Number Theory Summary

Divisibility and primes

The set $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ of integers, with its arithmetic operations of addition and multiplication, is the fundamental object of study in number theory. The structure of $\mathbb{Z}$ under addition is certainly easy to understand; it is an infinite cyclic group. The multiplicative structure of $\mathbb{Z}$ lies somewhat deeper. We have unique factorization of nonzero integers into primes - this is called the Fundamental Theorem of Arithmetic. The structure of $\mathbb{Z}$ under multiplication is then transparent given the primes, but the finer properties of the primes themselves are quite mysterious.

A *ring* is an abelian group, with the group operation called addition and written additively with $+$, on which there is another binary operation called multiplication which is written multiplicatively with a dot or without any sign. The multiplication is required to be associative, but does not need to be commutative. Addition and multiplication are related by distributivity; multiplication distributes over addition by the laws $a(b+c) = (ab)+(ac)$ and $(a+b)c = (ac)+(bc)$.

From the algebraic point of view, $\mathbb{Z}$ appears as a fundamental example of a commutative ring with multiplicative neutral element and without zero divisors. The latter property formalizes the observation that if $m, n \in \mathbb{Z}$ and $mn = 0$, then $m = 0$ or $n = 0$. So though division is not in general possible in $\mathbb{Z}$ without leaving its confines, nonzero common factors may be canceled, which is convenient when solving equations. Commutative rings with a multiplicative neutral element different from zero and without zero divisors are called *factorial rings* or *integral domains.* An integral domain in which every nonzero element has a multiplicative inverse is called a *field.*

The set $\mathbb{N}_0 \subset \mathbb{Z}$ of nonnegative integers $0, 1, 2, 3, \dots$, has an important property that is the basis for the principle of mathematical induction. The property may be formulated in various ways, but we use the one called Fermat's principle of infinite descent: If $A \subseteq \mathbb{N}_0$ is a set of nonnegative integers so that for every $a \in A$ there is some $b \in A$ with $b < a$, then $A = \emptyset$.

Division with remainder. *Suppose that m is a positive integer and n a nonnegative integer. Then there exist integers q and r with $0 \le r < m$, for which $n = mq + r$. The integer q is called the* quotient *and r the* remainder.

Proof. Given any positive integer m, let A_m be the set of nonnegative integers n for which division by m with remainder fails. Clearly $n \in A_m$ implies $n \ge m$ for otherwise $n = m\cdot 0 + n$ with $0 \le n < m$. But then $n - m$ is a nonnegative integer. And division of $n-m$ by m with remainder must fail, for $n-m = mq+r$ implies $n = m(q+1)+r$. Since $n-m < n$, Fermat's principle of infinite descent implies that $A_m = \emptyset$. □

Given two integers m and n we say that m *divides* n, which we write as $m|n$, if there is a third integer k such that $n = km$. Since $n = km$ and $m = jl$ implies $n = (kj)l$, we see that $l|m$ and $m|n$ implies $l|n$. Moreover if $m|n$ then $m|cn$ for any integer c, since $n = km$ implies $cn = (ck)m$. And if $m|n$ and $m|o$ then $m|(n \pm o)$ since $n = km$ and $o = lm$ gives $n \pm o = (k \pm l)m$.

The statement $0|n$ holds only if $n = 0$. Because this case is unimportant, and might sometimes require an exception, it is often excluded. If this case is excluded, in particular, $m|n$ is equivalent to n/m being an integer.

The *units* in $\mathbb{Z}$ are the integers u satisfying $u|1$. Clearly these are ± 1. Two elements a and b in $\mathbb{Z}$ are *associates* if $a = ub$ with u a unit. If $m|n$ without m and n being associates, we say that m *strictly divides* n. The nonzero elements in $\mathbb{Z}$ come in pairs $n, -n$ of associates. We note that if both $m|n$ and $n|m$ then $n = km = k(ln) = (kl)n$. Here $m = n = 0$, or $kl = 1$ so k is a unit. Hence m and n are associates.

An *ideal* $\mathfrak{a}$ in $\mathbb{Z}$ is a nonempty subset of $\mathbb{Z}$ such that $a, b \in \mathfrak{a}$ and $n \in \mathbb{Z}$ implies $a + b \in \mathfrak{a}$ and $na \in \mathfrak{a}$. It is clear that $\{0\}$ and $\mathbb{Z}$ are ideals. Moreover, if n is an integer, then $n\mathbb{Z}$ is an ideal. Such ideals generated by a single element are called *principal ideals*. Division with remainder implies that in $\mathbb{Z}$ all ideals are principal.

The zero ideal is clearly principal. If $\mathfrak{a} \neq \{0\}$ is an ideal in $\mathbb{Z}$ and $a \in \mathfrak{a}$, then $-a \in \mathfrak{a}$, so $\mathfrak{a}$ contains a smallest positive element m. Suppose that n is any element of $\mathfrak{a}$. We shall prove that $m|n$, so that $\mathfrak{a} = m\mathbb{Z}$. If $n = 0$ the statement is clear, and if $-n$ is a multiple of m, so is n. Thus we may assume that n is nonnegative. Divide n by m with remainder, so $n = qm + r$ where $0 \leq r < m$. Since $m, n \in \mathfrak{a}$, we also have $r = n - qm \in \mathfrak{a}$ because $\mathfrak{a}$ is an ideal. Because m is the smallest positive element in $\mathfrak{a}$, it follows that $r = 0$ and hence $m|n$. Thus every ideal in $\mathbb{Z}$ is principal.

A nonzero element $a \in \mathbb{Z}$ is an *irreducible element* if it is not a unit and if for any factorization $a = bc$, one of b and c is a unit. The irreducibles in $\mathbb{Z}$ are precisely the elements $\pm p$ where p runs through the prime numbers. A nonzero element $a \in \mathbb{Z}$ is a *prime element* if it is not a unit and if $a|bc$ implies $a|b$ or $a|c$. We see by induction that if a is a prime element and $a|b_1 b_2 \cdots b_s$ then $a|b_i$ for some i with $1 \leq i \leq s$. Any prime element is an irreducible element, for if $a = bc$ with a a prime element, then a divides one of b or c, say without of loss of generality that $a|b$. Then $b = ak$ so $a = akc$ and thus $1 = kc$, hence c is a unit.

Fundamental Theorem of Arithmetic. *Any nonzero integer that is not a unit is a product of irreducibles, unique up to order and associates.*

Proof. Any element in $\mathbb{Z}$ which is neither zero nor a unit has a factorization into irreducibles. For assume that a is such an element that has no factorization into irreducibles. Then a has some factorization $a = bc$ where neither b nor c is a unit, otherwise a would itself be an irreducible. Moreover at least one of the elements b and c has no factorization into irreducibles. Then we can find an infinite sequence $(c_i)_0^\infty$ of elements in $\mathbb{Z}$ such that $c_0 = a$ and $c_i = b_{i+1}c_{i+1}$ where b_{i+1} is never a unit. Now assume that (c_i) is any infinite sequence of nonzero

elements with $c_i = b_{i+1}c_{i+1}$. The principal ideals $c_i\mathbb{Z}$ form an ascending chain under inclusion, and hence

$$\mathfrak{c} = \bigcup_{i=0}^{\infty} c_i\mathbb{Z}$$

is an ideal. Since every ideal of $\mathbb{Z}$ is principal, there is some integer m such that $\mathfrak{c} = m\mathbb{Z}$. Because $m \in \mathfrak{c}$, there is some i such that $m \in c_i\mathbb{Z}$. Then $c_{i+1}\mathbb{Z} \subseteq \mathfrak{c} = c_i\mathbb{Z}$ and so $c_i|c_{i+1}$. On the other hand $c_{i+1}|c_i$ by the assumptions on the sequence (c_i). Hence c_i and c_{i+1} are associates, so b_{i+1} is a unit. We conclude that any element in $\mathbb{Z}$ which is neither zero nor a unit has a factorization.

If an integer n has a factorization into prime elements then this factorization is the unique factorization of that integer into irreducibles up to order and associates. For if $p_1 \cdots p_r = q_1 \cdots q_s$ where $p_1, \ldots, p_r$ are prime elements and $q_1, \ldots, q_s$ are irreducibles, then $p_r|q_k$ for some k. Since q_k is irreducible and p_r is not a unit, p_r and q_k are associates. We divide by p_r on both sides of $p_1 \cdots p_r = q_1 \cdots q_s$ leaving a unit q_k/p_r on the right-hand side, which we multiply into one of the other factors. Then we reindex the factors on the right-hand side leaving $p_1 \cdots p_{r-1} = q_1' \cdots q_{s-1}'$. Clearly we can continue the process, concluding that the p_i are pairwise associated to the q_j. This is the uniqueness statement.

We have so far obtained *existence* of factorizations into irreducibles and *uniqueness* of factorizations into prime elements in $\mathbb{Z}$. We will now prove that every irreducible in $\mathbb{Z}$ is a prime element.

Let a be an irreducible in $\mathbb{Z}$ such that $a|bc$. We shall show that $a|b$ or $a|c$ and hence that a is a prime element. If $a|b$ we are finished, so we may assume that $a\nmid b$ in which case a and b have no common factor, since a is irreducible. Consider now the ideal $\mathfrak{a} = a\mathbb{Z} + b\mathbb{Z}$. Since every ideal in $\mathbb{Z}$ is principal, there exists some integer m such that $\mathfrak{a} = m\mathbb{Z}$. Clearly $m|a$ and $m|b$ and so m is a unit, hence $\mathfrak{a} = \mathbb{Z}$. So the equation $ax + by = 1$ has a solution in integers x and y. Multiplying through by c gives $acx + bcy = c$. Since the left-hand side is divisible by a we obtain $a|c$ and so a is a prime element. Hence every nonzero element in $\mathbb{Z}$ that is not a unit has a factorization into primes, unique up to order and associates. □

The Fundamental Theorem of Arithmetic was first precisely formulated and proved by Gauss in the *Disquisitiones Arithmeticae.* The result that a prime dividing the product of two integers divides at least one of them occurs in the *Elements* of Euclid.

The concepts of divisibility, unit, associates, ideal, principal ideal, irreducible element, and prime element that played important roles in the proof of the Fundamental Theorem of Arithmetic carry over without change to any integral domain. The key fact that made the proof work is that every ideal in $\mathbb{Z}$ is principal. An integral domain in which every ideal is principal is called a *principal ideal domain* (abbreviated PID). An integral domain in which every nonzero element that is not a unit is a product of irreducibles, unique up to order and associates, is called a *unique factorization domain* (abbreviated UFD). We have thus proved the important result that any PID is a UFD.

That $\mathbb{Z}$ is a PID came from division with remainder, which is a result that is special to the integers. But for some integral domains there exists a substitute. An integral domain R is said to have a *gauge* $g : R \setminus \{0\} \to \mathbb{N}$ if $a|b$ implies $g(a) \leq g(b)$ and if for any $a \in R$ and $b \in R \setminus \{0\}$ there exist elements $q, r \in R$ such that $a = bq + r$ with $r = 0$ or $g(r) < g(b)$. An integral domain that carries a gauge is called a *Euclidean domain.* Every Euclidean domain R is a principal ideal domain. For if $\mathfrak{b}$ is a nonzero ideal in R then there exists some nonzero element $b \in \mathfrak{b}$ for which $g(b)$ is minimal. For any element $a \in R$, there exists elements $q, r \in R$ with $a = bq+r$ and $r = 0$ or $g(r) < g(b)$. The latter possibility is impossible by the choice of b, so $r = 0$ and thus $a = bq$. Hence $\mathfrak{b} = (b)$ and so R is a PID.

If k is a field and $a(x), b(x) \in k[x]$ with $b(x) \not\equiv 0$, there exist polynomials $q(x)$ and $r(x)$ over k with $a(x) = b(x)q(x) + r(x)$ and $\deg(r) < \deg(b)$. To prove this we may clearly assume that $\deg(a) \geq \deg(b)$. Then the method of undetermined coefficients applied with $q(x)$ a polynomial over k of degree at most $\deg(a) - \deg(b)$ and $r(x)$ a polynomial over k of degree at most $\deg(b) - 1$ leads to a linear system of $\deg(a) + 1$ equations over k and $\deg(a) - \deg(b) + 1 + \deg(b) = \deg(a) + 1$ unknowns. The system is square and choosing $a(x) \equiv 0$ we find the unique solution $q(x) \equiv 0$ and $r(x) \equiv 0$ by $b(x) \not\equiv 0$, so the system has a solution for any choice of $a(x)$. Since a polynomial that divides another polynomial has degree no larger than the other polynomial, the degree is a gauge and $k[x]$ is a principal ideal domain. Thus any polynomial over k factors uniquely up to order and scalar factors into a product of polynomials irreducible over k. For the units in $k[x]$ are the nonzero elements of k.

The ring $\mathbb{Z}[\sqrt{-1}] = \mathbb{Z} + \mathbb{Z}\sqrt{-1} = \mathbb{Z} + \mathbb{Z}i$ of *Gaussian integers* is also a Euclidean domain. We may choose $g(\alpha) = \alpha\overline{\alpha}$ as gauge. For given α and $\beta \neq 0$ Gaussian integers, there is a Gaussian integer σ with

$$\left|\mathrm{Re}\left(\frac{\alpha}{\beta}\right) - \mathrm{Re}(\sigma)\right| \leq 1/2 \quad \text{and} \quad \left|\mathrm{Im}\left(\frac{\alpha}{\beta}\right) - \mathrm{Im}(\sigma)\right| \leq 1/2.$$

Choosing $\rho = \alpha - \beta\sigma$ we see that

$$g(\rho) = |\alpha - \beta\sigma|^2 = |\beta|^2 \left|\frac{\alpha}{\beta} - \sigma\right|^2 \leq g(\beta)\left((1/2)^2 + (1/2)^2\right) < g(\beta).$$

If moreover $\alpha = \beta\gamma$ is an equation in Gaussian integers, $g(\alpha) = \alpha\overline{\alpha} = \beta\gamma\overline{\beta\gamma} = \beta\overline{\beta}\gamma\overline{\gamma} = g(\beta)g(\gamma)$ shows that $\beta|\alpha$ implies that $g(\beta) \leq g(\alpha)$, so g is a gauge. Hence the Gaussian integers constitute a PID and thus a UFD.

Unique factorization into primes yields a complete description of all the positive divisors of a positive integer. Any positive integer a has a factorization

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$$

into powers of distinct primes, where the product is empty if $a = 1$. If d is a positive integer with $d|a$, and p is a prime with $p|d$, then $p|a$. Since p is a

prime, this yields $p|p_i$ and hence $p = p_i$ for some i with $1 \leq i \leq r$. So the prime factorization of d is of the form

$$d = p_1^{\delta_1} p_2^{\delta_2} \cdots p_r^{\delta_r}$$

where the δ_i are nonnegative integers. Since $d|a$ we have $a = dc$ where c is also a divisor of a and hence of the form

$$c = p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_r^{\gamma_r}$$

where the γ_i are nonnegative integers. Now

$$p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r} = a = dc = p_1^{\gamma_1+\delta_1} p_2^{\gamma_2+\delta_2} \cdots p_r^{\gamma_r+\delta_r}$$

and since a prime cannot divide a product of primes unless it is one of the factors, we see that $\gamma_i + \delta_i = \alpha_i$ for all i with $1 \leq i \leq r$. Hence the divisors of

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$$

are precisely the integers of the form

$$d = p_1^{\delta_1} p_2^{\delta_2} \cdots p_r^{\delta_r}$$

with $0 \leq \delta_i \leq \alpha_i$ for all i with $1 \leq i \leq r$.

A positive integer s is *squarefree* if it is not divisible by any square $k^2 \geq 4$. The notation $p^\alpha || n$ signifies that $p^\alpha | n$, but that $p^{\alpha+1} \nmid n$, which means that p^α is the exact power to which p divides n. In this notation, an integer s is squarefree if and only if $p|s$ implies that $p||s$.

We end our discussion of unique factorization into primes with Euclid's proof that there are infinitely many primes. Let $p_1, p_2, \ldots, p_r$ be any finite collection of primes. Consider the nonzero integer $n = p_1 p_2 \cdots p_r + 1$. It is not a unit and so it is divisible by some prime q. But if q is contained in the collection $p_1, p_2, \ldots, p_r$ then $q|p_1 p_2 \cdots p_r$, which would imply that $q|1$, an impossibility. For a few arithmetic progressions the same argument may be used to show the existence of infinitely many primes in the progression, in particular for the arithmetic progression $3, 7, 11, 15, \ldots$ of integers of the form $4m - 1$. Let $p_1, p_2, \ldots, p_r$ be any finite collection of primes from the latter progression. Consider the nonzero integer $n = 4p_1 p_2 \cdots p_r - 1$. It is odd and it is not a unit and so it is divisible by some odd prime. Any product of integers of the form $4m + 1$ is itself of this form, so n must be divisible by some prime q of the form $4m - 1$. But if q is contained in the collection $p_1, p_2, \ldots, p_r$ then $q|p_1 p_2 \cdots p_r$, which would imply that $q|(-1)$, an impossibility.

Greatest common divisors

A *greatest common divisor* of integers $a_1, a_2, \ldots, a_n$ is an integer d such that $d|a_1, d|a_2, \ldots, d|a_n$, and if c is any integer such that $c|a_1, c|a_2, \ldots, c|a_n$ then $c|d$. We shall show that a greatest common divisor always exists. If all the a_i are zero, their greatest common divisor is zero. If not all the a_i are zero, we may remove all those a_i that are zero without changing the set of greatest common divisors. For every integer divides zero. We factor

$$a_i = u_i \prod_j p_j^{\alpha_{ij}}$$

into powers of distinct primes p_j and units u_i. The description of the divisors of an integer in terms of its prime factorization then implies that

$$d = \prod_j p_j^{\min\{\alpha_{1j}, \alpha_{2j}, \ldots, \alpha_{nj}\}}$$

is a greatest common divisor of the a_i that is unique up to associates.

We introduce the notation $\gcd(a_1, a_2, \ldots, a_n)$ for the nonnegative greatest common divisor of integers $a_1, a_2, \ldots, a_n$. The integers $a_1, a_2, \ldots, a_n$ are *coprime* if $\gcd(a_1, a_2, \ldots, a_n) = 1$, and are *pairwise coprime* if $\gcd(a_i, a_j) = 1$ for any i, j with $1 \leq i < j \leq n$. Note that the integers $4, 5, 6$ are coprime, but not pairwise coprime. Relatively prime is a synonym for coprime.

The formula $\gcd(ca_1, ca_2, \ldots, ca_n) = |c|\gcd(a_1, a_2, \ldots, a_n)$ follows directly from the definition of the greatest common divisor.

A *least common multiple* of integers $a_1, a_2, \ldots, a_n$ is an integer m such that $a_1|m, a_2|m, \ldots, a_n|m$, and if c is any integer such that $a_1|c, a_2|c, \ldots, a_n|c$ then $m|c$. Since any multiple of zero is zero, the least common multiple is zero if any of the integers a_i are zero. Hence we may assume that all the a_i are nonzero. We factor

$$a_i = u_i \prod_j p_j^{\alpha_{ij}}$$

into powers of distinct prime elements p_i and units u_i. The description of the divisors of an integer in terms of its prime factorization then implies that

$$m = \prod_j p_j^{\max\{\alpha_{1j}, \alpha_{2j}, \ldots, \alpha_{nj}\}}$$

is a least common multiple of the a_i that is unique up to associates.

We introduce the notation $\mathrm{lcm}[a_1, a_2, \ldots, a_n]$ for the nonnegative least common multiple of the integers $a_1, a_2, \ldots, a_n$. The formula $\mathrm{lcm}[ca_1, ca_2, \ldots, ca_n] = |c|\mathrm{lcm}[a_1, a_2, \ldots, a_n]$ follows directly from the definition of the least common multiple. The formula $\gcd(a, b)\mathrm{lcm}[a, b] = |ab|$ follows from the description of greatest common divisors and least common multiples in terms of prime factorizations. The analogue of this formula for three or more integers is not valid.

Since every ideal in $\mathbb{Z}$ is principal, for arbitrary integers $a_1, a_2, \ldots, a_n$ there exists an integer d such that

$$d\mathbb{Z} = a_1\mathbb{Z} + a_2\mathbb{Z} + \cdots + a_n\mathbb{Z}.$$

Clearly $a_i \in d\mathbb{Z}$ for each i with $1 \leq i \leq n$, and so $d|a_i$. On the other hand, if $c|a_i$ for $1 \leq i \leq n$, then c divides every element in $d\mathbb{Z}$, and $c|d$ in particular. Hence d is a greatest common divisor of the a_i. Thus the greatest common divisors are those integers d *minimal in the partial order of divisibility* for which the linear Diophantine equation

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = d$$

has a solution in integers x_i. Moreover the equation

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

has a solution in integers if and only if $d|b$ for some greatest common divisor d of the coefficients $a_1, a_2, \ldots, a_n$.

Finding all the divisors of a large integer without a huge number of trial divisions generally requires knowing its prime factorization. This can be a challenging problem, and the development of efficient factoring algorithms has drawn much attention. It is a remarkable fact that prime factorization can be short-circuited when finding a greatest common divisor of two integers. A procedure called the Euclidean Algorithm allows us to compute $\gcd(a, b)$ for integers a and b with great speed. The Euclidean algorithm is based on division with remainder. Assume without loss of generality that $a > b > 0$ and define sequences $\{a_k\}$, $\{b_k\}$, $\{q_k\}$ and $\{r_k\}$ by the requirements that $a_1 = a$ and $b_1 = b$, that $a_{k+1} = b_k$ and $b_{k+1} = r_k$, and that q_k is the quotient and r_k is the remainder on division of a_k by b_k. Hence

$$a_k = q_kb_k + r_k$$

where $0 \leq r_k < b_k$. The algorithm must terminate since r_k is nonnegative and $r_{k+1} = b_k < r_k$. We note that the common divisors of a_k and b_k coincide with the common divisors of b_k and r_k. But $a_{k+1} = b_k$ and $b_{k+1} = r_k$ so the common divisors of a_k and b_k coincide with the common divisors of a_{k+1} and b_{k+1}. Hence the common divisors of a_k and b_k are the same as the common divisors of a and b. The algorithm terminates when $r_m = 0$, and then $a_m = q_mb_m$. The common divisors of a_m and b_m are clearly the divisors of b_m. Hence b_m is a greatest common divisor of a and b. As an example we compute a greatest common divisor of 411 and 171 by the Euclidean algorithm:

$$\begin{aligned} 411 &= 2{\cdot}171 + 69 \\ 171 &= 2{\cdot}69 + 33 \\ 69 &= 2{\cdot}33 + 3 \\ 33 &= 11{\cdot}3 + 0. \end{aligned}$$

Hence $\gcd(411, 171) = 3$. We note that the maximal number of steps of the Euclidean algorithm is $O(\log(b))$. A good deal of work has been done on improvements and extensions of the Euclidean algorithm, and on their computational complexity.

The version of the Euclidean algorithm given by Euclid differs slightly from the version we use today, for it was based on repeated subtraction rather than repeated division with remainder. The basic insight underlying his version of the algorithm is that $\gcd(a, b) = \gcd(a - b, b)$.

The Euclidean algorithm can be used to solve the linear Diophantine equation

$$ax + by = c$$

in integers x and y. In order to have a nontrivial equation, we assume that a and b are nonzero. We run the Euclidean algorithm to find $\gcd(a, b)$, and then check whether $\gcd(a, b)|c$. If the answer is no, the equation has no solution in integers. If the answer is yes, we first use the information produced by the Euclidean algorithm to solve the equation

$$az + bw = \gcd(a, b)$$

in integers. We work backwards from the next to last equation produced by the Euclidean algorithm, expressing $\gcd(a, b)$ as an integer linear combination of a_k and b_k in each step. When we have obtained integers z_0 and w_0 that solve $az + bw = \gcd(a, b)$, we define $x_0 = (c/(a, b))z_0$ and $y_0 = (c/(a, b))w_0$. Then

$$\begin{aligned} ax_0 + by_0 &= a(c/\gcd(a, b))z_0 + b(c/\gcd(a, b))w_0 \\ &= (c/\gcd(a, b))(az_0 + bw_0) = (c/\gcd(a, b))\gcd(a, b) = c \end{aligned}$$

so we have a particular solution of $ax + by = c$. To find the general solution, put $x = x_0 + X$ and $y = y_0 + Y$ and substitute into the equation. This yields $aX + bY = 0$ and hence

$$\frac{a}{\gcd(a, b)}X = -\frac{b}{\gcd(a, b)}Y.$$

Since $a/\gcd(a, b)$ and $b/\gcd(a, b)$ are mutually prime, we see that $(a/\gcd(a, b))|Y$ and $(b/\gcd(a, b))|X$. Hence $X = t(b/\gcd(a, b))$ and $Y = -t(a/\gcd(a, b))$ where t is an integer parameter. The general solution is

$$x = x_0 + t\frac{b}{\gcd(a, b)} \quad \text{and} \quad y = y_0 - t\frac{a}{\gcd(a, b)}$$

where t runs through the integers.

