MA40238 NUMBER THEORY
2014/15 SEMESTER 1

ZIYU ZHANG

1. Unique Factorisation and Applications

We review the notion of unique factorisation and give some applications of unique factorisation in the ring of integers.

1.1. **Factorisation in integral domains.** We have studied this topic extensively in Algebra 2B. Here we review some important notions and results. In this lecture we always assume R is a commutative ring with 1, such that $0 \neq 1$. We say R is an *integral domain* if for $a, b \in R$ with $ab = 0$, we have either $a = 0$ or $b = 0$. We recall the definitions of Euclidean domains, principal ideal domains, unique factorisation domains, along with other relevant concepts and notations. (If you learned Algebra 2B in 2013, you have seen the mathematical content of these terminologies without knowing some of the names.)

Definition 1.1. Let R be an integral domain. A *Euclidean valuation* on R is a map

$$\nu : R \backslash \{0\} \to \{0, 1, 2, \cdots\}$$

such that if $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ with the property that $a = qb + r$ and either $r = 0$ or $\nu(r) < \nu(b)$. R is said to be a *Euclidean domain* if it has a Euclidean valuation.

Example 1.2. We recall some important examples of Euclidean domains

(1) The ring of integers $\mathbb{Z}$ is an Euclidean domain, with the absolute value function $\nu(n) = |n|$ being a Euclidean valuation.

(2) For $\Bbbk$ a field, the polynomial ring of a single variable $\Bbbk[x]$ is an Euclidean domain, with the degree function $\nu(f(x)) = \deg f(x)$ being a Euclidean valuation.

(3) The ring of Gaussian integers

$$\mathbb{Z}[i] = \{a + bi \in \mathbb{C} \mid a, b \in \mathbb{Z}\}$$

is an integral domain as it is a subring of the field of complex numbers $\mathbb{C}$. The function

$$\nu(a + bi) = a^2 + b^2$$

provides a Euclidean valuation. See Exercise 1.4.

Definition 1.3. Let R be an integral domain. An ideal I of R is a *principal ideal* if $I = (a)$ for some $a \in R$. R is a *principal ideal domain (PID)* if every ideal of R is principal.

Remark 1.4. Notice that we use a slightly different notation from the one you used in Algebra 2B. Here $(a) = Ra$ is the ideal generated by $a \in R$.

Theorem 1.5. *Every Euclidean domain is a PID.*

Date: January 8, 2015.

Proof. See Theorem 2.5 (2013) or Theorem 3.10 (2014) in Algebra 2B. (The 2013 version only proves this result in special cases, but some minor changes would make it into a complete proof for arbitrary Euclidean domains, which is given in the 2014 version.) □

By Theorem 1.5, all examples discussed in Example 1.2 are PIDs.

Before proceeding we review some basic definitions.

Definition 1.6. Let R be an integral domain. If $a, b \in R$ with $b \neq 0$, we say that b *divides* a if $a = bc$ for some $c \in R$. We denote it by $b \mid a$. (Otherwise we write $b \nmid a$.) An element $u \in R$ is called a *unit* if u divides 1. Two elements $a, b \in R$ are said to be *associated* if $a = bu$ for some unit u.

Remark 1.7. We can restate everything in the language of ideals: $b \mid a$ iff $(a) \subset (b)$; $u \in R$ is a unit iff $(u) = R$; a and b are associates iff $(a) = (b)$. See Lemma 2.9 (2013) or Lemmas 3.15 and 3.16 (2014) in Algebra 2B.

Definition 1.8. Let R be an integral domain. A non-unit $p \in R$ is said to be *irreducible* if $a \mid p$ implies that a is either a unit or an associate of p. A non-unit $p \in R$ is said to be *prime* if $p \neq 0$ and $p \mid ab$ implies that $p \mid a$ or $p \mid b$.

Proposition 1.9. *We have*

(1) *Let R be an integral domain. Then every prime element is irreducible.*

(2) *Let R be a PID. Then every irreducible element is prime.*

Proof. For (1), see Proposition 2.10 (2013) or Proposition 3.19 (2014) in Algebra 2B. For (2), see Proposition 2.12 (2013) or Proposition 3.21 (2014). □

Clearly, for all examples discussed in Example 1.2, the two notions "prime" and "irreducible" agree, so we can use them interchangeably. For historical reasons we usually say "primes" in $\mathbb{Z}$ and "irreducible polynomials" in $\Bbbk[x]$.

We move on to the definition of unique factorisation domains.

Definition 1.10. An integral domain R is a *unique factorisation domain (UFD)* if the following conditions are satisfied:

(1) Every non-zero non-unit element in R can be written as the product of finitely many irreducible elements in R;

(2) Given two such factorisations, say $r_1 r_2 \cdots r_s = r'_1 r'_2 \cdots r'_t$, we have $s = t$, and after renumbering if necessary, each r'_i is an associate of r_i for $1 \leqslant i \leqslant s$.

Theorem 1.11. *Every PID is a UFD.*

Proof. See Theorem 2.14 (2013) or Theorem 3.26 (2014) in Algebra 2B. □

By Theorem 1.11, all examples discussed in Example 1.2 are UFDs.

Remark 1.12. Sometimes we prefer to eliminate the ambiguity of the factorisations coming from units. The relation of being associated is an equivalence relation which partitions irreducible elements into equivalence classes. From each equivalence class we pick a representative and denote the set of all representatives (one from each class) by S. For instance, in $\mathbb{Z}$ we can take the set of all positive primes (irreducibles and primes agree in $\mathbb{Z}$); in $\Bbbk[x]$ we can take the set of all monic

(leading coefficient 1) irreducible polynomials. Then every non-zero element $a \in R$ can be written in the form

$$a = ur_1r_2 \cdots r_s$$

where u is a unit and $r_1, \cdots, r_s \in S$. Moreover u is unique and $r_1, r_2, \cdots, r_s$ are unique up to renumbering.

Corollary 1.13 (Fundamental Theorem of Arithmetic)**.** *Every non-zero integer n admits a prime factorisation*

$$n = (-1)^\epsilon p_1^{a_1} p_2^{a_2} \cdots p_s^{a_s}$$

where $\epsilon = 0$ or 1, s is a non-negative integer, $p_1, p_2, \cdots, p_s$ are distinct positive primes, $a_1, a_2, \cdots, a_s$ are positive integers. This factorisation is unique up to the order of factors.

Proof. We have seen that unique factorisation holds for $\mathbb{Z}$. By writing products of repeated factors as powers we get the desired form. □

Remark 1.14. Unique factorisation in the ring of integers has fundamental importance. However, unique factorisation fails for some other integral domains studied in number theory. Understanding why it fails and how to fix it, is an important topic in algebraic number theory. We will come back to this later.

The following famous result of Euclid is a nice application of the fundamental theorem of arithmetic. The proof is simple and clever.

Theorem 1.15. *There are infinitely many primes in $\mathbb{Z}$.*

Proof. It suffices to prove there are infinitely many positive primes in $\mathbb{Z}$. We prove by contradiction. Assume there are only finitely many positive primes. We can label all of them in increasing order $p_1, p_2, \cdots, p_n$. Let $N = p_1p_2 \cdots p_n + 1$. Then N is greater than 1 and not divisible by any p_i, $i = 1, 2, \cdots, n$. On the other hand, N can be factored into product of primes and hence is divisible by some prime p, which is different from any p_i. Contradiction! □

1.2. **Arithmetic functions.** An arithmetic function is a complex valued function defined on the set of positive integers, or in other words, simply a sequence of complex numbers.

Definition 1.16. An *arithmetic function* is a function

$$f : \mathbb{Z}^+ \to \mathbb{C}$$

where $\mathbb{Z}^+$ is the set of all positive integers.

In principle one could assign any complex number as the value of the function at an positive integer. We look at some examples.

Example 1.17. Here are some very simple examples.

(1) For every complex number $c \in \mathbb{C}$, we can define the constant function

$$f_c : \mathbb{Z}^+ \to \mathbb{C} \quad \text{given by} \quad f_c(n) = c \quad \text{for every } n \in \mathbb{Z}^+.$$

In particular, we denote the function which takes constant values 1 by I.

(2) Another function which will show up later will be the function $\mathbb{I}$ defined by

$$\mathbb{I}(n) = \begin{cases} 1 & \text{if } n = 1; \\ 0 & \text{if } n > 1. \end{cases}$$

However we are mainly interested in arithmetic functions with a meaningful assignment of values, most of which take values in integers.

Example 1.18. Here are some naturally defined arithmetic functions.

- For any $n \in \mathbb{Z}^+$, define $\nu(n)$ to be the number of positive divisors of n;
- For any $n \in \mathbb{Z}^+$, define $\sigma(n)$ to be the sum of the positive divisors of n.

By virtue of the unique factorisation, we can obtain the following formulas for the two functions:

Proposition 1.19. *Assume the integer $n > 1$ has the prime decomposition*

$$n = p_1^{a_1} p_2^{a_2} \cdots p_l^{a_l},$$

where $p_1, p_2, \cdots, p_l$ are distinct positive primes. Then we have

$$\nu(n) = (a_1 + 1)(a_2 + 1) \cdots (a_l + 1);$$
$$\sigma(n) = \frac{p_1^{a_1+1} - 1}{p_1 - 1} \cdot \frac{p_2^{a_2+1} - 1}{p_2 - 1} \cdots \frac{p_l^{a_l+1} - 1}{p_l - 1}.$$

Proof. To prove the first formula, we notice that $m \mid n$ iff

$$m = p_1^{b_1} p_2^{b_2} \cdots p_l^{b_l}$$

with $0 \leqslant b_i \leqslant a_i$ for every i. Thus the positive divisors of n are one-to-one correspondent to the n-tuples $(b_1, b_2, \cdots, b_l)$ with $0 \leqslant b_i \leqslant a_i$ for every i, and there are exactly

$$(a_1 + 1)(a_2 + 1) \cdots (a_l + 1)$$

such n-tuples.

To prove the second formula, we notice that

$$\sigma(n) = \sum_{1 \leqslant b_1 \leqslant a_1, 1 \leqslant b_2 \leqslant a_2, \cdots, 1 \leqslant b_l \leqslant a_l} p_1^{b_1} p_2^{b_2} \cdots p_l^{b_l}$$

where the sum is over the above set of n-tuples. Thus we can see that

$$\sigma(n) = (1 + p_1 + p_1^2 + \cdots + p_1^{a_1})(1 + p_2 + p_2^2 + \cdots + p_2^{a_2}) \cdots (1 + p_l + p_l^2 + \cdots + p_l^{a_l})$$

from which the result follows by applying the summation formula for geometric series. □

Next example is another arithmetic function which will play an important role in Möbius inversion theorem. For convenience, we say an integer n *square-free* if it is not divisible by the square of any integer greater than 1. An equivalent characterisation: n is square-free iff n does not have repeated prime factors in its prime decomposition. In other words, n is square-free iff n is the product of finitely many distinct primes.

Definition 1.20. For any positive integer n, we define the Möbius μ-function by

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1; \\ 0 & \text{if } n \text{ is not square-free;} \\ (-1)^l & \text{if } n = p_1 p_2 \cdots p_l \text{ is the product of } l \text{ distinct primes.} \end{cases}$$

We prove the following property of Möbius μ-function. Again, the unique factorisation is the key to the proof.

Proposition 1.21. *For any $n \in \mathbb{Z}^+$, we have*

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1, \end{cases}$$

where the summation runs over all positive divisors of n.

Proof. The case of $n = 1$ is clear. Now we assume $n \geqslant 2$. Let $n = p_1^{a_1} p_2^{a_2} \cdots p_l^{a_l}$ be the prime decomposition of n for some $l \in \mathbb{Z}^+$. The definition of μ-function shows that only those divisors d of n which do not have repeated prime factors contribute to the summation. For any i with $0 \leqslant i \leqslant l$, we consider the number of divisors d of n which are products of i distinct primes. Since the prime factors of d form a subset of those of n, there are exactly $\binom{l}{i}$ choices for such d, each of which contributes $(-1)^i$ to $\mu(n)$. Therefore we have

$$\begin{aligned} \sum_{d|n} \mu(d) &= \binom{l}{0} - \binom{l}{1} + \binom{l}{2} - \binom{l}{3} + \cdots + (-1)^l \binom{l}{l} \\ &= (1-1)^l = 0. \end{aligned}$$

□

The definition of the μ-function seems somewhat artificial at the first glance. However its significance will not be revealed until we introduce Dirichlet products of arithmetic functions.

1.3. **Dirichlet product and Möbius inversion.** Dirichlet product will be a handy tool for establishing Möbius inversion.

Definition 1.22. Let $f, g : \mathbb{Z}^+ \to \mathbb{C}$ be two arithmetic functions. The *Dirichlet product* (or *Dirichlet convolution*) of f and g is the arithmetic function $f * g$ defined by the formula

$$(f * g)(n) = \sum_{d_1 d_2 = n} f(d_1) g(d_2)$$

where the sum runs over all pairs (d_1, d_2) of positive integers such that $d_1 d_2 = n$.

Remark 1.23. Another equivalent way of writing the formula is

$$(f * g)(n) = \sum_{d|n} f(d) g(\frac{n}{d}),$$

where the sum is over all positive divisors d of n. We will use both formulas in the following discussion.

The Dirichlet product has many nice properties. In particular, it is commutative and associative, as we expect for any "product".

Lemma 1.24. *Let $f, g, h : \mathbb{Z}^+ \to \mathbb{C}$ be arithmetic functions, then*

$$\begin{aligned} f * g &= g * f \\ (f * g) * h &= f * (g * h). \end{aligned}$$

Proof. Commutativity is immediate. Indeed, for any $n \in \mathbb{Z}^+$, we have

$$(f * g)(n) = \sum_{d_1 d_2 = n} f(d_1) g(d_2) = \sum_{d_2 d_1 = n} g(d_2) f(d_1) = (g * f)(n).$$

Associativity requires some more manipulations. For any $n \in \mathbb{Z}^+$, we show that both expressions $((f*g)*h)(n)$ and $(f*(g*h))(n)$ can be transformed into the summation $\sum_{d_1 d_2 d_3 = n} f(d_1) g(d_2) h(d_3)$, where the sum runs over all 3-tuples (d_1, d_2, d_3) of positive integers such that $d_1 d_2 d_3 = n$.

For the left-hand side, we have

$$\begin{aligned}((f*g)*h)(n) &= \sum_{d_0d_3=n}(f*g)(d_0)h(d_3)\\ &= \sum_{d_0d_3=n}\left(\sum_{d_1d_2=d_0}f(d_1)g(d_2)\right)h(d_3)\\ &= \sum_{d_1d_2d_3=n}f(d_1)g(d_2)h(d_3).\end{aligned}$$

The computation for the right-hand side is similar and gives the same expression. So we are done. □

Example 1.25. Here are some simple examples of Dirichlet products:

(1) Let $\mathbb{I}$ be the function defined in Example 1.17 and f an arbitrary arithmetic function, then

$$\mathbb{I}*f = f*\mathbb{I} = f;$$

(2) Let I be the function with constant value 1 and f an arbitrary arithmetic function, then for every $n \in \mathbb{Z}^+$, we have

$$(f*I)(n) = \sum_{d\,|\,n} f(d);$$

(3) In particular, let f be the μ-function defined in Definition 1.20, then by Proposition 1.21 we have

$$\mu*I = I*\mu = \mathbb{I}.$$

We are ready to prove the following theorem:

Theorem 1.26 (Möbius Inversion Theorem)**.** *Let $f : \mathbb{Z}^+ \to \mathbb{C}$ be an arithmetic function. If we define the arithmetic function $F : \mathbb{Z}^+ \to \mathbb{C}$ by*

$$F(n) = \sum_{d\,|\,n} f(d),$$

then we have

$$f(n) = \sum_{d\,|\,n} \mu(d)F(\frac{n}{d}).$$

Proof. We use the full power of Lemma 1.24 and Example 1.25. The definition of F shows $F = f*I$. Then we have

$$f = f*\mathbb{I} = f*(I*\mu) = (f*I)*\mu = F*\mu = \mu*F,$$

which is what we want by Remark 1.23. □

As an immediate application of the theorem, we use it to obtain a formula for yet another important arithmetic function: the Euler ϕ-function.

Definition 1.27. The Euler ϕ-function is defined to be the following arithmetic function: for any $n \in \mathbb{Z}^+$, $\phi(n)$ is the number of integers m with $1 \leqslant m \leqslant n$ and $\mathrm{hcf}(m, n) = 1$.

We first prove the following simple property of the ϕ-function.

Proposition 1.28. *For any $n \in \mathbb{Z}^+$, the Euler ϕ-function satisfies the identity*

$$\sum_{d\,|\,n} \phi(d) = n.$$

Proof. Consider the n rational numbers

$$\frac{1}{n}, \frac{2}{n}, \cdots, \frac{n-1}{n}, \frac{n}{n}.$$

Reduce each to lowest terms; i.e. perform cancellations to express each number as a quotient of relatively prime integers. The denominators will all be divisors of n. If $d \mid n$, there are exactly $\phi(d)$ of our numbers whose denominators are equal to d after reducing to lowest terms. Thus they sum up to n, as desired. □

We can obtain a formula for the ϕ-function by Möbius inversion theorem.

Proposition 1.29. *Let $n = p_1^{a_1} p_2^{a_2} \cdots p_l^{a_l}$ be the factorisation of $n \in \mathbb{Z}^+$ where $p_1, p_2, \cdots, p_l$ are distinct primes, then*

$$\phi(n) = n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\cdots\left(1 - \frac{1}{p_l}\right).$$

Proof. By Theorem 1.26 and Proposition 1.28, we have that

$$\begin{aligned}\phi(n) &= \sum_{d \mid n} \mu(d)\frac{n}{d} \\ &= n - \sum_i \frac{n}{p_i} + \sum_{i<j} \frac{n}{p_i p_j} - \sum_{i<j<k} \frac{n}{p_i p_j p_k} + \cdots \\ &= n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\cdots\left(1 - \frac{1}{p_l}\right),\end{aligned}$$

as desired. □

Remark 1.30. Using the same factorisation of n, we can also write the formula for Euler ϕ-function in a slightly different form:

$$\phi(n) = p_1^{a_1-1} p_2^{a_2-1} \cdots p_l^{a_l-1}(p_1 - 1)(p_2 - 1)\cdots(p_l - 1).$$

Indeed, we can substitute n by its prime factorisation in the previous formula and cancel all denominators with the corresponding prime factors in n to get this formula. Caution: it does not imply that each p_i is still a prime factor of $\phi(n)$ because the exponent $a_i - 1$ could be zero.

Exercise Sheet 1

This sheet is due in the lecture on Tuesday 7th October, and will be discussed in the exercise class on Friday 10th October.

Exercise 1.1. *Review of highest common factors.*

(1) Use Euclidean algorithm to compute $\mathrm{hcf}(963, 657)$ and find a pair of integers m, n satisfying $963m + 657n = \mathrm{hcf}(963, 657)$.

(2) For non-zero integers a and b, let $d = \mathrm{hcf}(a, b)$, $a = da'$ and $b = db'$. Show that $\mathrm{hcf}(a', b') = 1$. (Hint: write $d = am + bn$ for some $m, n \in \mathbb{Z}$.)

Exercise 1.2. *Examples of arithmetic functions.*

(1) Compute the values of $\nu(n)$, $\sigma(n)$, $\mu(n)$, $\phi(n)$ for $n = 360$ and $n = 429$.

(2) For any integer $n \geqslant 3$, show that $\phi(n)$ is even.

(3) For any integer $n \geqslant 2$, show that the sum of all elements in the set $\{m \in \mathbb{Z} \mid 1 \leqslant m \leqslant n, \mathrm{hcf}(m, n) = 1\}$ is $\frac{1}{2}n\phi(n)$.

Exercise 1.3. *Applications of Möbius inversion.*

(1) Show that $\sum_{d|n} \mu(\frac{n}{d})\nu(d) = 1$ for any $n \in \mathbb{Z}^+$;

(2) Show that $\sum_{d|n} \mu(\frac{n}{d})\sigma(d) = n$ for any $n \in \mathbb{Z}^+$.

Exercise 1.4. *Unique factorisation in the ring of Gaussian integers.*

Consider the ring of Gaussian integers $\mathbb{Z}[i] = \{a + bi \in \mathbb{C} \mid a, b \in \mathbb{Z}\}$ and the function $\nu : \mathbb{Z}[i] \to \{0, 1, 2, \cdots\}$ given by $\nu(a + bi) = a^2 + b^2$ (the absolute value as a complex number).

(1) Verify that for all $\alpha, \beta \in \mathbb{Z}[i]$, $\nu(\alpha\beta) = \nu(\alpha)\nu(\beta)$. (Hint: either compute it directly, or use the fact that $\nu(\alpha) = \alpha \cdot \overline{\alpha}$.)

(2) Show that the function ν is a Euclidean valuation. (Hint: for $\alpha, \beta \in \mathbb{Z}[i]$, consider $\frac{\alpha}{\beta}$ as a complex number. Choose q to be the Gaussian integer which is the nearest to $\frac{\alpha}{\beta}$ in the complex plane.)

(3) Conclude that unique factorisation holds for $\mathbb{Z}[i]$.

(4) Show that $\alpha \in \mathbb{Z}[i]$ is a unit iff $\nu(\alpha) = 1$. Conclude that the only units in $\mathbb{Z}[i]$ are ± 1 and $\pm i$.

(5) For $\alpha \in \mathbb{Z}[i]$, suppose $\nu(\alpha)$ is a prime in $\mathbb{Z}$. Show that α is irreducible in $\mathbb{Z}[i]$.

(6) Show that $(2 + i)(2 - i) = 5 = (1 + 2i)(1 - 2i)$ are two factorisations of 5 into irreducible elements in $\mathbb{Z}[i]$. How is this consistent with unique factorisation?

Solutions to Exercise Sheet 1

We provide at least one solution to each problem. Other approaches are also possible for some problems.

Solution 1.1. *Review of highest common factors.*

(1) Using Euclidean algorithm, we have

$$\begin{aligned} 963 &= 657 \times 1 + 306; \\ 657 &= 306 \times 2 + 45; \\ 306 &= 45 \times 6 + 36; \\ 45 &= 36 \times 1 + 9; \\ 36 &= 9 \times 4 + 0. \end{aligned}$$

Hence we know $\mathrm{hcf}(963, 657) = 9$ which is the last non-zero remainder. Then we go backwards to find a linear combination which gives 1.

$$\begin{aligned} 9 &= 45 - 36 \\ &= 45 - (306 - 45) \\ &= 45 \times 7 - 306 \\ &= (657 - 306) - 306 \\ &= 657 \times 7 - 306 \times 15 \\ &= 657 \times 7 - (963 - 657) \times 15 \\ &= 657 \times 22 - 963 \times 15. \end{aligned}$$

So $m = -15$ and $n = 22$ is one solution.

(2) Since $d = \mathrm{hcf}(a, b)$, there exist some $m, n \in \mathbb{Z}$, such that $d = am + bn$. (For example, the Euclidean algorithm can always give such a pair of (m, n).) By substituting, we get $d = da'm + db'n$, hence $1 = a'm + b'n$. If $\mathrm{hcf}(a', b') = k$, then $k \mid a'$ and $k \mid b'$, thus $k \mid a'm + b'n = 1$, which implies $\mathrm{hcf}(a', b') = 1$.

Solution 1.2. *Examples of arithmetic functions.*

(1) We factor $360 = 2^3 \times 3^2 \times 5^1$. By the formulas in Proposition 1.19, Definition 1.20 and Proposition 1.29, we have

$$\begin{aligned}
\nu(360) &= (3+1)(2+1)(1+1) = 24;\\
\sigma(360) &= \frac{2^4-1}{2-1} \times \frac{3^3-1}{3-1} \times \frac{5^2-1}{5-1} = 15 \times 13 \times 6 = 1170;\\
\mu(360) &= 0 \qquad \text{since 360 is not square-free;}\\
\phi(360) &= 360 \times (1-\frac{1}{2})(1-\frac{1}{3})(1-\frac{1}{5}) = 96.
\end{aligned}$$

Similarly we have $429 = 3 \times 11 \times 13$. Therefore

$$\begin{aligned}
\nu(429) &= (1+1)(1+1)(1+1) = 8;\\
\sigma(429) &= \frac{3^2-1}{3-1} \times \frac{11^2-1}{11-1} \times \frac{13^2-1}{13-1} = 4 \times 12 \times 14 = 672;\\
\mu(429) &= (-1)^3 = -1 \qquad \text{since 429 is square-free;}\\
\phi(429) &= 429 \times (1-\frac{1}{3})(1-\frac{1}{11})(1-\frac{1}{13}) = 240.
\end{aligned}$$

(2) There are two different proofs. We show one of them here. The other proof will be given together with part (3). We consider two separate cases: if n has any odd prime factor p, then by Remark 1.30, $\phi(n)$ has a factor $p-1$ hence is even; if n has no odd prime factor, then we can write $n = 2^a$ for some $a \geqslant 2$, which implies $\phi(n) = 2^{a-1}$ by the same formula hence is even.

(3) Let $S = \{m \in \mathbb{Z} \mid 1 \leqslant m \leqslant n, \mathrm{hcf}(m,n) = 1\}$. When $n = 2$, the only element in S is 1, hence it is clear that the statement holds. From now on we assume $n \geqslant 3$. For every integer k with $k \leqslant \frac{n}{2}$, we consider the pair of integers $\{k, n-k\}$.

Let $m = \mathrm{hcf}(k,n)$ and $m' = \mathrm{hcf}(n-k,n)$. Then $m \mid k$ and $m \mid n$, hence $m \mid n-k$, which implies $m \mid m'$. A similar argument shows $m' \mid m$. Therefore $m = m'$, which implies either k and $n-k$ are both in S, or neither is in S.

The two integers k and $n-k$ in a pair are distinct unless $k = \frac{n}{2}$, which happens when n is even. However in such a case $\frac{n}{2} \notin S$ because $\mathrm{hcf}(n, \frac{n}{2}) = \frac{n}{2} > 1$. We conclude that S can be divided into pairs of distinct integers of the form $\{k, n-k\}$, which proves the number of elements in S, i.e. $\phi(n)$, is even. Moreover the sum of the two integers in a pair is n, and there are precisely $\frac{\phi(n)}{2}$ pairs in S (since there are $\phi(n)$ elements in S). This implies the sum of all elements in S is $n \cdot \frac{\phi(n)}{2}$, as required.

Solution 1.3. *Applications of Möbius inversion.*

By Example 1.18, for every $n \in \mathbb{Z}^+$, we can write

$$\begin{aligned}
\nu(n) &= \sum_{d \mid n} 1;\\
\sigma(n) &= \sum_{d \mid n} d.
\end{aligned}$$

Therefore we apply Theorem 1.26 for $f(n) = 1$ and $F(n) = \nu(n)$ to obtain

$$1 = \sum_{d \mid n} \mu(d) \nu\left(\frac{n}{d}\right) = \sum_{d \mid n} \mu\left(\frac{n}{d}\right) \nu(d)$$

which is the first statement. For $f(n) = n$ and $F(n) = \sigma(n)$ we obtain

$$n = \sum_{d|n} \mu(d)\sigma\left(\frac{n}{d}\right) = \sum_{d|n} \mu\left(\frac{n}{d}\right)\sigma(d)$$

which is the second statement.

Solution 1.4. *Unique factorisation in the ring of Gaussian integers.*

(1) The formula is in fact true for any complex numbers. For any $\alpha \in \mathbb{C}$, we have $\nu(\alpha) = \alpha\overline{\alpha}$. Hence for any $\alpha, \beta \in \mathbb{C}$, we have

$$\nu(\alpha\beta) = \alpha\beta \cdot \overline{\alpha}\overline{\beta} = \alpha\overline{\alpha} \cdot \beta\overline{\beta} = \nu(\alpha)\nu(\beta).$$

(2) The commutative ring $\mathbb{Z}[i]$ does not have zero divisors because it is a subring of $\mathbb{C}$ in which there is no zero divisor. Now we check that ν is a Euclidean valuation.

Let $\alpha = a + bi$ and $\beta = c + di \neq 0$. We can divide α by β as complex numbers and write $\frac{\alpha}{\beta} = r + si$ where r, s are real numbers. Choose integers m, n such that $|r - m| \leqslant \frac{1}{2}$ and $|s - n| \leqslant \frac{1}{2}$ (the choice may not be unique). Set $\gamma = m + ni$, then $\gamma \in \mathbb{Z}[i]$ and $\nu(\frac{\alpha}{\beta} - \gamma) = (r - m)^2 + (s - n)^2 \leqslant \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. Set $\delta = \alpha - \beta\gamma$, then $\delta \in \mathbb{Z}[i]$ and either $\delta = 0$ or $\nu(\delta) = \nu(\beta(\frac{\alpha}{\beta} - \gamma)) = \nu(\beta)\nu(\frac{\alpha}{\beta} - \gamma) \leqslant \frac{1}{2}\nu(\beta) < \nu(\beta)$. Hence ν defines a Euclidean valuation on $\mathbb{Z}[i]$, and $\mathbb{Z}[i]$ is a Euclidean domain.

(3) By Theorem 1.5 and Theorem 1.11, we know that a Euclidean domain is a UFD. Hence by part (2) we conclude that $\mathbb{Z}[i]$ is a UFD.

(4) Assume α is a unit, then there exists $\beta \in \mathbb{Z}[i]$, such that $ab = 1$. We apply the Euclidean valuation ν on both sides and use part (1) to get $\nu(\alpha)\nu(\beta) = \nu(1) = 1$. Since both $\nu(\alpha)$ and $\nu(\beta)$ are non-negative integer, the only possibility is $\nu(\alpha) = \nu(\beta) = 1$.

On the other hand, assume $\nu(\alpha) = 1$. Let $\alpha = a + bi$, then $a^2 + b^2 = 1$. This implies $(a + bi)(a - bi) = 1$. Since $a \pm bi \in \mathbb{Z}[i]$, we conclude that α divides 1, hence α is a unit.

To find all the units, we need to find all pairs of integers a, b such that $a^2 + b^2 = 1$. This is only possible when $a = \pm 1$ and $b = 0$, or $a = 0$ and $b = \pm 1$. In other words, $\alpha = \pm 1$ or $\pm i$.

(5) We prove by contradiction. Assume α is not irreducible. Then we can write $\alpha = \alpha_1\alpha_2$ where neither factor is zero or a unit. By part (4) we know $\nu(\alpha_1)$ and $\nu(\alpha_2)$ are both positive integers larger than 1. Therefore by part (1) we know $\nu(\alpha) = \nu(\alpha_1)\nu(\alpha_2)$ is composite, not a prime. Contradiction.

(6) We first show they are both irreducible factorisations of 5. We only need to check all factors are irreducible. This is true by part (5) because $\nu(2 \pm i) = \nu(1 \pm 2i) = 5$ is a prime integer.

We explain why this is consistent with unique factorisation. By Definition 1.10, unique factorisation means the number of irreducible factors agrees in two factorisations, and the corresponding factors are associated after reordering. In this example we have two irreducible factors in either factorisation. We can reorder the factors as $(2 + i)(2 - i) = 5 = (1 - 2i)(1 + 2i)$. Notice that $2 + i = i \cdot (1 - 2i)$ and i is a unit in $\mathbb{Z}[i]$, so $2 + i$ and $1 - 2i$ are associated. Similarly $2 - i = (-i) \cdot (1 + 2i)$ implies that $2 - i$ and $1 + 2i$ are also associated.

2. Congruences

We first recall the notion of congruence, then study how to solve linear congruence equations. The Chinese remainder theorem is important in solving simultaneous equations.

2.1. **Congruences and linear equations.** We recall the following definition from Discrete Mathematics and Programming:

Definition 2.1. If $a, b, m \in \mathbb{Z}$ and $m \neq 0$, we say that a is *congruent to b modulo m* if m divides $b - a$. This relation is written as

$$a \equiv b \pmod{m}.$$

For any $a \in \mathbb{Z}$, the set $\overline{a} = \{n \in \mathbb{Z} \mid n \equiv a \pmod{m}\}$ of integers congruent to a modulo m is called a *congruence class* modulo m. The set of congruence classes modulo m is denoted by $\mathbb{Z}_m$.

Remark 2.2. Although the notion of congruence is still well-defined for any non-zero integer m, we are usually only interested in positive values of m, as congruences modulo m and $-m$ coincide.

We have seen the following structure on $\mathbb{Z}_m$:

Proposition 2.3. *For any non-zero integer m, the set $\mathbb{Z}_m$ has the structure of a commutative ring with identity. In fact, it is the quotient ring $\mathbb{Z}/(m)$ where (m) is the principal ideal of $\mathbb{Z}$ generated by m.*

Proof. See Example (1) on Page 10 (2013) or Examples 1.20 and 1.35 (2014) in Algebra 2B. □

The cancellation law for congruences will be handy for solving congruence equations.

Proposition 2.4 (Cancellation Law)**.** *For any $a, b, k, m \in \mathbb{Z}$, $k \neq 0$, $m \neq 0$, assume* $\mathrm{hcf}(k, m) = d$*, then $ka \equiv kb \pmod{m}$ iff $a \equiv b \pmod{\frac{m}{d}}$.*

Proof. See Exercise 2.3. □

Now we turn to look at congruence equations. In general a congruence equation has the form

$$f(x) \equiv 0 \pmod{m},$$

where $f(x)$ is a polynomial with integer coefficients and m is a non-zero integer. We are only interested in solutions modulo m; i.e. solutions in $\mathbb{Z}_m$. *The number of solutions* is the number of congruence classes in $\mathbb{Z}_m$ which satisfy the given equation.

Proposition 2.5. *For any $a, b, m \in \mathbb{Z}$, $a \neq 0$, $m \neq 0$, assume* $\mathrm{hcf}(a, m) = d$*, then the congruence equation $ax \equiv b \pmod{m}$ has solutions iff $d \mid b$. In this case there are exactly d solutions in $\mathbb{Z}_m$. If x_0 is a solution, then the complete set of solutions is given by the congruence classes of $x_0, x_0 + m', x_0 + 2m', \cdots, x_0 + (d-1)m'$, where $m' = \frac{m}{d}$.*

Proof. If x_0 is a solution, then $ax_0 - b = my_0$ for some integer y_0. Thus $ax_0 - my_0 = b$. Since d divides $ax_0 - my_0$, we must have $d \mid b$.

Conversely, suppose that $d \mid b$ then $b = cd$ for some $c \in \mathbb{Z}$. Since $\mathrm{hcf}(a, m) = d$, there exist integers x_0' and y_0' such that $ax_0' - my_0' = d$. Multiply both sides of the equation by c. Then $a(x_0'c) - m(y_0'c) = b$. Let $x_0 = x_0'c$. Then $ax_0 \equiv b \pmod{m}$.

We have shown that $ax \equiv b \pmod{m}$ has a solution iff $d \mid b$.

Suppose that x_0 and x_1 are solutions. $ax_0 \equiv b \pmod{m}$ and $ax_1 \equiv b \pmod{m}$ imply that $ax_1 \equiv ax_0 \pmod{m}$. By Proposotion 2.4, it is equivalent to $x_1 \equiv x_0 \pmod{m'}$, hence x_1 is a solution iff $x_1 = x_0 + km'$ for some integer k. Moreover, for each $k \in \mathbb{Z}$ there are integers r and s such that $k = rd + s$ and $0 \leqslant s < d$. Thus $x_1 = x_0 + sm' + rm$, or equivalently, $x_1 \equiv x_0 + sm' \pmod{m}$. These solutions are in d distinct congruence classes modulo m. This completes the proof. □

We immediately have the following corollary:

Corollary 2.6. *If* $\mathrm{hcf}(a, m) = 1$*, then* $ax \equiv b \pmod{m}$ *has exactly one solution. In particular, if* p *is a prime and* $p \nmid a$*, then* $ax \equiv b \pmod{p}$ *has exactly one solution.*

Proof. In this caes $d = 1$ so clearly $d \mid b$, and there is exactly $d = 1$ solution. □

In practice, we can solve such equations by cancellations and the Euclidean algorithm.

Example 2.7. As an example we consider the congruence $9x \equiv 6 \pmod{15}$. Since $d = \mathrm{hcf}(9, 15) = 3$ divides 6, the equation has 3 solutions modulo 15. By Proposition 2.4 we can cancel 3 on both sides and reduce the equation to $3x \equiv 2 \pmod{5}$. Euclidean algorithm shows that $\mathrm{hcf}(3, 5) = 1$ and $3 \times 2 + 5 \times (-1) = 1$, thus $3 \times 2 \equiv 1 \pmod{5}$. Then we multiply both sides by 2 and get $x \equiv 4 \pmod{5}$. Therefore the solutions to the original equation are $x \equiv 4, 9, \text{ or } 14 \pmod{15}$.

