Number Theory I

E. Kowalski

Version of October 18, 2024
kowalski@math.ethz.ch

Contents

Preface

These are lecture notes for a first course in Number Theory. The contents are entirely standard, with an emphasis on keeping algebraic and analytic aspects as intertwined as they should be, and on encouraging an approach which uses computer software for various experiments to "guess" certain results (or marvel at the weird unpredictable facts that concrete numbers are made of).

Zürich, October 18, 2024

Acknowledgments. The first draft of these notes was prepared for the course "Number Theory I" that I taught at ETH Zürich during the Fall Semester 2024.

The computer experiments were performed mostly with PARI/GP (see [17]) and MAGMA (see [16]). Some basic scripts are listed in Appendix A.

Thanks to the students of the course for their interest and corrections, especially M. Verzasconi and J. Mo.

Prerequisites

The basic requirements for this text are standard introductory graduate courses in algebra, real analysis (Lebesgue integration theory) and complex analysis. In particular, the theory of finite fields and basic Galois theory will play an important role.

Notation

Number theory has a very rich history, and the statements involved remain often completely accessible, even after decades of work. However, some phenomena which are discovered at certains points in history may be re-interpreted, strengthened, etc, in such a way that the most natural version of a result cannot really be attributed to the first discoverer, *except in spirit*. In this case, we will sometimes give attributions with a dagger exponent, for instance Kronecker† for Theorem 1.4.2 – this indicates that some crucial insight came from the indicated author, even if he or she could not state or prove the stronger version which we display.

We will use the following notation:

(1) For a set X, $|\mathrm{X}| \in [0, +\infty]$ denotes its cardinal, with $|\mathrm{X}| = \infty$ if X is infinite. There is no distinction in this text between the various infinite cardinals.

(2) For any integer $n \geqslant 0$, we sometimes denote $[n] = \{1, \ldots, n\}$; if $n = 0$, this is the empty set. More generally, we write $[n; m] = \{n, n+1, \ldots, m\}$ for any integers $n \leqslant m$ in $\mathbf{Z}$.

(3) For subsets Y_1 and Y_2 of an arbitrary set X, we denote by $\mathrm{Y}_1 - \mathrm{Y}_2$ the difference set, i.e., the set of elements $x \in \mathrm{Y}_1$ such that $x \notin \mathrm{Y}_2$.

(4) If X is a set and f, g two complex-valued functions on X, then we write synonymously $f = \mathrm{O}(g)$ or $f \ll g$ to say that there exists a constant $\mathrm{C} \geqslant 0$ (sometimes called an "implied constant") such that $|f(x)| \leqslant \mathrm{C}g(x)$ for all $x \in \mathrm{X}$. Note that this implies that in fact $g \geqslant 0$. We also write $f \asymp g$ to indicate that $f \ll g$ and $g \ll f$.

(5) We write $a \mid b$ for the divisibility relation "a divides b"; we denote by (a, b) the gcd of two integers a and b, and by $[a, b]$ their lcm.

(6) If p is a prime number and $r \in \mathbf{Q}^\times$ is a rational number, we write $v_p(r)$ for the p-adic valuation of r: when r is written as a ratio a/b of coprime integers, $v_p(r)$ is the difference $n - m$ where n is the power of p dividing a and b the power of p dividing b. This function satisfies the relation

$$v_p(rs) = v_p(r) + v_p(s),$$

for any pairs (r, s) of non-zero rational numbers.

(7) Usually, the variable p will always refer to prime numbers. In particular, a series $\sum_p (\cdots)$ refers to a series over primes (summed in increasing order, in case it is not known to be absolutely convergent), and similarly for a product $\prod_p (\cdots)$ over primes.

(8) We denote by $\mathbf{F}_p$ the finite field $\mathbf{Z}/p\mathbf{Z}$, for p prime, and more generally by $\mathbf{F}_q$ a finite field with q elements, where $q = p^n$, $n \geqslant 1$, is a power of p.

(9) For a complex number z, we write $e(z) = e^{2i\pi z}$. If $q \geqslant 1$ and $x \in \mathbf{Z}/q\mathbf{Z}$, then $e(x/q)$ is then well-defined by taking any representative of x in $\mathbf{Z}$ to compute the exponential.

(10) If $q \geqslant 1$ and $x \in \mathbf{Z}$ (or $x \in \mathbf{Z}/q\mathbf{Z}$) is an integer which is coprime to q (or a residue class invertible modulo q), we sometimes denote by $\bar{x}$ the inverse class such that $x\bar{x} = 1$ in $\mathbf{Z}/q\mathbf{Z}$. This will always be done in such a way that the modulus q is clear from context, in the case where x is an integer.

CHAPTER 1

Introduction

1.1. What is number theory?

1.2. Looking at sums of two squares

We will start by considering sums of two squares of (positive) integers. This is an extremely classical subject, but there are still many research questions connected to it, some of which will be mentioned below. However, we begin with a natural beginning for current students: we experiment on a computer to look at these numbers.

Thus, one way or another (with PARI/GP or MAGMA for instace, or other suitable software), we might first get a list of the first integers which are sums of two squares; here is the beginning of that list with 200 numbers:

$$2, 5, 8, 10, 13, 17, 18, 20, 25, 26, 29, 32, 34, 37, 40, 41, 45, 50, 52, 53, 58, 61, 65, 68, 72, 73, 74,$$
$$80, 82, 85, 89, 90, 97, 98, 100, 101, 104, 106, 109, 113, 116, 117, 122, 125, 128, 130, 136, 137, 145,$$
$$146, 148, 149, 153, 157, 160, 162, 164, 169, 170, 173, 178, 180, 181, 185, 193, 194, 197, 200, 202,$$
$$205, 208, 212, 218, 221, 225, 226, 229, 232, 233, 234, 241, 242, 244, 245, 250, 257, 260, 261, 265,$$
$$269, 272, 274, 277, 281, 288, 289, 290, 292, 293, 296, 298, 305, 306, 313, 314, 317, 320, 325,$$
$$328, 333, 337, 338, 340, 346, 349, 353, 356, 360, 362, 365, 369, 370, 373, 377, 386, 388, 389,$$
$$392, 394, 397, 400, 401, 404, 405, 409, 410, 416, 421, 424, 425, 433, 436, 442, 445, 449, 450,$$
$$452, 457, 458, 461, 464, 466, 468, 477, 481, 482, 485, 488, 490, 493, 500, 505, 509, 512, 514, 520,$$
$$521, 522, 530, 533, 538, 541, 544, 545, 548, 549, 554, 557, 562, 565, 569, 577, 578, 580, 584, 585,$$
$$586, 592, 593, 596, 601, 605, 610, 612, 613, 617, 625, 626, 628, 629.$$

REMARK 1.2.1. There is already a noteworthy feature in this list: we can be sure that we have constructed the first sums of two squares, because of the lack of cancellation: if we wanted to know the first positive integers which are (say) of the form $a^2 - b^3$, the difficulty would be that even if, for a given $k \geqslant 1$, we check numerically that we cannot express k in this form with a and b bounded by some finite limit, it could be that some much bigger choices of a and b work.

As a concrete illustration, we will prove later on that the equation

$$x^2 - dy^2 = 1$$

always has a solution in non-negative integers if $d \geqslant 0$ is an integer which is not a square (a fact which was known to early Indian mathematicians like Brahmagupta and Bhaskar II; the established name "Pell equation" is a misnomer, due to Euler, alas). Even for small values of d, the sizes of the smallest solution (x, y) may be extremely large (we will also see that there are infinitely many solutions). For instance, for $d = 61$, one gets that the smallest solution is

$$(x, y) = (1766319049, 226153980).$$

The origin of these large solutions is understood in principle – but many questions remain open, as we will discuss later.

Staring at this list, or similar data, leads to many questions. This way (maybe), Fermat discovered the following remarkable fact, which is truly the tip of an iceberg.

THEOREM 1.2.2 (Fermat). *Any prime number p which is congruent to 1 modulo 4 is the sum of two squares of positive integers. Moreover, in a representation $p = a^2 + b^2$, the couple $\{a, b\}$ is unique; in other words, a and b are unique, except that they can be exchanged.*

EXAMPLE 1.2.3. In some cases, the representation of a prime as a sum of two squares is obvious: for instance, the integer $p = 2^{2^4} + 1 = 65537$ is prime[1] and congruent to 1 modulo 4, and $p = (2^8)^2 + 1^2$. But if we vary p, there is no apparent regularity in the values of a and b: for instance, the next few primes larger than 65537 which are congruent to 1 modulo 4 are

$$65557, \qquad 65581, \qquad 65609,$$

for which the representations are

$$65557 = 71^2 + 246^2, \qquad 65581 = 166^2 + 195^2, \qquad 65609 = 40^2 + 253^2.$$

One *can* say some things about a and b, but that requires a slightly different perspective, moving from the deterministic to the probabilistic.

We will see in these lectures a number of different proofs of Fermat's Theorem; according to one of these, it will become "obvious", and become part of a much larger picture as a consequence of basic facts of algebraic number theory. Other proofs will be equally well-motivated from slightly different perspectives, and some will even give a "formula" for a and b. But we begin with another proof of the existence of the representation (it doesn't give the uniqueness) – most people consider it as a really bad proof, and it is worth looking at just to understand why one would think so...[2]

This proof is due to Heath–Brown [11], and was simplified and popularized by Zagier [21]. The key ingredient is the following fact, where we recall that for an arbitrary set X, a map $f\colon \mathrm{X} \to \mathrm{X}$ is called an *involution* if $f(f(x)) = x$ for all $x \in \mathrm{X}$ (i.e., if $f \circ f$ is the identity map of X), and that (for any f), an element $x \in \mathrm{X}$ is called a *fixed point* if $f(x) = x$.

LEMMA 1.2.4. *Let* X *be a finite set and* $f\colon \mathrm{X} \to \mathrm{X}$ *an involution. The number of fixed points of f has the same parity as the size of* X*. In particular, if* $|\mathrm{X}|$ *is odd, then f has* at least *one fixed point, and if f has a* unique *fixed point, then* $|\mathrm{X}|$ *is odd.*

PROOF. Because f is an involution, the sets of the form $\mathrm{P}_x = \{x, f(x)\}$ with $x \in \mathrm{X}$ form a partition of X: they are either equal or disjoint. Indeed, if $\{x, f(x)\} \cap \{y, f(y)\}$ is not empty (the sets are not disjoint), then a common element z is either equal to $x = y$ or $x = f(y)$; in the first case, the sets are clearly the same, and in the second, the two sets are $\{x, f(x)\}$ and $\{x, y\} = \{x, f(f(y))\} = \{x, f(x)\}$. The fixed points of f correspond to those sets $\{x, f(x)\}$ with a single element. Writing

$$|\mathrm{P}_x| = (\text{number of fixed points}) + 2(\text{number of sets } \mathrm{P}_x \text{ with two elements}),$$

[1] In fact, Fermat conjectured that $2^{2^k} + 1$ is *always* prime – this is false, already $2^{2^5} + 1 = 4294967297 = 641 \cdot 6700417$ is not prime, as found out by Euler; it is not known if infinitely many such integers are prime, but this not expected to be true.

[2] Although Elsholtz [**?**] has discussed the underlying ideas with the aim of making them less miraculous.

the result follows. □

Proof of existence in Theorem 1.2.2. Let p be a prime number congruent to 1 modulo 4. Consider the set

$$\mathrm{X} = \{(a,b,c) \mid a,b,c \text{ positive integers and } a^2+4bc = p\}.$$

Note that b and c play symmetrical roles in X, and that a representation $p = a^2+b^2$ (with a, b positive integers) corresponds to an element (a,b,b) of X, so the existence question amounts to showing that the map $f\colon \mathrm{X} \to \mathrm{X}$ defined by $f(a,b,c) = (a,c,b)$ has at least one fixed-point. Since the size of X is finite (because of the requirement that a, b and c are non-negative), this will be true if $|\mathrm{X}|$ is odd according to the lemma. Now, to establish this last fact, we use the lemma again, applied to the map $g\colon \mathrm{X} \to \mathrm{X}$ defined by

$$g(a,b,c) = \begin{cases} (a+2c, c, b-a-c) & \text{if } a < b-c \\ (2b-a, b, a-b+c) & \text{if } b-c < a < 2b \\ (a-2b, a-b+c, b) & \text{if } a > 2b. \end{cases} \tag{1.1}$$

Precisely, we claim that (1) the map g is well-defined (i.e., the various right-hand side formulas always apply and always lead to an element of X if $(a,b,c) \in \mathrm{X}$); (2) the map g is an involution; (3) the map g has a unique fixed point. Assuming this, the conclusion that $|\mathrm{X}|$ is odd follows from the lemma!

All of these things are elementary.[3] Let us check a few. For instance, for (3), note that a fixed point (a,b,c) can only come from the second case in (1.1) (since b and c are positive), so it must come from an equation

$$(a,b,c) = (2b-a, b, a-b+c),$$

which holds if and only if $a = b$. But the fact that $(a,a,c) \in \mathrm{X}$ means that $a^2+4ac = p$, which because p is prime is only possible if $a = 1$, and then if $1+4c = p$, which has a solution *because* p is congruent to 1 modulo 4.

As far as the fact that g is well-defined, first of all note that if $(a,b,c) \in \mathrm{X}$, it is not possible that $a = 2b$ (it would imply that $p = a^2+4bc$ is divisible by 4) or $a = b-c$ (it would mean that $p = (b-c)^2+4bc = (b+c)^2$), so that $(a,b,c) \in \mathrm{X}$ falls necessarily in one of the three cases on the right-hand side of the definition. Then we have the algebraic identities

$$(a+2c)^2 + 4c(b-a-c) = (2b-a)^2 + 4b(a-b+c) = (a-2b)^2 + 4(a-b+c)b,$$

which show $g(a,b,c) \in \mathrm{X}$, since the conditions on the right-hand side also ensure that the images have positive integral coordinates.

Finally, we check that g is an involution. Let $x = (a,b,c) \in \mathrm{X}$, and assume it satisfies $a < b-c$, so $g(x)$ is given by the first case in (1.1). Let $(\alpha,\beta,\gamma) = g(x)$. We then see that $\beta-\gamma = c-(b-a-c) = a+2c-b < a+2c = \alpha$, and $2\beta = 2c < 2c+a = \alpha$, so $g(g(x))$ must be computed by the third case of (1.1). This gives

$$g(g(x)) = (\alpha-2\beta, \alpha-\beta+\gamma, \beta) = (a+2c-2c, a+2c-c+b-a-c, c) = (a,b,c) = x.$$

We leave the other two cases of the formula $g(g(x)) = x$ to check... □

[3] Leading Zagier to state that this was a "one-line proof".

1.3. Looking at primes

One of the first non-trivial facts about number theory was Euclid's proof that there are infinitely prime numbers (or, to phrase it according to his style: for any given prime number, there is one which is strictly greater). As soon as number theory was studied a bit deeper, there arose many questions with the intent of making this fact more precise. In particular, can one say *how many* primes there are below a certain number $x \geqslant 1$? This quantity is classically denoted $\pi(x)$, i.e.

$$\pi(x) = \sum_{p \leqslant x} 1$$

to display the convention which will be used throughout this book that sums or products (or set descriptions) involving a variable p will always assume (unless specified otherwise) that this is restricted to prime values.[4]

It is worth spending some time trying to make some conjecture or guess about the growth of $\pi(x)$ as $x \to +\infty$: clearly, $\pi(x) \leqslant x$ (because primes are integers – an observation that can be surprisingly efficient!), and Euclid's result is that $\pi(x) \to +\infty$ as $x \to +\infty$, but what is the order of magnitude?

Legendre, in somewhat imprecise way, and Gauss in an extremely sharp display of probabilistic thinking, were led to the conjecture that $\pi(x)$ should, for x large, be of the order of $x/\log(x)$. A beautiful argument of Chebychev succeeded in showing that this is not far from the truth, using ingenious elementary methods well before the more precise Prime Number Theorem was established in 1896 (as we will discuss later).

THEOREM 1.3.1 (Chebychev). *There exist positive real numbers $c_1 > 0$ and $c_2 > 0$ such that*

$$c_1 \frac{x}{\log(x)} \leqslant \pi(x) \leqslant c_2 \frac{x}{\log(x)}$$

for all $x \geqslant 2$.

Neither side of these inequalities is easy, but maybe the lower-bound is the more surprising (since it shows that, in some sense, there are many prime numbers, maybe more than a first naive guess might suggest).

PROOF. The proof below is based on the following fact: for any integer n, the middle binomial coefficient $\binom{2n}{n}$ is a (positive) integer for $n \geqslant 1$.[5] This property, which is "trivial" when the binomial coefficient $\binom{n}{k}$ is interpreted as counting subsets of size k in a set of size n, is much less obvious if one writes down the formula

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2},$$

because it is absolutely not obvious that the square of $n!$ should divide $(2n)!$.

We first use this to prove the upper bound in the theorem. We note for this that if p is a prime number with $n < p \leqslant 2n$, then p divides $(2n)!$ but not $(n!)^2$, and this tells us that

$$\binom{2n}{n} \geqslant \prod_{n < p \leqslant 2n} p.$$

[4] We will see later in practice why it is very convenient to defined this for arbitrary values of $x \geqslant 0$, and not just for integers.

[5] This use of binomial coefficients to "package" the ideas of Chebychev is due to Erdős [6].

However, the binomial coefficient cannot be too large: in fact, since

$$2^{2n} = (1+1)^{2n} = \sum_{k=0}^{2n} \binom{2n}{k},$$

positivity of the terms indicates that

$$\binom{2n}{n} \leqslant 2^{2n}$$

(which also follows immediately from the combinatorial interpretation: the number of subsets of size n in $\{1, \dots, 2n\}$ is certainly bounded by the total number 2^{2n} of all subsets). Combining the two inequalities and taking the logarithm, we obtain the inequality

$$\sum_{n<p\leqslant 2n} \log p \leqslant 2n \log(2).$$

It is now easy to deduce the Chebychev upper-bound, by general principles that we will discuss separately below (dyadic subdivisions and summation by parts), the key point being that $\log p$ is quite close to $\log(2n)$ for $n < p \leqslant 2n$.

Thus we go to the lower-bound, which is more involved. This may not be too surprising, since it amounts to showing that there are *many* prime numbers, which certainly seems difficult, given the fact that they are not specified by any formula. However, the direction of the argument is suggested by the proof of the upper-bound: we might now wish to combine a *lower-bound* for $\binom{2n}{n}$ with an *upper-bound* for $\binom{2n}{n}$ derived from information about the primes. But for such an idea to have a chance to work to prove a lower bound of size $n/\log(n)$, it must be the case that the first part of the proof was not too far from the truth. This means that the prime factorization of $\binom{2n}{n}$ cannot diverge too much from the simple product of the primes between n and $2n$.

Thus we are led to use the fundamental theorem of arithmetic to express $\binom{2n}{n}$ as a product of prime powers, and in then trying to determine or estimate the exponents that appear. Noting that only primes $p \leqslant 2n$ may appear in the factorization (because they have to divide the numerator $(2n)!$ in the factorial formula), we write

$$\binom{2n}{n} = \prod_{p\leqslant 2n} p^{v_p}$$

for some integers $v_p \geqslant 0$. There is a formula for v_p, following from a formula for $v_p(k!)$ for any $k \geqslant 1$ given in Lemma 1.3.2 below: we have

$$v_p = \sum_{j\geqslant 1} \Big(\Big\lfloor \frac{2n}{p^j} \Big\rfloor - 2 \Big\lfloor \frac{n}{p^j} \Big\rfloor \Big),$$

which follows from the lemma applied to $k = 2n$ and then to $k = n$, using the addition formula $v_p(ab) = v_p(a) + v_p(b)$ for all $a, b \geqslant 1$.

To estimate this valuation, we use two observations: first, that $\lfloor y \rfloor - 2\lfloor y/2 \rfloor \leqslant 1$ for all $y \geqslant 0$ (an elementary fact left to the reader), and second, that the sum over j above terminates after at most $\log(2n)/\log(p)$ terms (i.e., for j larger than this, the summand is always 0). It follows that

$$v_p \leqslant \frac{\log(2n)}{\log p}$$

for all primes p. Hence, taking once more the logarithm, we get

$$\log\left(\binom{2n}{n}\right)) = \sum_{p\leqslant 2n} v_p \log(p) \leqslant \pi(2n)\log(2n).$$

We finally need to get a lower bound for $\binom{2n}{n}$, and it should be of exponential size in order to get the right conclusion. This is indeed the case: we have $\binom{2n}{n} \geqslant 2^{2n}/(2n+1)$, because the middle binomial coefficient is the largest among the $2n+1$ values of $\binom{2n}{k}$ which sum to 2^{2n}, so we finally get

$$\pi(2n)\log(2n) \geqslant 2n\log(2) - \log(2n+1).$$

This clearly contains the gist of the Chebychev lower bound. To go from this to the precise statement in Theorem 1.3.1 is again a simple matter of general analytic principles, which we discuss separately later. □

We used the following lemma, which goes back to Legendre, and has quite a few other applications.

LEMMA 1.3.2. *Let $k \geqslant 1$ be an integer. The exponent $v_p(k!)$ such that $p^{v_p(k!)}$ divides $k!$ and $p^{v_p(k!)+1}$ does not is equal to*

$$v_p(k!) = \sum_{j\geqslant 1} \left\lfloor \frac{k}{p^j} \right\rfloor.$$

PROOF. We use one crucial property of the exponents $v_p(n)$ for arbitrary integers $n \geqslant 1$: they satisfy $v_p(mn) = v_p(n) + v_p(m)$ for arbitrary integers n and m. In particular, we have

$$v_p(k!) = \sum_{n=1}^{k} v_p(n).$$

To go further, we insert the following analytic expression for $v_p(n)$: we have

$$v_p(n) = \sum_{\substack{j\geqslant 1\\ p^j|n}} 1$$

(i.e., the exponent $v_p(n)$ is also the number of $j \geqslant 1$ such that p^j divides n). Inserting this in the sum defining $v_p(k!)$, we obtain

$$v_p(k!) = \sum_{n=1}^{k} \sum_{p^j|n} 1$$

and here we exchange the order of the two (finite) sums. We get

$$v_p(k!) = \sum_{j\geqslant 1} \sum_{p^j|n} 1,$$

where the second summation is now over those values of n, with $1 \leqslant n \leqslant k$, for which p^j divides n. So this inner sum counts the integers in the interval from 1 to k which are divisible by p^j: they are the integers p^j, $2p^j$, ..., mp^j, where m is allowed to increase as long as $mp^j \leqslant k$, i.e., m ranges from 1 to the largest integers $m \leqslant k/p^j$. This is the same as saying that $m = \lfloor k/p^j \rfloor$. In other words, we have shown that

$$v_p(k!) = \sum_{j\geqslant 1} \left\lfloor \frac{k}{p^j} \right\rfloor,$$

which concludes the proof. □

REMARK 1.3.3. (1) If we compare Theorem 1.3.1 with Fermat's Theorem, an immediate natural question arises: among the primes (whose density among integers we now have some idea about), how many are $\equiv 1 \pmod 4$? It is a priori not even clear if there are infinitely many of them! The proof of Chebychev's Theorem does not in any obvious way extend to handle this problem, and only by using essential new ideas, going back to Dirichlet and related to harmonic analysis, will we have robust methods available to attack such questions.

(2) One can ask whether the binomial coefficient can be replaced by other quantities to obtain better values of the constants c_1 and c_2. This is a delicate matter! Indeed, one can use for instance the fact that

$$\frac{(30n)!n!}{(15n)!(10n)!(6n)!}$$

is an integer for all $n \geqslant 1$ (by no means a clear fact!) to obtain slightly better estimates, and the question of trying to understand what similar expressions have this property is quite intricate – following observations of Rodriguez-Villegas, it is related to such topics as algebraicity of solutions of hypergeometric equations, and we refer to the paper [**?**] of Soundararajan for further discussion and results.

1.4. Roots of polynomial equations

The last topic for this introductory chapter concerns integral polynomial equations in one variable. Of course, a polynomial $f \in \mathbf{Z}[\mathrm{X}]$ of degree $\geqslant 2$ does not usually have integral or rational roots. The theory of field extensions and Galois theory show that the complex roots of such polynomials have subtle properties, and revealing these is one of the key goals of *algebraic number theory*.

On the other hand, even if $f \in \mathbf{Z}[\mathrm{X}]$ has no rational root, it may well be that it has some roots modulo q for certain integers q. For instance, from a relation like

$$4^2 + 1 = 17,$$

we see that $4 \pmod{17}$ is a "square root of -1" in $\mathbf{Z}/17\mathbf{Z}$. This immediately raises questions, among which maybe the most natural would be: for which $q \geqslant 1$ does there exist such a "modular i"? This, again, turns out to be an extraordinarily rich problem...