As an example we find the general solution to the Diophantine equation $411x + 171y = 21$. We already know that $\gcd(411, 171) = 3$, so the equation has integer solutions since $3|21$. Now

$$\begin{aligned} 3 &= 69 - 2{\cdot}33 \\ &= 69 - 2{\cdot}(171 - 2{\cdot}69) = (-2){\cdot}171 + 5{\cdot}69 \\ &= (-2){\cdot}171 + 5{\cdot}(411 - 2{\cdot}171) = 411{\cdot}5 + (-12){\cdot}171 \end{aligned}$$

so $z_0 = 5$ and $w_0 = -12$ solves the equation $411z + 171w = 3$. Since $x_0 = (21/\gcd(411, 171))5 = 35$ and $y_0 = (21/\gcd(411, 171))(-12) = -84$, the general solution is $x = 35 + 137t$ and $y = -84 - 57t$ as t runs through the integers.

The Euclidean algorithm can be used to find the greatest common divisor $\gcd(a_1, a_2, \ldots, a_n)$ of more than two integers by applying the formula

$$\gcd(b_1, b_2, \ldots, b_k) = \gcd(b_1, \gcd(b_2, \ldots, b_k))$$

repeatedly. This formula is easily deduced from the description of the divisors of integers in terms of their prime factorizations.

Congruences

Consider $\mathbb{Z}$ as a ring, and let m be a positive integer. The quotient $\mathbb{Z}/m\mathbb{Z}$ by the ideal $m\mathbb{Z}$ is a finite ring, with elements $1 + m\mathbb{Z}, 2 + m\mathbb{Z}, \ldots, m + m\mathbb{Z}$. These elements, considered as subsets of $\mathbb{Z}$, are called *residue classes* modulo m, because all integers in one residue class leaves the same remainder (residue) r (with $0 \leq r < m$) upon divison by m. The residue class of a modulo m may be denoted $\underline{a}$ when the modulus is known from the context. Gauss introduced a convenient notation for calculations with representatives of residue classes. We say that a and b are *congruent* modulo m, written as $a \equiv b \pmod{m}$, if $m|(a-b)$.

It is equivalent that a and b leave the same remainder r (with $0 \leq r < m$) upon division by m, that is to say, that they are representatives of the same residue class modulo m. The integer m is called the *modulus* of the congruence. Congruence is an equivalence relation, with $a \equiv a \pmod{m}$, $a \equiv b \pmod{m}$ implies $b \equiv a \pmod{m}$, and $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$ implies $a \equiv c \pmod{m}$. If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then

$$a \pm c \equiv b \pm d \pmod{m} \quad \text{and} \quad ac \equiv bd \pmod{m}.$$

This is just another way of stating that the operations $+$ and $\cdot$ are well-defined in $\mathbb{Z}/m\mathbb{Z}$.

There are also some relations between congruences to different moduli. If $a \equiv b \pmod{m}$ and $d|m$, then $a \equiv b \pmod{d}$. Another way to state this observation is that if $d|m$, then there is a well-defined epimorphism $\sigma : \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/d\mathbb{Z}$ given by $\sigma(a + m\mathbb{Z}) = a + d\mathbb{Z}$. Suppose that $a \equiv b \pmod{m}$, and put $c = \gcd(a, b)$ and $d = \gcd(c, m)$. Then $a = ca'$, $b = cb'$, $c = dc'$ and $m = dm'$ for some integers a', b', c', m', so $dc'(a' - b') = c(a' - b') = a - b = km = kdm'$. This yields $c'|km'$ and hence $c'|k$ by the Fundamental Theorem of Arithmetic, since $\gcd(c', m') = 1$. Then

$$\frac{a}{\gcd(a, b)} \equiv \frac{b}{\gcd(a, b)} \left(\mathrm{mod} \frac{m}{\gcd(m, \gcd(a, b))} \right)$$

on canceling the common factor.

Chinese Remainder Theorem. *If $m_1, \ldots, m_r$ are pairwise coprime positive integers and $b_1, \ldots, b_r$ are integers, then the simultaneous congruences*

$$\begin{aligned} n &\equiv b_1 \pmod{m_1} \\ n &\equiv b_2 \pmod{m_2} \\ &\vdots \\ n &\equiv b_r \pmod{m_r} \end{aligned}$$

have a unique solution $n \equiv n_0 \pmod{m_1 m_2 \cdots m_r}$.

Proof. By induction, it is enough to establish the case $r = 2$. The linear Diophantine equation $m_1 x + m_2 y = 1$ has a solution (x_0, y_0), because $\gcd(m_1, m_2) = 1$. Note that $m_2 x_0 \equiv 1 \pmod{m_1}$ and $m_1 y_0 \equiv 1 \pmod{m_2}$, and define $n_0 = b_1 m_2 x_0 + b_2 m_1 y_0$. Then

$$n_0 \equiv b_1 m_2 x_0 \equiv b_1 \pmod{m_1}$$

and

$$n_0 \equiv b_2 m_1 y_0 \equiv b_2 \pmod{m_2},$$

so the two simultaneous congruences have a common solution. This solution is unique modulo $m_1 m_2$. For if $n \equiv b_1 \pmod{m_1}$ and $n \equiv b_2 \pmod{m_2}$ while $n' \equiv b_1 \pmod{m_1}$ and $n' \equiv b_2 \pmod{m_2}$, then $n - n' \equiv b_1 - b_1 \equiv 0 \pmod{m_1}$ and $n - n' \equiv b_2 - b_2 \equiv 0 \pmod{m_2}$. But $m_1 | (n - n')$ and $m_2 | (n - n')$ and $\gcd(m_1, m_2) = 1$ imply $m_1 m_2 | (n - n')$ by the Fundamental Theorem of Arithmetic. □

The Chinese remainder theorem is thus termed because the method of calculation underlying the proof was first set forth in a handbook by the Chinese mathematician Sun Zi. He was such an obscure figure that even the century of his birth is not definitely known, but he may have lived about 1600 years ago.

Every element $\underline{a}$ in the ring $\mathbb{Z}/m\mathbb{Z}$ is either a zero divisor or has a multiplicative inverse. If $\gcd(m, a) = c \geq 2$, then $\underline{a}\,\underline{b} = \underline{d}\,\underline{m} = \underline{d}\,\underline{0} = \underline{0}$ with $b = m/c$ and $d = a/c$. If $\gcd(m, a) = 1$ on the other hand, then the linear Diophantine equation $ax + my = 1$ has some solution (x_0, y_0). But $ax_0 + my_0 = 1$ implies that $ax_0 \equiv 1 \pmod{m}$, or $\underline{a}\,\underline{x_0} = \underline{1}$ in $\mathbb{Z}/m\mathbb{Z}$. In particular $\mathbb{Z}/p\mathbb{Z}$ is a field for every prime p. Up to isomorphism it is the only field with p elements. For if F is another such field, there is a homomorphism $\vartheta : \mathbb{Z}/p\mathbb{Z} \to F$ defined by $\vartheta(\underline{1}) = 1$. Any nonzero homomorphism between fields is a monomorphism. Since the two fields both have p elements, ϑ is also an epimorphism.

An element of $\mathbb{Z}/m\mathrm{Z}$ that has a multiplicative inverse is called a *reduced residue class.* The set $(\mathbb{Z}/m\mathbb{Z})^\times$ of reduced residue classes forms a group under multiplication, for if a and b are both coprime with m, then so is ab. Note that $\underline{1}$ is the unit element. Define $\phi(m)$ to be the order of this group, that is to say

$$\phi(m) = |(\mathbb{Z}/m\mathbb{Z})^\times| = \sum_{\substack{1 \leq a \leq m \\ \gcd(m,a)=1}} 1.$$

It is a basic result in group theory that the order of an element of a finite group divides the order of the group, hence

$$a^{\phi(m)} \equiv 1 \pmod{m}$$

if $\gcd(m, a) = 1$. This congruence is due to Euler, and ϕ is called the Euler phi-function, or the Euler totient in the older literature. In the special case when the modulus is a prime, we obtain $\phi(p) = p - 1$ because $\mathbb{Z}/p\mathbb{Z}$ is a field for p prime. Then the congruence

$$x^p \equiv x \pmod{p}$$

holds for all integers x, for any prime p. For either p divides x or else p and x are coprime. The last congruence is known as the little theorem of Fermat.

Suppose that m_1 and m_2 are positive integers. Any integer a that is coprime with $m_1 m_2$ is coprime with m_1 and with m_2. Thus there are epimorphisms $\sigma_1 : (\mathbb{Z}/m_1 m_2\mathbb{Z})^\times \to (\mathbb{Z}/m_1\mathbb{Z})^\times$ given by $\sigma_1(a + m_1 m_2\mathbb{Z}) = a + m_1\mathbb{Z}$ and $\sigma_2 : (\mathbb{Z}/m_1 m_2\mathbb{Z})^\times \to (\mathbb{Z}/m_2\mathbb{Z})^\times$ given by $\sigma_2(a + m_1 m_2\mathbb{Z}) = a + m_2\mathbb{Z}$. Then there is an epimorphism $\sigma_1 \oplus \sigma_2 : (\mathbb{Z}/m_1 m_2\mathbb{Z})^\times \to (\mathbb{Z}/m_1\mathbb{Z})^\times \oplus (\mathbb{Z}/m_2\mathbb{Z})^\times$ given by $(\sigma_1 \oplus \sigma_2)(a + m_1 m_2\mathbb{Z}) = \sigma_1(a + m_1\mathbb{Z}) \oplus \sigma_1(a + m_2\mathbb{Z})$. This epimorphism is mono if m_1 and m_2 are coprime, for then $(\sigma_1 \oplus \sigma_2)(a + m_1 m_2\mathbb{Z}) = \underline{1}$ is equivalent to the simultaneous congruences

$$\begin{aligned} a &\equiv 1 \pmod{m_1} \\ a &\equiv 1 \pmod{m_2}, \end{aligned}$$

and these have only one solution $a \equiv 1$ modulo $m_1 m_2$, by the Chinese remainder theorem. Thus

$$\begin{aligned} \phi(m_1 m_2) &= |(\mathbb{Z}/m_1 m_2\mathbb{Z})^\times| = |(\mathbb{Z}/m_1\mathbb{Z})^\times \oplus (\mathbb{Z}/m_2\mathbb{Z})^\times| \\ &= |(\mathbb{Z}/m_1\mathbb{Z})^\times||(\mathbb{Z}/m_2\mathbb{Z})^\times| = \phi(m_1)\phi(m_2) \end{aligned}$$

if m_1 and m_2 are coprime. The relation $\phi(m_1 m_2) = \phi(m_1)\phi(m_2)$ for m_1 and m_2 coprime reduces the calculation of ϕ to the case of prime powers. This yields

$$\phi(p_1^{\alpha_1} \cdots p_r^{\alpha_r}) = (p_1 - 1) \cdots (p_r - 1) p_1^{\alpha_1 - 1} \cdots p_r^{\alpha_r - 1},$$

because $\phi(p^\alpha) = p^\alpha - p^{\alpha-1}$ if p is prime. For there are $p^{\alpha-1}$ integers a in the interval from 1 to p^α that are divisible by p.

The product of the elements of a finite abelian group equals the product of those elements that have order equal to one or two. For the others cancel in pairs, since they do not equal their inverses. Applying this observation to the group $(\mathbb{Z}/p\mathbb{Z})^\times$ for p prime, and noting that the congruence $x^2 \equiv 1 \pmod{p}$ has only the solutions $x \equiv \pm 1$ modulo p when p is prime, we obtain the congruence

$$(p - 1)! \equiv -1 \pmod{p}.$$

The result is named after J. Wilson, who rediscovered it in the eighteenth century but did not prove it. The Arab scholar Ibn al-Haytham (Alhacen) had

made use of it more than half a millennium before. It was first proved by J. L. Lagrange in 1773.

The linear congruence $ax \equiv b \pmod{m}$ in one unknown is equivalent to the linear Diophantine equation $ax + my = b$, and so does not call for particular comment. But if $P \in \mathbb{Z}[x]$ is a polynomial over the integers, we may consider the more general polynomial congruence $P(x) \equiv 0 \pmod{m}$ for m a positive integer. This may be considered as a polynomial equation in the ring $\mathbb{Z}/m\mathbb{Z}$. There is then a very substantial contrast between the case when m is composite, and the case when m is prime. In the former case, $\mathbb{Z}/m\mathbb{Z}$ has zero divisors, which greatly complicate the treatment of equations in the ring. But if m is prime, then this ring is a field, and the situation is much more transparent.

We consider a polynomial congruence modulo a prime p. By means of the little theorem of Fermat, the degree may always be reduced until one obtains a congruence with $\deg(P) \leq p - 1$. One is also free to reduce the coefficients of P modulo p without changing the set of solutions.

Lagrange's theorem. *The number of mutually incongruent solutions of a non-trivial polynomial congruence modulo a prime is at most equal to the degree of the polynomial.*

Proof. Suppose that p is a prime, $P(x)$ a polynomial over the field $\mathbb{Z}/p\mathbb{Z}$ and x_1 an arbitrary element of $\mathbb{Z}/p\mathbb{Z}$. Then $P(x) \equiv Q(x)(x - x_1) + r \pmod{p}$ where $Q(x)$ is a polynomial over $\mathbb{Z}/p\mathbb{Z}$ with degree one smaller than $P(x)$ and r is an element in $\mathbb{Z}/p\mathbb{Z}$. Supposing that x_1 is a solution of $P(x) \equiv 0$ modulo p, we see that $r \equiv 0$ and so $P(x) \equiv Q(x)(x - x_1)$ modulo p. If $Q(x)$ has a zero modulo p, we may repeat the procedure. But clearly we cannot split off more than $\deg(P)$ linear factors, since the degree decreases by one in each step and $P(x)$ was assumed not identically zero modulo p. □

Primitive roots

If the multiplicative group $(\mathbb{Z}/m\mathbb{Z})^\times$ of reduced residue classes modulo m is cyclic, then any integer g for which $g + m\mathbb{Z}$ is a generator of this group, is called a *primitive root* modulo m. In that case, if a is any integer coprime with m, then $g^j \equiv a \pmod{m}$ for some integer j. In particular, the powers $g, g^2, \ldots, g^{\phi(m)}$ are all distinct modulo m. Furthermore g^k is a primitive root modulo m if and only if k and $\phi(m)$ are coprime. Thus if there exists a primitive root modulo m, there are $\phi(\phi(m))$ distinct ones modulo $\phi(m)$.

Existence of primitive roots. *There exists a primitive root modulo $m \geq 2$ if and only if $m = 2, 4, p^\alpha$ or $2p^\alpha$ for some odd prime p.*

Proof. There exists a primitive root modulo p for any prime p. For it is a consequence of the structure theorem for finitely generated abelian groups that $(\mathbb{Z}/p\mathbb{Z})^\times \cong C_{d_1} \oplus C_{d_2} \oplus \cdots \oplus C_{d_r}$ where C_d denotes the cyclic group of order d, and $d_1, \ldots, d_r$ are positive integers with $d_1 | d_2 | \cdots | d_r$. Then clearly every element of $\mathbb{Z}/p\mathbb{Z}$ has order equal to a divisor of d_r, so $a^{d_r} = 1$ for a and p coprime, and

the polynomial $x^{d_r} - 1$ has at least $d_1 d_2 \cdots d_r$ distinct roots modulo p. But it also has at most d_r distinct roots modulo p by the theorem of Lagrange, hence $r = 1$ and $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic.

There is also a primitive root modulo any power p^α of an odd prime p. Let g be a primitive root modulo p and h an integer with

$$h \not\equiv \frac{g^p - g}{p} \pmod{p},$$

and put $r = g + hp$. We will show that r is a primitive root modulo p^α for every $\alpha \geq 2$. Denote by e the order of r as an element of the group $\mathbb{Z}/p^\alpha\mathbb{Z}$. We know that e divides $|\mathbb{Z}/p^\alpha\mathbb{Z}| = \phi(p^\alpha)$ and the task at hand is to show that $e = \phi(p^\alpha) = (p-1)p^{\alpha-1}$. Then r has the maximal number of distinct powers modulo p^α and is thus a primitive root. Now $r^{p-1} \equiv 1 \pmod{p}$ by Fermat's little theorem, while r is a primitive root modulo p since g is. Thus e has to be a multiple of $p - 1$. Hence it will be sufficient to show that

$$r^{(p-1)p^{\alpha-2}} \not\equiv 1 \pmod{p^\alpha},$$

and this is where most of the work lies.

Note that $r^p \equiv g^p + h^p p^p \equiv g^p \pmod{p}$ by the Binomial theorem, since the exponent is prime. Thus

$$r^p - r \equiv g^p - g - hp \equiv p\left(\frac{g^p - g}{p} - h\right) \not\equiv 0 \pmod{p^2}$$

and so $r^{p-1} = 1 + kp$ with k not divisible by p. Expanding $(1 + kp^j)^p$ for $j \geq 1$ by the Binomial theorem, the first two terms are 1 and kp^{j+1} while the other terms are either divisible by p^{2j+1} or p^{pj}. Thus all these terms are divisible by p^{j+2} since $p \geq 3$ is an odd prime, and

$$(1 + kp^j)^p \equiv 1 + kp^{j+1} \pmod{p^{j+2}}$$

follows. In particular,

$$(r^{p-1})^p \equiv 1 + kp^2 \pmod{p^3}$$

holds. Thus

$$(r^{p-1})^{p^s} \equiv 1 + kp^{s+1} \pmod{p^{s+2}}$$

holds for $s = 1$, and assuming it to hold for some s, the congruence

$$(1 + kp^{s+1})^p \equiv 1 + kp^{s+2} \pmod{p^{s+3}}$$

obtained by choosing $j = s + 1$ above, yields

$$(r^{p-1})^{p^{s+1}} \equiv (1 + kp^{s+1})^p \equiv 1 + kp^{s+2} \pmod{p^{s+3}}.$$

Hence

$$(r^{p-1})^{p^s} \equiv 1 + kp^{s+1} \pmod{p^{s+2}}$$

holds for all s by induction. Then

$$r^{(p-1)p^{\alpha-2}} \equiv 1 + kp^{\alpha-1} \not\equiv 1 \pmod{p^\alpha},$$

by substituting $s = \alpha - 2$.

Evidently 1 is a primitive root modulo 2 and 3 is a primitive root modulo 2^2. But $a^2 \equiv 1 \pmod{2^3}$ for any odd integer a, and if

$$a^{2^{\alpha-2}} \equiv 1 \pmod{2^\alpha},$$

then for $\alpha \geq 3$ we have

$$a^{2^{\alpha-1}} \equiv (1 + b2^\alpha)^2 \equiv 1 + b2^{\alpha+1} + b^2 2^{2\alpha} \equiv 1 \pmod{2^{\alpha+1}}$$

with b some integer. Hence

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

holds for $\alpha \geq 3$ by induction. Since the order of a is strictly smaller than $\phi(2^\alpha)$ for each odd integer, there is no primitive root modulo 2^α when $\alpha \geq 3$.

The Chinese remainder theorem implies that if $m = p_1^{\alpha_1} \cdots p_\ell^{\alpha_\ell}$ is the factorization of m into prime powers, then

$$(\mathbb{Z}/m\mathbb{Z})^\times \cong G_1 \oplus \cdots \oplus G_\ell$$

where G_i is isomorphic to $(\mathbb{Z}/p^{\alpha_i}\mathbb{Z})^\times$. Since any subgroup of a cyclic group is cyclic, we see that 2^α with $\alpha \geq 3$ cannot occur in the prime factorization of m. Moreover, a direct sum of cyclic groups is cyclic if and only if all but one summand is trivial. Since $(\mathbb{Z}/p^\alpha\mathbb{Z})^\times$ is trivial only when $p^\alpha = 2$, we see that $m = 2p^\alpha$ with p an odd prime are the only moduli that have primitive roots, beyond those that we have already found. □

Supposing g to be a primitive root modulo m and a to be an integer coprime with m, the congruence

$$g^e \equiv a \pmod{m}$$

has a unique solution e in the interval $0 \leq e\phi(m) - 1$. This unique exponent e is called the *index* of a to base g and is denoted by $\operatorname{ind}_g(a)$ (where m is assumed known from context.) The index of a depends only on the residue class of a modulo m. The formula

$$\operatorname{ind}_g(ab) \equiv \operatorname{ind}_g(a) + \operatorname{ind}_g(b) \pmod{\phi(m)}$$

reduces multiplication modulo m to addition modulo $\phi(m)$, and also permits the calculation of powers, and of k-th roots when k and $\phi(m)$ are coprime. This technique of calculation is called the *index calculus.* Clearly the index is analogous to the logarithm.

Quadratic residues

The Legendre symbol $(n|p)$ modulo an odd prime p may be defined by the congruence

$$\left(\frac{n}{p}\right) \equiv n^{(p-1)/2} \pmod{p}$$

and the inequality

$$-\frac{p-1}{2} \leq \left(\frac{n}{p}\right) \leq \frac{p-1}{2},$$

for all integers $n \not\equiv 0 \pmod{p}$. Fermat's little theorem implies that the Legendre symbol takes the values ± 1. Note that since there exists a primitive root modulo p, there exists some n with $(n|p) = -1$. In particular, the homomorphism $\sigma : (\mathbb{Z}/p\mathbb{Z})^\times \to \{\pm 1\}$ given by $n + p\mathbb{Z} \mapsto (n|p)$ is not epi, so the kernel of this homomorphism has index equal to 2. Evidently the sum of all the Legendre symbols modulo any fixed odd prime p is zero.

An integer n coprime with m is called a *quadratic residue* modulo m if the congruence

$$x^2 \equiv n \pmod{m}$$

has a solution. If not, n is called a *quadratic nonresidue* modulo m. Integers n with $\gcd(m, n) \geq 2$ are neither quadratic residues nor nonresidues modulo m.

Though the above definition of the Legendre symbol is essentially the one Legendre used, today it is customary to define $(n|p)$ to equal 1 if n is a quadratic residue modulo p and -1 if n is a quadratic nonresidue modulo p. Euler's criterion implies that the two definitions are equivalent. In the algebraical spirit that governs this Summary, we adopt Legendre's definition. For $(n|p) = -1$ if and only if the order of $\underline{n}$ as an element of the group $(\mathbb{Z}/p\mathbb{Z})^\times$ is maximal. It is obvious that the product of two quadratic residues modulo p is itself a quadratic residue modulo p, but the next result shows that more is true.

Euler's Criterion. *An integer n coprime with an odd prime p is a quadratic residue modulo p if and only if $(n|p) = 1$.*

Proof. The integers $1, 2^2, \ldots, (p-1)^2$ are the quadratic residues modulo p, but with repetitions modulo p. If a pair of these integers, say k^2 and m^2, are congruent modulo p, then $(m-k)(m+k) \equiv m^2 - k^2 \equiv 0 \pmod{p}$. Thus $m \equiv \pm k \pmod{p}$ since p is a prime, and so there are exactly $(p-1)/2$ quadratic residues modulo p. If n is a quadratic residue modulo p, then there is some integer m so that

$$n^{(p-1)/2} \equiv (m^2)^{(p-1)/2} \equiv m^{p-1} \equiv 1 \pmod{p},$$

and thus $(n|p) = 1$. Evidently every quadratic residue is a root of the polynomial $y^{(p-1)/2} - 1$ modulo p. But this polynomial has at most $(p-1)/2$ roots modulo p by Lagrange's theorem, so no quadratic nonresidue is a root of it. Hence n a quadratic nonresidue implies that $(n|p) = -1$. □

It follows from Euler's criterion that the quadratic residue classes modulo an odd prime p form a subgroup of $(\mathbb{Z}/p\mathbb{Z})^\times$ of index equal to 2. In particular, if the two congruences $x^2 \equiv m \pmod p$ and $x^2 \equiv n \pmod p$ have no solutions, then the congruence $x^2 \equiv mn \pmod p$ necessarily has a solution, and this is not at all obvious.

The Law of Quadratic Reciprocity. *The relation*

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

between Legendre symbols holds for all distinct odd primes p and q.