From $3x \equiv 2 \pmod{5}$ we can also try to add multiples of 5 to 2 until we can cancel the coefficient 3. In this case we have $3x \equiv 2 + 5 \times 2 \pmod{5}$. By Proposition 2.4 we still get $x \equiv 4 \pmod{5}$. Hence the solutions to the original equation are $x \equiv 4, 9, \text{ or } 14 \pmod{15}$.

Proposition 2.5 can also be used to solve linear Diophantine equations of the form $ax + by = c$, where $a, b, c \in \mathbb{Z}$. We explain it by the following example.

Example 2.8. We want to find all integer solutions to the equation $9x + 15y = 6$. We solve it by considering the congruence equation $9x \equiv 6 \pmod{15}$. The computation above has showed that the solution is given by $x \equiv 4 \pmod{5}$, i.e. $x = 5k + 4$ for any $k \in \mathbb{Z}$. By substitution we have $9(5k + 4) + 15y = 6$, so $y = -3k - 2$. Therefore all solutions are given by $x = 5k + 4, y = -3k - 2$ where k is an arbitrary integer.

Now we apply Proposition 2.5 to study the group of units in the ring $\mathbb{Z}_m$.

Proposition 2.9. *Let* m *be a positive integer. An element* $\overline{a} \in \mathbb{Z}_m$ *is a unit iff* $\mathrm{hcf}(a, m) = 1$*. There are exactly* $\phi(m)$ *units in* $\mathbb{Z}_m$*.* $\mathbb{Z}_m$ *is a field iff* m *is a prime.*

Proof. $\overline{a} \in \mathbb{Z}_m$ is a unit iff $ax \equiv 1 \pmod{m}$ is solvable. By Proposition 2.5, this is equivalent to $\mathrm{hcf}(a, m) \mid 1$, hence equivalent to a and m being coprime.

The number of units is precisely the number of such a's with $1 \leqslant a \leqslant m$ and $\mathrm{hcf}(a, m) = 1$. By Definition 1.27, there are precisely $\phi(m)$ units in $\mathbb{Z}_m$.

If p is a prime and $\overline{a} \neq 0$ in $\mathbb{Z}_p$, then $\mathrm{hcf}(a, p) = 1$. Thus every non-zero element of $\mathbb{Z}_p$ is a unit, which shows that $\mathbb{Z}_p$ is a field.

If m is not a prime, then we can write $m = m_1 m_2$, where $1 < m_1, m_2 < m$. Thus $\overline{m_1} \neq \overline{0}$ and $\overline{m_2} \neq \overline{0}$, but $\overline{m_1} \cdot \overline{m_2} = \overline{m} = \overline{0}$. Therefore $\mathbb{Z}_m$ is not a field. □

We immediately obtain the following corollaries, both of which have their own names:

Corollary 2.10 (Euler's Theorem)**.** *If* $\mathrm{hcf}(a, m) = 1$*, then we have* $a^{\phi(m)} \equiv 1 \pmod{m}$*.*

Proof. The units in $\mathbb{Z}_m$ form a group of order $\phi(m)$. If a and m are coprime, $\overline{a}$ is a unit. Thus $\overline{a}^{\phi(m)} = \overline{1}$, or equivalently, $a^{\phi(m)} \equiv 1 \pmod{m}$. □

Corollary 2.11 (Fermat's Little Theorem)**.** *If* p *is a prime and* $p \nmid a$*, then we have* $a^{p-1} \equiv 1 \pmod{p}$*.*

Proof. If $p \nmid a$, then a are p are relatively prime. Thus $a^{\phi(p)} \equiv 1 \pmod{p}$. The result follows, since for a prime p, we have $\phi(p) = p - 1$. □

2.2. **Chinese remainder theorem.** Sometimes we need to solve a system of congruence equations. The main result for this type of problems is the Chinese remainder theorem. We will continue to work in $\mathbb{Z}$ but this theorem is valid in more general situations; see Proposition 2.17 (2013) or Theorem 2.24 (2014) in Algebra 2B for two other versions.

Theorem 2.12. *Suppose that $m_1, m_2, \cdots, m_k$ are pairwise coprime (i.e. $\mathrm{hcf}(m_i, m_j) = 1$ for $i \neq j$) non-zero integers and $m = m_1 m_2 \cdots m_k$. Then the system of congruence equations*

$$\begin{aligned} x &\equiv b_1 \pmod{m_1}, \\ x &\equiv b_2 \pmod{m_2}, \\ &\cdots, \\ x &\equiv b_k \pmod{m_k}. \end{aligned}$$

has a solution, which is unique modulo m.

Proof. We prove it by induction on k. For $k = 1$ there is nothing to prove.

For $k = 2$, an integer solution to $x \equiv b_1 \pmod{m_1}$ is of the form $x = m_1 q + b_1$. So we need to have $m_1 q + b_1 \equiv b_2 \pmod{m_2}$, or $m_1 q \equiv b_2 - b_1 \pmod{m_2}$. Since $\mathrm{hcf}(m_1, m_2) = 1$, by Proposition 2.5, it has a unique solution for q, say $q \equiv q_0 \pmod{m_2}$. Or equivalently, $q = m_2 r + q_0$ for any $r \in \mathbb{Z}$. Hence $x = m_1 m_2 r + (m_1 q_0 + b_1)$ for any $r \in \mathbb{Z}$, which is the unique solution for x modulo $m = m_1 m_2$.

For general k, suppose we have proved the result for $k - 1$. That is, the first $k - 1$ congruence equations have a unique common solution $x \equiv s \pmod{m'}$ for some s, where $m' = m_1 m_2 \cdots m_{k-1}$. Then the problem reduces to a system of two congruences

$$\begin{aligned} x &\equiv s \pmod{m'}, \\ x &\equiv b_k \pmod{m_k}. \end{aligned}$$

By the case for $k = 2$ above, there is a unique solution for x modulo $m = m' m_k$. This finishes the induction. □

To use the theorem to make explicit computations, we just need to follow the proof. We illustrate the idea using the following example.

Example 2.13. Consider the system

$$\begin{aligned} x &\equiv 31 \pmod{41}, \\ x &\equiv 59 \pmod{26}. \end{aligned}$$

From the first equation we can write $x = 41q + 31$. We plug it into the second equation and get $41q + 31 \equiv 59 \pmod{26}$. By removing multiples of 26 we reduce it to $15q \equiv 2 \pmod{26}$. By Euclidean algorithm, we have $\mathrm{hcf}(15, 26) = 1$ and $15 \times 7 - 26 \times 4 = 1$, which implies $q \equiv 14 \pmod{26}$ is the unique solution for q. If we write $q = 26r + 14$, then $x = 41 \times 26r + (14 \times 41 + 31)$, i.e. $x \equiv 605 \pmod{1066}$.

Remark 2.14. We explain what to do in slightly more complicated situations.

(1) If there are more than two equations in the system, we need to find the common solution to the first two equations, then combine the result with the third equation to find a solution to all three equations, etc. This procedure is reflected by the inductive step in the proof.

(2) If the equations in the system are not in the form of $x \equiv b_i \pmod{m_i}$, we need to solve (at least) one equation before using substitution. See Example 2.15.

(3) In case the m_i's are not pairwise coprime, Theorem 2.12 does not apply any more. Therefore the existence and uniqueness of solutions may not hold. However the substitution method can still be used to solve the system. See Example 2.15.

Example 2.15. Consider the system

$$\begin{aligned} 5x &\equiv 7 \pmod{12}, \\ 7x &\equiv 1 \pmod{10}. \end{aligned}$$

Notice that the coefficients in front of x are not 1. Moreover 12 and 10 are not coprime. We can nevertheless solve it. Using the method in Example 2.7 we find the solution to the first equation $x \equiv 11 \pmod{12}$. Then we write $x = 12q + 11$ and substitute x in the second equation. We get $7(12q + 11) \equiv 1 \pmod{10}$, or $84q \equiv -76 \pmod{10}$. Using the method in Example 2.7 again, we remove multiples of 10 on both sides and cancel the common factor 2 to reduce the equation to $2q \equiv 2 \pmod{5}$, whose solution is $q \equiv 1 \pmod{5}$. Write $q = 5r + 1$ to get $x = 12(5r + 1) + 11 = 60r + 23$. Hence the solution to the original system is $x \equiv 23 \pmod{60}$.

We wish to interpret the Chinese remainder theorem in the language of rings. We need to recall the definition for the direct product of rings; see Definition on Page 27 (2013) or Definition 2.22 (2014) in Algebra 2B.

Definition 2.16. Let $R_1, R_2, \cdots, R_n$ be commutative rings with 1. The *direct product* is the ring

$$R_1 \times R_2 \times \cdots \times R_n = \big\{(a_1, a_2, \cdots, a_n) \mid a_i \in R_i \text{ for each } i\big\},$$

in which addition and multiplication are given component-wise by

$$\begin{aligned} (a_1, a_2, \cdots, a_n) + (b_1, b_2, \cdots, b_n) &= (a_1 + b_1, a_2 + b_2, \cdots, a_n + b_n), \\ (a_1, a_2, \cdots, a_n) \cdot (b_1, b_2, \cdots, b_n) &= (a_1 b_1, a_2 b_2, \cdots, a_n b_n). \end{aligned}$$

Remark 2.17. We make the following observations.

(1) All the algebraic laws hold in $R_1 \times R_2 \times \cdots \times R_n$ since they hold for every component. Clearly the element $(0_{R_1}, 0_{R_2}, \cdots, 0_{R_n})$ is the zero element, and the additive inverse of $(a_1, a_2, \cdots, a_n)$ is $(-a_1, -a_2, \cdots, -a_n)$. The element $(1_{R_1}, 1_{R_2}, \cdots, 1_{R_n})$ is the multiplicative identity. Thus $R_1 \times R_2 \times \cdots \times R_n$ is a commutative ring with 1.

(2) Notice that $(a_1, a_2, \cdots, a_n)$ is a unit in $R_1 \times R_2 \times \cdots \times R_n$ iff a_i is a unit in R_i for each i. We usually denote the group of units of a ring R by R^*, therefore we have

$$(R_1 \times R_2 \times \cdots \times R_n)^* = R_1^* \times R_2^* \times \cdots \times R_n^*.$$

See Remark on Page 27 (2013) or Remark 2.23 (2014) in Algebra 2B.

Now we restate the Chinese remainder theorem as follows:

Corollary 2.18. *Suppose that $m_1, m_2, \cdots, m_k$ are pairwise coprime non-zero integers and $m = m_1 m_2 \cdots m_k$. Then there is a ring isomorphism*

$$\mathbb{Z}_m \cong \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_k}.$$

Proof. For each i there is a natural ring homomorphism $\psi_i : \mathbb{Z} \to \mathbb{Z}_{m_i}$ which maps every integer n to the congruence class modulo m_i containing n. We construct a map $\psi : \mathbb{Z} \to \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_k}$ by $\psi(n) = (\psi_1(n), \psi_2(n), \cdots, \psi(n))$. We can see ψ respects additions and multiplications, because each component ψ_i does. Therefore ψ is a ring homomorphism.

We apply Theorem 2.12. The existence of solutions shows that ψ is surjective; in other words, $\operatorname{im}\psi = \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_k}$. The uniqueness of solutions modulo m shows that $\ker\psi = (m)$. By the fundamental isomorphism theorem of rings (Theorem 1.8 (2013) or Theorem 2.13 (2014) in

Algebra 2B), ψ induces a ring isomorphism $\mathbb{Z}/(m) \cong \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_k}$. By Proposition 2.3, the left-hand side is precisely $\mathbb{Z}_m$. □

We have the following immediate consequence concerning the groups of units.

Corollary 2.19. *Suppose that $m_1, m_2, \cdots, m_k$ are pairwise coprime non-zero integers and $m = m_1 m_2 \cdots m_k$. Then there is a group isomorphism*

$$\mathbb{Z}_m^* \cong \mathbb{Z}_{m_1}^* \times \mathbb{Z}_{m_2}^* \times \cdots \times \mathbb{Z}_{m_k}^*.$$

Proof. We apply Remark 2.17 and Corollary 2.18 and obtain

$$\mathbb{Z}_m^* \cong (\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_k})^* = \mathbb{Z}_{m_1}^* \times \mathbb{Z}_{m_2}^* \times \cdots \times \mathbb{Z}_{m_k}^*$$

as desired. □

Remark 2.20. This result is very helpful in studying the group of units in $\mathbb{Z}_m^*$ for an arbitrary positive integer m. More precisely, let $m = 2^a p_1^{a_1} p_2^{a_2} \cdots p_l^{a_l}$ be the prime decomposition of m, where $p_1, p_2, \cdots p_l$ are distinct odd primes. Since $2^a, p_1^{a_1}, p_2^{a_2}, \cdots, p_l^{a_l}$ are pairwise coprime, we get

$$\mathbb{Z}_m^* \cong \mathbb{Z}_{2^a}^* \times \mathbb{Z}_{p_1^{a_1}}^* \times \mathbb{Z}_{p_2^{a_2}}^* \times \cdots \times \mathbb{Z}_{p_l^{a_l}}^*.$$

Therefore, to understand the group structure of $\mathbb{Z}_m^*$ for an arbitrary m, it suffices to understand it for m being powers of primes. This is what we are going to study next.

Exercise Sheet 2

This sheet is due in the lecture on Tuesday 14th October, and will be discussed in the exercise class on Friday 17th October.

Exercise 2.1. *Solving linear equations.*

(1) Solve the equation $140x \equiv 98 \pmod{84}$.

(2) Solve the equation $28x \equiv 124 \pmod{116}$.

(3) Find all integer solutions to the equation $12x + 7y = 17$.

(4) Let $a, b, c \in \mathbb{Z}$ where a and b are not simultaneously zero. Show that the equation $ax+by = c$ has solutions in integers iff $\mathrm{hcf}(a, b) \mid c$.

Exercise 2.2. *Solving systems of linear equations.*

(1) Solve the system $x \equiv 1 \pmod{7}$, $x \equiv 4 \pmod{9}$, $x \equiv -2 \pmod{5}$.

(2) Solve the system $4x \equiv 6 \pmod{13}$, $6x \equiv 4 \pmod{8}$.

(3) Solve the system $x \equiv 7 \pmod{15}$, $x \equiv 5 \pmod{9}$.

Exercise 2.3. *Cancellation law for congruences.*

Let $a, b, k, m \in \mathbb{Z}$, $k \neq 0$, $m \neq 0$.

(1) Assume $k \mid m$. Show that $ka \equiv kb \pmod{m}$ iff $a \equiv b \pmod{\frac{m}{k}}$;

(2) Assume $\mathrm{hcf}(k, m) = 1$. Show that $ka \equiv kb \pmod{m}$ iff $a \equiv b \pmod{m}$;

(3) In general, assume $\mathrm{hcf}(k, m) = d$. Show that $ka \equiv kb \pmod{m}$ iff $a \equiv b \pmod{\frac{m}{d}}$. (Hint: use parts (1) and (2).)

Exercise 2.4. *Wilson's theorem and beyond.*

(1) Let p be an odd prime. If $k \in \{1, 2, \cdots, p-1\}$, show that there is a unique b_k in this set such that $kb_k \equiv 1 \pmod{p}$.

(2) Show that $k = b_k$ iff $k = 1$ or $k = p - 1$.

(3) Use parts (1) and (2) to prove that $(p-1)! \equiv -1 \pmod{p}$. This is known as Wilson's theorem.

(4) If $n \in \mathbb{Z}$, $n > 1$, is not a prime, show that $(n-1)! \equiv 0 \pmod{n}$ unless $n = 4$.

(5) Let $n \in \mathbb{Z}$, $n > 1$. Conclude that $(n-1)! \equiv -1 \pmod{n}$ iff n is a prime.

Solutions to Exercise Sheet 2

Solution 2.1. *Solving linear equations.*

(1) We use the Euclidean algorithm to compute $\mathrm{hcf}(140, 84)$ and decide if the equation has a solution.

$$\begin{aligned} 140 &= 84 \times 1 + 56; \\ 84 &= 56 \times 1 + 28; \\ 56 &= 28 \times 2 + 0. \end{aligned}$$

Hence $\mathrm{hcf}(140, 84) = 28$, which does not divide 98. By Proposition 2.5, the equation has no solution.

(2) By Euclidean algorithm, we can find $\mathrm{hcf}(28, 116)$.

$$\begin{aligned} 116 &= 28 \times 4 + 4; \\ 28 &= 4 \times 7 + 0. \end{aligned}$$

Hence $\mathrm{hcf}(28, 116) = 4$, which divides 124. So the equation has 4 solutions modulo 116. We can solve it first by cancelling 4 to get $7x \equiv 31 \pmod{29}$, which reduces to $7x \equiv 2 \pmod{29}$. Now we use Euclidean algorithm for the pair 7 and 29.

$$\begin{aligned} 29 &= 7 \times 4 + 1; \\ 7 &= 1 \times 7 + 0. \end{aligned}$$

So we simply have $1 = 29 - 7 \times 4$ hence $7 \times (-4) \equiv 1 \pmod{29}$. Multiply both sides by 2 to get $7 \times (-8) \equiv 2 \pmod{29}$. Since we usually prefer to use positive numbers as representatives of congruence classes, we add 29 to -8 to get 21. Hence $x \equiv 21 \pmod{29}$. To get solutions modulo 116, we keep adding 29 to 21 until we get repeated congruence classes. So we have $x \equiv 21, 50, 79$ or $108 \pmod{116}$, which are all solutions to the original equation.

(3) We write it as a congruence equation $12x \equiv 17 \pmod{7}$. Since $\mathrm{hcf}(12, 7) = 1$, we should have a unique solution to it. To solve the equation we can add multiples of 7 to 17 until we can cancel the coefficient 12. Hence we have $12x \equiv 24 \pmod{7}$, then $x \equiv 2 \pmod{7}$. We write $x = 7k + 2$ for an arbitrary $k \in \mathbb{Z}$, then substitute x in the original equation to get $12(7k + 2) + 7y = 17$. Therefore we have $7y = -84k - 7$ thus $y = -12k - 1$. The solutions to the original equation is $x = 7k + 2, y = -12k - 1$ for an arbitrary $k \in \mathbb{Z}$.

(4) For simplicity we write $d = \mathrm{hcf}(a, b)$. For one direction, assume that $ax + by = c$ has a solution $x = x_0$ and $y = y_0$. Then $ax_0 + by_0 = c$. Since $d \mid a$ and $d \mid b$, we know $d \mid (ax + by)$, which gives $d \mid c$. For the other direction, assume $d \mid c$, then we can write $c = dc'$ for some integer c'. Since $d = \mathrm{hcf}(a, b)$, we can find integers x_0' and y_0', such that $ax_0' + by_0' = d$ (for example, by Euclidean algorithm). Multiply both sides by c', then we get $ax_0'c' + by_0'c' = dc' = c$. Therefore $x = x_0'c'$ and $y = y_0'c'$ is a solution.

Solution 2.2. *Solving systems of linear equations.*

(1) We find a common solution to the first two equations. From the first equation we can write $x = 7q + 1$. Substituting x in the second equation to get $7q + 1 \equiv 4 \pmod 9$, hence $7q \equiv 3 \pmod 9$. Adding 18 to 3 and we get $7q \equiv 21 \pmod 9$, hence $q \equiv 3 \pmod 9$. Write $q = 9r + 3$ to get $x = 7(9r + 3) + 1 = 63r + 22$. So the solution to the first two equations is $x \equiv 22 \pmod{63}$. Now we bring the third equation into the question. By substitution we get $63r + 22 \equiv -2 \pmod 5$, hence $63r \equiv -24 \pmod 5$. We reduce it to $3r \equiv 1 \pmod 5$, hence $3r \equiv 6 \pmod 5$, which gives $r \equiv 2 \pmod 5$. Write $r = 5s + 2$ to get $x = 63(5s + 2) + 22 = 315s + 148$. So the solution to the original system is $x \equiv 148 \pmod{315}$.

(2) Since $\mathrm{hcf}(4, 13) = 1$ divides 6 and $\mathrm{hcf}(6, 8) = 2$ divides 4, both equations have solutions. From $4x \equiv 6 \pmod{13}$ we get $4x \equiv 32 \pmod{13}$ hence $x \equiv 8 \pmod{13}$. Write $x = 13q + 8$ and substitute x in the second equation to get $6(13q + 8) \equiv 4 \pmod 8$. We write it as $78q \equiv -44 \pmod 8$ and reduce it to $6q \equiv 4 \pmod 8$. By cancelling 2 we get $3q \equiv 2 \pmod 4$. By adding 4 to 2 we get $3q \equiv 6 \pmod 4$ hence $q \equiv 2 \pmod 4$. We write $q = 4r + 2$, then $x = 13(4r + 2) + 8 = 52r + 34$. So the solution is $x \equiv 34 \pmod{52}$.

Remark: you might ask if the result is consistent with the Chinese remainder theorem because the modulus is not $13 \times 8 = 104$. In fact, the solution to the first equation is $x \equiv 8 \pmod{13}$. And the second equation has two solutions $x \equiv 2 \pmod 8$ and $x \equiv 6 \pmod 8$. By the Chinese remainder theorem, they combine to give two solutions to the original system, which are $x \equiv 34 \pmod{104}$ and $x \equiv 86 \pmod{104}$. They can be represented by a single congruence $x \equiv 34 \pmod{52}$.

(3) From the first equation we can write $x = 15q + 7$. We substitute x in the second equation to get $15q + 7 \equiv 5 \pmod 9$. That is $15q \equiv -2 \pmod 9$, which reduces to $6q \equiv 7 \pmod 9$. Notice that $\mathrm{hcf}(6, 9) = 3$ which does not divide 7. By Proposition 2.5, this equation has no solution. Hence so is the original system.

Solution 2.3. *Cancellation law for congruences.*

(1) Since $k \mid m$, we can write $m = km'$ for some integer m'. For one direction, assume $ka \equiv kb \pmod m$. Then there exists some $c \in \mathbb{Z}$ such that $ka - kb = cm$. We divide both sides by k to get $a - b = cm'$, which implies $a \equiv b \pmod{m'}$, as required.

For the other direction, assume $a \equiv b \pmod{m'}$. Then there exists some $c \in \mathbb{Z}$ such that $a - b = cm'$. We multiply both sides by k to get $ka - kb = ckm' = cm$, which implies $ka \equiv kb \pmod m$.

(2) Since $ka \equiv kb \pmod m$, we know $m \mid (ka - kb) = k(a - b)$. Since $\mathrm{hcf}(k, m) = 1$, we claim that we have $m \mid (a - b)$. Indeed, using the condition $\mathrm{hcf}(k, m) = 1$, we can find some $\alpha, \beta \in \mathbb{Z}$, such that $k\alpha + m\beta = 1$. Multiply both sides by $a - b$ to get $k(a-b)\alpha + m(a-b)\beta = a - b$. Since m divides both terms on the left-hand side, we conclude that m divides the right-hand side; i.e. $m \mid (a - b)$. It follows that $a \equiv b \pmod m$.

For the other direction, assume $a \equiv b \pmod m$. Then we know $m \mid (a - b)$, hence $m \mid k(a - b) = ka - kb$. It follows that $ka \equiv kb \pmod m$.

(3) Since $\mathrm{hcf}(k, m) = d$, we can write $k = dk'$ and $m = dm'$. By Exercise 1.1 (2), we know $\mathrm{hcf}(k', m') = 1$. The condition $ka \equiv kb \pmod m$ is equivalent to $dk'a \equiv dk'b \pmod{dm'}$, which is equivalent to $k'a \equiv k'b \pmod{m'}$ by part (1), which is further equivalent to $a \equiv b \pmod{m'}$ by part (2). This proves the equivalence required in question.

Solution 2.4. *Wilson's theorem and beyond.*

(1) We write $S = \{1, 2, \cdots, p - 1\}$. For any $k \in S$, $p \nmid k$ hence $\mathrm{hcf}(k, p) = 1$, which implies $kx \equiv 1 \pmod p$ has a unique solution modulo p by Proposition 2.5. Since the congruence class $\overline{0}$ is not the solution, this solution must be a congruence class $\overline{b}$ for some b not divisible

by p. This congruence contains exactly one element in the set S, which we call b_k. Therefore this b_k is the unique solution in S to the equation $kx \equiv 1 \pmod{p}$.

(2) When $k = 1$, it is clear that $b_k = 1$ does satisfy the equation $kb_k \equiv 1 \pmod{p}$. When $k = p - 1$, it is also clear that $b_k = p - 1$ satisfy the same equation because $kb_k = (p-1)(p-1) \equiv (-1)(-1) = 1 \pmod{p}$.

It remains to show that these are the only values of k which make $k = b_k$. In other words, if $k^2 \equiv 1 \pmod{p}$ is satisfied by some $k \in S$, we want to show that $k = 1$ or $k = p - 1$. Indeed, the equation $k^2 \equiv 1 \pmod{p}$ is equivalent to $p \mid (k^2 - 1) = (k + 1)(k - 1)$, which implies that either $p \mid k + 1$ or $p \mid k - 1$ because p is a prime. If $p \mid k + 1$, then $k \equiv -1 \pmod{p}$, so the only value in S is $k = p - 1$. If $p \mid k - 1$, then $k \equiv 1 \pmod{p}$, so the only value in S is $k = 1$. This shows that the only values for k which make $k = b_k$ are $k = 1$ and $k = p - 1$.

(3) By parts (1) and (2), the set $S\backslash\{1, p-1\}$ can be divided into pairs, such that the product of the two elements in each pair is congruent to 1 modulo p. Hence the product of all elements in $S\backslash\{1, p-1\}$ is congruent to 1 modulo p. Taking the remaining two elements 1 and $p - 1$ into consideration, the product of all elements in S is congruent to $p - 1$ modulo p, or equivalently, -1 modulo p.

(4) Assume n is composite and $n \neq 4$, then we can write $n = ab$ for some $a, b \in \mathbb{Z}$, $1 < a, b < n$. There are two cases. If $a \neq b$, then a and b appear as distinct factors in $(n-1)!$. Hence $(n-1)!$ is a multiple of ab. In other words, $(n-1)! \equiv 0 \pmod{n}$. If $a = b$, then the assumption implies $a = b \geqslant 3$, hence $2a < ab = n$. Now a and $2a$ appear as distinct factors in $(n-1)!$. Hence $(n-1)!$ is a multiple of $a \cdot 2a = 2ab = 2n$, which implies $(n-1)! \equiv 0 \pmod{n}$. When $n = 4$, we have $(4-1)! = 3! = 6 \equiv 2 \pmod{4}$.

(5) The "if" part is proved in part (3) for odd primes, and is clear for $n = 2$. The contrapositive of the "only if" part is proved in part (4). Therefore the condition $(n-1)! \equiv -1 \pmod{n}$ is equivalent to n being a prime.

3. Primitive Roots

We study the group structure of $\mathbb{Z}_m^*$ for any integer $m \geqslant 2$. In particular, we wish to know when it is a cyclic group. This leads to the notion of the primitive root.

3.1. The cases of primes and powers of 2.

We start with the definition of primitive roots.

Definition 3.1. Let $a, m \in \mathbb{Z}$, $m \geqslant 2$, $\mathrm{hcf}(a, m) = 1$. a is said to be a *primitive root* modulo m if the group of units $\mathbb{Z}_m^*$ is cyclic and the congruence class $\overline{a}$ is a generator.

Remark 3.2. We make some comments about this definition.

(1) Assume a and m and coprime. The *order* of a modulo m is defined to be the order of $\overline{a}$ in the group of units $\mathbb{Z}_m^*$. For any integer n, $a^n \equiv 1 \pmod{m}$ iff n is a multiple of the order of a modulo m. In this terminology, a is a primitive root modulo m iff a is coprime to m and the order of a modulo m is $\phi(m)$.

(2) Knowing that a is a primitive root modulo m allows us to write

$$\mathbb{Z}_m^* = \{\, \overline{a}^{\,k} \mid k \in \mathbb{Z}, 0 \leqslant k < \phi(m) \,\}.$$

In other words, every integer coprime to m is congruent to a^k for some $k \in \mathbb{Z}$. This will be extremely helpful in many different situations. See Exercises 3.2 and 3.3.

(3) If a is a primitive root modulo m, then $\mathbb{Z}_m^*$ is cyclic of order $\phi(m)$ hence has $\phi(\phi(m))$ generators. More precisely, any primitive root modulo m lies in the congruence class $\overline{a}^k$ for some k with $0 \leqslant k < \phi(m)$ and $\mathrm{hcf}(k, \phi(m)) = 1$.

We have seen in Remark 2.20 that it is essential to understand $\mathbb{Z}_m^*$ when m is a power of a prime in order to understand the general case. We first consider the situation when m is a prime. We need the following lemma:

Lemma 3.3. *Let $f(x) \in \Bbbk[x]$ where $\Bbbk$ is a field. Suppose that $\deg f(x) = n$. Then f has at most n distinct roots in $\Bbbk$.*

Proof. The proof goes by induction on n. For $n = 0$ the assertion is trivial. Assume that the statement is true for polynomials of degree $n-1$. If $f(x)$ has no roots in k, we are done. If α is a root, since $k[x]$ is a Euclidean domain, we can write $f(x) = (x-\alpha)q(x) + r$, where r is a constant. Setting $x = \alpha$ we see that $r = 0$. Thus $f(x) = (x-\alpha)q(x)$ and $\deg q(x) = n-1$. If $\beta \neq \alpha$ is another root of $f(x)$, then $0 = f(\beta) = (\beta - \alpha)q(\beta)$, which implies that $q(\beta) = 0$. Since by induction $q(x)$ has at most $n-1$ distinct roots, $f(x)$ has at most n distinct roots. □

The following theorem is useful in many situations.

Theorem 3.4. *Let K be a field and K^* the group of non-zero elements under multiplication. Suppose G is a finite subgroup of K^*, then G is cyclic.*

Proof. We prove by strong induction on $n = |G|$. If $n = 1$ there is nothing to prove. Now we assume any subgroup of K^* with order smaller than n is cyclic.

For any d with $d \mid n$ and $d < n$, we write $G_d = \{g \in G \mid g^d = 1\}$. We claim G_d is a subgroup of G. Indeed, $1 \in G_d$ because $1^d = 1$. If $g_1, g_2 \in G_d$, then $(g_1 g_2)^d = g_1^d g_2^d = 1$ because multiplication is commutative in the field K. Therefore G_d is closed under multiplication. Moreover, if $g \in G_d$, then $(g^{-1})^d = (g^d)^{-1} = 1$, hence G_d is closed under taking inverse. These conclude that G_d is a group, thus a subgroup of G. Each element of G_d is a solution to $x^d - 1 = 0$ in K, so $|G_d| \leqslant d$ by Lemma 3.3. By induction hypothesis we know G_d is a cyclic group.

Let $\psi(d)$ be the number of elements of order d in G. Each such element is contained in G_d, so $\psi(d)$ is also the number of elements of order d in G_d. If $|G_d| < d$ then $\psi(d) = 0$. Otherwise G_d is a cyclic group of order d and $\psi(d) = \phi(d)$. So we always have $\psi(d) \leqslant \phi(d)$.

On one hand $\psi(n) + \sum_{d \mid n, d<n} \psi(d) = n$ since the order of any element of G is a divisor of n. On the other hand $\phi(n) + \sum_{d \mid n, d<n} \phi(d) = n$ by Proposition 1.28. Since for each $d < n$ we have $\psi(d) \leqslant \phi(d)$, we must have $\psi(n) \geqslant \phi(n) > 0$. In other words, there are elements of order n in G, hence G is cyclic. □

The following immediate consequence has fundamental importance. It was first proved by Gauss.

Corollary 3.5. *Let p be a prime, then $\mathbb{Z}_p^*$ is a cyclic group; i.e. there exist primitive roots modulo p.*

Proof. By Proposition 2.9, $\mathbb{Z}_p$ is a field. Then the result follows from Theorem 3.4. □

Next we study the case of prime powers. We will show that primitive roots exist for powers of odd primes, but the situation is completely different for powers of 2. The necessity of treating 2 differently from the other primes occurs repeatedly in number theory.

Proposition 3.6. *Let l be a positive integer. Then $\mathbb{Z}_{2^l}^*$ is not cyclic unless $l = 1$ or 2.*

Proof. It is easy to see that 1 is a primitive root modulo 2, and 3 is a primitive root modulo 4. From now on we assume that $l \geqslant 3$. We claim that

$$a^{2^{l-2}} \equiv 1 \pmod{2^l}$$

for every odd integer a. It means that the order of every element in $\mathbb{Z}_{2^l}^*$ is strictly smaller than $\phi(2^l)$, hence $\mathbb{Z}_{2^l}^*$ cannot be cyclic.

We prove this claim by induction on l. When $l = 3$, $\mathbb{Z}_8^* = \{\overline{1}, \overline{3}, \overline{5}, \overline{7}\}$. We can check them one by one and conclude $a^2 \equiv 1 \pmod 8$ for any odd integer a. Now we assume the claim holds for l, then we can write $a^{2^{l-2}} = 1 + b \cdot 2^l$, thus

$$a^{2^{l-1}} = (1 + b \cdot 2^l)^2 = 1 + b \cdot 2^{l+1} + b^2 \cdot 2^{2l}.$$

The last two terms are divisible by 2^{l+1}, hence $a^{2^{l-1}} \equiv 1 \pmod{2^{l+1}}$, i.e. the claim holds for $l + 1$. □

Remark 3.7. For enthusiasts: for any $l \geqslant 3$, we actually have $\mathbb{Z}_{2^l}^* \cong \mathbb{Z}_2 \times \mathbb{Z}_{2^{l-2}}$ which is the direct product of two cyclic groups. We do not prove this fact but it is not difficult.

3.2. **The case of odd prime powers and the general case.** We first show that primitive roots always exist for powers of odd primes. After that we wrap up and give a list of all values of $m \geqslant 2$ which possess primitive roots.

Proposition 3.8. *Let p be an odd prime and $l \geqslant 2$ an integer. Then $\mathbb{Z}_{p^l}^*$ is cyclic; i.e. there exist primitive roots modulo p^l.*

Proof. We prove the result in three steps. We first produce a candidate, then prove that it is indeed a primitive root modulo p^l.

Step 1. By Corollary 3.5, we assume g is a primitive root modulo p. Then we have $g^{p-1} \equiv 1 \pmod p$. We claim that we can choose g such that $g^{p-1} \not\equiv 1 \pmod{p^2}$.

In fact, if g satisfies $g^{p-1} \equiv 1 \pmod{p^2}$, we can consider $g + p$, which is still a primitive root modulo p. However we have

$$\begin{aligned}(g+p)^{p-1} &\equiv g^{p-1} + (p-1)g^{p-2}p \pmod{p^2} \\ &\equiv 1 + (p-1)g^{p-2}p \pmod{p^2} \\ &\not\equiv 1 \pmod{p^2},\end{aligned}$$

which shows that we can replace g by $g + p$ and achieve our claim.

Step 2. By Step 1 we can write $g^{p-1} \equiv 1 + ap \pmod{p^2}$ for some $a \in \mathbb{Z}$ not divisible by p. We claim that for each $l \geqslant 2$, we similarly have

$$g^{\phi(p^{l-1})} \equiv 1 + a \cdot p^{l-1} \pmod{p^l}. \tag{3.1}$$

We prove it by induction on l. When $l = 2$, the claim follows from Step 1. Assume the claim is true for some $l \geqslant 2$, then we can write

$$g^{\phi(p^{l-1})} = 1 + b \cdot p^{l-1}$$

for some $b \in \mathbb{Z}$ with $a \equiv b \pmod p$. Then

$$g^{\phi(p^l)} = (1 + b \cdot p^{l-1})^p = 1 + b \cdot p^l + \sum_{i=2}^{p-1} \binom{p}{i} b^i \cdot p^{i(l-1)} + b^p \cdot p^{p(l-1)}.$$

We know $\binom{p}{i}$ is divisible p. (Indeed, we have $p! = i!(p-i)!\binom{p}{i}$ by the definition of binomial coefficients. The left-hand side is divisible by p, hence so is the right-hand side. But p does not

divide $i!(p-i)!$ since it is a product of integers less than, and thus coprime to p. Hence p divides $\binom{p}{i}$.) Therefore for each $i \geqslant 2$, the corresponding term in the summation is divisible by $p^{1+i(l-1)}$, where $1+i(l-1) \geqslant 1+2(l-1) \geqslant l+1$. The term after the summation is divisible by $p^{p(l-1)}$, where $p(l-1) \geqslant 3(l-1) \geqslant l+1$ since p is an odd prime. Also notice that the difference of a and b is a multiple of p. All this together implies

$$g^{\phi(p^l)} \equiv 1 + a \cdot p^l \pmod{p^{l+1}}. \tag{3.2}$$

Therefore the claim is true for $l+1$.

Step 3. We show that for each $l \geqslant 2$, the order of g modulo p^l is $\phi(p^l)$; i.e. g is a primitive root modulo p^l.

