) = P^2$. Find an explicit lattice basis for P. In which cases is P a principal ideal?
10. A prime integer might be of the form $a^2 + b^2d$, with $a, b \in \mathbb{Z}$. Discuss carefully how this is related to the prime factorization of (p) in R.

*11. Prove Proposition (9.1).

10. Ideal Classes in Imaginary Quadratic Fields

1. Prove that the ideals A and A' are similar if and only if there is a nonzero ideal C such that AC and $A'C$ are principal ideals.
2. The estimate of Corollary (10.12) can be improved to $|\alpha|^2 \leq 2\Delta(L)/\sqrt{3}$, by studying lattice points in a circle rather than in an arbitrary centrally symmetric convex set. Work this out.
3. Let $R = \mathbb{Z}[\delta]$, where $\delta^2 = -6$.
   (a) Prove that the lattices $P = (2, \delta)$ and $Q = (3, \delta)$ are prime ideals of R.
   (b) Factor the principal ideal (6) into prime ideals explicitly in R.
   (c) Prove that the ideal classes of P and Q are equal.
   (d) The Minkowski bound for R is $[\mu] = 3$. Using this fact, determine the ideal class group of R.
4. In each case, determine the ideal class group and draw the possible shapes of the lattices.
   (a) $d = -10$ **(b)** $d = -13$ **(c)** $d = -14$ **(d)** $d = -15$ **(e)** $d = -17$
   (f) $d = -21$
5. Prove that the values of d listed in Theorem (7.7) have unique factorization.
6. Prove Lemma (10.13).
7. Derive Corollary (10.14) from Lemma (10.13).
8. Verify Table (10.24).

11. Real Quadratic Fields

1. Let $R = \mathbb{Z}[\delta]$, $\delta = \sqrt{2}$. Define a size function on R using the lattice embedding (11.2): $\sigma(a + b\delta) = a^2 - 2b^2$. Prove that this size function makes R into a Euclidean domain.
2. Let R be the ring of integers in a real quadratic number field, with $d \equiv 2$ or 3 (modulo 4). According to (6.14), R has the form $\mathbb{Z}[x]/(x^2 - d)$. We can also consider the ring $R' = \mathbb{R}[x]/(x^2 - d)$, which contains R as a subring.
   (a) Show that the elements of R' are in bijective correspondence with points of $\mathbb{R}^2$ in such a way that the elements of R correspond to lattice points.

(b) Determine the group of units of R'. Show that the subset U' of R' consisting of the points on the two hyperbolas $xy = \pm 1$ forms a subgroup of the group of units.
(c) Show that the group of units U of R is a discrete subgroup of U', and show that the subgroup U_0 of units which are in the first quadrant is an infinite cyclic group.
(d) What are the possible structures of the group of units U?

3. Let U_0 denote the group of units of R which are in the first quadrant in the embedding (11.2). Find a generator for U_0 when (a) $d = 3$, (b) $d = 5$.
4. Prove that if d is a square >1 then the equation $x^2 - y^2 d = 1$ has no solution except $x = \pm 1, y = 0$.
5. Draw a figure showing the hyperbolas and the units in a reasonable size range for $d = 3$.

12. Some Diophantine Equations

1. Determine the primes such that $x^2 + 5y^2 = 2p$ has a solution.
2. Express the assertion of Theorem (12.10) in terms of congruence modulo 20.
3. Prove that if $x^2 \equiv -5$ (modulo p) has a solution, then there is an integer point on one of the two ellipses $x^2 + 5y^2 = p$ or $2x^2 + 2xy + 3y^2 = p$.
4. Determine the conditions on the integers a, b, c such that the linear Diophantine equation $ax + by = c$ has an integer solution, and if it does have one, find all the solutions.
5. Determine the primes p such that the equation $x^2 + 2y^2 = p$ has an integer solution.
6. Determine the primes p such that the equation $x^2 + xy + y^2 = p$ has an integer solution.
7. Prove that if the congruence $x^2 \equiv -10$ (modulo p) has a solution, then the equation $x^2 + 10y^2 = p^2$ has an integer solution. Generalize.
8. Find all integer solutions of the equation $x^2 + 2 = y^3$.
9. Solve the following Diophantine equations.
(a) $y^2 + 10 = x^3$ (b) $y^2 + 1 = x^3$ (c) $y^2 + 2 = x^3$