To be more precise, we denote by $\mathrm{Z}_f(\mathbf{Z}/q\mathbf{Z})$ the set of $x \in \mathbf{Z}/q\mathbf{Z}$ such that $f(x) = 0 \pmod q$. We first make a reduction: we express q as a product of prime powers, say

$$q = \prod_p p^{v_p}$$

where $v_p \geqslant 0$ and $v_p \geqslant 1$ for finitely many primes only. Then because of the Chinese Remainder Theorem (the fact that the map

$$\mathbf{Z}/q\mathbf{Z} \to \prod_p \mathbf{Z}/p^{v_p}\mathbf{Z}$$

induced by reducing an integer modulo q modulo the different divisors p^{v_p} of q is an isomorphism *of rings*), we have a bijection

$$\mathrm{Z}_f(\mathbf{Z}/q\mathbf{Z}) \to \prod_p \mathrm{Z}_f(\mathbf{Z}/p^{v_p}\mathbf{Z}),$$

which allows us to restrict our attention to those q which are powers of primes.[6] Since the existence of a solution of $f(x) = 0$ modulo p^2 or p^3, or higher powers, certainly implies the existence of a solution modulo p, we will in fact restrict our attention to solutions modulo primes.

This will more fully be justified a bit later by other general principles which, essentially, state that for all but finitely many primes (depending on f), the number of solutions modulo p^k, for any $k \geqslant 2$, is the same as the number of solutions modulo p. Moreover, a good reason *a priori* to focus on primes is that $\mathbf{Z}/p\mathbf{Z}$ is a field, whereas $\mathbf{Z}/p^k\mathbf{Z}$ is not if $k \geqslant 2$, and this has many algebraic advantages. As is customary, the emphasize the fact that $\mathbf{Z}/p\mathbf{Z}$ is a field, we denote it by $\mathbf{F}_p$.

EXAMPLE 1.4.1. (1) Taking $f = \mathrm{X}^2 + 1$ (so we are looking for "modular" versions of the imaginary unit $i \in \mathbf{C}$), here is a list for the first 100 primes[7] of the number of solutions of the equation $x^2 + 1 = 0 \pmod{p}$.

$$(2,1),(3,0),(5,2),(7,0),(11,0),(13,2),(17,2),(19,0),(23,0),(29,2),(31,0),(37,2),$$
$$(41,2),(43,0),(47,0),(53,2),(59,0),(61,2),(67,0),(71,0),(73,2),(79,0),(83,0),$$
$$(89,2),(97,2),(101,2),(103,0),(107,0),(109,2),(113,2),(127,0),(131,0),(137,2),$$
$$(139,0),(149,2),(151,0),(157,2),(163,0),(167,0),(173,2),(179,0),(181,2),(191,0),$$
$$(193,2),(197,2),(199,0),(211,0),(223,0),(227,0),(229,2),(233,2),(239,0),(241,2),$$
$$(251,0),(257,2),(263,0),(269,2),(271,0),(277,2),(281,2),(283,0),(293,2),(307,0),$$
$$(311,0),(313,2),(317,2),(331,0),(337,2),(347,0),(349,2),(353,2),(359,0),(367,0),$$
$$(373,2),(379,0),(383,0),(389,2),(397,2),(401,2),(409,2),(419,0),(421,2),(431,0),$$
$$(433,2),(439,0),(443,0),(449,2),(457,2),(461,2),(463,0),(467,0),(479,0),(487,0),$$
$$(491,0),(499,0),(503,0),(509,2),(521,2),(523,0),(541,2).$$

The reader is already invited to search for patterns and regularities.

(2) Take $f = \mathrm{X}^3 + \mathrm{X} - 1$. We make the same experiment, and we get:

$$(2,0),(3,1),(5,0),(7,0),(11,1),(13,1),(17,1),(19,0),(23,1),(29,1),(31,2),(37,1),$$
$$(41,0),(43,1),(47,3),(53,1),(59,0),(61,1),(67,3),(71,0),(73,1),(79,1),(83,1),$$
$$(89,1),(97,0),(101,0),(103,0),(107,0),(109,0),(113,0),(127,1),(131,3),(137,1),$$
$$(139,1),(149,3),(151,1),(157,0),(163,0),(167,1),(173,3),(179,1),(181,1),(191,0),$$
$$(193,0),(197,1),(199,1),(211,0),(223,1),(227,3),(229,1),(233,0),(239,1),(241,1),$$
$$(251,1),(257,0),(263,1),(269,1),(271,1),(277,1),(281,0),(283,3),(293,3),(307,0),$$
$$(311,0),(313,1),(317,0),(331,1),(337,1),(347,1),(349,3),(353,1),(359,0),(367,1),$$
$$(373,0),(379,3),(383,1),(389,1),(397,0),(401,1),(409,1),(419,0),(421,0),(431,3),$$
$$(433,1),(439,0),(443,0),(449,1),(457,1),(461,1),(463,1),(467,0),(479,0),(487,1),$$
$$(491,1),(499,1),(503,0),(509,1),(521,3),(523,1),(541,0).$$

[6] In other words, we use the fact that for $x \in \mathbf{Z}/q\mathbf{Z}$ corresponding to the family (x_p) modulo p^{v_p}, the value $f(x) \pmod{q}$ corresponds to the family $(f(x_p))$.

[7] It is sometimes useful to remember that 541 is the hundredth prime, 1223 the two-hundredth, 10007 the first prime larger than 10000, and to know a few other prime numbers, especially 163, 691 and 196561.

Again, what patterns (if any) do seem to emerge? And how does this compare with the first example?

(3) Here we take $f = \mathrm{X}^8 - 16$ and list only the number of roots.

$$1, 2, 4, 2, 2, 4, 8, 2, 2, 4, 2, 4, 8, 2, 2, 4, 2, 4, 2, 2, 8, 2, 2, 8, 8, 4, 2, 2, 4, 8, 2, 2, 8, 2,$$
$$4, 2, 4, 2, 2, 4, 2, 4, 2, 8, 4, 2, 2, 2, 2, 4, 8, 2, 8, 2, 8, 2, 4, 2, 4, 8, 2, 4, 2, 2, 8, 4, 2, 8,$$
$$2, 4, 8, 2, 2, 4, 2, 2, 4, 4, 8, 8, 2, 4, 2, 8, 2, 2, 8, 8, 4, 2, 2, 2, 2, 2, 2, 2, 4, 8, 2, 4, 2, 4,$$
$$2, 8, 2, 8, 2, 8, 2, 8, 2, 4, 8, 2, 2, 8, 2, 2, 4, 2, 4, 8, 4, 2, 2, 4, 4, 2, 2, 4, 2, 2, 2, 4, 8, 8,$$
$$4, 2, 4, 8, 2, 4, 2, 2, 4, 2, 4, 8, 2, 2, 4, 8, 2, 2, 2, 2, 2, 8, 8, 4, 2, 8, 2, 2, 8, 2, 2, 4, 8, 4,$$
$$2, 4, 2, 8, 2, 8, 2, 4, 2, 4, 2, 2, 4, 8, 2, 4, 4, 2, 8, 2, 8, 2, 2, 4, 2, 8, 8, 4, 8, 2.$$

(4) Now with $f = \mathrm{X}^2 - \mathrm{X} + 41$, again listing only the number of roots.

$$0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 2, 2, 2, 2, 0, 2, 0, 2, 0, 0, 2, 0, 2, 0, 0, 0, 0, 2, 0, 2, 0, 0, 0,$$
$$2, 0, 1, 2, 2, 2, 0, 0, 0, 2, 2, 0, 2, 2, 0, 0, 0, 0, 2, 0, 2, 0, 0, 0, 2, 0, 0, 2, 0, 2, 0, 0, 0, 2,$$
$$0, 0, 2, 2, 2, 2, 2, 0, 2, 0, 2, 2, 2, 0, 0, 2, 0, 0, 2, 2, 0, 0, 0, 2, 0, 2, 2, 0, 0, 2, 0, 2, 0, 2, 0,$$
$$0, 2, 0, 2, 0, 0, 2, 0, 0, 0, 0, 2, 0, 2, 2, 0, 2, 2, 2, 0, 2, 2, 2, 0, 0, 2, 2, 2, 0, 0, 0, 0, 2, 2,$$
$$2, 0, 0, 2, 0, 0, 2, 2, 2, 0, 0, 0, 2, 0, 0, 0, 0, 2, 2, 0, 0, 2, 2, 0, 2, 2, 0, 0, 0, 0, 0, 2, 2, 2, 2,$$
$$2, 2, 2, 0, 2, 2, 2, 0, 2, 2, 2, 0, 2, 0, 2, 2, 2, 0, 2, 0, 2, 2, 0, 2, 0, 0, 0.$$

This last example is related, strangely enough, to the fact that

$$e^{\pi\sqrt{163}} = 262537412640768743.99999999999925007259\ldots$$

is extremely close to an integer (but it isn't one...)

We come back to the general discussion. As in these examples, we first concentrate our attention on the size of $\mathrm{Z}_f(\mathbf{F}_p)$ (a tricky enough problem without trying to address the actual values of the roots when they exist). We denote $\nu_f(p) = |\mathrm{Z}_f(\mathbf{F}_p)|$. There is a fundamental fact which illustrates in the simplest instance a connection between properties of the complex roots (equivalently, of the polynomial f, viewed as a "global" object with integral coefficients) and those modulo primes (where each reduction modulo a prime p is interpreted as giving "local" information); it was discovered by Kronecker [15] in the late 19th Century (though really only proved a bit later by Dedekind).

THEOREM 1.4.2 (Kronecker†). *The "average value" of $\nu_f(p)$ as p ranges over primes is the number of irreducible factors of f over $\mathbf{Q}$, counted without multiplicity. More precisely, if f is irreducible over $\mathbf{Q}$, then we have the limit formula*

$$\lim_{x\to+\infty} \frac{1}{\pi(x)} \sum_{p\leqslant x} \nu_f(p) = 1.$$

In contrast to the results of the previous sections, this theorem does not currently have a simple proof, and it is very doubtful that such a proof could exist (except for rather special polynomials). We will see later that it is really essentially equivalent to the fact that the so-called Dedekind ζ function attached to the splitting field of f has a simple pole at 1.

EXAMPLE 1.4.3. (1) Let $f = \mathrm{X}^2 + 1$. It is relatively easy to compute $\nu_f(p)$ here, because $f(x) = 0$ means that $x^2 = -1 \pmod p$, and is therefore equivalent to saying that x is a fourth root of unity (as it should, since it is a version of i modulo p), and it is not of order 2, *unless* $p = 2$. Excluding this case, in order to have such an element in $\mathbf{F}_p$,

the order $p-1$ of the group must be divisible by 4. But conversely, since $\mathbf{F}_p^\times$ is known to be a *cyclic* group of order $p-1$, we know that it contains a unique cyclic subgroup of any order dividing $p-1$, and thus $p \equiv 1 \pmod 4$ implies the existence of some x with $x^2 = -1 \pmod p$. Therefore we get

$$\nu_f(p) = \begin{cases} 0 & \text{if } p \equiv 3 \pmod 4, \\ 1 & \text{if } p = 2, \\ 2 & \text{if } p \equiv 1 \pmod 4. \end{cases}$$

Comparing with Fermat's Theorem, we see that the primes with $\nu_f(p) = 2$ are exactly those which are sums of two primes. One direction of this equality is easy: if $a^2 + b^2 = p$, then reducing modulo p, we see that $a^2 + b^2 \equiv 0 \pmod p$, and since a is coprime to p (otherwise the left-hand side is larger than p), we get $x^2 = -1 \pmod p$ with $x = b/a \pmod p$ (the inverse being computed modulo p). The other direction would mean a different proof of Fermat's Theorem, where the existence of (a, b) would be deduced from the existence of x. We will see two or three proofs of this kind later on.

(2) The previous example naturally extends to cyclotomic polynomials, those whose roots are roots of unity. Indeed, let $q \geqslant 1$ be an integer and let $f = \Phi_q \in \mathbf{Z}[\mathrm{X}]$ be the q-th cyclotomic polynomial, namely the monic polynomial whose roots are all primitive q-th roots of unity. Concretely, this is

$$\Phi_q = \prod_{\substack{1 \leqslant a < q \\ (a,q)=1}} (\mathrm{X} - e^{2ia\pi/q}),$$

since the $e^{2ia\pi/q}$ with a coprime to q are the primitive q-th roots of unity. Thus, this polynomial has degree $\varphi(q) = |(\mathbf{Z}/q\mathbf{Z})^\times|$ (whis is called the *Euler function*).

If q is prime, then this is the simple explicit polynomial

$$\Phi_q = \mathrm{X}^{q-1} + \cdots + \mathrm{X} + 1,$$

but in general this polynomial can be more surprisingly complicated. For instance

$$\begin{aligned} \Phi_{210} = \mathrm{X}^{48} - \mathrm{X}^{47} + \mathrm{X}^{46} + \mathrm{X}^{43} - \mathrm{X}^{42} + 2\mathrm{X}^{41} - \mathrm{X}^{40} + \mathrm{X}^{39} + \mathrm{X}^{36} - \mathrm{X}^{35} + \mathrm{X}^{34} - \\ \mathrm{X}^{33} + \mathrm{X}^{32} - \mathrm{X}^{31} - \mathrm{X}^{28} - \mathrm{X}^{26} - \mathrm{X}^{24} - \mathrm{X}^{22} - \mathrm{X}^{20} - \mathrm{X}^{17} + \mathrm{X}^{16} - \mathrm{X}^{15} + \mathrm{X}^{14} - \\ \mathrm{X}^{13} + \mathrm{X}^{12} + \mathrm{X}^9 - \mathrm{X}^8 + 2\mathrm{X}^7 - \mathrm{X}^6 + \mathrm{X}^5 + \mathrm{X}^2 - \mathrm{X} + 1 \end{aligned}$$

(where one sees coefficients different from 0, -1 or 1; but note that the first occurrence of such a coefficient is for Φ_{105}).

The same properties of finite cyclic groups used in the first example (which is the same as the case $q = 4$) show that if p does not divide q, then the number of roots of $f = \Phi_q$ modulo p is given by

$$\nu_f(p) = \begin{cases} \varphi(q) & \text{if } p \equiv 1 \pmod q, \\ 0 & \text{otherwise.} \end{cases}$$

The conclusion of Kronecker's Theorem is then the assertion that

$$\frac{\varphi(q)}{\pi(x)} \sum_{\substack{p \leqslant x \\ p \equiv 1 \pmod q}} 1 \to m,$$

where m is the number of irreducible factors of f over $\mathbf{Q}$ (note that since f visibly has no multiple root in $\mathbf{C}$, this number is the same whether we count the irreducible factors

with or without multiplicity). It is more customary to write the equivalent formula

$$\pi(x;q,1) \sim \frac{m}{\varphi(q)}\pi(x),$$

where by definition, for any integer $q \geqslant 1$ and any integer a coprime to q, we denote by $\pi(x;q,a)$ the number of prime numbers $p \leqslant x$ which are $\equiv a \pmod q$.

We can view this in two different ways. First, it is an earlier theorem of Kronecker (1854) that the cyclotomic polynomial Φ_q is irreducible over $\mathbf{Q}$ (this had been proved earlier for q prime by Gauss, and simpler proofs were known of that case; although there are now every simple-looking arguments, such as the one we will review in the next chapter, this is a deep arithmetic fact, and is a prototype for another very important type of studies in number theory, nowadays known as "Galois representations"). Thus we know that $m = 1$, and we see that Kronecker's Theorem implies that

$$\pi(x;q,1) \sim \frac{1}{\varphi(q)}\pi(x),$$

and in particular there are infinitely many primes congruent to 1 modulo q (even better, Chebychev's estimate shows that the number of those $\leqslant x$ is growing approximately like $\frac{1}{\varphi(q)}\frac{x}{\log(x)}$; this is not as precise as stronger forms of the Prime Number Theorem in arithmetic progressions will provided later, but it is already far from obvious).

There is however another interpretation: one can in fact prove *independently* of the irreducibility of the cyclotomic polynomials that

$$\pi(x;q,a) \sim \frac{1}{\varphi(q)}\pi(x),$$

as $x \to +\infty$, for all a coprime to q (in a weaker form, this is the celebrated theorem of Dirichlet on primes in arithmetic progressions, which was in fact proved in 1837, before Kronecker's result on the irreducibility of Φ_q). Hence we deduce the non-trivial fact that Φ_q is irreducible! Interestingly, in the introduction to the paper where he states Theorem 1.4.2, Kronecker very clearly states the possibility of such an "analytic" proof as a motivation for his work exploring the properties of $\nu_f(p)$.

(3) Taking the first example of $f = \mathrm{X}^2 + 1$ as the "simplest" irreducible polynomial over $\mathbf{Q}$, there is another natural generalization which might seem simpler than that of cyclotomic polynomials: let's consider $f = \mathrm{X}^2 - d$, where $d \in \mathbf{Z}$ in an arbitrary integer, and to avoid the case where f is reducible, assume that d is not the square of an integer. Then $\nu_f(p)$ counts the number of square roots of d modulo p. The study of Kronecker's problem in that particular case actually goes back to Fermat, Euler, Lagrange, Legendre and Gauss. It turns out to be incredibly rich, and even more than Fermat's Theorem, it is the starting point of algebraic number theory.

One feature is clear: note that $0 \leqslant \nu_f(p) \leqslant 2$ (since f is never identically zero modulo p), and that $-x$ is a square root of d whwnever x is. So, if $\nu_f(p)$ is non-zero, it will be equal to 2 unless there is an $x \in \mathbf{F}_p$ such that $x^2 = d$ and $x = -x$. The last can only happen if $2x = 0$ in $\mathbf{F}_p$, i.e., if $p = 2$ or $x = 0$; in the first case, we have $\nu_f(2) = 1$ (either d is even and $x = 0$ is the unique root or d is odd and $x = 1$ is the unique root), and the second case will occur only if $p \mid d$, in which case indeed $\nu_f(p) = 1$.

We are therefore led to concentrate on the case of primes p not dividing d, where $\nu_f(p)$ is either 0 or 2, and the question is to determine when these two cases occur.

Here is the data for the first hundred primes when $d = 12$ and $d = -13$ for instance, where we keep track of the primes again as this is crucial to understand the structure.

For $f = \mathrm{X}^2 - 12$:

$$(2,1),(3,1),(5,0),(7,0),(11,2),(13,2),(17,0),(19,0),(23,2),(29,0),(31,0),(37,2),(41,0),$$
$$(43,0),(47,2),(53,0),(59,2),(61,2),(67,0),(71,2),(73,2),(79,0),(83,2),(89,0),(97,2),$$
$$(101,0),(103,0),(107,2),(109,2),(113,0),(127,0),(131,2),(137,0),(139,0),(149,0),(151,0),$$
$$(157,2),(163,0),(167,2),(173,0),(179,2),(181,2),(191,2),(193,2),(197,0),(199,0),(211,0),$$
$$(223,0),(227,2),(229,2),(233,0),(239,2),(241,2),(251,2),(257,0),(263,2),(269,0),(271,0),$$
$$(277,2),(281,0),(283,0),(293,0),(307,0),(311,2),(313,2),(317,0),(331,0),(337,2),(347,2),$$
$$(349,2),(353,0),(359,2),(367,0),(373,2),(379,0),(383,2),(389,0),(397,2),(401,0),(409,2),$$
$$(419,2),(421,2),(431,2),(433,2),(439,0),(443,2),(449,0),(457,2),(461,0),(463,0),(467,2),$$
$$(479,2),(487,0),(491,2),(499,0),(503,2),(509,0),(521,0),(523,0),(541,2).$$

For $f = \mathrm{X}^2 + 13$:

$$(2,1),(3,0),(5,0),(7,2),(11,2),(13,1),(17,2),(19,2),(23,0),(29,2),(31,2),(37,0),$$
$$(41,0),(43,0),(47,2),(53,2),(59,2),(61,2),(67,2),(71,2),(73,0),(79,0),(83,2),(89,0),$$
$$(97,0),(101,2),(103,0),(107,0),(109,0),(113,2),(127,0),(131,0),(137,0),(139,0),(149,0),$$
$$(151,2),(157,2),(163,2),(167,2),(173,2),(179,0),(181,2),(191,0),(193,0),(197,0),$$
$$(199,0),(211,0),(223,2),(227,2),(229,0),(233,2),(239,2),(241,0),(251,0),(257,2),$$
$$(263,0),(269,2),(271,2),(277,2),(281,0),(283,0),(293,0),(307,2),(311,0),(313,2),$$
$$(317,0),(331,2),(337,2),(347,0),(349,0),(353,0),(359,2),(367,0),(373,2),(379,2),$$
$$(383,2),(389,2),(397,0),(401,0),(409,0),(419,0),(421,0),(431,2),(433,2),(439,0),$$
$$(443,0),(449,0),(457,0),(461,0),(463,2),(467,0),(479,2),(487,2),(491,0),(499,2),$$
$$(503,0),(509,0),(521,2),(523,0),(541,0).$$

It is probably based on experiments that Legendre was led to the discovery of some remarkably regularity in this data concerning modular square roots. In particular, he found (as you may if considering enough evidence) that, except for finitely many exceptions, the set of primes for which $\mathrm{X}^2 - d$ has a root modulo p is determined by *arithmetic progressions* modulo d. This appeared in the case of $\mathrm{X}^2 + 1$ (where the progression is modulo 4, and the exception is $p = 2$), and also – and this may seem very intriguing – for the higher-degree cyclotomic polynomials. This may lead you back to look at some of the previous data – for instance, can one see such a pattern for $f = \mathrm{X}^3 + \mathrm{X} - 1$?

Legendre went far enough to make a precise conjecture on *which* arithmetic progressions described those primes where $\mathrm{X}^2 - d$ has a root. We will consider this, but we first formally close this chapter, as we will start to build the required theory to understand better these problems.

Theorem 1.4.2 is a fitting conclusion to this introductory chapter, because it shows how the dichotomies that are sometimes created between "algebraic" or "analytic" number theory, or the one we proposed between "deterministic" and "probabilistic" number theory, are always artificial: in Kronecker's Theorem, if we are given a concrete polynomial f for which we ask whether it is irreducible over $\mathbf{Q}$ or not (e.g., Kronecker was

especially interested in the case of cyclotomic polynomials), the question is certainly deterministic, and of algebraic nature. But the answer is in terms of *average behavior* of $\nu_f(p)$, which is of probabilistic nature.

CHAPTER 2

Elementary algebraic number theory

2.1. Introduction

REMARK 2.1.1. When we do not give full proofs, we will use as a reference the textbook of Ireland and Rosen [12], which is very accessible.

2.2. Quadratic reciprocity: the statement

We begin by stating and explaining a proof of what is certainly the most important result in algebraic number theory from the historical point of view – the Quadratic Reciprocity Law, discovered by Legendre and proved first by Gauss. Recall that the underlying question is the following: given an integer d, not a square, can one describe the set of primes p such that $X^2 - d$ has a root modulo p?

The answer will be provided by the *quadratic reciprocity law*. This involves the so-called *Legendre symbol*, which Legendre introduced to keep track of which residue classes are squares and which are not modulo a prime number p:

DEFINITION 2.2.1 (Legendre symbol). Let p be a prime number. The quadratic residue symbol modulo p is the function

$$\mathbf{F}_p \to \{-1, 0, 1\}$$

denoted $x \mapsto \left(\frac{x}{p}\right)$ which is defined by

$$\left(\frac{x}{p}\right) = \begin{cases} -1 & \text{if } x \text{ is not a square of an element of } \mathbf{F}_p^\times, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x \text{ is a square of an element of } \mathbf{F}_p^\times. \end{cases}$$

More generally, if $a \in \mathbf{Z}$, we denote by $\left(\frac{a}{p}\right)$ the Legendre symbol of $a \pmod p \in \mathbf{F}_p$.

For instance, $\left(\frac{-1}{p}\right) = 1$ whenever $p \equiv 1 \pmod 4$: this is saying that there is a square-root of -1 modulo p in that case. Note that although it seems somehow related to the original question, the Legendre symbol seems to be going in the wrong direction: instead of thinking of finding *varying* primes p for which a *given* integer d is a square (i.e., admits a modular square-root), we are apparently fixing p and looking at all the squares modulo this prime.