Denote the order of g modulo p^l by d. First of all, $g^d \equiv 1 \pmod{p^l}$ implies $g^d \equiv 1 \pmod{p}$. Since we chose g to be a primitive root modulo p in Step 1, we know that $\phi(p)$ divides d. Then by (3.2) we have $g^{\phi(p^l)} \equiv 1 \pmod{p^l}$, hence d divides $\phi(p^l)$. Finally by (3.1) we have $g^{\phi(p^{l-1})} \not\equiv 1 \pmod{p^l}$, hence d does not divide $\phi(p^{l-1})$. These requirements leave $d = \phi(p^l)$ as the only possibility. □

Remark 3.9. Notice that Steps 2 and 3 in the proof actually shows that: if g is a primitive root modulo p and $g^{p-1} \not\equiv 1 \pmod{p^2}$, then g is a primitive root modulo p^l for any integer $l \geqslant 2$. This sufficient condition will be handy in looking for primitive roots modulo higher powers of odd primes; see Exercise 3.1 for an example. In fact, this condition is also necessary; see Exercise 3.4.

Finally we put all our existing results together and get:

Theorem 3.10. *An integer $m \geqslant 2$ possesses primitive roots iff m is of the form 2, 4, p^k or $2p^k$, where p is an odd prime and k is a positive integer.*

Proof. This proof is not covered in lecture and is non-examinable.

We first show that m possesses primitive roots if it has one of the given forms. We already know this for 2, 4 and p^k. In the last case, by Remark 2.20 we have

$$\mathbb{Z}_{2p^k}^* \cong \mathbb{Z}_2^* \times \mathbb{Z}_{p^k}^* \cong \mathbb{Z}_{p^k}^*,$$

it follows that $\mathbb{Z}_{2p^k}^*$ is cyclic; i.e. $2p^k$ possesses primitive roots.

We then show that n does not possess primitive roots in all other cases. We already know this for $m = 2^l$ with $l \geqslant 3$, so we can now assume m is not a power of 2.

We claim that m can be written as a product $m_1 m_2$, where m_1 and m_2 are coprime, $m_1 > 2$ and $m_2 > 2$. Indeed, assume $m = 2^a p_1^{a_1} p_2^{a_2} \cdots p_l^{a_l}$ is the prime factorisation of m, where $p_1, p_2, \cdots, p_l$ are distinct odd primes, $a \geqslant 0$ and $a_i \geqslant 1$ for each i. If $l \geqslant 2$, then we can take $m_1 = p_1^{a_1}$ and $m_2 = 2^a p_2^{a_2} \cdots p_l^{a_l}$. Otherwise $l = 1$, hence by assumption $a \geqslant 2$, then we can take $m_1 = 2^a$ and $m_2 = p_1^{a_1}$.

We then have that $\phi(m_1)$ and $\phi(m_2)$ are both even by Exercise 1.2 and that $\mathbb{Z}_m^* \cong \mathbb{Z}_{m_1}^* \times \mathbb{Z}_{m_2}^*$ by Remark 2.20. Since every group of even order has an element of order 2, both factors have elements of order 2, which implies that $\mathbb{Z}_m^*$ has at least two elements of order 2. Therefore it is not cyclic since a cyclic group contains at most one element of order 2. Thus m does not possess primitive roots. □

Exercise Sheet 3

This sheet is due in the lecture on Tuesday 21st October, and will be discussed in the exercise class on Friday 24th October.

Exercise 3.1. *Examples of primitive roots.*

(1) Show that 2 is a primitive root modulo 29. How many generators does $\mathbb{Z}_{29}^*$ have?

(2) Show that 2 is a primitive root modulo $1331 = 11^3$. How many generators does $\mathbb{Z}_{1331}^*$ have? (Hint: Remark 3.9.)

(3) Find all primitive roots modulo 10, 11 and 12 respectively, if there is any.

Exercise 3.2. *Applications in solving non-linear equations.*

Let p be an odd prime and g a primitive root modulo p.

(1) For any $d \mid (p-1)$, show that $g^{\frac{p-1}{d}}$ has order d modulo p.

(2) Show that $g^{\frac{p-1}{2}} \equiv -1 \pmod{p}$.

(3) Use the primitive root in Exercise 3.1 (1) to find all solutions to $x^7 \equiv 1 \pmod{29}$.

Exercise 3.3. *Applications in higher order residues.*

Let p be an odd prime and g a primitive root modulo p. Assume $d \mid (p-1)$ and $p \nmid a$.

(1) Show that $x^d \equiv a \pmod{p}$ has solutions iff $a \equiv g^{dk} \pmod{p}$ for some $k \in \mathbb{Z}$.

(2) Show that $x^d \equiv a \pmod{p}$ has solutions iff $a^{\frac{p-1}{d}} \equiv 1 \pmod{p}$.

(3) Find all values of a with $0 < a < 29$ such that $x^4 \equiv a \pmod{29}$ has solutions. (Hint: you can use Exercise 3.1 (1) or Exercise 3.2 (3).)

Exercise 3.4. *Characterisation of primitive roots modulo higher powers of odd primes.*

Let p be an odd prime.

(1) For any positive integer l, if $a \equiv b \pmod{p^l}$, show that $a^p \equiv b^p \pmod{p^{l+1}}$. (Hint: write $a = b + c \cdot p^l$ for some $c \in \mathbb{Z}$ and compute a^p.)

(2) For any positive integers $m < n$, if g is a primitive root modulo p^n, show that g is a primitive root modulo p^m. (Hint: prove by contradiction and use part (1).)

(3) For any integer $l \geqslant 2$, conclude that a necessary and sufficient condition for g being a primitive root modulo p^l is that g is a primitive root modulo p and $g^{p-1} \not\equiv 1 \pmod{p^2}$. (Hint: use part (2) to prove necessity. Sufficiency has been proved in Proposition 3.8; see Remark 3.9.)

Solutions to Exercise Sheet 3

Solution 3.1. *Examples of primitive roots.*

(1) Since the group $\mathbb{Z}_{29}^*$ has $\phi(29) = 28$ elements, we need to show that 2 has order 28 modulo 29. All positive divisors of 28 are $1, 2, 4, 7, 14$ and 28. Since the order of 2 must be a positive divisor of 28, it suffices to show that $2^k \not\equiv 1 \pmod{29}$ for $k = 1, 2, 4, 7, 14$. This can be done by direct computation. $2^1 \equiv 2 \pmod{29}$, $2^2 \equiv 4 \pmod{29}$, $2^4 \equiv 16 \pmod{29}$, $2^7 = 128 \equiv 12 \pmod{29}$, $2^{14} \equiv 12^2 = 144 \equiv 28 \equiv -1 \pmod{29}$. None of these remainders is $1 \pmod{29}$, hence the order of 2 must be 28. In other words, 2 is a primitive root modulo 29. The number of generators of $\mathbb{Z}_{29}^*$ is $\phi(28) = 28(1 - \frac{1}{2})(1 - \frac{1}{7}) = 12$.

(2) By Remark 3.9, it suffices to show that 2 is a primitive root modulo 11 and the condition $2^{10} \not\equiv 1 \pmod{11^2}$. To show 2 is a primitive root modulo 11, we need to show 2 has order 10 modulo 11. In other words, its order is not $1, 2$ or 5. Indeed, $2^1 \equiv 2 \pmod{11}$, $2^2 \equiv 4 \pmod{11}$, $2^5 = 32 \equiv 10 \pmod{11}$. None of them is congruent to 1 modulo 29, hence 2 is a primitive root modulo 11. To show the second condition $2^{10} \not\equiv 1 \pmod{11^2}$, we simply

compute $2^{10} = 1024 \equiv 56 \not\equiv 1 \pmod{121}$. Hence 2 is a primitive root modulo 11^3. The number of generators in $\mathbb{Z}_{11^3}^*$ is given by $\phi(\phi(11^3)) = \phi(10 \times 11^2) = 440$.

(3) We consider primitive roots modulo 10. We have $\phi(10) = 4$ and we can even write down $\mathbb{Z}_{10}^* = \{\overline{1}, \overline{3}, \overline{7}, \overline{9}\}$. We show 3 is a primitive root (in other words $\overline{3}$ is a generator of $\mathbb{Z}_{10}^*$). Indeed, $3 \equiv 3 \pmod{10}$, $3^2 \equiv 9 \pmod{10}$, so the order of 3 modulo 10 is not 1 or 2, hence must be 4. By Remark 3.2 (3), the generators of $\mathbb{Z}_{10}^*$ are $\overline{3}$ and $\overline{3}^3 = \overline{27} = \overline{7}$. Hence $a \in \mathbb{Z}$ is a primitive root modulo 10 iff $a \equiv 3$ or $7 \pmod{10}$.

We consider primitive roots modulo 11. We have found in part (2) that 2 is a primitive root modulo 11. By Remark 3.2 (3), we need to compute the congruence classes of 2^k modulo 11, where $1 \leqslant k \leqslant 10$ and $\mathrm{hcf}(k, 10) = 1$; i.e., $k = 1, 3, 7, 9$. So we have $2^1 \equiv 2 \pmod{11}$, $2^3 \equiv 8 \pmod{11}$, $2^7 = 128 \equiv 7 \pmod{11}$, $2^9 \equiv 7 \times 4 \equiv 6 \pmod{11}$. Therefore $a \in \mathbb{Z}$ is a primitive root modulo 11 iff $a \equiv 2, 6, 7$ or $8 \pmod{11}$.

We finally consider primitive roots modulo 12. We have the factorisation $12 = 2^2 \times 3$. We compare it with the list of forms in Theorem 3.10, but it does not match any of the given forms. Therefore there are no primitive roots modulo 12.

Solution 3.2. *Applications in solving non-linear equations.*

(1) Since g is a primitive root modulo p, we know that the order of g modulo p is $\phi(p) = p - 1$. In other words, $g^{p-1} \equiv 1 \pmod{p}$ and $g^l \not\equiv 1 \pmod{p}$ for any $1 \leqslant l < p - 1$. Let $a = g^{\frac{p-1}{d}}$. We want to show a has order d. In other words, $a^d \equiv 1 \pmod{p}$ and $a^k \not\equiv 1 \pmod{p}$ for any $1 \leqslant k < d - 1$.

On one hand, $a^d = g^{p-1} \equiv 1 \pmod{p}$. On the other hand, for any k with $1 \leqslant k < d$, $a^k \equiv g^{k \cdot \frac{p-1}{d}} \pmod{p}$. Since $0 < k \cdot \frac{p-1}{d} < p - 1$, $a^k \not\equiv 1 \pmod{p}$. Therefore we conclude a has order d modulo p.

(2) Let $b = g^{\frac{p-1}{2}}$. By part (1) we know b has order 2 modulo p. In other words, $b^2 \equiv 1 \pmod{p}$ and $b \not\equiv 1 \pmod{p}$. The first condition implies $p \mid (b^2 - 1) = (b + 1)(b - 1)$, hence either $p \mid b + 1$ or $p \mid b - 1$, or equivalently, $b \equiv -1 \pmod{p}$ or $b \equiv 1 \pmod{p}$. The second condition rules out the second possibility. Hence $g^{\frac{p-1}{2}} = b \equiv -1 \pmod{p}$ is the only possibility.

(3) Let $g = 2$ be the primitive root modulo 29 found in Exercise 3.1 (1), then $g^{28} \equiv 1 \pmod{29}$. Therefore for any $k \in \mathbb{Z}$, $x \equiv g^{4k} \pmod{29}$ is a solution to the equation $x^7 \equiv 1 \pmod{29}$ because $(g^{4k})^7 = g^{28k} \equiv 1^k = 1 \pmod{29}$. In particular, the congruence classes of g^{4k} for $0 \leqslant k \leqslant 6$ are distinct solutions because g has order 28 modulo 29 (indeed, the congruence classes of g^l for $0 \leqslant l < 28$ modulo 29 are all distinct). On the other hand, since $\mathbb{Z}_{29}$ is a field by Proposition 2.9, the equation $x^7 = 1$ has at most 7 distinct solutions in $\mathbb{Z}_{29}$; in other words, at most 7 distinct congruence classes. Therefore $x \equiv g^{4k} \pmod{29}$ for $0 \leqslant k \leqslant 6$ are all solutions. We do explicit computation: $2^0 \equiv 1 \pmod{29}$, $2^4 \equiv 16 \pmod{29}$, $2^8 \equiv 16^2 \equiv 24 \equiv -5 \pmod{29}$, $2^{12} = 2^4 2^8 \equiv 16 \times (-5) \equiv 7 \pmod{29}$, $2^{16} = (2^8)^2 \equiv (-5)^2 = 25 \equiv -4 \pmod{29}$, $2^{20} \equiv 2^4 2^{16} \equiv 16 \times (-4) \equiv 23 \equiv -6 \pmod{29}$, $2^{24} \equiv (2^{12})^2 \equiv 7^2 \equiv 20 \pmod{29}$. Therefore all solutions to the equation $x^7 \equiv 1 \pmod{29}$ are $x \equiv 1, 16, 24, 7, 25, 23$ or $20 \pmod{29}$.

Solution 3.3. *Applications in higher order residues.*

(1) For the "if" part, we assume $a \equiv g^{dk} \pmod{p}$. Then $x \equiv g^k \pmod{p}$ is clearly a solution to $x^d \equiv a \pmod{p}$. For the "only if" part, assume $x^d \equiv a \pmod{p}$ has a solution $x \equiv x_0 \pmod{p}$. Then $p \nmid x_0$ because $x_0^d \equiv a \pmod{p}$ and $p \nmid a$. Therefore $\overline{x}_0$ is an element in $\mathbb{Z}_p^*$ hence $x_0 \equiv g^k \pmod{p}$ for some $k \in \mathbb{Z}$ (because $\overline{g}$ is a generator of $\mathbb{Z}_p^*$). Therefore $a \equiv x_0^d \equiv g^{dk} \pmod{p}$.

(2) By part (1), it suffices to show that $a \equiv g^{dk} \pmod{p}$ is equivalent to $a^{\frac{p-1}{d}} \equiv 1 \pmod{p}$. We first assume $a \equiv g^{dk} \pmod{p}$. Then $a^{\frac{p-1}{d}} \equiv (g^{dk})^{\frac{p-1}{d}} = g^{k(p-1)} \equiv 1 \pmod{p}$ since $g^{p-1} \equiv 1 \pmod{p}$. For the other direction, since $p \nmid a$, $\overline{a} \in \mathbb{Z}_p^*$. Hence $a \equiv g^l \pmod{p}$ for some $l \in \mathbb{Z}$. Then $a^{\frac{p-1}{d}} \equiv g^{l \cdot \frac{p-1}{d}} \equiv 1 \pmod{p}$. Since g has order $p-1$ modulo p, we conclude that $l \cdot \frac{p-1}{d}$ must be a multiple of $p-1$. (This uses a fact in group theory: assume an element g in a group G has order q, then $g^r = e$ is the identity of the group iff $q \mid r$.) In other words, there exists some $k \in \mathbb{Z}$, such that $l \cdot \frac{p-1}{d} = k(p-1)$. This simplifies to $l = dk$, hence $a \equiv g^{dk} \pmod{p}$ for some $k \in \mathbb{Z}$.

(3) We use the result from part (1). $x^4 \equiv a \mod 29$ has solutions iff $a \equiv g^{4k} \pmod{29}$. We know from Exercise 3.1 (1) that $g = 2$ is a primitive root modulo 29. Therefore $a \equiv 2^{4k} \pmod{29}$ for $k \in \mathbb{Z}$. For $0 \leqslant k \leqslant 6$ the formula gives distinct congruence classes. Therefore $x^4 \equiv a \pmod{29}$ has solutions iff $a \equiv 2^{4k}$ for $0 \leqslant k \leqslant 6$. To find the corresponding values of a within the range $0 < a < 29$, we need to find the remainder of each 2^{4k} modulo 29. This calculation has been done in Exercise 3.3 (3); i.e. $a = 1, 16, 24, 7, 25, 23$ or 20.

Solution 3.4. *Characterisation of primitive roots modulo higher powers of odd primes.*

(1) Since $a \equiv b \pmod{p^l}$, we can write $a = b + c \cdot p^l$ for some $c \in \mathbb{Z}$. We then take p-th power on both sides and expand the right-hand side. We get

$$a^p = (b + c \cdot p^l)^p = b^p + p \cdot b^{p-1} c p^l + \sum_{i=2}^{p} \binom{p}{i} b^{p-i} c^i p^{il}.$$

We claim that every term on the right-hand side except b^p is divisible by p^{l+1}. Indeed, the second term $p \cdot b^{p-1} c p^l$ is clearly divisible by p^{l+1}. For every term in the summation, the exponent in the power p^{il} is at least $il \geqslant 2l = l + l \geqslant l + 1$, hence p^{l+1} divides the term $\binom{p}{i} b^{p-i} c^i p^{il}$ for each $i \geqslant 2$. Therefore, modulo p^{l+1}, the above equation can be written as $a^p \equiv b^p \pmod{p^{l+1}}$.

(2) We assume the order of g modulo p^m is d. We need to show $d = \phi(p^m)$. It suffices to prove that $d \mid \phi(p^m)$ and $\phi(p^m) \mid d$. For the first division, notice that $\mathbb{Z}_{p^m}^*$ has order $\phi(p^m)$, hence the order d of any element $\overline{g}$ is a positive divisor of $\phi(p^m)$; that is $d \mid \phi(p^m)$. For the second division, we apply the statement in part (1) on the congruence $g^d \equiv 1 \pmod{p^m}$ for $n - m$ times. Step by step we will get $g^{dp} \equiv 1 \pmod{p^{m+1}}, g^{dp^2} \equiv 1 \pmod{p^{m+2}}, \cdots, g^{dp^{n-m}} \equiv 1 \pmod{p^n}$. Since g has order $\phi(p^n)$ modulo p^n, the last congruence implies $\phi(p^n) \mid dp^{n-m}$. (This uses again the fact in group theory: assume an element g in a group G has order q, then $g^r = e$ is the identity of the group iff $q \mid r$.) Hence $dp^{n-m} = c\phi(p^n) = c(p-1)p^{n-1}$ for some $c \in \mathbb{Z}$. It follows that $d = c(p-1)p^{m-1} = c\phi(p^m)$, hence $\phi(p^m) \mid d$ which is the second division. The two divisions guarantee $d = \phi(p^m)$.

(3) The sufficiency is stated in Remark 3.9 and proved in Proposition 3.8. We still need to prove the necessity of the two given conditions. Since g is a primitive root modulo p^l, using the statement in part (2), we know g is a primitive root modulo p and p^2 because $l \geqslant 2$, which prove the two conditions respectively. Indeed, the first condition is clear. For the second condition, since g has order $\phi(p^2)$ modulo p^2, we know that for any integer d, $1 \leqslant d < \phi(p^2)$, $g^d \neq 1 \pmod{p^2}$. In particular, it holds for $d = p - 1$.

4. Quadratic Residues

We study quadratic residues and non-residues. In this part we are mainly interested in deciding whether a given integer a is a quadratic residue modulo an odd prime p. We will introduce quadratic reciprocity, whose proof will be given in next part.

4.1. **Quadratic residues and the Legendre symbol.** First we recall the definition of quadratic residues and non-residues.

Definition 4.1. For integers a and m, $m \neq 0$, $\mathrm{hcf}(a, m) = 1$, a is called a *quadratic residue* modulo m if the congruence $x^2 \equiv a \pmod{m}$ has a solution. Otherwise a is called a *quadratic non-residue* modulo m.

Given any fixed positive integer m, it is possible to determine the quadratic residues by simply listing the positive integers less than and coprime to m, squaring them, and reducing modulo m. But we prefer to have a more convenient way to determine whether a given integer a coprime to m is a quadratic residue modulo m. At the moment we are mostly interested in the case that m is an odd prime p. An example of a composite m will be given in Exercise 5.3.

The Legendre symbol is a very simple yet powerful tool in studying this problem. Roughly speaking, it is the indication function for quadratic residues. We recall its definition:

Definition 4.2. Let p be an odd prime. The *Legendre symbol* $\left(\frac{a}{p}\right)$ takes value 1 if a is a quadratic residue modulo p, or -1 if a is a quadratic non-residue modulo p, or 0 if p divides a.

Therefore the problem reduces to the computation of the Legendre symbol. There are a series of rules which help with the computation. We introduce them in four groups.

The first group of properties are simple consequences of the definition.

Proposition 4.3. *Let p be an odd prime.*

(1) *If $a \equiv b \pmod{p}$, then $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.*

(2) *If $p \nmid a$, then $\left(\frac{a^2}{p}\right) = 1$.*

Proof. Both statements are clear by definition. □

Next group of properties are more interesting. The proof essentially use the existence of primitive roots.

Proposition 4.4. *Let p be an odd prime.*

(1) (Euler's criterion). $\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}$.

(2) $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$.

Proof. For (1), both sides are congruent to 0 if $p \mid a$. Now we assume $p \nmid a$. Notice that $a^{p-1} \equiv 1 \pmod{p}$ by Corollary 2.11. Hence $(a^{\frac{p-1}{2}} + 1)(a^{\frac{p-1}{2}} - 1) \equiv 0 \pmod{p}$, so $a^{\frac{p-1}{2}} \equiv 1$ or $-1 \pmod{p}$.

If a is a quadratic residue modulo p, assume $a \equiv x^2 \pmod{p}$. Then $p \nmid x$, and $a^{\frac{p-1}{2}} \equiv x^{p-1} \equiv 1 \pmod{p}$ by Corollary 2.11 again. If a is a quadratic non-residue modulo p, it suffices to show $a^{\frac{p-1}{2}} \not\equiv 1 \pmod{p}$. Let g be a primitive root modulo p, then $a \equiv g^r \pmod{p}$ for some $r \in \mathbb{Z}$. We observe that r must be odd, otherwise $a \equiv (g^{\frac{r}{2}})^2 \pmod{p}$ is a quadratic residue. Hence we can write $r = 2k + 1$ for some $k \in \mathbb{Z}$. Then we have $a^{\frac{p-1}{2}} \equiv g^{(2k+1)\cdot\frac{p-1}{2}} \equiv g^{(p-1)k} \cdot g^{\frac{p-1}{2}} \equiv g^{\frac{p-1}{2}} \not\equiv 1 \pmod{p}$ because the order of g modulo p is $p - 1$.

For (2), by (1) we can get $(\frac{ab}{p}) \equiv (ab)^{\frac{p-1}{2}} = a^{\frac{p-1}{2}} b^{\frac{p-1}{2}} \equiv (\frac{a}{p})(\frac{b}{p}) \pmod{p}$. Since both sides can only take values in $\{-1, 0, 1\}$, they must be equal. □

We characterise those primes for which -1 or 2 is a quadratic residue by the follow proposition. We remind the reader that if n is an odd integer, then $n-1$ is always a multiple of 2 and n^2-1 is always a multiple of 8 (we have seen this fact in the proof of Proposition 3.6).

Proposition 4.5. *Let p be an odd prime.*

$$(1)\ \left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = \begin{cases} 1 & \text{if } p \equiv 1 \pmod 4 \\ -1 & \text{if } p \equiv -1 \pmod 4. \end{cases}$$

$$(2)\ \left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod 8 \\ -1 & \text{if } p \equiv \pm 3 \pmod 8. \end{cases}$$

Proof. Part (1) follows immediately from of Proposition 4.4 (1). There are different ways of proving part (2). We provide an elementary proof here. Consider the following $\frac{p-1}{2}$ congruences

$$\begin{aligned} p-1 &\equiv 1\cdot(-1)^1 \pmod p \\ 2 &\equiv 2\cdot(-1)^2 \pmod p \\ p-3 &\equiv 3\cdot(-1)^3 \pmod p \\ &\vdots \\ \frac{p-1}{2} \text{ or } p-\frac{p-1}{2} &\equiv \frac{p-1}{2}\cdot(-1)^{\frac{p-1}{2}} \pmod p. \end{aligned}$$

The pattern on the left-hand side: for every $i = 1, 2, \cdots, \frac{p-1}{2}$, we put i if i is even, or $p-i$ if i is odd. So the left-hand side of the above congruences has exhausted all even numbers between 1 and p. We multiply all of the congruences together to get

$$2\cdot 4\cdot 6\cdots(p-3)\cdot(p-1) \equiv \left(\frac{p-1}{2}\right)!\cdot(-1)^{1+2+\cdots+\frac{p-1}{2}} \pmod p.$$

Therefore we have

$$2^{\frac{p-1}{2}}\cdot\left(\frac{p-1}{2}\right)! \equiv \left(\frac{p-1}{2}\right)!\cdot(-1)^{\frac{p^2-1}{8}} \pmod p.$$

Since p does not divide $\left(\frac{p-1}{2}\right)!$, we can cancel it on both sides to get

$$2^{\frac{p-1}{2}} \equiv (-1)^{\frac{p^2-1}{8}} \pmod p.$$

By Proposition 4.4 (1) we get

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$$

since they both take values 1 or -1.

Finally, if $p \equiv \pm 1 \pmod 8$, then we can write $p = 8k \pm 1$ for some $k \in \mathbb{Z}$. Hence $\frac{p^2-1}{8} = 8k^2 \pm 2k$ is an even number. If $p \equiv \pm 3 \pmod 8$, then we can write $p = 8k \pm 3$ for some $k \in \mathbb{Z}$. Hence $\frac{p^2-1}{8} = 8k^2 \pm 6k + 1$ is an odd number. This proves

$$(-1)^{\frac{p^2-1}{8}} = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod 8 \\ -1 & \text{if } p \equiv \pm 3 \pmod 8, \end{cases}$$

as desired. □

Finally, we state the law of quadratic reciprocity. This is a deep result which has great influence in the modern number theory. The proof will be postponed to next part.

Theorem 4.6 (Law of Quadratic Reciprocity)**.** *Let p and q be distinct odd primes. Then*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}.$$

Remark 4.7. We can state the quadratic reciprocity in a slightly different way: for odd primes p and q, we have $(\frac{q}{p}) = \pm(\frac{p}{q})$. We take the positive sign if either p or q is congruent to 1 modulo 4, or the negative sign if both p and q are congruent to -1 modulo 4.

The law of quadratic reciprocity can be used in conjunction with the previous propositions to compute the Legendre symbol. Very roughly speaking, given a Legendre symbol $(\frac{a}{p})$, after replacing a by the remainder of a modulo p if possible, we use the prime factorisation of a to write $(\frac{a}{p})$ as the product of several Legendre symbols, some of which can be immediately evaluated. Then we use the quadratic reciprocity for the other factors and repeat this process. We give an example:

Example 4.8. We calculate $(\frac{79}{101})$. Since $101 \equiv 1 \pmod 4$ we have $(\frac{79}{101}) = (\frac{101}{79}) = (\frac{22}{79})$. Then we factor as $(\frac{22}{79}) = (\frac{2}{79})(\frac{11}{79})$. Now $79 \equiv 7 \pmod 8$, thus $(\frac{2}{79}) = 1$. Since both 11 and 79 are congruent to 3 modulo 4 we have $(\frac{11}{79}) = -(\frac{79}{11}) = -(\frac{2}{11})$. Finally $11 \equiv 3 \pmod 8$ implies that $(\frac{2}{11}) = -1$. Therefore $(\frac{79}{101}) = 1$; i.e. 79 is a quadratic residue modulo 101. Indeed, we can check $33^2 \equiv 79 \pmod{101}$.

4.2. **The Jacobi symbol.** The Legendre symbol $(\frac{a}{p})$ indicates whether an integer a not divisible by an odd prime p is a quadratic residue modulo p. We have seen how to use quadratic reciprocity to compute it. However this requires to factor a into primes, which is in general a hard problem when a is large. To make the computation easier, we introduce the following generalisation of the Legendre symbol:

Definition 4.9. Let a be any integer and b be a positive odd integer. Let $b = p_1p_2\cdots p_m$ be its prime factorisation, where $p_1, p_2, \cdots, p_m$ are not necessarily distinct primes. The symbol $\left(\frac{a}{b}\right)$ defined by

$$\left(\frac{a}{b}\right) = \left(\frac{a}{p_1}\right)\left(\frac{a}{p_2}\right)\cdots\left(\frac{a}{p_m}\right)$$

is called the *Jacobi symbol*.

The notation for the Jacobi symbol is identical to that for the Legendre symbol. Indeed, when b is an odd prime, the Jacobi symbol $(\frac{a}{b})$ is precisely the corresponding Legendre symbol by definition. Moreover, the Jacobi symbol has properties that are remarkably similar to the Legendre symbol. However, we should also be aware of their difference. We immediately point out some important differences between the two symbols before we show their similarities.

Remark 4.10. We illustrate the following differences between the two symbols by examples.

(1) For $(\frac{a}{b})$, as a Legendre symbol we require that b is a positive odd prime, while as a Jacobi symbol we only require that b is a positive odd integer. So $(\frac{6}{11})$ can be interpreted either as a Legendre symbol or a Jacobi symbol, while $(\frac{14}{45})$ must be a Jacobi symbol.

(2) The Jacobi symbol is in general not an indicator for quadratic residues. That is, $(\frac{a}{b})$ may equal 1 without a being a quadratic residue modulo b. For example, $(\frac{2}{15}) = (\frac{2}{3})(\frac{2}{5}) = (-1)(-1) = 1$, but 2 is not a quadratic residue modulo 15, because if $x^2 \equiv 2 \pmod{15}$ has a solution, then the same integer value of x is a solution to $x^2 \equiv 2 \pmod 3$, which is impossible. It is true, however, that if $(\frac{a}{b}) = -1$, then a is a quadratic non-residue modulo b; see Exercise 4.3 (1).

(3) In comparison to Proposition 4.4 (1), $(\frac{a}{b})$ and $a^{\frac{b-1}{2}}$ are in general not congruent modulo b in case of the Jacobi symbol. For example, $(\frac{2}{15}) \not\equiv 2^{\frac{15-1}{2}} \pmod{15}$ because $(\frac{2}{15}) = 1$ while $2^7 \equiv 8 \pmod{15}$.

Now we turn to the properties of the Jacobi symbol. Apart from what was mentioned above, most of the properties of the Jacobi symbol are extremely similar to those of the Legendre symbol. As a result, the computation of Jacobi symbols are also very similar to that of Legendre symbols, even easier.

The basic idea behind all their proofs is to use the definition to rewrite everything in terms of the Legendre symbol and apply the corresponding properties of the Legendre symbol. We list all the properties and prove only the first one as a sample of the proofs. One more proof will be left as an exercise for you to try yourself; see Exercise 4.3 (2).

Proposition 4.11. *Let b be a positive odd integer.*

(1) *If* $a_1 \equiv a_2 \pmod{b}$, *then* $\left(\frac{a_1}{b}\right) = \left(\frac{a_2}{b}\right)$.

(2) *If* $\gcd(a, b) = 1$, *then* $\left(\frac{a^2}{b}\right) = 1$.

Proof. Both statements are consequences of Definition 4.9 and Proposition 4.3. We assume $b = p_1 p_2 \cdots p_m$ is the prime factorisation of b, where $p_1, p_2, \cdots, p_m$ are not necessarily distinct primes. For (1), by definition and Proposition 4.3 (1) we have

$$\left(\frac{a_1}{b}\right) = \left(\frac{a_1}{p_1}\right)\left(\frac{a_1}{p_2}\right)\cdots\left(\frac{a_1}{p_m}\right) = \left(\frac{a_2}{p_1}\right)\left(\frac{a_2}{p_2}\right)\cdots\left(\frac{a_2}{p_m}\right) = \left(\frac{a_2}{b}\right).$$

For (2), since $p_i \nmid a$ for each i, by definition and Proposition 4.3 (2) we have

$$\left(\frac{a^2}{b}\right) = \left(\frac{a^2}{p_1}\right)\left(\frac{a^2}{p_2}\right)\cdots\left(\frac{a^2}{p_m}\right) = 1.$$

□

Proposition 4.12. *Let b, b_1, b_2 be positive odd integers.*

(1) $\left(\frac{a_1 a_2}{b}\right) = \left(\frac{a_1}{b}\right)\left(\frac{a_2}{b}\right)$.

(2) $\left(\frac{a}{b_1 b_2}\right) = \left(\frac{a}{b_1}\right)\left(\frac{a}{b_2}\right)$.

Proposition 4.13. *Let b be a positive odd integer.*

(1) $\left(\frac{-1}{b}\right) = (-1)^{\frac{b-1}{2}} = \begin{cases} 1 & \text{if } b \equiv 1 \pmod 4 \\ -1 & \text{if } b \equiv -1 \pmod 4. \end{cases}$

(2) $\left(\frac{2}{b}\right) = (-1)^{\frac{b^2-1}{8}} = \begin{cases} 1 & \text{if } b \equiv \pm 1 \pmod 8 \\ -1 & \text{if } b \equiv \pm 3 \pmod 8. \end{cases}$

Proposition 4.14 (Quadratic Reciprocity for the Jacobi symbol)**.** *Let a, b be coprime positive odd integers. Then*

$$\left(\frac{a}{b}\right)\left(\frac{b}{a}\right) = (-1)^{\frac{a-1}{2}\cdot\frac{b-1}{2}}.$$

The Jacobi symbol is very useful. We are mainly interested in using it to calculate Legendre symbols. Roughly speaking, aside from pulling out factors of -1 and 2 as they arise, one can proceed with quadratic reciprocity without worrying about whether or not the numerator is a prime. We show the procedure in the following example.

Example 4.15. Given 1151 is a prime, we compare the two calculations for the Legendre symbol $(\frac{1003}{1151})$.

Without using the Jacobi symbol, we need to factor the numerator $1003 = 17 \times 59$ (it takes some effort to get this!). Hence $(\frac{1003}{1151}) = (\frac{17}{1151})(\frac{59}{1151})$. Since 17 is congruent to 1 modulo 4, we have $(\frac{17}{1151}) = (\frac{1151}{17}) = (\frac{12}{17}) = (\frac{4}{17})(\frac{3}{17}) = (\frac{3}{17})$. By the same reason $(\frac{3}{17}) = (\frac{17}{3}) = (\frac{2}{3}) = -1$. On the other hand since both 59 and 1151 are congruent to -1 modulo 4, we have $(\frac{59}{1151}) = -(\frac{1151}{59}) = -(\frac{30}{59}) = -(\frac{2}{59})(\frac{3}{59})(\frac{5}{59})$. Since $59 \equiv 3 \pmod 8$ we get $(\frac{2}{59}) = -1$. Since $3 \equiv 59 \equiv -1 \pmod 4$ we get $(\frac{3}{59}) = -(\frac{59}{3}) = -(\frac{2}{3}) = 1$. Since $5 \equiv 1 \pmod 4$ we get $(\frac{5}{59}) = (\frac{59}{5}) = (\frac{4}{5}) = 1$. All this together shows $(\frac{59}{1151}) = -1$.

Using the Jacobi symbol, we can avoid the prime factorisation, so the calculation is much simpler. Since 1003 and 1151 are both congruent to -1 modulo 4, the quadratic reciprocity gives $(\frac{1003}{1151}) = -(\frac{1151}{1003}) = -(\frac{148}{1003}) = -(\frac{4}{1003})(\frac{37}{1003}) = -(\frac{37}{1003})$. Since $37 \equiv 1 \pmod 4$, we have $(\frac{37}{1003}) = (\frac{1003}{37}) = (\frac{4}{37}) = 1$. Hence $(\frac{1003}{1151}) = -1$. Works like a charm!

Now we switch gears and discuss a more significant application of the Legendre and Jacobi symbols. From Proposition 4.5 we noticed that -1 is a quadratic residue for primes of the form $4k+1$ and that 2 is a quadratic residue for primes of the form $8k \pm 1$. If a is an arbitrary integer, for what odd primes p is a a quadratic residue modulo p? We illustrate this type of questions using the following example.

Example 4.16. To find all odd primes p for which 3 is a quadratic residue, we need to compute $(\frac{3}{p})$ for all $p \neq 3$. To apply quadratic reciprocity, we need to consider two cases.

If $p \equiv 1 \pmod 4$, then $(\frac{3}{p}) = (\frac{p}{3})$, which is 1 if $p \equiv 1 \pmod 3$, or -1 if $p \equiv 2 \pmod 3$. We can solve the system of the congruences modulo 3 and 4 to obtain: $(\frac{p}{3}) = 1$ if $p \equiv 1 \pmod{12}$, or -1 if $p \equiv 5 \pmod{12}$. (Please fill in the details of the computation.)

On the other hand, if $p \equiv 3 \pmod 4$, then $(\frac{3}{p}) = -(\frac{p}{3})$. Still, $(\frac{p}{3}) = 1$ if $p \equiv 1 \pmod 3$, or -1 if $p \equiv 2 \pmod 3$. We can solve the system of the congruences modulo 3 and 4 to obtain: $(\frac{p}{3}) = 1$ if $p \equiv 7 \pmod{12}$, or -1 if $p \equiv 11 \pmod{12}$. (Please fill in the details of the computation.)