Miscellaneous Problems

1. Prove that there are infinitely many primes congruent 1 modulo 4.
2. Prove that there are infinitely many primes congruent to -1 (modulo 6) by studying the factorization of the integer $p_1 p_2 \cdots p_r - 1$, where $p_1, \ldots, p_r$ are the first r primes.
3. Prove that there are infinitely many primes congruent to -1 (modulo 4).
4. (a) Determine the prime ideals of the polynomial ring $\mathbb{C}[x, y]$ in two variables.
(b) Show that unique factorization of ideals does not hold in the ring $\mathbb{C}[x, y]$.
5. Relate proper factorizations of elements in an integral domain to proper factorizations of principal ideals. Using this relation, state and prove unique factorization of ideals in a principal ideal domain.
6. Let R be a domain, and let I be an ideal which is a product of distinct maximal ideals in two ways, say $I = P_1 \cdots P_r = Q_1 \cdots Q_s$. Prove that the two factorizations are the same, except for the ordering of the terms.
7. Let R be a ring containing $\mathbb{Z}$ as a subring. Prove that if integers m, n are contained in a proper ideal of R, then they have a common integer factor > 1.

*8. (a) Let θ be an element of the group $\mathbb{R}^+/\mathbb{Z}^+$. Use the Pigeonhole Principle [Appendix (1.6)] to prove that for every integer n there is an integer $b \leq n$ such that $|b\theta| \leq 1/bn$.

(b) Show that for every real number r and every $\epsilon > 0$, there is a fraction m/n such that $|r - m/n| \leq \epsilon/n$.

(c) Extend this result to the complex numbers by showing that for every complex number α and every real number $\epsilon > 0$, there is an element of $\mathbb{Z}(i)$, say $\beta = (a + bi)/n$ with $a, b, n \in \mathbb{Z}$, such that $|\alpha - \beta| \leq \epsilon/n$.

(d) Let ϵ be a positive real number, and for each element $\beta = (a + bi)/n$ of $\mathbb{Q}(i)$, $a, b, n \in \mathbb{Z}$, consider the disc of radius ϵ/n about β. Prove that the interiors of these discs cover the complex plane.

(e) Extend the method of Proposition (7.9) to prove the finiteness of the class number for any imaginary quadratic field.

*9. (a) Let R be the ring of functions which are polynomials in $\cos t$ and $\sin t$, with real coefficients. Prove that $R \approx \mathbb{R}[x, y]/(x^2 + y^2 - 1)$.

(b) Prove that R is not a unique factorization domain.

*(c) Prove that $\mathbb{C}[x, y]/(x^2 + y^2 - 1)$ is a principal ideal domain and hence a unique factorization domain.

*10. In the definition of a Euclidean domain, the size function σ is assumed to have as range the set of nonnegative integers. We could generalize this by allowing the range to be some other ordered set. Consider the product ring $R = \mathbb{C}[x] \times \mathbb{C}[y]$. Show that we can define a size function $R - \{0\} \longrightarrow S$, where S is the ordered set $\{0, 1, 2, 3, \ldots; \omega, \omega + 1, \omega + 2, \omega + 3, \ldots\}$, so that the division algorithm holds.

*11. Let φ: $\mathbb{C}[x, y] \longrightarrow \mathbb{C}[t]$ be a homomorphism, defined say by $x \rightsquigarrow x(t), y \rightsquigarrow y(t)$. Prove that if $x(t)$ and $y(t)$ are not both constant, then $\ker \varphi$ is a nonzero principal ideal.