Proof. For every odd positive integer n denote by N_n the number of solutions in $(\mathbb{Z}/q\mathbb{Z})^n$ of the equation $x_1^2 - x_2^2 + x_3^2 - \cdots + x_n^2 = 1$. Making the change of variable $x_1 = y + x_2$ yields $y^2 + x_3^2 - \cdots + x_n^2 - 1 = 2yx_2$. For each $y \neq 0$ and any choice of $x_3, \ldots, x_n$ there is a unique corresponding value of x_2 which also determines x_1. This yields a total of $(q-1)q^{n-2}$ solutions. Together with $y = 0$ the original equation is equivalent to the system

$$x_1 = x_2 \quad \text{and} \quad x_3^2 - \cdots + x_n^2 = 1,$$

and so there are qN_{n-2} solutions for $y = 0$. Thus $N_n = (q-1)q^{n-2} + qN_{n-2}$ and so

$$\begin{aligned} N_n &= (q-1)q^{n-2} + qN_{n-2} = (q-1)q^{n-2} + q((q-1)q^{n-4} + qN_{n-4}) \\ &= (q-1)q^{n-2} + (q-1)q^{n-3} + q^2N_{n-4} = \cdots \\ &= (q-1)q^{n-2} + \cdots + (q-1)q^{n-1-k} + q^kN_{n-2k}. \end{aligned}$$

Choosing $k = (n-1)/2$ yields

$$\begin{aligned} N_n &= (q-1)q^{n-2} + \cdots + (q-1)q^{(n-3)/2} + q^{(n-1)/2}N_1 \\ &= q^{n-1} + q^{(n-1)/2}(N_1 - 1) = q^{n-1} + q^{(n-1)/2}, \end{aligned}$$

and thus $N_p \equiv 1 + (q|p) \pmod p$ on taking $n = p$ and using the definition of the Legendre symbol.

We obtain the law of quadratic reciprocity by computing N_p in another way. Let $N(x^2 = t)$ denote the number of solutions of $x^2 = t$ in $\mathbb{Z}/q\mathbb{Z}$. Then

$$N_p = \sum_{t_1+\cdots+t_p=1} N(x_1^2 = t_1)N(x_2^2 = -t_2)N(x_2^2 = t_3)\cdots N(x_p^2 = t_p)$$

where $t_1, t_2, \ldots, t_p$ range over $\mathbb{Z}/q\mathbb{Z}$, and so

$$N_p = \sum_{t_1+\cdots+t_p=1} \left(1 + \left(\frac{t_1}{q}\right)\right)\left(1 + \left(\frac{-t_2}{q}\right)\right)\left(1 + \left(\frac{t_3}{q}\right)\right)\cdots\left(1 + \left(\frac{t_p}{q}\right)\right),$$

for $x^2 = t$ has two solutions if t is a quadratic residue class modulo q, one solution if $t = 0$ and no solutions if t is a quadratic nonresidue class. Multiplying out the product under the last summation sign, only the terms 1 and $(t_1|q)(-t_2|q)(t_3|q)\cdots(t_p|q)$ in the inner sum contribute to the outer sum. For if some but not not all the factors of the form $(\pm t_j|q)$ are present in a term, then at least one t_j will run unrestrictedly over $\mathbb{Z}/q\mathbb{Z}$, and summing over this t_j yields zero, since the sum of the Legendre symbol over a complete collection of residue classes modulo q is zero. Thus

$$N_p = q^{p-1} + \sum_{t_1+\cdots+t_p=1} \left(\frac{t_1}{q}\right)\left(\frac{-t_2}{q}\right)\left(\frac{t_3}{q}\right)\cdots\left(\frac{t_p}{q}\right)$$

since $t_1 + t_2 + \cdots + t_p = 1$ has q^{p-1} solutions. Now

$$N_p = q^{p-1} + \left(\frac{(-1)^{(p-1)/2}}{q}\right) \sum_{t_1+\cdots+t_p=1} \left(\frac{t_1\cdots t_p}{q}\right).$$

Each tuple $(u_1, \ldots, u_p)$ with $u_1 + \cdots + u_p = 1$ belongs to an equivalence class under permutation, and all members of this equivalence class are also solutions of $t_1 + \cdots t_p = 1$ and the value of the term $(t_1\cdots t_p|q)$ is the same for all of them. If all u_j are equal, that is to say if $u_j = p^{-1}$, then this equivalence class is a singleton, but otherwise the number of elements in the equivalence class is divisible by p. For this number is of the form $p!/k_1!\cdots k_r!$ where $k_1, \ldots, k_r$ are the numbers of elements of $(u_1, \ldots, u_p)$ that are identical in groups. Thus

$$\sum_{t_1+\cdots+t_p=1} \left(\frac{t_1\cdots t_p}{q}\right) \equiv \left(\frac{p^{-p}}{q}\right) \pmod{p},$$

and so

$$\begin{aligned} N_p &\equiv 1 + \left(\frac{(-1)^{(p-1)/2}}{q}\right)\left(\frac{p^{-p}}{q}\right) \equiv 1 + \left((-1)^{(p-1)/2}\right)^{(q-1)/2}\left(\frac{p}{q}\right)^{-p} \\ &\equiv 1 + (-1)^{(p-1)(q-1)/4}\left(\frac{p}{q}\right) \pmod{p}. \end{aligned}$$

Now comparison with the other congruence

$$N_p \equiv 1 + \left(\frac{q}{p}\right) \pmod{p}$$

yields the law of quadratic reciprocity. □

This is the central result on quadratic congruences. It was first conjectured by Euler, and first proved by Gauss. The proof presented here was found recently by W. Castryck.

There are many proofs of the quadratic reciprocity law. Gauss found eight proofs, some based on the following criterion for quadratic residuacy.

Gauss' Lemma. *Let p be an odd prime and n an integer not divisible by p. Denote by μ the number of the integers $n, 2n, \ldots, (p-1)n/2$ that are congruent to some integer in the interval $[-(p-1)/2, -1]$. Then $(n|p) = (-1)^\mu$.*

Proof. The integers $n, 2n, \ldots, (p-1)n/2$ are pairwise incongruent modulo p and each is congruent to one of the integers $-(p-1)/2, \ldots, -1, 1, \ldots, (p-1)/2$ modulo p. Moreover, $jn \equiv -kn \pmod{p}$ implies that $(j+k)n \equiv 0 \pmod{p}$, which is impossible. Thus

$$n \cdot 2n \cdots (p-1)n/2 \equiv (-1)^\mu 1 \cdot 2 \cdots (p-1)/2 \pmod{p}$$

and so $(n|p) \equiv n^{(p-1)/2} \equiv (-1)^\mu \pmod{p}$ on canceling the common factor $1 \cdot 2 \cdots (p-1)/2$. □

The quadratic reciprocity law allows us to determine if n is a quadratic residue modulo p if n is an odd positive integer. The case of odd negative integers is also easily handled since $(-1|p) = (-1)^{(p-1)/2}$ by our definition of the Legendre symbol. This leaves the case of even integers. Choosing $n = 2$ in Gauss' lemma, we must find the number μ of even integers $2, 4, \ldots, p-1$ congruent modulo p to integers in the interval $[-(p-1)/2, -1)$. But this is obviously equal to the number of even integers in the interval $[(p+1)/2, p-1]$. Counting the number of integers k for which $(p+1)/2 \leq 2k \leq p-1$ yields

$$\mu = \frac{p-1}{2} - \left[\frac{p+1}{4}\right]$$

if $(p+1)/4$ is not an integer, and

$$\mu = \frac{p-1}{2} - \left[\frac{p+1}{4}\right] + 1$$

otherwise. Considering the possibilities for μ as p ranges through the odd residue classes modulo 8, one sees that $\mu \equiv (p^2-1)/8 \pmod{2}$, and so $(2|p) = (-1)^{(p^2-1)/8}$. The formulas for $(-1|p)$ and $(2|p)$ are called the *supplementary laws,* for together with the law of quadratic reciprocity, they permit the efficient calculation of $(n|p)$ for any integer n.

The *Jacobi symbol* $(n|m)$ generalizes the Legendre symbol to odd positive integer moduli m. Supposing m to have the prime factorization

$$m = p_1^{\alpha_1} \cdots p_r^{\alpha_r},$$

and $\gcd(m, n) = 1$, the Jacobi symbol is defined by

$$\left(\frac{n}{m}\right) = \left(\frac{n}{p_1}\right)^{\alpha_1} \cdots \left(\frac{n}{p_r}\right)^{\alpha_r}$$

in terms of Legendre symbols. Supposing n also to be an odd positive integer, with prime factorization

$$n = q_1^{\beta_1} \cdots q_s^{\beta_s},$$

we have

$$\left(\frac{n}{m}\right) = \left(\frac{n}{p_1}\right)^{\alpha_1} \cdots \left(\frac{n}{p_r}\right)^{\alpha_1} = \left(\frac{q_1}{p_1}\right)^{\alpha_1\beta_1} \cdots \left(\frac{q_k}{p_j}\right)^{\alpha_j\beta_k} \cdots \left(\frac{q_s}{p_r}\right)^{\alpha_r\beta_s}$$

and similarly

$$\left(\frac{m}{n}\right) = \left(\frac{p_1}{q_1}\right)^{\alpha_1\beta_1} \cdots \left(\frac{p_j}{q_k}\right)^{\alpha_j\beta_k} \cdots \left(\frac{p_r}{q_s}\right)^{\alpha_r\beta_s}.$$

Thus

$$\left(\frac{n}{m}\right)\left(\frac{m}{n}\right) = \cdots \left(\left(\frac{q_k}{p_j}\right)\left(\frac{p_j}{q_k}\right)\right)^{\alpha_j\beta_k} \cdots = \cdots \left((-1)^{\frac{p_j-1}{2}\frac{q_k-1}{2}}\right)^{\alpha_j\beta_k} \cdots$$

if $\gcd(m,n) = 1$, by the Law of Quadratic Reciprocity. Now $(p_j - 1)^m \equiv 0 \pmod 4$ for odd primes p_j and integers $m \geq 2$, so

$$\begin{aligned} p_1^{\alpha_1} \cdots p_r^{\alpha_r} &\equiv (1 + (p_1 - 1))^{\alpha_1} \cdots (1 + (p_r - 1))^{\alpha_r} \\ &\equiv 1 + (p_1 - 1)\alpha_1 + \cdots + (p_r - 1)\alpha_r \pmod 4 \end{aligned}$$

by the Binomial Theorem. Since this also holds for the q_k and β_k we obtain

$$\frac{(m-1)(n-1)}{4} \equiv \frac{1}{4}\left(\sum_{j=1}^{r}(p_j - 1)\alpha_j\right)\left(\sum_{k=1}^{s}(q_k - 1)\beta_k\right) \pmod 2$$

and thus

$$\left(\frac{n}{m}\right)\left(\frac{m}{n}\right) = (-1)^{\frac{m-1}{2}\frac{n-1}{2}}$$

if m and n are odd coprime integers. This relation generalizes the Law of Quadratic Reciprocity to Jacobi symbols. The supplementary laws also generalize to odd positive moduli by similar calculations.

If n is an integer and m an odd positive integer with $(n|m) = -1$, there must be some prime $p|m$ with $(n|p) = -1$, so n is a quadratic nonresidue modulo p. But then n is also a quadratic nonresidue modulo m. There is no implication in the other direction, as we see by an example. The calculation

$$\left(\frac{3}{7}\right) = \left(\frac{3}{7}\right)\left(\frac{1}{3}\right) = \left(\frac{3}{7}\right)\left(\frac{7}{3}\right) = -1,$$

shows that 3 is a quadratic nonresidue modulo 7. Clearly 3 is also a quadratic nonresidue modulo 7^2, but $(3|7^2) = (3|7)^2 = 1$. Thus both quadratic residues and nonresidues may have a positive Jacobi symbol.

It is not possible to carry over both the relationship with quadratic residuacy and the reciprocity law when generalizing the Legendre symbol to general odd positive moduli. One or the other has to be sacrificed, and it is generally felt that it is most useful to keep the reciprocity law.

An integer D is called a *fundamental discriminant* if either D is squarefree and $D \equiv 1 \pmod 4$ or else $4|D$ with $D/4$ squarefree and $D/4 \equiv 2$ or $3 \pmod 4$. This terminology comes from algebraic number theory. When D is a fundamental discriminant, and only then, we shall generalize the Jacobi symbol further, to the Kronecker symbol $(D|m)$ for m a positive integer that may be even. Note that we do not require D and m to be coprime in the Kronecker symbol.

The Kronecker symbol is defined by

$$\left(\frac{D}{m}\right) = \left(\frac{D}{p_1}\right)^{\alpha_1} \cdots \left(\frac{D}{p_r}\right)^{\alpha_r}$$

if $m = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$ is a positive integer. Here $(D|p_j)$ is the Legendre symbol if p_j is an odd prime that does not divide D. We set $(D|p_j) = 0$ if $p_j|D$, and

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

if $p_j = 2$.

The multiplicativity property $(D|mn) = (D|m)(D|n)$ is a direct consequence of the definition. It is also clear that $(D|m) = 0$ whenever $\gcd(D, m) \geq 2$.

The fundamental discriminant D is a period of the Kronecker symbol. Note that if $\gcd(D, m) \geq 2$, then $\gcd(D, m + D) \geq 2$, so we may suppose D and m coprime. Write $m = 2^\alpha m'$ with m' odd if $D \equiv 1 \pmod 4$. Then

$$\begin{aligned}
\left(\frac{D}{m}\right) &= \left(\frac{D}{2}\right)^{\alpha} \left(\frac{D/|D|}{m'}\right) \left(\frac{|D|}{m'}\right) \\
&= (-1)^{\alpha \frac{D^2-1}{8}} \left(\frac{\operatorname{sgn}(D)}{m'}\right) \left(\frac{m'}{|D|}\right) (-1)^{\frac{m'-1}{2}\frac{|D|-1}{2}} \\
&= \left(\frac{2}{|D|}\right)^{\alpha} \left(\frac{m'}{|D|}\right) \left(\frac{\operatorname{sgn}(D)}{m'}\right) \left(\frac{-1}{m'}\right)^{\frac{|D|-1}{2}} \\
&= \left(\frac{m}{|D|}\right) \left(\frac{\operatorname{sgn}(D)(-1)^{\frac{|D|-1}{2}}}{m'}\right) = \left(\frac{m}{|D|}\right)
\end{aligned}$$

by quadratic reciprocity, the supplementary laws for the Jacobi symbol, and

$$\operatorname{sgn}(D)(-1)^{\frac{|D|-1}{2}} = (-1)^{\frac{D-1}{2}}.$$

If $D = 4D'$ with $D' \equiv 3 \pmod 4$ then

$$\left(\frac{D}{m}\right) = \left(\frac{m}{|D'|}\right) \left(\frac{-1}{m}\right)$$

for m odd by the same kind of calculation. The second factor is periodic with period 4. If $D = 8D'$ with $D' \equiv 1 \pmod 2$ then

$$\left(\frac{D}{m}\right) = \left(\frac{m}{|D'|}\right) \left(\frac{2(-1)^{\frac{D'-1}{2}}}{m}\right)$$

The second factor is one of $(\pm 2|m)$ and is therefore periodic with period 8. So D is a period of $(D|m)$ in all three cases.

Sums of two squares

A positive integer b is a *sum of two squares* if there exists integers k and m, not necessarily positive, such that $b = k^2 + m^2$. The Brahmagupta-Fibonacci identity

$$(x^2 + y^2)(u^2 + v^2) = (xu + yv)^2 + (yu - xv)^2$$

shows that the product of two sums of two squares is itself a sum of two squares. Since any square is congruent to 0 or 1 modulo 4, it is clear that no prime congruent to 3 modulo 4 can be a sum of two squares. The prime $2 = 1^2 + 1^2$ is a sum of two squares, which leaves the primes congruent to 1 modulo 4 to be accounted for.

Fermat's two squares theorem. *Every prime p congruent to 1 modulo 4 is a sum of two squares.*

Proof. The involution $(x, y, z) \overset{\mathrm{f}}{\mapsto} (-x, z, y)$ on the finite set

$$S = \{(x, y, z) \in \mathbb{Z}^3 \mid x^2 + 4yz = p\,,\ y > 0\,,\ z > 0\}$$

has no fixpoint since $x = 0$ is impossible. In particular $f(T) = S \setminus T$ where

$$T = \{(x, y, z) \in S \mid x > 0\}.$$

Furthermore $f(U) = S \setminus U$ where

$$U = \{(x, y, z) \in S \mid x - y + z > 0\}.$$

For there are no elements in S with $x - y + z = 0$ since this would give $p = x^2 + 4yz = (y - z)^2 + 4yz = (y + z)^2$. Now f is a bijection and so

$$f(T \setminus U) = f(T) \setminus f(U) = (S \setminus T) \setminus (S \setminus U) = U \setminus T$$

shows that $T \setminus U$ and $U \setminus T$ have the same number of elements. Since $T = (T \setminus U) \cup (T \cap U)$ and $U = (U \setminus T) \cup (T \cap U)$ it follows that T and U have the same number of elements.

The mapping $(x, y, z) \mapsto (2y - x, y, x - y + z)$ is an involution on U, since $(2y - x)^2 + 4y(x - y + z) = 4y^2 - 4xy + x^2 + 4xy - 4y^2 + 4yz = x^2 + 4yz = p$. The condition for (x, y, z) to be a fixpoint of this involution is that $x = y$. This gives $x^2 + 4xz = p$, and since p is a prime, $x = \pm 1, \pm p$ are the only possibilities. Now $x = \pm p$ is excluded because $x^2 + 4yz = p$ with $y > 0$ and $z > 0$. If $x = -1$ then $y = -1$ and $z = (1 - p)/4$, and this contradicts the condition $x - y + z > 0$. This leaves $(x, y, z) = (1, 1, (p - 1)/4)$, which is a fixpoint because $(p - 1)/4$ is an integer. So the involution has precisely one fixpoint, thus the number of elements of U is odd, hence the number of elements of T is also odd.

The mapping $(x, y, z) \mapsto (x, z, y)$ is an involution on T. Since the number of elements of T is odd, this involution has an odd number of fixpoints, thus at least one, so there exists some element $(x, y, z) \in T$ with $y = z$. But then $p = x^2 + 4yz = x^2 + (2z)^2$ so p is a sum of two squares. □

The two-squares theorem is one of the most famous in number theory. It was first stated by A. Girard in 1625, and Fermat claimed a proof in a letter to M. Mersenne dated Christmas Day 1640. In a letter to P. de Carcavi he stated that he obtained the proof by means of his method of infinite descent. The first published proof is due to Euler, also using descent. The above proof is due to Heath-Brown.

Since $p^2 = p^2 + 0^2$ for the primes p congruent to 3 modulo 4, all products of powers of 2, powers of primes congruent to 1 modulo 4 and powers of squares p^2 of primes p congruent to 3 modulo 4 are sums of two squares. On the other hand, if a sum of two squares b is divisible by some odd prime p to an odd power, then p is congruent to 1 modulo 4. For if $u^2 + v^2 = p^k m$ with p an odd prime, k odd and p prime to m, we may assume that p is prime to uv. Otherwise p^2 is a common factor of u^2 and v^2 and may be divided out, repeatedly if necessary, leaving a relation of the same form with a smaller exponent k that is still an odd number, and with p prime to uv. Then v has a multiplicative inverse $\overline{v}$ modulo p, and $(u\overline{v})^2$ is congruent to -1 modulo p. Now

$$(-1)^{(p-1)/2} \equiv ((u\overline{v})^2)^{(p-1)/2} \equiv (u\overline{v})^{p-1} \equiv 1 \pmod{p}$$

by the little theorem of Fermat, and so $(p-1)/2$ must be even. At this point we have obtained a characterization in multiplicative terms: The sums of two squares are multiplicatively generated by 2, the primes $p \equiv 1 \pmod 4$, and the squares p^2 of primes $p \equiv 3 \pmod 4$.

We shall now determine the number of representations $r(n)$ of a positive integer n as a sum of two squares, by investigating the ring of Gaussian integers $\mathbb{Z}[i] = \mathbb{Z} + \mathbb{Z}i$ where $i^2 = -1$. Observing the factorization

$$k^2 + m^2 = (k + mi)(k - mi),$$

we define the *norm*

$$\mathrm{N}(k + mi) = (k + mi)\overline{(k + mi)} = k^2 + m^2$$

of a Gaussian integer, and note that this coincides with the gauge that we used to show that the ring of Gaussian integers is a Euclidean domain.

Denoting Gaussian integers by Greek letters, we have

$$\mathrm{N}(\alpha\beta) = \alpha\beta\overline{\alpha\beta} = \alpha\overline{\alpha}\beta\overline{\beta} = \mathrm{N}(\alpha)\mathrm{N}(\beta)$$

so that $\alpha|\beta$ implies $\mathrm{N}(\alpha)|\mathrm{N}(\beta)$. In particular $\mathrm{N}(\upsilon) = \pm 1$ if υ is a unit. But if $\mathrm{N}(\upsilon) = \pm 1$, then $\upsilon\overline{(\pm\upsilon)} = 1$ so that υ is a unit. Solving $k^2 + m^2 = \mathrm{N}(k + mi) = \pm 1$ in rational integers k and m, we obtain the units $\upsilon = \pm 1, \pm i$.

Every Gaussian integer α divides some rational integer by $\alpha|\alpha\overline{\alpha}$, so in particular every Gaussian prime π divides some rational integer, say $\pi|n$ with $n = p_1^{a_1} \cdots p_r^{a_r}$. Since the rational integers are contained among the Gaussian integers, and the ring of Gaussian integers is a UFD, we conclude that there is some rational prime p with $\pi|p$. If q is another rational prime, then there are integers x and y for which $px + qy = 1$, and thus $\pi|q$ would imply that $\pi|1$, which

is impossible. So every Gaussian prime is found as a divisor of some unique rational prime. If a nonzero Gaussian integer α has a factorization $\alpha = \beta\gamma$ into Gaussian integers β and γ, neither of which is a unit, then $\mathrm{N}(\beta) = \mathrm{N}(\beta)\mathrm{N}(\gamma)$, where neither $\mathrm{N}(\beta)$ nor $\mathrm{N}(\gamma)$ is a unit in $\mathbb{Z}$. So if α is a composite Gaussian integer, $\mathrm{N}(\alpha)$ is a composite rational integer, and $\mathrm{N}(\alpha)$ prime in $\mathbb{Z}$ is a sufficient condition for α to be prime in the Gaussian integers.

We have a factorization $2 = (1+i)(1-i) = (-i)(1+i)^2$, and $\mathrm{N}(1+i) = 2$ is prime, so $1+i$ is a Gaussian prime, and the only one up to associates that divides 2. We say that 2 *ramifies* in the Gaussian integers since its factorization into Gaussian primes has a repeated factor.

Any prime $p \equiv 1 \pmod 4$ has a representation $p = k^2 + m^2$ as a sum of two squares, and hence a factorization $p = (k+mi)(k-mi)$ with $\mathrm{N}(k+mi) = N(k-mi) = k^2+m^2 = p$ prime, so $k+mi$ and $k-mi$ are Gaussian primes, and the only ones up to associates that divide p. These two primes are not associates, so every rational prime $p \equiv 1 \pmod 4$ *splits* into two distinct Gaussian primes. But if π is a Gaussian prime that divides a rational prime $p \equiv 3 \pmod 4$, then $\alpha\pi = p$ with some Gaussian integer α shows that $\mathrm{N}(\alpha)\mathrm{N}(\pi) = p^2$, so $\mathrm{N}(\pi) = 1$ or $\mathrm{N}(\pi) = p$ or $\mathrm{N}(\pi) = p^2$. The first possibility is excluded since π is not a unit, and the second is impossible since $p \equiv 3 \pmod 4$ is not a sum of two squares. The third possibility implies that α is a unit and thus π and p are associates. We conclude that the rational primes $p \equiv 3 \pmod 4$ remain prime in the ring of Gaussian integers, and say that they stay *inert* when passing from the rational to the Gaussian integers.

We have used Fermat's two squares theorem, as well as the fact that $\mathbb{Z}[i]$ is a UFD, in determining the Gaussian primes. By means of an argument of Dedekind, we may dispense with the two squares theorem while leaning more heavily on the fact that $\mathbb{Z}[i]$ is a UFD. This yields an independent proof of the two squares theorem.

If $p \equiv 1 \pmod 4$ is a prime, then the number of integers $1, 2, \ldots, (p-1)/2$ is even, and these integers are pairwise congruent to the integers $(p-1)/2 + 1, (p-1)/2 + 2, \ldots, p-1$ modulo p. Thus

$$-1 \equiv (p-1)! \equiv \left(\frac{p-1}{2}\right)!(-1)^{(p-1)/2}\left(\frac{p-1}{2}\right)! \equiv \left[\left(\frac{p-1}{2}\right)!\right]^2 \pmod p$$

by Wilson's theorem, so $p|(k^2+1)$ for some integer k. Factoring $k^2 + 1 = (k+i)(k-i)$ we see that p cannot be a Gaussian prime, for p would have to divide $k+i$ or $k-i$ since $\mathbb{Z}[i]$ is a UFD, but

$$\frac{k}{p} \pm \frac{1}{p}i$$

are not Gaussian integers. Thus $p = \alpha\beta$ with α and β Gaussian integers, neither of them a unit. And then $p^2 = \mathrm{N}(p) = \mathrm{N}(\alpha)\mathrm{N}(\beta)$ so $p = \mathrm{N}(\alpha) = \alpha\overline{\alpha}$. Furthermore this factorization exhibits p as a sum of two squares, and the Fermat two squares theorem is proved again.