	$p \equiv 1 \pmod 4$	$p \equiv 3 \pmod 4$
$p \equiv 1 \pmod 3$	$(\frac{3}{p}) = 1, p \equiv 1 \pmod{12}$	$(\frac{3}{p}) = -1, p \equiv 7 \pmod{12}$
$p \equiv 2 \pmod 3$	$(\frac{3}{p}) = -1, p \equiv 5 \pmod{12}$	$(\frac{3}{p}) = 1, p \equiv 11 \pmod{12}$

Summarising the above results, we get

$$\left(\frac{3}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1 \text{ or } 11 \pmod{12} \\ -1 & \text{if } p \equiv 5 \text{ or } 7 \pmod{12}. \end{cases}$$

In other words, 3 is a quadratic residue for an odd prime p iff $p \equiv \pm 1 \pmod{12}$.

Exercise Sheet 4

This sheet is due in the lecture on Tuesday 28th October, and will be discussed in the exercise class on Friday 31st October.

Exercise 4.1. *Computation of the Legendre symbol.*

(1) Evaluate the Legendre symbol $(\frac{474}{733})$ without using the Jacobi symbol.

(2) Evaluate the Legendre symbol $(\frac{-113}{997})$.

(3) Evaluate the Legendre symbol $(\frac{514}{1093})$.

Exercise 4.2. *Primes for which a given number is a quadratic residue.*

(1) Find all odd primes for which 5 is a quadratic residue.

(2) Find all odd primes for which -3 is a quadratic residue.

Exercise 4.3. *Properties of Jacobi symbols.*

(1) Let b be a positive odd integer and $\mathrm{hcf}(a,b)=1$. If a is a quadratic residue modulo b, show that the Jacobi symbol $(\frac{a}{b})=1$.

(2) Use Definition 4.9 and Proposition 4.4 (1) to give a proof of Proposition 4.12 (1); i.e. for any positive odd integer b, show that

$$\Big(\frac{a_1a_2}{b}\Big)=\Big(\frac{a_1}{b}\Big)\Big(\frac{a_2}{b}\Big).$$

Exercise 4.4. *Quadratic residues and the Legendre symbol.*

(1) Find all quadratic residues and non-residues modulo 13.

(2) Let p be an odd prime and a any integer. Show that the number of solutions to the congruence $x^2\equiv a \pmod p$ is given by $1+(\frac{a}{p})$.

(3) Use part (2) to show that $\displaystyle\sum_{a=0}^{p-1}\Big(\frac{a}{p}\Big)=0$. (Hint: each congruence class modulo p is a solution to $x^2\equiv a \pmod p$ for a unique $a\in\{0,1,\cdots,p-1\}$.)

(4) Use part (3) to show that, in the set $\{1,2,\cdots,p-1\}$, there are as many quadratic residues as non-residues modulo p. Is your answer to part (1) consistent with this result?

Solutions to Exercise Sheet 4

Solution 4.1. *Computation of the Legendre symbol.*

(1) We factor 474 into primes as $474=2\times3\times79$. Hence $(\frac{474}{733})=(\frac{2}{733})(\frac{3}{733})(\frac{79}{733})$. We have $(\frac{2}{733})=-1$ since $733\equiv5 \pmod 8$. We use quadratic reciprocity to compute the other two factors. Notice that $733\equiv1 \pmod 4$, therefore $(\frac{3}{733})=(\frac{733}{3})=(\frac{1}{3})=1$. For the same reason we have $(\frac{79}{733})=(\frac{733}{79})=(\frac{22}{79})=(\frac{2}{79})(\frac{11}{79})$. Since $79\equiv-1 \pmod 8$ we have $(\frac{2}{79})=1$. Since $11\equiv79\equiv3 \pmod 8$, by quadratic reciprocity we get $(\frac{11}{79})=-(\frac{79}{11})=-(\frac{2}{11})=1$, where the last equality is due to $11\equiv3 \pmod 8$. Hence we have $(\frac{79}{733})=1$. It follows that $(\frac{474}{733})=(-1)\times1\times1=-1$.

(2) The computation is always easier if we use Jacobi symbols. We just need to remember pulling out -1 and 2 from the numerators.

In this problem we have $(\frac{-113}{997})=(\frac{-1}{997})(\frac{113}{997})$. The first factor $(\frac{-1}{997})=1$ since $997\equiv1 \pmod 4$. The second factor $(\frac{113}{997})=(\frac{997}{113})$ by quadratic reciprocity since $113\equiv1 \pmod 4$ (or $997\equiv1 \pmod 4$). Then $(\frac{997}{113})=(\frac{93}{113})=(\frac{113}{93})=(\frac{20}{93})=(\frac{4}{93})(\frac{5}{93})=(\frac{5}{93})=(\frac{93}{5})=(\frac{3}{5})=(\frac{5}{3})=(\frac{2}{3})=-1$, where the second, sixth and eighth equalities are consequences of quadratic reciprocity since $113\equiv1 \pmod 4$ and $5\equiv1 \pmod 4$. Finally we conclude $(\frac{-113}{997})=-1$.

(3) For this one we have $(\frac{514}{1093})=(\frac{2}{1093})(\frac{257}{1093})$. Since $1093\equiv5 \pmod 8$ we get $(\frac{2}{1093})=-1$. Realising $257\equiv1 \pmod 4$ and using quadratic reciprocity, we have $(\frac{257}{1093})=(\frac{1093}{257})=(\frac{65}{257})=(\frac{257}{65})=(\frac{62}{65})$. At this point we can of course factor 62 and do the computation as usual. But there is a shortcut. We write $(\frac{62}{65})=(\frac{-3}{65})=(\frac{-1}{65})(\frac{3}{65})$. Since $65\equiv1 \pmod 4$, we have $(\frac{-1}{65})=1$, and by quadratic reciprocity $(\frac{3}{65})=(\frac{65}{3})=(\frac{2}{3})=-1$. Finally we conclude that $(\frac{514}{1093})=(-1)\times(-1)=1$.

Solution 4.2. *Primes for which a given number is a quadratic residue.*

(1) To find all the odd primes p for which 5 is a quadratic residue, we need to compute $(\frac{5}{p})$ for any odd prime $p \neq 5$ (because p has to be coprime with 5 for being a quadratic residue). Since $5 \equiv 1 \pmod 4$, $(\frac{5}{p}) = (\frac{p}{5})$. By direct computation we know that $(\frac{1}{5}) = (\frac{4}{5}) = 1$, $(\frac{2}{5}) = -1$ and $(\frac{3}{5}) = (\frac{5}{3}) = (\frac{2}{3}) = -1$. Hence

$$\left(\frac{5}{p}\right) = \begin{cases} 1 & \text{if} \quad p \equiv 1 \text{ or } 4 \pmod 5 \\ -1 & \text{if} \quad p \equiv 2 \text{ or } 3 \pmod 5. \end{cases}$$

In other words, 5 is a quadratic residue modulo an odd prime p iff $p \equiv \pm 1 \pmod 5$.

(2) Let p be an odd prime and $p \neq 3$ (because p has to be coprime with -3). We compute $(\frac{-3}{p})$. We know $(\frac{-3}{p}) = (\frac{-1}{p})(\frac{3}{p})$. The first factor $(\frac{-1}{p}) = 1$ if $p \equiv 1 \pmod 4$ and -1 if $p \equiv 3 \pmod 4$. We apply quadratic reciprocity for the second factor; i.e. $(\frac{3}{p}) = (\frac{p}{3})$ if $p \equiv 1 \pmod 4$ and $-(\frac{p}{3})$ if $p \equiv 3 \pmod 4$. No matter whether $p \equiv 1$ or $3 \pmod 4$, we always have $(\frac{-3}{p}) = (\frac{p}{3})$. Since $(\frac{1}{3}) = 1$ and $(\frac{2}{3}) = -1$, we have

$$\left(\frac{-3}{p}\right) = \begin{cases} 1 & \text{if} \quad p \equiv 1 \pmod 3 \\ -1 & \text{if} \quad p \equiv 2 \pmod 3. \end{cases}$$

In other words, -3 is a quadratic residue modulo an odd prime p iff $p \equiv 1 \pmod 3$.

Solution 4.3. *Properties of Jacobi symbols.*

(1) Let $b = p_1 p_2 \cdots p_m$ be its prime factorisation, where $p_1, p_2, \cdots, p_m$ are not necessarily distinct. Since a is a quadratic residue modulo b, there exists some integer $x \in \mathbb{Z}$, such that $x^2 \equiv a \pmod b$. It follows that $x^2 \equiv a \pmod{p_i}$ for each $i = 1, 2, \cdots, m$. Since $\mathrm{hcf}(a, b) = 1$, we know $p_i \nmid a$, therefore a is a quadratic residue modulo p_i for each $i = 1, 2, \cdots, m$. By Definition 4.2, $(\frac{a}{p_i}) = 1$ for each i, hence by Definition 4.9, we have $(\frac{a}{b}) = (\frac{a}{p_1})(\frac{a}{p_2}) \cdots (\frac{a}{p_m}) = 1$.

(2) Let $b = p_1 p_2 \cdots p_m$ be its prime factorisation, where $p_1, p_2, \cdots, p_m$ are not necessarily distinct primes. By Definition 4.9 and Proposition 4.4 (2) we have

$$\left(\frac{a_1 a_2}{b}\right) = \left(\frac{a_1 a_2}{p_1}\right) \cdots \left(\frac{a_1 a_2}{p_m}\right) = \left(\frac{a_1}{p_1}\right)\left(\frac{a_2}{p_1}\right) \cdots \left(\frac{a_1}{p_m}\right)\left(\frac{a_2}{p_m}\right)$$
$$\left(\frac{a_1}{b}\right)\left(\frac{a_2}{b}\right) = \left(\frac{a_1}{p_1}\right) \cdots \left(\frac{a_1}{p_m}\right) \cdot \left(\frac{a_2}{p_1}\right) \cdots \left(\frac{a_2}{p_m}\right).$$

The right-hand sides of the above two equations are products of the same factors (although in different orders), Hence they are equal. It follows that the left-hand sides of these two equations are also equal.

Solution 4.4. *Quadratic residues and the Legendre symbol.*

(1) We do it in the most naive way. We could try to compute the square of all integers from 1 to 12 to get all quadratic residues modulo 13. In fact we only need to compute the first six of them, because for every $k \in \mathbb{Z}$, $1 \leqslant k \leqslant 6$, we have $13 - k \equiv -k \pmod{13}$, hence $(13 - k)^2 \equiv k^2 \pmod{13}$. In other words, the square of any integer between 7 and 12 would not produce any new congruence class. The squares of $1, 2, 3, 4, 5, 6$ are $1, 4, 9, 16, 25, 36$, which reduce to $1, 4, 9, 3, 12, 10$ modulo 13. So a is a quadratic residue modulo 13 iff $a \equiv 1, 3, 4, 9, 10$ or $12 \pmod{13}$, and a quadratic non-residue modulo 13 iff $a \equiv 2, 5, 6, 7, 8$ or $11 \pmod{13}$.

(2) Recall that a solution to such a congruence equation is a congruence class modulo p. There are three cases. If $p \mid a$, then the congruence equation becomes $x^2 \equiv 0 \pmod p$. It follows

that $p \mid x$ and $x \equiv 0 \pmod{p}$ is the only solution to the equation. In this case we do have $(\frac{a}{p}) + 1 = 1$ which is the number of solutions.

If a is a quadratic residue modulo p, then there exists some $x_0 \in \mathbb{Z}$ such that $x_0^2 \equiv a \pmod{p}$. Since $p \nmid a$, we also have $p \nmid x_0$. We claim that the congruence $x^2 \equiv a \pmod{p}$ has two solutions, which are given by $x \equiv x_0 \pmod{p}$ and $x \equiv -x_0 \pmod{p}$. Obviously both are solutions to the congruence equation. They must be distinct. Indeed, if they were the same solution, then $x_0 \equiv -x_0 \pmod{p}$, hence $2x_0 \equiv 0 \pmod{p}$. Since p is an odd prime, this implies $p \mid x_0$. Contradiction. Therefore we have found two solutions to the congruence equation $x^2 \equiv a \pmod{p}$. We can interpret this congruence as an equation $x^2 = \overline{a}$ in $\mathbb{Z}_p$. Since $\mathbb{Z}_p$ is a field by Proposition 2.9, this equation has at most two solutions by Lemma 3.3. Hence we have found all solutions. In this case, $(\frac{a}{p}) + 1 = 2$ which is indeed the number of solutions.

If a is a quadratic non-residue modulo p, then there is no solution to the congruence $x^2 \equiv a \pmod{p}$. And we do have $(\frac{a}{p}) + 1 = 0$ in this case. We proved our result in all three possible cases.

(3) We consider the congruence equations $x^2 \equiv a \pmod{p}$ for $a = 0, 1, \cdots, p-1$. There are p equations in total. The sum of numbers of solutions to these p equations is given by $\sum_{a=0}^{p-1}((\frac{a}{p}) + 1)$.

On the other hand, every congruence class modulo p is precisely a solution to one of these equations. (In other words, for every $0 \leqslant x_0 \leqslant p-1$, the congruence class $x \equiv x_0 \pmod{p}$ is a solution to the unique congruence equation $x^2 \equiv a \pmod{p}$ for a being the residue of x_0^2 modulo p.) Therefore the sum of numbers of solutions to all p congruence equations is p.

It follows that $\sum_{a=0}^{p-1}((\frac{a}{p}) + 1) = p$. The left-hand side is $\sum_{a=0}^{p-1}(\frac{a}{p}) + p$, hence we conclude that $\sum_{a=0}^{p-1}(\frac{a}{p}) = 0$.

(4) We look at the left-hand side of the equation $\sum_{a=0}^{p-1}(\frac{a}{p}) = 0$. For $a = 0$, we have $(\frac{a}{p}) = 0$. For all other values of a, $(\frac{a}{p}) = \pm 1$. Since they add up to 0, there should be the same number of 1's and -1's. In other words, in the set $\{1, 2, \cdots, p-1\}$, there are the same number of quadratic residues and non-residues.

The answer to part (1) is consistent with this conclusion, because among all positive integers less than 13, we found 6 quadratic residues modulo 13 and 6 quadratic non-residues.

5. Quadratic Reciprocity

We introduce yet another way of computing Legendre symbol due to Gauss and give a proof of the law of quadratic reciprocity.

5.1. **Gauss' lemma.** For any odd prime p and any integer a not divisible by p, Euler's criterion Proposition 4.4 (1) gives a characterisation of the Legendre symbol. Next we introduce another characterisation of the Legendre symbol due to Gauss, usually named as Gauss' lemma.

For simplicity we write $r = \frac{p-1}{2}$. We consider the set

$$S = \{-r, -(r-1), \cdots, -2, -1, 1, 2, \cdots, r-1, r\}.$$

Any integer n not divisible by p is congruent to one element in S, which is called the *least residue* of n modulo p. If $p \nmid a$, let μ be the number of integers among $a, 2a, \cdots, ra$ which have negative least residues modulo p. For example, let $p = 7$ and $a = 4$. Then $r = 3$, and the residues of $1 \cdot 4, 2 \cdot 4, 3 \cdot 4$ are $-3, 1, -2$ respectively. Thus in this case $\mu = 2$.

Gauss' lemma is the following very simple yet very powerful result:

Lemma 5.1 (Gauss' Lemma)**.** *Let p be an odd prime, $r = \frac{p-1}{2}$, $p \nmid a$, and μ the number of integers among $a, 2a, \cdots, ra$ which have negative least residues modulo p. Then $\left(\frac{a}{p}\right) = (-1)^\mu$.*

Proof. Let m_l or $-m_l$ be the least residue of la modulo p, where m_l is positive. As l ranges between 1 and r, μ is clearly the number of minus signs that occur in this way. We claim that $m_l \neq m_k$ for any $l \neq k$ and $1 \leqslant l, k \leqslant r$. For, if $m_l = m_k$, then $la \equiv \pm ka \pmod{p}$, and since $p \nmid a$ this implies that $l \pm k \equiv 0 \pmod{p}$. The latter congruence is impossible since $l \neq k$ and $|l \pm k| \leqslant |l| + |k| \leqslant p-1$. It follows that the sets $\{1, 2, \cdots, r\}$ and $\{m_1, m_2, \cdots, m_r\}$ coincide. Multiply the congruences

$$\begin{aligned} 1 \cdot a &\equiv \pm m_1 \pmod{p}, \\ 2 \cdot a &\equiv \pm m_2 \pmod{p}, \\ &\vdots, \\ r \cdot a &\equiv \pm m_r \pmod{p}. \end{aligned}$$

Notice that the number of negative signs on the right hand sides is μ, we obtain

$$r! \cdot a^r \equiv (-1)^\mu \cdot r! \pmod{p}.$$

Since $p \nmid r!$, this yields

$$a^r \equiv (-1)^\mu \pmod{p}.$$

By Euler's criterion $a^r = a^{\frac{p-1}{2}} \equiv (\frac{a}{p}) \pmod{p}$ and the result follows. □

We use Gauss' lemma to give another characterisation of the Legendre symbol, which will be used in the proof of quadratic reciprocity.

For later convenience, we introduce the so-called *floor function*. For any real number x, we define the symbol $[x]$ to be the largest integer less than or equal to x, which is sometimes also called the *integral part* of x. But pay attention when x is negative. For example, $[3] = [3.2] = 3$, $[-3] = -3$ but $[-3.2] = -4$.

If $a, b \in \mathbb{Z}$ and $b \neq 0$, we know that there is a unique way to write $a = bq + c$ for some $q, c \in \mathbb{Z}$ and $0 \leqslant c < |b|$, where q is called the *quotient* and c is called the *remainder* (or *Euclidean residue*). If we assume $b > 0$, then q is the integral part of the fraction $\frac{a}{b}$; i.e. $\left[\frac{a}{b}\right] = q$. In other words we can write $a = b\left[\frac{a}{b}\right] + c$.

Lemma 5.2. *Let p be an odd prime, a an odd integer not divisible by p. Let*

$$t = \sum_{l=1}^{\frac{p-1}{2}} \left[\frac{la}{p}\right].$$

Then $\left(\frac{a}{p}\right) = (-1)^t$.

Proof. For simplicity we write $r = \frac{p-1}{2}$. For each $l = 1, 2, \cdots, r$, we can write

$$la = p\left[\frac{la}{p}\right] + c_l,$$

where $0 \leqslant c_l \leqslant p-1$. We take the sum of the l equations and get

$$a \cdot \sum_{l=1}^{r} l = pt + \sum_{l=1}^{r} c_l. \tag{5.1}$$

Recall we wrote $\pm m_l$ for the least residue in the proof of Lemma 5.1. It is clear that

$$c_l = \begin{cases} m_l & \text{if the sign in front of } m_l \text{ is positive;} \\ -m_l + p & \text{if the sign in front of } m_l \text{ is negative.} \end{cases}$$

Modulo 2 we get

$$c_l \equiv \begin{cases} m_l \pmod 2 & \text{if the sign in front of } m_l \text{ is positive;} \\ m_l + p \pmod 2 & \text{if the sign in front of } m_l \text{ is negative.} \end{cases}$$

Now we take the sum of the l congruences and keep in mind that the negative sign in front of m_l appears exactly μ times:

$$\sum_{l=1}^{r} c_l \equiv \sum_{l=1}^{r} m_l + p\mu \pmod 2.$$

We also know that $\{m_1, m_2, \cdots, m_r\}$ is simply a permutation of $\{1, 2, \cdots, r\}$, hence

$$\sum_{l=1}^{r} c_l \equiv \sum_{l=1}^{r} l + p\mu \pmod 2. \tag{5.2}$$

Now we use (5.2) to rewrite (5.1) as

$$a \cdot \sum_{l=1}^{r} l \equiv pt + \sum_{l=1}^{r} l + p\mu \pmod 2.$$

Since a is odd, we get $pt + p\mu \equiv 0 \pmod 2$. Since p is also odd, we get $t + \mu \equiv 0 \pmod 2$; that is $t \equiv \mu \pmod 2$. By Lemma 5.1 we have

$$\left(\frac{a}{p}\right) = (-1)^{\mu} = (-1)^{t},$$

as desired. □

5.2. **A proof of quadratic reciprocity.** The law of quadratic reciprocity is so fundamentally important that many people tried to prove it in different ways. Gauss gave the first proof in 1796 and found eight separate proofs in his life. There are over a hundred now in existence. In these proofs, a lot of new techniques were developed which have become standard in modern number theory. Here we present an elementary but ingenious proof due to Gauss. It relies on a clever geometric observation which we explain now.

Lemma 5.3. *Let p and q be distinct odd primes. Then*

$$\sum_{l=1}^{\frac{p-1}{2}} \left[\frac{lq}{p}\right] + \sum_{k=1}^{\frac{q-1}{2}} \left[\frac{kp}{q}\right] = \frac{p-1}{2} \cdot \frac{q-1}{2}. \tag{5.3}$$

Proof. For simplicity we write $r = \frac{p-1}{2}$ and $s = \frac{q-1}{2}$. In the (x, y)-plane, we consider the number of integral points in the interior of the rectangle with four vertices at $(0, 0)$, $(\frac{p}{2}, 0)$, $(\frac{p}{2}, \frac{q}{2})$ and $(0, \frac{q}{2})$. Any such integral point is given by a pair of integers (x, y) with $1 \leqslant x \leqslant r$ and $1 \leqslant y \leqslant s$. Therefore the number of such integral points is rs, which is the right-hand side of (5.3).

We want to count the number of integral points in a different way to obtain the left hand side of (5.3). We connect the points $(0, 0)$ and $(\frac{p}{2}, \frac{q}{2})$ by a line segment to cut the rectangle into two triangles. We notice that there is no interior integral point lying on this line segment. Indeed, if there is an interior integral point (x, y) on this line segment, then we will have $qx = py$, which implies $p \mid x$, contradicting the requirement $1 \leqslant x \leqslant r$. Hence any integral point in the interior of the rectangle lies in the interior of one of the triangles.

We count the number of interior integral points in the triangle with vertices at $(0,0)$, $(\frac{p}{2},0)$ and $(\frac{p}{2},\frac{q}{2})$. For each $l=1,2,\cdots,r$, we fix $x=l$ and think how many integral points of the form (l,y) lie in the interior this triangle. The intersection of the vertical line $x=l$ and the diagonal of the rectangle is the point $(l,\frac{lq}{p})$. We find that y can only take positive integral values not larger than $\frac{lq}{p}$, hence has $\left[\frac{lq}{p}\right]$ choices. Therefore the number of integral points in the whole triangle is given by $\sum_{l=1}^{r}\left[\frac{lq}{p}\right]$. Similarly, the number of integral points in the other triangle is given by $\sum_{k=1}^{s}\left[\frac{kp}{q}\right]$. They add up to the left-hand side of (5.3). □

Finally we explain why the above lemmas prove the quadratic reciprocity.

Proof of the Law of Quadratic Reciprocity. For distinct odd primes p and q, by Lemma 5.2 we can write

$$\begin{aligned}\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) &= (-1)^{\sum_{l=1}^{\frac{p-1}{2}}\left[\frac{lq}{p}\right]}\cdot(-1)^{\sum_{k=1}^{\frac{q-1}{2}}\left[\frac{kp}{q}\right]}\\ &= (-1)^{\sum_{l=1}^{\frac{p-1}{2}}\left[\frac{lq}{p}\right]+\sum_{k=1}^{\frac{q-1}{2}}\left[\frac{kp}{q}\right]}\\ &= (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}},\end{aligned}$$

where the last equality follows from Lemma 5.3. □

To close this topic, we show how quadratic residues can help to give a refinement of Theorem 1.15 on infinitely many primes. The following results are special cases of the so-called Dirichlet's theorem.

Proposition 5.4. *The following statements hold*

(1) *There are infinitely many primes which are congruent to* -1 *modulo* 4.

(2) *There are infinitely many primes which are congruent to* 1 *modulo* 4.

Proof. We follow the idea in the proof of Theorem 1.15.

For (1), we assume by contradiction that the set of all primes congruent to -1 modulo 4 is finite, say, $S=\{p_1,p_2,\cdots,p_n\}$. Then we consider $N=4p_1p_2\cdots p_n-1$. Obviously $p_i\nmid N$ for each i. Let p be any prime factor of N. Then $p\notin S$ hence $p\equiv 1 \pmod 4$. This implies N is the product of primes which are all congruent to 1 modulo 4, hence $N\equiv 1 \pmod 4$. Contradiction.

For (2), we similarly assume by contradiction that the set of all primes congruent to 1 modulo 4 is finite, say, $T=\{q_1,q_2,\cdots,q_m\}$. Then we consider $M=(2q_1q_2\cdots q_m)^2+1$. Obviously $q_j\nmid M$ for each j. Let q be any prime factor of M. Then $q\notin T$ hence $q\equiv -1 \pmod 4$. However $q\mid M$ implies $(2q_1q_2\cdots q_m)^2\equiv -1 \pmod q$, i.e. -1 is a quadratic residue modulo q. By Proposition 4.5 (1) we get $q\equiv 1 \pmod 4$. Contradiction. □

Corollary 5.5. *There are infinitely many odd primes* p *for which* -1 *is a quadratic residue. There are also infinitely many odd primes* p *for which* -1 *is a quadratic non-residue.*

Proof. This is a immediate consequence of Propositions 4.5 (1) and 5.4. □

Exercise Sheet 5

This sheet is due in the lecture on Tuesday 4th November, and will be discussed in the exercise class on Friday 7th November.

Exercise 5.1. *Evaluating Legendre symbols by Gauss' lemma.*

(1) Use Gauss' lemma to determine $(\frac{5}{7})$, $(\frac{3}{11})$, $(\frac{6}{13})$.

(2) For any odd prime p, use Gauss' lemma to determine $(\frac{-1}{p})$ and $(\frac{2}{p})$.

(3) For any odd prime p, use Lemma 5.2 to determine $(\frac{-1}{p})$.

Exercise 5.2. *Special cases of Dirichlet's theorem.*

(1) Show that there are infinitely many primes which are congruent to -1 modulo 6. (Hint: follow the proof of Proposition 5.4 (1) to design the formula for N.)

(2) Show that there are infinitely many primes which are congruent to -1 modulo 8. (Hint: follow the proof of Proposition 5.4 (2) to design the formula for N. You need Proposition 4.5 (2) to analyse prime factors of N.)

Exercise 5.3. *Quadratic residues for powers of odd primes.*

Let p be an odd prime, $e > 0$ and $p \nmid a$.

(1) Assume a is a quadratic residue modulo p^{e+1}. Show that a is a quadratic residue modulo p^e.

(2) Assume a is a quadratic residue modulo p^e. Show that a is a quadratic residue modulo p^{e+1}. (Hint: if $x^2 \equiv a \pmod{p^e}$, then we can write $x^2 = a + bp^e$. Set $y = x + cp^e$ and show that we can find c such that $y^2 \equiv a \pmod{p^{e+1}}$.)

(3) Conclude by induction that a is a quadratic residue modulo p^e iff $(\frac{a}{p}) = 1$.

Exercise 5.4. *Fermat's two square problem.*

Let p be an odd prime. Recall the ring of Gaussian integers $\mathbb{Z}[i]$ from Exercise 1.4.

(1) Suppose $p \equiv 1 \pmod 4$. Show that there exist integers s and t such that $pt = s^2 + 1$. Conclude that p is not a prime in $\mathbb{Z}[i]$. (Hint: -1 is a quadratic residue modulo p; remember that $\mathbb{Z}[i]$ has unique factorisation as in Exercise 1.4 (3).)

(2) Suppose $p \equiv 1 \pmod 4$. Use part (1) to show that p is the sum of two squares; i.e. $p = a^2 + b^2$ for some $a, b \in \mathbb{Z}$. (Hint: part (1) implies $p = \alpha\beta$ for some non-units α and β in $\mathbb{Z}[i]$. Then use Exercise 1.4 (1) and (4).)

(3) Suppose $p \equiv 3 \pmod 4$. Show that p cannot be written as the sum of two squares.

Solutions to Exercise Sheet 5

Solution 5.1. *Evaluating Legendre symbols by Gauss' lemma.*

- For $(\frac{5}{7})$, since $p = 7$ and $r = 3$, we need to consider the least residues of 5, 10 and 15, which are -2, 3 and 1. There is only one negative least residue, hence $(\frac{5}{7}) = -1$.

  For $(\frac{3}{11})$, since $p = 11$ and $r = 5$, we consider the least residues of 3, 6, 9, 12 and 15, which are 3, -5, -2, 1 and 4. There are two negative least residues, hence $(\frac{3}{11}) = 1$.

  For $(\frac{6}{13})$, since $p = 13$ and $r = 6$, we consider the least residue of 6, 12, 18, 24, 30 and 36, which are 6, -1, 5, -2, 4 and -3. There are three negative ones, hence $(\frac{6}{13}) = -1$.

- We consider $(\frac{-1}{p})$. Let $r = \frac{p-1}{2}$. We need to look at the least residues of $-1, -2, \cdots, -r$. But they are already least residues themselves. Since there are r of them, by Gauss' Lemma, we get $(\frac{-1}{p}) = (-1)^r = (-1)^{\frac{p-1}{2}}$.

  Now we consider $(\frac{2}{p})$. Let $r = \frac{p-1}{2}$. We look at the least residues of $2, 4, \cdots, 2r$. We deal with four cases $p \equiv 1, 3, 5$ or $7 \pmod 8$ separately. If $p \equiv 1 \pmod 8$, then we can

assume $p = 8m+1$ for some $m \geqslant 0$, and $r = 4m$. The number $2k$ has positive least residue for $1 \leqslant k \leqslant 2m$ and negative least residue for $2m+1 \leqslant k \leqslant 4m$. Hence by Gauss' Lemma, $(\frac{2}{p}) = (-1)^{2m} = 1$. If $p \equiv 3 \pmod 8$, then we write $p = 8m+3$, and $r = 4m+1$. The number $2k$ has positive least residue for $1 \leqslant k \leqslant 2m$ and negative least residue for $2m+1 \leqslant k \leqslant 4m+1$. Hence $(\frac{2}{p}) = (-1)^{2m+1} = -1$. If $p \equiv 5 \pmod 8$, then we write $p = 8m+5$ and $r = 4m+2$. The number $2k$ has positive least residue for $1 \leqslant k \leqslant 2m+1$ and negative least residue for $2m+2 \leqslant k \leqslant 4m+2$, hence $(\frac{2}{p}) = (-1)^{2m+1} = -1$. If $p \equiv 7 \pmod 8$, then we write $p = 8m+7$ and $r = 4m+3$. The number $2k$ has positive least residue for $1 \leqslant k \leqslant 2m+1$ and negative least residue for $2m+2 \leqslant k \leqslant 4m+3$, hence $(\frac{2}{p}) = (-1)^{2m+2} = 1$. In summary, we have $(\frac{2}{p}) = 1$ if $p \equiv 1$ or $7 \pmod 8$ and -1 if $p \equiv 3$ or $5 \pmod 8$.

- Since $a = -1$, for any $1 \leqslant l \leqslant \frac{p-1}{2}$, $-1 < \frac{la}{p} < 0$, hence $[\frac{la}{p}] = -1$. Then $t = \sum_{l=1}^{\frac{p-1}{2}} [\frac{la}{p}] = \sum_{l=1}^{\frac{p-1}{2}} -1 = -\frac{p-1}{2}$. By Lemma 5.2, $(\frac{-1}{p}) = (-1)^t = (-1)^{-\frac{p-1}{2}} = (-1)^{\frac{p-1}{2}}$, where the last equality is due to the fact that n and $-n$ always have the same parity (both odd or both even) for any integer n. Or equivalently, $(\frac{-1}{p}) = 1$ if $p \equiv 1 \pmod 4$ and -1 if $p \equiv -1 \pmod 4$.

Solution 5.2. *Special cases of Dirichlet's theorem.*

(1) Assume there are only finitely many primes congruent to -1 modulo 6, say, $S = \{p_1, p_2, \cdots, p_n\}$. Then we consider $N = 6p_1p_2\cdots p_n - 1 > 1$. It is clear that $p_i \nmid N$ for each $p_i \in S$, hence $p \notin S$ for each prime factor p of N. It follows that $p \not\equiv 5 \pmod 6$. Moreover, p must be odd since N is odd, so $p \not\equiv 0, 2$ or $4 \pmod 6$. Furthermore, the only prime congruent to 3 modulo 6 is 3. However $3 \nmid N$, hence $p \not\equiv 3 \pmod 6$. Therefore the only possibility is $p \equiv 1 \pmod 6$. It follows that N is a product of primes congruent to 1 modulo 6, hence $N \equiv 1 \pmod 6$, which contradicts the formula of N, from which we can see $N \equiv 5 \pmod 6$. It follows that there are infinitely many primes congruent to -1 modulo 6.

(2) Assume there are only finitely many primes congruent to -1 modulo 8, say, $T = \{q_1, q_2, \cdots, q_m\}$. Then we consider $M = (4q_1q_2\cdots q_m)^2 - 2 > 1$. Since each $q_j \in T$ is an odd prime, $q_j \nmid 2$, hence $q_j \nmid M$. If follows that if q is an odd prime factor of M, then $q \notin T$, hence $q \not\equiv -1 \pmod 8$. On the other hand, $q \mid M$ implies that 2 is a quadratic residue modulo q, hence $q \equiv 1$ or $-1 \pmod 8$. It follows that $q \equiv 1 \pmod 8$. In other words, every odd prime factor of M is congruent to 1 modulo 8. If we write $M = 2(8q_1^2q_2^2\cdots q_m^2 - 1)$, then the second factor $8q_1^2q_2^2\cdots q_m^2 - 1$ must be a product of primes congruent to 1 modulo 8, which is itself congruent to 1 modulo 8. Contradiction. This contradiction shows that there are infinitely many primes congruent to -1 modulo 8.

Solution 5.3. *Quadratic residues for powers of odd primes.*

(1) Since a is a quadratic residue modulo p^{e+1}, there exists some $x \in \mathbb{Z}$, such that $x^2 \equiv a \pmod{p^{e+1}}$. Equivalently, $x^2 - a$ is a multiple of p^{e+1}, which implies $x^2 - a$ is a multiple of p^e. Or equivalently, $x^2 \equiv a \pmod{p^e}$. Since $p \nmid a$, we have $\gcd(a, p^e) = 1$. We conclude that a is a quadratic residue modulo p^e.

(2) Since a is a quadratic residue modulo p^e, we have $x^2 \equiv a \pmod{p^e}$ for some $x \in \mathbb{Z}$. Equivalently, we can write $x^2 = a + bp^e$ for some $b \in \mathbb{Z}$. Set $y = x + cp^e$ for some $c \in \mathbb{Z}$, then we consider $y^2 - a$. We have $y^2 - a = (x + cp^e)^2 - a = x^2 - a + 2xcp^e + c^2p^{2e} = (b + 2xc)p^e + c^2p^{2e}$.

Now we claim that we can choose c such that $b + 2xc$ is a multiple of p. Indeed, since $p \nmid a$, we have $p \nmid x$, hence $\gcd(2x, p) = 1$. It follows by Proposition 2.5 that the congruence equation $2xz \equiv -b \pmod p$ (think of it as an equation of z) has a solution for z. Let $z = c$ be such a solution, then $2xc + b$ is a multiple of p, hence $(b + 2xc)p^e$ is a multiple of p^{e+1}. On

the other hand c^2p^{2e} is also a multiple of p^{e+1} because $2e \geqslant e+1$. It follows that $y^2 - a$ is a multiple of p^{e+1}, or equivalently, $y^2 \equiv a \pmod{p^{e+1}}$. Since $p \nmid a$, we have $\mathrm{hcf}(a, p^{e+1}) = 1$. Therefore a is a quadratic residue modulo p^{e+1}.

(3) By parts (1) and (2), a is a quadratic residue modulo p^e iff a is a quadratic residue modulo p^{e+1}. Using this result inductively, we can conclude that a is a quadratic residue modulo p^e for any positive integer e iff p is a quadratic residue modulo p, which is equivalent to $(\frac{a}{p}) = 1$.

Solution 5.4. *Fermat's two-square problem.*

(1) Since $p \equiv 1 \pmod 4$, -1 is a quadratic residue modulo p. In other words, $x^2 \equiv -1 \pmod p$ has a solution. Let $x = s$ be one such solution, then $s^2 + 1$ is a multiple of p. We can then write $s^2 + 1 = pt$, where $s, t \in \mathbb{Z}$. It follows that p divides $s^2 + 1 = (s + i)(s - i)$ in $\mathbb{Z}[i]$. If p could divide $s + i$ in $\mathbb{Z}[i]$, then we can write $s + i = p(x + yi)$ for some $x, y \in \mathbb{Z}$. It follows that $py = 1$. Contradiction. Therfore p does not divide $s + i$. Similar one can show that p does not divide $s - i$. Hence p is not a prime, because p divides the product of $s + i$ and $s - i$ but neither of the factors.