THEOREM 2.2.2 (Gauss). *Let p and q be distinct odd primes. We then have*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4}.$$

Before talking about the proof, we will spend some time discussing this result. First, let us check that it provides (essentially) the full solution to the original problem of modular square roots. This is clear if we consider the square-roots of d modulo primes when $d = q$ is an odd prime number. Then, except for the single case of $p = q$, the

integer q has a square-root modulo p *if and only if* $\left(\frac{q}{p}\right) = 1$, by definition, and by Quadratic Reciprocity, this is true if and only if

$$\left(\frac{p}{q}\right) = (-1)^{(p-1)(q-1)/4}.$$

We claim that this is a condition depending only on the residue class of p modulo $4q$. Indeed, if $q \equiv 1 \,(\mathrm{mod}\, 4)$, the right-hand side is 1, so we are looking to the even simpler set of primes p which are squares modulo q. On the other hand, if $q \equiv 3 \,(\mathrm{mod}\, 4)$, then $(q-1)/2$ is odd so the right-hand side is $(-1)^{(p-1)/2}$. So the set in question is the union of (1) the set of primes $p \equiv 1 \,(\mathrm{mod}\, 4)$ which are also squares modulo q; (2) the set of primes $p \equiv 3 \,(\mathrm{mod}\, 4)$ which are also non-squares modulo q. By the Chinese Remainder Theorem, either of these describes uniquely a set of classes modulo $4q$.

Suppose now that d is more general than an odd prime number. Then we can use (in principle) the multiplicativity property

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

of the Legendre symbol, which is one justification for its definition (using the characteristic function of the set of squares would carry the same information of distinguishing squares and non-squares, but would lack this very useful property). We prove this below; for the moment, observe that if we factor d in the form

$$d = \varepsilon p_1^{n_1} \cdots p_k^{n_k}$$

where $k \geqslant 0$, the p_i's are distinct primes and $n_i \geqslant 1$, and where $\varepsilon \in \{-1, 1\}$ indicates the sign of d (remember that even the case $d = -1$ is interesting), then we have

$$\left(\frac{d}{p}\right) = \left(\frac{\varepsilon}{p}\right)\left(\frac{p_1}{p}\right)^{n_1} \cdots \left(\frac{p_k}{p}\right)^{n_k}.$$

If n_i is even, then $\left(\frac{p_i}{p}\right)^{n_i} = 1$, and if n_i is odd, then $\left(\frac{p_i}{p}\right)^{n_i} = \left(\frac{p_i}{p}\right)$. So, with the exception of the possibility p might divide d (in which case $d \equiv 0 \,(\mathrm{mod}\, p)$, so it is a square) or that 2 might be one of the primes dividing d, what is needed is to compute $\left(\frac{q}{p}\right)$ when q is an odd prime distinct from p, and Quadratic Reciprocity computes this in terms of $\left(\frac{p}{p}\right)$. Thus the following additional formulas give access to the full solution of the problem of existence of modular square-roots modulo odd primes.

Proposition 2.2.3. *For all odd primes p, the formulas*

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

hold.

Note that the first of these refers again to the problem of square-roots of -1 modulo primes, and we noticed already that it amounted to have $p \equiv 1 \,(\mathrm{mod}\, 4)$ because of its interpretation in terms of fourth roots of unity. The second case, however, will need a separate proof. Note that, concretely, we can check by looking at the various cases that it means that 2 is a square modulo primes which are either 1 or 7 modulo 8, and is not a square modulo primes which are 3 or 5 modulo 8.

Before going further, we prove the multiplicativity of the Legendre symbol.

Lemma 2.2.4. *Let p be an odd prime number. For all a and b modulo p, we have*

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

PROOF. Note that the only case which is not obvious from the definition is when both $\left(\frac{a}{p}\right)$ and $\left(\frac{b}{p}\right)$ is -1: then we need to show that the product of two *non-squares* is a square. But we use algebra to deal with the general case in one stroke: we consider the squaring morphism

$$s\colon \mathbf{F}_p^\times \to \mathbf{F}_p^\times$$

defined by $s(x) = x^2$. Its image, denoted $(\mathbf{F}_p^\times)^2$, is the subgroup of non-zero squares modulo p; its kernel is $\{-1, 1\}$, since these are the roots of $\mathrm{X}^2 = 1$ modulo p, and has order 2 (here we use the assumption that p is odd). It follows that the group of squares has order $(p-1)/2$, and the quotient $\mathrm{Q} = \mathbf{F}_p^\times/(\mathbf{F}_p^\times)^2$ has order 2 by elementary group theory. But now observe that if we identify Q with the group $\{-1, 1\}$ (which we can without ambiguity since both have order 2, and there is only one way to make the identification!), then the Legendre symbol $\left(\frac{a}{p}\right)$, for $a \in \mathbf{F}_p^\times$, "is" the value at a of the quotient morphism $\mathbf{F}_p^\times \to \mathrm{Q} = \{-1, 1\}$. (Indeed, being a square means being the trivial element in the quotient group, a non-square being the non-trivial element -1). So the multiplicativity of the Legendre symbol is just the fact that the quotient map is a group homomorphism. □

Finally, we want to explain that Quadratic Reciprocity is absolutely amazing. There are at least three reasons *why*:

(1) It is often said (including in these notes) that distinct primes tend to behave "independently" of each other, and this is often justified. The Quadratric Reciprocity Law shows that in fact they are not independent: knowing whether one is a square modulo the other gives information on the reciprocal situation, which wouldn't be true with exact independence.

(2) The statement is, indeed symmetric in terms of p and q, but this symmetry is completely absent from the way the original problem was phrased: here q was thought as the fixed source of the polynomial $\mathrm{X}^2 - q$, whose modular roots are studied, while p was varying in the infinite set of primes distinct from q.

(3) The Legendre symbol $\left(\frac{a}{p}\right)$ is defined for a modulo p, so the Legendre symbol $\left(\frac{q}{p}\right)$ doesn't seem to care whether q is prime or not. But $\left(\frac{p}{q}\right)$ does, since the Legendre symbol works modulo primes!

2.3. Quadratic reciprocity: sketch of a proof

We will now describe a proof of the Quadratic Reciprocity law. It will not be quite complete, but it will be easy to finalize it as soon as we have set up the basic outlines of algebraic number theory. Moreover, this proof, again, opens up an almost infinite horizon for wider discussion.

The focus of this proof is to keep in mind the original problem. So we are thinking of the odd prime q as fixed, and we want to find a criterion, in terms of q, for which primes p admit a modular square-root of q. (It is in some sense irrelevant that the criterion will take the remarkable form of the Reciprocity Law – in some sense, this might even be misleading!)

The first step is to reword the problem. Let p be a prime different from q. Whether q is or not a square modulo p, we can always consider a square-root of q modulo p, which may simply lie in some extension of the field $\mathbf{F}_p$. Let α be such a root and $\mathrm{E} = \mathbf{F}_p(\alpha)$. By the theory of finite fields, the element α belongs to $\mathbf{F}_p$ if and only if $\alpha^p = \alpha$. If this is not the case, then since α^p is also a root of $\mathrm{X}^2 = q$ modulo p, we have $\alpha^p = -\alpha$. Noting

the sign, we conclude that the Legendre symbol $\left(\frac{q}{p}\right)$, which we try to compute, is also characterized by

$$\alpha^p = \left(\frac{q}{p}\right)\alpha, \tag{2.1}$$

for α a square-root of q in some extension of $\mathbf{F}_p$.

This reformulation gives an opening for going further using the following idea: we will try, in general, to find a *second* expression for α, for which the value of α^p can be determined. The way to do this is to take seriously the following naive point of view: α should be "reduction" of the actual number $\sqrt{q}$. In other words, we want to explicitly link the modular problem with "characteristic zero" information.

This is provided by a remarkable identity, which also goes back to Lagrange and especially Gauss, which expresses $\sqrt{q}$ in terms of expressions now known as *Gauss sums*.

PROPOSITION 2.3.1. *Let q be an odd prime number. The complex number*

$$\tau_q = \sum_{a=1}^{q-1}\left(\frac{a}{q}\right)e^{2i\pi a/q}$$

satisfies

$$\tau_q^2 = (-1)^{(q-1)/2}q.$$

This means that τ_q is either a square-root of q (it might not be clear if it is $\sqrt{q}$ or $-\sqrt{q}$) or one of $-q$, depending on whether $q \equiv 1 \pmod 4$ or $q \equiv 3 \pmod 4$ (so it might be $i\sqrt{q}$ or $-i\sqrt{q}$). If we define $q^* = q$ and $\widetilde{\tau}_q = \tau_q$ if $q \equiv 1 \pmod 4$, whereas $q^* = 4q$ and $\widetilde{\tau}_q = i\tau_q$ if $q \equiv 3 \pmod 4$, then in all cases, we get $\widetilde{\tau}_q^2 = q$, and $\widetilde{\tau}_q$ is a linear combination with integral coefficients of roots of unity of order q^*.

Note that, a priori, this has nothing to do with modular square-roots. However, the formula for τ_q (or $i\tau_q$) represents $\sqrt{q}$ (or $-\sqrt{q}$ or $i\sqrt{q}$ or $-i\sqrt{q}$) as an element in the field $\mathbf{Q}(e^{2i\pi/q^*})$ generated over $\mathbf{Q}$ by the q^*-th roots of unity. In fact, and this is essential, it gives an expression as an element of the *ring* $\mathbf{Z}[e^{2i\pi/q^*}]$. This ring has properties analogue to those of $\mathbf{Z}$, and in particular this will provide us with reduction maps modulo various ideals; the images of $\widetilde{\tau}_q$ will be modular square-roots of q in the respective quotients.

Now the final link is provided by the fact that, for a square-root α of q modulo a prime $p \neq q$, computing α^p means applying to α the Frobenius automorphism $x \mapsto x^p$ of an extension in which α lies. A crucial compatibility property is that there is, similarly, an automorphism of the field $\mathbf{Q}(e^{2i\pi/q^*})$ which "represents" this Frobenius automorphism modulo p, in the sense that applying it, then reducing, is the same as reducing, then applying the Frobenius. It is maybe not too surprising that this automorphism, say σ_p, is characterized by the fact that

$$\sigma_p\left(e^{2i\pi/q}\right) = e^{2i\pi p/q} \qquad i \mapsto i^p$$

(raising to the p-th power the roots of unity!).

The outcome of the previous discussion (which requires a proof, which will come later) is that $\sigma_p(\widetilde{\tau}_q)$ will be either $\widetilde{\tau}_q$ or $-\widetilde{\tau}_q$ (because the square of $\sigma_p(\widetilde{\tau}_q)$ is the same as that of $\widetilde{\tau}_q$), and the sign which appears is the same as the one which appears modulo p under the Frobenius action, which is $\left(\frac{q}{p}\right)$ by (2.1).

Now what is the image of τ_q under the automorphism σ_p? We compute

$$\sigma_p(\tau_q) = \sigma_p\left(\sum_{a=1}^{q-1}\left(\frac{a}{q}\right)e^{2i\pi a/q}\right) = \sum_{a=1}^{q-1}\left(\frac{a}{q}\right)\sigma_p\left(e^{2i\pi a/q}\right) = \sum_{a=1}^{q-1}\left(\frac{a}{q}\right)e^{2i\pi ap/q}.$$

To evaluate further, we note that $e^{2i\pi x/q}$, for $x \in \mathbf{Z}$, only depends on x modulo q, as does $\left(\frac{x}{q}\right)$. Since p is different from q, the classes of ap modulo q, as a varies from 1 to $q-1$, are exactly the $q-1$ invertible residue classes modulo q. Writing $b = ap$, we obtain

$$\left(\frac{a}{q}\right)e^{2i\pi ap/q} = \left(\frac{bp}{q}\right)e^{2i\pi b/q} = \left(\frac{p}{q}\right)\left(\frac{b}{q}\right)e^{2i\pi b/q}$$

(because $\left(\frac{bp}{q}\right) = \left(\frac{b}{q}\right)\left(\frac{p}{q}\right) = \left(\frac{b}{q}\right)^2\left(\frac{a}{q}\right)$, by Lemma 2.2.4), and therefore

$$\sigma_p(\tau_q) = \sum_{b=1}^{q-1}\left(\frac{p}{q}\right)\left(\frac{b}{q}\right)e^{2i\pi b/q} = \left(\frac{p}{q}\right)\tau_q.$$

On the other hand, we have $\sigma_p(i) = i^p = (-1)^{(p-1)/2}i$, hence finally

$$\sigma_p(\widetilde{\tau}_q) = \left(\frac{p}{q}\right)\widetilde{\tau}_q \quad \text{if } q \equiv 1 \pmod 4,$$
$$\sigma_p(\widetilde{\tau}_q) = (-1)^{(p-1)/2}\left(\frac{p}{q}\right)\widetilde{\tau}_q \quad \text{if } q \equiv 3 \pmod 4.$$

We said (and this is, with Proposition 2.3.1, the only unproved part) that the sign must be the same as in (2.1). Hence we get

$$\left(\frac{q}{p}\right) = \left(\frac{p}{q}\right) \quad \text{if } q \equiv 1 \pmod 4,$$
$$\left(\frac{q}{p}\right) = (-1)^{(p-1)/2}\left(\frac{p}{q}\right) \quad \text{if } q \equiv 3 \pmod 4.$$

This statement is equivalent to the Quadratic Reciprocity formula: if $q \equiv 1 \pmod 4$, we have $(-1)^{(p-1)(q-1)/4} = 1$, and otherwise $(-1)^{(p-1)(q-1)/4} = (-1)^{(p-1)/2}$.

We will prove Proposition 2.3.1 right away, and leave the final compatibility property to a later point where it will become straightforward. First, to simplify notation (and make certain things look less "transcendental"), we will use the notation

$$e(z) = e^{2i\pi z}$$

for $z \in \mathbf{C}$. This has the property that $e(z+w) = e(z)e(w)$ and $e(n) = 1$ if and only if $n \in \mathbf{Z}$; moreover, as already observed, for any positive integer m and any $a \in \mathbf{Z}$, the quantity $e(a/m)$ only depends on the class of a modulo m, and we will often consider $e(a/m)$ where $a \in \mathbf{Z}/m\mathbf{Z}$.

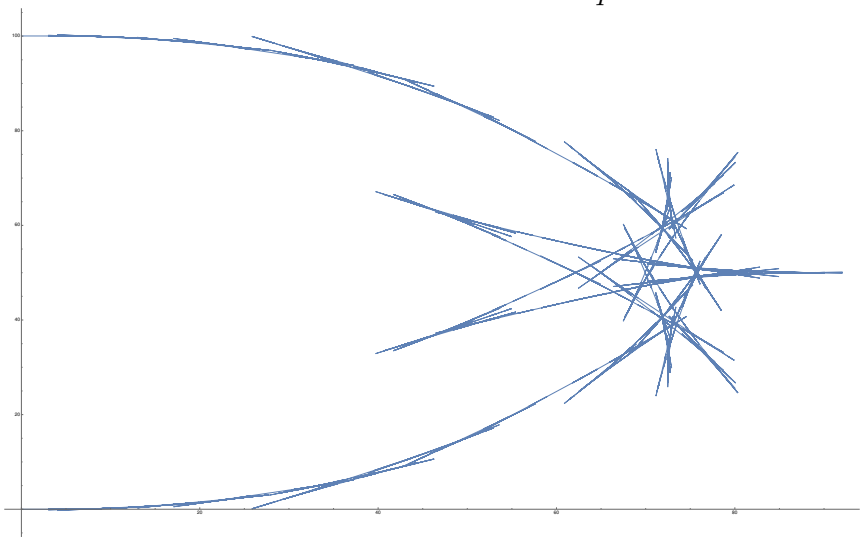
The square of the Gauss sum. By multiplying the sums, and then using the multiplicativity of Legendre symbols, we obtain

$$\tau_q^2 = \sum_{1\leqslant a,b\leqslant q-1}\left(\frac{a}{q}\right)\left(\frac{b}{q}\right)e\left(\frac{a+b}{q}\right) = \sum_{1\leqslant a,b\leqslant q-1}\left(\frac{ab}{q}\right)e\left(\frac{a+b}{q}\right).$$

To go further, we will make a change of variable, and it is convenient to think of a and b as ranging over $\mathbf{F}_q^\times$ when doing this. Then we define x so that $b = ax$; the substitution $(a,x) \mapsto (a,ax)$ is bijective from $\mathbf{F}_q^\times \times \mathbf{F}_q^\times$ to itself (the inverse is $(a,b) \mapsto (a,b/a)$), so

$$\tau_q^2 = \sum_{a,x\in\mathbf{F}_q^\times}\left(\frac{x}{q}\right)e\left(\frac{a(1+x)}{q}\right).$$

FIGURE 2.1. The Gauss sum for $q = 10007$.



This change of variable has "separated" the contribution of the Legendre symbol, and we can now first sum over a for a fixed valued of x. Thus, we have the expression

$$\tau_q^2 = \sum_{x\in\mathbf{F}_q^\times} \Big(\frac{x}{q}\Big) \sum_{a\in\mathbf{F}_q^\times} e\Big(\frac{a(1+x)}{q}\Big).$$

In the inner sum over a, two cases may arise; if $x = -1$, then all terms are equal to 1, and the sum is equal to $q-1$, whereas if $x \neq -1$, we are summing all q-th roots of unity, except 1, and the sum is then equal to -1 (since we know that the sum of all q-th roots of unity is zero). So

$$\tau_q^2 = -\sum_{\substack{x\in\mathbf{F}_q^\times \\ x\neq -1}} \Big(\frac{x}{q}\Big) + (q-1)\Big(\frac{-1}{q}\Big) = \Big(\frac{-1}{q}\Big)q + \sum_{x\in\mathbf{F}_q^\times} \Big(\frac{x}{q}\Big).$$

The last sum is simply the difference between the number of squares in $\mathbf{F}_q^\times$ and the number of non-squares, so it is zero. And the Legendre symbol $\left(\frac{-1}{q}\right)$ is, as always, the sign $(-1)^{(q-1)/2}$ detecting the congruence class of q modulo 4... This completes the proof. □

REMARK 2.3.2. (1) The formula of Proposition 2.3.1 is by itself a very interesting statement! One can think of a geometric interpretation: the Gauss sum τ_q describes a specific "path" in the complex plane which starts at the origin and ends at some square-root of $(-1)^{(q-1)/2}q$ (which one?), where the steps are limited to be of length one and roots of unity of order $2q$. How does this path look like? Figure 2.3.2 shows this $q = 10007$, where the sum is done in the order indicated in the definition: the variable ranges from 1 to 10006.

The existence of this path is quite fascinating –it is very far from clear that it should exist (we will discuss later the Kronecker–Weber Theorem, and see that it predicts *a priori* only that there is an expression for $\sqrt{q}$ in terms of q-th roots of unity, but only with some unspecified integral coefficients), but one can observe that it is unique up to reordering the terms of the sum, simply because the primitive q-th roots of unity are linearly independent over $\mathbf{Q}$ (when q is prime).

(2) Again, we ask the natural question: *which* of the roots of $(-1)^{(q-1)/2}q$ does the Gauss sum give? The answer is far from clear, so we resort to experiments, and here is

the data the first few primes, with 5 digits of precision:

$$\begin{array}{cccc} \tau_{101} = 10.050, & \tau_{103} = 10.149\ i, & \tau_{107} = 10.344\ i, & \tau_{109} = 10.440 \\ \tau_{113} = 10.630, & \tau_{127} = 11.269\ i, & \tau_{131} = 11.446\ i, & \tau_{137} = 11.705 \\ \tau_{139} = 11.790\ i, & \tau_{149} = 12.207, & \tau_{151} = 12.288\ i, & \tau_{157} = 12.530 \\ \tau_{163} = 12.767\ i, & \tau_{167} = 12.923\ i, & \tau_{173} = 13.153, & \tau_{179} = 13.379\ i \\ \tau_{181} = 13.454, & \tau_{191} = 13.820\ i, & \tau_{193} = 13.892, & \tau_{197} = 14.036 \end{array}$$

$$\tau_{199} = 14.107\ i$$

The evidence, which can go much further, leads to the conclusion that: *the sign is always* +; in other words, τ_q is always the positive square root of q if $q \equiv 1 \pmod 4$, and is always $i\sqrt{q}$ if $q \equiv 3 \pmod 4$. This may seem innocuous, but it is a very deep property, and it is strikingly difficult to prove (which was first done by Gauss, who had seen the evidence, and struggled for a very long time on finding a proof).

We conclude this section by explaining another way to make this proof rigorous, without waiting for the basic constructions of algebraic number theory to be available. This consists simply in constructing the square-root α of q modulo p by imitating the Gauss sum over $\mathbf{F}_p$. Precisely, for m coprime to p, let ω_m denote a primitive m-th root of unity in some fixed algebraic closure of $\mathbf{F}_p$. In $\widetilde{\mathrm{E}}_q = \mathbf{F}_p(\omega_q, \omega_4)$, we can define

$$\eta_q = \sum_{a=1}^{q-1} \left(\frac{a}{q}\right) \omega_q^a, \quad \widetilde{\eta}_q = \begin{cases} \eta_q & \text{if } q \equiv 1 \pmod 4, \\ i\eta_q & \text{if } q \equiv 3 \pmod 4. \end{cases}$$

By just reproducing the proof of Proposition 2.3.1, we get $\widetilde{\eta}_q^2 = q$, where the equality is as elements of $\mathbf{F}_p$. The computation of $\widetilde{\eta}_q^p$ is also straightforward, since raising to the p-th power is an automorphism in characteristic p, and gives exactly as before the formula

$$\widetilde{\eta}_q^p = \left(\frac{p}{q}\right) \widetilde{\eta}_q \quad \text{if } q \equiv 1 \pmod 4,$$

$$\widetilde{\eta}_q^p = (-1)^{(p-1)/2} \left(\frac{p}{q}\right) \widetilde{\eta}_q \quad \text{if } q \equiv 3 \pmod 4.$$

Using $\widetilde{\eta}_q$ as the square-root α of q, we obtain the Quadratic Reciprocity Formula, as an equality

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4}$$

in $\mathbf{F}_p$. But since p is odd and both sides of this formula are equal to 1 or -1 modulo p, the congruence implies the equality in $\mathbf{Z}$.

2.4. Number fields and their rings of integers

We now begin to develop the theory of *number fields*, which is one of the main subject of algebraic number theory. These will be essential tools to understand properly statements like Quadratic Reciprocity and its generalizations.

Definition 2.4.1. A number field K is a finite (necessarily algebraic) field extension of $\mathbf{Q}$.

Example 2.4.2. (1) One important construction of number fields is provided by a field $\mathbf{Q}(\alpha)$ obtained by adjoining to $\mathbf{Q}$ a root α of a fixed non-constant polynomial $f \in \mathbf{Q}[\mathrm{X}]$. Indeed, all number fields arise this way, but the polynomial f is far from unique, and the number field turns out to be a more useful object for further study than the polynomial f.

Similarly, starting from f, we can construct the *splitting field* K of f, which is another number field; this time, it is generated by *all* roots of f in some field which contains them all, for instance in $\mathbf{C}$. The number fields obtained this way are characterized as all the *normal* finite extensions of $\mathbf{Q}$, or equivalently, all the finite *Galois* extensions of $\mathbf{Q}$.

(2) Concretely, some of the most classical and important number fields are:

(i) Quadratic fields, of the form $\mathrm{K} = \mathbf{Q}(\sqrt{d})$, where $d \in \mathbf{Q}$ is not a square;

(ii) Cyclotomic fields, of the form $\mathrm{K} = \mathbf{Q}(\omega)$ for some root of unity $\omega \in \mathbf{C}$ (of some order).

These examples are Galois extensions of $\mathbf{Q}$; they are far from typical, however, because in both cases, their Galois group is abelian.

(3) In some sense, there doesn't seem to be "really nice" *simple* families of number fields with non-abelian Galois groups. However, for testing purposes, one can consider examples such that $\mathrm{K}_n = \mathbf{Q}(\theta_n)$, where θ_n is a root of

$$1 + \mathrm{X} + \frac{\mathrm{X}^2}{2} + \cdots + \frac{\mathrm{X}^n}{n} = 0.$$

It is known that f_n is irreducible over $\mathbf{Q}$ (this is due to Schur), and that the Galois group of the splitting field of K_n is the symmetric group $\mathfrak{S}_n$.

To do number theory, however, one needs not only elements of number fields, which are analogues of rational numbers, but most importantly the analogues of *integers*. The most naive approach to defining integers in a number field would be to take, for instance, $\mathbf{Z}[\alpha]$ instead of $\mathbf{Q}(\alpha)$. This turns out to be a bad definition in general, and the correct one is more subtle. We give the general definition (since this can be useful) before considering its specialization to number fields.