To determine the number $r(n)$ of representations of n as a sum of two squares, it is enough to count the Gaussian divisors $\delta|n$ that satisfy $\delta\overline{\delta} = n$. We write

$$n = 2^{a_0} p_1^{a_1} \cdots p_r^{a_r} q_1^{c_1} \cdots q_s^{c_s}$$

as a product of prime powers, where a_0 is a nonnegative integer, the primes $p_1, \ldots, p_r$ are congruent to 1 modulo 4 and $a_1, \ldots, a_r$ are positive integers, and the primes q_k are congruent to 3 modulo 4 and $c_1, \ldots, c_s$ are positive integers. Either one or both of the nonnegative integers r and s may be zero. The above factorization over the rational integers yields a factorization

$$n = (1+i)^{a_0}(1-i)^{a_0}\pi_1^{a_1}\overline{\pi}_1^{a_1}\cdots\pi_r^{a_r}\overline{\pi}_r^{a_r} q_1^{c_1}\cdots q_s^{c_s}$$

over the Gaussian primes. Here $\pi_j, \overline{\pi}_j$ are distinct Gaussian primes into which p_j splits. The general Gaussian divisor $\delta|n$ is of the form

$$\delta = \upsilon(1+i)^{a_0'}(1-i)^{a_0''}\pi_1^{a_1'}\overline{\pi}_1^{a_1''}\cdots\pi_r^{a_r'}\overline{\pi}_r^{a_r''} q_1^{c_1'}\cdots q_s^{c_s'}$$

where υ is a unit, a_j' and a_j'' are nonnegative integers with $a_j' \le a_j$ and $a_j'' \le a_j$, and c_j' and c_j'' are nonnegative integers with $c_j' \le c_j$ and $c_j'' \le c_j$. Since $1+i$ and $1-i$ are associates, only the value of $a_0' + a_0''$ is significant for the factorization of δ up to associates. In determining the form of the general divisor, we rely on the fact that $\mathbb{Z}[i]$ is a UFD.

We now count Gaussian divisors $\delta|n$ for which the complementary divisor n/δ equals the conjugate $\overline{\delta}$. Since $\upsilon\overline{\upsilon} = 1$ for each unit ε, there are four possibilities for υ. We must have $a_0' = a_0$, so there is only 1 possibility for a_0'. Seeing that $a_j - a_j' = a_j''$ and $a_j - a_j'' = a_j'$ hold by comparing exponents, there are $a_j + 1$ possibilities for the pair a_j', a_j''. Similarly $c_1 - c_k' = c_k'$ must hold, so there is 1 possibility for c_k if $c_k = 2b_k$ is even, and none otherwise. Thus we obtain a formula

$$r(n) = 4(a_1+1)\cdots(a_r+1)\cdot\begin{cases} 1 & \text{if } c_1 \equiv \cdots \equiv c_s \equiv 0 \pmod 2 \\ 0 & \text{otherwise} \end{cases}$$

for the number of representations of n as a sum of two squares.

The formula shows that $r(n)/4$ is a multiplicative function, and that $r(2^a)/4 = 1$ while $r(p^a)/4 = a+1$ if $p \equiv 1 \pmod 4$ and $r(q^c)/4 = 1-1+\cdots+(-1)^c$ if $p \equiv 3 \pmod 4$. The arithmetic function $\chi(n)$ that equals zero on the even integers, 1 on the integers congruent to 1 modulo 4, and -1 on the integers congruent to 3 modulo 4, is multiplicative. This is in fact the unique nonprincipal Dirichlet character modulo 4. It has the property that

$$\sum_{d|2^a}\chi(d) = 1 \quad , \quad \sum_{d|p^a}\chi(d) = a+1 \quad , \quad \sum_{d|q^c}\chi(d) = 1-1+\cdots+(-1)^c$$

if $p \equiv 1 \pmod 4$ and $q \equiv 3 \pmod 4$. This yields Jacobi's formula

$$r(n) = 4\sum_{d|n}\chi(d) = 4\sum_{\substack{d|n \\ d\equiv 1(4)}} 1 - 4\sum_{\substack{d|n \\ d\equiv 3(4)}} 1,$$

or $r = 4 * \chi$, which he deduced from his theory of theta functions.

Number fields

A *field extension* K/k consists of a field K together with some choice of subfield k of K. The dimension of K considered as a vector space over k is called the *degree* of K/k and is denoted by $n_{K/k}$, or by n_K if $k = \mathbb{Q}$. Another notation $[K : k]$ for the degree is also in common use. An extension of finite degree is just called a *finite extension.* Degree is multiplicative in towers $L/K/k$ of extensions, i. e. $n_{L/k} = n_{L/K} \cdot n_{K/k}$. This is clear if L/K or K/k is of infinite degree. If K/k and L/K are finite extensions, then K has a basis ω_i over k of $n_{K/k}$ elements and L has a basis ψ_j over K of $n_{L/K}$ elements. Then $\omega_i\psi_j$ is a basis of L over k of $n_{L/K} \cdot n_{K/k}$ elements.

Suppose that K/k is a field extension, and $A \subseteq K$ an arbitrary set. Then the extension of k by A is the intersection field

$$k(A) = \bigcap F$$

of all fields F with $K/F/k$ and $A \subseteq F$. An extension of k by a single element α is called a *simple extension* and is denoted by $k(\alpha)$. If M/k is a field extension and $M/K/k$ and $M/L/k$ are towers of extensions, then the field $KL = k(K \cup L)$ is called the *compositum* of K and L.

An element β of an extension K/k is *algebraic* over k if β is a root of a nonconstant polynomial in $k[x]$. Complex numbers algebraic over $\mathbb{Q}$ are called *algebraic numbers.* The polynomials $p(x) \in k[x]$ vanishing on β form an ideal I. Since $k[x]$ is a PID, this ideal is generated by a polynomial $m(x)$ over k which we may require to be monic. It is then unique, by consideration of the difference of two such polynomials. It is termed the *minimal polynomial* of β over k, because it is the monic polynomial over k of minimal degree that vanishes on β. The minimal polynomial is irreducible over k, for otherwise one of its factors would vanish on β but not be in $I = (m(x))$ by consideration of degrees, a contradiction. For $\beta \neq 0$ the equation $m(\beta) = 0$ may be rewritten as

$$\frac{1}{\beta} = -\frac{1}{a_n}\beta^{n-1} - \frac{a_1}{a_n}\beta^{n-2} - \cdots - \frac{a_{n-1}}{a_n},$$

where $m(x) = x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n$ is the minimal polynomial of β over k. Note that $a_n \neq 0$ because $m(x)$ is irreducible and $\beta \neq 0$. Thus $1/\beta \in k[\beta]$ for every $\beta \neq 0$ that is algebraic over k. Clearly $k[\beta] \subseteq k[\alpha]$ if $\beta \in k[\alpha]$, so $k[\alpha]$ is a field if α is algebraic over k. But by its very definition, $k(\alpha)$ does not contain any proper subfield containing both k and α, and so $k[\alpha] = k(\alpha)$. The elements $1, \alpha, \ldots, \alpha^{\deg(m(x))-1}$ span $k(\alpha)$ over k by $m(\alpha) = 0$, while these elements are linearly independent over k by the minimal degree property of $m(x)$. Thus $n_{k(\alpha)/k} = \deg(m(x))$.

A *number field* is a finite extension of $\mathbb{Q}$. The arithmetic of number fields is the central concern of algebraic number theory. Clearly any finite extension K/k of a number field is itself a number field, and this more general *relative case* is also of great interest. We make repeated use of two properties of number fields that follow from the fact that they contain $\mathbb{Q}$: They have infinitely many

elements and $1 + 1 + \cdots + 1 \neq 0$ for any positive number of terms in the sum. Neither of these properties hold in all fields, and fields with the latter property are said to have *characteristic zero.*

For every finite extension K/k of a number field of degree $n = n_{K/k}$ there exists some element $\alpha \in K$ such that $\alpha, \ldots, \alpha^n$ is a basis for K over k. For let $\omega_1, \cdots, \omega_n$ be a basis for K over k and set $\alpha = x\omega_1 + \cdots + x^n\omega_n$. We shall choose $x \in k$ so that $\alpha, \ldots, \alpha^n$ are linearly independent over k. Assume $\gamma_1\alpha + \cdots + \gamma_n\alpha^n = 0$ with $\gamma_1, \ldots, \gamma_n \in k$. Each product of powers of basis elements $\omega_1, \ldots, \omega_n$ is a k-linear combination of these basis elements. Thus

$$\begin{aligned}\gamma_1\alpha + \cdots + \gamma_n\alpha^n &= \gamma_1(x\omega_1 + \cdots + x^n\omega_n) + \cdots + \gamma_n(x\omega_1 + \cdots + x^n\omega_n)^n \\ &= (\gamma_1 x + \cdots)\omega_1 + \cdots = p_1(x)\omega_1 + \cdots\end{aligned}$$

where $p_1(x) \in k[x]$ is a polynomial whose lowest degree term is $\gamma_1 x$, since the characteristic of k is zero. This polynomial must be nonzero for all but finitely many $x \in k$ unless $\gamma_1 = 0$. We are not going to choose x among these finitely many exceptional values, and conclude that $\gamma_1 = 0$. Now ω_1^2 is a k-linear combination of the basis elements, with at least one coefficient different from zero. Say that $\omega_1^2 = \beta_j\omega_j + \cdots$ with $\beta_j \neq 0$. Then

$$0{\cdot}\omega_1 + \gamma_2\omega_2 + \cdots + \gamma_n\omega_n = (\gamma_2\beta_j x^2 + \cdots)\omega_j + \cdots = p_2(x)\omega_j + \cdots$$

where $p_2(x) \in k[x]$ is a polynomial whose lowest degree term is $\gamma_2\beta_j x^2$. This polynomial must be nonzero for all but finitely many $x \in k$ unless $\gamma_2 = 0$. We are not going to choose x among these finitely many exceptional values, and conclude that $\gamma_2 = 0$. Note that the exceptional values of x in this round may be different from those in the last round. Continuing like this we obtain $\gamma_1 = \cdots = \gamma_n = 0$ unless x lies in a finite set of exceptional values. But k is infinite, and we choose $x \in k$ outside the finite set of exceptional values.

From now on we make use of the Fundamental Theorem of Algebra, by adopting the standing assumption that all our number fields are subfields of $\mathbb{C}$. Nothing is lost thereby, for every number field is isomorphic to a subfield of $\mathbb{C}$. The result that every nonconstant complex polynomial factors into complex linear factors is easy to prove by simple complex analysis, and enables us to avoid the concept of algebraic closure.

The minimal polynomial $m(x)$ of an element α algebraic over a number field k has simple roots. For $m(x)$ and $m'(x)$ have no nontrivial common factor in $k[x]$, since $m(x)$ is irreducible over k and $m'(x)$ is not the zero polynomial. The roots of the minimal polynomial of α over k are called the *algebraic conjugates* of α over k and are denoted by $\alpha^{(1)} = \alpha, \alpha^{(2)}, \ldots, \alpha^{(n)}$ where $n = n_{k(\alpha)/k} = \deg(m(x))$. We have already seen that any finite extension K/k of a number field is a simple extension $K = k(\alpha)$. If $m(x)$ is the minimal polynomial of α over k, the quotient ring $k[x]/(m(x))$ is a field and the isomorphism $\varphi : k[x]/(m(x)) \to k[\alpha] = k(\alpha) = K$ given by $p(x)+(m(x)) \mapsto p(\alpha)$ shows that K/k may be specified up to isomorphism simply by prescribing a suitable irreducible polynomial over k.

In general a single finite extension of a number field $k \subset \mathbb{C}$ may be realized inside $\mathbb{C}$ in several different ways. Define monomorphisms $f_j : k[x]/(m(x)) \to \mathbb{C}$ by $p(x) + (m(x)) \mapsto p(\alpha^{(j)})$ for $1 \leq j \leq n = n_{K/k} = \deg(m(x))$. The n monomorphisms $\sigma_j : K \to \mathbb{C}$ given by $\sigma_j = f_j \circ \varphi^{-1}$ are called the *embeddings* of K/k into $\mathbb{C}$. Since $\sigma_j(a) = f_j(\varphi^{-1}(a)) = f_j(a + (m(x))) = a$ for $a \in k$, the embeddings fix k pointwise. Suppose that $\sigma : K \to \mathbb{C}$ is a monomorphism that fixes k pointwise, and put $\beta = \varphi(x+(m(x)))$. Then $0 = (\sigma\circ\varphi)(m(x)+(m(x))) = \sigma(m(\beta)) = m(\sigma(\beta))$ and so $\sigma(\beta)$ is one of the conjugates $\alpha^{(1)}, \ldots, \alpha^{(n)}$. The calculation $(\sigma \circ \varphi)(p(x)+(m(x))) = \sigma(p(\beta)) = p(\sigma(\beta))$ shows that $\sigma \circ \varphi$ is equal to one of the f_j.

It is time to pull together the information that we have gathered. Every finite extension K/k of a number field $k \subset \mathbb{C}$ comes with $n_{K/k}$ distinct embeddings $\sigma_j : K \to \mathbb{C}$ that fix k pointwise. Their images are subfields of $\mathbb{C}$, called *conjugate fields,* each of which is isomorphic to K over k. Though the embeddings are distinct, their images may not be distinct. The number field $\mathbb{Q}(\sqrt{2})$ has two distinct embeddings $a + b\sqrt{2} \mapsto a + b\sqrt{2}$ and $a + b\sqrt{2} \mapsto a - b\sqrt{2}$ into $\mathbb{C}$, but their images are the same field $\mathbb{Q}(\sqrt{2})$. For us the most important case of embeddings will be embeddings of a number field $K/\mathbb{Q}$.

For a tower $L/K/k$ of extensions of number fields, every embedding $\sigma : K \to \mathbb{C}$ over k extends to $n_{L/K}$ embeddings $\tilde{\sigma} : L \to \mathbb{C}$ over k. There is some element $\alpha \in L$ with $L = K(\alpha)$ and α has a minimal polynomial $m(x)$ over K. Extend σ to a monomorphism from $K[x]$ into $\mathbb{C}[x]$ by $\sigma(x) = x$ and define $\tilde{\sigma} = f \circ \varphi^{-1}$ where $f : K[x]/(m(x)) \to \mathbb{C}$ is given by $p(x) + ((m(x)) \mapsto \sigma(p)(\alpha)$ and $\varphi : K[x]/(m(x)) \to L$ is given by $p(x) + (m(x)) \mapsto p(\alpha)$. Then $\tilde{\sigma}(c) = f(c + (m(x))) = \sigma(c)(\alpha) = \sigma(c)$ for $c \in K$. Thus each embedding of K over k has an extension to an embedding of L over k. Composing such an embedding with the $n_{L/K}$ embeddings of L over K, we obtain $n_{K/L}$ distinct embeddings of L over k extending the single embedding of K over k, since the embeddings over K equal the identity on K. Since there are $n_{K/k}$ embeddings of K over k, we have accounted for the $n_{L/k} = n_{L/K} \cdot n_{K/k}$ embeddings of L over k, so there are exactly $n_{L/K}$ embeddings of L over k extending an embedding of K over k.

A number field $K = \mathbb{Q}(\alpha)$ is said to be *totally real* if the image of every embedding of K into $\mathbb{C}$ lies in $\mathbb{R}$. It is clearly equivalent that α and all its conjugates be real. The field $\mathbb{Q}(\sqrt{2})$ is totally real, while the field $\mathbb{Q}(\sqrt[3]{2})$ is contained in $\mathbb{R}$ but not totally real, since the minimal polynomial $m(x) = x^3 - 2$ has complex roots. For a finite extension K/k of a totally real number field, it makes sense to distinguish among the embeddings of K according to whether their images are contained in $\mathbb{R}$ or not. There will be r_1 real embeddings and an even number $2r_2$ of non-real, commonly called complex, embeddings that come in complex conjugate pairs $\sigma_j, \overline{\sigma_j}$. Clearly $r_1 + 2r_2 = n_{K/k}$. The extension is called *totally real* if $r_2 = 0$ and *totally complex* if $r_1 = 0$. In the latter case the degree of K/k must be even. As examples, we note that $r_1 = 2$ and $r_2 = 0$ for $\mathbb{Q}(\sqrt{2})$, that $r_1 = 0$ and $r_2 = 1$ for $\mathbb{Q}(\sqrt{-1})$, and that $r_1 = r_2 = 1$ for $\mathbb{Q}(\sqrt[3]{2})$. It will be convenient to order the embeddings so that the first r_1 ones are real, and the last $2r_2$ ones are complex and listed in such a way that $\sigma_{j+r_2} = \overline{\sigma_j}$ for $r_1 + 1 \leq j \leq r_1 + r_2$.

Consider a finite extension K/k of a number field, and the $n = n_{K/k}$ embeddings $\sigma_j : K \to \mathbb{C}$ over k. As usual we regard all fields involved as subfields of $\mathbb{C}$, as we may by the Fundamental Theorem of Algebra. There is at least one embedding σ_j with $\sigma_j(K) = K$, namely the identity map. But there may well be other such embeddings, called *automorphisms* of K over k. In any case, the embeddings with this property form a group, termed the automorphism group of the extension. This group had its origin in the theory of polynomial equations, but it also has great significance for the arithmetic of number fields, and it is from this angle that we shall view it.

The extension $\mathbb{Q}(\sqrt{2}, \sqrt{-3})/\mathbb{Q}$ has four embeddings generated by the assignments $\sqrt{2} \mapsto \pm\sqrt{2}$ and $\sqrt{-3} \mapsto \pm\sqrt{-3}$ of conjugates of the generators of the extension, all of which preserve $\mathbb{Q}(\sqrt{2}, \sqrt{-3})$. There are no more, for the degree of the extension is 4 by the multiplicativity of degrees in the tower $\mathbb{Q}(\sqrt{2}, \sqrt{-3})/\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ and the fact that $\sqrt{-3} \notin \mathbb{Q}(\sqrt{2})$. Thus the automorphism group is isomorphic to $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$. This group is called the *Klein Viergruppe,* and has no element of order greater than two. The fifth root of unity $\zeta_5 = \exp(2\pi i/5)$ generates a cyclotomic extension (extension by roots of unity) $\mathbb{Q}(\zeta_5)/\mathbb{Q}$ and the assignments $\zeta_5 \mapsto \zeta_5^j$ for $j = 1, \ldots, 4$ yield four distinct embeddings, all of which preserve $\mathbb{Q}(\zeta_5)$. There are no more, for the minimal polynomial of ζ_5 is the fourth degree polynomial $m(x) = x^4 + x^3 + x^2 + x + 1$. Since the embedding generated by the assignment $\zeta_5 \mapsto \zeta_5^2$ is a generator for the automorphism group, we see that it is isomorphic to $\mathbb{Z}/4\mathbb{Z}$, which has elements of order four. Two extensions K/k and L/k are said to be *isomorphic* if there is an isomorphism $\varphi : K \to L$ with $\varphi|_k = \mathrm{id}|_k$. Note that the extensions $\mathbb{Q}(\sqrt{2}, \sqrt{-3})/\mathbb{Q}$ and $\mathbb{Q}(\zeta_5)/\mathbb{Q}$ are non-isomorphic since they have non-isomorphic automorphism groups. We could not draw this conclusion merely by consideration of degrees.

An extension K/k of number fields is *normal* if all the embeddings of K over k have the same image. Normal extensions are rather special; note that if k is totally real and K/k a finite normal extension, then K/k must be either totally real or totally complex. Every extension K/k of number fields is contained in a unique minimal normal extension L/k given by

$$L = k\left(\bigcup_{j=1}^{n} \sigma_j(K)\right)$$

where $\sigma_1, \ldots, \sigma_n$ are the embeddings of K over k. The extension K/k is called the *normal closure* of K/k. Any finite extension K/k of a number field is a simple extension $K = k(\alpha)$. Clearly the extension is normal if and only if it contains all the roots of the minimal polynomial $m(x)$ of α over k. In particular the normal closure of K/k is obtained by adjoining to k all the roots of $m(x)$, and every finite normal extension K/k is obtained by adjoining all the roots of some irreducible polynomial over k. It is in fact more common to define an extension K/k to be normal if every irreducible polynomial in $k[x]$ that has a root in K splits completely into linear factors over K.

If K/k is an extension with automorphism group G and H a subgroup of G, the field

$$K^H = \bigcap_{\sigma \in H} \ker(\sigma - \mathrm{id})$$

is called the *fixed field* of H. We have a mapping $H \mapsto K^H$ taking subgroups H of the automorphism group to *intermediate fields* F with $K/F/k$. For any intermediate field F the automorphism group of K/F may be viewed as a subgroup of the automorphism group K/k, since any automorphism of K that restricts to the identity on F naturally also restricts to the identity on k. Thus we have two maps, one taking subgroups of the automorphism group to intermediate fields of the extension, and the other taking intermediate fields to subgroups.

The automorphism group of an extension K/k of number fields has special properties if the extension is normal. In this case the automorphism group of K over k is called the *Galois group* and is denoted by $\mathrm{Gal}(K/k)$.

Fundamental Theorem of Galois Theory. *Let K/k be a finite normal extension of a number field. The maps $H \mapsto K^H$ and $F \mapsto \mathrm{Gal}(K/F)$ are inverses.*

Proof. Let F be an arbitrary intermediate field of K/k and put $H = \mathrm{Gal}(K/F)$. Then K/F is a normal extension, for every embedding of K over F is an embedding of K over k, and the latter embeddings all have the same image by the normality of K/k. The extension K/K^H has $|H|$ automorphisms, since K^H is the fixed field of H. Then $n_{K/K^H} \geq |H|$, for n_{K/K^H} equals the total number of embeddings of K over K^H, and every automorphism is an embedding. But $|H| = n_{K/F}$ since K/F is normal, hence $n_{K/K^H} \geq n_{K/F}$, so $K^H = F$.

Let H be an arbitrary subgroup of $\mathrm{Gal}(K/k)$ and put $F = K^H$. Then H is a subgroup of $\mathrm{Gal}(K/F)$. There is an element $\alpha \in K$ such that $K = F(\alpha)$, and the coefficients of the polynomial

$$f(x) = \prod_{\sigma \in H} (x - \sigma(\alpha))$$

are elementary symmetric functions in the $\sigma(\alpha)$ for $\sigma \in H$. The polynomial is in $F[x]$ because these coefficients are fixed by the homomorphisms $\eta \in H$. Then

$$|H| = \deg(f) \geq n_{K/F} = |\mathrm{Gal}(K/F)|,$$

since $f(\alpha) = 0$ and K/F is normal, so $H = \mathrm{Gal}(K/F)$. □

This result is a special case of the full theorem, due to the restriction on the ground field k that it be a number field. In the general case, one must require the property of separability, which we can and do avoid here. For inseparability cannot occur for extensions of fields of characteristic zero. The mappings $H \mapsto K^H$ and $F \mapsto \mathrm{Gal}(K/F)$ are called the *Galois correspondence.* They reverse inclusions, in that if H_1 is a proper subgroup of H_2, then K^{H_2} is a proper subfield of K^{H_1}, while if F_1 is a proper subfield of F_2, then $\mathrm{Gal}(K/F_2)$ is a proper subgroup of $\mathrm{Gal}(K/F_1)$.

Often the Galois correspondence is used to draw conclusions about the intermediate fields from knowledge of the subgroups of the Galois group. For example, suppose that K/k is a normal extension of number fields. It is obvious that $\mathrm{Gal}(K/k)$ has only finitely many subgroups, immediately implying by the Galois correspondence the decidedly less obvious conclusion that K/k has only finitely many intermediate fields.