(2) We know from Exercise 1.4 (2) that $\mathbb{Z}[i]$ is a Euclidean domain, hence a PID. By Proposition 1.9 (2), every irreducible element in $\mathbb{Z}[i]$ is a prime. By part (1), p is not a prime in $\mathbb{Z}[i]$ hence is not irreducible. It follows that we can write $p = \alpha\beta$, such that α and β are non-units. We apply Exercise 1.4 (1) and get $\nu(p) = \nu(\alpha)\nu(\beta)$. By the formula of the valuation ν, the left-hand side is p^2. By Exercise 1.4 (4), neither of the factor on the right-hand side is 1. Therefore the only possibility is $\nu(\alpha) = \nu(\beta) = p$. Let $\alpha = a + bi$ for some $a, b \in \mathbb{Z}$. Then $\nu(\alpha) = a^2 + b^2 = p$.

(3) We show that $a^2 \equiv 0$ or $1 \pmod 4$ for every $a \in \mathbb{Z}$. Indeed, if a is even, say $a = 2k$ for some $k \in \mathbb{Z}$, then $a^2 = 4k^2 \equiv 0 \pmod 4$. If a is odd, say $a = 2k + 1$ for some $k \in \mathbb{Z}$, then $a^2 = (2k + 1)^2 = 4k^2 + 4k + 1 \equiv 1 \pmod 4$. The same is true for b^2. We consider all the combinations and conclude that $a^2 + b^2 \equiv 0$ or 1 or $2 \pmod 4$. By assumption $p \equiv 3 \pmod 4$, hence $p = a^2 + b^2$ is never possible.

6. Number Fields

So far we have mainly focused on the ring of integers $\mathbb{Z}$. But modern number theory is not only about integers. From this point on we will enlarge our vision and study the so-called algebraic integers.

6.1. Algebraic numbers and algebraic integers.

We first introduce the following terminologies, which will be convenient for our discussions:

Definition 6.1. An *algebraic number* is a complex number that is a root of a non-zero polynomial $f(x)$ with coefficients in $\mathbb{Q}$. An *algebraic integer* is a complex number that is a root of a non-zero monic (leading coefficient 1) polynomial $f(x)$ with coefficients in $\mathbb{Z}$.

Example 6.2. For example, $\sqrt{2}$ is an algebraic integer because it is a root of the polynomial $x^2 - 2$; $i = \sqrt{-1}$ is also an algebraic integer because it is a root of $x^2 + 1$; so is a fifth root of unity $\cos\frac{2\pi}{5} + i\sin\frac{2\pi}{5}$ because it is a root of $x^5 - 1$. They are also algebraic numbers.

Remark 6.3. We make the following remarks about this definition.

(1) It is clear that every $a \in \mathbb{Z}$ is an algebraic integer. To avoid any potential confusion, we sometimes call any element $a \in \mathbb{Z}$ a *rational integer* (especially when both notions appear in the same sentence).

(2) Clearly every algebraic integer is an algebraic number. But the converse is not true; see Exercise 6.1.

(3) There are complex numbers which are not algebraic numbers. They are usually called *transcendental* numbers. Typical examples include the ratio of the circumference and diameter of a circle $\pi = 3.14159\ldots$, and the base of the natural logarithm $e = 2.71828\ldots$. We will not explain why they are not algebraic, but it is a standard topic in transcendental number theory.

Example 6.4. A slightly more complicated example is $\sqrt{2} + \sqrt{3}i$. We show it is an algebraic integer by definition. Let $x = \sqrt{2} + \sqrt{3}i$. We rewrite it as $x - \sqrt{2} = \sqrt{3}i$ and square both sides to get $x^2 - 2\sqrt{2}x + 2 = -3$. We rewrite it as $x^2 + 5 = 2\sqrt{2}x$ and square both sides again to get $x^4 + 10x^2 + 25 = 8x^2$. Hence $x^4 + 2x^2 + 25$ is a monic polynomial in $\mathbb{Z}[x]$ for which $\sqrt{2} + \sqrt{3}i$ is a root.

We want to have a more straightforward way to understand the above example. Given two algebraic integers, we ask whether their sum and product are still algebraic integers. We will see the answer is yes. Before proving it we establish the following criterion:

Lemma 6.5. *Let $V = \{\gamma_1, \gamma_2, \cdots, \gamma_n\}$ be a finite set of non-zero complex numbers. Suppose a complex number α has the property that for each $i = 1, 2, \cdots, n$, the product $\alpha\gamma_i$ can be written as an integral linear combination of elements in the set V. Then α is an algebraic integer.*

Proof. By assumption, for each $i = 1, 2, \cdots, n$, we can write

$$\alpha\gamma_i = \sum_{j=1}^{n} a_{ij}\gamma_j,$$

where each $a_{ij} \in \mathbb{Z}$.

Using the language of linear algebra, we have

$$\alpha \cdot \mathbf{v} = \mathbf{M} \cdot \mathbf{v},$$

where

$$\mathbf{M} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{pmatrix}.$$

Since $\mathbf{v} \neq 0$, we see that α is an eigenvalue of the square matrix $\mathbf{M}$. In other words, α is a solution of the equation

$$\det(x \cdot \mathbf{I} - \mathbf{M}) = 0.$$

Since all entries of $\mathbf{M}$ are integers, it is clear that the left-hand side of the equation is a polynomial with integer coefficients, whose leading term is x^n. Therefore α is an algebraic integer, as desired. □

Now we can prove the following important result:

Proposition 6.6. *The sum and product of two algebraic integers are algebraic integers.*

Proof. Suppose α and β are algebraic integers. If either $\alpha = 0$ or $\beta = 0$, the statement is clear. From now on we assume $\alpha \neq 0$ and $\beta \neq 0$. We want to apply Lemma 6.5 to show that $\alpha + \beta$ and $\alpha\beta$ are also algebraic integers. Suppose α and β satisfy

$$\alpha^n + a_1\alpha^{n-1} + a_2\alpha^{n-2} + \cdots + a_{n-1}\alpha + a_n = 0$$
$$\beta^m + b_1\beta^{m-1} + b_2\beta^{m-2} + \cdots + b_{m-1}\beta + b_m = 0,$$

where each a_i and b_j are integers. Let

$$V = \{\alpha^i\beta^j \mid 0 \leqslant i < n, 0 \leqslant j < m\}.$$

For each element $\alpha^i\beta^j \in V$, we claim that $(\alpha+\beta)\cdot\alpha^i\beta^j$ and $\alpha\beta\cdot\alpha^i\beta^j$ can both be written as integral linear combinations of elements in V. Indeed, we have

$$(\alpha+\beta)\cdot\alpha^i\beta^j = \alpha^{i+1}\beta^j + \alpha^i\beta^{j+1} \tag{6.1}$$

$$\alpha\beta\cdot\alpha^i\beta^j = \alpha^{i+1}\beta^{j+1}. \tag{6.2}$$

If $0 \leqslant i \leqslant n-2$ and $0 \leqslant j \leqslant m-2$, then our claim is already true. Otherwise, if $i = n-1$ and/or $j = m-1$, we can replace α^n by $-(a_1\alpha^{n-1} + a_2\alpha^{n-2} + \cdots + a_{n-1}\alpha + a_n)$ and/or β^m by $-(b_1\beta^{m-1} + b_2\beta^{m-2} + \cdots + b_{m-1}\beta + b_m)$ in the right-hand sides of (6.1) and (6.2), then their expansions are still integral linear combinations of elements of V. Therefore our claim is true. By Lemma 6.5, we conclude that $\alpha+\beta$ and $\alpha\beta$ are both algebraic integers. □

Corollary 6.7. *The set of all algebraic integers forms a commutative ring with* 1.

Proof. We have to check that the addition, the additive inverse and the multiplication are all well-defined in the set of algebraic integers, and all algebraic laws required in the definition of a ring hold in this set.

Proposition 6.6 proved that the addition and the multiplication are both well-defined. The existence of additive inverse is given by Exercise 6.1. All algebraic laws related concerning the addition and the multiplication hold because they hold for complex numbers. Hence the set of algebraic integers forms a ring. □

6.2. **Number fields.** We introduce the notion of number fields as follows:

Definition 6.8. An *(algebraic) number field* is a field K, such that $\mathbb{Q} \subseteq K \subseteq \mathbb{C}$, and K has finite degree (dimension as a vector space) over $\mathbb{Q}$.

Example 6.9. Simple example: the field $\mathbb{Q}$ itself is a number field of degree 1 over $\mathbb{Q}$.

We recall a useful result in Algebra 2B which gives a lot of examples of number fields.

Proposition 6.10. *Let $\Bbbk \subseteq K$ be a field extension, and let $\alpha \in K$ be a root of some non-zero polynomial $g(x) \in \Bbbk[x]$. Then the set $\{f(\alpha) \in K \mid f \in \Bbbk[x]\}$ is a field, denoted by $\Bbbk[\alpha]$ or $\Bbbk(\alpha)$, satisfying $\Bbbk \subseteq \Bbbk(\alpha) \subseteq K$.*

Moreover, assume $g(x)$ is irreducible and $\deg g(x) = n$, then $\Bbbk(\alpha)$ has degree n over $\Bbbk$ and $\{1, \alpha, \alpha^2, \cdots, \alpha^{n-1}\}$ is a basis of $\Bbbk(\alpha)$ over $\Bbbk$.

Proof. See Proposition 2.23 (2013) or Theorem 4.8 (2014) in Algebra 2B. □

Remark 6.11. We point out two things.

(1) In Algebra 2B, we used the notation $\Bbbk[\alpha]$. But in literature (especially in literature on field theory) the notation $\Bbbk(\alpha)$ seems to be used more often. We will use the latter.

(2) Roughly speaking, if an element α in the large field is the root of a polynomial with coefficients in the small field, then we can "add" α to the small field to generate an intermediate field, which has a finite degree over the small field, with a basis given by powers of α. If the small and large fields are $\mathbb{Q}$ and $\mathbb{C}$ respectively, we can get lots of examples of number fields.

Example 6.12. In Proposition 6.10, we take $\Bbbk = \mathbb{Q}$ and $K = \mathbb{C}$.

(1) For any square-free integer $d \neq 1$, $\sqrt{d}$ is a root of the irreducible polynomial $x^2 - d \in \mathbb{Q}[x]$. Therefore $\mathbb{Q}(\sqrt{d}) = \{a + b\sqrt{d} \mid a, b \in \mathbb{Q}\}$ is number field of degree 2 over $\mathbb{Q}$. A number field of this form is called a *quadratic field*. It is called a *real quadratic field* if $d > 0$, or an *imaginary quadratic field* if $d < 0$. For instance, $\mathbb{Q}(\sqrt{2})$ is a real quadratic fields and $\mathbb{Q}(i)$ is an imaginary quadratic field.

(2) We have that $\sqrt[3]{2}$ is a root of the irreducible polynomial $x^3 - 2 \in \mathbb{Q}[x]$. Therefore $\mathbb{Q}(\sqrt[3]{2}) = \{a + b\sqrt[3]{2} + c\sqrt[3]{4} \mid a, b, c \in \mathbb{Q}\}$ is a number field of degree 3 over $\mathbb{Q}$. This is an example of the so-called *cubic field*.

(3) We have that $\zeta = \cos\frac{2\pi}{5} + i\sin\frac{2\pi}{5}$ is the root of an irreducible polynomial $x^4+x^3+x^2+x+1$. Therefore $\mathbb{Q}(\zeta)$ is a number field of degree 4 over $\mathbb{Q}$. This is an example of the so-called *cyclotomic field*.

The following lemma justifies the name.

Lemma 6.13. *Every element in a number field is an algebraic number.*

Proof. Let K be a number field of degree n over $\mathbb{Q}$ and $\alpha \in K$. Then $1, \alpha, \alpha^2, \cdots, \alpha^n$ must be linearly dependent over $\mathbb{Q}$; i.e. $a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_n\alpha^n = 0$, where $a_i \in \mathbb{Q}$ for each i and all a_i's are not simultaneously zero. This implies α is a root of a polynomial with rational coefficients, hence is an algebraic number. □

We then introduce the notions of traces and norms.

Definition 6.14. Let K be a number field. Every $\alpha \in K$ defines a $\mathbb{Q}$-linear transformation

$$L_\alpha : K \to K, \quad \gamma \mapsto \alpha\gamma.$$

The trace of the linear transformation L_α is called the *trace* of α in K, denoted by $T_K(\alpha)$. The determinant of the linear transformation L_α is called the *norm* of α in K, denoted by $N_K(\alpha)$.

Remark 6.15. We make the following comments about this definition.

(1) The $\mathbb{Q}$-linearity of L_α can be easily checked by observing $\alpha(\gamma_1 + \gamma_2) = \alpha\gamma_1 + \alpha\gamma_2$ for any $\gamma_1, \gamma_2 \in K$, and $\alpha(\lambda\gamma) = \lambda(\alpha\gamma)$ for any $\gamma \in K$ and $\lambda \in \mathbb{Q}$.

(2) The trace and norm depends on both K and α. The same algebraic number α, when considered as an element of different number fields, could have different traces and norms. If there is only one number field K in consideration, we often omit the reference to K and write $T(\alpha)$ and $N(\alpha)$ for simplicity.

(3) In practice we can choose any $\mathbb{Q}$-basis of K and write the linear transformation L_α as a matrix to compute $T(\alpha)$ and $N(\alpha)$. We know that the trace and determinant of a linear transformation are independent of the choice of the basis, but choosing the basis wisely can make the computation easier.

The following properties can be easily proved using the language of linear transformations and matrices.

Lemma 6.16. *Let K be a number field of degree n over $\mathbb{Q}$, $\alpha, \beta \in K$ and $a \in \mathbb{Q}$. Then*

(1) $T(\alpha + \beta) = T(\alpha) + T(\beta)$, $N(\alpha\beta) = N(\alpha)N(\beta)$;

(2) $T(a\alpha) = aT(\alpha)$, $N(a\beta) = a^n N(\beta)$;

(3) $T(1) = n$, $N(1) = 1$;

(4) $N(\alpha) = 0$ *iff* $\alpha = 0$.

Proof. We leave them as exercises. See Exercise 6.3. □

We show two examples of computation of traces and norms.

Example 6.17. Consider the number field $K = \mathbb{Q}$. For any $\alpha \in K$, we compute its trace and norm. We choose a $\mathbb{Q}$-basis $\{1\}$ for K, then the matrix of L_α under this basis is a 1×1 matrix with the only entry α. Hence $T(\alpha) = \alpha$ and $N(\alpha) = \alpha$.

Example 6.18. Consider the quadratic field $K = \mathbb{Q}(\sqrt{d})$ where $d \neq 1$ is a square-free integer. For any $\alpha = a + b\sqrt{d} \in K$, we compute its trace and norm. We choose a $\mathbb{Q}$-basis $\{1, \sqrt{d}\}$ for K. Since $L_\alpha(1) = a + b\sqrt{d}$ and $L_\alpha(\sqrt{d}) = bd + a\sqrt{d}$, the matrix of L_α under this basis is $\begin{pmatrix} a & bd \\ b & a \end{pmatrix}$. Therefore $T(\alpha) = 2a$ and $N(\alpha) = a^2 - b^2 d$.

A crucial property of the trace and the norm is the following:

Proposition 6.19. *Let K be a number field and α an algebraic integer in K, then $T(\alpha), N(\alpha) \in \mathbb{Z}$.*

Sketch of proof. The proof of this result will be left in Exercise 6.4. Here we explain briefly the motivation and main idea in the proof and give some hints step by step.

By Definition 6.14, if we can find a $\mathbb{Q}$-basis for K, under which the matrix of the linear transformation L_α has integral entries, then $T(\alpha)$ and $N(\alpha)$ are integers. Therefore the proof contains two steps: find a $\mathbb{Q}$-basis for K; show that the matrix of L_α under this basis has integer entries.

More precisely, we consider an intermediate field $\mathbb{Q} \subseteq \mathbb{Q}(\alpha) \subseteq K$ as in Proposition 6.10. Then for some $m > 0$, we know $\{1, \alpha, \alpha^2, \cdots, \alpha^{m-1}\}$ is a basis of $\mathbb{Q}(\alpha)$ over $\mathbb{Q}$. On the other hand, we choose any basis of K over $\mathbb{Q}(\alpha)$, say $\{\beta_0, \beta_1, \cdots, \beta_{n-1}\}$. We can prove that the set

$$S = \{\alpha^i \beta_j \mid 0 \leqslant i \leqslant m-1, 0 \leqslant j \leqslant n-1\}$$

is a basis of K over $\mathbb{Q}$. For this purpose, we need to show that S is a spanning set and elements in S are independent. Then we write down the matrix for L_α under this basis and conclude all entries are integers. □

Exercise Sheet 6

This sheet is due in the lecture on Tuesday 11th November, and will be discussed in the exercise class on Friday 14th November.

Exercise 6.1. *Examples of algebraic integers.*

(1) Show that $\frac{1}{2}(1 + \sqrt{5})$ is an algebraic integer by definition; i.e. by writing down a monic polynomial in $\mathbb{Z}[x]$ for which it is a root. Do the same for $3 + i$ and $\sqrt{2} + \sqrt[3]{3}$.

(2) Show that $\frac{1}{2}$ is an algebraic number but not an algebraic integer by definition.

(3) Suppose that α is an algebraic integer. Show that $-\alpha$ is also an algebraic integer.

Exercise 6.2. *Examples of traces and norms.*

(1) Let K be the cubic field $\mathbb{Q}(\sqrt[3]{2})$. For any $\alpha = a + b\sqrt[3]{2} + c\sqrt[3]{4} \in K$ with $a, b, c \in \mathbb{Q}$, write down the matrix for the linear transformation L_α under the basis $\{1, \sqrt[3]{2}, \sqrt[3]{4}\}$. Compute the trace and norm of α in K.

(2) Let K be the cyclotomic field $\mathbb{Q}(\zeta)$ where $\zeta = \cos\frac{2\pi}{5} + i \sin\frac{2\pi}{5}$. Write down the matrix for the linear transformation L_ζ under the basis $\{1, \zeta, \zeta^2, \zeta^3\}$. Compute the trace and norm of ζ in K.

Exercise 6.3. *Elementary properties of the trace and norm.*

Let K be a number field of degree n over $\mathbb{Q}$, $\alpha, \beta \in K$ and $a \in \mathbb{Q}$. Prove the following

(1) $T(\alpha+\beta) = T(\alpha) + T(\beta)$, $N(\alpha\beta) = N(\alpha)N(\beta)$;
(2) $T(a\alpha) = aT(\alpha)$, $N(a\beta) = a^n N(\beta)$;
(3) $T(1) = n$, $N(1) = 1$;
(4) $N(\alpha) = 0$ iff $\alpha = 0$.

Exercise 6.4. *Traces and norms of algebraic integers.*

Supply the details in the proof of Proposition 6.19 in the following steps. The set S is defined in the sketch of proof in the lecture notes.

(1) Show that S spans K over $\mathbb{Q}$, i.e. every element in K is a $\mathbb{Q}$-linear combination of elements in S with rational coefficients;
(2) Show that elements in S are linearly independent over $\mathbb{Q}$;
(3) Write down the matrix for L_α under the basis S. Conclude that all entries are in $\mathbb{Z}$, and $T(\alpha), N(\alpha) \in \mathbb{Z}$.

Solutions to Exercise Sheet 6

Solution 6.1. *Examples of algebraic integers.*

(1) $\frac{1}{2}(1+\sqrt{5})$ is an algebraic integer because it is a root of the polynomial $x^2 - x - 1$. For $3+i$, we let $x = 3+i$, rewrite it as $x - 3 = i$ and square both sides to get $x^2 - 6x + 9 = -1$, hence $3+i$ is the root of the polynomial $x^2 - 6x + 10$.

For $\sqrt{2} + \sqrt[3]{3}$, we let $x = \sqrt{2} + \sqrt[3]{3}$, rewrite it as $x - \sqrt{2} = \sqrt[3]{3}$, take the third powers to get $x^3 - 3\sqrt{2}x^2 + 6x - 2\sqrt{2} = 3$. We rewrite it as $x^3 + 6x - 3 = (3x^2+2)\sqrt{2}$ and square both sides to get $(x^3 + 6x - 3)^2 = 2(3x^2 + 2)^2$. Then we conclude that $\sqrt{2} + \sqrt[3]{3}$ is the root of the polynomial $(x^3 + 6x - 3)^2 - 2(3x^2 + 2)^2 = x^6 - 6x^4 - 6x^3 + 12x^2 - 36x + 1$. Notice that all coefficients are integers, and the leading term x^6 has coefficient 1. This shows $\sqrt{2} + \sqrt[3]{3}$ is an algebraic integer.

(2) $\frac{1}{2}$ is an algebraic number because it is the root of $2x - 1$. We show it is not an algebraic integer by contradiction. Assume it is the root of a monic polynomial

$$x^n + a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_{n-2}x^2 + a_{n-1}x + a_n.$$

By substitution $x = \frac{1}{2}$ we have

$$\frac{1}{2^n} + \frac{a_1}{2^{n-1}} + \frac{a_2}{2^{n-2}} + \cdots + \frac{a_{n-2}}{2^2} + \frac{a_{n-1}}{2} + a_n = 0.$$

Now we multiply 2^n on both sides to clear the denominators and obtain

$$1 + 2a_1 + 2^2 a_2 + \cdots + 2^{n-2}a_{n-2} + 2^{n-1}a_{n-1} + 2^n a_n = 0.$$

The left-hand side is an odd number. Contradiction. Therefore $\frac{1}{2}$ is not an algebraic integer.

(3) Since α is an algebraic integer, it is a root of a polynomial $f(x) = x^n + a_1 x^{n-1} + a_2 x^{n-2} + a_3 x^{n-3} + \cdots + a_{n-1}x + a_n \in \mathbb{Z}[x]$. We consider the polynomial $g(x) = x^n - a_1 x^{n-1} + a_2 x^{n-2} - a_3 x^{n-3} + \cdots + (-1)^{n-1} a_{n-1} x + (-1)^n a_n$, which is a monic polynomial with integer

coefficients. We claim that $-\alpha$ is a root of $g(x)$. Indeed, we have

$$\begin{aligned} g(-\alpha) &= (-\alpha)^n - a_1(-\alpha)^{n-1} + a_2(-\alpha)^{n-2} - a_3(-\alpha)^{n-3} + \cdots \\ &\qquad \cdots + (-1)^{n-1}a_{n-1}(-\alpha) + (-1)^n a_n \\ &= (-1)^n(\alpha^n + a_1\alpha^{n-1} + a_2\alpha^{n-2} + a_3\alpha^{n-3} + \cdots + a_{n-1}\alpha + a_n) \\ &= 0. \end{aligned}$$

Hence $-\alpha$ is an algebraic integer.

The following is another proof. I would like to thank people who provided this much better proof in their submitted solutions.

Since both α and -1 are algebraic integers, and the product of two algebraic integers is still an algebraic integer, we immediately know $-\alpha$ is an algebraic integer.

Solution 6.2. *Examples of traces and norms.*

(1) We have $L_\alpha(1) = a + b\sqrt[3]{2} + c\sqrt[3]{4}$, $L_\alpha(\sqrt[3]{2}) = 2c + a\sqrt[3]{2} + b\sqrt[3]{4}$, $L_\alpha(\sqrt[3]{4}) = 2b + 2c\sqrt[3]{2} + a\sqrt[3]{4}$. We write the coefficients as column vectors and get the matrix

$$M = \begin{pmatrix} a & 2c & 2b \\ b & a & 2c \\ c & b & a \end{pmatrix}.$$

Therefore we have $T(\alpha) = \operatorname{tr}(M) = 3a$ and $N(\alpha) = \det(M) = a^3 + 2b^3 + 4c^3 - 6abc$.

(2) We have $L_\zeta(1) = \zeta$, $L_\zeta(\zeta) = \zeta^2$, $L_\zeta(\zeta^2) = \zeta^3$, $L_\zeta(\zeta^3) = \zeta^4 = -\zeta^3 - \zeta^2 - \zeta - 1$. Hence the matrix is

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

Therefore we have $T(\zeta) = \operatorname{tr}(M) = -1$ and $N(\zeta) = \det(M) = 1$.

Solution 6.3. *Elementary properties of the trace and norm.*

(1) For any $\gamma \in K$, $L_{\alpha+\beta}(\gamma) = (\alpha+\beta)\gamma = \alpha\gamma + \beta\gamma = L_\alpha(\gamma) + L_\beta(\gamma)$. Hence the linear transformation $L_{\alpha+\beta}$ is the sum of the two linear transformations L_α and L_β. Under any fixed basis, if the matrices for L_α and L_β are A and B respectively, then their sum $L_{\alpha+\beta}$ corresponds to the matrix $A + B$. Since we have $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$, we get $T(\alpha + \beta) = T(\alpha) + T(\beta)$.

For any $\gamma \in K$, $L_{\alpha\beta}(\gamma) = (\alpha\beta)\gamma = \alpha(\beta\gamma) = L_\alpha(L_\beta(\gamma))$. Hence the linear transformation $L_{\alpha\beta}$ is the composition of the two linear transformations L_α and L_β. Under any fixed basis, if the matrices for L_α and L_β are A and B respectively, then their composition $L_{\alpha\beta}$ corresponds to the matrix AB. Since we have $\det(AB) = \det(A)\det(B)$, we get $N(\alpha\beta) = N(\alpha)N(\beta)$.

(2) For any $\gamma \in K$, $L_{a\alpha}(\gamma) = (a\alpha)\gamma = a(\alpha\gamma) = aL_\alpha(\gamma)$. Hence the linear transformation $L_{a\alpha}$ is the linear transformation $a \cdot L_\alpha$. Under any fixed basis, if the matrices for L_α is A, then the matrix corresponds to $L_{a\alpha}$ is aA. Since we have $\operatorname{tr}(aA) = a\operatorname{tr}(A)$, we get $T(a\alpha) = aT(\alpha)$. Similarly, since we have $\det(aA) = a^n \det(A)$ as A is an $n \times n$ matrix, we get $N(a\alpha) = a^n N(\alpha)$.

(3) For any $\gamma \in K$, $L_1(\gamma) = \gamma$. Hence the linear transformation L_1 is the identity map. Under any basis, its matrix is the $n \times n$ identity matrix I_n. Therefore $T(1) = \operatorname{tr}(I_n) = n$ and $N(1) = \det(I_n) = 1$.

(4) If $\alpha = 0$, then L_α is the zero linear transformation, hence $N(0) = 0$. Now we prove the other direction. We assume that $N(\alpha) = 0$ for some $\alpha \in K$. Under a fixed basis, we assume the matrix for L_α is A. Then $\det(A) = 0$, which means that A has a non-trivial null space. In other words, there is a non-zero vector $\mathbf{v}$ such that $A\mathbf{v} = 0$. But $\mathbf{v}$ is the vector form of some non-zero element $\gamma \in K$. Hence we have $L_\alpha(\gamma) = 0$. In other words, $\alpha\gamma = 0$. Since $\gamma \neq 0$, we must have $\alpha = 0$.

Solution 6.4. *Traces and norms of algebraic integers.*

(1) We first check S is a spanning set. For any $x \in K$, since $\{\beta_j \mid 0 \leqslant j \leqslant n-1\}$ is a spanning set for K over $\mathbb{Q}(\alpha)$, there exist $a_j \in \mathbb{Q}(\alpha)$ for $0 \leqslant j \leqslant n-1$ such that

$$x = \sum_{j=0}^{n-1} a_j \beta_j.$$

Since $\{\alpha^i \mid 0 \leqslant i \leqslant m-1\}$ is a spanning set for $\mathbb{Q}(\alpha)$ over $\mathbb{Q}$, for every j there exists $b_{ij} \in \mathbb{Q}$ for $0 \leqslant i \leqslant m-1$ such that

$$a_j = \sum_{i=0}^{m-1} b_{ij} \alpha^i.$$

Therefore we have

$$x = \sum_{j=0}^{n-1} \sum_{i=0}^{m-1} b_{ij} \alpha^i \beta_j,$$

which implies that S is a spanning set for K over $\mathbb{Q}$.

(2) We then check elements in S are independent over $\mathbb{Q}$. Assume we have

$$\sum_{j=0}^{n-1} \sum_{i=0}^{m-1} b_{ij} \alpha^i \beta_j = 0$$

for some $b_{ij} \in \mathbb{Q}$. We can group the terms as

$$\sum_{j=0}^{n-1} \left(\sum_{i=0}^{m-1} b_{ij} \alpha^i \right) \beta_j = 0.$$

Since $\sum_{i=0}^{m-1} b_{ij} \alpha^i \in \mathbb{Q}(\alpha)$ for each j, and $\{\beta_j\}$ are independent over $\mathbb{Q}(\alpha)$, we conclude that

$$\sum_{i=0}^{m-1} b_{ij} \alpha^i = 0$$

for each j. Moreover by the linear independence of $\{\alpha^i\}$, we conclude that

$$b_{ij} = 0$$

for every pair (i, j), which implies that elements in S are independent over $\mathbb{Q}$.

(3) Now we compute the matrix of L_α under the basis S for K over $\mathbb{Q}$. We assume that α is a root of a monic irreducible polynomial $g(x) \in \mathbb{Z}[x]$ of degree m, and we write

$$g(x) = x^m + c_1 x^{m-1} + \cdots + c_{m-1} x + c_m$$

where $c_1, \cdots, c_l \in \mathbb{Z}$. For every pair of (i, j), we have

$$L_\alpha(\alpha^i \beta_j) = \begin{cases} \alpha^{i+1} \beta_j & \text{if } 0 \leqslant i \leqslant l-2 \\ \alpha^l \beta_j = -c_1 \alpha^{l-1} \beta_j - \cdots - c_{l-1} \alpha \beta_j - c_n \beta_j & \text{if } i = l-1. \end{cases}$$

We observe that all coefficients are integers, hence the matrix M associated to the linear transformation L_α under the basis S is a matrix with integer entries. It follows that $T(\alpha)$ and $N(\alpha)$, as the trace and determinant of M, are also integers.

More precisely, the matrix M can be written in the following block diagonal form

$$M = \begin{pmatrix} D & & & \\ & D & & \\ & & \ddots & \\ & & & D \end{pmatrix},$$

where each block along the diagonal is given by

$$D = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -c_m \\ 1 & 0 & 0 & \cdots & 0 & -c_{m-1} \\ 0 & 1 & 0 & \cdots & 0 & -c_{m-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -c_2 \\ 0 & 0 & 0 & \cdots & 1 & -c_1 \end{pmatrix}.$$

Hence $T(\alpha) = \mathrm{tr}(M) = -nc_1 \in \mathbb{Z}$ and $N(\alpha) = \det(M) = ((-1)^m c_m)^n \in \mathbb{Z}$.

7. The Ring of Integers in a Number Field

We introduce the ring of integers $\mathcal{O}_K$ in a number field K and determine the additive structure of $\mathcal{O}_K$.

7.1. The ring of integers.

We first introduce the central object that we will study.

Let K be a number field. We consider the set of all algebraic integers in K. By Corollary 6.7 and the fact that K is a field, this set is closed under addition, multiplication and inverse, hence is a subring of the ring of all algebraic integers. This ring is called the *ring of (algebraic) integers* in K, denote by $\mathcal{O}_K$. The remaining part of this course will be devoted to study various properties of this ring.

The first obvious question, is to understand the elements in $\mathcal{O}_K$. We study this question in two concrete examples.

Proposition 7.1. *A rational number $\alpha \in \mathbb{Q}$ is an algebraic integer iff $\alpha \in \mathbb{Z}$.*

Proof. If $\alpha \in \mathbb{Z}$, it is clearly an algebraic integer. For the other direction, if α is an algebraic integer, by Proposition 6.19, we have $T(\alpha) \in \mathbb{Z}$ and $N(\alpha) \in \mathbb{Z}$. By Example 6.17, in this case $T(\alpha) = N(\alpha) = \alpha$, hence $\alpha \in \mathbb{Z}$. □

Proposition 7.2. *Let $d \neq 1$ be a square-free integer and $K = \mathbb{Q}(\sqrt{d})$ the corresponding quadratic field. The elements in the ring of integers $\mathcal{O}_K$ is given by $\{a + b\omega \mid a, b \in \mathbb{Z}\}$, where*

$$\omega = \begin{cases} \sqrt{d} & \text{if } d \equiv 2 \text{ or } 3 \pmod 4; \\ \frac{1}{2}(1 + \sqrt{d}) & \text{if } d \equiv 1 \pmod 4. \end{cases}$$

Proof. We first show that for any $a, b \in \mathbb{Z}$, $a + b\omega$ is an algebraic number. By Proposition 6.6, it suffices to show ω is an algebraic integer. If $d \equiv 2$ or $3 \pmod 4$, ω is a root of $x^2 - d$ hence is an algebraic integer. If $d \equiv 1 \pmod 4$, ω is a root of $x^2 - x - \frac{d-1}{4}$ hence is also an algebraic integer.

It remains to show that every algebraic integer in K has the given form. Let $\alpha = r + s\sqrt{d}$ is an algebraic integer for some $r, s \in \mathbb{Q}$. By Example 6.17 and Proposition 6.19, we know $T(r + s\sqrt{d}) = 2r \in \mathbb{Z}$ and $N(r + s\sqrt{d}) = r^2 - s^2 d \in \mathbb{Z}$. Thus $(2r)^2 - (2s)^2 d \in 4\mathbb{Z}$ and $(2s)^2 d \in \mathbb{Z}$. Since d is square-free, this implies $2s \in \mathbb{Z}$.

Now we consider the case $d \equiv 2$ or $3 \pmod 4$. If both $2r$ and $2s$ are odd, then $(2r)^2 \equiv 1 \pmod 4$ and $(2s)^2 d \equiv d \pmod 4$, which contradicts $(2r)^2 - (2s)^2 d \in 4\mathbb{Z}$. Hence at least one of them is even. Then by $(2r)^2 \equiv (2s)^2 d \pmod 4$ again and $4 \nmid d$ we conclude that both $2r$ and $2s$ are even; i.e. $r, s \in \mathbb{Z}$. So $\alpha = r + s\sqrt{d}$ has the given form.

Now we consider the other case $d \equiv 1 \pmod 4$. By $(2r)^2 \equiv (2s)^2 d \equiv (2s)^2 \pmod 4$ we know that $2r$ and $2s$ are either both even or both odd; i.e. $r - s \in \mathbb{Z}$. Then $\alpha = r + s\sqrt{d} = (r-s) + s(1+\sqrt{d}) = (r-s) + 2s \cdot \omega$ has the given form. □

Now we turn to the notion of the discriminant.

Definition 7.3. Let K be a number field of degree n over $\mathbb{Q}$ and $\alpha_1, \alpha_2, \cdots, \alpha_n$ an n-tuple of elements of K. We define the *discriminant* of the n-tuple to be

$$\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) = \det \begin{pmatrix} T(\alpha_1\alpha_1) & T(\alpha_1\alpha_2) & \cdots & T(\alpha_1\alpha_n) \\ T(\alpha_2\alpha_1) & T(\alpha_2\alpha_2) & \cdots & T(\alpha_2\alpha_n) \\ \vdots & \vdots & \ddots & \vdots \\ T(\alpha_n\alpha_1) & T(\alpha_n\alpha_2) & \cdots & T(\alpha_n\alpha_n) \end{pmatrix}. \tag{7.1}$$

Remark 7.4. If $\alpha_1, \alpha_2, \cdots, \alpha_n \in \mathcal{O}_K$, then each entry of the matrix is an integer by Proposition 6.19, hence the discriminant $\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) \in \mathbb{Z}$.

Proposition 7.5. *The n-tuple $\alpha_1, \alpha_2, \cdots, \alpha_n$ is a $\mathbb{Q}$-basis for K iff $\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) \neq 0$.*

Proof. We first show that if $\{\alpha_i \mid 1 \leqslant i \leqslant n\}$ are linearly dependent over $\mathbb{Q}$, then $\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) = 0$. By assumption we can find $a_1, a_2, \cdots, a_n \in \mathbb{Q}$, not all zero, such that $\sum_{i=1}^n a_i\alpha_i = 0$. Multiply this equation by α_j and take the trace. By Lemma 6.16 we get $\sum_{i=1}^n a_i T(\alpha_i\alpha_j) = 0$ for each $j = 1, 2, \cdots, n$. This shows that the rows of the matrix in (7.1) are linearly dependent, so its determinant is zero.