Definition 2.4.3 (Integral element, integral extension). Let A and B be commutative rings and $\varphi\colon \mathrm{A} \to \mathrm{B}$ a morphism of rings, preserving the unit. An element $b \in \mathrm{B}$ is *integral over* A if there exists a monic polynomial $f \in \mathrm{A}[\mathrm{X}]$, say

$$f = \mathrm{X}^d + a_{d-1}\mathrm{X}^{d-1} + \cdots + a_1\mathrm{X} + a_0, \qquad a_i \in \mathrm{A},$$

such that b is "a root of f", i.e., such that

$$b^d + \varphi(a_{d-1})b^{d-1} + \cdots + \varphi(a_1)b + \varphi(a_0) = 0.$$

The set of elements of B which are integral over A is called the integral closure of A in B.

The *ring of integers* $\mathbf{Z}_\mathrm{K}$ of a number field K is the integral closure of $\mathbf{Z}$ in K.

Although this definition is the correct one, it is not immediately clear why, but this will appear later. We will also see how, in a natural sense, an element x of a number field K is integral (i.e., is in $\mathbf{Z}_\mathrm{K}$) exactly when "it has no denominator".

Remark 2.4.4. Many (in fact, most) writers use the notation $\mathscr{O}_\mathrm{K}$ for the ring of integers. The notation $\mathbf{Z}_\mathrm{K}$ is more common in the French community, and seems more logical.

Remark 2.4.5. Let K be a number field. Since x is algebraic over $\mathbf{Q}$, there is a well-defined monic minimal polynomial $f \in \mathbf{Q}[\mathrm{X}]$ for f, namely the unique monic generator of the prime ideal

$$\mathrm{I} = \{g \in \mathbf{Q}[\mathrm{X}] \mid g(x) = 0\} \subset \mathbf{Q}[\mathrm{X}].$$

The generator f is an irreducible polynomial. We claim that if x is integral over $\mathbf{Z}$, then in fact $f \in \mathbf{Z}[\mathrm{X}]$, so that $f(x) = 0$ gives an integral equation satisfied by x. Indeed,

there is an element $g \in \mathbf{Z}[\mathrm{X}]$, irreducible in $\mathbf{Z}[\mathrm{X}]$, such that $g(x) = 0$ (taking a factorization of an integral equation satisfied by x, we only need take one of its irreducible factors). It is known that g is a primitive polynomial (the gcd of the coefficients is equal to 1) and is irreducible in $\mathbf{Q}[\mathrm{X}]$ (see, e.g, [2, Prop. 2.6.5]). It follows that g is a generator of I, so $f = \alpha g$ for some $\alpha \in \mathrm{K}^\times$. Looking at the leading coefficient, it follows that α is a unit in $\mathbf{Z}$, so is -1 or 1, and then $f \in \mathbf{Z}[\mathrm{X}]$.

Example 2.4.6. (1) Let $a \in \mathbf{Z}$. For any integer $k \geqslant 1$, any k-th root α of a is an element of $\mathbf{C}$ which is integral over $\mathbf{Z}$: we can take the polynomial $f = \mathrm{X}^k - a$.

(2) Let $k \geqslant 1$. Any k-th root of unity is integral over $\mathbf{Z}$.

(3) The element

$$\alpha = \frac{1+\sqrt{5}}{2} \in \mathbf{Q}(\sqrt{5})$$

seems to "have a denominator". However, it is in fact an integer: indeed, α and $\alpha' = (1-\sqrt{5})/2$ are the two real solutions of the quadratic equation

$$\mathrm{X}^2 - \mathrm{X} - 1 = 0.$$

Proposition 2.4.7. *Let* K *be a number field of degree* $d = [\mathrm{K} : \mathbf{Q}] \geqslant 1$. *The ring of integers* $\mathbf{Z}_\mathrm{K}$ *is a subring of* K, *containing* $\mathbf{Z}$, *with fraction field equal to* K *and satisfying* $\mathbf{Q} \cap \mathbf{Z}_\mathrm{K} = \mathbf{Z}$.

In fact, for any $x \in \mathrm{K}$, *there exists some integer* $d \geqslant 1$ *such that* $dx \in \mathbf{Z}_\mathrm{K}$, *and there exists a* $\mathbf{Q}$*-basis* $(\omega_1, \dots, \omega_d)$ *of* K *such that*

$$\mathbf{Z}_\mathrm{K} = \omega_1\mathbf{Z} \oplus \cdots \oplus \omega_d\mathbf{Z}.$$

Proof. The first key fact is that $\mathbf{Z}_\mathrm{K}$ is, indeed, a ring. Since it contains $\mathbf{Z}$ (using $\mathrm{X} - m$ as equation to see that $m \in \mathbf{Z}_\mathrm{K}$), one needs to check that the product and the sum of two integral elements is also integral, and this is done similarly to the proof that the sum or product of algebraic numbers is still algebraic; see, e.g., [12, Prop. 6.1.5] for the proof.

We now establish the remaining properties. First, let $x \in \mathbf{Z}_\mathrm{K} \cap \mathbf{Q}$, and let

$$x^k + a_{k-1}x^{k-1} + \cdots + a_1 x + a_0 = 0$$

with $a_i \in \mathbf{Z}$ be an equation of smallest degree satisfied by x. Write $x = a/b$ with a and b coprime integers. By multiplying with b^k, we obtain

$$a^k + a_{k-1}a^{k-1}b + \cdots + a_1 a b^{k-1} + a_0 b^k = 0,$$

and reducing modulo b, we deduce that $a^k \equiv 0 \pmod{b}$. Since a and b are coprime, this is only possible if $b \in \{-1, 1\}$, so that $x \in \mathbf{Z}$.

Now, let $x \in \mathrm{K}$. It satisfies an equation

$$x^k + a_{k-1}x^{k-1} + \cdots + a_1 x + a_0 = 0$$

with coefficients $a_i \in \mathbf{Q}$, for some $k \leqslant d$. Pick a common denominator $d \geqslant 1$ such that $b_i = da_i \in \mathbf{Z}$ for all i. Then, multiplying by d^k, we see that $y = dx$ satisfies

$$y^k + b_{k-1}y^{k-1} + \cdots + d^{k-2}b_1 y + d^{k-1}b_0 = 0,$$

which implies that $y \in \mathbf{Z}_\mathrm{K}$. So any element x of K has a "denominator" d such that $dx \in \mathbf{Z}_\mathrm{K}$. It follows that K is the fraction field of $\mathbf{Z}_\mathrm{K}$.

Fix now a basis of K over $\mathbf{Q}$; we may assume that its components, say $(\omega_1, \dots, \omega_d)$, are in $\mathbf{Z}_{\mathrm{K}}$, by the previous fact. For any $x \in \mathbf{Z}_{\mathrm{K}}$, we can write

$$x = \sum_{i=1}^{d} \lambda_i(x)\omega_i$$

where $\lambda_i(x) \in \mathbf{Q}$. The key property we finally need is: the λ_i are "almost" integral, in the sense that there exists some fixed integer D such that $\mathrm{D}\lambda_i(x) \in \mathbf{Z}$ for all i and all $x \in \mathbf{Z}_{\mathrm{K}}$. Once we know this, we note that

$$\omega_1\mathbf{Z} \oplus \cdots \oplus \omega_d\mathbf{Z} \subset \mathbf{Z}_{\mathrm{K}} \subset \frac{\omega_1}{\mathrm{D}}\mathbf{Z} \oplus \cdots \oplus \frac{\omega_d}{\mathrm{D}}\mathbf{Z}.$$

The second inclusion shows that $\mathbf{Z}_{\mathrm{K}}$, as an abelian group, is a subgroup of a free abelian group of rank d, hence is free of some finite rank, and since the first inclusion finds a free subgroup of rank d in $\mathbf{Z}_{\mathrm{K}}$, it follows that $\mathbf{Z}_{\mathrm{K}}$ is isomorphic to $\mathbf{Z}^d$ as an abelian group. A $\mathbf{Z}$-basis of $\mathbf{Z}_{\mathrm{K}}$ is then necessarily a $\mathbf{Q}$-basic of K, so the final conclusion follows.

To prove the claim, in view of what was already done, we try to "represent" λ_i by duality by elements of K, using some non-degenerate bilinear form on K. One such form is

$$\mathrm{B}(x, y) = \mathrm{Tr}(xy),$$

where the trace map from $\mathrm{K} \to \mathbf{Q}$ is defined to map $x \in \mathrm{K}$ to the trace of the $\mathbf{Q}$-linear endomorphism $z \mapsto xz$ of K.

This bilinear map is non-degenerate (because if $x \neq 0$, then $\mathrm{B}(x, x^{-1}) = \mathrm{Tr}(1) = d$), and induces a bilinear map $\mathbf{Z}_{\mathrm{K}} \times \mathbf{Z}_{\mathrm{K}} \to \mathbf{Z}$ (because $\mathrm{Tr}(x) \in \mathbf{Z}$ for $x \in \mathbf{Z}_{\mathrm{K}}$: one checks that $\mathrm{Tr}(x) \in \mathbf{Z}_{\mathrm{K}} \cap \mathbf{Q} = \mathbf{Z}$ since $z \mapsto xz$ maps $\mathbf{Z}_{\mathrm{K}}$ to $\mathbf{Z}_{\mathrm{K}}$). So, in particular, for every i, we find some element $\alpha_i \in \mathrm{K}$ such that $\lambda_i(x) = \mathrm{B}(\alpha_i, x)$. Fixing $\mathrm{D} \geqslant 1$ such that $\mathrm{D}\alpha_i \in \mathbf{Z}_{\mathrm{K}}$ for all i, we get

$$\mathrm{D}\lambda_i(x) = \mathrm{DB}(\alpha_i, x) = \mathrm{B}(\mathrm{D}\alpha_i, x) \in \mathbf{Z}$$

whenever $x \in \mathbf{Z}_{\mathrm{K}}$. □

Remark 2.4.8. (1) Of course, there is no unique $\mathbf{Z}$-basis of $\mathbf{Z}_{\mathrm{K}}$.

(2) Another elementary consequence of the definition of integral elements, generalizing the fact that $\mathbf{Q} \cap \mathbf{Z}_{\mathrm{K}} = \mathbf{Z}$, is that $\mathbf{Z}_{\mathrm{K}}$ is integrally closed in K, in the sense that any element of K which is integral over $\mathbf{Z}_{\mathrm{K}}$ is already in $\mathbf{Z}_{\mathrm{K}}$.

Example 2.4.9. (1) Let $d = a/b$ be a non-zero rational number, with a, b coprime integers. We want to compute the ring of integers of the quadratic field $\mathrm{K} = \mathbf{Q}(\sqrt{d})$. To do this, we first observe that $\mathrm{K} = \mathbf{Q}(b\sqrt{d}) = \mathbf{Q}(\sqrt{ab})$, which means that we may assume that $d \in \mathbf{Z}$. In this case, since $\sqrt{d}$ is a root of $\mathrm{X}^2 - d$, we see that $\sqrt{d} \in \mathbf{Z}_{\mathrm{K}}$, hence also $\mathbf{Z}[\sqrt{d}] \subset \mathbf{Z}_{\mathrm{K}}$. We have already seen that (e.g. for $d = 5$) the inclusion may be strict, so determining $\mathbf{Z}_{\mathrm{K}}$ requires some care.

We first make a further simplification: for any prime factor p of d which occurs with even exponent v_p, we may replace d by d/p^{v_p} without changing K, and similarly if v_p is odd, we may replace d by d/p^{v_p-1}. This allows us to assume that any $p \mid d$ occurs with exponent 1 exactly (such integers are called *squarefree*). For instance, from $d = 1316684 = 2^2 \cdot 17^3 \cdot 67$, we can reduce to $d = 17 \cdot 67$.

Let $x \in \mathbf{Z}_{\mathrm{K}}$. We can certainly write $x = m + n\sqrt{d}$, with m and n in $\mathbf{Q}$. If $n = 0$, then $x \in \mathbf{Z}$ (as we have seen), so assume that $n \neq 0$. Then the minimal polynomial f of x, normalized to be monic, has integral coefficients (Remark 2.4.5), and has degree 2.

In fact, denoting $x' = m - n\sqrt{d}$, we have

$$f = (\mathrm{X} - x)(\mathrm{X} - x') = \mathrm{X}^2 - (x + x')\mathrm{X} + xx' = \mathrm{X}^2 - 2m\mathrm{X} + m^2 - dn^2.$$

The condition that $f \in \mathbf{Z}[\mathrm{X}]$ means that $2m \in \mathbf{Z}$ and $m^2 - dn^2 \in \mathbf{Z}$. The first condition says that m is either an integer or half an integer. In the first case ($m \in \mathbf{Z}$), the second condition becomes $dn^2 \in \mathbf{Z}$. Our assumption on d ensures that this is only possible if $n \in \mathbf{Z}$ also: for a prime number p, the p-adic valuation of dn^2 is $v_p(d) + 2v_p(n)$, and must be $\geqslant 0$. But $v_p(d)$ is either 0 or 1; in the first case, we get $v_p(n) \geqslant 0$, and in the second $v_p(n) \geqslant -1/2$, which gives the same thing since $v_p(n) \in \mathbf{Z}$.

There remains the case where $m = \mu/2$ for some $\mu \in \mathbf{Z}$. The second condition is $\mu^2 - 4dn^2 \in 4\mathbf{Z}$, so in particular $4dn^2 \in \mathbf{Z}$. The same argument previously used still applies to odd primes p, so that n can only have a power of two in the denominator. For the prime 2, the condition is

$$2 + v_2(d) + 2v_2(n) \geqslant 0.$$

Using $v_2(d) \leqslant 1$, it follows that $v_2(n) \geqslant -1$, so either n is an integer, or half an integer. So the final question, for given μ, $\nu \in \mathbf{Z}$, with $\nu \neq 0$, is when

$$x = \frac{\mu + \nu\sqrt{d}}{2}$$

is an element in $\mathbf{Z}_{\mathrm{K}}$. This is now equivalent with $\mu^2 - d\nu^2 \in 4\mathbf{Z}$. If ν is even, we deduce that μ is also even, and then $x \in \mathbf{Z}[\sqrt{d}]$. If ν is odd, we deduce that d is a square modulo 4. Since d cannot be 0 modulo 4, we see that $d \equiv 1 \pmod 4$. And finally, assuming this, we find that the condition is that $\mu^2 - \nu^2 \in 4\mathbf{Z}$, and this is equivalent to $\mu \equiv \nu \pmod 2$. In other words:

(1) If d is not $\equiv 1 \pmod 4$, then $\mathbf{Z}_{\mathrm{K}} = \mathbf{Z}[\sqrt{d}]$. (For instance, for $d = -1$, we get the ring of integers $\mathbf{Z}[i]$ of $\mathbf{Q}(i)$.) A $\mathbf{Z}$-basis of $\mathbf{Z}_{\mathrm{K}}$ is $(1, \sqrt{d})$.

(2) If d is $\equiv 1 \pmod 4$, then

$$\mathbf{Z}_{\mathrm{K}} = \left\{ \frac{\mu + \nu\sqrt{d}}{2} \mid \mu, \nu \in \mathbf{Z} \text{ and } \mu - \nu \in 2\mathbf{Z} \right\},$$

and a $\mathbf{Z}$-basis of $\mathbf{Z}_{\mathrm{K}}$ is given by

$$\left(1, \frac{1 + \sqrt{d}}{2}\right).$$

Note that, at least, there is always some integral element $\alpha \in \mathbf{Z}_{\mathrm{K}}$ such that $\mathbf{Z}_{\mathrm{K}} = \mathbf{Z}[\alpha]$, namely $\alpha = \sqrt{d}$ or $\alpha = (1+\sqrt{d})/2$. This fact is *not* true for all number fields; the simplest example where such an α cannot be found is the field $\mathrm{K}(\theta)$ where θ is a root of the cubic polynomial

$$\mathrm{X}^3 - \mathrm{X}^2 - 2\mathrm{X} - 8$$

(an example of Dedekind).

(2) The cyclomotic fields are very well behaved from the point of view of the computation of their ring of integers: for any positive integer $m \geqslant 1$, one can show that the field $\mathbf{Q}(e^{2i\pi/m})$ has ring of integers $\mathbf{Z}[e^{2i\pi/m}]$. However, this is not an easy result (see for instance the proof in Washington's book [20, Th. 2.6]).

(3) Consider as a random example the polynomial

$$f = 1 + \mathrm{X} + \frac{\mathrm{X}^2}{2} + \frac{\mathrm{X}^3}{6} + \frac{\mathrm{X}^4}{24}.$$

Using PARI/GP, we learn that the number field K generated by a root α of f has ring of integers with $\mathbf{Z}$-basis

$$1, 3 + 2\alpha + \frac{\alpha^2}{2} + \frac{\alpha^3}{4}, -2 - \alpha - \frac{\alpha^2}{2} - \frac{\alpha^3}{4}, 2 + \alpha + \frac{\alpha^2}{2}.$$

The index $[\mathbf{Z}_{\mathrm{K}} : \mathbf{Z}[\alpha]]$ is equal to 8 here.

Before starting the next topic, we note that the proof of Proposition 2.4.7 has used the bilinear form $(x, y) \mapsto \mathrm{Tr}(xy)$ on a number field K, where $\mathrm{Tr}(x)$ is the trace of the $\mathbf{Q}$-linear map $z \mapsto xz$. This bilinear form will reappear often, and one of the most crucial invariants of a number field K is related to it.

DEFINITION 2.4.10. Let K be a number field of degree d over $\mathbf{Q}$. Let

$$(\omega_1, \ldots, \omega_d)$$

be a $\mathbf{Z}$-basis of $\mathbf{Z}_{\mathrm{K}}$. The *discriminant* of K is the determinant

$$\det(\mathrm{T}r(\omega_i\omega_j))_{1 \leqslant i,j \leqslant d}.$$

The discriminant is well-defined, because for a different $\mathbf{Z}$-basis (η_i), we have

$$\det(\mathrm{Tr}(\omega_i\omega_j)) = \det(\mathrm{Tr}(\eta_i\eta_j)) \det(\mathrm{A})^2$$

where A is the base-change matrix, and the latter has determinant either -1 or 1, because it and its inverse have integral coefficients.

EXAMPLE 2.4.11. For $\mathrm{K} = \mathbf{Q}(\sqrt{d})$, with d assumed to be squarefree, we have computed a $\mathbf{Z}$-basis above, so we can compute the discriminant. We have $\mathrm{Tr}(1) = 2$. The matrix of multiplication by $\sqrt{d}$ in the $\mathbf{Q}$-basis $(1, \sqrt{d})$ is

$$\begin{pmatrix} 0 & d \\ 1 & 0 \end{pmatrix},$$

so that $\mathrm{Tr}(\sqrt{d}) = 0$, and in general we get $\mathrm{Tr}(a + b\sqrt{d}) = 2a$ (since the trace is $\mathbf{Q}$-linear).

We have two cases: if d is not 1 modulo 4, then we can use the $\mathbf{Z}$-basis $(1, \sqrt{d})$, and the corresponding matrix whose determinant gives the discriminant is

$$\begin{pmatrix} \mathrm{Tr}(1) & \mathrm{Tr}(\sqrt{d}) \\ \mathrm{Tr}(\sqrt{d}) & \mathrm{Tr}(d) \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2d \end{pmatrix},$$

hence $\mathrm{disc}(\mathbf{Z}_{\mathrm{K}}) = 4d$. On the other hand, if $d \equiv 1 \pmod 4$, then we can take the basis $(1, (1 + \sqrt{d})/2)$; since

$$\Big(\frac{1 + \sqrt{d}}{2}\Big)^2 = \frac{d + 1 + \sqrt{d}}{4},$$

the matrix to consider is now

$$\begin{pmatrix} 2 & 1 \\ 1 & \frac{d+1}{2} \end{pmatrix},$$

hence $\mathrm{disc}(\mathbf{Z}_{\mathrm{K}}) = d$.

2.5. Ideal structure

We intend to use the rings of integers in a number field for arguments similar to those one does with "ordinary" integers. In particular, we want to study the divisibility relation in $\mathbf{Z}_{\mathrm{K}}$, and the existence (or not) of unique factorization. The following questions are therefore relevant:

(1) What can be said about ideals in $\mathbf{Z}_{\mathrm{K}}$? What can be said about the quotient rings $\mathbf{Z}_{\mathrm{K}}/\boldsymbol{n}$ when $\boldsymbol{n} \subset \mathbf{Z}_{\mathrm{K}}$ is an ideal?

(2) What can be said about the prime ideals and the maximal ideals in $\mathbf{Z}_{\mathrm{K}}$?

(3) What can be said about the *units* of $\mathbf{Z}_{\mathrm{K}}$?

(4) Is $\mathbf{Z}_{\mathrm{K}}$ is Unique Factorization Domain? A Principal Ideal Domain?

The answers to these questions give the foundations for all of algebraic number theory. Even here, however, we will quickly see problems which are still unsolved.

The first important fact for number theory, and the first parallel with $\mathbf{Z}$, is the following:

PROPOSITION 2.5.1. *Let* K *be a number field. For any non-zero ideal* $\boldsymbol{n} \subset \mathbf{Z}_{\mathrm{K}}$*, the quotient ring* $\mathbf{Z}_{\mathrm{K}}/\boldsymbol{n}$ *is finite.*

PROOF. Since $\boldsymbol{n}$ is non-zero, there is some non-zero element $\alpha \in \boldsymbol{n}$. Writing an equation

$$\alpha^d + a_{d-1}\alpha^{d-1} + \cdots + a_1\alpha + a_0 = 0$$

with $d \geqslant 1$ satisfied by α, with $a_0 \neq 0$, we observe that since $\alpha \in \boldsymbol{n}$, we have

$$a_0 = -\alpha(\alpha^{d-1} + a_{d-1}\alpha^{d-2} + \cdots + a_1) \in \boldsymbol{n}.$$

Since the inclusion $a_0\mathbf{Z}_{\mathrm{K}} \subset \boldsymbol{n}$ gives an injective map $\mathbf{Z}_{\mathrm{K}}/\boldsymbol{n} \to \mathbf{Z}_{\mathrm{K}}/a_0\mathbf{Z}_{\mathrm{K}}$, we are reduced to showing that $\mathbf{Z}_{\mathrm{K}}/a_0\mathbf{Z}_{\mathrm{K}}$ is finite. To see this, fix a $\mathbf{Z}$-basis (ω_i) of $\mathbf{Z}_{\mathrm{K}}$, and note that we have an isomorphism

$$(\mathbf{Z}/a_0\mathbf{Z})^{[\mathrm{K}:\mathbf{Q}]} \to \mathbf{Z}_{\mathrm{K}}/a_0\mathbf{Z}_{\mathrm{K}}$$

induced by the $\mathbf{Z}$-linear isomorphism $\mathbf{Z}^{[\mathrm{K}:\mathbf{Q}]} \to \mathbf{Z}_{\mathrm{K}}$ given by the chosen basis. □

DEFINITION 2.5.2. Let K be a number field. For a non-zero ideal $\boldsymbol{n} \subset \mathbf{Z}_{\mathrm{K}}$, the *norm* of $\boldsymbol{n}$ is the size of the quotient ring $\mathbf{Z}_{\mathrm{K}}/\boldsymbol{n}$; it is denoted $|\boldsymbol{n}|$.

REMARK 2.5.3. It is more traditional to use gothic fonts, such as $\mathfrak{n}$ or $\mathfrak{q}$ for ideals in number fields. Similarly, the usual notation for the norm of an ideal $\boldsymbol{n}$ would be $\mathrm{N}(\boldsymbol{n})$. We try to use a lighter notation which emphasizes the similarity with the case of $\mathbf{Z}$.

EXAMPLE 2.5.4. The proof of the proposition shows that if $\alpha \in \mathbf{Z}$, we have $|\alpha\mathbf{Z}_{\mathrm{K}}| = |\alpha|^{[\mathrm{K}:\mathbf{Q}]}$, where $|\alpha|$ on the right-hand side refers to the usual absolute value.