A complement to the Fundamental Theorem of Galois Theory identifies those intermediate fields of a finite normal extension K/k of a number field that are normal over the ground field: The fixed field K^H is normal over k if and only if H is a normal subgroup of $\mathrm{Gal}(K/k)$. Put $F = K^H$ and note that F is normal over k if and only if F is preserved by all embeddings of F over k. Each such embedding is the restriction of an automorphism of K over k, since K contains the conjugates over k of every element of F. Thus F/k is normal if and only if $\sigma(F) = F$ for all $\sigma \in \mathrm{Gal}(K/k)$. Now

$$\sigma(F)^{\sigma H \sigma^{-1}} = (\sigma^{-1}\sigma(F))^{\sigma H} = F^{\sigma H} = \sigma(F^H) = \sigma(F)$$

so $\sigma(F)$ is the fixed field of $\sigma H \sigma^{-1}$. Hence F/k is normal if and only if $\sigma H \sigma^{-1} = H$ for all $\sigma \in \mathrm{Gal}(K/k)$, i. e. if and only if H is a normal subgroup of $\mathrm{Gal}(K/k)$.

Given an element α of an extension K/k of number fields, the mapping $M_\alpha : K \to K$ given by $\gamma \mapsto \alpha\gamma$ is a linear operator on K as a vector space over k. The *norm* of α is defined as $N_{K/k}(\alpha) = \det(M_\alpha)$. The norm is a mapping $N_{K/k} : K \to k$ possessing the crucial multiplicativity property $N_{K/k}(\alpha\beta) = N_{K/k}(\alpha)N_{K/k}(\beta)$ because $(\alpha\beta)\gamma = \alpha(\beta\gamma)$. When $k = \mathbb{Q}$, we write N_K for the norm, and observe that it is rational.

The norm of an element α of an extension K/k of number fields can be expressed in terms of its embeddings. The characteristic polynomial $f_\alpha(x) = \det(x\,\mathrm{id} - M_\alpha)$ of M_α is monic and has coefficients in k. Note that

$$N_{K/k}(\alpha) = (-1)^{n_{K/k}} f_\alpha(0)$$

since the determinant has $n_{K/k}$ rows. The characteristic polynomial of M_α is called the *field polynomial* of α. It depends on the extension K/k and not just on α as an algebraic number. Supposing L/K to be a finite extension, α also has a field polynomial $g_\alpha(x)$ relative to L/k and it turns out that

$$g_\alpha(x) = f_\alpha(x)^{n_{L/K}}.$$

Express multiplication by α relative to some ordered basis $(\omega_1, \ldots, \omega_n)$ for K over k by

$$\alpha\omega_i = a_{i1}\omega_1 + \cdots + a_{in}\omega_n \quad , \quad 1 \le i \le n \quad , \quad n = n_{K/k}.$$

Denote the representing matrix by $A = [a_{ih}]$. Also choose an ordered basis $(\psi_1, \ldots, \psi_l)$ for L over K and express multiplication by α relative to the ordered basis $(\omega_1\psi_1, \omega_2\psi_1, \ldots, \omega_n\psi_1, \omega_1\psi_2, \ldots, \omega_n\psi_l)$ for L over k by

$$\alpha\omega_i\psi_j = a_{i1}\omega_1\psi_j + \cdots + a_{in}\omega_n\psi_j \quad , \quad 1 \le j \le l \quad , \quad l = n_{L/K}.$$

Reading off the representing matrix B, we see that B is a block matrix with l copies of A along the main diagonal, and zeros elsewhere. Thus $\det(B) = \det(A)^l$ and so g_α is the l-th power of f_α. Note that if f_α is a field polynomial for α, then $f_\alpha(\alpha) = 0$. For $(\alpha\,\mathrm{id} - M_\alpha)\gamma = \alpha\gamma - \alpha\gamma = 0$, so $x\,\mathrm{id} - M_\alpha$ is the zero operator when $x = \alpha$.

Choosing $K = k(\alpha)$, the field polynomial $f_\alpha(x)$ is a monic polynomial of degree $n_{K/k}$ with coefficients in k, and vanishes in α. The minimal polynomial $m(x)$ of α over k is also a monic polynomial of degree $n_{K/k}$ with coefficients in k, vanishing in α, and is moreover irreducible over k. But then $f_\alpha(x) = m(x)$ and

$$\begin{aligned} N_{L/k}(\alpha) &= (-1)^{n_{L/k}} g_\alpha(0) = (-1)^{n_{L/k}} m(0)^{n_{L/K}} \\ &= (-1)^{n_{L/k}}((-1)^{n_{K/k}}\alpha^{(1)}\cdots\alpha^{(n)})^{n_{L/K}} = (\alpha^{(1)}\cdots\alpha^{(n)})^l. \end{aligned}$$

Now $N_{L/k}(\alpha) = \eta_1(\alpha)\cdots\eta_{ln}(\alpha)$ where $\eta_1, \ldots, \eta_{ln}$ are the embeddings of L over k. For each embedding of K over k extends to l embeddings of L over k, and the image of α under the various embeddings of K over k are the algebraic conjugates of α over k.

For the embeddings $\sigma_1, \ldots, \sigma_n$ of K over k choose extensions $\tilde{\sigma}_1, \ldots, \tilde{\sigma}_n$ to embeddings of L over k. Denote by $\theta_1, \ldots, \theta_l$ the embeddings of L over K. Then the embeddings of L over k are $\tilde{\sigma}_1 \circ \theta_1, \ldots, \tilde{\sigma}_n \circ \theta_l$, so

$$\begin{aligned} N_{L/k}(\beta) &= (\tilde{\sigma}_1 \circ \theta_1)(\beta)\cdots(\tilde{\sigma}_n \circ \theta_l)(\beta) \\ &= \tilde{\sigma}_1(\theta_1(\beta)\cdots\theta_l(\beta))\cdots\tilde{\sigma}_n(\theta_1(\beta)\cdots\theta_l(\beta)) \\ &= \sigma_1(N_{L/K}(\beta))\cdots\sigma_n(N_{L/K}(\beta)) = N_{K/k}(N_{L/K}(\beta)) \end{aligned}$$

and thus $N_{L/k} = N_{K/k} \circ N_{L/K}$ in a tower $L/K/k$ of extensions of number fields.

Given an element α of an extension K/k of number fields, the *trace* of α is defined as $\mathrm{tr}_{K/k}(\alpha) = \mathrm{tr}(M_\alpha)$. The trace has properties somewhat analogous to those of the norm. Proceeding as above we obtain

$$\mathrm{tr}_{K/k}(\alpha) = l(\alpha^{(1)} + \cdots + \alpha^{(n)}),$$

so that $\mathrm{tr}_{L/k}(\alpha) = \eta_1(\alpha) + \cdots + \eta_{ln}(\alpha)$. Thus the trace equals the sum of the embeddings. Moreover

$$\begin{aligned} \mathrm{tr}_{L/k}(\beta) &= (\tilde{\sigma}_1 \circ \theta_1)(\beta) + \cdots + (\tilde{\sigma}_n \circ \theta_l)(\beta) \\ &= \tilde{\sigma}_1(\theta_1(\beta) + \cdots + \theta_l(\beta)) + \cdots + \tilde{\sigma}_n(\theta_1(\beta) + \cdots + \theta_l(\beta)) \\ &= \sigma_1(\mathrm{tr}_{L/K}(\beta)) + \cdots + \sigma_n(\mathrm{tr}_{L/K}(\beta)) = \mathrm{tr}_{K/k}(\mathrm{tr}_{L/K}(\beta)) \end{aligned}$$

and thus $\mathrm{tr}_{L/k} = \mathrm{tr}_{K/k} \circ \mathrm{tr}_{L/K}$ in a tower of extensions of number fields.

The expression $\mathrm{tr}_{K/k}(\alpha\beta)$ defines a nondegenerate bilinear form on K considered as a vector space over k. Bilinearity is obvious by the linearity of the trace, and if $\alpha \neq 0$ choose $\beta = \alpha^{-1}$ to see that $\mathrm{tr}_{K/k}(\alpha\beta) = \mathrm{tr}_{K/k}(1) = n_{K/k} \neq 0$.

Algebraic integers

A complex number is an *algebraic integer* if it is a root of a monic polynomial whose coefficients lie in $\mathbb{Z}$. An algebraic integer α that is a rational number is a rational integer. For if $\alpha = b/c$ with b and c coprime rational integers, the equation $(b/c)^n + a_1(b/c)^{n-1} + \cdots + a_{n-1}(b/c) + a_n = 0$ with $a_1, \ldots, a_n$ rational integers implies $b^n + a_1 b^{n-1} c + \cdots + a_{n-1} b c^{n-1} + a_n c^n = 0$. Thus $c|b^n$, which implies $c = \pm 1$ since b and c are coprime.

Suppose that α is an algebraic integer and let $f(x)$ be the unique monic polynomial of least degree with coefficients in $\mathbb{Z}$ for which $f(\alpha) = 0$. Clearly $f(x)$ is irreducible over $\mathbb{Z}$ by consideration of degree. Then it is also irreducible over $\mathbb{Q}$ by Gauss' Lemma from polynomial algebra. The minimal polynomial $m(x)$ of α over $\mathbb{Q}$ divides $f(x)$ in $\mathbb{Q}[x]$, so $f(x) = m(x)$ by the irreduciblity. We conclude that the minimal polynomial over $\mathbb{Q}$ of an algebraic integer has coefficients in $\mathbb{Z}$. This has the very important consequence that the algebraic conjugates of an algebraic integer are themselves algebraic integers. Consequently embeddings over $\mathbb{Q}$ take algebraic integers to algebraic integers.

Denote the set of algebraic integers in a number field K by $\mathcal{O}_K$. We are going to show that $\mathcal{O}_K$ is a subring of K and that K is the field of fractions of $\mathcal{O}_K$. For the first statement it will be enough to show that if α and β are algebraic integers, so is $\alpha+\beta$ and $\alpha\beta$. The abelian group of $\mathbb{Z}[\alpha]$ under addition is finitely generated since α is an algebraic integer, and similarly $\mathbb{Z}[\beta]$ is also finitely generated under addition. All powers of $\alpha+\beta$ and $\alpha\beta$ lie in $\mathbb{Z}[\alpha, \beta]$, which is finitely generated under addition because α and β commute. The rest of the argument runs in parallel in the two cases, so we proceed with $\gamma = \alpha + \beta$. The powers of γ generate an abelian group A under addition, which is a subgroup of $\mathbb{Z}[\alpha, \beta]$ under addition. A subgroup of a finitely generated abelian group is itself finitely generated, so A is finitely generated. Multiplication by γ preserves A, so γ is an eigenvalue of a matrix with integer entries, hence an algebraic integer. That K is the field of fractions of $\mathcal{O}_K$ is a consequence of the stronger statement that any element $\alpha \in K$ is of the form $\alpha = \beta/a_0$ where β is an algebraic integer in K and a_0 is a rational integer. Let $a_0\alpha^n + a_1\alpha^{n-1} + a_2\alpha^{n-2} + \cdots + a_n = 0$ be a nontrivial equation for α over $\mathbb{Z}$. Then $(a_0\alpha)^n + a_1(a_0\alpha)^{n-1} + a_2 a_0(a_0\alpha)^{n-2} + \cdots + a_n a_0^{n-1} = 0$, and thus $\beta = a_0\alpha$ is an algebraic integer.

Any number field K has a basis over $\mathbb{Q}$ consisting of algebraic integers. For if $\omega_1, \ldots, \omega_n$ is any basis, we may write $\omega_j = \psi_j/c_j$ for $1 \leq j \leq n$ with ψ_j an algebraic integer and c_j a rational integer. But then $\psi_1, \ldots, \psi_n$ is a basis consisting of algebraic integers. In fact a stronger statement holds: Any number field K has a basis $\omega_1, \ldots, \omega_n$ consisting of algebraic integers such that $\mathcal{O}_K = \mathbb{Z}\omega_1 + \cdots + \mathbb{Z}\omega_n$. Such a basis is called an *integral basis* for K. To prove the existence of an integral basis, we make use of the discriminant.

For a basis $\omega_1, \cdots, \omega_n$ of an extension K/k, the *discriminant* of the basis is

$$\Delta_{K/k}(\omega_1, \ldots, \omega_n) = \det([\sigma_i(\omega_j)]_{1 \leq i,j \leq n})^2$$

where $\sigma_1, \ldots, \sigma_n$ are the embeddings of K over k. The discriminant is independent of the ordering of the basis and of the ordering of the embeddings, because

a reordering interchanges the columns or rows of the determinant, just changing its sign, but the discriminant is the square of the determinant. Supposing A to be an invertible n by n matrix over k, and $(\omega_1, \cdots, \omega_n)$ and $(\psi_1, \cdots, \psi_n)$ ordered bases for K over k related by $[\psi_1, \cdots, \psi_n]^t = A[\omega_1, \cdots, \omega_n]^t$. Then the discriminants of the bases are related by

$$\Delta_{K/k}(\psi_1, \ldots, \psi_n) = \det(A)^2 \Delta_{K/k}(\omega_1, \ldots, \omega_n).$$

The discriminant of a basis over k is a nonzero element of k. To see this, change the basis to one of the form $1, \alpha, \ldots, \alpha^{n-1}$, for which the discriminant is the square of a necessarily nonzero Vandermonde determinant in the pairwise distinct elements $\sigma_1(\alpha), \ldots, \sigma_n(\alpha)$. This exhibits the discriminant as a symmetric polynomial in the algebraic conjugates of α over k, so the discriminant is a polynomial over $\mathbb{Q}$ in the coefficients of the minimal polynomial of α over k, and these are in k. We may express the discriminant

$$\begin{aligned}
\Delta_{K/k}(\omega_1, \ldots, \omega_n) &= \det([\sigma_i(\omega_j)]_{1 \le i,j \le n}) \det([\sigma_l(\omega_m)]_{1 \le l,m \le n}) \\
&= \det([\sigma_i(\omega_j)]_{1 \le i,j \le n} [\sigma_l(\omega_m)]_{1 \le l,m \le n}) \\
&= \det([\sigma_j(\omega_i)]_{1 \le i,j \le n} [\sigma_j(\omega_m)]_{1 \le j,m \le n}) \\
&= \det([\sigma_1(\omega_i)\sigma_1(\omega_m) + \cdots + \sigma_n(\omega_i)\sigma_n(\omega_m)]_{1 \le i,m \le n}) \\
&= \det([\sigma_1(\omega_i\omega_m) + \cdots + \sigma_n(\omega_i\omega_m)]_{1 \le i,m \le n}) \\
&= \det([\mathrm{tr}_{K/k}(\omega_i\omega_m)]_{1 \le i,m \le n})
\end{aligned}$$

of a basis $\omega_1, \ldots, \omega_n$ of an extension K/k of number fields in terms of the trace form $\mathrm{tr}_{K/k}(\alpha\beta)$. This is another way to see that the value of the discriminant is in k.

Suppose $\omega_1, \ldots, \omega_n$ is a basis for K over $\mathbb{Q}$ consisting of algebraic integers. Since embeddings take algebraic integers to algebraic integers, the discriminant $\Delta_{K/\mathbb{Q}}(\omega_1, \ldots, \omega_n)$ is the square of a sum of products of algebraic integers, thus itself an algebraic integer. But it is also a nonzero rational number, so it must be a nonzero integer. Now choose such a basis for which $|\Delta_{K/\mathbb{Q}}(\omega_1, \ldots, \omega_n)|$ is minimal. We claim that $\omega_1, \ldots, \omega_n$ is an integral basis. Let $\alpha \in \mathcal{O}_K$ be arbitrary and $\alpha = c_1\omega_1 + \cdots + c_n\omega_n$ over $\mathbb{Q}$. We must show that $c_1, \ldots, c_n \in \mathbb{Z}$, and it is clearly sufficient to consider c_1. Write $c_1 = c + \delta$ where c is a rational integer and $0 < \delta \le 1$, and define a new basis $\psi, \omega_2, \ldots, \omega_n$ for K over $\mathbb{Q}$ by $\psi = \alpha - c\omega_1$. Clearly ψ is an algebraic integer. Now

$$\begin{aligned}
|\Delta_{K/\mathbb{Q}}(\omega_1, \ldots, \omega_n)| &\le |\Delta_{K/\mathbb{Q}}(\psi, \omega_2, \ldots, \omega_n)| \\
&= |\Delta_{K/\mathbb{Q}}(\delta\omega_1 + c_2\omega_2 + \cdots + c_n\omega_n, \omega_2, \ldots, \omega_n)| \\
&= |\Delta_{K/\mathbb{Q}}(\delta\omega_1, \ldots, \omega_n)| = \delta^2 |\Delta_{K/\mathbb{Q}}(\omega_1, \ldots, \omega_n)|
\end{aligned}$$

implies that $\delta = 1$ and thus c_1 is a rational integer.

The coordinates $\mathbf{x}(\alpha)$ and $\mathbf{y}(\alpha)$ of an element $\alpha \in K$ relative to two different ordered bases, considered as column matrices, are related by $\mathbf{y}(\alpha) = A\mathbf{x}(\alpha)$ where A is a nonsingular rational matrix. If the bases are integral bases, then

multiplication by A must take arbitrary vectors with coordinates in $\mathbb{Z}$ to vectors with coordinates in $\mathbb{Z}$. Thus the entries of A are rational integers, and the same argument in the other direction shows that A^{-1} has rational integer entries. But then $\det(A)\det(A^{-1}) = \det(AA^{-1}) = \det(I) = 1$ implies that $\det(A) = \pm 1$ since $\det(A^{-1})$ is an integer. Thus the discriminants of all integral bases of K are the same. Their common value is denoted by d_K and is termed the *discriminant* of the number field K. Supposing that $\varphi : L \to K$ is an isomorphism of number fields K and L over $\mathbb{Q}$, and σ_j are the embeddings of K, then $\sigma_j \circ \varphi$ are the embeddings of L. Thus the discriminant Δ that realizes the minimum value of $|\Delta|$ over all bases consisting of algebraic integers is the same for the two fields, so $d_K = d_L$ if K and L are isomorphic over $\mathbb{Q}$. The discriminant is an invariant of number fields that can be used to distinguish them, but is also of great importance in other contexts.

The ring of algebraic integers $\mathcal{O}_K$ of a number field K is of course an integral domain, and the concepts of divisibility, units, associates, irreducibles and prime elements discussed earlier apply to $\mathcal{O}_K$. To obtain a supply of examples we determine the ring of algebraic integers in each number field of degree two. Every such field is obtained by adjoining to $\mathbb{Q}$ a root of a quadratic polynomial with coefficients in $\mathbb{Z}$. For this reason these fields are called *quadratic number fields.* The formula for the roots of a quadratic polynomial shows that such a root may be expressed rationally in terms of the square root of a rational integer. So the fields in question are of the form $\mathbb{Q}(\sqrt{d})$ with $d \in \mathbb{Z}$, and we may without loss of generality assume that $d \neq 0, 1$ and that d is squarefree. Assuming that $\alpha = a + b\sqrt{d}$ with $a, b \in \mathbb{Q}$ is an algebraic integer, only the case $b \neq 0$ is interesting. Then $\alpha^2 = a^2 + 2ab\sqrt{d} + db^2 = a^2 + 2a\alpha - 2a^2 + db^2$ so $m(x) = x^2 - 2ax + a^2 - db^2$ is the unique monic minimal polynomial of α. Thus $2a, a^2 - db^2 \in \mathbb{Z}$ by assumption. If a is an integer, then so is b since d is squarefree. If $a = c/2$ for some odd integer c, we obtain $(2a)^2 \equiv 1 \pmod 4$. Then $d(2b)^2 \equiv 1 \pmod 4$, so $2b$ must be an odd integer, for d is squarefree. Thus $(2b)^2 \equiv 1 \pmod 4$, which is equivalent to $d \equiv 1 \pmod 4$. Hence

$$\mathcal{O}_{\mathbb{Q}(\sqrt{d})} = \mathbb{Z} + \mathbb{Z}\frac{1+\sqrt{d}}{2},$$

for $2a$ and $2b$ are either both even or both odd integers. The above argument also shows that if $d \equiv 2 \pmod 4$ or $d \equiv 3 \pmod 4$, then the ring of algebraic integers of $\mathbb{Q}(\sqrt{d})$ is $\mathbb{Z} + \mathbb{Z}\sqrt{d}$, since $2a$ and $2b$ must be even integers. As $-1 \equiv 3 \pmod 4$, we see that the ring of algebraic integers in $\mathbb{Q}(\sqrt{-1})$ coincides with the familiar Gaussian integers. We now calculate the discriminants of quadratic fields. The embeddings of $\mathbb{Q}(\sqrt{d})$ are the identity and $a + b\sqrt{d} \mapsto a - b\sqrt{d}$. Thus

$$d_K = \begin{vmatrix} 1 & \frac{1+\sqrt{d}}{2} \\ 1 & \frac{1-\sqrt{d}}{2} \end{vmatrix}^2 = d \quad \text{and} \quad d_K = \begin{vmatrix} 1 & \sqrt{d} \\ 1 & -\sqrt{d} \end{vmatrix}^2 = 4d$$

for $d \equiv 1 \pmod 4$ and $d \equiv 2$ or $3 \pmod 4$ respectively. Since the discriminants are different for different values of d, distinct quadratic fields are non-isomorphic over $\mathbb{Q}$.

The discriminants of quadratic number fields are all congruent to either 0 or 1 modulo 4. This is no coincidence, for it is a result of L. Stickelberger that the discriminant of any number field K satisfies a congruence $d_K \equiv 0$ or $1 \pmod 4$. Let $\omega_1, \ldots, \omega_n$ be an integral basis for K and write $d_K = \Delta_{K/\mathbb{Q}}(\omega_1, \ldots, \omega_n) = (A+B)^2 - 4AB$ where A is the sum of the products in the determinant of $[\sigma_i(\omega_j)]_{1 \leq i,j \leq n}$ corresponding to even permutations and B is the sum of the products corresponding to odd permutations. Clearly $A+B$ and AB are algebraic integers since they are polynomials in the embeddings $\sigma_i(\omega_j)$ with rational integer coefficients. But in fact these polynomials are symmetric functions in the $\sigma_i(\omega_j)$ and thus rational numbers. For the σ_i may be extended to embeddings of the normal closure L of K and the symmetric functions of $\sigma_i(\omega_j)$ are preserved by the full Galois group of the normal extension $L/\mathbb{Q}$, so they are rational numbers by the Fundamental Theorem of Galois Theory. Since $A+B$ and AB are rational integers, the Stickelberger criterion follows.

The norm $N_K(\alpha)$ of an algebraic integer α in a number field K is a rational integer, since it is both a rational number and an algebraic integer. For it equals the product $\sigma_1(\alpha) \cdots \sigma_n(\alpha)$ where $\sigma_1, \ldots, \sigma_n$ are the embeddings of K over $\mathbb{Q}$, thus it is a product of algebraic integers. In particular the important inequality $|N_K(\alpha)| \geq 1$ holds if $\alpha \neq 0$ is an algebraic integer.

A q-th *root of unity* is any complex number ζ with $\zeta^q = 1$ for some positive integer q. If ζ is a q-th root of unity but not a q'-th root of unity for any $q' < q$, then it is a *primitive* q-th root of unity. The q-th roots of unity are $\zeta = e(m/q)$ with $1 \leq m \leq q$, and there are q of them. The primitive q-th roots of unity are given by the same formula, but with the condition $\gcd(q, m) = 1$, and there are $\phi(q)$ of those.

If all the roots of a monic polynomial in $\mathbb{Z}[x]$ lie on the unit circle, they must be roots of unity. For if $p(x)$ is such a polynomial and m a positive integer, the coefficients of the monic polynomial $q_m(x)$ whose roots are the m-th powers of the roots of $p(x)$ are symmetric functions of the roots of $p(x)$ and thus rational integers since the elementary symmetric functions of the roots of $p(x)$ are obviously rational integers. Because all the powers of the roots of $p(x)$ have absolute value 1 by assumption, there exists a constant C so that any coefficient a of any of the polynomials $q_m(x)$ for $m = 1, 2, 3, \ldots$ satisfies the bound $|a| \leq C$. So there are only finitely many such polynomials, hence there is some repetition $q_m(x) = q_n(x)$ with $m < n$. Thus the m-th power of any root ζ of $p(x)$ is equal to the n-th power of some root η of $p(x)$. Moreover there is some root η_2 with $\eta^m = {\eta_2}^n$ and so

$$\zeta^{m^2} = (\zeta^m)^m = (\eta^n)^m = (\eta^m)^n = ({\eta_2}^n)^n = {\eta_2}^{n^2}.$$

By induction there are roots η_k for $k = 1, 2, 3, \ldots$ such that $\zeta^{m^k} = {\eta_k}^{n^k}$. Since there are only finitely many roots, there is some repetition $\eta_k = \eta_l$ with $k < l$ in the sequence, and then

$$\zeta^{m^l} = {\eta_l}^{n^l} = {\eta_k}^{n^l} = ({\eta_k}^{n^k})^{n^{l-k}} = (\zeta^{m^k})^{n^{l-k}} = \zeta^{m^k n^{l-k}}$$

so ζ is a root of unity.