We then show that if $\{\alpha_i \mid 1 \leqslant i \leqslant n\}$ is a $\mathbb{Q}$-basis for K, then $\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) \neq 0$. Assume on the contrary that $\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) = 0$, then the rows of the matrix in (7.1) are linearly dependent, so we can find $a_1, a_2, \cdots, a_n \in \mathbb{Q}$, not all zero, such that $\sum_{i=1}^n a_i T(\alpha_i\alpha_j) = 0$ for each $j = 1, 2, \cdots, n$. Let $\alpha = \sum_{i=1}^n a_i\alpha_i$. By Lemma 6.16 we get $T(\alpha\alpha_j) = 0$ for each $j = 1, 2, \cdots, n$. Assume on the contrary that $\{\alpha_i \mid 1 \leqslant i \leqslant n\}$ is a basis, then $\alpha \neq 0$, and there exist $b_1, b_2, \cdots, b_n \in \mathbb{Q}$ such that $\alpha^{-1} = \sum_{j=1}^n b_j\alpha_j$. By Lemma 6.16 again we have $T(\alpha\alpha^{-1}) = \sum_{j=1}^n b_j T(\alpha\alpha_j) = 0$. Contradiction to $T(1) = n \neq 0$. □

Proposition 7.6. *Suppose $\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$ and $\{\beta_1, \beta_2, \cdots, \beta_n\}$ are both n-tuples in K. Assume that for each j, $\alpha_j = \sum_{i=1}^n a_{ij}\beta_i$ for some $a_{ij} \in \mathbb{Q}$ and $M = (a_{ij})$ the transition matrix, then*

$$\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) = (\det M)^2 \, \Delta(\beta_1, \beta_2, \cdots, \beta_n).$$

Proof. (This proof is not covered in lecture and is non-examinable.) We have $\alpha_j\alpha_l = \sum_i \sum_k a_{ij}a_{kl}\beta_i\beta_k$. Taking the traces of both sides we get $T(\alpha_j\alpha_l) = \sum_i \sum_k a_{ij}a_{kl}T(\beta_i\beta_k)$. Let $A = (T(\alpha_j\alpha_l))$, $B = (T(\beta_i\beta_k))$ be $n \times n$ matrices. Then we find the matrix identity $A = M'BM$ where M' is the transpose of M. Take the determinant on both sides to get $\det A = (\det M)^2 \det B$, as desired. □

7.2. **Integral bases of ideals.** We focus on the additive structure of the ring $\mathcal{O}_K$, then $\mathcal{O}_K$ is an (additive) abelian group, and every ideal I of $\mathcal{O}_K$ is an abelian subgroup. We are aiming to show that every ideal I is a free abelian group.

Lemma 7.7. *For any $\beta \in K$, there exists some $b \in \mathbb{Z}$, $b \neq 0$, such that $b\beta \in \mathcal{O}_K$.*

Proof. By Lemma 6.13, β is an algebraic number. Therefore β satisfies an equation

$$a_0\beta^m + a_1\beta^{m-1} + a_2\beta^{m-2} + \cdots + a_m = 0$$

where $a_i \in \mathbb{Z}$ for each i and $a_0 \neq 0$. Multiply both sides by a_0^{m-1} to get

$$(a_0\beta)^m + a_1(a_0\beta)^{m-1} + a_2a_0(a_0\beta)^{m-2} + \cdots + a_ma_0^{m-1} = 0.$$

This shows that $a_0\beta$ is an algebraic integer since $a_ia_0^{i-1} \in \mathbb{Z}$ for each i. □

Lemma 7.8. *Every non-zero ideal I of $\mathcal{O}_K$ contains a basis for K over $\mathbb{Q}$.*

Proof. Assume the degree of K over $\mathbb{Q}$ is n. Pick any $\mathbb{Q}$-basis $\beta_1, \beta_2, \cdots, \beta_n$ of K. By Lemma 7.7 we can find some $b \in \mathbb{Z}$, $b \neq 0$, such that $b\beta_1, b\beta_2, \cdots, b\beta_n \in \mathcal{O}_K$. Indeed, there is some non-zero $b_i \in \mathbb{Z}$ for each β_i such that $b_i\beta_i \in \mathcal{O}_K$. Then take b to be any common multiple all b_i's.

We choose any $\alpha \in I$, $\alpha \neq 0$. Then $b\beta_1\alpha, b\beta_2\alpha, \cdots, b\beta_n\alpha$ are in I and form a $\mathbb{Q}$-basis of K. Indeed, for any $a_1, a_2, \cdots, a_n \in \mathbb{Q}$, if

$$a_1b\beta_1\alpha + a_2b\beta_2\alpha + \cdots + a_nb\beta_n\alpha = 0,$$

then since $b\alpha \neq 0$ we have

$$a_1\beta_1 + a_2\beta_2 + \cdots + a_n\beta_n = 0,$$

which implies $a_i = 0$ for each i. Hence $b\beta_1\alpha, b\beta_2\alpha, \cdots, b\beta_n\alpha$ are $\mathbb{Q}$-independent and is a $\mathbb{Q}$-basis for K. □

In other words, the above proposition says we can find a $\mathbb{Q}$-basis for K which entirely consists of algebraic integers. There are in general many choices for the $\mathbb{Q}$-basis of K in $\mathcal{O}_K$, but the follow result shows that some of them are much preferred.

Proposition 7.9. *Let I be a non-zero ideal of $\mathcal{O}_K$. Then we can find $\alpha_1, \alpha_2, \cdots, \alpha_n \in I$ such that they form a $\mathbb{Q}$-basis for K, and for every element α in the field K, $\alpha \in I$ iff $\alpha = a_1\alpha_1 + a_2\alpha_2 + \cdots + a_n\alpha_n$ for some $a_1, a_2, \cdots, a_n \in \mathbb{Z}$.*

Proof. By Lemma 7.8, I contains $\mathbb{Q}$-bases for K. By Remark 7.4 and Proposition 7.5, the discriminant of any such basis is a non-zero integer. Therefore we can always find a $\mathbb{Q}$-basis for $\mathcal{O}_K$ in I such that $|\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n)|$ minimal.

It is clear that every integral linear combination of $\alpha_1, \alpha_2, \cdots, \alpha_n$ is in I since I is an ideal. For the other direction, for any $\alpha \in I$, we can write $\alpha = \gamma_1\alpha_1 + \gamma_2\alpha_2 + \cdots + \gamma_n\alpha_n$ with $\gamma_i \in \mathbb{Q}$. We need to show that every $\gamma_i \in \mathbb{Z}$. If not, then some $\gamma_i \notin \mathbb{Z}$ and by relabeling if necessary we can assume $\gamma_1 \notin \mathbb{Z}$. We write $\gamma_1 = m + \theta$ where $m \in \mathbb{Z}$ and $0 < \theta < 1$. Let $\beta_1 = \alpha - m\alpha_1, \beta_2 = \alpha_2, \cdots, \beta_n = \alpha_n$. Then $\beta_1, \beta_2, \cdots, \beta_n \in I$ and is a $\mathbb{Q}$-basis of K. And the transition matrix between the two basis is

$$\begin{pmatrix} \theta & 0 & \cdots & 0 \\ \gamma_2 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_n & 0 & \cdots & 1 \end{pmatrix}.$$

By Proposition 7.6, we find $\Delta(\beta_1, \beta_2, \cdots, \beta_n) = \theta^2\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n)$, which contradicts the minimality of $|\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n)|$ since $0 < \theta < 1$. Therefore $\gamma_i \in \mathbb{Z}$ for every i, which means every element in I is an integral linear combination of $\alpha_1, \alpha_2, \cdots, \alpha_n$. □

Remark 7.10. We make some comments.

(1) For $\alpha_1, \alpha_2, \cdots, \alpha_n$ satisfying the conditions in Proposition 7.9, we say they form an *integral basis* for I. This is very useful in the sense that every element in K can be uniquely written as a rational linear combination of them, and every element in I can be uniquely written as an integral linear combination of them. We sometimes write $I = \mathbb{Z}\alpha_1 \oplus \mathbb{Z}\alpha_2 \oplus \cdots \oplus \mathbb{Z}\alpha_n$ to indicate the second condition.

(2) As a special case of Proposition 7.9, we think of $\mathcal{O}_K$ as a non-zero ideal in itself. Then there is a $\mathbb{Q}$-basis of K, $\omega_1, \omega_2, \cdots, \omega_n$, such that every element $\alpha \in K$ is a $\mathbb{Q}$-linear combination of $\omega_1, \omega_2, \cdots, \omega_n$, and α is an algebraic integer iff all coefficients in this linear combination are in $\mathbb{Z}$. As an example, if K is a quadratic field, we can choose $\omega_1 = 1$ and $\omega_2 = \omega$ as in Proposition 7.2.

Proposition 7.9 shows the existence of an integral basis for I, but the integral basis for I may not be unique. Although there could be many choices, they all have the same discriminants. We look at the following result:

Lemma 7.11. *Suppose* $\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$ *and* $\{\beta_1, \beta_2, \cdots, \beta_n\}$ *are two integral bases for* I. *Then* $\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) = \Delta(\beta_1, \beta_2, \cdots, \beta_n)$.

Proof. We leave it as an exercise. See Exercise 7.2. □

By Lemma 7.11, the discriminant of an integral basis of an ideal I in $\mathcal{O}_K$ is independent of the choice of the integral basis. We have the following definition:

Definition 7.12. For any non-zero ideal I in $\mathcal{O}_K$, the discriminant of any integral basis of I is called the *discriminant of the ideal* I, written as $\Delta(I)$. In particular, the discriminant of $\mathcal{O}_K$ is called the *discriminant of the number field* K, written as $\Delta(\mathcal{O}_K)$, or simply Δ_K.

Remark 7.13. By Remark 7.4 and Proposition 7.5, we know that $\Delta(I)$ (hence Δ_K) is always a non-zero integer.

The discriminant of a number field is an important quantity associated to a number field. In the following example we give the values for quadratic fields. We need to remember them because they will be used extensively later.

Proposition 7.14. *Let* $d \neq 1$ *be a square-free integer and* $K = \mathbb{Q}(\sqrt{d})$ *a quadratic field. Then*

$$\Delta_K = \begin{cases} 4d & \text{if } d \equiv 2 \text{ or } 3 \pmod 4; \\ d & \text{if } d \equiv 1 \pmod 4. \end{cases}$$

Proof. We leave it as an exercise. See Exercise 7.3. □

Exercise Sheet 7

This sheet is due in the lecture on Tuesday 18th November, and will be discussed in the exercise class on Friday 21st November.

Exercise 7.1. *Examples of discriminants.*

(1) Let K be the cubic field $\mathbb{Q}(\sqrt[3]{2})$. Compute the discriminant $\Delta(1, \sqrt[3]{2}, \sqrt[3]{4})$.

(2) Let K be the cyclotomic field $\mathbb{Q}(\zeta)$ where $\zeta = \cos\frac{2\pi}{5} + i\sin\frac{2\pi}{5}$. Compute the discriminant $\Delta(1, \zeta, \zeta^2, \zeta^3)$.

Exercise 7.2. *The discriminant of an ideal is independent of the choice of integral basis.*

Supply the proof of Lemma 7.11 in the following steps.

(1) Show that there exist $n \times n$ matrices M and N with integer entries, such that

$$\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) = (\det M)^2 \Delta(\beta_1, \beta_2, \cdots, \beta_n),$$
$$\Delta(\beta_1, \beta_2, \cdots, \beta_n) = (\det N)^2 \Delta(\alpha_1, \alpha_2, \cdots, \alpha_n).$$

(2) Show that $(\det M)^2 (\det N)^2 = 1$. Conclude that $(\det M)^2 = (\det N)^2 = 1$ and $\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) = \Delta(\beta_1, \beta_2, \cdots, \beta_n)$.

Exercise 7.3. *The discriminant of a quadratic field.*

Supply the proof of Proposition 7.14 in the following two cases.

(1) Suppose $d \neq 1$ is a square-free integer, $d \equiv 2$ or $3 \pmod 4$ and $K = \mathbb{Q}(\sqrt{d})$. Compute $\Delta(1, \sqrt{d})$. What is the value for Δ_K in this case?

(2) Suppose $d \neq 1$ is a square-free integer, $d \equiv 1 \pmod 4$ and $K = \mathbb{Q}(\sqrt{d})$. Compute $\Delta(1, \frac{1+\sqrt{d}}{2})$. What is the value for Δ_K in this case?

Exercise 7.4. *Integral basis for a principal ideal.*

Let K be a number field of degree n over $\mathbb{Q}$. Assume $\{\omega_1, \omega_2, \cdots, \omega_n\}$ is an integral basis for $\mathcal{O}_K$. Let $\alpha \in \mathcal{O}_K$, $\alpha \neq 0$ and $I = (\alpha)$. Show that $\{\alpha\omega_1, \alpha\omega_2, \cdots, \alpha\omega_n\}$ is an integral basis for I in the following steps.

(1) Show that $\alpha\omega_i \in I$ for each i, $1 \leqslant i \leqslant n$.

(2) Show that $\{\alpha\omega_1, \alpha\omega_2, \cdots, \alpha\omega_n\}$ are $\mathbb{Q}$-linearly independent. Conclude that it is a $\mathbb{Q}$-basis for K.

(3) Show that every $\gamma \in I$ is a linear combination of elements in $\{\alpha\omega_1, \alpha\omega_2, \cdots, \alpha\omega_n\}$ with integer coefficients. Conclude that it is an integral basis for I.

(4) As an example, suppose $K = \mathbb{Q}(\sqrt{3})$. Let $\alpha = \sqrt{3}$ and $I = (\alpha)$ an ideal in $\mathcal{O}_K$. Write down an integral basis for I, and use it to compute the discriminant $\Delta(I)$.

Solutions to Exercise Sheet 7

Solution 7.1. *Examples of discriminants.*

(1) By the definition of the discriminants, we need to compute

$$\Delta(1, \sqrt[3]{2}, \sqrt[3]{4}) = \det \begin{pmatrix} T(1) & T(\sqrt[3]{2}) & T(\sqrt[3]{4}) \\ T(\sqrt[3]{2}) & T(\sqrt[3]{4}) & T(2) \\ T(\sqrt[3]{4}) & T(2) & T(2\sqrt[3]{2}) \end{pmatrix}.$$

By Exercise 6.2 (1), if $\alpha = a + b\sqrt[3]{2} + c\sqrt[3]{4} \in K$ for some $a, b, c \in \mathbb{Q}$, then the trace of α in $\mathbb{Q}(\sqrt[3]{2})$ is given by $T(\alpha) = 3a$. Hence we have $T(1) = 3$, $T(2) = 6$, while $T(\sqrt[3]{2}) = T(\sqrt[3]{4}) = T(2\sqrt[3]{2}) = 0$. Therefore

$$\Delta(1, \sqrt[3]{2}, \sqrt[3]{4}) = \det \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 6 \\ 0 & 6 & 0 \end{pmatrix} = -108.$$

(2) The discriminant that we need to compute is given by

$$\Delta(1, \zeta, \zeta^2, \zeta^3) = \det \begin{pmatrix} T(1) & T(\zeta) & T(\zeta^2) & T(\zeta^3) \\ T(\zeta) & T(\zeta^2) & T(\zeta^3) & T(\zeta^4) \\ T(\zeta^2) & T(\zeta^3) & T(\zeta^4) & T(1) \\ T(\zeta^3) & T(\zeta^4) & T(1) & T(\zeta) \end{pmatrix}.$$

Following the method in Exercise 6.2 (2), we can write down the matrices corresponding to L_1, L_ζ, L_{ζ^2}, L_{ζ^3} and L_{ζ^4} under the basis $\{1, \zeta, \zeta^2, \zeta^3\}$ to compute the corresponding traces. More precisely, we have $T(1) = 4$ by Lemma 6.16 (3) and

$$T(\zeta) = \operatorname{tr}\begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} = -1; \qquad T(\zeta^2) = \operatorname{tr}\begin{pmatrix} 0 & 0 & -1 & 1 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \end{pmatrix} = -1;$$

$$T(\zeta^3) = \operatorname{tr}\begin{pmatrix} 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{pmatrix} = -1; \qquad T(\zeta^4) = \operatorname{tr}\begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \end{pmatrix} = -1.$$

Therefore, the discriminant can be computed as

$$\Delta(1, \zeta, \zeta^2, \zeta^3) = \det\begin{pmatrix} 4 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 4 \\ -1 & -1 & 4 & -1 \end{pmatrix} = \det\begin{pmatrix} 5 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 5 \\ 0 & -1 & 5 & 0 \end{pmatrix} = 125.$$

Solution 7.2. *The discriminant of an ideal.* Since $\{\beta_1, \beta_2, \cdots, \beta_n\}$ is an integral basis for I, for each i we can write $\alpha_i = \sum_{j=1}^n a_{ij}\beta_j$, such that all entries of the transition matrix $M = (a_{ij})$ are integers. By Proposition 7.6, we get

$$\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) = (\det M)^2 \Delta(\beta_1, \beta_2, \cdots, \beta_n). \tag{7.2}$$

Similarly we can write $\beta_i = \sum_{j=1}^n b_{ij}\alpha_j$ and all entries of the transition matrix $N = (b_{ij})$ are also integers. By Proposition 7.6 we also get

$$\Delta(\beta_1, \beta_2, \cdots, \beta_n) = (\det N)^2 \Delta(\alpha_1, \alpha_2, \cdots, \alpha_n). \tag{7.3}$$

By (7.2) and (7.3), we get

$$\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) = (\det M)^2 (\det N)^2 \Delta(\alpha_1, \alpha_2, \cdots, \alpha_n).$$

Since $\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) \neq 0$ by Proposition 7.5, we get $(\det M)^2(\det N)^2 = 1$. Since all entries of M and N are integers, $(\det M)^2$ and $(\det N)^2$ are both non-negative integers, hence $(\det M)^2 = (\det N)^2 = 1$, and the statement we want to prove follows.

Solution 7.3. *The discriminant of a quadratic field.*

(1) In Example 6.18 we know that, for any $\alpha = a + b\sqrt{d} \in K$ for $a, b \in \mathbb{Q}$, its trace in K is given by $T(\alpha) = 2a$. Therefore we have

$$\Delta(1, \sqrt{d}) = \det\begin{pmatrix} T(1) & T(\sqrt{d}) \\ T(\sqrt{d}) & T(d) \end{pmatrix} = \det\begin{pmatrix} 2 & 0 \\ 0 & 2d \end{pmatrix} = 4d.$$

Since $\{1, \sqrt{d}\}$ is an integral basis by Proposition 7.2, we conclude that $\Delta_K = 4d$ for the quadratic field $K = \mathbb{Q}(\sqrt{d})$ when $d \equiv 2$ or $3 \pmod 4$.

(2) We still use the same formula as in part (1). Notice that $T(\frac{1+\sqrt{d}}{2}) = 2 \cdot \frac{1}{2} = 1$ and $T((\frac{1+\sqrt{d}}{2})^2) = T(\frac{1+d+2\sqrt{d}}{4}) = 2 \cdot \frac{1+d}{4} = \frac{1+d}{2}$. Then we have

$$\Delta\left(1, \frac{1+\sqrt{d}}{2}\right) = \det\begin{pmatrix} T(1) & T\left(\frac{1+\sqrt{d}}{2}\right) \\ T\left(\frac{1+\sqrt{d}}{2}\right) & T\left(\left(\frac{1+\sqrt{d}}{2}\right)^2\right) \end{pmatrix} = \det\begin{pmatrix} 2 & 1 \\ 1 & \frac{1+d}{2} \end{pmatrix} = d.$$

Since $\{1, \frac{1+\sqrt{d}}{2}\}$ is an integral basis by Proposition 7.2, we conclude that $\Delta_K = d$ for the quadratic field $K = \mathbb{Q}(\sqrt{d})$ when $d \equiv 1 \pmod 4$.

Solution 7.4. *Integral basis of a principal ideal.*

(1) Since $\alpha \in I$ and $\omega_i \in \mathcal{O}_K$ for each i, by the definition of an ideal, we get $\alpha\omega_i \in I$ for each i.

(2) Assume $b_1\alpha\omega_1 + b_2\alpha\omega_2 + \cdots + b_n\alpha\omega_n = 0$ for some $b_1, b_2, \cdots, b_n \in \mathbb{Q}$. Since $\alpha \neq 0$ we get $b_1\omega_1 + b_2\omega_2 + \cdots + b_n\omega_n = 0$. It follows that $b_i = 0$ for each i, since $\{\omega_1, \omega_2, \cdots, \omega_n\}$ is a $\mathbb{Q}$-basis for K. Therefore $\{\alpha\omega_1, \alpha\omega_2, \cdots, \alpha\omega_n\}$ are $\mathbb{Q}$-independent. Thus they form a $\mathbb{Q}$-basis for K.

(3) Every $\gamma \in I = (\alpha)$ can be written as $\gamma = \alpha\beta$ for some $\beta \in \mathcal{O}_K$. Since $\{\omega_1, \omega_2, \cdots, \omega_n\}$ is an integral basis for $\mathcal{O}_K$, we can write $\beta = b_1\omega_1 + b_2\omega_2 + \cdots + b_n\omega_n$ for $b_1, b_2, \cdots, b_n \in \mathbb{Z}$. Hence $\gamma = b_1\alpha\omega_1 + b_2\alpha\omega_2 + \cdots + b_n\alpha\omega_n$ is an integral linear combination of $\{\alpha\omega_1, \alpha\omega_2, \cdots, \alpha\omega_n\}$. Together with the result in part (2), we conclude that $\alpha\omega_1, \alpha\omega_2, \cdots, \alpha\omega_n$ is an integral basis for I.

(4) By Proposition 7.2, an integral basis for $\mathcal{O}_K$ is given by $\{\omega_1 = 1, \omega_2 = \sqrt{3}\}$. By the conclusion in part (3), an integral basis for I is given by $\{\alpha\omega_1 = \sqrt{3}, \alpha\omega_2 = 3\}$. Therefore we have

$$\Delta(I) = \Delta(\sqrt{3}, 3) = \det \begin{pmatrix} T(3) & T(3\sqrt{3}) \\ T(3\sqrt{3}) & T(9) \end{pmatrix} = \det \begin{pmatrix} 6 & 0 \\ 0 & 18 \end{pmatrix} = 108.$$

8. Unique Factorisation of Ideals

8.1. **Finiteness of Quotient Rings.** We will look at the norm of an ideal I in $\mathcal{O}_K$. There are several descriptions of this notion. We will use the first description as the definition and prove the other descriptions are all equivalent to this one.

Definition 8.1. Let K be a number field and $\mathcal{O}_K$ its ring of integers. The *norm* of any non-zero ideal I of $\mathcal{O}_K$ is defined by

$$N(I) = \left| \frac{\Delta(I)}{\Delta_K} \right|^{\frac{1}{2}}.$$

Remark 8.2. It is worth pointing out the following things about this notion.

(1) This definition is not to be confused with the norm of an element α in the number field K; see Definition 6.14. Although they share the same terminology and notation, whether the argument is an element of an ideal should tell us which definition is in use. On the other hand, the two notions do have very close relation. We will explain that in Proposition 8.9.

(2) By Remark 7.13, we know that both $\Delta(I)$ and Δ_K are non-zero, hence the norm of the ideal I is always well-defined and a positive number. We will show that it is in fact always a positive integer; see Proposition 8.3.

Proposition 8.3. *Suppose $\omega_1, \omega_2, \cdots, \omega_n$ is an integral basis for $\mathcal{O}_K$ and $\alpha_1, \alpha_2, \cdots, \alpha_n$ is an integral basis for I. For each j, suppose $\alpha_j = \sum_{i=1}^n a_{ij}\omega_i$ and $M = (a_{ij})$ is the transition matrix. Then $N(I) = |\det(M)|$. In particular, $N(I)$ is a positive integer.*

Proof. Using Proposition 7.6, we have $\Delta(\alpha_1, \alpha_2, \cdots, \alpha_n) = (\det(M))^2\Delta(\omega_1, \omega_2, \cdots, \omega_n)$. By Definition 7.12, this is equivalent to $\Delta(I) = (\det(M))^2\Delta_K$. By Remark 7.13, $\Delta(I) \neq 0$ and $\Delta_K \neq 0$, hence we get $|\det(M)| = \left|\frac{\Delta(I)}{\Delta_K}\right|^{\frac{1}{2}} = N(I)$. Since $\omega_1, \omega_2, \cdots, \omega_n$ is an integral basis for $\mathcal{O}_K$ and each $\alpha_j \in \mathcal{O}_K$, we know that the coefficients $a_{ij} \in \mathbb{Z}$. Therefore $\det(M)$ is an integer. Since $N(I) \neq 0$, we conclude $N(I) = |\det(M)|$ is a positive integer. □

We give the third description of the norm of the ideal I. It also reveals a special property of the ring $\mathcal{O}_K$, namely, the finiteness of quotient rings.

Proposition 8.4. *For any non-zero ideal I of $\mathcal{O}_K$, the quotient ring $\mathcal{O}_K/I$ is finite and has order $N(I)$.*

Proof. (This proof is not covered in lectures and is non-examinable.) Since I is an ideal in $\mathcal{O}_K$, by forgetting the multiplication on them we know I is a subgroup of $\mathcal{O}_K$. By Proposition 7.9, $\mathcal{O}_K$ and I are both free abelian groups of rank n. By the structure theorem of finitely generated free abelian groups in group theory, we can find an integral basis $\omega_1, \omega_2, \cdots, \omega_n$ for $\mathcal{O}_K$, such that $d_1\omega_1, d_2\omega_2, \cdots, d_n\omega_n$ is an integral basis for I, where each d_i is a positive integer. We write $d = d_1 d_2 \cdots d_n$.

We now show that the quotient ring $\mathcal{O}_K/I$ is finite of order d. In other words, there are precisely d cosets of I in $\mathcal{O}_K$. For this purpose, we will show that

$$S = \{\lambda_1\omega_1 + \lambda_2\omega_2 + \cdots + \lambda_n\omega_n \mid 0 \leqslant \lambda_i < d_i \text{ for } i = 1, 2, \cdots, n\}$$

is a complete set of representatives for cosets of I in $\mathcal{O}_K$. On one hand, for each $\beta \in \mathcal{O}_K$, let $\beta = a_1\omega_1 + a_2\omega_2 + \cdots + a_n\omega_n$ for some $a_1, a_2, \cdots, a_n \in \mathbb{Z}$. For each i, we can write $a_i = q_i d_i + r_i$ for some $0 \leqslant r_i < d_i$. Let $\gamma = r_1\omega_1 + r_2\omega_2 + \cdots r_n\omega_n$, then $\beta - \gamma = q_1 d_1\omega_1 + q_2 d_2\omega_2 + \cdots + q_n d_n\omega_n \in I$. Since $\gamma \in S$, this shows every coset is represented by some element in S. On the other hand, we need to show that elements in S represent distinct cosets. Assume $\lambda = \lambda_1\omega_1 + \lambda_2\omega_2 + \cdots + \lambda_n\omega_n \in S$ and $\delta = \delta_1\omega_1 + \delta_2\omega_2 + \cdots + \delta_n\omega_n \in S$ are in the same coset, then $\lambda - \delta \in I$, which implies $d_i \mid \lambda_i - \delta_i$ for each i. However we also have $-d_i < \lambda_i - \delta_i < d_i$, hence $\lambda_i - \delta_i = 0$ for each i, which implies $\lambda = \delta$. This concludes S is a complete set of representatives for all cosets of I in $\mathcal{O}_K$, hence $\mathcal{O}_K/I$ is finite of order $d = d_1 d_2 \cdots d_n$.

It remains to show that $d = N(I)$. We apply Proposition 8.3 for the particular bases we chose at the beginning of the proof. Under these bases the matrix M is diagonal with diagonal entries $d_1, d_2, \cdots, d_n$ which are positive integers, hence $N(I) = |\det(M)| = d_1 d_2 \cdots d_n = d$. It follows that the order of $\mathcal{O}_K/I$ is $N(I)$. □

The following is an interesting consequence. $N(I) \in \mathbb{Z}$ implies $N(I) \in \mathcal{O}_K$. In fact, we have

Corollary 8.5. *For any non-zero ideal I in $\mathcal{O}_K$, $N(I) \in I$.*

Proof. Since $1 \in \mathcal{O}_K$, we consider the coset $1 + I$. By Proposition 8.4, the sum of $N(I)$ copies of $1 + I$ is the zero element in $\mathcal{O}_K/I$; i.e. the coset $N(I) + I$ is $0 + I$. It follows $N(I) \in I$. □

Corollary 8.6. *For any non-zero ideal I in $\mathcal{O}_K$, $N(I) = 1$ iff $I = \mathcal{O}_K$.*

Proof. Both conditions $N(I) = 1$ and $I = \mathcal{O}_K$ are equivalent to the condition that there is only one coset of I in $\mathcal{O}_K$, hence they are equivalent. □

In other words, the norm of any other non-zero ideal is a positive integer larger than 1.

Remark 8.7. We have understood the norm of an ideal $N(I)$ from three points of views: in terms of discriminants (Definition 8.1); in terms of integral basis and transition matrix (Proposition 8.3); in terms of the quotient ring (Proposition 8.4).

The following consequence of Proposition 8.4 is called the *ascending chain condition*. Recall that a similar result was required to show that every PID is a UFD.

Proposition 8.8 (Ascending Chain Condition)**.** *Let K be a number field. In the ring of integers $\mathcal{O}_K$, every ascending chain of ideals $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ stabilises. In other words, there is a positive integer N such that $I_m = I_{m+1}$ for all $m \geqslant N$.*

Proof. For each $m \in \mathbb{Z}^+$, suppose $d_m = N(I_m)$ which is the order of $\mathcal{O}_K/I_m$ by Proposition 8.4. If $I_m \subsetneq I_{m+1}$, then for any $a \in \mathcal{O}_K$, we have $a + I_m \subsetneq a + I_{m+1}$; i.e. every coset of I_m is contained in some coset of I_{m+1} while every coset of I_{m+1} contains more than one coset of I_m. It follows that $d_m \geqslant d_{m+1}$ and the equality holds iff $I_m = I_{m+1}$. The increasing chain of ideals gives $d_1 \geqslant d_2 \geqslant d_3 \geqslant \cdots$. Since all d_m's are positive integers, there exists some $N > 0$ such that $d_m = d_{m+1}$ for $m \geqslant N$, hence $I_m = I_{m+1}$ for every $m \geqslant N$. □

To provide a convenient tool for computing the norm of a principal ideal, we will explain the relation between the two norms: the norm of an element and the norm of an ideal. If the ideal $I = (\alpha)$ is generated by a single element α, it is natural to expect that $N(\alpha)$ and $N(I)$ are closely related. It is true by the following result.

Proposition 8.9. *Let $I = (\alpha)$ for some non-zero element $\alpha \in \mathcal{O}_K$. Then $N(I) = |N(\alpha)|$.*

Proof. We will follow the definitions to interpret the two norms by determinants of certain matrices. We fix an integral basis $\omega_1, \omega_2, \cdots, \omega_n$ for $\mathcal{O}_K$. It is also a $\mathbb{Q}$-basis for K. For each $j = 1, 2, \cdots, n$, write $\alpha\omega_j = \sum_{i=1}^n a_{ij}\omega_i$, then the linear transformation L_α under this basis is given by the matrix $M = (a_{ij})$. Hence $N(\alpha) = \det(M)$.

To compute $N(I)$, we first need to write down an integral basis for I. By Exercise 7.4, we know that $\alpha\omega_1, \alpha\omega_2, \cdots, \alpha\omega_n$ is such an integral basis. Using this integral basis, we apply Proposition 8.3 and get that $N(I) = |\det(M)|$. It follows that $N(I) = |N(\alpha)|$. □

8.2. **Unique factorisation of ideals.** We review operations of ideals from Algebra 2B.

Let R be a commutative ring with identity 1. Let I and J be ideals of R, then the *sum* of I and J is define to be

$$I + J = \{a + b \in R \mid a \in I, b \in J\},$$

and the *product* of I and J is defined to be

$$IJ = \left\{ \sum_{i=1}^{k} a_i b_i \in R \mid k \in \mathbb{Z}^+, a_i \in I, b_i \in J \text{ for all } 1 \leqslant i \leqslant k \right\}.$$

The sum $I + J$ and product IJ are both ideals of R. This fact is Lemma 2.4 (2013) or Lemma 2.20 (2014) in Algebra 2B.

In particular, for any $\alpha \in R$ and ideal I, we can easily verify that $(\alpha)I = \{\alpha a \mid a \in I\}$.

It is easy to check that under the assumption that R is commutative, both operations are commutative and associative. Namely, for ideals I and J of R, we have $I + J = J + I$ and $IJ = JI$; for ideals I_1, I_2 and I_3 of R, we have $(I_1 + I_2) + I_3 = I_1 + (I_2 + I_3)$ and $(I_1I_2)I_3 = I_1(I_2I_3)$. Therefore, we can simply write $I_1 + I_2 + I_3$ or $I_1I_2I_3$ without specifying the order of the operations.

The building blocks in the factorisation of integers are prime numbers. To study factorisation of ideals, we also need to understand the building blocks first.

Definition 8.10. Let R be a commutative ring with 1. An ideal I of R is a *proper ideal* if $I \neq R$. An ideal $\mathfrak{p}$ of R is a *prime ideal* if $\mathfrak{p}$ is proper, and $ab \in \mathfrak{p}$ implies $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$. An ideal $\mathfrak{m}$ of R is a *maximal ideal* if $\mathfrak{m}$ is proper, and there is no ideal I strictly between $\mathfrak{m}$ and R; i.e. $\mathfrak{m} \subseteq I \subseteq R$ implies $I = \mathfrak{m}$ or $I = R$.

Example 8.11. Let $R = \mathbb{Z}$. (6) is not a prime ideal because $2 \cdot 3 \in (6)$ but $2 \notin (6)$ and $3 \notin (6)$. It is not a maximal idea because $(6) \subsetneq (2) \subsetneq \mathbb{Z}$. But (2) is a prime ideal, because if $ab \in (2)$, then ab is even, hence either a or b is even. (2) is also a maximal ideal because any ideal of $\mathbb{Z}$ has the form (d). If $(2) \subseteq (d) \subseteq \mathbb{Z}$, then $d \mid 2$, hence $(d) = (1)$ or (2).

The notions of prime ideals and maximal ideals lie in the heart of the study of algebraic number theory and algebraic geometry. In general they are distinct notions, but in the context of number fields, we have the following nice agreement.

Proposition 8.12. *Let K be a number field, $\mathcal{O}_K$ its ring of integers, and I a non-zero ideal in $\mathcal{O}_K$. Then I is a prime ideal iff I is a maximal ideal.*

Sketch of Proof. This is a standard fact in commutative ring theory. For any commutative ring R with 1, one can prove that I is a prime ideal iff R/I is an integral domain, and I is a maximal ideal iff R/I is a field. A field is always an integral domain, hence a maximal ideal is a prime ideal. This direction holds for any R. The other direction requires $R = \mathcal{O}_K$. But by Proposition 8.4, $\mathcal{O}_K/I$ is a finite commutative integral domain, hence a field. This shows a non-zero prime ideal is also a maximal ideal. □

We study the unique factorisation of ideals in the ring of integers $\mathcal{O}_K$ of a number field K and its consequences.

Proposition 8.13. *Let I be a non-zero ideal in $\mathcal{O}_K$. Then there exists an ideal J such that IJ is a non-zero principal ideal.*

Proof. This proof is omitted and non-examinable due to the limitation of time. It is technical but does not use anything beyond what have learned so far. □

We have the following two useful consequences. The first one is the cancellation law for ideals in $\mathcal{O}_K$. The second one can be phrased as "to contain is to divide".

Corollary 8.14. *Let I, J_1, J_2 be ideals in $\mathcal{O}_K$, $I \neq 0$. If $IJ_1 = IJ_2$, then $J_1 = J_2$.*

Corollary 8.15. *Let I_1, I_2 be ideals in $\mathcal{O}_K$. If $I_1 \subseteq I_2$, then there exists an ideal J in $\mathcal{O}_K$, such that $I_1 = I_2 J$.*

Proof of Corollaries 8.14 and 8.15. Both statements are simple consequences of Proposition 8.13. We leave them as exercises. See Exercise 8.4. □

Now we are ready to establish the unique factorisation for ideals in $\mathcal{O}_K$.

Theorem 8.16 (Unique Factorisation of Ideals in $\mathcal{O}_K$)**.** *Let K be a number field and $\mathcal{O}_K$ its ring of integers. Then every non-zero proper ideal in $\mathcal{O}_K$ can be uniquely written as a finite product of prime ideals up to reordering factors.*

Proof. The proof consists of two parts: existence and uniqueness of prime factorisations.

First we prove the existence. Let I be a non-zero proper ideal of $\mathcal{O}_K$. We claim that I is contained in some maximal ideal P_1. If I is not contained in any maximal ideal of $\mathcal{O}_K$, then in particular, I itself is not maximal. Hence there is an ideal I_1 with $I \subsetneq I_1 \subsetneq \mathcal{O}_K$. Since I_1 is not maximal, we can find I_2 with $I_1 \subsetneq I_2 \subsetneq \mathcal{O}_K$. The same procedure can be repeated to obtain a strictly increasing chain of infinitely many ideals $I \subsetneq I_1 \subsetneq I_2 \subsetneq I_3 \subsetneq \cdots$, which contradicts Proposition 8.8.