The finiteness of quotient rings has immediate consequences on the structure of the ring $\mathbf{Z}_{\mathrm{K}}$:

COROLLARY 2.5.5. *Let* K *be a number field. The ring* $\mathbf{Z}_{\mathrm{K}}$ *is a noetherian ring, and all its non-zero prime ideals are maximal.*

PROOF. Recall that a ring A is noetherian if and only if all increasing sequences of ideals

$$\mathrm{I}_1 \subset \mathrm{I}_2 \subset \cdots \subset \mathrm{I}_n \subset \cdots$$

are stationary: there exist m such that $\mathrm{I}_m = \mathrm{I}_n$ for $n \geqslant m$. In the case of $\mathbf{Z}_\mathrm{K}$, we observe that such a chain gives rise to a decreasing sequence of quotient rings

$$\mathbf{Z}_\mathrm{K}/\mathrm{I}_1 \supset \mathbf{Z}_\mathrm{K}/\mathrm{I}_2 \supset \cdots \supset \mathbf{Z}_\mathrm{K}/\mathrm{I}_n \supset \cdots$$

and since the starting point $\mathbf{Z}_\mathrm{K}/\mathrm{I}_1$ is finite, there is some m such that $\mathbf{Z}_\mathrm{K}/\mathrm{I}_m = \mathbf{Z}_\mathrm{K}/\mathrm{I}_n$ for all $n \geqslant m$. But then it follows that $\mathrm{I}_m = \mathrm{I}_n$ for $n \geqslant m$ (because $\mathrm{I}_n = \ker(\mathbf{Z}_\mathrm{K} \to \mathbf{Z}_\mathrm{K}/\mathrm{I}_n)$).

For the second statement, let $\boldsymbol{p} \subset \mathbf{Z}_\mathrm{K}$ be a non-zero prime ideal. Then the quotient ring is a finite integral domain, and it is known that such a finite ring is a field; this means that $\boldsymbol{p}$ is a maximal ideal. □

Remark 2.5.6. (1) The fact that $\mathbf{Z}_\mathrm{K}$ is noetherian can be proved in other ways: for instance, any ideal $\boldsymbol{n}$ is, as an abelian group, a subgroup of $\mathbf{Z}_\mathrm{K}$, which has finite rank. So $\boldsymbol{n}$ is a finitely-generated abelian group, and *a fortiori* a finitely-generated $\mathbf{Z}_\mathrm{K}$-module.

(2) In the terminology of commutative algebra and algebraic geometry, the fact that the non-zero prime ideals of $\mathbf{Z}_\mathrm{K}$ are prime, but $\mathbf{Z}_\mathrm{K}$ is not a field, means that $\mathbf{Z}_\mathrm{K}$ has *Krull dimension* equal to 1. It is a very important property, that leads in modern algebraic geometry (the theory of *schemes*) to the viewpoint that $\mathbf{Z}_\mathrm{K}$ should be seen as a geometric object, and that this object is a *curve*.

Up to now, most of the properties we have discussed are shared by the rings $\mathbf{Z}[\alpha]$, for $\alpha \in \mathbf{Z}_\mathrm{K}$, even when they do not coincide with $\mathbf{Z}_\mathrm{K}$. This is *not the case* anymore for the following essential result, which justifies the definition of $\mathbf{Z}_\mathrm{K}$. Before stating it, recall that for any commutative ring A, and any ideals I_1 and I_2 in A, the product ideal $\mathrm{I}_1\mathrm{I}_2$ is defined as the ideal generated by elements of the form a_1a_2 with $a_1 \in \mathrm{I}_1$ and $a_2 \in \mathrm{I}_2$. We have $\mathrm{I}_1\mathrm{I}_2 \subset \mathrm{I}_1 \cap \mathrm{I}_2$ (by the ideal property), but there is no equality in general. It is important to note that associativity and commutativity hold for these products: we have $\mathrm{I}_1(\mathrm{I}_2\mathrm{I}_3) = (\mathrm{I}_1\mathrm{I}_2)\mathrm{I}_3$ and $\mathrm{I}_1\mathrm{I}_2 = \mathrm{I}_2\mathrm{I}_1$ (both because these have the same generating sets, using associativity and commutativity of the multiplication in A). Moreover, we can generalize the product to consider the product $\mathrm{I}_1\mathrm{I}_2$ of $\mathbf{Z}_\mathrm{K}$-submodules of the field K (for instance $\frac{1}{2}\mathbf{Z} \cdot \frac{1}{2}\mathbf{Z} = \frac{1}{4}\mathbf{Z} \subset \mathbf{Q}$). Finally, we recall that $\mathrm{I}_1 \subset \mathrm{I}_2$ is also denoted $\mathrm{I}_2 \mid \mathrm{I}_1$, and corresponds to divisibility of ideals.

Theorem 2.5.7 (Dedekind). *Let* K *be a number field. The ring of integers* $\mathbf{Z}_\mathrm{K}$ *is a* Dedekind domain*, i.e., the non-zero ideals of* $\mathbf{Z}_\mathrm{K}$ *admit unique factorizations as products of prime ideals.*

In concrete terms, this result can be rephrased as the existence of $\boldsymbol{p}$-adic valuations for all non-zero prime ideals $\boldsymbol{p}$ of $\mathbf{Z}_\mathrm{K}$: denoting by $\mathscr{I}(\mathbf{Z}_\mathrm{K})$ the set of all non-zero ideals of $\mathbf{Z}_\mathrm{K}$ there are well-defined maps

$$v_{\boldsymbol{p}} \colon \mathscr{I}(\mathbf{Z}_\mathrm{K}) \to \mathbf{Z}$$

such that for any non-zero ideal $\boldsymbol{n}$, all but finitely many of the values $v_{\boldsymbol{p}}(\boldsymbol{n})$ are non-zero, and

$$\boldsymbol{n} = \prod_{\boldsymbol{p}} \boldsymbol{p}^{v_{\boldsymbol{p}}(\boldsymbol{n})},$$

and these $\boldsymbol{p}$-adic valuations satisfy the usual properties

$$v_{\boldsymbol{p}}(\boldsymbol{nm}) = v_{\boldsymbol{p}}(\boldsymbol{n}) + v_{\boldsymbol{p}}(\boldsymbol{m}), \tag{2.2}$$

$$v_{\boldsymbol{p}}(\boldsymbol{n} + \boldsymbol{m}) \geqslant \min(v_{\boldsymbol{p}}(\boldsymbol{n}), v_{\boldsymbol{p}}(\boldsymbol{m})). \tag{2.3}$$

The uniqueness of prime factorization means that

$$\boldsymbol{m} = \boldsymbol{n} \text{ if and only if } v_{\boldsymbol{p}}(\boldsymbol{m}) = v_{\boldsymbol{p}}(\boldsymbol{n})$$

for all non-zero prime ideals $\boldsymbol{p}$. (In particular, note that this implies that $v_{\boldsymbol{p}}(\boldsymbol{p}) = 1$ and $v_{\boldsymbol{q}}(\boldsymbol{p}) = 0$ if $\boldsymbol{q}$ is a prime ideal different from $\boldsymbol{p}$, using the known factorization $\boldsymbol{p} = \boldsymbol{p}^1$ for $\boldsymbol{p}$.

It also follows that for non-zero ideals $\boldsymbol{m}$ and $\boldsymbol{n}$, we have $\boldsymbol{m} \subset \boldsymbol{n}$ (equivalently, $\boldsymbol{n} \mid \boldsymbol{m}$) if and only if

$$v_{\boldsymbol{p}}(\boldsymbol{n}) \leqslant v_{\boldsymbol{p}}(\boldsymbol{m})$$

for all non-zero prime ideals $\boldsymbol{p} \subset \mathbf{Z}_{\mathrm{K}}$. We then have a product decomposition $\boldsymbol{m} = \boldsymbol{n}\boldsymbol{n}'$ with

$$\boldsymbol{n}' = \prod_{\boldsymbol{p}} \boldsymbol{p}^{v_{\boldsymbol{p}}(\boldsymbol{m}) - v_{\boldsymbol{p}}(\boldsymbol{n})}.$$

Sketch of proof of Theorem 2.5.7. We sketch a fairly algebraic proof, which relies only on the fact that $\mathbf{Z}_{\mathrm{K}}$ is noetherian, integrally closed, and that its non-zero prime ideals are maximal, and follows Samuel's account in [18, §3.4]. A more elementary argument is given in [12, §12.2], relying on first proving Theorem 2.5.12 below. We should emphasize that all the small non-obvious details in the proof are simple consequences of the result itself – in this sense, knowing the proof is not really necessary to go further in understanding number fields (see Example 2.5.10 below).

Step 1. Let $\boldsymbol{p} \subset \mathbf{Z}_{\mathrm{K}}$ be a non-zero prime ideal. Then one shows that the $\mathbf{Z}_{\mathrm{K}}$-submodule

$$\widetilde{\boldsymbol{p}} = \{x \in \mathrm{K} \mid x\boldsymbol{p} \subset \mathbf{Z}_{\mathrm{K}}\}$$

of K has the property that $\boldsymbol{p}\widetilde{\boldsymbol{p}} = \mathbf{Z}_{\mathrm{K}}$. It is this step which fails for rings $\mathbf{Z}[\alpha] \subset \mathbf{Z}_{\mathrm{K}}$ in general, because they are not always integrally closed.

More precisely, the $\mathbf{Z}_{\mathrm{K}}$-module $\boldsymbol{p}\widetilde{\boldsymbol{p}}$ is contained in $\mathbf{Z}_{\mathrm{K}}$, so is an ideal in $\mathbf{Z}_{\mathrm{K}}$. Since $1 \in \widetilde{\boldsymbol{p}}$, this ideal contains $\boldsymbol{p}$. By maximality of the non-zero prime ideal $\boldsymbol{p}$, either $\boldsymbol{p}\widetilde{\boldsymbol{p}} = \boldsymbol{p}$ or $\boldsymbol{p}\widetilde{\boldsymbol{p}} = \mathbf{Z}_{\mathrm{K}}$, as we are trying to prove.

Assume then for contradiction that $\boldsymbol{p}\widetilde{\boldsymbol{p}} = \boldsymbol{p}$. The next claim, where the fact that $\mathbf{Z}_{\mathrm{K}}$ is integrally closed appears, is that this implies that $\widetilde{\boldsymbol{p}} = \mathbf{Z}_{\mathrm{K}}$. Indeed, assume that $\boldsymbol{p}\widetilde{\boldsymbol{p}} = \boldsymbol{p}$; for any element $x \in \widetilde{\boldsymbol{p}}$, we find that $\mathbf{Z}[x]\boldsymbol{p} \subset \boldsymbol{p}$. It follows that $\mathbf{Z}[x]$ is a finitely-generated abelian group, and this implies that x is integral over $\mathbf{Z}$. Hence $x \in \mathbf{Z}_{\mathrm{K}}$ by definition of $\mathbf{Z}_{\mathrm{K}}$.

From there, we get a contradiction by checking that $\widetilde{\boldsymbol{p}}$, which contains $\mathbf{Z}_{\mathrm{K}}$ because $\boldsymbol{p}$ is an ideal of $\mathbf{Z}_{\mathrm{K}}$, is not equal to $\mathbf{Z}_{\mathrm{K}}$. To see this, we pick some non-zero element $a \in \boldsymbol{p}$. One shows that the ideal $a\mathbf{Z}_{\mathrm{K}}$ *divides* (in other words, contains) some ideal $\boldsymbol{n}$ which is a product of non-zero prime ideals (the argument applies to any non-zero ideal: considering the set X of non-zero ideals which do not contain a product of prime ideals, one checks that if X were not empty, there would be no maximal element of X for inclusion, which would contradict the fact that $\mathbf{Z}_{\mathrm{K}}$ is noetherian). Then $\boldsymbol{p}$, which divides $a\mathbf{Z}_{\mathrm{K}}$, divides one of the prime ideals into which $\boldsymbol{n}$ factors, and by maximality is equal to one of them; this gives the relation $a\mathbf{Z}_{\mathrm{K}} \mid \boldsymbol{p}\boldsymbol{m}$ for some non-zero ideal $\boldsymbol{m}$. One may assume that $a\mathbf{Z}_{\mathrm{K}}$ does not divide $\boldsymbol{m}$ (otherwise, use the fact that $\boldsymbol{m}$ is by assumption a product of prime ideals to repeat the argument with $\boldsymbol{n}$ replaced by $\boldsymbol{m}$). So $a\mathbf{Z}_{\mathrm{K}} \mid b\,\boldsymbol{p}$ for some $b \notin a\mathbf{Z}_{\mathrm{K}}$ (any element of $\boldsymbol{m} - a\mathbf{Z}_{\mathrm{K}}$), and therefore $b/a \in \widetilde{\boldsymbol{p}}$ but $b/a \notin \mathbf{Z}_{\mathrm{K}}$.

Step 2. We prove the existence of prime factorization of ideals using a common technique in noetherian rings: let X be the set of ideals $\boldsymbol{n} \subset \mathbf{Z}_{\mathrm{K}}$ which are *not* products of primes; then if X is not empty (which we try to exclude), it has a maximal element by another characterization of noetherian rings, say $\boldsymbol{n}$. Then $\boldsymbol{n}$ is a proper ideal, so is contained in some maximal ideal $\boldsymbol{p}$, and is different from $\boldsymbol{p}$. Then $\boldsymbol{n} \subset \boldsymbol{n}\widetilde{\boldsymbol{p}} \subset \mathbf{Z}_{\mathrm{K}}$, and

the first inclusion is strict (because $\boldsymbol{np}$ is strictly contained in $\boldsymbol{n}$, this being as in Step 1 a consequence of the fact that $\mathbf{Z}_{\mathrm{K}}$ is integrally closed). So $\boldsymbol{n}\widetilde{\boldsymbol{p}}$ is not in X, and thus is a product of prime ideals; this is a contradiction, since then we would have $\boldsymbol{n} = (\boldsymbol{n}\widetilde{\boldsymbol{p}})\boldsymbol{p} \in \mathrm{X}$.

Step 3. Finally, we prove uniqueness. Suppose

$$\boldsymbol{p}_1 \cdots \boldsymbol{p}_k = \boldsymbol{q}_1 \cdots \boldsymbol{q}_l$$

with $\boldsymbol{p}_i$ and $\boldsymbol{q}_j$ non-zero prime ideals. Then $\boldsymbol{p}_1$ contains the product of the ideals $\boldsymbol{q}_j$, which by a standard property of prime ideals means that $\boldsymbol{p}_1 \supset \boldsymbol{q}_j$ for some j. Since the prime ideals in question are maximal, we get $\boldsymbol{p}_1 = \boldsymbol{q}_j$. Now multiplying both sides with $\widetilde{\boldsymbol{p}}_1$, we obtain an equality

$$\boldsymbol{p}_2 \cdots \boldsymbol{p}_k = \boldsymbol{q}_1 \cdots \boldsymbol{q}_{j-1}\boldsymbol{q}_{j+1} \cdots \boldsymbol{q}_l,$$

with fewer terms, and we conclude by induction. □

This has already one very useful consequence, which again is not always valid for the simpler rings $\mathbf{Z}[\alpha]$.

Corollary 2.5.8. *Let* K *be a number field. For any ideal* $\boldsymbol{n}$ *and* $\boldsymbol{m}$ *of* $\mathbf{Z}_{\mathrm{K}}$*, we have* $|\boldsymbol{nm}| = |\boldsymbol{n}||\boldsymbol{m}|$.

Proof. Using prime factorization for $\boldsymbol{m}$ and induction, it suffices to prove that $|\boldsymbol{np}| = |\boldsymbol{n}||\boldsymbol{p}|$ when $\boldsymbol{p}$ is a non-zero prime ideal of $\mathbf{Z}_{\mathrm{K}}$. We have a standard isomorphism of finite abelian groups

$$(\mathrm{A}/\boldsymbol{np})/(\boldsymbol{n}/\boldsymbol{pn}) \to \mathrm{A}/\boldsymbol{n},$$

so that

$$|\boldsymbol{np}| = |\boldsymbol{n}|\ |\boldsymbol{np}/\boldsymbol{n}|,$$

where one must be somewhat careful that the right-hand side involves the norm of an ideal and the size of a quotient group. The abelian group $\mathrm{E} = \boldsymbol{n}/\boldsymbol{pn}$ is stable under multiplication by $\boldsymbol{p}$, so it is naturally a vector space over the finite quotient field $k = \mathbf{Z}_{\mathrm{K}}/\boldsymbol{p}$. To compute its size, it suffices to compute its k-dimension; for this, we observe that a k-linear subspace $\mathrm{F} \subset \mathrm{E}$ corresponds (by taking inverse image under the projection $\boldsymbol{n} \to k$) to ideals $\boldsymbol{m}$ such that $\boldsymbol{np} \subset \boldsymbol{m} \subset \boldsymbol{n}$. Looking at the $\boldsymbol{p}$-adic valuation of $\boldsymbol{m}$, it is either 0 or 1 because of (2.2), and it follows that F is either 0 or equal to E. Since $\mathrm{E} \neq \{0\}$ (because $v_{\boldsymbol{p}}(\boldsymbol{np}) = v_{\boldsymbol{p}}(\boldsymbol{n})+1$), we deduce that $\dim_k(\mathrm{F}) = 1$, and then that $|\mathrm{F}| = |k| = |\boldsymbol{p}|$. The product formula follows. □

Example 2.5.9. We illustrate the failure of Theorem 2.5.7, through that of Step 1 of its proof, in one of the simplest examples. Let $\mathrm{K} = \mathbf{Q}(\sqrt{5})$, whose ring of integers is $\mathbf{Z}_{\mathrm{K}} = \mathbf{Z}[(1+\sqrt{5})/2]$. Let $\mathrm{A} = \mathbf{Z}[\sqrt{5}]$.

We note that we can also express A in the form $\mathrm{A} = \mathbf{Z} + 2\mathbf{Z}_{\mathrm{K}}$ (e.g., $a + b\sqrt{5} = (a-2b) + 2b(1+\sqrt{5})/2$). In particular, $2\mathbf{Z}_{\mathrm{K}}$ is contained in A, and so is an ideal in A. The quotient $\mathrm{A}/2\mathbf{Z}_{\mathrm{K}}$ is $(\mathbf{Z} + 2\mathbf{Z}_{\mathrm{K}})/2\mathbf{Z}_{\mathrm{K}}$ and is isomorphic to $\mathbf{Z}/(2\mathbf{Z}_{\mathrm{K}} \cap \mathbf{Z}) = \mathbf{Z}/2\mathbf{Z}$. In particular, this ideal $\boldsymbol{p} = 2\mathbf{Z}_{\mathrm{K}} \subset \mathrm{A}$ is a prime ideal of A. (Concretely, we just have

$$2\mathbf{Z}_{\mathrm{K}} = \{a + b\sqrt{5} \in \mathbf{Z}[\sqrt{5}] \mid a \equiv b \,(\mathrm{mod}\, 2)\},$$

so the isomorphism with $\mathbf{Z}/2\mathbf{Z}$ is given by $a + b\sqrt{5} \mapsto a \,(\mathrm{mod}\, 2)$.)

Let $\widetilde{\boldsymbol{p}} = \{x \in \mathrm{K} \mid x\boldsymbol{p} \subset \mathrm{A}\}$. If we had $\boldsymbol{p}\widetilde{\boldsymbol{p}} = \mathrm{A}$, then we would obtain

$$\{x \in \mathrm{K} \mid x\boldsymbol{p} \subset \boldsymbol{p}\} = \mathrm{A},$$

(by multiplying both sides of an inclusion $x\boldsymbol{p} \subset \boldsymbol{p}$ by $\widetilde{\boldsymbol{p}}$ to check that the left-hand side is contained in A, the other inclusion being elementary).

But we can check directly that

$$\{x \in \mathrm{K} \mid x\boldsymbol{p} \subset \boldsymbol{p}\} = \mathbf{Z}_{\mathrm{K}}.$$

Indeed, $\mathbf{Z}_{\mathrm{K}}$ is contained in the right-hand side, since $\boldsymbol{p}$ is an ideal of $\mathbf{Z}_{\mathrm{K}}$ also. Conversely, if $x\boldsymbol{p} \subset \boldsymbol{p}$, then $2x \in 2\mathbf{Z}_{\mathrm{K}}$, which when writing $x = \alpha + \beta\sqrt{5}$ with α, $\beta \in \mathbf{Q}$, translates to

$$2\alpha + 2\beta\sqrt{5} \in \{a + b\sqrt{5} \in \mathbf{Z}[\sqrt{5}] \mid a \equiv b \,(\mathrm{mod}\, 2)\},$$

so $\alpha = \mu/2$ and $\beta = \nu/2$ for some integers μ and ν with $\mu \equiv \nu \,(\mathrm{mod}\, 2)$, which is exactly the description of $\mathbf{Z}_{\mathrm{K}}$.

The conclusion is that Step 1 in the proof of Theorem 2.5.7 fails for this prime ideal $\boldsymbol{p}$. Elaborating these arguments, one can see that $2\mathrm{A} \subset \mathrm{A}$ is an ideal which is *not* a product of prime ideals, and moreover one can check that $|(2\mathrm{A})^2| = 8 \neq |2\mathrm{A}|^2 = 4$, the norm for non-zero ideals I of A being defined again as the size of the quotient ring A/I (which is finite), which shows concretely that Corollary 2.5.8 fails.

Example 2.5.10. We illustrate how, assuming Theorem 2.5.7, one can recover the property of the key Step 1 in the proof. So let $\boldsymbol{p} \subset \mathbf{Z}_{\mathrm{K}}$ be a non-zero prime ideal, and define $\widetilde{\boldsymbol{p}} = \{x \in \mathrm{K} \mid x\boldsymbol{p} \subset \mathbf{Z}_{\mathrm{K}}\}$. We have $\boldsymbol{p} \subset \boldsymbol{p}\widetilde{\boldsymbol{p}} \subset \mathbf{Z}_{\mathrm{K}}$, or in other words $\boldsymbol{p} \mid \boldsymbol{p}\widetilde{\boldsymbol{p}} \mid \mathbf{Z}_{\mathrm{K}}$. The only possibility if $\boldsymbol{p}\widetilde{\boldsymbol{p}} \neq \mathbf{Z}_{\mathrm{K}}$ is $\boldsymbol{p}\widetilde{\boldsymbol{p}} = \boldsymbol{p}$. But there is some $\alpha \in \mathbf{Z}_{\mathrm{K}} - \{0\}$ such that $\boldsymbol{q} = \alpha\widetilde{\boldsymbol{p}}$ is an ideal of $\mathbf{Z}_{\mathrm{K}}$; multiplying by α would give $\boldsymbol{p}\boldsymbol{q} = \alpha\boldsymbol{p}$, and therefore $\alpha\widetilde{\boldsymbol{p}} = \boldsymbol{q} = \alpha\mathbf{Z}_{\mathrm{K}}$, which implies that $\widetilde{\boldsymbol{p}} = \mathbf{Z}_{\mathrm{K}}$.

In fact, something more general holds in $\mathbf{Z}_{\mathrm{K}}$: given a non-zero ideal $\boldsymbol{n} \subset \mathbf{Z}_{\mathrm{K}}$, there is a unique $\mathbf{Z}_{\mathrm{K}}$-submodule $\boldsymbol{q}$ of K such that $\boldsymbol{n}\boldsymbol{q} = \mathbf{Z}_{\mathrm{K}}$, and it is given by

$$\boldsymbol{q} = \{x \in \mathrm{K} \mid x\boldsymbol{n} \subset \mathbf{Z}_{\mathrm{K}}\}.$$

This is even more general than Step 1, since it applies to *any* non-zero ideal, and not just to prime ideals. To prove this, we first show that some $\boldsymbol{q}$ exists with the desired property: namely, write $\boldsymbol{n} = \boldsymbol{p}_1 \cdots \boldsymbol{p}_k$, and take $\boldsymbol{q} = \widetilde{\boldsymbol{p}}_1 \cdots \widetilde{\boldsymbol{p}}_k$. To get the uniqueness, observe that if $\boldsymbol{n}\boldsymbol{q} = \mathbf{Z}_{\mathrm{K}}$, then the $\mathbf{Z}_{\mathrm{K}}$-module defined by

$$\widetilde{\boldsymbol{n}} = \{x \in \mathrm{K} \mid x\boldsymbol{n} \subset \mathbf{Z}_{\mathrm{K}}\}$$

satisfies $\boldsymbol{q} \subset \widetilde{\boldsymbol{n}}$. To get the converse inclusion, note that if $x \in \widetilde{\boldsymbol{n}}$, then $x\boldsymbol{n} \subset \mathbf{Z}_{\mathrm{K}}$, which implies by multiplying by $\boldsymbol{q}$ that $x\mathbf{Z}_{\mathrm{K}} \subset \boldsymbol{q}$, i.e. $x \in \boldsymbol{q}$.

Theorem 2.5.7 applies in particular to principal ideals $\alpha\mathbf{Z}_{\mathrm{K}}$, for $\alpha \in \mathbf{Z}_{\mathrm{K}}$, but in a factorization

$$\alpha\mathbf{Z}_{\mathrm{K}} = \boldsymbol{p}_1 \cdots \boldsymbol{p}_k,$$

the prime ideals $\boldsymbol{p}_j$ are usually not principal. This is one of the major differences between the arithmetic of $\mathbf{Q}$ and $\mathbf{Z}$ and that of number fields. However, a crucial fact is that, in some sense, the difference is always under control. Before stating the precise form of this result, we consider an example which indicates that the theory we are discussing has some depth: we will recover Fermat's Theorem on sums of two squares.