The field $\mathbb{Q}(\zeta_q)$ with $\zeta_q = e(1/q)$ is called a *cyclotomic field.* Note that $\mathbb{Q}(\zeta_q^m) \subseteq \mathbb{Q}(\zeta_q)$ and $\mathbb{Q}(\zeta_q) = \mathbb{Q}(\zeta_q^{qx+my}) = \mathbb{Q}((\zeta_q^m)^y) \subseteq \mathbb{Q}(\zeta_q^m)$ for some integers x and y if $\gcd(q,m) = 1$. So $\mathbb{Q}(\zeta_q^m) = \mathbb{Q}(\zeta_q)$ for $\gcd(q,m) = 1$. Put $G = \mathrm{Gal}(\mathbb{Q}(\zeta_q)/\mathbb{Q})$ and let $\sigma \in G$. If ζ is a q-th root of unity, then $\sigma(\zeta)^q = \sigma(\zeta^q) = \sigma(1) = 1$, thus $\sigma(\zeta)$ is a q-th root of unity. We conclude that the extension $\mathbb{Q}(\zeta_q)/\mathbb{Q}$ is normal. The so-called *cyclotomic polynomial*

$$\Phi_q(x) = \prod_{\substack{1 \le m \le q \\ \gcd(q,m)=1}} (x - \zeta_q^m)$$

has rational coefficients, for the coefficients are symmetric functions of the primitive q-th roots of unity, thus preserved by all $\sigma \in G$, hence they are in $\mathbb{Q}$ by the Fundamental Theorem of Galois Theory. Since any root of unity is visibly an algebraic integer, the coefficients of the polynomials $\Phi_q(x)$ are algebraic integers, hence they are rational integers. Let $f(x) \in \mathbb{Z}[x]$ be a monic polynomial with $f(\zeta_q) = 0$. Choose m with $\gcd(q,m) = 1$ and note that there exist arbitrarily large primes $p \equiv m \pmod{q}$. For any such m and p we have

$$f(\zeta_q)^p = (\zeta_q^N + a_1\zeta_q^{N-1} + \cdots + a_N)^p = (\zeta_q^p)^N + a_1^p(\zeta_q^p)^{N-1} + \cdots + a_N^p + pS_1$$

by expanding out using the multinomial theorem. Here S_1 is a sum of integer multiples of powers of ζ_q. For every multinomial coefficient different from 1 of prime order p is divisible by p. Furthermore

$$f(\zeta_q)^p = (\zeta_q^p)^N + a_1(\zeta_q^p)^{N-1} + \cdots + a_N + pS_2 = f(\zeta_q^p) + pS_2 = f(\zeta_q^m) + pS_2$$

with S_2 a sum of integer multiples of powers of ζ_q, by the small theorem of Fermat and the assumption on m and p. Now $f(\zeta_q) = 0$ implies that $f(\zeta_q^m) = -pS_2$ with S_2 an algebraic integer depending on p. Then

$$|N_K(f(\zeta_q^m))| = |N_K(-pS_2)| = p^{n_K}|N_K(S_2)| \ge p^{n_K}$$

if $f(\zeta_q^m) \neq 0$, which is impossible since the prime p can be taken arbitrarily large. So ζ_q^m is a root of $f(x)$ for every m with $\gcd(q,m) = 1$, thus $\Phi_q(x) | f(x)$ and so Φ_q is the minimal polynomial of ζ_q.

The degree of the extension $\mathbb{Q}(\zeta_q)$ is $\phi(q)$ since it equals the degree of Φ_q. This information enables us to determine the Galois group G of the cyclotomic extension. Any automorphism $\sigma \in G$ takes ζ_q to some root of Φ_q, say

$$\sigma(\zeta_q) = \zeta_q^{e(\sigma)}$$

with $\gcd(q, e(\sigma)) = 1$. Here the exponent $e(\sigma)$ is well defined modulo q so $\sigma \mapsto e(\sigma)$ gives a map from G into $(\mathbb{Z}/q\mathbb{Z})^\times$. Now

$$(\eta \circ \sigma)(\zeta_q) = \eta(\zeta_q^{e(\sigma)}) = \eta(\zeta_q)^{e(\sigma)} = \zeta_q^{e(\eta)e(\sigma)} \quad \text{and} \quad (\eta \circ \sigma)(\zeta_q) = \zeta_q^{e(\eta\circ\sigma)}$$

so this map is a homomorphism. The map is mono, for $\sigma(\zeta_q) = \zeta_q$ implies $\sigma(\zeta_q^m) = \zeta_q^m$ and thus σ is the identity. So $G \cong (\mathbb{Z}/q\mathbb{Z})^\times$ since both groups have order $\phi(q)$.

It is now possible to gain some preliminary information about units in algebraic number fields. Suppose that K is a number field, and υ a unit in K of finite order, so that $\upsilon^m = 1$. Clearly υ is a root of unity, and if ζ is a root of unity in K with $\zeta^m = 1$, then ζ^{m-1} is an algebraic integer and $\zeta \cdot \zeta^{m-1} = 1$ so ζ is a unit in K of finite order. Every root of unity ζ is a primitive root of unity to some modulus q and of these there are $\phi(q)$. The degree of $\mathbb{Q}(\zeta)$ is $\phi(q)$ and

$$\phi(q) = q \prod_{p|q} \left(1 - \frac{1}{p}\right) \geq q \prod_{p \leq q} \left(1 - \frac{1}{p}\right) \gg \frac{q}{\log(q)}$$

by a result of Mertens, so there are only a finite number of roots of unity in any number field K. Their number is an important invariant of K and is denoted by w_K. Note that $w_K = 2$ if K is totally real, since ± 1 are the only real roots of unity.

The units of finite order form a finite subgroup of the group of all units, consisting of roots of unity. This group is actually cyclic, for if $\zeta_1 = \exp(2\pi i a_1/q_1)$ and $\zeta_2 = \exp(2\pi i a_2/q_2)$ are elements, with $\gcd(a_1, q_1) = \gcd(a_2, q_2) = 1$, we can solve the linear Diophantine equation $a_1 q_2 x + a_2 q_1 y = c$ with $c = \gcd(a_1 q_2, a_1 q_2))$ in integers x and y. Then $\zeta = \zeta_1^x \zeta_2^y = \exp(2\pi i c/q_1 q_2)$ is also an element of the group, and $\zeta_1 = \zeta^{e_1}$ and $\zeta_2 = \zeta^{e_2}$ with $e_1 = a_1 q_2/c$ and $e_2 = a_2 q_1/c$. But if a finite group is not cyclic, there has to be two elements that are not both powers of some third element.

To find the units of infinite order in an algebraic number field is a more difficult problem. If υ is a unit in K, then $\upsilon\alpha = 1$ for some algebraic integer α in K, so $N_K(\upsilon)N_K(\alpha) = N_K(\upsilon\alpha) = N_K(1) = 1$. As the norms are rational integers, we see that $N_K(\upsilon) = \pm 1$ is a necessary condition for υ to be a unit. But this condition is also sufficient, for $N_K(\upsilon) = \upsilon\sigma_2(\upsilon)\cdots\sigma_n(\upsilon)$ where $\sigma_1 = \mathrm{id}, \sigma_2, \ldots, \sigma_n$ are the embeddings, and the $\sigma_j(\upsilon)$ are algebraic integers. Thus $\upsilon \cdot (\sigma_2(\upsilon)\cdots\sigma_n(\upsilon)N_K(\upsilon)) = 1$ exhibits υ as a unit. Supposing $\omega_1, \ldots, \omega_n$ to be an integral basis for K, we may write down an associated *norm form* $P(x_1, \ldots, x_n) = N_K(x_1\omega_1 + \cdots + x_n\omega_n)$, which is a homogenous polynomial of degree n_K in n_K variables with rational integer coefficients. Norm forms are not unique by any means, since they depend on some arbitrary choice of integral basis. But any two norm forms P and Q for the same number field are related by $Q(\mathbf{x}) = P(A\mathbf{x})$ for some matrix A with rational integer entries and $\det(A) = \pm 1$, since any two integral bases for the number field are similarly related. The two forms P and Q are said to be *equivalent*, and *properly equivalent* if $\det(A) = 1$. The determination of the units of a number field is equivalent to solving the two Diophantine equations $P(x_1, \ldots, x_n) = \pm 1$ over $\mathbb{Z}$ for some norm form P for the number field, and this can be a difficult problem.

As an example we find the units of quadratic number fields $K = \mathbb{Q}(\sqrt{d})$ with $d < 0$. The embeddings are the identity and $a + b\sqrt{d} \mapsto a - b\sqrt{d}$, so the norm is

$$N_K(a + b\sqrt{d}) = (a + b\sqrt{d})(a - b\sqrt{d}) = a^2 - db^2.$$

If $d \equiv 2$ or $3 \pmod 4$ the condition for $\upsilon = x + y\sqrt{d}$ to be a unit is

$$x^2 - dy^2 = \pm 1.$$

For $d = -1$ the equation has the four solutions

$$(x, y) = (\pm 1, 0), (0, \pm 1)$$

yielding four units

$$\upsilon = \pm 1, \pm i$$

among the Gaussian integers in $\mathbb{Q}(\sqrt{-1})$. If $d \leq -2$ the only solutions are $(x, y) = (\pm 1, 0)$, giving the two units $\upsilon = \pm 1$.

If $d \equiv 1 \pmod 4$ the algebraic integers α in $\mathbb{Q}(\sqrt{d})$ may expressed as

$$\alpha = \frac{x}{2} + \frac{y}{2}\sqrt{d}$$

with x and y rational integers satisfying $x \equiv y \pmod 2$. Then

$$\upsilon = \frac{x}{2} + \frac{y}{2}\sqrt{d}$$

is a unit if

$$x^2 - dy^2 = \pm 4 \quad \text{and} \quad x \equiv y \pmod 2$$

For $d = -3$ the equation has the six solutions

$$(x, y) = (\pm 2, 0), (\pm 1, \pm 1),$$

yielding six units

$$\upsilon = \pm 1, \pm\frac{1}{2} \pm \frac{\sqrt{-3}}{2},$$

in $\mathbb{Q}(\sqrt{-3})$. If $d \leq -7$ the only solutions are $(x, y) = (\pm 1, 0)$, giving the two units $\upsilon = \pm 1$.

A quadratic number field K is called an *imaginary quadratic field* if the discriminant d_K is negative. It is called a *real quadratic field* if the discriminant d_K is positive. We have determined the units in all imaginary quadratic fields: They are ± 1 except for $\mathbb{Q}(\sqrt{-1})$, for which they are $\pm 1, \pm i$, and $\mathbb{Q}(\sqrt{-3})$ for which they are $\pm 1, \pm \exp(2\pi i/3), \pm \exp(4\pi i/3)$. For the real quadratic fields, the situation is more complicated. In particular norm forms are indefinite. We shall exhibit one example, and then comment on the general case of real quadratic fields. For the field $\mathbb{Q}(\sqrt{3})$ the condition to determine the units is

$$x^2 - 3y^2 = \pm 1,$$

yielding among others the unit $\upsilon = 2 + \sqrt{3}$. This is of course a unit of infinite order since it is not a root of unity. So υ^m are all distinct units for $m \in \mathbb{Z}$, and $-\upsilon^m$ likewise since -1 is also a unit. It can be shown that υ is the smallest unit larger than 1, called the *fundamental unit* of $\mathbb{Q}(\sqrt{3})$. Every real quadratic field $\mathbb{Q}(\sqrt{d})$ has a fundamental unit ε_d, defined as the least unit $\upsilon > 1$, in terms of which every unit is expressed as $(\pm 1)\varepsilon_d^m$ for some $m \in \mathbb{Z}$ and a choice of sign. The fundamental unit is an important invariant of real quadratic fields, and it fluctuates widely and erratically in size as a function of d. This behavior is related to an unsolved problem of Gauss about binary quadratic forms.

Factorization of ideals

The ring of integers of an algebraic number field need not be a unique factorization domain. The field $K = \mathbb{Q}(\sqrt{-5})$ affords an example, for 6 has two distinct factorizations

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

into irreducibles. The irreducibility of the factors is easily established by means of the norm $N_K(a+b\sqrt{-5}) = a^2+5b^2$. For if α is a nontrivial factor of $1\pm\sqrt{-5}$, then $N_K(\alpha)$ is a nontrivial factor of $N_K(1\pm\sqrt{-5}) = 6$, implying $N_K(\alpha) = 2$ or $N_K(\alpha) = 3$, which is clearly impossible. If α is a nontrivial factor of 2 or 3, then $N_K(\alpha)$ is a nontrivial factor of 4 or 9, so again $N_K(\alpha) = 2$ or $N_K(\alpha) = 3$.

If the ring of algebraic integers $\mathcal{O}_K$ could be extended to a larger ring in which $2, 3, 1\pm\sqrt{-5}$ have unique factorizations into irreducibles, then the above example of nonunique factorization could be interpreted as the result of grouping irreducibles in a finer factorization together in different ways. Dedekind realized that by introducing ideals the benefit of such an extension may be obtained without stepping outside the confines of $\mathcal{O}_K$. Observe that if $2 = \alpha\beta$ and $1 + i\sqrt{-5} = \alpha\gamma$ in a larger ring, where α is a highest common factor of 2 and $1 + \sqrt{-5}$ in that ring, then α may be specified up to associates by the ideal

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

without reference to the larger ring. We now throw away the crutch of thinking in terms of a larger ring, and resolve to discuss factorization in terms of ideals.

An *ideal* in a commutative ring R with 1 is a nonempty set $I \subseteq R$ such that if $a, b \in I$ and $s, t \in R$ then $sa + tb \in I$. The ideal

$$(a_1, \ldots, a_m) = a_1R + \cdots + a_mR$$

is *generated* by $a_1, \ldots, a_m$ and an ideal $(a) = aR$ is called a *principal ideal.* The *zero ideal* (0) and the whole ring $R = (1)$ are principal ideals. A *proper ideal* is any ideal different from these two. A *maximal ideal* is an ideal $M \nsubseteq R$ so that if $M \subseteq I \nsubseteq R$ for some ideal I, then $I = M$. A *prime ideal* is an ideal $P \nsubseteq R$ so that if $ab \in P$ then $a \in P$ or $b \in P$. If $IJ \subseteq P$ with I and J ideals and P a prime ideal, then $I \subseteq P$ or $J \subseteq P$. For if $I \nsubseteq P$ we can choose some $a \in I \setminus P$. Then for arbitrary $b \in J$ we obtain $ab \in IJ \subseteq P$, so $b \in J$ since P is prime.

Any intersection of ideals is an ideal. The *sum* $I + J$ of two ideals is the intersection of all ideals containing I and J, and thus

$$I + J = \{a + b \mid a \in I \text{ and } b \in J\}.$$

The *product* IJ is the intersection of all ideals that contain the products ab with $a \in I$ and $b \in J$. The typical element c of IJ is of the form

$$c = a_1b_1 + \cdots + a_mb_m$$

with $a_1, \ldots, a_m \in I$ and $b_1, \ldots, b_m \in J$. Obviously a product of ideals is contained in the intersection of the same ideals.

Two elements $s, t \in R$ are *congruent* modulo an ideal I if $s - t \in I$, and we write $s \equiv t \pmod{I}$. Congruence modulo an ideal is a straight generalization of congruence modulo an integer q, the latter being the same as congruence modulo the principal ideal (q). Congruence modulo an ideal is an equivalence relation, with equivalence classes $\underline{s} = s + I = \{s + a \mid a \in I\}$ that are called residue classes modulo I. The operations $\underline{s} + \underline{t} = \underline{s + t}$ and $\underline{s} \cdot \underline{t} = \underline{st}$ of addition and multiplication of residue classes are well defined. With these operations the residue classes form a ring R/I called the *quotient ring* of R modulo I.

It turns out that R/I is a field if and only if I is a maximal ideal, and R/I is an integral domain if and only if I is a prime ideal. In particular any maximal ideal is a prime ideal. Suppose first that I is a maximal ideal. Let $\underline{0} \neq \underline{a} \in R/I$ and note that $(a) + I = R$ since I is maximal. There is some $b \in R$ so that $ab \equiv 1 \pmod{I}$, and thus $\underline{a} \cdot \underline{b} = \underline{1}$, so R/I is a field. Assume that R/I is a field and $I \subseteq J$ for some ideal J. Then J/I is an ideal in R/I so $J/I = (\underline{0})$ or $J/I = (\underline{1})$ because R/I is a field. Thus $J = I$ or $J = R$, and I must be a maximal ideal. Suppose that I is a prime ideal, and let $\underline{a}, \underline{b} \in R/I$ with $\underline{a} \cdot \underline{b} = \underline{0}$. Then $ab \in I$ so $a \in I$ or $b \in I$, meaning that $\underline{a} = \underline{0}$ or $\underline{b} = \underline{0}$ in R/I, which must therefore be an integral domain. Assuming that R/I is an integral domain and $ab \in I$, then $\underline{a} \cdot \underline{b} = \underline{0}$ in R/I. Thus $\underline{a} = \underline{0}$ or $\underline{b} = \underline{0}$ so $a \in I$ or $b \in I$ which means that I is a prime ideal.

We are going to establish special properties of the ideals in rings $\mathcal{O}_K$ of algebraic integers of number fields K. These efforts will culminate in the statement that any nonzero ideal in $\mathcal{O}_K$ has a factorization as a product of prime ideals in $\mathcal{O}_K$ that is unique up to order of the factors. In number theory the zero ideal is seldom of any importance, and we adopt the convention that ideal shall mean nonzero ideal unless the opposite is stated.

An ideal $\mathfrak{a}$ in $\mathcal{O}_K$ is a subgroup of the additive group of $\mathcal{O}_K$. The latter is a free abelian group of finite rank $n = n_K$ as seen from the existence of an integral basis for $\mathcal{O}_K$. Thus $\mathfrak{a}$ is a free abelian group of rank at most n. Choose some nonzero algebraic integer $\alpha \in \mathfrak{a}$ and an integral basis $\omega_1, \ldots, \omega_n$ for $\mathcal{O}_K$. Then the principal ideal

$$\alpha\mathcal{O}_K = \mathbb{Z}\alpha\omega_1 + \cdots + \mathbb{Z}\alpha\omega_n$$

is contained in $\mathfrak{a}$ and is a free abelian group of rank n. So the additive group of $\mathfrak{a}$ has rank n and there are elements $\alpha_1, \ldots, \alpha_n \in \mathfrak{a}$ for which

$$\mathfrak{a} = \mathbb{Z}\alpha_1 + \cdots + \mathbb{Z}\alpha_n.$$

Such a collection $\alpha_1, \ldots, \alpha_n$ is called an *ideal basis* for $\mathfrak{a}$. Note that since $\mathfrak{a}$ is finitely generated over $\mathbb{Z}$, it is certainly finitely generated over $\mathcal{O}_K$. A ring all of whose ideals are finitely generated is called *Noetherian.* We have established that $\mathcal{O}_K$ is Noetherian for any number field K.

Any nonzero ideal in $\mathcal{O}_K$ contains a nonzero rational integer, for the minimal equation of a nonzero element $\alpha \in \mathcal{O}_K$ may be rewritten to express its constant term as an element of (α). We now show that $\mathcal{O}_K/\mathfrak{a}$ is finite for any nonzero ideal in $\mathcal{O}_K$. Since there is some positive rational integer $m \in \mathfrak{a}$ and $(m) \subseteq \mathfrak{a}$, it

will be enough to show that $\mathcal{O}_K/(m)$ is finite. If $\omega_1, \ldots, \omega_n$ is an integral basis for $\mathcal{O}_K$ and $\alpha \in \mathcal{O}_K$ is arbitrary, the calculation

$$\begin{aligned}\alpha &\equiv a_1\omega_1 + \cdots + a_n\omega_n \equiv (q_1m + r_1)\omega_1 + \cdots + (q_nm + r_n)\omega_n \\ &\equiv r_1\omega + \cdots + r_n\omega_n + m(q_1\omega_1 + \cdots + q_n\omega_n) \\ &\equiv r_1\omega + \cdots + r_n\omega_n \pmod{(m)}\end{aligned}$$

shows that $\mathcal{O}_K$ has m^n residue classes modulo (m).

Any nonzero prime ideal $\mathfrak{p}$ in $\mathcal{O}_K$ is a maximal ideal. For $\mathcal{O}_K/\mathfrak{p}$ is a finite integral domain. The powers of an arbitrary element $r \neq 0$ of a finite ring cannot all be distinct, so there must be two equal powers of r with different exponents. If the ring is an integral domain, a common factor may be canceled between the two equal powers to show that $r^j = 1$ for some positive integer j. But then $r \cdot r^{j-1} = 1$ implies that the finite integral domain is a field. Thus $\mathcal{O}_K/\mathfrak{p}$ is a field, and $\mathfrak{p}$ is a maximal ideal.

Every nonzero ideal of $\mathcal{O}_K$ contains a product of prime ideals. Otherwise the set of nonzero ideals containing no product of prime ideals would be nonempty. An arbitrary ascending infinite chain $\mathfrak{a}_1 \subseteq \mathfrak{a}_2 \subseteq \cdots$ of such ideals under inclusion becomes stationary. For $|\mathcal{O}_K/\mathfrak{a}_1| \geq |\mathcal{O}_K/\mathfrak{a}_2| \geq \cdots$ is a decreasing sequence of positive integers. Thus by Zorn's Lemma there is an ideal $\mathfrak{a}$ containing no product of prime ideals that is not properly contained in another such ideal. Obviously $\mathfrak{a}$ cannot be a prime ideal, so there exist $\alpha, \beta \in \mathcal{O}_K$ with $\alpha\beta \in \mathfrak{a}$ but $\alpha, \beta \notin \mathfrak{a}$. The ideals $(\alpha) + \mathfrak{a}$ and $(\beta) + \mathfrak{a}$ strictly contain $\mathfrak{a}$ so each contains a product of prime ideals. Then their product $\mathfrak{b} = ((\alpha) + \mathfrak{a})((\beta) + \mathfrak{a})$ contains a product of prime ideals. But $\mathfrak{b} \subseteq \mathfrak{a}$, contradicting the statement that $\mathfrak{a}$ contains no product of prime ideals.

To develop the theory of divisibility for ideals it is convenient, though by no means indispensable, to make use of fractional ideals. A *fractional ideal* $\mathfrak{F}$ of K is a set $\mathfrak{F} \subseteq K$ for which there exists some $\alpha \in K^\times$ such that $\alpha\mathfrak{F}$ is a nonzero ordinary ideal of $\mathcal{O}_K$. Fractional ideals generalize nonzero ideals of $\mathcal{O}_K$, and we shall reserve the term ideal for ideals in the sense of ring theory. In some books the term integral ideal is used for ideals to distinguish them from fractional ideals. As an example of a fractional ideal we note $7\mathbb{Z}/3$ in $\mathbb{Q}$. A fractional ideal of the form $\alpha\mathcal{O}_K$ with $\alpha \in K^\times$ is called a *principal fractional ideal.* Fractional ideals may be multiplied together in the same way as ideals to yield new fractional ideals. For a product of ideals is an ideal and if $\mathfrak{F}_1$ and $\mathfrak{F}_2$ are fractional ideals with $\alpha_1\mathfrak{F}_1$ and $\alpha_2\mathfrak{F}_2$ ideals, then $\alpha_1\alpha_2\mathfrak{F}_1\mathfrak{F}_2 = (\alpha_1\mathfrak{F}_1)(\alpha_2\mathfrak{F}_2)$ is an ideal. Every fractional ideal $\mathfrak{F}$ has an *inverse*

$$\mathfrak{F}^{-1} = \{\alpha \in K \mid \alpha\mathfrak{F} \subseteq \mathcal{O}_K\}$$

which is a fractional ideal. The inverse is closed under addition, and under multiplication by elements of $\mathcal{O}_K$. Since there exists some $\alpha \in K^\times$ with $\alpha\mathfrak{F}$ a necessarily nonzero ideal of $\mathcal{O}_K$, clearly $\mathfrak{F}^{-1}$ contains a nonzero element. If $\beta \in \mathfrak{F}$ then $\beta\mathfrak{F}^{-1} \subseteq \beta\mathfrak{F}\mathfrak{F}^{-1} \subseteq \mathcal{O}_K$, so $\mathfrak{F}^{-1}$ is a fractional ideal. Note that the inverse of a nonzero ideal is a fractional ideal that contains the element 1.