By Corollary 8.15, we have $I = P_1 J_1$ for some ideal J_1. It is clear that $I \subseteq J_1$. We claim $I \neq J_1$. Indeed, if $I = J_1$, then by Corollary 8.14, we have $\mathcal{O}_K = P_1$, which contradicts the properness of P_1.

If $J_1 \neq \mathcal{O}_K$, then the same argument shows that $J_1 = P_2 J_2$ for some maximal ideal P_2 and some ideal J_2 strictly larger than J_1. If $J_2 \neq \mathcal{O}_K$ then we can continue the process to get P_3 and J_3. We claim that we can get $J_r = \mathcal{O}_K$ for some r. If not, this process goes on forever and we get a strictly

increasing chain of infinitely many ideals $I \subsetneq J_1 \subsetneq J_2 \subsetneq J_3 \subsetneq \cdots$, which contradicts Proposition 8.8.

Assume $J_l = \mathcal{O}_K$, then the process terminates here and we get

$$I = P_1 J_1 = P_1 P_2 J_2 = P_1 P_2 P_3 J_3 = \cdots = P_1 P_2 \cdots P_r J_r = P_1 P_2 \cdots P_r,$$

where each P_i is a maximal ideal, hence is also a prime ideal by Proposition 8.12.

Then we prove the uniqueness. Suppose $P_1 P_2 \cdots P_r = I = Q_1 Q_2 \cdots Q_s$ where P_i's and Q_j's are prime ideals. Then $P_1 \supseteq Q_1 Q_2 \cdots Q_s$. We claim that $P_1 \supseteq Q_j$ for some Q_j. If not, then for each $j = 1, 2, \cdots, s$, we can find $a_j \in Q_j \backslash P_1$. Since P_1 is a prime ideal, $a_1 a_2 \cdots a_s \notin P_1$. However $a_1 a_2 \cdots a_s \in Q_1 Q_2 \cdots Q_s \subseteq P_1$. Contradiction.

Therefore, by renumbering the Q_j's if necessary, we can assume that $P_1 \supseteq Q_1$. Since Q_1 is a maximal ideal by Proposition 8.12, we conclude that $P_1 = Q_1$.

Using Corollary 8.14 we obtain $P_2 \cdots P_r = Q_2 \cdots Q_s$. Continuing in the same way we eventually find that $r = s$ and $P_i = Q_i$ for all i after renumbering. □

Exercise Sheet 8

This sheet is due in the lecture on Tuesday 25th November, and will be discussed in the exercise class on Friday 28th November.

Exercise 8.1. *Examples of norms of ideals.*

(1) Let $d \neq 1$ be a square-free integer and $K = \mathbb{Q}(\sqrt{d})$. For any algebraic integer $\alpha = a + b\sqrt{d} \in \mathcal{O}_K$, let $I = (\alpha)$. Find the norm $N(I)$. (Hint: Proposition 8.9.)

(2) Let K be a number field of degree n over $\mathbb{Q}$, $a \in \mathbb{Z}$. Let $I = (a)$ be the principal ideal in $\mathcal{O}_K$ generated by a. Find the norm $N(I)$. (Hint: Proposition 8.9.)

Exercise 8.2. *Examples of sums and products of ideals.*

Let R be a commutative ring with 1.

(1) Let I and J be ideals in R. Show that $IJ \subseteq I$ and $I \subseteq I + J$.

(2) Let I be an ideal in R, $\alpha \in R$. Show that $(\alpha)I = \{\alpha\gamma \mid \gamma \in I\}$.

(3) Let $\alpha, \beta \in R$. Show that $(\alpha)(\beta) = (\alpha\beta)$.

(4) Let $\Bbbk$ be a field. The ideal (x, y) in $\Bbbk[x, y]$ is defined to be the sum of the two principal ideals $(x) + (y)$. Show that (x, y) consists of all polynomials in $\Bbbk[x, y]$ whose constant terms are 0.

Exercise 8.3. *Examples of prime and maximal ideals.*

(1) Let $p \in \mathbb{Z}$ be prime. Show that the principal ideal (p) in $\mathbb{Z}$ is prime and maximal.

(2) Let $\Bbbk$ be a field. Show that the principal ideal (x) in $\Bbbk[x]$ is prime and maximal.

(3) Let $\Bbbk$ be a field. Show that the ideal (x, y) in $\Bbbk[x, y]$ is prime and maximal. Show that the principal ideal (x) in $\Bbbk[x, y]$ is prime but not maximal.

Exercise 8.4. *Cancellation law and "to contain is to divide".*

(1) Prove Corollary 8.14. (Hint: by Proposition 8.13, there is an ideal J such that $IJ = (\gamma)$ is a non-zero principal ideal. Multiply $IJ_1 = IJ_2$ on both sides by J to get $(\gamma)J_1 = (\gamma)J_2$. Then show that $J_1 \subseteq J_2$ and similarly $J_2 \subseteq J_1$ to conclude.)

(2) Prove Corollary 8.15. (Hint: first explain why the statement is clear if $I_2 = 0$. If $I_2 \neq 0$, then by Proposition 8.13, there is an ideal I_3 and $\gamma \neq 0$ such that $I_2 I_3 = (\gamma)$. Hence we have $I_1 I_3 \subseteq I_2 I_3 = (\gamma)$. Define the set $J = \{\alpha \in \mathcal{O}_K \mid \gamma\alpha \in I_1 I_3\}$. Show that J is an ideal, and that $I_1 I_3 = (\gamma)J = I_2 I_3 J$. Then apply the cancellation law proved in part (1) to conclude $I_1 = I_2 J$.)

Solutions to Exercise Sheet 8

Solution 8.1. *Examples of norms of ideals.*

(1) Using the formula in Example 6.18, we have $N(\alpha) = a^2 - b^2 d$. By Proposition 8.9, $N(I) = |N(\alpha)| = |a^2 - b^2 d|$.

(2) By Lemma 6.16, we have $N(a) = a^n N(1) = a^n$. (We can also do it by writing down a matrix for L_a, which is a diagonal matrix with a's along the diagonal.) By Proposition 8.9, $N(I) = |N(a)| = |a^n|$.

Solution 8.2. *Examples of sums and products of ideals.*

(1) We show $IJ \subseteq I$. Every element in IJ has the form $a_1 b_1 + a_2 b_2 + \cdots + a_k b_k$ for some positive integer k, where $a_i \in I$, $b_i \in J$ for each $i = 1, 2, \cdots, k$. Since $a_i \in I$ and $b_i \in J \subseteq R$, we have $a_i b_i \in I$ for each i. Hence their sum $a_1 b_1 + a_2 b_2 + \cdots + a_k b_k \in I$. We then show $I \subseteq I + J$. For every element $a \in I$, we have $a = a + 0 \in I + J$ since $0 \in J$. Both claims are proved.

(2) We need to show mutual inclusions. First we show $(\alpha)I \supseteq \{\alpha a \mid a \in I\}$. This is clear because $\alpha \in (\alpha)$ and $a \in I$ imply $\alpha a \in (\alpha)I$. Then we show the other inclusion $(\alpha)I \subseteq \{\alpha a \mid a \in I\}$. Every element in (α) has the form $r\alpha$ for some $r \in R$. By the definition of the product of two ideals, every element in $(\alpha)I$ can be written as a finite sum $r_1\alpha a_1 + r_2\alpha a_2 + \cdots + r_k\alpha a_k$ for some positive integer k, where $r_1, \cdots, r_k \in R$ and $a_1, \cdots, a_k \in I$. Since I is an ideal, we know that $r_i a_i \in I$ for each $i = 1, \cdots, k$, hence $\gamma = r_1 a_1 + \cdots + r_k a_k \in I$. Therefore $r_1\alpha a_1 + r_2\alpha a_2 + \cdots + r_k\alpha a_k = \alpha(r_1 a_1 + \cdots + r_k a_k) = \alpha\gamma$ has the required form.

(3) We need to show mutual inclusions. First we show $(\alpha)(\beta) \supseteq (\alpha\beta)$. Every element in $(\alpha\beta)$ has the form $r\alpha\beta$ for some $r \in R$. since $r\alpha \in (\alpha)$ and $\beta \in (\beta)$, we know that $r\alpha\beta \in (\alpha)(\beta)$. We then show the other inclusion $(\alpha)(\beta) \subseteq (\alpha\beta)$. By part (2) we know $(\alpha)(\beta) = \{\alpha\gamma \mid \gamma \in (\beta)\}$, hence every element in $(\alpha)(\beta)$ has the form $\alpha\gamma$ for some $\gamma \in (\beta)$. We write $\gamma = \beta\delta$ for some $\delta \in R$, then $\alpha\gamma = \alpha\beta\delta \in (\alpha\beta)$.

(4) We need to show two directions. First we show every element in (x, y) is a polynomial with zero constant term. Since (x, y) is defined to be the sum of ideals $(x) + (y)$, every element in it has the form $xf + yg$ for some $f, g \in \Bbbk[x, y]$. Every term in the expansion of $xf + yg$ has either a factor of x (if it comes from xf) or a factor of y (if it comes from yg). Hence the expansion of $xf + yg$ is a polynomial with zero constant term. Now we show that every polynomial $h \in \Bbbk[x, y]$ with zero constant term is an element in (x, y). Since h has zero constant term, every non-zero term in h has a factor x or y (possibly both). Now we write h as the sum of two polynomials $h = h_1 + h_2$ as follows: if a term in h is divisible by x but not divisible by y, then it becomes a term in h_1; if it is divisible by y but not by x, then it becomes a term in h_2; if it is divisible by both x and y, then it becomes a term in either h_1 or h_2 (the one of your choice). Now we realise that every term in h_1 is divisible by x, hence we can write $h_1 = xf$ for some $f \in \Bbbk[x, y]$. Similarly every term in h_2 is divisible by y, hence we can write $h_2 = yg$ for some $g \in \Bbbk[x, y]$. Therefore $h = xf + yg \in (x) + (y) = (x, y)$.

Solution 8.3. *Examples of prime and maximal ideals.*

(1) It is clear that (p) is a proper ideal since $1 \notin (p)$. We first show (p) is a prime ideal. If $ab \in (p)$ for some $a, b \in \mathbb{Z}$, then $p \mid ab$. Since p is a prime, $p \mid a$ or $p \mid b$, which means either $a \in (p)$ or $b \in (p)$. Hence (p) is a prime ideal. We then show (p) is a maximal ideal. Assume there is an ideal I such that $(p) \subseteq I \subseteq \mathbb{Z}$. Since $\mathbb{Z}$ is a PID, $I = (a)$ is a principal ideal generated by some $a \in \mathbb{Z}$. Then we have $(p) \subseteq (a) \subseteq \mathbb{Z}$, which implies that $p \in (a)$, hence $a \mid p$. It follows that $a = \pm 1$ or $\pm p$. In other words, $I = (a) = (1) = \mathbb{Z}$ or $I = (a) = (p)$. Hence (p) is a maximal ideal.

(2) It is clear that (x) is a proper ideal since the constant polynomial $1 \notin (x)$. We first show (x) is a prime ideal. If $fg \in (p)$ for some $f, g \in \Bbbk[x]$, then $x \mid fg$. Hence either f or g has a factor x, which means either $f \in (x)$ or $g \in (x)$. Hence (x) is a prime ideal. We then show (x) is a maximal ideal. Assume there is an ideal I such that $(x) \subseteq I \subseteq \Bbbk[x]$. Since $\Bbbk[x]$ is a PID, $I = (h)$ is a principal ideal generated by some $h \in \Bbbk[x]$. Then we have $(x) \subseteq (h) \subseteq \Bbbk[x]$, which implies that $x \in (h)$, hence h is a factor of x. It follows that h is a non-zero constant polynomial or a non-zero constant multiple of x. Since every non-zero constant polynomial is a unit in $\Bbbk[x]$, if h is a non-zero constant, then $I = (h) = \Bbbk[x]$; if h is a non-zero constant multiple of x, then $I = (h) = (x)$. Hence (x) is a maximal ideal.

(3) By Exercise 8.2 (4), every element in (x, y) is a polynomial with zero constant term. Hence the constant polynomial $1 \notin (x, y)$, and (x, y) is a proper ideal. We first show that (x, y) is a prime ideal. Assume $fg \in (x, y)$ for some $f, g \in \Bbbk[x, y]$. Then fg has a zero constant term. It follows that either f or g has a zero constant term (otherwise the constant term of fg, as a product of two non-zero constant terms, is non-zero). This shows that either f or g is an element in (x, y), hence (x, y) is a prime ideal. We then show that (x, y) is a maximal ideal. Assume $(x, y) \subseteq I \subseteq \Bbbk[x, y]$. Then either $I = (x, y)$ or I contains some polynomial h with a non-zero constant term. In the second possibility, we write $h = h_0 + c$ where c is the constant term of h while h_0 is the sum of all other terms in h. Since $h \in I$ and $h_0 \in (x, y) \subseteq I$, we know that $c = h - h_0 \in I$. However c is a unit in $\Bbbk[x, y]$, hence $I = \Bbbk[x, y]$. We have proved that an intermediate ideal I is either (x, y) or $\Bbbk[x, y]$. Therefore (x, y) is a maximal ideal.

We now look at the ideal (x). Clearly $1 \notin (x)$, hence (x) is a proper ideal. We first show (x) is a prime ideal. If $fg \in (p)$ for some $f, g \in \Bbbk[x, y]$, then $x \mid fg$. Hence either f or g has a factor x, which means either $f \in (x)$ or $g \in (x)$. Hence (x) is a prime ideal. Then we show that (x) is not a maximal ideal. Indeed, it is clear that every polynomial in (x) is a multiple of x, hence has zero constant term. It follows that $(x) \subseteq (x, y)$. Since y is a polynomial in (x, y) but not in (x), we get the strict inclusions $(x) \subsetneq (x, y) \subsetneq \Bbbk[x, y]$, which shows that (x) is not maximal.

Solution 8.4. *Cancellation law and "to contain is to divide".*

(1) By Proposition 8.13, there is an ideal J such that $IJ = (\gamma)$ is a non-zero principal ideal. Multiply $IJ_1 = IJ_2$ on both sides by J. We find $(\gamma)J_1 = (\gamma)J_2$.

We show that $J_1 \subseteq J_2$. For any element $\alpha \in J_1$, we know that $\gamma\alpha \in (\gamma)J_1$ hence $\gamma\alpha \in (\gamma)J_2$. By Exercise 8.2 (2), we know that every element in $(\gamma)J_2$ can be written as $\gamma\beta$ for some $\beta \in J_2$. It follows that $\gamma\alpha = \gamma\beta$. Since $\gamma \neq 0$, we have $\alpha = \beta \in J_2$. This shows $J_1 \subseteq J_2$. By switching subscripts we can show that $J_2 \subseteq J_1$ using the same argument. Hence $J_1 = J_2$.

(2) If $I_2 = 0$, then $I_1 = 0$, hence we can choose J to be any ideal in $\mathcal{O}_K$. If $I_2 \neq 0$, then by Proposition 8.13, there is an ideal I_3 and $\gamma \neq 0$ such that $I_2 I_3 = (\gamma)$. Hence we have $I_1 I_3 \subseteq I_2 I_3 = (\gamma)$. We define $J = \{\alpha \in \mathcal{O}_K \mid \gamma\alpha \in I_1 I_3\}$.

We show that J is an ideal in $\mathcal{O}_K$. For any $\alpha_1, \alpha_2 \in J$, we have $\gamma\alpha_1, \gamma\alpha_2 \in I_1 I_3$. Since $I_1 I_3$ is an ideal, we get $\gamma(\alpha_1 + \alpha_2) = \gamma\alpha_1 + \gamma\alpha_2 \in I_1 I_3$. By the definition of J, $\alpha_1 + \alpha_2 \in J$. On the other hand, for any $\alpha \in J$ and any $\beta \in \mathcal{O}_K$, since $\alpha\gamma \in I_1 I_3$ and $I_1 I_3$ is an ideal, we

know that $\beta\alpha\gamma \in I_1I_3$. It follows that $\beta\alpha \in J$ by the definition of J. These two conditions prove J is an ideal in $\mathcal{O}_K$.

We claim that $(\gamma)J = I_1I_3$. First we show that $(\gamma)J \subseteq I_1I_3$. By Exercise 8.2 (2), every element in $(\gamma)J$ can be written as $\gamma\alpha$ for some $\alpha \in J$. By the definition of J, we have $\gamma\alpha \in I_1I_3$. Hence $(\gamma)J \subseteq I_1I_3$. To show the other inclusion, assume we have $\beta \in I_1I_3$. Since $I_1I_3 \subseteq (\gamma)$, we know $\beta \in (\gamma)$ hence $\beta = \gamma\alpha$ for some $\alpha \in \mathcal{O}_K$. In fact, by the definition of J we actually have $\alpha \in J$. Hence $\beta = \gamma\alpha \in (\gamma)J$, which shows that $I_1I_3 \subseteq (\gamma)J$. The mutual inclusions show that $(\gamma)J = I_1I_3$. It follows that $I_1I_3 = (\gamma)J = I_2I_3J$. By Corollary 8.14 which we have proved in part (1), we can cancel I_3 on both sides and conclude $I_1 = I_2J$.

9. The Ideal Class Group and Minkowski's Theorem

We introduce the notions of the ideal class group and the class number, and prove Minkowski's Theorem, which will be used later to compute class numbers explicitly.

9.1. **The ideal class group.** We show some important applications of the theorem of unique factorisation of ideals. The following definition plays a major role in algebraic number theory.

Definition 9.1. Let K be a number field and $\mathcal{O}_K$ its ring of integers. Two non-zero ideals I, J in $\mathcal{O}_K$ are said to be equivalent, $I \sim J$, if there exist non-zero $\alpha, \beta \in \mathcal{O}_K$, such that $(\alpha)I = (\beta)J$. This is an equivalence relation. Each equivalence class is called an *ideal class*.

We leave it in Exercise 9.3 to verify that $I \sim J$ is an equivalence relation.

Theorem 9.2. *For any number field K, the set of ideal classes in $\mathcal{O}_K$ form an abelian group.*

Proof. For any non-zero ideal I of $\mathcal{O}_K$, let $\overline{I}$ denote the ideal class containing I. For two ideals I and J of $\mathcal{O}_K$, we define the product of the ideal classes $\overline{I}$ and $\overline{J}$ to be the ideal class $\overline{IJ}$. The product is closed since IJ is an ideal. We need to check the product is well-defined; that is, the product of two ideal classes does not depend on the choice of the ideals in the two classes. This is Exercise 9.3. The commutativity and associativity follow from those of multiplications of ideals. The ideal class containing $\mathcal{O}_K$ serves as the identity for the multiplication. For any non-zero ideal I of $\mathcal{O}_K$, by Proposition 8.13 there exists some ideal J in $\mathcal{O}_K$ such that IJ is a non-zero principle ideal, hence the inverse of $\overline{I}$ is given by $\overline{J}$. Therefore the ideal classes form an abelian group. □

Based on the above theorem, we make the following definitions.

Definition 9.3. Let K be a number field and $\mathcal{O}_K$ its ring of integers. The group of ideal classes in $\mathcal{O}_K$ under multiplication is called the *ideal class group* of K. The order of the ideal class group is called the *class number* of K, denoted by h_K.

Remark 9.4. It can be proved that there are only finitely many ideal classes for every number field, hence the class number is always finite. However, we will only prove the finiteness for quadratic fields. And we will also show how to compute the class number in some explicit examples.

In some sense, the class number measures how far $\mathcal{O}_K$ is from being a PID.

Proposition 9.5. *Let K be a number field and $\mathcal{O}_K$ its ring of integers. Then $h_K = 1$ iff $\mathcal{O}_K$ is a PID.*

Proof. It is clear that that $h_K = 1$ iff every non-zero ideal I is equivalent to $\mathcal{O}_K$, and $\mathcal{O}_K$ is a PID iff every non-zero ideal is principal. Therefore it suffices to show that, for any non-zero ideal I, we have $I \sim \mathcal{O}_K$ iff I is principal.

For one direction, assume that I is a principal ideal (α). Then we have $(1)I = (\alpha)\mathcal{O}_K$, hence $I \sim \mathcal{O}_K$.

For the other direction, assume that $I \sim \mathcal{O}_K$. Then there are non-zero $\alpha, \beta \in \mathcal{O}_K$, such that $(\alpha)I = (\beta)\mathcal{O}_K = (\beta)$. From $\beta \in (\alpha)I$ we know $\beta = \alpha\gamma$ for some $\gamma \in I$. We claim $I = (\gamma)$. It is clear that $I \supseteq (\gamma)$ since $\gamma \in I$. For any $a \in I$, $\alpha a \in (\beta)$ hence $\alpha a = \beta b$ for some $b \in \mathcal{O}_K$. Therefore $a = \gamma b \in (\gamma)$, from which we conclude $I \subseteq (\gamma)$. □

In this proof we have actually showed

Corollary 9.6. *Let I be a non-zero ideal in $\mathcal{O}_K$, then $I \sim \mathcal{O}_K$ iff I is a principal ideal.*

Proof. The proof is already contained in that of Proposition 9.5. □

Corollary 9.7. *Let K be a number field and $\mathcal{O}_K$ its ring of integers. If $h_K = 1$, then $\mathcal{O}_K$ is a UFD.*

Proof. This is an immediate consequence of Proposition 9.5 and Theorem 1.11. □

Example 9.8. If $K = \mathbb{Q}[i]$, then $\mathcal{O}_K = \mathbb{Z}[i]$ by Proposition 7.2. From Exercise 1.4 we know $\mathbb{Z}[i]$ is a Euclidean domain, hence a PID and UFD. Then we know the class number of $K = \mathbb{Q}[i]$ is 1. In many other examples, the opposite direction could be more useful: if we can show the class number $h_K = 1$, then $\mathcal{O}_K$ is a UFD. Hence it is important to find a systematic way to compute class numbers. We will see it later.

Our next goal is to prove Minkowski's Theorem, which is the main tool for computing class numbers. We need to introduce some terminologies before stating the theorem.For the moment we forget number theory and think about some geometry.

Definition 9.9. Let e_1, e_2 be two linearly independent vectors in $\mathbb{R}^2$. The abelian group $L = \{m_1e_1 + m_2e_2 \mid m_1, m_2 \in \mathbb{Z}\}$ is called a *lattice* of rank 2 in $\mathbb{R}^2$. The set $\{e_1, e_2\}$ is called a *generator* of L. The *fundamental domain* of L with respect to the generator $\{e_1, e_2\}$ is the set $T = \{a_1e_1 + a_2e_2 \mid a_1, a_2 \in \mathbb{R}, 0 \leqslant a_1 < 1, 0 \leqslant a_2 < 1\}$.

Using the standard metric on $\mathbb{R}^2$, we can define the *volume* (or *area*) of a measurable subset $X \subseteq \mathbb{R}^2$ in the usual way, more precisely by $\int_X \mathrm{d}x\mathrm{d}y$, denoted by $\mathrm{vol}(X)$. However the only examples that we are interested in are the volumes of rectangles, disks, and parallelograms, which are familiar. For instance, let $e_i = (x_i, y_i)$ for $i = 1, 2$, then the volumn of the fundamental domain of the lattice L is given by

$$\mathrm{vol}(T) = \left| \det \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} \right|.$$

Definition 9.10. A subset $X \subseteq \mathbb{R}^2$ is *convex* if, whenever $p, q \in X$, the point $\lambda p + (1 - \lambda)q \in X$ for all real λ, $0 \leqslant \lambda \leqslant 1$. A subset $X \subset \mathbb{R}^2$ is *centrally symmetric* if $p \in X$ implies $-p \in X$.

In other words, if X is convex, then the straight line segment joining two points in X completely lies in X. For example a disk, a square, a triangle is convex, but an annulus is not. A disk is centrally symmetric only when its centre is at $(0, 0)$.

9.2. **Minkowski's theorem.** From now on we will focus on quadratic fields $\mathbb{Q}(\sqrt{d})$ for any square-free integer d and prove their class numbers are finite; see Example 6.12 (1). Now we state the famous Minkowski's Theorem in dimension 2, which is the main tool in studying this problem.

Theorem 9.11 (Minkowski's Theorem)**.** *Let L be a lattice of rank 2 in $\mathbb{R}^2$ with fundamental domain T. Let X be a centrally symmetric convex subset of $\mathbb{R}^2$. If $\operatorname{vol}(X) > 4\operatorname{vol}(T)$, then X contains a non-zero point of L.*

Proof. We first shrink X to half of its size in length; precisely speaking, we consider $Y = \{p \in \mathbb{R}^2 \mid 2p \in X\}$. Then $\operatorname{vol}(Y) = \frac{1}{4}\operatorname{vol}(X) > \operatorname{vol}(T)$.

For every $h \in L$, we define $h+T = \{h+p \mid p \in T\}$ which is the transport of the fundamental domain along the vector h. It is clear that $\mathbb{R}^2$ becomes the disjoint union of these parallelograms. Let $Y_h = Y \cap (h+T)$ is the part of Y which lies in the parallelogram $h+T$ for each $h \in L$, then Y becomes the disjoint union of all Y_h's, hence $\sum_{h\in L}\operatorname{vol}(Y_h) = \operatorname{vol}(Y) > \operatorname{vol}(T)$. We transport each Y_h back to the fundamental domain, say $Y_h' = \{q \in T \mid h+q \in Y_h\}$. Then $\sum_{h\in L}\operatorname{vol}(Y_h') = \sum_{h\in L}\operatorname{vol}(Y_h) > \operatorname{vol}(T)$. Since each $Y_h' \subseteq T$, this inequality implies they are not disjoint. Therefore there exist $h_1, h_2 \in L$, $h_1 \neq h_2$, such that we can find some $q \in Y_{h_1}' \cap Y_{h_2}'$. That implies $p_1 = h_1 + q \in Y_{h_1} \subseteq Y$ and $p_2 = h_2 + q \in Y_{h_2} \subseteq Y$, hence we found $p_1, p_2 \in Y$, such that $p_1 - p_2 = h_1 - h_2 \in L$.

Since $p_1, p_2 \in Y$, we have $2p_1, 2p_2 \in X$. Since X is centrally symmetric, $-2p_2 \in X$. Since X is convex, $\frac{1}{2}(2p_1) + \frac{1}{2}(-2p_2) \in X$. And $\frac{1}{2}(2p_1) + \frac{1}{2}(-2p_2) = h_1 - h_2$ is a non-zero point in L. □

Corollary 9.12. *Let L be a lattice of rank 2 in $\mathbb{R}^2$ with fundamental domain T. Let X be a centrally symmetric convex subset of $\mathbb{R}^2$. If X is compact (i.e. closed and bounded), and $\operatorname{vol}(X) \geqslant 4\operatorname{vol}(T)$, then X contains a non-zero point of L.*

Proof. We do not prove this corollary rigorously because it requires some understanding of topology. Intuitively, we can enlarge X a little bit so that we can apply Theorem 9.11 and obtain lattice points in the enlarged X. Since this enlargement can be arbitrarily tiny, there must be lattice points within the boundary of X. □

As an indication on how geometry can be used to study number fields, we construct lattices from some familiar objects. Here we consider a quadratic number field $K = \mathbb{Q}(\sqrt{d})$ for any square-free integer d. As usual, its ring of integers is denoted by $\mathcal{O}_K$ and let I be any non-zero ideal of $\mathcal{O}_K$.

Proposition 9.13. *Let $d < 0$ be a square-free integer and $K = \mathbb{Q}(\sqrt{d})$ a quadratic field. For any non-zero ideal I in $\mathcal{O}_K$, the set $L_I = \big\{(\operatorname{Re}\alpha, \operatorname{Im}\alpha) \in \mathbb{R}^2 \mid \alpha \in I\big\}$ is a lattice of rank 2 in $\mathbb{R}^2$. Let T_I be the fundamental domain of L_I, then $\operatorname{vol}(T_I) = \frac{1}{2}N(I)\,|\Delta_K|^{\frac{1}{2}}$.*

Proof. This proposition can be proved in three steps.

Step 1. We prove that L_I is a lattice of rank 2 in $\mathbb{R}^2$. By Proposition 7.9, assume α_1, α_2 is an integral basis for I, then we can write $I = \{m_1\alpha_1 + m_2\alpha_2 \mid m_1, m_2 \in \mathbb{Z}\}$. Let $e_1 = (\operatorname{Re}\alpha_1, \operatorname{Im}\alpha_1)$ and $e_2 = (\operatorname{Re}\alpha_2, \operatorname{Im}\alpha_2)$, then for every $\alpha = m_1\alpha_1 + m_2\alpha_2 \in I$, $(\operatorname{Re}\alpha, \operatorname{Im}\alpha) = m_1(\operatorname{Re}\alpha_1, \operatorname{Im}\alpha_1) + m_2(\operatorname{Re}\alpha_2, \operatorname{Im}\alpha_2) = m_1e_1 + m_2e_2$. Hence $L_I = \{m_1e_1 + m_2e_2 \mid m_1, m_2 \in \mathbb{Z}\}$ is a rank 2 lattice in $\mathbb{R}^2$.

Step 2. We calculate the volume of the fundamental domain in a special case, i.e. $T_{\mathcal{O}_K}$. By Proposition 7.2, we can write $\mathcal{O}_K = \{m_1\omega_1 + m_2\omega_2 \mid m_1, m_2 \in \mathbb{Z}\}$, where $\omega_1 = 1$, and $\omega_2 = \sqrt{d}$ if $d \equiv 2$ or $3 \pmod 4$ and $\frac{1}{2}(1+\sqrt{d})$ if $d \equiv 1 \pmod 4$.

When $d \equiv 2$ or $3 \pmod 4$, we have $e_1 = (\operatorname{Re}\omega_1, \operatorname{Im}\omega_1) = (1, 0)$ and $e_2 = (\operatorname{Re}\omega_2, \operatorname{Im}\omega_2) = (0, \sqrt{-d})$. Hence the volume of the fundamental domain is

$$\operatorname{vol}(T_{\mathcal{O}_K}) = \left|\det\begin{pmatrix} 1 & 0 \\ 0 & \sqrt{-d} \end{pmatrix}\right| = \sqrt{-d} = \frac{1}{2}\,|\Delta_K|^{\frac{1}{2}},$$

where the last equality follows from Proposition 7.14.

When $d \equiv 1 \pmod 4$, we have $e_1 = (\operatorname{Re}\omega_1, \operatorname{Im}\omega_1) = (1,0)$ and $e_2 = (\operatorname{Re}\omega_2, \operatorname{Im}\omega_2) = (\frac{1}{2}, \frac{1}{2}\sqrt{-d})$. Hence the volume of the fundamental domain is

$$\operatorname{vol}(T_{\mathcal{O}_K}) = \left| \det \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2}\sqrt{-d} \end{pmatrix} \right| = \frac{1}{2}\sqrt{-d} = \frac{1}{2} |\Delta_K|^{\frac{1}{2}},$$

where the last equality still follows from Proposition 7.14.

Step 3. We calculate the volume of the fundamental domain T_I in general. For an arbitrary ideal I with an integral basis α_1, α_2, we can write $\alpha_1 = a_{11}\omega_1 + a_{21}\omega_2$ and $\alpha_2 = a_{12}\omega_1 + a_{22}\omega_2$, as well as the transition matrix $M = (a_{ij})$, where $a_{ij} \in \mathbb{Z}$. By taking real parts and imaginary parts of α_1 and α_2, we realise that they can be organised into the following matrix

$$\begin{pmatrix} \operatorname{Re}\alpha_1 & \operatorname{Re}\alpha_2 \\ \operatorname{Im}\alpha_1 & \operatorname{Im}\alpha_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \operatorname{Re}\omega_1 & \operatorname{Re}\omega_2 \\ \operatorname{Im}\omega_1 & \operatorname{Im}\omega_2 \end{pmatrix}.$$

Taking determinants and absolute values on both sides, we get

$$\operatorname{vol}(T_I) = |\det M| \operatorname{vol}(T_{\mathcal{O}_K}).$$

By Proposition 8.3 and step 2, we conclude that

$$\operatorname{vol}(T_I) = \frac{1}{2} N(I) |\Delta_K|^{\frac{1}{2}}$$

as required. □

A parallel statement can be established as follows

Proposition 9.14. *Let $d > 1$ be square-free and $K = \mathbb{Q}(\sqrt{d})$ a quadratic field. For any non-zero ideal I of $\mathcal{O}_K$, the set $L_I = \left\{(a + b\sqrt{d}, a - b\sqrt{d}) \in \mathbb{R}^2 \mid a + b\sqrt{d} \in I, a, b \in \mathbb{Q}\right\}$ is a lattice of rank 2 in $\mathbb{R}^2$. Let T_I be the fundamental domain of L_I, then $\operatorname{vol}(T_I) = N(I) |\Delta_K|^{\frac{1}{2}}$.*

Proof. We leave it as an exercise. See Exercise 9.4. □

Exercise Sheet 9

This sheet is due in the lecture on Tuesday 2nd December, and will be discussed in the exercise class on Friday 5th December.

Exercise 9.1. *Card games and non-card games.*

Answer the following questions. You do not need to justify your answers.

(1) Which of the following shape(s) is/are convex? (i) a spade; (ii) a heart; (iii) a club; (iv) a diamond; (v) a joker. (Hint: think of these shapes as in a standard 52-card deck, but pretend that the four sides of the diamond are straight line segments.)

(2) Which of the following shape(s) is/are centrally symmetric? (i) a square with vertices $(0,0), (1,0), (1,1), (0,1)$; (ii) a rhombus with vertices $(1,0)$, $(0,2)$, $(-1,0)$, $(0,-2)$; (iii) a triangle with vertices $(1,-1), (0,1), (-1,-1)$; (iv) a parallelogram with vertices $(2,3)$, $(3,4)$, $(-2,-3)$, $(-3,-4)$; (v) a disk $\{(x,y) \in \mathbb{R}^2 \mid (x-1)^2 + (y-1)^2 \leqslant 1\}$; (vi) an annulus $\{(x,y) \in \mathbb{R}^2 \mid 1 \leqslant x^2 + y^2 \leqslant 2\}$.

Exercise 9.2. *Applications of Minkowski's Theorem.*

(1) Assume we have a lattice L of rank 2 in $\mathbb{R}^2$ whose fundamental domain has volume A. For which positive values of r is the disk $D = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leqslant r^2\}$ guaranteed to contain at least one non-zero point of L?

(2) Assume we have a lattice L of rank 2 in $\mathbb{R}^2$ whose fundamental domain has volume A. For which positive values of r is the square $S = \{(x, y) \in \mathbb{R}^2 \mid |x| + |y| \leqslant r\}$ guaranteed to contain at least one non-zero point of L?

Exercise 9.3. *Basic properties of ideal classes.*

(1) Show that the relation $\sim$ in Definition 9.1 is an equivalence relation. (Hint: an equivalence relation requires (i) reflexivity: $I \sim I$; (ii) symmetry: if $I \sim J$ then $J \sim I$; (iii) transitivity: if $I_1 \sim I_2$ and $I_2 \sim I_3$ then $I_1 \sim I_3$.)

(2) Show that the product of ideal classes is well-defined; i.e. if $I_1 \sim I_2$ and $J_1 \sim J_2$, then $I_1 J_1 \sim I_2 J_2$.

Exercise 9.4. *Volume of the fundamental domain for real quadratic fields.*

Supply the proof of Proposition 9.14 in the following steps.

(1) Prove L_I is a lattice of rank 2 in $\mathbb{R}^2$ by writing down a pair of generators e_1, e_2.

(2) Use the integral basis of $\mathcal{O}_K$ given in Proposition 7.2 to compute $\mathrm{vol}(T_{\mathcal{O}_K})$.

(3) Use a matrix M to relate $\mathrm{vol}(T_I)$ and $\mathrm{vol}(T_{\mathcal{O}_K})$, and prove the formula for $\mathrm{vol}(T_I)$.

Solutions to Exercise Sheet 9

Solution 9.1. *Card games and non-card games.*

(1) A diamond is convex assuming the four sides are all line segments (despite that they look a little curved on any playing cards). All the other shapes are non-convex.

(2) The shapes (ii), (iv) and (vi) are centrally symmetric. The other shapes are not.

Solution 9.2. *Applications of Minkowski's Theorem.*

(1) D is centrally symmetric, convex and compact. Hence Corollary 9.12 applies. If $\mathrm{vol}(D) \geqslant 4A$, then D is guaranteed to contain a non-zero point in L. This condition can be written as $\pi r^2 \geqslant 4A$. When $r > 0$, it is equivalent to $r \geqslant \left(\frac{4A}{\pi}\right)^{\frac{1}{2}}$.