Example 2.5.11. Let $\mathrm{K} = \mathbf{Q}(i) \subset \mathbf{C}$. The ring of integers is $\mathbf{Z}[i]$ (see Example 2.4.9, (1)). Its elements are called *gaussian integers*. The ring $\mathbf{Z}[i]$ is (exceptionnally) a principal ideal domain. In fact, it is a euclidean ring: for any x and $y \in \mathbf{Z}[i]$, provided $y \neq 0$, we

can find unique elements q and r in $\mathbf{Z}[i]$, with[1] $\|r\| < \|y\|$, such that

$$x = qy + r$$

(geometrically, take r to be the gaussian integer closest to x/y in $\mathbf{C}$, with respect to the usual distance in the plane, and note that $\|r - x/y\| < 1$). Using this, the proof that an ideal $\boldsymbol{n} \subset \mathbf{Z}[i]$ is principal proceeds identically to the proof that $\mathbf{Z}$ is principal (namely, if $\boldsymbol{n} = \{0\}$, there is nothing to prove; otherwise, take an element in $\boldsymbol{n}$ of minimal modulus, and show using euclidean division that it generates $\boldsymbol{n}$).

Let p be an odd prime number. We claim that p is a sum of two squares of integers if and only if p factors non-trivially in $\mathbf{Z}[i]$, i.e., if $p = (a + ib)(c + id)$ where $a + ib$ and $c + id$ are not units. To check this, we first compute the units: the inverse

$$\frac{1}{a + ib} = \frac{a}{a^2 + b^2} - i\frac{b}{a^2 + b^2}$$

in $\mathbf{Q}(i)$ of a non-zero element of $\mathbf{Z}[i]$ is in $\mathbf{Z}[i]$ if and only if $a^2 + b^2$ divides a and b, which means that one of a and b must be zero, and the other is -1 or 1; in other words, the group of units of $\mathbf{Z}[i]$ is $\{1, i, -1, -i\}$.

Then, first of all, if $p = a^2 + b^2$ (with a and b in $\mathbf{Z}$), then $p = (a + ib)(a - ib)$ is a non-trivial factorization. Conversely, from $p = (a + ib)(c + id)$, we get $p^2 = (a^2 + b^2)(c^2 + d^2)$, which by unique factorization in $\mathbf{Z}$ implies that $p = a^2 + b^2 = c^2 + d^2$. This gives the characterization we claimed. Furthermore, p factors non-trivially in $\mathbf{Z}[i]$ if and only if $p\mathbf{Z}[i]$ is not a prime ideal in $\mathbf{Z}[i]$.

If $p \equiv 1 \pmod 4$, there exists an integer n with $n^2 + 1 \equiv 0 \pmod p$. Then $(n + i)(n - i)$ is in $p\mathbf{Z}[i]$ without either of the two factors being there; hence $p\mathbf{Z}[i]$ is not a prime ideal, and consequently p must be a sum of two squares.

So from very basic facts (which do not require the full theory since existence and uniqueness of prime factorizations in $\mathbf{Z}[i]$ follow easily from the fact that it is a Principal Ideal Domain), we have recovered Fermat's Theorem. Moreover, the uniqueness of prime factorization allows us to go beyond our first proof of Theorem 1.2.2: since writing $p = a^2 + b^2$ is equivalent to saying that

$$p\mathbf{Z}[i] = (a + ib)\mathbf{Z}[i]\ (a - ib)\mathbf{Z}[i],$$

and the ideals $(a + ib)\mathbf{Z}[i]$ and $(a - ib)\mathbf{Z}[i]$ are prime (their norms have to be equal to p since $p^2 = |p\mathbf{Z}[i]|$, so the quotients $\mathbf{Z}[i]/(a + ib)\mathbf{Z}[i]$ and $\mathbf{Z}[i]/(a - ib)\mathbf{Z}[i]$ have order p, and therefore are isomorphic to $\mathbf{F}_p$), any representation $p = c^2 + d^2$ must satisfy either $(c + id)\mathbf{Z}[i] = (a + ib)\mathbf{Z}[i]$ or $(c + id)\mathbf{Z}[i] = (a - ib)\mathbf{Z}[i]$. This means that $c + id$ differs from one of $a + ib$ and $a - ib$ by multiplication by an element in the group $\{-1, -i, i, 1\}$ of units of $\mathbf{Z}[i]$. So the possibilities for $c + id$ are

$$(c, d) \in \{(a, b), (a, -b), (-a, -b), (-a, b), (-b, a), (b, a), (b, -a), (-b, -a)\},$$

and if we insist that a and b be positive, the only possibility is exchanging a and b, as announced in the original statement of Theorem 1.2.2.

THEOREM 2.5.12 (Dedekind). *Let* K *be a number field.*

(1) *There exists an integer* $h \geqslant 1$ *such that* $\boldsymbol{n}^h$ *is principal for* all *ideals* $\boldsymbol{n} \subset \mathbf{Z}_{\mathrm{K}}$.

[1] To avoid confusion with the norm of ideals and elements of $\mathbf{Q}(i)$, we use $\|z\|$ for the usual modulus of a complex number; it is the euclidean norm in $\mathbf{C}$.

(2) *More precisely, let* $\mathscr{P}(\mathbf{Z}_K)$ *denote the set of non-zero principal ideals in* $\mathbf{Z}_K$*. Define the equivalence relation* $\sim$ *on* $\mathscr{I}(\mathbf{Z}_K)$ *by* $\boldsymbol{n} \sim \boldsymbol{m}$ *if and only if there exists* $\alpha\mathbf{Z}_K$ *and* $\beta\mathbf{Z}_K$ *in* $\mathscr{P}(\mathbf{Z}_K)$ *such that*

$$(\alpha\mathbf{Z}_K)\boldsymbol{n} = (\beta\mathbf{Z}_K)\boldsymbol{m}.$$

Then the product of ideals induces on the quotient set $\mathscr{I}(\mathbf{Z}_K)/\mathscr{P}(\mathbf{Z}_K)$ *a structure of finite abelian group.*

Sketch of proof. We first observe that it is straightforward that the relation $\boldsymbol{m} \sim \boldsymbol{n}$ is an equivalence relation.

To prove the finiteness of the number of equivalence classes, we follow the argument of Hurwitz which is also given in [12, §12.2, Th.1]; it is natural from the point of view of trying to go as far as possible when trying the imitate the proof that euclidean domains are principal ideal domains, using euclidean division.

Hurwitz's statement is the following: there exists an integer $m \geqslant 1$ such that for any $x \in \mathbf{Z}_K$ and $y \in \mathbf{Z}_K - \{0\}$, there exist q and r in $\mathbf{Z}_K$ and an integer k with $1 \leqslant k \leqslant m$ such that

$$kx = qy + r, \qquad \text{with either } r = 0 \text{ or } |r\mathbf{Z}_K| < |y\mathbf{Z}_K|.$$

(In other words: we might not have a "good" division of x by y, but some small multiple of x has a "good" division.)

Assuming this property (which we will prove later – it is a clever application of the *pigeon-hole principle*), we first show that the quotient set $\mathscr{H}(K)$ is finite by finding an explicit set of ideals which contains a representative of any ideal class.

Let $\boldsymbol{n}$ be any non-zero ideal of $\mathbf{Z}_K$. Let $y \in \boldsymbol{n}$ be a non-zero element with $|y\mathbf{Z}_K|$ the smallest possible (in the euclidean case, this would be a generator of $\boldsymbol{n}$). For any $x \in \boldsymbol{n}$, we apply Hurwitz's division property: writing $kx = qy + r$ for some integer k with $1 \leqslant k \leqslant m$, we see that $r = kx - qy \in \boldsymbol{n}$, so we cannot have $|r\mathbf{Z}_K| < |y\mathbf{Z}_K|$ by definition of y. This means that $r = 0$, or in other words, there exists some positive integer $k \leqslant m$ such that $kx = qy \in y\mathbf{Z}_K$. This applies to all x, but with k possibly depending on x. However, in all cases, we get $m!x \in y\mathbf{Z}_K$, so that $m!\boldsymbol{n} \subset y\mathbf{Z}_K$, or equivalently $y\mathbf{Z}_K \mid m!\boldsymbol{n}$.

There is therefore a non-zero ideal $\boldsymbol{m}$ of $\mathbf{Z}_K$ such that $m!\boldsymbol{n} = y\boldsymbol{m}$, which means that $\boldsymbol{n} \sim \boldsymbol{m}$. But since $y \in \boldsymbol{n}$, we have

$$y\boldsymbol{m} = m!\boldsymbol{n} \mid ym!\mathbf{Z}_K,$$

and hence $\boldsymbol{m} \mid m!\mathbf{Z}_K$. So the ideal $\boldsymbol{m}$ is one of the finitely many[2] dividing the non-zero ideal $m!\mathbf{Z}_K$, and hence the finite set of these ideals exhaust the quotient set $\mathscr{H}(K)$.

We next check the first property in the statement. Let $\boldsymbol{n}$ be a non-zero ideal of $\mathbf{Z}_K$. Among the powers $\boldsymbol{n}^j$ with $j \geqslant 1$, there are only finitely many equivalence classes, so there exists integers $1 \leqslant j < k$ such that $\boldsymbol{n}^j \sim \boldsymbol{n}^k$, say

$$\alpha\boldsymbol{n}^j = \beta\boldsymbol{n}^k$$

for some non-zero elements α and β of $\mathbf{Z}_K$. It follows that $\alpha\mathbf{Z}_K = \beta\boldsymbol{n}^{k-j}$. In particular, this gives $\alpha/\beta \in \mathbf{Z}_K$, and $\boldsymbol{n}^{k-j} = (\alpha/\beta)\mathbf{Z}_K$ is principal.

To see that the quotient set is a group, we need here to be a bit careful since $\mathscr{I}(K)$, with the product of ideals, is not a group: although multiplication of ideals is commutative, associative and has the neutral element $\mathbf{Z}_K$, there is no inverse in general. However,

[2] The fact that there are only finitely many ideals dividing a given non-zero ideal is another quick consequence of the existence and uniqueness of factorization in prime ideals: there are only finitely many possible prime ideals that can appear, and each can only appear with exponent bounded by the valuation of the original ideal – we will say more about this question in the next chapter.

given a non-zero ideal $\boldsymbol{n}$, and an integer $h \geqslant 1$ such that $\boldsymbol{n}^h$ is principal, it follows that the class of $\boldsymbol{n}$ in $\mathscr{H}(\mathrm{K})$ has for inverse the class of $\boldsymbol{n}^{h-1}$. The remaining properties of an abelian group follow straightforwardly from the associativity and commutativity of the product of ideals. □

The quotient group in the second part of the theorem is another among the fundamental invariants of a number field. Note for instance that this group is trivial if and only if every ideal is principal, i.e., if and only if $\mathbf{Z}_\mathrm{K}$ is a principal ideal domain. Indeed, we can quickly see that the neutral element in $\mathscr{H}(\mathrm{K})$ is the class of the ideal $\mathbf{Z}_\mathrm{K}$: by definition, to say that $\boldsymbol{n} \sim \mathbf{Z}_\mathrm{K}$ means that there exist non-zero elements α and $\beta \in \mathbf{Z}_\mathrm{K}$ with $\alpha\mathbf{Z}_\mathrm{K} = \beta\boldsymbol{n}$, and this implies that $\alpha/\beta \in \mathbf{Z}_k$ and $\boldsymbol{n} = (\alpha/\beta)\mathbf{Z}_\mathrm{K}$ is principal.

Definition 2.5.13 (Class group). Let K be a number field. The quotient group $\mathscr{I}(\mathbf{Z}_\mathrm{K})/\mathscr{P}(\mathbf{Z}_\mathrm{K})$ is called the *ideal class group* of K, or just *class group* of K. It is denoted $\mathscr{H}(\mathrm{K})$. The order of $\mathscr{H}(\mathrm{K})$ is called the *class number* of K, and is denoted $h(\mathrm{K})$.

Remark 2.5.14. Historically, Gauss proved Theorem 2.5.12 in the case of quadratic fields, but in the language of integral binary quadratic forms instead of quadratic fields. His investigations went quite deep, including computations by hand of the class group of many quadratic fields $\mathbf{Q}(\sqrt{d})$ (both for d positive and negative).

Example 2.5.15. It is clear from the proof of Example 2.5.11 that an interesting theorem about representability of certain primes by certain polynomial expressions will arise whenever a number field has trivial class group. For instance, if d is a square-free integer which is not congruent to 1 modulo 4, so that the ring of integers of $\mathbf{Q}(\sqrt{d})$ is equal to $\mathbf{Z}[\sqrt{d}]$, this will concern the representations $p = a^2 - db^2$. Are the corresponding class groups (typically) trivial?

Here is a list of data $(d, h(\mathbf{Q}(\sqrt{d})))$ for squarefree positive integers $d \leqslant 100$ which are not 1 modulo 4:

$$(2,1),(3,1),(6,1),(7,1),(10,2),(11,1),(14,1),(15,2),(19,1),(221),(231),$$
$$(262),(302),(311),(342),(352),(381),(39,2),(42,2),(43,1),(46,1),(47,1),$$
$$(51,2),(55,2),(58,2),(59,1),(62,1),(66,2),(67,1),(70,2),(71,1),(74,2),(78,2),$$
$$(79,3),(82,4),(83,1),(86,1),(87,2),(91,2),(94,1),(95,2)$$

and here are ten random values for such $d \leqslant 10000$:

$$(7094,1),(3559,1),(6491,1),(4330,4),(939,4),(3910,4),$$
$$(8931,8),(2042,2),(9991,8),(7331,5).$$

Here is the same for negative values of d:

$$(-1,1),(-2,1),(-5,2),(-6,2),(-10,2),(-13,2),(-14,4),(-17,4),(-21,4),$$
$$(-22,2),(-26,6),(-29,6),(-30,4),(-33,4),(-34,4),(-37,2),(-38,6),$$
$$(-41,8),(-42,4),(-46,4),(-53,6),(-57,4),(-58,2),(-61,6),(-62,8),$$
$$(-65,8),(-66,8),(-69,8),(-70,4),(-73,4),(-74,10),(-77,8),(-78,4),$$
$$(-82,4),(-85,4),(-86,10),(-89,12),(-93,4),(-94,8),(-97,4),$$

and

$$(-5405, 64), (-6681, 80), (-8193, 52), (-5242, 42), (-854, 44), (-8574, 52),$$
$$(-4017, 48), (-6518, 42), (-8893, 46), (-2717, 32).$$

2.6. Factoring primes

Let again K be a number field. We now consider in general the way principal ideals generated by primes in $\mathbf{Z}$ factor in $\mathbf{Z}_{\mathrm{K}}$, which we saw appear in Example 2.5.11.

Let p be a prime number. We write a factorization

$$p\mathbf{Z}_{\mathrm{K}} = \boldsymbol{p}_1^{e_1} \cdots \boldsymbol{p}_g^{e_g} \tag{2.4}$$

where the $\boldsymbol{p}_i$'s are distinct non-zero prime ideals in $\mathbf{Z}_{\mathrm{K}}$ and $e_i \geqslant 1$ (note that since $p\mathbf{Z}_{\mathrm{K}} \neq \mathbf{Z}_{\mathrm{K}}$, we have $g \geqslant 1$). The ideals which appear are, by definition, those which divide $p\mathbf{Z}_{\mathrm{K}}$. In general, we will denote by $\mathbf{F}_{\boldsymbol{p}}$, the quotient finite field $\mathbf{Z}_{\mathrm{K}}/\boldsymbol{p}$ for any non-zero prime ideal $\boldsymbol{p}$ (this is called the *residue field* modulo $\boldsymbol{p}$).

LEMMA 2.6.1. (1) *The ideals $\boldsymbol{p}_i$ are characterized as the prime ideals $\boldsymbol{p}$ in $\mathbf{Z}_{\mathrm{K}}$ such that $\boldsymbol{p} \cap \mathbf{Z} = p\mathbf{Z}$.*

(2) *For $1 \leqslant i \leqslant g$, there exists an integer $f_i \geqslant 1$ such that $|\boldsymbol{p}_i| = p^{f_i}$. This integer is also characterized by the fact that the finite field $\mathbf{F}_{\boldsymbol{p}_i}$ is an extension of degree f_i of $\mathbf{F}_p$.*

(3) *We have the relation*

$$[\mathrm{K} : \mathbf{Q}] = \sum_{i=1}^{g} e_i f_i, \tag{2.5}$$

and in particular $g \leqslant [\mathrm{K} : \mathbf{Q}]$, with equality if and only if $e_i = f_i = 1$ for $1 \leqslant i \leqslant g$.

PROOF. (1) The ideals $\boldsymbol{p}_i$ are exactly those prime ideals which divide $p\mathbf{Z}_{\mathrm{K}}$, or in other words which contain it. In particular, $\boldsymbol{p}_i \cap \mathbf{Z} \supset p\mathbf{Z}_{\mathrm{K}} \cap \mathbf{Z} = p\mathbf{Z}$. Moreover, since $\boldsymbol{p}_i \cap \mathbf{Z}$ is a prime ideal in $\mathbf{Z}$, and $p\mathbf{Z}$ is a maximal ideal, we must have $\boldsymbol{p}_i \cap \mathbf{Z} = p\mathbf{Z}$.

Conversely, suppose that $\boldsymbol{p} \subset \mathbf{Z}_{\mathrm{K}}$ is an ideal with $\boldsymbol{p} \cap \mathbf{Z} = p\mathbf{Z}$. Then $p\mathbf{Z} \subset \boldsymbol{p}$, so the ideal $p\mathbf{Z}_{\mathrm{K}}$ generated by $p\mathbf{Z}$ is also contained in $\boldsymbol{p}$, which means that $\boldsymbol{p}$ divides $p\mathbf{Z}_{\mathrm{K}}$, and therefore is one of the $\boldsymbol{p}_i$'s.

We prove (2) and (3) at the same time: taking the norm of both sides of (2.4), we obtain

$$p^{[\mathrm{K}:\mathbf{Q}]} = \prod_{i=1}^{g} |\boldsymbol{p}_i|^{e_i}$$

(by Corollary 2.5.8). It follows that $|\boldsymbol{p}_i|$ must be a power of p, say $|\boldsymbol{p}_i| = p^{f_i}$, and this leads to (2.5). The inclusion $p\mathbf{Z}_{\mathrm{K}} \subset \boldsymbol{p}_i$ gives an induced injective morphism

$$\mathbf{F}_p \to \mathbf{Z}_{\mathrm{K}}/\boldsymbol{p}_i,$$

which means that $\mathbf{F}_{p_i}$ is a finite field of characteristic p. It has size p^{f_i}, so it has degree f_i as an extension of $\mathbf{F}_p$. □

Some special cases are particularly important, and have separate terminology.

DEFINITION 2.6.2. Let p be a prime number and write $p\mathbf{Z}_{\mathrm{K}} = \boldsymbol{p}_1^{e_1} \cdots \boldsymbol{p}_g^{e_g}$ as above.

(1) The integer e_i is called the *ramification index* of $\boldsymbol{p}_i$, and the integer f_i is called the *residual degree* of $\boldsymbol{p}_i$.

(2) If $g = [\mathrm{K} : \mathbf{Q}]$, then p is said to be *totally split* in K, or in $\mathbf{Z}_{\mathrm{K}}$.

(3) If $g = 1$ and $e_1 = 1$, then p is said to be *inert* in K, or in $\mathbf{Z}_{\mathrm{K}}$.

(4) If some e_i is $\geqslant 2$, then p is said to be *ramified* in K, or in $\mathbf{Z}_{\mathrm{K}}$. Otherwise, p is said to be *unramified*.

We come back to the general discussion. From the definitions, note in particular that a ramified prime p is one which generates an ideal of $\mathbf{Z}_{\mathrm{K}}$ which is not squarefree: it has some prime (ideal) factor with exponent $\geqslant 2$. These ramified primes are very important, but also very constrained, as shown by the next proposition.

PROPOSITION 2.6.3. *Let* K *be a number field.*

(1) *A prime number p is ramified in* K *if and only if the quotient ring $\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$ contains some non-zero nilpotent element, i.e., some non-zero element x such that $x^k = 0$ for some $k \geqslant 1$.*

(2) *The set of primes which are ramified in $\mathbf{Z}_{\mathrm{K}}$ is finite, and coincides with the set of prime numbers dividing the discriminant of* K.

The usual terminology for a commutative ring A without non-zero nilpotent elements is that A is *reduced*. So the first part of this proposation states that $\boldsymbol{p}$ is unramified if and only if $\mathbf{Z}_{\mathrm{K}}/\boldsymbol{p}$ is reduced.

PROOF. According to the previous discussion, we will use the quotient ring $\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$.
(1) By the Chinese Remainder Theorem, there is an isomorphism

$$\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}} \to \prod_{i=1}^{g} \mathbf{Z}_{\mathrm{K}}/\boldsymbol{p}_i^{e_i}.$$

It is elementary that if A_1 and A_2 are commutative rings, then $\mathrm{A}_1 \times \mathrm{A}_2$ is reduced if and only if A_1 and A_2 are both reduced. This reduces the assertion to the fact that, given a prime ideal $\boldsymbol{p}$ and an integer $e \geqslant 1$, the ring $\mathbf{Z}_{\mathrm{K}}/\boldsymbol{p}^e$ has non-zero nilpotent elements if and only if $e \geqslant 2$. Indeed, if $e = 1$, this ring is a field, so it doesn't have non-zero nilpotent elements, whereas if $e \geqslant 2$ and $x \in \boldsymbol{p} - \boldsymbol{p}^e$, then the class of x is non-zero in $\mathbf{Z}_{\mathrm{K}}/\boldsymbol{p}^e$ and satisfies $x^e \equiv 0 \pmod{\boldsymbol{p}^e}$, so $\mathbf{Z}_{\mathrm{K}}/\boldsymbol{p}^e$ is not reduced.
(2) Recall that the discriminant was defined as the determinant of the Gram matrix associated to the bilinear form $(x, y) \mapsto \mathrm{Tr}(xy)$ on $\mathbf{Z}_{\mathrm{K}} \times \mathbf{Z}_{\mathrm{K}}$ (see Definition 2.4.10). The discriminant modulo a prime p will then be the corresponding determinant for the "same" bilinear form on $\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}} \times \mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$. Thus it will vanish (i.e., p divides the discriminant) if and only if this reduced bilinear form is degenerate. According to (1), we are claiming that the bilinear form is degenerate if and only if the quotient $\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$ is not reduced.
Indeed, first of all, if $\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$ contains a non-zero nilpotent element x, then for any $y \in \mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$, the group homomorphism $m_{xy} \colon z \mapsto xyz$ is nilpotent (some power m_{xy}^k is zero with $k \geqslant 1$), so its trace is zero. Conversely, if $p\mathbf{Z}_{\mathrm{K}}$ is squarefree, the fact that the trace bilinear form on finite fields (see Lemma A.1.2) is non-degenerate quickly implies that the trace form is non-degenerate for $\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$. □

REMARK 2.6.4. (1) Computing the discriminant of a number field can be quite involved. It is often easier to find an integer Δ such that disc(K) divides Δ, and one then knows at least that the ramified primes are among the divisors of Δ, which may be sufficient for certain purposes.
(2) The exponents of primes in the factorization of the discriminant are also quite important, and can be difficult to compute. For instance, one can show that the discriminant

of $\mathbf{Q}(e^{2i\pi/m})$ is

$$(-1)^{\varphi(m)}m^{\varphi(m)}\prod_{p\mid m}p^{-\varphi(m)/(p-1)}$$

(see, e.g., [20, Prop. 2.7]). Thus only $p \mid m$ are ramified, and the corresponding p-adic valuation of the discriminant is

$$\varphi(m)\Big(v_p(m)-\frac{1}{p-1}\Big)$$

(which is indeed an integer because $p-1 \mid \varphi(m)$ if $p \mid m$, and is $\geqslant 0$ because $v_p(m) \geqslant 1$).

Example 2.6.5. (1) We consider again a quadratic field $\mathrm{K} = \mathbf{Q}(\sqrt{d})$, where d is a squarefree integer. In this case, we have $[\mathrm{K} : \mathbf{Q}] = 2$, and the equation (2.5) for a given prime number p leaves only three possibilites:

(1) We can have $g = 2$, $f_1 = f_2 = 1$, $e_1 = e_2 = 1$: this corresponds to totally split primes.

(2) We can have $g = 1$, $f_1 = 2$ and $e_1 = 1$: these are the inert primes.