If $\mathfrak{p}$ is a prime ideal, then $\mathfrak{p}\mathfrak{p}^{-1} = \mathcal{O}_K$. For $\mathfrak{p} \subseteq \mathfrak{p}\mathfrak{p}^{-1} \subseteq \mathcal{O}_K$ and $\mathfrak{p}$ is a maximal ideal, so $\mathfrak{p}\mathfrak{p}^{-1} = \mathcal{O}_K$ or $\mathfrak{p}\mathfrak{p}^{-1} = \mathfrak{p}$. If the latter case holds, $\alpha\mathfrak{p} \subseteq \mathfrak{p}$ for all $\alpha \in \mathfrak{p}^{-1}$. Let $\alpha_1, \ldots, \alpha_n$ be an ideal basis for $\mathfrak{p}$ and write

$$\alpha\alpha_i = \sum_{j=1}^{n} a_{ij}\alpha_j \quad , \quad A = [a_{ij}] \quad , \quad a_{ij} \in \mathbb{Z}$$

for $1 \leq i \leq n$. Then α is a root of $\det(x\mathrm{id} - A) = 0$ so α is an algebraic integer, and $\mathfrak{p}^{-1}$ is an ideal. Since $\mathcal{O}_K \subseteq \mathfrak{p}^{-1}$, this implies that $\mathfrak{p}^{-1} = \mathcal{O}_K$. But this is false. For any nonzero $\alpha \in \mathfrak{p}$ there exists a collection $\mathfrak{p}_1, \ldots, \mathfrak{p}_m$ of prime ideals, with m minimal, for which

$$\mathfrak{p}_1 \cdots \mathfrak{p}_m \subseteq \alpha\mathcal{O}_K \subseteq \mathfrak{p}.$$

Then $\mathfrak{p}_1 \subseteq \mathfrak{p}$ after reindexing. Otherwise there are $\alpha_j \in \mathfrak{p}_j \setminus \mathfrak{p}$ for $1 \leq j \leq m$, but $\alpha_1 \cdots \alpha_m \in \mathfrak{p}$, which contradicts the statement that $\mathfrak{p}$ is a prime ideal. Thus $\mathfrak{p}_1 = \mathfrak{p}$ since $\mathfrak{p}_1$ is a prime ideal. Now $\mathfrak{p}_2 \cdots \mathfrak{p}_m \nsubseteq \alpha\mathcal{O}_K$ by the minimality of m, so there is some $\beta \in \mathfrak{p}_2 \cdots \mathfrak{p}_m$ with $\beta \notin \alpha\mathcal{O}_K$. Then $\alpha^{-1}\beta \notin \mathcal{O}_K$, but on the other hand $\alpha^{-1}\beta \in \mathfrak{p}^{-1}$. For $\beta\mathfrak{p} \subseteq \alpha\mathcal{O}_K$ since $\beta\mathfrak{p} \subseteq \mathfrak{p}_2 \cdots \mathfrak{p}_m\mathfrak{p}_1$, so $\alpha^{-1}\beta\mathfrak{p} \subseteq \mathcal{O}_K$.

We define a relation $\mathfrak{a}|\mathfrak{b}$ of divisibility between nonzero ideals of $\mathcal{O}_K$ by $\mathfrak{b} = \mathfrak{a}\mathfrak{c}$ for some nonzero ideal $\mathfrak{c}$. Then $\mathfrak{p}|\mathfrak{a}\mathfrak{b}$ implies $\mathfrak{p}|\mathfrak{a}$ or $\mathfrak{p}|\mathfrak{b}$ if $\mathfrak{p}$ is a prime ideal, in analogy with the definition of prime elements in an integral domain. For $\mathfrak{a}\mathfrak{b} = \mathfrak{p}\mathfrak{c} \subseteq \mathfrak{p}\mathcal{O}_K = \mathfrak{p}$ implies that $\mathfrak{a} \subseteq \mathfrak{p}$ or $\mathfrak{b} \subseteq \mathfrak{p}$, say the former. Then $\mathfrak{a}\mathfrak{p}^{-1} \subseteq \mathfrak{p}\mathfrak{p}^{-1} = \mathcal{O}_K$, so $\mathfrak{a}\mathfrak{p}^{-1}$ is an ideal with $\mathfrak{a} = (\mathfrak{a}\mathfrak{p}^{-1})\mathfrak{p}$, thus $\mathfrak{p}|\mathfrak{a}$.

Unique factorization into prime ideals. *Every nonzero ideal of $\mathcal{O}_K$ is a product of prime ideals that is unique up to order of the factors.*

Proof. The ideal (1) factors as the empty product. Assume the set of nonzero ideals that do not factor into prime ideals to be nonempty. An arbitrary ascending infinite chain $\mathfrak{a}_1 \subseteq \mathfrak{a}_2 \subseteq \cdots$ of such ideals under inclusion becomes stationary. For $|\mathcal{O}_K/\mathfrak{a}_1| \geq |\mathcal{O}_K/\mathfrak{a}_2| \geq \cdots$ is a decreasing sequence of positive integers. Thus there is an ideal $\mathfrak{a}$ that does not factor into prime ideals that is not properly contained in another such ideal. There is a maximal ideal, hence a prime ideal, $\mathfrak{p} \supseteq \mathfrak{a}$. Then $\mathfrak{a} \subseteq \mathfrak{a}\mathfrak{p}^{-1} \subseteq \mathfrak{p}\mathfrak{p}^{-1} = \mathcal{O}_K$. Obviously $\mathfrak{p} \neq \mathfrak{a}$, so $\mathfrak{a}\mathfrak{p}^{-1} \neq \mathcal{O}_K$. Then $\mathfrak{a}\mathfrak{p}^{-1}$ has a factorization into prime ideals since every ideal properly containing $\mathfrak{a}$ has such a factorization. But then $\mathfrak{a} = \mathfrak{a}\mathcal{O}_K = (\mathfrak{a}\mathfrak{p}^{-1})\mathfrak{p}$ also has such a factorization, a contradiction.

It remains to prove uniqueness of factorization into prime ideals. Suppose that

$$\mathfrak{p}_1 \cdots \mathfrak{p}_s = \mathfrak{q}_1 \cdots \mathfrak{q}_t$$

are two factorizations into prime ideals. Then there is some j with $1 \leq j \leq t$ such that $\mathfrak{p}_s|\mathfrak{q}_j$. So $\mathfrak{q}_j \subseteq \mathfrak{p}_s$, but both ideals are maximal, therefore $\mathfrak{p}_s = \mathfrak{q}_j$. Multiplying through by $\mathfrak{p}_s^{-1}$ cancels the factor $\mathfrak{p}_s$ on both sides, and we are left with the same kind of equal factorizations, with $s-1$ and $t-1$ prime ideals on the two sides. Continuing like this shows that $s = t$ and that the prime ideals in one of the factorizations equal the prime ideals in the other in some order. $\square$

Every nonzero ideal $\mathfrak{a}$ of $\mathcal{O}_K$ has an inverse $\mathfrak{a}^{-1}$ under multiplication of fractional ideals. For there is a factorization

$$\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_s$$

into prime ideals, and

$$\mathfrak{a} = \mathfrak{p}_1^{-1} \cdots \mathfrak{p}_s^{-1}$$

is an inverse, as one sees by multiplying the two products together and using $\mathfrak{p}_j \mathfrak{p}_j^{-1} = \mathcal{O}_K$.

Suppose $\mathfrak{a}$ and $\mathfrak{b}$ are nonzero ideals of $\mathcal{O}_K$. If $\mathfrak{a}|\mathfrak{b}$, then $\mathfrak{b} = \mathfrak{a}\mathfrak{c} \subseteq \mathfrak{a}\mathcal{O}_K = \mathfrak{a}$. While if $\mathfrak{b} \subseteq \mathfrak{a}$, then $\mathfrak{a}^{-1}\mathfrak{b} \subseteq \mathfrak{a}^{-1}\mathfrak{a} = \mathcal{O}_K$, so $\mathfrak{a}^{-1}\mathfrak{b}$ is an ideal with $\mathfrak{a}(\mathfrak{a}^{-1}\mathfrak{b}) = \mathfrak{b}$. Thus $\mathfrak{a}|\mathfrak{b}$ and $\mathfrak{b} \subseteq \mathfrak{a}$ are equivalent for nonzero ideals. This observation has the succinct formulation 'to divide is to contain.'

If $\mathfrak{a}$ and $\mathfrak{b}$ are nonzero ideals, and $\mathfrak{p}_1, \ldots, \mathfrak{p}_s$ the distinct prime ideals that divide one or the other of them, we may write

$$\mathfrak{a} = \mathfrak{p}_1^{a_1} \cdots \mathfrak{p}_s^{a_s} \quad , \quad \mathfrak{b} = \mathfrak{p}_1^{b_1} \cdots \mathfrak{p}_s^{b_s}$$

with $a_1, \ldots, a_s, b_1, \ldots, b_s$ nonnegative integers. Then one shows that $\mathfrak{a}|\mathfrak{b}$ if and only if $a_j \leq b_j$ for $1 \leq j \leq s$ just as in the case of rational integers.

There is an important result that is central to some proofs of the theorem on factorization into prime ideals, which for us is rather a consequence: For every nonzero ideal $\mathfrak{a}$ of $\mathcal{O}_K$, there is some $\alpha \in \mathcal{O}_K$ with $\mathfrak{a}|(\alpha)$. Briefly, every ideal divides some principal ideal. It is obviously enough to show that every nonzero prime ideal divides some principal ideal. If $\mathfrak{p}$ is prime ideal, choose some nonzero $\beta \in \mathfrak{p}$. Then $(\beta) \subseteq \mathfrak{p}$ so $(\beta)\mathfrak{p}^{-1} \subseteq \mathfrak{p}\mathfrak{p}^{-1} = \mathcal{O}_K$, thus $(\beta)\mathfrak{p}^{-1}$ is an ideal with $((\beta)\mathfrak{p}^{-1})\mathfrak{p} = (\beta)$, which implies $\mathfrak{p}|(\beta)$.

Every fractional ideal $\mathfrak{a}$ has an inverse, for if $\alpha \in K^\times$ with $\alpha\mathfrak{a} = \mathfrak{b}$ an ideal, we have $\mathfrak{a}^{-1} = (\alpha)\mathfrak{b}^{-1}$. Thus the fractional ideals of $\mathcal{O}_K$ form a group, which we denote by J_K. Since $\mathfrak{a} = (m)(\beta)^{-1}\mathfrak{b}$ with $\alpha = \beta/m$, β an algebraic integer and m a positive rational integer, factorization of ideals into prime ideals implies that fractional ideals are of the form

$$\mathfrak{a} = \mathfrak{p}_1^{a_1} \cdots \mathfrak{p}_s^{a_s}$$

where $\mathfrak{p}_1, \ldots, \mathfrak{p}_s$ are prime ideals and $a_1, \ldots, a_s \in \mathbb{Z}$. Thus J_K is the free abelian group generated by the prime ideals of $\mathcal{O}_K$ under multiplication. This group has a subgroup P_K isomorphic with $K^\times$ consisting of the principal fractional ideals.

The *norm* of a nonzero ideal of $\mathcal{O}_K$ is

$$N(\mathfrak{a}) \stackrel{\text{def}}{=} |\mathcal{O}_K/\mathfrak{a}|.$$

This is also called the *absolute norm.* Recall that we already showed that $\mathcal{O}_K/\mathfrak{a}$ is finite if $\mathfrak{a}$ is a nonzero ideal of $\mathcal{O}_K$. The most important property of the norm on ideals is that it is totally multiplicative: If $\mathfrak{a}$ and $\mathfrak{b}$ are nonzero ideals, then $N(\mathfrak{a}\mathfrak{b}) = N(\mathfrak{a})N(\mathfrak{b})$. To prove this we establish a generalization of the Chinese Remainder Theorem.

Chinese Remainder Theorem for ideals. *Suppose that R is a commutative ring with 1 and that $I_1, \ldots, I_m$ are ideals with $I_i + I_j = R$ whenever $i \neq j$. Then*

$$R/I \cong (R/I_1) \times \cdots \times (R/I_m)$$

with $I = I_1 \cdots I_m$.

Proof. The homomorphism

$$\vartheta : R \to (R/I_1) \times \cdots \times (R/I_m)$$

given by

$$a \mapsto (a + I_1, \ldots, a + I_m)$$

has kernel $\ker(\vartheta) = I_1 \cap \cdots \cap I_m$. We need to show that ϑ is epi and that $I_1 \cdots I_m = I_1 \cap \cdots \cap I_m$. Now

$$\prod_{j \neq i} (I_i + I_j) = R$$

by $I_i + I_j = R$ for $1 \leq i \neq j \leq m$. Thus

$$I_i + \prod_{j \neq i} I_j = R$$

because all the terms in the expansion of the product of the $(I_i + I_j)$-s except one are contained in I_i. Then there exist $a_i \in I_i$ and

$$b_i \in \prod_{j \neq i} I_j \subseteq \bigcap_{j \neq i} I_j$$

such that $a_i + b_i = 1$. Supposing

$$(c_1 + I_1, \ldots, c_m + I_m) \in (R/I_1) \times \cdots \times (R/I_m)$$

to be arbitrary, we have

$$\begin{aligned} \vartheta(b_1 c_1 + \cdots + b_m c_m) &= (b_1 c_1 + \cdots + b_m c_m + I_1, \ldots, b_1 c_1 + \cdots + b_m c_m + I_m) \\ &= (c_1 + I_1, \ldots, c_m + I_m) \end{aligned}$$

since $a_i \in I_i$ implies $b_i - 1 \in I_i$, and $b_i \in I_j$ for $j \neq i$ by the obvious inclusion between products and intersections of ideals. Thus ϑ is an epimorphism.

To establish the non-obvious inclusion

$$I_1 \cap \cdots \cap I_m \subseteq I_1 \cdots I_m$$

we use the condition that $I_i + I_j = R$ whenever $i \neq j$, proceeding by induction. Suppose that $m = 2$, and choose $a_1 \in I_1$ and $a_2 \in I_2$ such that $a_1 + a_2 = 1$. If $a \in I_1 \cap I_2$, then $a = aa_1 + aa_2 \in I_1 I_2$. Now assume that the inclusion has been proved for some $m \geq 2$. Then $I_1 \cap \cdots \cap I_m = I_1 \cdots I_m$, so $I_1 \cap \cdots \cap I_{m+1} = (I_1 \cdots I_m) \cap I_{m+1}$. But $I_1 \cdots I_m + I_{m+1} = R$ was established above, so we can choose $a_1 \in I_1 \cdots I_m$ and $a_2 \in I_{m+1}$ such that $a_1 + a_2 = 1$. If $a \in (I_1 \cdots I_m) \cap I_{m+1}$, then $a = aa_1 + aa_2 \in (I_1 \cdots I_m) I_{m+1}$. Since the other inclusion is obvious, we obtain $I_1 \cdots I_m = I_1 \cap \cdots \cap I_m$. $\square$

We have

$$\begin{aligned} N(\mathfrak{a}) = |\mathcal{O}_K/\mathfrak{a}| &= |(\mathcal{O}_K/\mathfrak{p}_1^{a_1}) \oplus \cdots \oplus (\mathcal{O}_K/\mathfrak{p}_s^{a_s})| \\ &= |\mathcal{O}_K/\mathfrak{p}_1^{a_1}| \cdots |\mathcal{O}_K/\mathfrak{p}_s^{a_s}| = N(\mathfrak{p}_1^{a_1}) \cdots N(\mathfrak{p}_s^{a_s}) \end{aligned}$$

by the factorization $\mathfrak{a} = \mathfrak{p}_1^{a_1} \cdots \mathfrak{p}_s^{a_s}$ into powers of distinct prime ideals, and the Chinese Remainder Theorem. For

$$\mathfrak{p}_i^{a_i} + \mathfrak{p}_j^{b_j} = \mathcal{O}_K$$

if $\mathfrak{p}_i$ and $\mathfrak{p}_j$ are distinct prime ideals, otherwise they would be divisible by a common prime ideal $\mathfrak{p}$, an impossibility.

We now show that $|\mathcal{O}_K/\mathfrak{p}^a| = |\mathcal{O}_K/\mathfrak{p}|^a$ to conclude that $N(\mathfrak{p}^a) = N(\mathfrak{p})^a$. This implies the total multiplicativity of the absolute norm. If $a \geq 2$ then $\mathfrak{p}^{a-1}/\mathfrak{p}^a$ is a subgroup of $\mathcal{O}_K/\mathfrak{p}^a$, and $\mathcal{O}_K/\mathfrak{p}^{a-1} \cong (\mathcal{O}_K/\mathfrak{p}^a)/(\mathfrak{p}^{a-1}/\mathfrak{p}^a)$. So by induction it will be sufficient to establish that $|\mathfrak{p}^{a-1}/\mathfrak{p}^a| = |\mathcal{O}_K/\mathfrak{p}|$. Since $\mathfrak{p}^a \neq \mathfrak{p}^{a-1}$ there is some $\alpha \in \mathfrak{p}^{a-1} \setminus \mathfrak{p}^a$. The inclusions $\mathfrak{p}^a \subseteq (\alpha) + \mathfrak{p}^a \subseteq \mathfrak{p}^{a-1}$ imply that $(\alpha) + \mathfrak{p}^a = \mathfrak{p}^{a-1}$ since $\mathfrak{p}$ is a prime ideal. The map $\vartheta : \mathcal{O}_K \to \mathfrak{p}^{a-1}/\mathfrak{p}^a$ given by $\beta \mapsto \alpha\beta + \mathfrak{p}^a$ is an epimorphism of abelian groups. For

$$\vartheta(\beta + \gamma) = \alpha(\beta + \gamma) + \mathfrak{p}^a = \alpha\gamma\mathfrak{p}^a + \alpha\gamma + \mathfrak{p}^a = \vartheta(\beta) + \vartheta(\gamma)$$

and $\mathrm{im}(\vartheta) = ((\alpha) + \mathfrak{p}^a) + \mathfrak{p}^a = \mathfrak{p}^{a-1}/\mathfrak{p}^a$. Furthermore $\beta \in \ker(\vartheta)$ if and only if $\alpha\beta \in \mathfrak{p}^a$, which is equivalent to $\mathfrak{p}^a | (\alpha)(\beta)$. While $\alpha \in \mathfrak{p}^{a-1} \setminus \mathfrak{p}^a$ is equivalent to $\mathfrak{p}^{a-1} || (\alpha)$, and thus to $\mathfrak{p} | (\beta)$ by the theorem on unique factorization into prime ideals. Thus $\ker(\vartheta) = \mathfrak{p}$ so that $\mathfrak{p}^{a-1}/\mathfrak{p}^a = \mathrm{im}(\vartheta) \cong \mathcal{O}_K/\mathfrak{p}$.

We now show that if α is a nonzero element of $\mathcal{O}_K$, then $N((\alpha)) = |N_{K/\mathbb{Q}}(\alpha)|$. That is to say, the absolute norm of a principal ideal equals the absolute value of the norm of any generator. Let $\mathcal{B} = (\omega_1, \ldots, \omega_n)$ be an ordered integral basis for $\mathcal{O}_K$. The linear operator M_α used to define the norm $N_{K/\mathbb{Q}}(\alpha)$ may be represented relative to $\mathcal{B}$ by a matrix A with rational integer entries, and

$$N_{K/\mathbb{Q}}(\alpha) = \det(M_\alpha) = \det(A).$$

Moreover $\mathcal{O}_K/(\alpha) = \mathcal{O}_K/M_\alpha\mathcal{O}_K \cong \mathbb{Z}^n/\mathrm{im}(\vartheta)$ where the homomorphism $\vartheta : \mathbb{Z}^n \to \mathbb{Z}^n$ is given by $\vartheta(\mathbf{x}) = A\mathbf{x}$. The image of ϑ consists of those $\mathbf{b} \in \mathbb{Z}^n$ for which $A\mathbf{x} = \mathbf{b}$ is solvable in integers, and they can be determined by means of the adjugate matrix. We have $\det(A)\mathbf{x} = \mathrm{adj}(A)A\mathbf{x} = \mathrm{adj}(A)\mathbf{b}$, so the condition on $\mathbf{b}$ is that $\mathrm{adj}(A)\mathbf{b} \equiv \mathbf{0} \pmod{\det(A)}$. Then $|\mathbb{Z}^n/\mathrm{im}(\varphi)| = |\det(A)|$ since there is a single solution modulo $\det(A)$, and

$$N((\alpha)) = |\mathcal{O}_K/(\alpha)| = |\mathbb{Z}^n/\mathrm{im}(\vartheta)| = |\det(A)| = |N_{K/\mathbb{Q}}(\alpha)|.$$

We define the absolute norm of the zero ideal to be zero so as to make the relationship between the two kinds of norm hold for all principal ideals.

The material about norms is vitally important to us, because the very definition of the Dedekind zeta function hinges on it.

Finite fields

At this point we require some information about finite fields. By congruential arithmetic we already know that $\mathbb{Z}/p\mathbb{Z}$ is a field with p elements. Denote its isomorphism class by $\mathbb{F}_p$. We shall soon see that $\mathbb{F}_p$ is the only field with p elements.

If a field F does not contain a subfield isomorphic to $\mathbb{Q}$, then there must exist some positive integer m so that $1+\cdots+1=0$ with m summands. For otherwise the field contains a subring isomorphic to $\mathbb{Z}$, whose field of fractions is isomorphic to $\mathbb{Q}$. The smallest such m is called the *characteristic* of the field and is denoted by $\mathrm{char}(F)$. We write $\mathrm{char}(F)=0$ if F contains a subfield isomorphic to $\mathbb{Q}$. If F has positive characteristic, the characteristic is a prime number p. For m has a factorization into primes, thus $0=1+\cdots+1=(1+\cdots+1)\cdots(1+\cdots+1)$ where the first sum of 1-s in F has m terms, and each of the other sums has a prime number of terms. But at least one of these sums must be zero, since F is a field. If F has characteristic p, the ring homomorphism $\vartheta:\mathbb{Z}\to F$ defined by $\vartheta(n)=n\cdot 1$ has kernel $\ker(\vartheta)=p\mathbb{Z}$. Thus $\mathrm{im}(\vartheta)$ is isomorphic to $\mathbb{Z}/p\mathbb{Z}$ as a ring. But then F contains a subfield isomorphic to $\mathbb{F}_p$. So if F itself has p elements, it is isomorphic to $\mathbb{F}_p$.

Every finite subgroup G of the multiplicative group $F^\times$ of a field is cyclic. For the structure theorem for finitely generated abelian groups implies that $G\cong(\mathbb{Z}/d_1\mathbb{Z})\oplus\cdots\oplus(\mathbb{Z}/d_s\mathbb{Z})$ with $d_1|\cdots|d_s$. Then $x^{d_s}=1$ for all $x\in G$, so this equation has at least $|G|$ roots in F, hence $d_s\geq|G|$. But then $d_s=|G|$ and $d_1=\cdots=d_{s-1}=1$, so G is cyclic.

Up to isomorphism every field is an extension of $\mathbb{Q}$ or $\mathbb{F}_p$ for some prime p. In particular every finite field F is a finite extension of $\mathbb{F}_p$ where $p=\mathrm{char}(F)$. Suppose F/E is any finite extension of finite fields. Since $F^\times$ is cyclic, it has a generator α, and clearly $F=E(\alpha)$. So finite extensions of finite fields are simple. But the only special property of the field extensions that we used in our proof of the Fundamental Theorem of Galois Theory for finite extensions of number fields was that they are simple. Hence the Fundamental Theorem of Galois Theory holds for finite extensions of finite fields.

The map $\sigma_p:F\to F$ given by $\sigma_p(x)=x^p$ is an endomorphism. This is clear for multiplication and follows for addition by the Binomial Theorem

$$(x+y)^p=x^p+\binom{p}{1}x^{p-1}y+\cdots+\binom{p}{p-1}xy^{n-1}+y^p$$

since the binomial coefficients

$$\binom{p}{1},\binom{p}{2},\ldots,\binom{p}{p-1}$$

are divisible by p. The map σ_p is called the *Frobenius endomorphism.* Clearly $\sigma_p(1)=1$. Every nonzero homomorphism between fields is a monomorphism, and since F has finitely many elements, σ_p is also an epimorphism, thus an automorphism.

An extension F/E of finite fields has a basis of $n_{F/E}$ elements over E, so

$$|F| = |E|^{n_{F/E}}.$$

Choosing $E = \mathbb{F}_p$ in particular, we see that $|F| = p^n$ with $n = n_{F/\mathbb{F}_p}$. Since $E^\times$ has order $|E| - 1$, we have $x^{|E|} = x$ for all $x \in E$. Thus σ_p^m with $m = n_{E/\mathbb{F}_p}$ is an automorphism of F over E. Supposing α to be a generator of $F^\times$, we have

$$(\sigma_p^m)^k(\alpha) = \alpha^{p^{km}}.$$

But α has order $p^n - 1$, so all these images of α are distinct for $1 \leq k \leq n_{F/E}$. This yields as many distinct automorphisms $(\sigma_p^m)^k$ of F over E as the degree of F/E, so the extension F/E is normal. Thus the Fundamental Theorem of Galois Theory applies to any extension F/E of finite fields. We also see that $\mathrm{Gal}(F/E) = \langle \sigma_p^m \rangle \cong \mathbb{Z}/n_{F/E}\mathbb{Z}$ so the Galois group of an extension of finite fields is cyclic, with a canonical generator given in terms of the Frobenius endomorphism.