(2) S is centrally symmetric, convex and compact. Hence Corollary 9.12 applies. If $\mathrm{vol}(S) \geqslant 4A$, then S is guaranteed to contain a non-zero point in L. Note that $\mathrm{vol}(S) = 2r^2$, hence this condition becomes $2r^2 \geqslant 4A$. When $r > 0$, it is equivalent to $r \geqslant (2A)^{\frac{1}{2}}$.

Solution 9.3. *Basic properties of ideal classes.*

(1) The reflexivity is clear, as for any non-zero principal ideal (α), we have $(\alpha)I = (\alpha)I$, hence $I \sim I$. The symmetry is also easy. If $I \sim J$, then there exist non-zero principal ideals (α) and (β), such that $(\alpha)I = (\beta)J$. We switch the two sides and write the equation as $(\beta)J = (\alpha)I$, then by definition we get $J \sim I$.

Now we prove the transitivity. By $I_1 \sim I_2$, we can find non-zero principal ideals (α_1) and (α_2), such that $(\alpha_1)I_1 = (\alpha_2)I_2$. By $I_2 \sim I_3$, we can find non-zero principal ideals (β_2) and (β_3), such that $(\beta_2)I_2 = (\beta_3)I_3$. We multiply both sides of the first identity by (β_2) and get $(\alpha_1)(\beta_2)I_1 = (\alpha_2)(\beta_2)I_2$. By Exercise 8.2 (3), we can rewrite it as $(\alpha_1\beta_2)I_1 = (\alpha_2\beta_2)I_2$. Similarly, we can multiply both sides of the second identity by (α_2) to get $(\alpha_2)(\beta_2)I_2 = (\alpha_2)(\beta_3)I_3$, which can be rewritten as $(\alpha_2\beta_2)I_2 = (\alpha_2\beta_3)I_3$. Now we get $(\alpha_1\beta_2)I_1 = (\alpha_2\beta_2)I_2 = (\alpha_2\beta_3)I_3$. We need to show that $(\alpha_1\beta_2)$ and $(\alpha_2\beta_3)$ are both non-zero principal ideals. Since α_1 and β_2 are both non-zero complex numbers, their product $\alpha_1\beta_2$ is also non-zero, hence $(\alpha_1\beta_2)$ is also a non-zero principal ideal. For the same reason $(\alpha_2\beta_3)$ is a non-zero principal ideal. Hence we conclude that $I_1 \sim I_3$.

(2) From $I_1 \sim I_2$, we know that for some non-zero principal ideals (α_1) and (α_2), we have $(\alpha_1)I_1 = (\alpha_2)I_2$. $J_1 \sim J_2$, we know that for some non-zero principal ideals (β_1) and (β_2), we have $(\beta_1)J_1 = (\beta_2)J_2$. We multiply the two identities to get $(\alpha_1)(\beta_1)I_1J_1 = (\alpha_2)(\beta_2)I_2J_2$. By Exercise 8.2 (3), we can rewrite it as $(\alpha_1\beta_1)I_1J_1 = (\alpha_2\beta_2)I_2J_2$. For similar reasons as in part (1), both $(\alpha_1\beta_1)$ and $(\alpha_2\beta_2)$ are non-zero principal ideals. Hence we have $I_1J_1 \sim I_2J_2$.

Solution 9.4. *Volume of the fundamental domain for real quadratic fields.*

(1) We prove that L_I is a lattice of rank 2 in $\mathbb{R}^2$. By Proposition 7.9, assume α_1, α_2 is an integral basis for I, then we can write $I = \{m_1\alpha_1 + m_2\alpha_2 \mid m_1, m_2 \in \mathbb{Z}\}$. We write $\alpha_1 = a_1 + b_1\sqrt{d}$ and $\alpha_2 = a_2 + b_2\sqrt{d}$ for some $a_1, b_1, a_2, b_2 \in \mathbb{Q}$. Let $e_1 = (a_1 + b_1\sqrt{d}, a_1 - b_1\sqrt{d})$ and $e_2 = (a_2 + b_2\sqrt{d}, a_2 - b_2\sqrt{d})$, then for every $\alpha = m_1\alpha_1 + m_2\alpha_2 = (m_1a_1 + m_2a_2) + (m_1b_1 + m_2b_2)\sqrt{d} \in I$, the corresponding point in L_I is given by $((m_1a_1 + m_2a_2) + (m_1b_1 + m_2b_2)\sqrt{d}, (m_1a_1 + m_2a_2) - (m_1b_1 + m_2b_2)\sqrt{d}) = m_1(a_1 + b_1\sqrt{d}, a_1 - b_1\sqrt{d}) + m_2(a_2 + b_2\sqrt{d}, a_2 - b_2\sqrt{d}) = m_1e_1 + m_2e_2$. Hence $L_I = \{m_1e_1 + m_2e_2 \mid m_1, m_2 \in \mathbb{Z}\}$ is a rank 2 lattice in $\mathbb{R}^2$.

(2) We calculate $T_{\mathcal{O}_K}$. By Proposition 7.2, we can write $\mathcal{O}_K = \{m_1\omega_1 + m_2\omega_2 \mid m_1, m_2 \in \mathbb{Z}\}$, where $\omega_1 = 1$, and $\omega_2 = \sqrt{d}$ if $d \equiv 2$ or $3 \pmod 4$ and $\frac{1}{2}(1 + \sqrt{d})$ if $d \equiv 1 \pmod 4$.

When $d \equiv 2$ or $3 \pmod 4$, we have $e_1 = (1, 1)$ and $e_2 = (\sqrt{d}, -\sqrt{d})$. Hence

$$\mathrm{vol}(T_{\mathcal{O}_K}) = \left|\det \begin{pmatrix} 1 & \sqrt{d} \\ 1 & -\sqrt{d} \end{pmatrix}\right| = \left|-2\sqrt{d}\right| = 2\sqrt{d} = |\Delta_K|^{\frac{1}{2}},$$

where the last equality follows from Proposition 7.14.

When $d \equiv 1 \pmod 4$, we have $e_1 = (1, 1)$ and $e_2 = (\frac{1}{2}(1 + \sqrt{d}), \frac{1}{2}(1 - \sqrt{d}))$. Hence the volume of the fundamental domain is

$$\mathrm{vol}(T_{\mathcal{O}_K}) = \left|\det \begin{pmatrix} 1 & \frac{1}{2}(1 + \sqrt{d}) \\ 1 & \frac{1}{2}(1 - \sqrt{d}) \end{pmatrix}\right| = \left|-\sqrt{d}\right| = \sqrt{d} = |\Delta_K|^{\frac{1}{2}},$$

where the last equality still follows from Proposition 7.14.

(3) We calculate the volume of the fundamental domain T_I in general. For an arbitrary ideal I with an integral basis α_1, α_2, we can write $\alpha_1 = a_{11}\omega_1 + a_{21}\omega_2$ and $\alpha_2 = a_{12}\omega_1 + a_{22}\omega_2$, as well as the transition matrix $M = (a_{ij})$, where $a_{ij} \in \mathbb{Z}$. For simplicity, we write the points in L_I corresponding to α_i by (α_i, α_i') for $i = 1, 2$. Similarly, we write the points in L_I corresponding to ω_i by (ω_i, ω_i') for $i = 1, 2$. Then they can be organised into the following matrix

$$\begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_1' & \alpha_2' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \omega_1 & \omega_2 \\ \omega_1' & \omega_2' \end{pmatrix}.$$

Taking determinants and absolute values on both sides, we get

$$\mathrm{vol}(T_I) = |\det M| \, \mathrm{vol}(T_{\mathcal{O}_K}).$$

By Proposition 8.3 and part (2), we conclude that

$$\mathrm{vol}(T_I) = N(I) \, |\Delta_K|^{\frac{1}{2}}.$$

10. Computation of Class Numbers

We will establish the Minkowski bound for class numbers, and show how to use it to make explicit computations in examples.

10.1. **Minkowski bound.** We will show an upper bound for class numbers due to Minkowski. The formula is still a little different for real quadratic fields and imaginary quadratic fields.

Proposition 10.1. *Let $K = \mathbb{Q}(\sqrt{d})$ be a quadratic field with $d < 0$. For any non-zero ideal I of $\mathcal{O}_K$, there exists a non-zero element $\alpha \in I$ such that $|N(\alpha)| \leqslant \frac{2}{\pi} N(I) \, |\Delta_K|^{\frac{1}{2}}$.*

Proof. By Proposition 9.13, we know that L_I is a rank 2 lattice in $\mathbb{R}^2$ with the volume of the fundamental domain $\operatorname{vol}(T_I) = \frac{1}{2} N(I) \, |\Delta_K|^{\frac{1}{2}}$.

Now we consider the closed disk D with centre $(0,0)$ and radius $r = \left(\frac{2}{\pi} N(I) \, |\Delta_K|^{\frac{1}{2}} \right)^{\frac{1}{2}}$. D is centrally symmetric, convex, compact, with volume $\operatorname{vol}(D) = \pi r^2 = 2N(I) \, |\Delta_K|^{\frac{1}{2}} = 4 \operatorname{vol}(T_I)$. By Corollary 9.12, D contains non-zero lattice point in L_I. In other words, there exists some $\alpha \in I$, such that the point $(\operatorname{Re} \alpha, \operatorname{Im} \alpha) \in D$. Hence $(\operatorname{Re} \alpha)^2 + (\operatorname{Im} \alpha)^2 \leqslant r^2$. If we write $\alpha = a + b\sqrt{d}$, then $\operatorname{Re} \alpha = a$ and $\operatorname{Im} \alpha = b\sqrt{-d}$, hence $(\operatorname{Re} \alpha)^2 + (\operatorname{Im} \alpha)^2 = a^2 - b^2 d = N(\alpha)$ by Example 6.18. In particular, $N(\alpha) \geqslant 0$. It follows that $|N(\alpha)| = N(\alpha) \leqslant r^2 = \frac{2}{\pi} N(I) \, |\Delta_K|^{\frac{1}{2}}$. □

To prove next result we need the following lemma

Lemma 10.2. *For any numeber field K, let I and J be non-zero ideals in $\mathcal{O}_K$. Then $N(IJ) = N(I)N(J)$.*

Proof. The proof is omitted and non-examinable. It is a consequence of Theorem 8.16. □

Proposition 10.3. *Let $K = \mathbb{Q}(\sqrt{d})$ be a quadratic field with $d < 0$. Then every ideal class $\mathcal{C}$ of $\mathcal{O}_K$ contains an ideal I with $N(I) \leqslant \frac{2}{\pi} \, |\Delta_K|^{\frac{1}{2}}$.*

Proof. By Theorem 9.2, the ideal class $\mathcal{C}$ has an inverse in the ideal class group. We denote this inverse ideal class by $\overline{J}$ where J is any representative. Then by Proposition 10.1, there exists a non-zero element $\beta \in J$ such that $|N(\beta)| \leqslant \frac{2}{\pi} N(J) \, |\Delta_K|^{\frac{1}{2}}$. Since we have $(\beta) \subseteq J$, there exists some ideal I such that $IJ = (\beta)$ by Corollary 8.15. Since the ideal class containing (β) is the identity element in the ideal class group, $\overline{I}$ and $\overline{J}$ are inverse of each other, hence I is an ideal in $\mathcal{C}$. It remains to show $N(I)$ satisfies the given bound.

By Lemma 10.2 and Proposition 8.9, we have the following calculation

$$N(I)N(J) = N(IJ) = N((\beta)) = |N(\beta)| \leqslant \frac{2}{\pi} N(J) \, |\Delta_K|^{\frac{1}{2}} .$$

Since $N(J)$ is a positive integer by Proposition 8.3, we cancel it to get $N(I) \leqslant \frac{2}{\pi} \, |\Delta_K|^{\frac{1}{2}}$ as required. □

We can get the following parallel results for real quadratic fields. We leave the proofs as exercises.

Proposition 10.4. *Let $K = \mathbb{Q}(\sqrt{d})$ be a quadratic field with $d > 0$. For any ideal I of $\mathcal{O}_K$, there exists a non-zero element $\alpha \in I$ such that $|N(\alpha)| \leqslant \frac{1}{2} N(I) \, |\Delta_K|^{\frac{1}{2}}$.*

Proof. The proof is similar to that of Proposition 10.1. See Exercise 10.3. □

Proposition 10.5. *Let $K = \mathbb{Q}(\sqrt{d})$ be a quadratic field with $d > 0$. Then every ideal class $\mathcal{C}$ of $\mathcal{O}_K$ contains an ideal I with $N(I) \leqslant \frac{1}{2} \, |\Delta_K|^{\frac{1}{2}}$.*

Proof. The proof is similar to that of Proposition 10.3. See Exercise 10.3. □

Summarising the above results, we get the following definition:

Definition 10.6. Let d be a square-free integer, $d \neq 1$, and $K = \mathbb{Q}(\sqrt{d})$ a quadratic field. The *Minkowski bound* M_K is defined by

$$M_K = \begin{cases} \frac{2}{\pi} |\Delta_K|^{\frac{1}{2}} & \text{if } d < 0, \\ \frac{1}{2} |\Delta_K|^{\frac{1}{2}} & \text{if } d > 0, \end{cases}$$

with the property that every ideal class in $\mathcal{O}_K$ contains an ideal whose norm is at most M_K.

This allows us to prove the following important result:

Theorem 10.7. *Let d be a square-free integer, $d \neq 1$, and $K = \mathbb{Q}(\sqrt{d})$ a quadratic field. The class number h_K is finite.*

Proof. By Definition 10.6, every ideal class contains an ideal with norm not larger than M_K. Hence it remains to show there are only finitely many ideals with norm not larger than M_K. By Proposition 8.3, every such norm is a positive integer not larger than M_K, hence there are only finitely many choices for such norms. It suffices to show that for every fixed positive integer $q \leqslant M_K$, there are only finitely many ideals I with $N(I) = q$.

By Corollary 8.5, we know $q \in I$, hence $(q) \subseteq I$. By Corollary 8.15, we can find some ideal J such that $(q) = IJ$. By Theorem 8.16, the ideal (q) has a unique factorisation into finitely many prime ideals, say $(q) = P_1 P_2 \cdots P_r$. Since I is a factor of (q), it must be the product of some prime ideals in the factorisation of (q), hence there are at most finitely many choices for such I. This completes the proof. □

Remark 10.8. This proof not only shows the finiteness of class numbers, but also provide a recipe for computation. Namely, we can factor all ideals (q) for positive integers $q \leqslant M_K$ to find all ideals with norm q. Then every ideal class is representated by some of these ideals. By eliminating repeated ideal classes and analysing the multiplicative structure, we should in principle understand the ideal class group.

We give one simple example as follows:

Example 10.9. Consider the quadratic field $\mathbb{Q}(i)$. By Proposition 7.2, we know its ring of integers is $\mathcal{O}_K = \mathbb{Z}[i]$. Since $d = -1$, we have $\Delta_K = -4$ by Proposition 7.14. The Minkowski bound for this field is $M_K = \frac{4}{\pi} < 2$. Therefore every ideal class contains an ideal I of norm $N(I) = 1$. By Corollary 8.6, the only possibility is $I = \mathcal{O}_K$. So there is only one ideal class, and $h_K = 1$. By Proposition 9.5 and Corollary 9.7, the ring $\mathcal{O}_K = \mathbb{Z}[i]$ is a PID and UFD. This is consistent with the result in Exercise 1.4. The same argument works for every quadratic field K with $M_K < 2$.

10.2. **Computing class numbers.** We compute class numbers for quadratic fields in some concrete examples.

In Example 10.9, we have seen that, if the Minkowski bound is smaller than 2, then the class number $h_K = 1$ and the class group is a trivial group. In general, we need to use the strategy mentioned in Remark 10.8. More precisely, we need to first factor ideals of the form (q) for all positive integers $q \leqslant M_K$ to find all ideals of norm q, then analyse the relation among these ideals.

There is, in fact, a systematic way to factor any ideal of the form (p) for any prime p in $\mathcal{O}_K$ when K is a quadratic field.

Proposition 10.10. *Let $d \neq 1$ be a square-free integer and $K = \mathbb{Q}(\sqrt{d})$. Then we can factor (2) into prime ideals as follows*

(1) *If $d \not\equiv 1 \pmod 4$, then $(2) = \mathfrak{p}^2$ for some prime ideal $\mathfrak{p}$, which is the only ideal of norm 2;*

(2) *If* $d \equiv 1 \pmod 8$, *then* $(2) = \mathfrak{p}_1\mathfrak{p}_2$ *for distinct prime ideals* $\mathfrak{p}_1$ *and* $\mathfrak{p}_2$, *which are the only ideals of norm* 2*;*

(3) *If* $d \equiv 5 \pmod 8$, *then* (2) *is a prime ideal itself, and there is no ideal of norm* 2.

Proposition 10.11. *Let* $d \neq 1$ *be a square-free integer and* $K = \mathbb{Q}(\sqrt{d})$. *For any odd prime* p, *we can factor* (p) *into prime ideals as follows*

(1) *If* $p \mid d$, *then* $(p) = \mathfrak{p}^2$ *for some prime ideal* $\mathfrak{p}$, *which is the only ideal of norm* p*;*

(2) *If* $(\frac{d}{p}) = 1$, *then* $(p) = \mathfrak{p}_1\mathfrak{p}_2$ *for distinct prime ideals* $\mathfrak{p}_1$ *and* $\mathfrak{p}_2$, *which are the only ideals of norm* p*;*

(3) *If* $(\frac{d}{p}) = -1$, *then* (p) *is a prime ideal itself, and there is no ideal of norm* p.

Proof of Propositions 10.10 and 10.11. In both propositions, we can in fact write down the prime ideals in the factorisations explicitly. Parts (1) and (2) can be proved by verifying the mutual inclusions of the two sides of the equations. Part (3) can be proved by showing the quotient ring is a field (hence an integral domain). The details of the proofs are omitted due to limitation of time. This proof is non-examinable. □

If we want to factor (q) for some composite q, we can factor q into primes in $\mathbb{Z}$, say $q = p_1p_2\cdots p_r$, then write $(q) = (p_1)(p_2)\cdots(p_r)$ and factor each (p_i) using Propositions 10.10 and 10.11.

The following examples show how to compute class numbers using the general strategy mentioned above.

Example 10.12. Let $K = \mathbb{Q}(\sqrt{-19})$. We want to compute h_K. Since $d = -19$, we have $\Delta_K = -19$ by Proposition 7.14. The Minkowski bound for this field is $M_K = \frac{2\sqrt{19}}{\pi} < 3$. By Definition 10.6, every ideal class contains an ideal of norm at most 2. By Corollary 8.6, an ideal of norm 1 must be $\mathcal{O}_K$. Since $d = -19 \equiv 5 \pmod 8$, by Proposition 10.10, there is no ideal of norm 2. We conclude that $h_K = 1$. By Proposition 9.5 and Corollary 9.7, the ring $\mathcal{O}_K$ is a PID and UFD when $K = \mathbb{Q}(\sqrt{-19})$.

Example 10.13. Let $K = \mathbb{Q}(\sqrt{-5})$. We want to compute h_K. Since $d = -5$, we have $\Delta_K = -20$ by Proposition 7.14. The Minkowski bound for this field is $M_K = \frac{2\sqrt{20}}{\pi} < 3$. By Definition 10.6, every ideal class contains an ideal of norm at most 2. By Corollary 8.6, an ideal of norm 1 must be $\mathcal{O}_K$. Since $d = -5 \not\equiv 1 \pmod 4$, by Proposition 10.10, $(2) = \mathfrak{p}$ for some prime ideal $\mathfrak{p}$ which is the only ideal of norm 2. Therefore there are at most 2 ideal classes, represented by $\mathcal{O}_K$ and $\mathfrak{p}$. We still need to know whether they are the same ideal class or distinct ideal classes.

Assume $\mathcal{O}_K$ and $\mathfrak{p}$ are in the same ideal class, then $\mathfrak{p}$ is a principal ideal. Say, $\mathfrak{p} = (\alpha)$ for some $\alpha \in \mathcal{O}_K$. By Proposition 7.2, we can write $\alpha = a + b\sqrt{-5}$ for some $a, b \in \mathbb{Z}$. By Proposition 8.9, we know that $|N(\alpha)| = N((\alpha)) = 2$, hence $N(\alpha) = \pm 2$. By Example 6.18, we know that $N(\alpha) = a^2 + 5b^2$. Therefore we have $a^2 + 5b^2 = \pm 2$ for some $a, b \in \mathbb{Z}$. This equation has no integer solutions. Contradiction. It follows that $\mathfrak{p}$ cannot be a principal ideal. By Corollary 9.6, $\mathfrak{p}$ and $\mathcal{O}_K$ are in different ideal classes, hence $\mathcal{O}_K$ does have two distinct ideal classes. We conclude that $h_K = 2$ for $K = \mathbb{Q}(\sqrt{-5})$.

Example 10.14. Let $K = \mathbb{Q}(\sqrt{10})$. We want to compute h_K. Since $d = 10$, we have $\Delta_K = 40$ by Proposition 7.14. The Minkowski bound for this field is $M_K = \frac{\sqrt{40}}{2} < 4$. By Definition 10.6, every ideal class contains an ideal of norm at most 3. By Corollary 8.6, an ideal of norm 1 must be $\mathcal{O}_K$. Since $d = 10 \not\equiv 1 \pmod 4$, by Proposition 10.10, $(2) = \mathfrak{p}_0^2$ for some prime ideal $\mathfrak{p}$ which is the only ideal of norm 2. Since $(\frac{10}{3}) = (\frac{1}{3}) = 1$, by Proposition 10.11, $(3) = \mathfrak{p}_1\mathfrak{p}_2$ for prime ideals $\mathfrak{p}_1$ and $\mathfrak{p}_2$ which are the only ideals of norm 3. Therefore we have at most 4 ideal classes, represented by

$\mathcal{O}_K$, $\mathfrak{p}_0$, $\mathfrak{p}_1$ and $\mathfrak{p}_2$. However, some of them might be in the same ideal class. So we still need to understand their relations.

We first show that $\mathfrak{p}_0$ is not a principal ideal, thus $\mathcal{O}_K$ and $\mathfrak{p}_0$ are in two different ideal classes. If $\mathfrak{p}_0 = (\alpha)$ for some $\alpha \in \mathcal{O}_K$. By Proposition 7.2, we can write $\alpha = a + b\sqrt{10}$ for some $a, b \in \mathbb{Z}$. By Proposition 8.9, we know that $|N(\alpha)| = N((\alpha)) = 2$, hence $N(\alpha) = \pm 2$. By Example 6.18, we know that $N(\alpha) = a^2 - 10b^2$. Therefore we have $a^2 - 10b^2 = \pm 2$ for some $a, b \in \mathbb{Z}$. This would imply $a^2 \equiv \pm 2 \pmod 5$, hence either 2 or -2 must be a quadratic residue modulo 5. However, $(\frac{2}{5}) = (\frac{-2}{5}) = -1$. Contradiction. It follows that $\mathfrak{p}_0$ cannot be a principal ideal. Therefore we have at least two distinct ideal classes, given by $\overline{\mathcal{O}_K}$ and $\overline{\mathfrak{p}_0}$.

Finally we analyse $\mathfrak{p}_1$ and $\mathfrak{p}_2$. We will show that they are in the same ideal class as $\mathfrak{p}_0$. For this purpose we look at $\alpha = 2 + \sqrt{10} \in \mathcal{O}_K$. By Example 6.18, $N(\alpha) = -6$. By Proposition 7.2, $N((\alpha)) = |N(\alpha)| = 6$. By Corollary 8.5, we know $6 \in (\alpha)$, hence $(6) \subseteq (\alpha)$. By Corollary 8.15, we can find some ideal I such that $(6) = I(\alpha)$. By Theorem 8.16, the ideal (6) has a unique factorisation into finitely many prime ideals. Indeed, we can find it as $(6) = (2)(3) = \mathfrak{p}_0^2\mathfrak{p}_1\mathfrak{p}_2$. Since (α) is a factor of (6), it must be the product of some prime ideals in the factorisation of (6). On the other hand, $N((\alpha)) = 6$, so it has to be the product of an ideal of norm 2 and an ideal of norm 3, i.e., $(\alpha) = \mathfrak{p}_0\mathfrak{p}_1$ or $(\alpha) = \mathfrak{p}_0\mathfrak{p}_2$. If the first case happens, then the ideal classes $\overline{\mathfrak{p}_1} = \overline{\mathfrak{p}_0}^{-1}$ in the ideal class group because (α) is a principal ideal. Similarly from $(2) = \mathfrak{p}_0^2$ and $(3) = \mathfrak{p}_1\mathfrak{p}_2$, we also know $\overline{\mathfrak{p}_0}^{-1} = \overline{\mathfrak{p}_0}$ and $\overline{\mathfrak{p}_2} = \overline{\mathfrak{p}_1}^{-1} = \overline{\mathfrak{p}_0}$. It follows $\overline{\mathfrak{p}_0} = \overline{\mathfrak{p}_1} = \overline{\mathfrak{p}_2}$. If the second case happens, then we can prove the same result by switching the subscripts in $\mathfrak{p}_1$ and $\mathfrak{p}_2$. Hence the only distinct ideal classes are the ones represented by $\mathcal{O}_K$ and $\mathfrak{p}_0$. We conclude $h_K = 2$ for $K = \mathbb{Q}(\sqrt{10})$.

Exercise Sheet 10

This sheet is NOT due in the lecture on Tuesday 9th December, and will be discussed in an exercise class in the same week.

Exercise 10.1. *Some computation of class numbers.*

(1) Compute the class number of $K = \mathbb{Q}(\sqrt{2})$.

(2) Compute the class number of $K = \mathbb{Q}(\sqrt{6})$. (Hint: what is norm of $I = (2 + \sqrt{6})$?)

(3) Compute the class number of $K = \mathbb{Q}(\sqrt{-13})$. (Hint: at some point you need to explain why the only ideal of norm 4 is the principal ideal (2).)

Exercise 10.2. *Fermat's two square problem (revisited).*

Let p be a positive prime such that $p \equiv 1 \pmod 4$.

(1) Show that there exists some $u \in \mathbb{Z}$, such that $u^2 + 1 \equiv 0 \pmod p$.

(2) Consider the lattice $L = \{m_1e_1 + m_2e_2 \mid m_1, m_2 \in \mathbb{Z}\}$ where $e_1 = (1, u)$ and $e_2 = (0, p)$. Compute the volume of its fundamental domain.

(3) Show that the disk $D = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < \frac{3}{2}p\}$ contains at least one non-zero point $(a, b) \in L$. Show that $0 < a^2 + b^2 < 2p$.

(4) Use the generators of L to show $b \equiv ua \pmod p$, then show $a^2 + b^2 \equiv 0 \pmod p$.

(5) Conclude from parts (3) and (4) that $p = a^2 + b^2$.

Exercise 10.3. *Minkowski bound for real quadratic fields.*

Supply the proofs of Propositions 10.4 and 10.5 in the following steps.

(1) For any real numbers $x, y \in \mathbb{R}$, show that $|xy| \leqslant \frac{1}{4}(|x| + |y|)^2$.

(2) Let $r = \left(2N(I)\,|\Delta_K|^{\frac{1}{2}}\right)^{\frac{1}{2}}$. Show that the square $S = \{(x, y) \in \mathbb{R}^2 \mid |x| + |y| \leqslant r\}$ contains at least one non-zero point in L_I. (Hint: Proposition 9.14.)

(3) Reinterpret the result in part (2) as follows: there exists some non-zero $\alpha = a + b\sqrt{d} \in I$, such that for $x = a + b\sqrt{d}$ and $y = a - b\sqrt{d}$, we have $|x| + |y| \leqslant r$.

(4) Use parts (1) and (3) to show that $|N(\alpha)| = |xy| \leqslant \frac{1}{4}r^2 = \frac{1}{2}N(I)\,|\Delta_K|^{\frac{1}{2}}$. This proves Proposition 10.4.

(5) Prove Proposition 10.5. (Hint: almost identical to the proof of Proposition 10.3.)

Exercise 10.4. *Review and reinforce knowledge.*

If you have any questions, or you want to see more examples of anything in this course, please write them down.

Solutions to Exercise Sheet 10

Solution 10.1. *Some computation of class numbers.*

(1) We have $d = 2$, hence $\Delta_K = 4d = 8$, and $M_K = \frac{1}{2}\sqrt{8} = \sqrt{2} < 2$. Therefore every ideal class contains an ideal of norm 1, which must be $\mathcal{O}_K$. It follows that $h_K = 1$.

(2) We have $d = 6$, hence $\Delta_K = 4d = 24$, and $M_K = \frac{1}{2}\sqrt{24} = \sqrt{6} < 3$. Therefore every ideal class contains an ideal of norm 1 or 2. An ideal of norm 1 must be $\mathcal{O}_K$. By Proposition 10.10, since $d \not\equiv 1 \pmod 4$, we have $(2) = \mathfrak{p}^2$ and $\mathfrak{p}$ is the only ideal of norm 2. Therefore every ideal class contains $\mathcal{O}_K$ or $\mathfrak{p}$.

It remains to determine whether $\mathcal{O}_K$ and $\mathfrak{p}$ belong to the same ideal class, or equivalently, whether $\mathfrak{p}$ is a principal ideal. Since $\mathfrak{p}$ is the only ideal of norm 2, if we can find a principal ideal (α) of norm 2, then $\mathfrak{p} = (\alpha)$ is a principal ideal. If we assume $\alpha = a + b\sqrt{6}$, then $N((\alpha)) = |N(\alpha)| = |a^2 - 6b^2|$. Hence $N((\alpha)) = 2$ if and only if $a^2 - 6b^2 = \pm 2$. We observe that $a = 2$ and $b = 1$ satisfy $a^2 - 6b^2 = -2$. Therefore the norm of the principal ideal $(2 + \sqrt{6})$ is 2. By the above analysis we know that $\mathfrak{p} = (2 + \sqrt{6})$ is a principal ideal, hence $\mathcal{O}_K$ and $\mathfrak{p}$ are in the same ideal class. It follows that $h_K = 1$.

(3) We have $d = -13$, hence $\Delta_K = 4d = -52$, and $M_K = \frac{2}{\pi}\sqrt{52} < 5$. Therefore every ideal class contains an ideal of norm 1, 2, 3 or 4. An ideal of norm 1 must be $\mathcal{O}_K$. By Proposition 10.10, since $d \not\equiv 1 \pmod 4$, we have $(2) = \mathfrak{p}^2$ where $\mathfrak{p}$ is the only ideal of norm 2. By Proposition 10.11, since $(\frac{-13}{3}) = (\frac{-1}{3}) = -1$, (3) itself is a prime ideal and there is no ideal of norm 3. By the proof of Theorem 10.7, every ideal of norm 4 must be the product of some prime factors of the principal ideal (4). We realise that $(4) = (2)(2) = \mathfrak{p}^4$, hence the only ideals which divide (4) are $\mathfrak{p}^i$ for $0 \leqslant i \leqslant 4$. Since $N(\mathfrak{p}) = 2$, by Lemma 10.2, the only one among them which has norm 4 is $\mathfrak{p}^2 = (2)$. In other words, the ideal of norm 4 is (2). So we conclude that every ideal class contains an ideal among $\mathcal{O}_K$, $\mathfrak{p}$ and (2).

It is clear that (2) is a principal ideal, hence is in the same ideal class as $\mathcal{O}_K$. We claim that $\mathfrak{p}$ is not a prime ideal. If $\mathfrak{p} = (\alpha)$ for some non-zero $\alpha \in \mathcal{O}_K$, we assume $\alpha = a + b\sqrt{-13}$, then $N((\alpha)) = |N(\alpha)| = |a^2 + 13b^2|$. On the other hand $N((\alpha)) = N(\mathfrak{p}) = 2$, hence $a^2 + 13b^2 = \pm 2$. It is clear that $a^2 + 13b^2 = -2$ has no integer solutions, as the left-hand side is non-negative. It is also easy to see that $a^2 + 13b^2 = 2$ has no integer solutions, since $a^2 \leqslant 2$ implies $a^2 = 0$ or 1, and $13b^2 \leqslant 2$ implies $b^2 = 0$, which cannot add up to 2. We conclude that $\mathfrak{p}$ is not a principal ideal, hence it is not in the same ideal class as $\mathcal{O}_K$. Therefore $h_K = 2$.

Solution 10.2. *Fermats two square problem (revisited).*

(1) Since $p \equiv 1 \pmod 4$, -1 is a quadratic residue modulo p. It follows that there exists some $u \in \mathbb{Z}$, such that $u^2 \equiv -1 \pmod p$; or equivalently, $u^2 + 1 \equiv 0 \pmod p$.

(2) Assume the fundamental domain is T, then

$$\operatorname{vol}(T) = \left|\det \begin{pmatrix} 1 & 0 \\ u & p \end{pmatrix}\right| = p.$$

(3) The volume of the disk is $\operatorname{vol}(D) = \pi \cdot \frac{3}{2}p = \frac{3}{2}\pi p > 4p = 4\operatorname{vol}(T)$. By Theorem 9.11, D contains at least one non-zero point in L, say $(a, b) \in L$. Since a and b are not simultaneously zero, we have $a^2 + b^2 > 0$. On the other hand $(a, b) \in D$ implies $a^2 + b^2 < \frac{3}{2}p < 2p$.

(4) Since $(a, b) \in L$, we have that $(a, b) = m_1(1, u) + m_2(0, p)$ for some $m_1, m_2 \in \mathbb{Z}$. Therefore $a = m_1$ and $b = m_1u + m_2p = ua + pm_2 \equiv ua \pmod p$. It follows that $a^2 + b^2 \equiv a^2 + u^2a^2 = a^2(u^2 + 1) \equiv 0 \pmod p$, where the last congruence is due to part (1).

(5) From part (4) we know that $a^2 + b^2$ is a multiple of p, while within the range given in part (3), the only multiple of p is p itself. Hence $a^2 + b^2 = p$.

Solution 10.3. *Minkowski bound for real quadratic fields.*

(1) The inequality $|xy| \leqslant \frac{1}{4}(|x| + |y|)^2$ is equivalent to $4|xy| \leqslant (|x| + |y|)^2$, which is further equivalent to $(|x|+|y|)^2-4|xy| \geqslant 0$. However the left-hand side is $|x|^2+2|xy|+|y|^2-4|xy| = |x|^2 - 2|xy| + |y|^2 = (|x| - |y|)^2 \geqslant 0$. Hence the inequality holds.

(2) By Proposition 9.14, the volume of the fundamental domain is $\operatorname{vol}(T_I) = N(I)|\Delta_K|^{\frac{1}{2}}$. On the other hand, the volume of the square S is given by $\operatorname{vol}(S) = 2r^2 = 4N(I)|\Delta_K|^{\frac{1}{2}} = 4\operatorname{vol}(T_I)$. By Corollary 9.12, S contains at least one non-zero point in L_I.

(3) By part (2) and the definition of L_I in Proposition 9.14, S contains a non-zero point in L_I, which is given by $(a + b\sqrt{d}, a - b\sqrt{d})$ for some non-zero $\alpha = a + b\sqrt{d} \in I$. We write $x = a + b\sqrt{d}$ and $y = a - b\sqrt{d}$, then by the definition of S we have $|x| + |y| \leqslant r$.

(4) For the α chosen in part (3), we have $N(\alpha) = a^2 - b^2d = (a + b\sqrt{d})(a - b\sqrt{d}) = xy$. Hence $|N(\alpha)| = |xy| \leqslant \frac{1}{4}(|x| + |y|)^2 \leqslant \frac{1}{4}r^2 = \frac{1}{2}N(I)|\Delta_K|^{\frac{1}{2}}$, in which the first inequality follows from part (1) and the second inequality follows from part (3).

(5) By Theorem 9.2, the ideal class $\mathcal{C}$ has an inverse in the ideal class group. We denote this inverse ideal class by $\overline{J}$ where J is any representative. Then by part (4) (which is Proposition 10.4), there exists a non-zero element $\beta \in J$ such that $|N(\beta)| \leqslant \frac{1}{2}N(J)\,|\Delta_K|^{\frac{1}{2}}$. Since we have $(\beta) \subseteq J$, there exists some ideal I such that $IJ = (\beta)$ by Corollary 8.15. Since the ideal class containing (β) is the identity element in the ideal class group, $\overline{I}$ and $\overline{J}$ are inverse of each other, hence I is an ideal in $\mathcal{C}$. It remains to show $N(I)$ satisfies the given bound.

By Lemma 10.2 and Proposition 8.9, we have the following calculation

$$N(I)N(J) = N(IJ) = N((\beta)) = |N(\beta)| \leqslant \frac{1}{2}N(J)\,|\Delta_K|^{\frac{1}{2}}.$$

Since $N(J)$ is a positive integer by Proposition 8.3, we cancel it to get $N(I) \leqslant \frac{1}{2}\,|\Delta_K|^{\frac{1}{2}}$ as required.