(3) We have have $g = 1$, $f_1 = 1$ and $e_1 = 2$: these are the ramified primes.

We can also express these conditions more concretely if $p \nmid d$, using the Legendre symbol. The key point is that since $\sqrt{d} \in \mathbf{Z}_{\mathrm{K}}$, the class of $\sqrt{d}$ modulo any prime ideal $\boldsymbol{p}$ is a square-root of d in the quotient field $\mathbf{Z}_{\mathrm{K}}/\boldsymbol{p}$.

If $\boldsymbol{p}$ is split, then the two prime ideals dividing $p\mathbf{Z}_{\mathrm{K}}$ have norm p, so the quotient field is isomorphic to $\mathbf{F}_p$, and this gives a root of $\sqrt{d}$ modulo p, which means that $\left(\frac{d}{p}\right) = 1$ if $p \nmid d$.

If $\boldsymbol{p}$ is inert, then this means that the square root of d belongs to a quadratic extension of $\mathbf{F}_p$, so that $\left(\frac{d}{p}\right) = -1$. This statement is in fact a characterization of inert primes (in all other cases, there is a root of d modulo p).

Recall that the discriminant of $\mathbf{Z}_{\mathrm{K}}$ is either d or $4d$, the first case corresponding to $d \equiv 1 \pmod 4$. But instead of using this with Proposition 2.6.3 to determine the ramified primes, we can recover by hand the results, and this is quite enlightening.

First, any inert prime is unramified, as we already observed.

Second, if p does not divide d and $\left(\frac{d}{p}\right) = 1$, then we claim that p is unramified and totally split. Indeed, let a be an integer with $a^2 \equiv d \pmod p$. Then note that we have an isomorphism

$$\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}} \simeq \mathbf{Z}[\mathrm{X}]/(p, \mathrm{X}^2 - d) = \mathbf{F}_p[\mathrm{X}]/(\mathrm{X} - a)(\mathrm{X} + a),$$

and since a and $-a$ are different modulo p (here we use the assumption that $p \nmid d$), the polynomials $\mathrm{X} - a$ and $\mathrm{X} + a$ are coprime in $\mathbf{F}_p[\mathrm{X}]$. The Chinese Remainder Theorem shows that $\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$ is then isomorphic to a product of two fields, hence is reduced, which is equivalent to our claim by the elementary first part of Proposition 2.6.3.

Third, if p is *odd* and $p \mid d$, then we can directly compute $\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$: if d is not 1 modulo 4, then we have isomorphisms

$$\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}} \simeq \mathbf{Z}[\mathrm{X}]/(p, \mathrm{X}^2 - d) = \mathbf{F}_p[\mathrm{X}]/(\mathrm{X}^2 - d) = \mathbf{F}_p[\mathrm{X}]/(\mathrm{X}^2),$$

while, similarly, if $d \equiv 1 \pmod 4$, then we get

$$\begin{aligned}\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}} &\simeq \mathbf{Z}[\mathrm{X}]/(p, \mathrm{X}^2 - \mathrm{X} + (1-d)/4)\\ &= \mathbf{F}_p[\mathrm{X}]/(\mathrm{X}^2 - \mathrm{X} + (1-d)/4) = \mathbf{F}_p[\mathrm{X}]/((\mathrm{X} - 1/2)^2),\end{aligned}$$

(using the fact that 2 is invertible modulo the odd prime p). Either of these isomorphism implies that p is ramified in K.

Finally, we determine when 2 is ramified. The first isomorphism still applies to show that $p = 2$ is ramified if d is not $\equiv 1 \pmod 4$. So the only remaining question is whether the prime $p = 2$ is unramified when $d \equiv 1 \pmod 4$ (it may then be either inert or split). We start again with the isomorphism

$$\mathbf{Z}_\mathrm{K}/2\mathbf{Z}_\mathrm{K} \simeq \mathbf{Z}[\mathrm{X}]/(2, \mathrm{X}^2 - \mathrm{X} + (1-d)/4) = \mathbf{F}_2[\mathrm{X}]/(\mathrm{X}^2 + \mathrm{X} + (1-d)/4)$$

(which makes sense since $(1-d)/4$ is an integer). Two cases arise: if $d \equiv 1 \pmod 8$, then we get

$$\mathbf{Z}_\mathrm{K}/2\mathbf{Z}_\mathrm{K} \simeq \mathbf{F}_2[\mathrm{X}]/(\mathrm{X}(\mathrm{X}+1)),$$

which (by the Chinese Remainder Theorem again) is isomorphic to $\mathbf{F}_2 \times \mathbf{F}_2$, so we see as before that 2 is unramified and split in that case. On the other hand, if $d \equiv 5 \pmod 8$, then

$$\mathbf{Z}_\mathrm{K}/2\mathbf{Z}_\mathrm{K} \simeq \mathbf{F}_2[\mathrm{X}]/(\mathrm{X}^2 + \mathrm{X} + 1),$$

and the key point is that $\mathrm{X}^2 + \mathrm{X} + 1$ is irreducible in $\mathbf{F}_2[\mathrm{X}]$ (this is easily checked, since it means that there is no root in $\mathbf{F}_2$).[3] This means that 2 is unramified and inert in K.

(2) Let's look at some concrete examples for $p = 2$ and $d \equiv 1 \pmod 4$, in which case we have seen that 2 is unramified in $\mathbf{Q}(\sqrt{d})$.

For $d = 5$, one can check that $\mathbf{Z}_{\mathbf{Q}(\sqrt{5})}$ is principal. Then we have the factorization

$$2 = -\frac{1+\sqrt{5}}{2} \cdot (1 - \sqrt{5}),$$

where $(1+\sqrt{5})/2$ is a unit, and $1 - \sqrt{5}$ is irreducible. This means that 2 is inert in the field $\mathbf{Q}(\sqrt{5})$.

For $d = 17$, one can also check that $\mathbf{Q}(\sqrt{17})$ is principal. Here, we get by playing around the factorization

$$2 = \frac{5+\sqrt{17}}{2} \cdot \frac{5-\sqrt{17}}{2},$$

and the elements $(5+\sqrt{17})/2$ and $(5-\sqrt{17})/2$ are both irreducible; their ratio is not a unit, so

$$2\mathbf{Z}\Big[\frac{1+\sqrt{17}}{2}\Big] = \Big(\frac{5+\sqrt{17}}{2}\Big)\mathbf{Z}\Big[\frac{1+\sqrt{17}}{2}\Big] \cdot \Big(\frac{5-\sqrt{17}}{2}\Big)\mathbf{Z}\Big[\frac{1+\sqrt{17}}{2}\Big]$$

gives the factorization which shows that the prime 2 is totally split.

(3) We look at a few "random" examples: the data below gives a few polynomials f, the discriminant Δ of the ring of integers of the field $\mathbf{Q}(\alpha)$, where α is a root of f, and

[3] In fact, $\mathrm{X}^2 + \mathrm{X} + 1$ is the only irreducible polynomial of degree 2 in $\mathbf{F}_2[\mathrm{X}]$.

the factorization of the discriminant:

$$f = 64x^6 - 480x^4 + 720x^2 - 120,$$
$$\Delta = 503884800000 = 2^{13} \cdot 3^9 \cdot 5^5,$$
$$f = x^7 + 2x^6 + 3x^5 + 2x^4 - 5x^3 + 2x^2 + 3x + 2,$$
$$\Delta = -60035466240 = -2^{10} \cdot 3^2 \cdot 5 \cdot 199 \cdot 6547,$$
$$f = 6435x^8 - 12012x^6 + 6930x^4 - 1260x^2 + 35,$$
$$\Delta = 2^{24} \cdot 3^{10} \cdot 5^6 \cdot 7^7 \cdot 11^5 \cdot 13^3,$$
$$f = x^9 + x^8 + 10x^7 + 13x^6 + 2x^5 + 14x^4 + 16x^3 + 10x^2 + 4x + 5,$$
$$\Delta = 213287391425295766669 = 683 \cdot 312280221706143143.$$

These give an indication of the complexity involved in the discriminant.

Before starting the next section, we address a practical question which arises from the definitions and results concerning ideals in number field, and factorization of primes in particular: how does one determine the factorization (or *splitting type*) of a given prime number? Although the multiplication of ideals may seem at first a bit complicated, the key point is that the information is also present in the *quotient ring* $\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$, and computing this ring will be the main approach to understanding the factorizations of ideals. Indeed, note the following:

(1) The prime ideals $\boldsymbol{p}_i$ in (2.4) are, by construction, the prime ideals containing $p\mathbf{Z}_{\mathrm{K}}$, and by elementary algebra, they are in *explicit* bijection with the prime ideals of the (finite) ring $\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$. Thus, given a prime ideal I in $\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$, we obtain one of the primes $\boldsymbol{p}_i$ as the set of $x \in \mathbf{Z}_{\mathrm{K}}$ whose class modulo $p\mathbf{Z}_{\mathrm{K}}$ is in I.

(2) Assuming we know one of the prime ideals $\boldsymbol{p}_i$, and its image $\mathrm{I} \subset \mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$, the exponent e_i is determined as the largest positive integer e such that $\boldsymbol{p}^e$ contains $p\mathbf{Z}_{\mathrm{K}}$ and is the number of distinct ideals in the sequence

$$\mathrm{I} \supset \mathrm{I}^2 \supset \cdots \supset \mathrm{I}^k \supset \cdots.$$

Example 2.6.6. These facts are not specific to the factorization of ideals of the form $p\mathbf{Z}_{\mathrm{K}}$. They are indeed perfectly visible already in the case of $\mathbf{Z}$: if $n \geqslant 1$ factors as

$$n = p_1^{e_1} \cdots p_k^{e_k},$$

with distinct primes and exponents $e_i \geqslant 1$, we get by the Chinese Remainder Theorem the isomorphism

$$\mathbf{Z}/n\mathbf{Z} \to \prod_{i=1}^{k} \mathbf{Z}/p_k^{e_k}\mathbf{Z}.$$

The prime ideals in the finite product ring are

$$\mathrm{I}_j = \{(x_i)_{1 \leqslant i \leqslant k} \mid x_j \in p\mathbf{Z}/p_j^{e_j}\mathbf{Z}\}$$

for $1 \leqslant j \leqslant k$. Taking I_1 as example, we get for $e \geqslant 1$ the equality

$$\mathrm{I}_1^e = p_1^e\mathbf{Z}/p_1^{e_1}\mathbf{Z} \times \mathbf{Z}/p_2^{e_2}\mathbf{Z} \times \cdots$$

and the distinct powers that appear are I_1, I_1^2, …, I_1^e.

We will illustrate this principle with a theorem which shows how relate the factorization of $p\mathbf{Z}_{\mathrm{K}}$ with that of polynomial over finite fields. Before giving the statement, we record a very useful lemma which allows to avoid the difference between $\mathbf{Z}_{\mathrm{K}}$ and the simpler rings $\mathbf{Z}[\alpha]$ in some cases.

LEMMA 2.6.7. *Let $f \in \mathbf{Z}[\mathrm{X}]$ be an irreducible monic polynomial, let $\alpha \in \mathbf{C}$ be a root of f and $\mathrm{K} = \mathbf{Q}(\alpha)$ the number field it generates. The set*

$$\boldsymbol{c} = \{x \in \mathbf{Z}_{\mathrm{K}} \mid x\mathbf{Z}_{\mathrm{K}} \subset \mathbf{Z}[\alpha]\},$$

is a non-zero ideal of $\mathbf{Z}_{\mathrm{K}}$ contained in $\mathbf{Z}[\alpha]$.

For any prime number p such that $p\mathbf{Z}_{\mathrm{K}}$ is coprime to $\boldsymbol{c}$, the inclusion $\mathbf{Z}[\alpha] \to \mathbf{Z}_{\mathrm{K}}$ induces by passing to the quotients an isomorphism

$$\mathbf{Z}[\alpha]/p\mathbf{Z}[\alpha] \to \mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}.$$

PROOF. It is straightforward that $\boldsymbol{c}$ is an ideal in $\mathbf{Z}_{\mathrm{K}}$; it is contained in $\mathbf{Z}[\alpha]$ (since any $x \in \boldsymbol{c}$ satisfies $x \cdot 1 = x \in \mathbf{Z}[\alpha]$ by definition), and it is non-zero because $\mathbf{Z}_{\mathrm{K}}$ is a finitely-generated abelian group (this is the existence of a non-zero "common denominator" for all the integers of K).

Since $p\mathbf{Z}[\alpha] \subset p\mathbf{Z}_{\mathrm{K}}$, the composite $\mathbf{Z}[\alpha] \to \mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$ always gives the induced quotient morphism $j \colon \mathbf{Z}[\alpha]/p\mathbf{Z}[\alpha] \to \mathbf{Z}_{\mathrm{L}}/p\mathbf{Z}_{\mathrm{K}}$.

We now assume that $\boldsymbol{c}$ is coprime to $p\mathbf{Z}_{\mathrm{K}}$. Then $p\mathbf{Z}_{\mathrm{K}} + \boldsymbol{c} = \mathbf{Z}_{\mathrm{K}}$, so any class modulo $p\mathbf{Z}_{\mathrm{K}}$ has a representative in $\boldsymbol{c}$, and hence in $\mathbf{Z}[\alpha]$, and this shows that the morphism j is surjective.

The kernel of j is by definition $(\mathbf{Z}[\alpha] \cap \mathbf{Z}_{\mathrm{K}})/p\mathbf{Z}[\alpha]$. We claim that $\mathbf{Z}[\alpha] \cap \mathbf{Z}_{\mathrm{K}} = p\mathbf{Z}[\alpha]$, which will prove that j is injective. Indeed, note that $\boldsymbol{c} \cap \mathbf{Z}$ is a non-zero ideal of $\mathbf{Z}$ which is coprime to $p\mathbf{Z}$. So we can find $a \in \mathbf{Z}$ and $b \in \boldsymbol{c} \cap \mathbf{Z}$ such that $a + b = 1$. Now let $x \in \mathbf{Z}[\alpha \cap p\mathbf{Z}_{\mathrm{K}}$; we get $x = pax + bx$. But since $x \in \mathbf{Z}[\alpha]$, we get $pax \in p\mathbf{Z}[\alpha]$, and since $b \in \boldsymbol{c}$, we have $bx \in p\mathbf{Z}_{\mathrm{K}} \cdot \boldsymbol{c} \subset p\mathbf{Z}[\alpha]$ by definition of $\boldsymbol{c}$. □

And now for the theorem.

THEOREM 2.6.8. *Let $f \in \mathbf{Z}[\mathrm{X}]$ be an irreducible monic polynomial, let $\alpha \in \mathbf{C}$ be a root of f and $\mathrm{K} = \mathbf{Q}(\alpha)$ the number field it generates.*

Let

$$\boldsymbol{c} = \{x \in \mathbf{Z}_{\mathrm{K}} \mid x\mathbf{Z}_{\mathrm{K}} \subset \mathbf{Z}[\alpha]\},$$

and let p be a prime number coprime to $\boldsymbol{c}$. Write

$$f \,(\mathrm{mod}\, p) = h_1^{e_1} \cdots h_g^{e_g}$$

where $g \geqslant 1$, the h_i's are distinct irreducible monic polynomials and $e_i \geqslant 1$ are positive integers. Denote $f_i = \deg(h_i)$.

Then the ideal $p\mathbf{Z}_{\mathrm{K}}$ factors as a product of g prime ideals $\boldsymbol{p}_i$ in $\mathbf{Z}_{\mathrm{K}}$ with ramification indices e_i and residual degrees f_i. In fact, if $\widetilde{h}_i \in \mathbf{Z}[\mathrm{X}]$ is a monic polynomial with $\widetilde{h}_i \equiv h_i \,(\mathrm{mod}\, p)$, then the prime ideal $\boldsymbol{p}_i$ is generated by p and $\widetilde{h}_i(\alpha)$.

PROOF. We determine the factorization of $p\mathbf{Z}_{\mathrm{K}}$ by computing the ring $\mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$, as explained above. By Lemma 2.6.7, the assumption tells us that the inclusion of $\mathbf{Z}[\alpha]$ in $\mathbf{Z}_{\mathrm{K}}$ gives an isomorphism $\mathbf{Z}[\alpha]/p\mathbf{Z}[\alpha] \to \mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}$. Since evaluation at α gives an isomorphism $\mathbf{Z}[\mathrm{X}]/f\mathbf{Z}[\mathrm{X}] \to \mathbf{Z}[\alpha]$, we get an (explicit) isomorphism

$$\mathbf{F}_p[\mathrm{X}]/f\mathbf{F}_p[\mathrm{X}] \to \mathbf{Z}[\mathrm{X}]/(p, f\mathbf{Z}[\mathrm{X}]) \to \mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}.$$

The Chinese Remainder Theorem in $\mathbf{F}_p[\mathrm{X}]$ provides us finally with an isomorphism

$$\prod_{i=1}^{g} \mathbf{F}_p[\mathrm{X}]/h_i^{e_i}\mathbf{F}_p[\mathrm{X}] \to \mathbf{F}_p[\mathrm{X}]/f\mathbf{F}_p[\mathrm{X}] \to \mathbf{Z}_{\mathrm{K}}/p\mathbf{Z}_{\mathrm{K}}.$$

The product ring

$$\mathrm{A} = \prod_{i=1}^{g} \mathbf{F}_p[\mathrm{X}]/h_i^{e_i}\mathbf{F}_p[\mathrm{X}]$$

contains g distinct prime ideals, namely the ideals I_j for $1 \leqslant j \leqslant g$ where the only condition is that the j-th coordinate in this decomposition lies in $h_j\mathbf{F}_p[\mathrm{X}]/h_j^{e_j}\mathbf{F}_p[\mathrm{X}]$. Thus $p\mathbf{Z}_\mathrm{K}$ is divisible by g distinct prime ideals $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_g$. Moreover, we see that the sequence

$$\mathrm{I}_j \supset \mathrm{I}_j^2 \supset \cdots \supset \mathrm{I}_j^k \supset \cdots$$

stabilizes at $e = e_j$, so the ramification index for $\boldsymbol{p}_j$ is equal to e_j. By the compatibility of the previous isomorphisms with computing residue fields, the residue field at $\boldsymbol{p}_j$ is isomorphic to

$$\mathrm{A}/\mathrm{I}_j = \Big(\prod_{i=1}^{g} \mathbf{F}_p[\mathrm{X}]/h_i^{e_i}\mathbf{F}_p[\mathrm{X}]\Big)/\mathrm{I}_j = \mathbf{F}_p[\mathrm{X}]/h_j\mathbf{F}_p[\mathrm{X}],$$

which is a finite extension of degree $\deg(h_j) = f_j$ of $\mathbf{F}_p[\mathrm{X}]$, so the residual degree is f_j.

Finally, to conclude the proof, we track back explicitly the isomorphism from I_j to $\boldsymbol{p}_j$, and see that $\boldsymbol{p}_j$ is generated by p and an element $\widetilde{h}_j(\alpha)$ for some monic polynomial $\widetilde{h}_j \in \mathbf{Z}[\mathrm{X}]$ which reduces to h_j modulo p. □

The next corollary shows that the study of splitting of primes is a generalization of the questions surrounding Kronecker's Theorem in Section 1.4.

Corollary 2.6.9. *Let $f \in \mathbf{Z}[\mathrm{X}]$ be an irreducible monic polynomial, let $\alpha \in \mathbf{C}$ be a root of f and $\mathrm{K} = \mathbf{Q}(\alpha)$ the number field it generates. For all but finitely many prime numbers p, the number $\nu_f(p)$ of roots of f in $\mathbf{F}_p$ is equal to the number of non-zero prime ideals dividing $p\mathbf{Z}_\mathrm{K}$ which have residual degree equal to* 1.

Proof. Indeed, with the notation of the theorem, $\nu_f(p)$ is equal to the number of irreducible factors h_i of $f \pmod p$ which have degree 1, and this is (according to the theorem) the same as the number of prime divisors of $p\mathbf{Z}_\mathrm{K}$ of residuel degree 1, at least when p is coprime to $\boldsymbol{c}$. □

Example 2.6.10. (1) Let (somewhat randomly)

$$f = \mathrm{X}^5 - 12\mathrm{X}^4 + \mathrm{X}^3 - 163,$$

and consider $\mathrm{K} = \mathbf{Q}(\alpha)$ where $f(\alpha) = 0$. It turns out here that $\mathbf{Z}_\mathrm{K} = \mathbf{Z}[\alpha]$ and the discriminant is $11^2 \cdot 157 \cdot 163 \cdot 527453$. Factoring the defining polynomial modulo 2, 3, 5, 17, 1009 and 2689, we get

$$\begin{aligned}
f \pmod 2 &= \mathrm{X}^5 + \mathrm{X}^3 + 1 \quad \text{(which is irreducible)} \\
f \pmod 3 &= (\mathrm{X}+1)(\mathrm{X}^4 + 2\mathrm{X}^3 + 2\mathrm{X}^2 + \mathrm{X} + 2) \\
f \pmod 5 &= (\mathrm{X}+1)(\mathrm{X}^4 + 2\mathrm{X}^3 + 2\mathrm{X}^2 + \mathrm{X} + 2) \\
f \pmod{17} &= (\mathrm{X}^2 - 3\mathrm{X} - 7)(\mathrm{X}^3 + 8\mathrm{X}^2 - 2\mathrm{X} - 1) \\
f \pmod{1009} &= \mathrm{X}^5 - 12\mathrm{X}^4 + \mathrm{X}^3 - 163 \quad \text{(which is irreducible)} \\
f \pmod{2689} &= (\mathrm{X}+194)(\mathrm{X}+504)(\mathrm{X}+1024)(\mathrm{X}+1514)(\mathrm{X}+2130).
\end{aligned}$$

Hence we can read off the prime factorization of the corresponding principal ideals $p\mathbf{Z}_\mathrm{K}$: 2689 is totally split, 2 and 1009 are inert, 3 and 5 are both the product of a prime with residue degree 1 and another with residue degree 3, and 17 is the product of a prime

with residue degree 2 and another with residue degree 3. Even better, one can give the generators of the dividing prime ideals: for instance,

$$17\mathbf{Z}_{\mathrm{K}} = (17\mathbf{Z}_{\mathrm{K}} + (\alpha^2 - 3\alpha - 7)\mathbf{Z}_{\mathrm{K}}) \cdot (17\mathbf{Z}_{\mathrm{K}} + (\alpha^3 + 8\alpha^2 - 2\alpha - 1)\mathbf{Z}_{\mathrm{K}}).$$

(One also checks that, maybe a bit surprisingly, the class group is trivial here.)

(2) Let $\mathrm{K} = \mathbf{Q}(\sqrt{-14})$. We then have $\mathbf{Z}_{\mathrm{K}} = \mathbf{Z}[\sqrt{-14}]$, so the ideal $\boldsymbol{c}$ is $\mathbf{Z}_{\mathrm{K}}$. Modulo 3, we find that

$$\mathrm{X}^2 + 14 \equiv \mathrm{X}^2 - 1 = (\mathrm{X} - 1)(\mathrm{X} + 1),$$

hence it follows from the theorem, or from Example 2.6.5, that $3\mathbf{Z}_{\mathrm{K}} = \boldsymbol{p}_1\boldsymbol{p}_2$ for two distinct prime ideals $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$. These have norm 3. On the other hand, one can prove that $|(a + b\sqrt{-14})\mathbf{Z}_{\mathrm{K}}| = a^2 + 14b^2$. Since the equation $a^2 + 14b^2 = 3$ has no integral solution, we conclude that $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ cannot be principal ideals. In fact, one can check that $\mathscr{H}(\mathbf{Q}(\sqrt{-14}))$ is cyclic of order 4, generated by $\boldsymbol{p}_1$ (or $\boldsymbol{p}_2$).

(3) The next example shows that the condition that p is coprime with $\boldsymbol{c}$ is necessary. Let $f = \mathrm{X}^2 - 5$ and $\alpha = \sqrt{5}$, so that $\mathrm{K} = \mathbf{Q}(\sqrt{5})$. Then $\mathbf{Z}_{\mathrm{K}} = \mathbf{Z}[(1 + \sqrt{5})/2]$, and we see that $2\mathbf{Z}_{\mathrm{K}} \subset \boldsymbol{c}$. We also have $\boldsymbol{c} \neq \mathbf{Z}_{\mathrm{K}}$, since $\mathbf{Z}_{\mathrm{K}}$ is not contained in $\mathbf{Z}[\sqrt{5}]$.

This implies that $p = 2$ is not coprime with $\boldsymbol{c}$. We have seen that 2 is unramified and inert in K (Example 2.6.5), but we have $f \pmod 2 = \mathrm{X}^2 + 1 = (\mathrm{X} + 1)^2$, which (in the setting of Theorem 2.6.8) would have suggested that 2 is ramified.

2.7. Galois action and Frobenius automorphism

We have not really used Galois theory up to now, and it is time to do so. The simple observation which allows us to use it for algebraic integers is that if K is a number field and $\sigma\colon \mathrm{K} \to \mathrm{K}'$ a field morphism (recall that such a morphism is always injective), then the image by σ of some integral element $x \in \mathbf{Z}_{\mathrm{K}}$ belongs to $\mathbf{Z}_{\mathrm{K}'}$, simply because from an equation

$$x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0 = 0$$

with $a_i \in \mathbf{Z}$, we deduce that

$$\sigma(x)^d + a_{d-1}\sigma(x)^{d-1} + \cdots + a_1\sigma(x) + a_0 = 0$$

(since σ is always the identity on $\mathbf{Q}$). Hence we have an induced morphism of rings from $\mathbf{Z}_{\mathrm{K}} \to \mathbf{Z}_{\mathrm{K}'}$. In particular, if $\boldsymbol{n} \subset \mathbf{Z}_{\mathrm{K}}$ is an ideal, then $\sigma(\boldsymbol{n})$ is an ideal in $\mathbf{Z}_{\mathrm{K}'}$. If $\sigma\colon \mathrm{K} \to \mathrm{K}$ is a field automorphism of K, we deduce that σ induces by restriction a ring automorphism $\mathbf{Z}_{\mathrm{K}} \to \mathbf{Z}_{\mathrm{K}}$.

We now look at the interaction of the automorphisms with prime factorizations, in the case of a Galois extension.

Proposition 2.7.1. *Let* K *be a number field which is a Galois extension of* $\mathbf{Q}$*, and denote by* G *the Galois group of* K *over* $\mathbf{Q}$*. Let* p *be a prime number and* S_p *the set of non-zero prime ideals* $\boldsymbol{p} \subset \mathbf{Z}_{\mathrm{K}}$ *which divide* $p\mathbf{Z}_{\mathrm{K}}$*.*

(1) *The group* G *acts on* S_p *by* $\sigma \cdot \boldsymbol{p} = \sigma(\boldsymbol{p})$*, i.e., by taking the set-theoretic image.*

(2) *This action is* transitive*, i.e., for any* $\boldsymbol{p}$ *and* $\boldsymbol{q}$ *in* S_p*, there exists* $\sigma \in \mathrm{G}$ *such that* $\boldsymbol{q} = \sigma \cdot \boldsymbol{p}$*.*

Proof. The first statement is straightforward from the definitions and the previous discussion, since σ is an automorphism. But the second is a fundamental and non-trivial fact.

The proof is very ingenious. We fix $\boldsymbol{q} \in \mathrm{S}_p$, and wish to prove that all other prime ideals dividing $p\mathbf{Z}_{\mathrm{K}}$ are obtained by taking the image of $\boldsymbol{q}$ by some Galois automorphism.

Let $\boldsymbol{p} \in \mathrm{S}_p$ be one such ideal. Let $x \in \boldsymbol{p}$ be any element of $\boldsymbol{p}$. We form the element

$$y = \prod_{\sigma \in \mathrm{G}} \sigma(x).$$

We then observe that y belongs to $\boldsymbol{p}$ (since the identity is an element of G, so that x appear in the product), but also that y is invariant under the action of G on K (i.e., we have $\tau(y) = y$ for all $\tau \in \mathrm{G}$). This means first that $y \in \mathbf{Q} \cap \mathbf{Z}_\mathrm{K} = \mathbf{Z}$. But $\mathbf{Z} = \mathbf{Z}_\mathrm{K} \cap \boldsymbol{q}$ (by Lemma 2.6.1), so

$$y = \prod_{\sigma \in \mathrm{G}} \sigma(x) \in \boldsymbol{q}.$$

Since $\boldsymbol{q}$ is a prime ideal, some term in the product belongs to $\boldsymbol{q}$, which means that there exists some $\sigma \in \mathrm{G}$ (depending a priori on x) such that $x \in \sigma^{-1} \cdot \boldsymbol{q}$. We have therefore shown that

$$\boldsymbol{p} \subset \bigcup_{\sigma \in \mathrm{G}} \sigma \cdot \boldsymbol{q}.$$

Since the union of ideals is very rarely an ideal, it should not be too surprising that this implies that $\boldsymbol{p} \subset \sigma \cdot \boldsymbol{q}$ for some single $\sigma \in \mathrm{G}$, and this is explained in Lemma A.1.1. Since $\boldsymbol{p}$ and $\sigma \cdot \boldsymbol{q}$ are both maximal ideals, we then have in fact $\boldsymbol{p} = \sigma \cdot \boldsymbol{q}$. This finishes the proof. □

Intuitively, the transitivity of the action means that all the ideals dividing $p\mathbf{Z}_\mathrm{K}$ "look the same". In particular, the general factorization pattern is drastically simplified.

Corollary 2.7.2. *Let* K *be a number field which is a Galois extension of* $\mathbf{Q}$. *Let* p *be a prime number. In the factorization*

$$p\mathbf{Z}_\mathrm{K} = \boldsymbol{p}_1^{e_1} \cdots \boldsymbol{p}_g^{e_g}$$

of $p\mathbf{Z}_\mathrm{K}$, *the ramification indices* e_i *and the residual degrees* $f_i = [\mathbf{Z}_\mathrm{K}/\boldsymbol{p}_i : \mathbf{F}_p]$ *are independent of* i. *If we denote them by* e *and* f, *respectively, then we have*

$$efg = [\mathrm{K} : \mathbf{Q}].$$

Any transitive action of a group on a non-empty set is isomorphic to that of the group on the cosets of a subgroup, which is the stabilizer of some fixed element of the set. Applying this principle to a finite Galois extension K of $\mathbf{Q}$, a prime number p and the action of the Galois group of K on the prime ideals dividing $p\mathbf{Z}_\mathrm{K}$, we fix some such prime ideal $\boldsymbol{p}$, and define the *decomposition group* at $\boldsymbol{p}$ to be

$$\mathrm{D}_{\boldsymbol{p}} = \{\sigma \in \mathrm{G} \mid \sigma \cdot \boldsymbol{p} = \boldsymbol{p}\}.$$

We then have a bijection from the cosets $\mathrm{G}/\mathrm{D}_{\boldsymbol{p}}$ to the set of prime ideals dividing $p\mathbf{Z}_\mathrm{K}$ by mapping a coset $\sigma\mathrm{D}_{\boldsymbol{p}}$ to $\sigma \cdot \boldsymbol{p}$ (which is well-defined by construction of $\mathrm{D}_{\boldsymbol{p}}$).

Because of the specific nature of the action involved, we can go further: for any $\sigma \in \mathrm{D}_{\boldsymbol{p}}$, the fact that $\sigma \cdot \boldsymbol{p} = \boldsymbol{p}$ means that σ defines, by passing to the quotient, a field morphism

$$\widetilde{\sigma} \colon \mathbf{Z}_\mathrm{K}/\boldsymbol{p} \to \mathbf{Z}_\mathrm{K}/\boldsymbol{p}$$

which must therefore be a field automorphism of the finite extension $\mathbf{F}_{\boldsymbol{p}} = \mathbf{Z}_\mathrm{K}/\boldsymbol{p}$ of $\mathbf{F}_p$. This procedure is compatible with restriction, hence we have in fact a group homomorphism from $\mathrm{D}_{\boldsymbol{p}}$ to the Galois group of the extension $\mathbf{F}_{\boldsymbol{p}}$ of $\mathbf{F}_p$ (recall that any finite extension of finite fields is a Galois extension). The following is the final[4] fundamental basic fact about algebraic number theory.

[4] Except, admittedly, for Dirichlet's Unit Theorem, which will be discussed later.

THEOREM 2.7.3 (Dedekind). *Let* K *be a number field which is a Galois extension of* $\mathbf{Q}$*, and let* G *be its Galois group. Let* p *be a prime number and let*

$$\varphi_{\boldsymbol{p}} \colon \mathrm{G} \to \mathrm{Gal}(\mathbf{F}_{\boldsymbol{p}}/\mathbf{F}_p)$$

be the group homomorphism defined above.

(1) *The morphism* $\varphi_{\boldsymbol{p}}$ *is surjective.*

(2) *The morphism* $\varphi_{\boldsymbol{p}}$ *is injective if and only if* p *is unramified in* K.

Recall that for any finite extension F/E of finite fields, the Galois group of F over E is cyclic, and is generated by the Frobenius automorphism $x \mapsto x^{|\mathrm{E}|}$. Dedekind's Theorem therefore allows to state the following definition, which remarkably only goes back to Artin in the 1920's.

DEFINITION 2.7.4 (Frobenius automorphism). Let K be a number field which is a Galois extension of $\mathbf{Q}$, and let G be its Galois group. Let p be a prime number and let $\boldsymbol{p}$ be a prime ideal dividing $p\mathbf{Z}_{\mathrm{K}}$.

(1) The kernel of the surjective morphism $\varphi_{\boldsymbol{p}} \colon \mathrm{G} \to \mathrm{Gal}(\mathbf{F}_{\boldsymbol{p}}/\mathbf{F}_p)$ is called the *inertia group* at $\boldsymbol{p}$ and is denoted $\mathrm{I}_{\boldsymbol{p}}$.

(2) The *Frobenius automorphism* at $\boldsymbol{p}$ is the unique element σ of $\mathrm{D}_{\boldsymbol{p}}/\mathrm{I}_{\boldsymbol{p}}$ such that $\varphi_{\boldsymbol{p}}(\sigma)$ is the Frobenius automorphism $x \mapsto x^p$ of $\mathbf{F}_{\boldsymbol{p}}$. It is denoted $\mathrm{Fr}_{\boldsymbol{p}}$ if no confusion is likely.

REMARK 2.7.5. (1) Concretely, if p is unramified (which holds for all but finitely many primes), then $\mathrm{Fr}_{\boldsymbol{p}}$ is the unique element of G such that the congruence

$$\sigma(x) \equiv x^p \,(\mathrm{mod}\, \boldsymbol{p})$$

holds for all $x \in \mathbf{Z}_{\mathrm{K}}$. Indeed, this relation first implies that σ belongs to the decomposition group (since for $x \in \boldsymbol{p}$ it gives $\sigma(x) \equiv 0 \,(\mathrm{mod}\, \boldsymbol{p})$, so $\sigma(x) \in \boldsymbol{p}$), and then by definition means that $\varphi_{\boldsymbol{p}}(\sigma)$ is the Frobenius automorphism of $\mathbf{F}_{\boldsymbol{p}}$.

(2) If $\sigma \in \mathrm{Gal}(\mathrm{K}/\mathbf{Q})$ is any element such that $\sigma \in \mathrm{D}_{\boldsymbol{p}}$ and $\varphi_{\boldsymbol{p}}(\sigma)$ is the Frobenius automorphism of $\mathbf{F}_{\boldsymbol{p}}$, then one says that σ is "a" Frobenius automorphism at $\boldsymbol{p}$. Similarly, if $\sigma \in \mathrm{Gal}(\mathrm{K}/\mathbf{Q})$ is a Frobenius automorphism at *some* prime ideal $\boldsymbol{p} \mid p\mathbf{Z}_{\mathrm{K}}$, then one says that it is a Frobenius automorphism at p (the original prime number). We will see soon that all such elements are conjugate in the Galois group when p is unramified.

SKETCH OF PROOF OF THEOREM 2.7.3. We will explain the proof in the case where we have $\mathbf{Z}_{\mathrm{K}} = \mathbf{Z}[\alpha]$ for some element $\alpha \in \mathbf{Z}_{\mathrm{K}}$ (and hence also $\mathrm{K} = \mathbf{Q}(\alpha)$). The general case can be deduced from this using a general technique called "localization", which roughly speaking exploits the fact that the problem only concerns a single prime.

Let $f \in \mathbf{Z}[\mathrm{X}]$ be the minimal monic irreducible polynomial of α and $d = \deg(f) = [\mathrm{K} : \mathbf{Q}]$. There are two key points: (1) the set $\mathrm{Z}_f \subset \mathbf{Z}_{\mathrm{K}}$ of roots of f maps surjectively to the set $\mathrm{Z}_f(\mathbf{F}_{\boldsymbol{p}})$ of roots of f in $\mathbf{Z}_{\mathrm{K}}/\boldsymbol{p}$; (2) the Galois group $\mathrm{Gal}(\mathrm{K}/\mathbf{Q})$ acts transitively on the roots of f in $\mathbf{Z}_{\mathrm{K}}$ (by elementary Galois theory, this is equivalent to the fact that f is irreducible over $\mathbf{Q}$). Thus, denoting $\widetilde{\alpha} = \alpha \,(\mathrm{mod}\, \boldsymbol{p}) \in \mathbf{F}_{\boldsymbol{p}}$, there exists first a root $\beta \in \mathrm{Z}_f$ of f such that $\beta \equiv \alpha^p \,(\mathrm{mod}\, p)$ because $\widetilde{\alpha}^p \in \mathrm{Z}_f(\mathbf{F}_{\boldsymbol{p}})$, and second there exists an element $\sigma \in \mathrm{Gal}(\mathrm{K}/\mathbf{Q})$ such that $\sigma(\alpha) = \beta$. We claim that this σ belongs to the decomposition group $\mathrm{D}_{\boldsymbol{p}}$ and "is" a Frobenius automorphism at $\boldsymbol{p}$. This will prove the first part of the theorem.

Indeed, we simply observe that for all $a_i \in \mathbf{Z}$ for $0 \leqslant i < d$, we have

$$\sigma\Big(\sum_{i=0}^{d-1} a_i \alpha^i\Big) = \sum_{i=0} a_i \sigma(\alpha)^i \equiv \sum_{i=0} a_i \alpha^{pi} \,(\mathrm{mod}\, \boldsymbol{p}),$$

which implies that

$$\sigma\Big(\sum_{i=0}^{d-1} a_i \alpha^i\Big) \equiv \Big(\sum_{i=0} a_i \alpha^i\Big)^p \,(\mathrm{mod}\, \boldsymbol{p}),$$

since $\mathbf{F}_{\boldsymbol{p}}$ has characteristic p. By the preceeding remark, this proves the claim.

We now consider the second statement. Here, the point is that for a given $\sigma \in \mathrm{D}_{\boldsymbol{p}}$, the element $\varphi_{\boldsymbol{p}}$ of the Galois group of the extension $\mathbf{F}_{\boldsymbol{p}}/\mathbf{F}_p$ is entirely determined by the value of $\sigma(\alpha)$, which is a root of f. The discriminant of K is divisible by a prime p if and only if there are two distinct roots α_1 and α_2 of f which are equal modulo p. If this happens, then distinct automorphisms σ_1 and σ_2 with $\sigma_i(\alpha) = \alpha_i$ will satisfy $\varphi_{\boldsymbol{p}}(\sigma_1) = \varphi_{\boldsymbol{p}}(\sigma_2)$, so $\varphi_{\boldsymbol{p}}$ is not injective when p is ramified. The converse also follows for the same reason. □

Example 2.7.6. As usual, we discuss some basic examples.

The next result generalizes some of the observations of this example, and gives a very concrete interpretation of the Frobenius automorphism, in the spirit of Theorem 2.6.8.

Theorem 2.7.7. *Let $f \in \mathbf{Z}[\mathrm{X}]$ be an irreducible monic polynomial, let $\alpha \in \mathbf{C}$ be a root of f and $\mathrm{K} = \mathbf{Q}(\alpha)$ the number field it generates. Denote by $\mathrm{L} \supset \mathrm{K}$ the splitting field of f.*

Let

$$\boldsymbol{c} = \{x \in \mathbf{Z}_{\mathrm{K}} \mid x\mathbf{Z}_{\mathrm{K}} \subset \mathbf{Z}[\alpha]\},$$

and let p be a prime number coprime to $\boldsymbol{c}$ which is unramified in K. Write

$$f \,(\mathrm{mod}\, p) = h_1 \cdots h_g$$

where $g \geqslant 1$ and the h_i's are distinct irreducible monic polynomials of degree f_i.

Let $\boldsymbol{p}$ be a prime ideal dividing $p\mathbf{Z}_{\mathrm{K}}$. The Frobenius automorphism at $\boldsymbol{p}$, viewed as a permutation of the roots of f in $\mathbf{C}$, has cycle type given by a product of cycles of length f_i.

Proof. TODO □

CHAPTER 3

Elementary analytic number theory

APPENDIX A

Reminders and scripts

We summarize here, with precise references when needed, a number of elementary facts that are used in the rest of the text. We also provide for information some basic PARI/GP scripts that can be used for some of the numerical experiments which appeared in the text.

A.1. Some algebraic facts

We recall some simple definitions and results from algebra. Below, all rings are assumed to have a unit.

LEMMA A.1.1. *Let* A *be a commutative ring and let* $n \geqslant 1$ *be an integer. If* $(\boldsymbol{p}, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_n)$ *are prime ideals in* A *such that*

$$\boldsymbol{p} \subset \bigcup_{1 \leqslant i \leqslant n} \boldsymbol{p}_i,$$

then there exists some j *such that* $\boldsymbol{p} \subset \boldsymbol{p}_j$.

PROOF. We argue by induction on n, where the case $n = 1$ is tautological. Assume that $n \geqslant 2$, and that the property is valid for $n - 1$. Assume further by contradiction that $\boldsymbol{p}$ is not contained in $\boldsymbol{p}_i$ for any i.

It follows then, by induction, that $\boldsymbol{p}$ is not contained in the union of $\boldsymbol{p}_2$, ..., $\boldsymbol{p}_n$, so there is some element $x_1 \in \boldsymbol{p} - (\boldsymbol{p}_2 \cup \cdots \cup \boldsymbol{p}_n)$, and similarly for any j with $1 \leqslant j \leqslant n$, there is some $x_j \in \boldsymbol{p}$ which is not in $\boldsymbol{p}_i$ if $i \neq j$. We necessarily have $x_j \in \boldsymbol{p}_j$ for $1 \leqslant j \leqslant n$, since $\boldsymbol{p}$ is contained in the union of the $\boldsymbol{p}_i$ by assumption.

We then consider

$$y = x_1 + x_2 \cdots x_n.$$

Note that $y \in \boldsymbol{p}$ as combination of elements of $\boldsymbol{p}$. But $y \notin \boldsymbol{p}_1$, as this would imply that $x_2 \cdots x_n \in \boldsymbol{p}_1$, and therefore that some x_i with $2 \leqslant i \leqslant n$ is in $\boldsymbol{p}_1$, since this is a prime ideal, and this is impossible. For $2 \leqslant j \leqslant n$, we also have $y \notin \boldsymbol{p}_j$, as this would imply that $x_1 \in \boldsymbol{p}_j$, which is again not true. This means that we have a contradiction, so $\boldsymbol{p}$ had to be contained in some $\boldsymbol{p}_j$. □

We will use the basic theory of finite fields. In particular, we recall the non-degeneracy of the trace. We recall that (as in Chapter 2), the trace of an element x of a finite field E of characteristic p is the trace of the linear map $y \mapsto xy$, which is $\mathbf{F}_p$-linear.

LEMMA A.1.2. *Let* E *be a finite field of characteristic* p. *The* $\mathbf{F}_p$*-bilinear form* $(x, y) \mapsto \mathrm{Tr}(xy)$ *is non-degenerate.*

PROOF. When the degree of E over $\mathbf{F}_p$ is not divisible by p, this is straightforward: giving $x \in \mathrm{E}^\times$, we have $\mathrm{Tr}(x \cdot x^{-1}) = \mathrm{T}r(1) = [\mathrm{E} : \mathbf{F}_p]$, which is then non-zero.

Suppose then that $p \mid \dim[\mathrm{E} : \mathbf{F}_p]$. In this case, the simplest argument uses the formula

$$\mathrm{Tr}(z) = z + z^p + \cdots + z^{[\mathrm{E}:\mathbf{F}_p]-1}$$

for $z \in \mathrm{E}$. This shows that the trace is a polynomial function on E, represented by a polynomial of degree $p^{[\mathrm{E}:\mathbf{F}_p]-1}$. In particular, such a polynomial has at most that many roots, and since this number is strictly less than the size of E, there exists some z_0 such that $\mathrm{Tr}(z_0) \neq 0$. For a given $x \in \mathrm{E}^\times$, we then have $\mathrm{Tr}(xy) \neq 0$ when $y = z_0/x$. □

A.2. Pari/GP scripts

The scripts below have no pretention to being anything but simple tools to obtain some numerical evidence. In particular, there is no attempt to optimization of any kind.

(1) DENSITIES: given a polynomial f and an upper-bound x, returns a vector of length $1 + \deg(f)$ which contains the number of primes $p \leqslant x$ such that the number $\nu_f(p)$ of roots of f in $\mathbf{F}_p$ is equal to a given i is $v[i-1]$ (the shift is due to the fact that vectors in PARI/GP are indexed from 1 to the length).

```
densities(f,x)=
{
  local(j,v=vector(1+poldegree(f)));
  forprime(p=2,x,j=length(polrootsmod(f,p));v[j+1]++);
  v
}
```

(2) GAUSSSUM: given q, computes the quadratic Gauss sum of Proposition 2.3.1. Note that the Legendre symbol is computed by the function KRONECKER, referring to the Kronecker symbol, which generalizes the Legendre symbol.

```
gaussum(q)=
{
  sum(a=1,q-1,exp(2*I*Pi*a/q)*kronecker(a,q))
}
```

Bibliography

[1] N. Bourbaki: *Éléments de mathématique: Théories spectrales*, chapitres I et II, Springer, 2019.

[2] A. Chambert-Loir: *(Mostly) commutative algebra*, Universitext, Springer, 2021.

[3] P. Deligne: *La conjecture de Weil, II*, Publ. Math. IHÉS 52 (1980), 137–252.

[4] P.G.L. Dirichlet: *Verallgemeinerung eines Satzes aus der Lehre von den Kettenbrüchen nebst einigen Anwendungen auf die Theorie der Zahlen*, 1842; in G. Lejeune Dirichlet's Werke, Vol. 1, G. Reimer, Berlin, 1889; https://archive.org/details/glejeunedirichl01dirigoog/page/635/mode/1up?view=theater.

[5] M. Einsiedler and T. Ward: *Ergodic theory: with a view towards number theory*, Grad. Texts in Math. 259, Springer, 2011.

[6] P. Erdős: *Beweis eines Satzes von Tschebyschef*, Acta Scientiarum Mathematicarum 5 (1932), 194–198; https://old.renyi.hu/~p_erdos/1932-01.pdf.

[7] K. Ford: *The distribution of integers with a divisor in a given interval*, Annals of Math. 168 (2008), 367–433.

[8] J. Friedlander and H. Iwaniec: *Opera de cribro*, Colloquium Publ. 57, A.M.S, 2010.

[9] B.J. Green and T. Tao: *The primes contain arbitrarily long arithmetic progressions*, Annals of Math. 167 (2008), 481–547.

[10] G.H. Hardy and E.M. Wright: *An introduction to the theory of numbers*, 5th edition, Oxford, 1979.

[11] R. Heath–Brown: *Fermat's two squares theorem*, Invariant (1984), 3–5.

[12] K. Ireland and M. Rosen: *A classical introduction to modern number theory*, 2nd edition, Grad. Texts in Math. 84, Springer, 1992.

[13] H. Iwaniec and E. Kowalski: *Analytic Number Theory*, Colloquium Publ. 53, A.M.S, 2004.

[14] E. Kowalski: *An introduction to probabilistic number theory*, Cambridge Studies in Advanced Math. 192, Cambridge University Press, 2021.

[15] L. Kronecker: *Über die Irreductibilität von Gleichungen*, Monatsberichte der Königlich Preuss. Akad. der Wiss. Berlin (1880), 155–162.

[16] W. Bosma, J. Cannon and C. Playoust: *The Magma algebra system, I. The user language*, J. Symbolic Comput. 24 (1997), 235–265; also http://magma.maths.usyd.edu.au/magma/

[17] PARI/GP, version `2.6.0`, Bordeaux, 2011, http://pari.math.u-bordeaux.fr/.

[18] P. Samuel: *Théorie algébrique des nombres*, Hermann, 1967.

[19] K. Soundararajan: *Finite fields, with applications to combinatorics*, Student Math. Library 99, American Math. Soc., 2022.

[20] L. Washington: *Introduction to cyclotomic fields*, 2nd edition, Grad. Textes in Math. 83, Springer, 1997.

[21] D. Zagier: *A one-sentence proof that every prime* $p \equiv 4 \pmod 4$ *is a sum of two squares*, American Math. Monthly 97 (1990), 144.