We now prove that $\mathbb{F}_p$ has at most one extension of each degree up to isomorphism. If $E = \mathbb{F}_p(\alpha)$ and $F = \mathbb{F}_p(\beta)$ are finite extensions of $\mathbb{F}_p$, then the field $\mathbb{F}_p(\alpha, \beta)$ may be constructed by means of the minimal polynomial of β over E, and E and F are both subfields of the finite field $\mathbb{F}_p(\alpha, \beta)$. The Galois group of the extension $\mathbb{F}_p(\alpha, \beta)/\mathbb{F}_p$ is isomorphic to $\mathbb{Z}/n\mathbb{Z}$ where n is the degree of the extension. The latter group has a single subgroup of order d for each $d|n$ and no other subgroups, so the extension $\mathbb{F}_p(\alpha, \beta)/\mathbb{F}_p$ has a single intermediate field of degree d for each $d|n$ by the Fundamental Theorem of Galois Theory. Thus E and F are equal if they have the same number of elements, and finite fields with the same number of elements must be isomorphic.

It remains to show that there exists a finite field with p^n elements for every integer $n \geq 1$. For this purpose it will be enough to establish the existence of an irreducible polynomial over $\mathbb{F}_p$ of each degree n. Define $N_p(n)$ to be the number of monic irreducible polynomials of degree n over $\mathbb{F}_p$. Following Gauss we calculate $N_p(n)$, and show that it is always positive. This will complete our brief account of finite fields.

We begin by establishing the result

$$\gcd(x^a - 1, x^b - 1) = x^{\gcd(a,b)} - 1$$

from polynomial algebra. The formula is trivial if $a = b = 1$ or $a = b$, so we may assume that $a < b$ and prove it by induction on b. We have

$$\begin{aligned}\gcd(x^a - 1, x^b - 1) &= \gcd(x^a - 1, x^{b-a}(x^a - 1) + x^{b-a} - 1) \\ &= \gcd(x^a - 1, x^{b-a} - 1) = x^{\gcd(a,b-a)} - 1 \\ &= x^{\gcd(a,b)} - 1\end{aligned}$$

by the induction hypothesis, since $b - a < b$.

Fix the prime p and let $V_{d,p}(x)$ denote the product of the monic irreducible polynomials over $\mathbb{F}_p$ of degree d. Let $f(x)$ be one of these polynomials and

assume that $d|n$. Let α be a root of α. Since there are $p^d - 1$ elements of $\mathbb{F}_p(\alpha)^\times$, we have $\alpha^{p^d} = \alpha$, and thus $f(x)$ divides $x^{p^d} - x$. Now

$$\begin{aligned}\gcd\left(x^{p^d} - x, x^{p^n} - x\right) &= x \cdot \gcd\left(x^{p^d-1} - 1, x^{p^n-1} - 1\right) \\ &= x\left(x^{\gcd(p^d-1,p^n-1)} - 1\right) \\ &= x\left(x^{p^{\gcd(d,n)}-1} - 1\right) = x^{p^d} - x\end{aligned}$$

on applying the above result in polynomial algebra twice. We conclude that if $d|n$, then every monic irreducible polynomial over $\mathbb{F}_p$ of degree d divides $x^{p^n} - 1$.

On the other hand, none of these irreducible polynomials divide $x^{p^n} - x$ to a higher power than the first. For

$$\frac{d}{dx}\left(x^{p^n} - x\right) = p^n x^{p^n-1} - 1$$

so that $x^{p^n} - 1$ has no multiple zeros in any field of characteristic p. For the derivative is constant, and equal to -1, in any such field.

If $f(x)$ is an irreducible divisor of $x^{p^n} - x$ of degree d and α a root of $f(x)$, then $\alpha^{p^n} = \alpha$. An arbitrary element β of $\mathbb{F}_p(\alpha)$ is a polynomial in α with coefficients in $\mathbb{F}_p$, and so $\sigma_p^n(\beta) = \beta$ since σ_p is an automorphism over $\mathbb{F}_p$ and $\alpha^{p^n} = \alpha$. Thus every element β of $\mathbb{F}_p(\alpha)$ is a root of $x^{p^n} - x$. On the other hand $\mathbb{F}_p(\alpha)$ has p^d elements, so

$$(x^{p^d} - x)|(x^{p^n} - x),$$

because $x^{p^d} - x$ is the product of the linear factors $x - \beta$ with $\beta \in \mathbb{F}_p(\alpha)$. Consequently $d|n$ by the above result from polynomial algebra.

The above conclusions imply that

$$x^{p^n} - x = \prod_{d|n} V_{d,p}(x)$$

and a comparison of degrees yields

$$p^n = \sum_{d|n} dN_p(d).$$

Then

$$N_p(n) = \frac{1}{n}\sum_{d|n} \mu\left(\frac{n}{d}\right) p^d$$

by Möbius inversion. And

$$\left|N_p(n) - \frac{p^n}{n}\right| \leq \frac{p^{n/2}}{n(p-1)}$$

since $d|n$ and $d \neq n$ implies $d \leq n/2$. In particular $N_p(n) > 0$ for all primes p and positive integers n.

Decomposition and ramification

We consider the decomposition of prime ideals in extensions K/k of number fields. If $\mathfrak{p}$ is a nonzero prime ideal of $\mathcal{O}_k$, the extension $\mathfrak{p}\mathcal{O}_K$ in $\mathcal{O}_K$ generated by $\mathfrak{p}$ may not be a prime ideal, but it can be factored as a product

$$\mathfrak{p}\mathcal{O}_K = \prod_{j=1}^{g} \mathfrak{P}_j^{e_j}$$

of powers of prime ideals $\mathfrak{P}_j$ of $\mathcal{O}_K$. We say that the $\mathfrak{P}_j$ *lie over* $\mathfrak{p}$ in K/k. The $\mathfrak{P}_j$ in the factorization are precisely those prime ideals $\mathfrak{P}$ for which $\mathfrak{P} \cap \mathcal{O}_k = \mathfrak{p}$. For if $\mathfrak{P} \cap \mathcal{O}_k = \mathfrak{p}$ then $\mathfrak{p}\mathcal{O}_k \subseteq \mathfrak{P}$, so $\mathfrak{P}$ divides $\mathfrak{p}\mathcal{O}_K$, and must be one of the $\mathfrak{P}_j$. While $\mathfrak{P}_j \cap \mathcal{O}_k \supseteq \mathfrak{p}\mathcal{O}_K \cap \mathcal{O}_k = \mathfrak{p}$, so $\mathfrak{P}_j \cap \mathcal{O}_k = \mathfrak{p}$ since $\mathfrak{p}$ is a maximal ideal. In particular a prime ideal $\mathfrak{P}$ in $\mathcal{O}_K$ cannot lie over more than one prime ideal in $\mathcal{O}_k$, and in fact every nonzero prime ideal $\mathfrak{P}$ of $\mathcal{O}_K$ lies over exactly one nonzero prime ideal in $\mathcal{O}_k$. For the ideal $\mathfrak{P} \cap \mathcal{O}_k$ is nonzero, since it contains the positive integer $N(\mathfrak{P})$. And it is a prime ideal in $\mathcal{O}_k$, for if $\alpha, \beta \in \mathcal{O}_k$ with $\alpha\beta \in \mathfrak{P} \cap \mathcal{O}_k$, then $\alpha\beta \in \mathfrak{P}$ implies $\alpha \in \mathfrak{P}$ or $\beta \in \mathfrak{P}$ since $\mathfrak{P}$ is a prime ideal. Moreover, there is at least one nonzero prime ideal $\mathfrak{P}$ of $\mathcal{O}_K$ over every nonzero prime ideal $\mathfrak{p}$ of $\mathcal{O}_k$. For $\mathfrak{p}\mathcal{O}_K \neq \mathcal{O}_K$ since equality would yield $\mathcal{O}_K = \mathfrak{p}^{-1}\mathfrak{p}\mathcal{O}_K = \mathfrak{p}^{-1}\mathcal{O}_K$, and $\mathfrak{p}^{-1}$ contains an element of k that is not an algebraic integer, as we saw during the discussion of the inverse of a prime ideal.

The finite field $\mathcal{O}_K/\mathfrak{P}_j$ may be viewed as an extension of $\mathcal{O}_k/\mathfrak{p}$, for the map $\alpha + \mathfrak{p} \mapsto \alpha + \mathfrak{P}_j$ is well defined by the inclusions $\mathfrak{p} \subseteq \mathfrak{p}\mathcal{O}_K \subseteq \mathfrak{P}_j$, and it is obviously a homomorphism, thus a monomorphism since $\mathcal{O}_k/\mathfrak{p}$ and $\mathcal{O}_K/\mathfrak{P}_j$ are fields. The degree of $\mathcal{O}_K/\mathfrak{P}_j$ over $\mathcal{O}_k/\mathfrak{p}$ is denoted f_j and is called the *residue class degree* of $\mathfrak{P}_j$ over $\mathfrak{p}$. We shall allow ourselves to use the notation $\deg(\mathfrak{p})$ for the residue class degree of $\mathfrak{p}$ over the rational prime that $\mathfrak{p}$ lies over. The exponent e_j is called the *ramification index* of $\mathfrak{P}_j$ over $\mathfrak{p}$. If at least one ramification index is larger than one, the prime ideal $\mathfrak{p}$ is said to be *ramified* in the extension, otherwise it is unramified. An unramified prime ideal $\mathfrak{p}$ *remains inert* in the extension if $\mathfrak{p}\mathcal{O}_K$ is a prime ideal, otherwise it *splits.* It *splits completely* if all the residue class degrees equal one.

The residue class rings $\mathcal{O}_K/\mathfrak{p}\mathcal{O}_K$ and $\mathcal{O}_K/\mathfrak{P}_j^{e_j}$ are vector spaces over the finite field $\mathcal{O}_k/\mathfrak{p}$. We establish the second statement; the first goes in the same way. The multiplication

$$(\alpha + \mathfrak{p})(\beta + \mathfrak{P}_j^{e_j}) = \alpha\beta + \mathfrak{P}_j^{e_j} \quad , \quad \alpha \in \mathcal{O}_k \quad , \quad \beta \in \mathcal{O}_K$$

is well defined. For if $\alpha - \alpha' \in \mathfrak{p}$ and $\beta - \beta' \in \mathfrak{P}_j^{e_j}$, then

$$\alpha\beta - \alpha'\beta' = \alpha(\beta - \beta') + (\alpha - \alpha')\beta' \in \mathfrak{P}_j^{e_j}$$

since $\mathfrak{P}_j^{e_j}$ contains $\mathfrak{p}\mathcal{O}_K$.

Now

$$\mathcal{O}_K/\mathfrak{p}\mathcal{O}_K = \mathcal{O}_K/\prod_{j=1}^{g} \mathfrak{P}_j^{e_j} \cong \bigoplus_{j=1}^{g} \left(\mathcal{O}_K/\mathfrak{P}_j^{e_j}\right)$$

by the Chinese Remainder Theorem for ideals, so

$$\dim(\mathcal{O}_K/\mathfrak{p}\mathcal{O}_K) = \sum_{j=1}^{g} \dim\left(\mathcal{O}_K/\mathfrak{P}_j^{e_j}\right),$$

where the dimensions are over $\mathcal{O}_k/\mathfrak{p}$. From our discussion of the absolute norm we have

$$|\mathcal{O}_K/\mathfrak{P}_j^{e_j}| = |\mathcal{O}_K/\mathfrak{P}_j|^{e_j} = |\mathcal{O}_k/\mathfrak{p}|^{e_j f_j},$$

and thus

$$\dim\left(\mathcal{O}_K/\mathfrak{P}_j^{e_j}\right) = e_j f_j.$$

We shall show that $\dim(\mathcal{O}_K/\mathfrak{p}\mathcal{O}_K) = n = n_{K/k}$, but this requires a little more work. Firstly any $n+1$ elements of $\mathcal{O}_K/\mathfrak{p}\mathcal{O}_K$ are linearly dependent over $\mathcal{O}_k/\mathfrak{p}$. For any $n+1$ elements $\beta_1, \ldots, \beta_{n+1} \in \mathcal{O}_K$ there are $\alpha_1, \ldots, \alpha_{n+1} \in \mathcal{O}_k$, not all zero, such that

$$\alpha_1\beta_1 + \cdots + \alpha_{n+1}\beta_{n+1} = 0.$$

The ideal $\mathfrak{a} = (\alpha_1, \ldots, \alpha_{n+1})$ in $\mathcal{O}_k$ is nonzero, so the inverse $\mathfrak{a}^{-1}$ exists and $\mathfrak{a}^{-1}\mathfrak{a} = \mathcal{O}_k$. Thus there exists some $\gamma \in \mathfrak{a}^{-1}$ with $\gamma\mathfrak{a} \nsubseteq \mathfrak{p}$, which implies that $\gamma\alpha_j \notin \mathfrak{p}$ for some j with $1 \le j \le n+1$. Then

$$(\gamma\alpha_1 + \mathfrak{p})(\beta_1 + \mathfrak{p}\mathcal{O}_K) + \cdots + (\gamma\alpha_{n+1} + \mathfrak{p})(\beta_{n+1} + \mathfrak{p}\mathcal{O}_K) = \mathfrak{p}\mathcal{O}_K$$

with not all $\gamma\alpha_j + \mathfrak{p}$ zero in $\mathcal{O}_k/\mathfrak{p}\mathcal{O}_k$, and the linear dependence is proved.

Next we show that if $\beta_1 + \mathfrak{p}\mathcal{O}_K, \ldots, \beta_m + \mathfrak{p}\mathcal{O}_K$ span $\mathcal{O}_K/\mathfrak{p}\mathcal{O}_K$ over $\mathcal{O}_k/\mathfrak{p}$, then $\beta_1, \ldots, \beta_m$ span K over k. For this purpose it is convenient to make use of the concept of module. An *R-module* M is an abelian group under addition acted on multiplicatively by a ring R such that $r(a+b) = ra + rb$, $(r+s)a = ra + sa$, $(rs)a = r(sa)$ and $1 \cdot a = a$ for $a, b \in M$ and $r, s \in R$. We consider only the case where the ring R is commutative with 1, though the module concept is also extremely important for noncommutative rings. From our point of view the most illuminating examples of modules over a ring are its ideals, i. e. the modules over the ring that are contained in the ring.

Define a $\mathcal{O}_k$-module M by $M = \mathcal{O}_k\beta_1 + \cdots + \mathcal{O}_k\beta_m$, and a factor module $N = \mathcal{O}_K/M$, noting that M is a submodule of the $\mathcal{O}_k$-module $\mathcal{O}_K$. We already know from the discussion leading up to unique factorization of ideals into prime ideals that $\mathcal{O}_K$ is finitely generated over $\mathbb{Z}$, so it is certainly finitely generated over $\mathcal{O}_k$. Any factor module of a finitely generated module is finitely generated, since the images of a set of generators under the quotient map are a set of generators, and thus N is finitely generated over $\mathcal{O}_k$. We now make use of the assumption on $\beta_1, \ldots, \beta_m$, obtaining

$$\begin{aligned} M + \mathfrak{p}\mathcal{O}_K &= \mathcal{O}_k\beta_1 + \cdots + \mathcal{O}_k\beta_m + \mathfrak{p}\mathcal{O}_K \\ &= (\mathcal{O}_k + \mathfrak{p})(\beta_1 + \mathfrak{p}\mathcal{O}_K) + \cdots + (\mathcal{O}_k + \mathfrak{p})(\beta_1 + \mathfrak{p}\mathcal{O}_K) + \mathfrak{p}\mathcal{O}_K = \mathcal{O}_K \end{aligned}$$

and thus $N = \mathfrak{p}N$. This implies that if $\alpha_1, \cdots, \alpha_l$ generate N, then

$$\alpha_j = \sum_{i=1}^{l} a_{ij}\alpha_j$$

with $a_{ij} \in \mathfrak{p}$. Defining a matrix $A = I - [a_{ij}]_{1 \le i,j \le l}$ we have $A[\alpha_1 \ldots \alpha_l]^t = 0$, and

$$0 = \mathrm{adj}(A)A[\alpha_1 \ldots \alpha_l]^t = \det(A)I[\alpha_1 \ldots \alpha_l]^t = [\det(A)\alpha_1 \ldots \det(A)\alpha_l]^t,$$

with $\mathrm{adj}(A)$ the adjugate matrix of A. Thus $\det(A)N = 0$ and $\det(A)\mathcal{O}_K \subseteq M$. Now

$$\det(A) = \det(I - [a_{ij}]_{1 \le i,j \le l}) \equiv 1 \pmod{\mathfrak{p}}$$

because all terms in the expansion of $\det(A)$ are in $\mathfrak{p}$ with the exception of the term corresponding to the entries on the main diagonal of the $l \times l$ matrix I, so $\det(A) \neq 0$. Then

$$K = \det(A)K = \det(A)\mathcal{O}_K k \subseteq (\mathcal{O}_k\beta_1 + \cdots + \mathcal{O}_k\beta_n)k = k\beta_1 + \cdots + k\beta_m,$$

so $\beta_1, \ldots, \beta_m$ span K over k. Hence $\dim(\mathcal{O}_K/\mathfrak{p}\mathcal{O}_K) = n_{K/k}$ and the fundamental identity

$$n = e_1 f_1 + \cdots + e_g f_g$$

follows. This identity implies a very important bound: The number of prime ideals of $\mathcal{O}_K$ lying above any prime ideal of $\mathcal{O}_k$ is bounded by $n_{K/k}$. We could however have obtained this bound with less work if we had not sought the full identity.

If K/k is an extension of number fields, $\mathfrak{p}$ a prime ideal of $\mathcal{O}_k$ and $\mathfrak{P}$ a prime ideal of $\mathcal{O}_K$, we denote by $e(\mathfrak{P}|\mathfrak{p})$ and $f(\mathfrak{P}|\mathfrak{p})$ respectively the ramification index and residue class degree of $\mathfrak{P}$ in K/k. Suppose that $L/K/k$ is a tower of extensions of number fields, $\mathfrak{p}$ a prime ideal of $\mathcal{O}_k$, $\mathfrak{P}$ a prime ideal of $\mathcal{O}_K$ lying above $\mathfrak{p}$, and $\mathfrak{Q}$ a prime ideal of $\mathcal{O}_L$ lying above $\mathfrak{P}$. Then the *tower laws*

$$e(\mathfrak{Q}|\mathfrak{p}) = e(\mathfrak{Q}|\mathfrak{P})e(\mathfrak{P}|\mathfrak{p})$$
$$f(\mathfrak{Q}|\mathfrak{p}) = f(\mathfrak{Q}|\mathfrak{P})f(\mathfrak{P}|\mathfrak{p})$$

hold. For the first law, $\mathfrak{P}$ divides $\mathfrak{p}\mathcal{O}_K$ to the $e(\mathfrak{P}|\mathfrak{p})$-th power and $\mathfrak{Q}$ divides $\mathfrak{P}\mathcal{O}_L$ to the $e(\mathfrak{Q}|\mathfrak{P})$-th power, so $\mathfrak{Q}$ divides $\mathfrak{p}\mathcal{O}_L$ to at least the $e(\mathfrak{Q}|\mathfrak{P})e(\mathfrak{P}|\mathfrak{p})$-th power. Writing

$$\mathfrak{p}\mathcal{O}_K = \mathfrak{P}^{e(\mathfrak{P}|\mathfrak{p})}\mathfrak{N}$$

with $\mathfrak{P} \nmid \mathfrak{N}$, we see that

$$\mathfrak{Q} \mid \mathfrak{N}\mathcal{O}_L \Longrightarrow \mathfrak{Q} \cap \mathcal{O}_K \supseteq \mathfrak{N}\mathcal{O}_L \cap \mathcal{O}_K \Longrightarrow \mathfrak{P} \supseteq \mathfrak{N} \Longrightarrow \mathfrak{P} \mid \mathfrak{N}.$$

Thus $\mathfrak{Q}$ divides $\mathfrak{p}\mathcal{O}_L$ to at most the $e(\mathfrak{Q}|\mathfrak{P})e(\mathfrak{P}|\mathfrak{p})$-th power. The second law follows from the multiplicativity of degree in towers of extensions, applied to $(L/\mathfrak{Q})/(K/\mathfrak{P})/(k/\mathfrak{p})$.

For the rest of this section we make a standing assumption that K/k is a normal extension of number fields. Then the decomposition theory of prime ideals simplifies a good deal. The key fact is that for a normal extension the Galois group acts transitively on the prime ideals lying above a fixed prime ideal.

If $\mathfrak{A}$ is an ideal of $\mathcal{O}_K$ and $\sigma \in G = \mathrm{Gal}(K/k)$, then $\sigma\mathfrak{A}$ is also an ideal of $\mathcal{O}_K$, for σ is an epimorphism. Moreover $\mathcal{O}_K/\sigma\mathfrak{A} = \sigma\mathcal{O}_K/\sigma\mathfrak{A} \cong \mathcal{O}_K/\mathfrak{A}$ since σ is an isomorphism. So the elements of the Galois group take prime ideals to prime ideals, for $\mathfrak{A}$ is a prime ideal if and only if $\mathcal{O}_K/\mathfrak{A}$ is an integral domain. We show that the Galois group of K/k acts transitively on the prime ideals of $\mathcal{O}_K$ lying above a prime ideal in $\mathcal{O}_k$. If $\mathfrak{P}$ lies above $\mathfrak{p}$ and $\sigma \in G$, then $(\sigma\mathfrak{P}) \cap \mathcal{O}_k = \sigma(\mathfrak{P} \cap \mathcal{O}_k) = \sigma\mathfrak{p} = \mathfrak{p}$, for σ restricts to the identity on $\mathcal{O}_k$. Thus the Galois group acts on the set of prime ideals lying above $\mathfrak{p}$. It remains to show that the action is transitive. Suppose $\mathfrak{P}_0$ and $\mathfrak{P}_1$ are distinct prime ideals lying above $\mathfrak{p}$ but that $\mathfrak{P}_0$ is not an image of $\mathfrak{P}_1$ under the action of the Galois group on prime ideals. By the Chinese Remainder Theorem for ideals, there exists some $\alpha \in \mathcal{O}_K$ such that $\alpha \equiv 0 \pmod{\mathfrak{P}_0}$ and $\alpha \equiv 1 \pmod{\sigma\mathfrak{P}_1}$ for $\sigma \in G$. Now

$$N_{K/k}(\alpha) = \prod_{\sigma \in G} \sigma(\alpha) \equiv 0 \pmod{\mathfrak{P}_0}$$

while $N_{K/k}(\alpha) \in \mathcal{O}_k$, so that $N_{K/k}(\alpha) \in \mathfrak{P}_0 \cap \mathcal{O}_k = \mathfrak{p}$. Then $N_{K/k}(\alpha) \in \mathfrak{P}_1$ since $\mathfrak{P}_1$ lies above $\mathfrak{p}$. But this implies that $\sigma(\alpha) \in \mathfrak{P}_1$ for some $\sigma \in G$ since $\mathfrak{P}_1$ is a prime ideal, which leads to a contradiction, for $\alpha \in \sigma^{-1}\mathfrak{P}_1$ while $\alpha \equiv 1 \pmod{\sigma^{-1}\mathfrak{P}_1}$.

Since the Galois group acts transitively on the prime ideals $\mathfrak{P}_j$ lying above $\mathfrak{p}$, the quotients $\mathcal{O}_K/\mathfrak{P}_j$ are isomorphic, thus the residue class degrees f_j are equal, say $f_j = f$ for all j. Furthermore

$$\prod_{j=1}^{g} \mathfrak{P}_j^{e_j} = \mathfrak{p}\mathcal{O}_K = \sigma(\mathfrak{p})\mathcal{O}_K = \sigma(\mathfrak{p}\mathcal{O}_K) = \sigma\left(\prod_{j=1}^{g} \mathfrak{P}_j^{e_j}\right) = \prod_{j=1}^{g} \sigma(\mathfrak{P}_j)^{e_j}$$

for each $\sigma \in G$, since σ restricts to the identity on $\mathcal{O}_k$. Then uniqueness of factorization into powers of prime ideals implies that the ramification indices e_j are equal, say $e_j = e$, since G acts transitively on the prime ideals lying above $\mathfrak{p}$. So finally we see that

$$n = efg$$

when K/k is a normal extension of number fields.

Consider a normal extension K/k of number fields, with Galois group $G = \mathrm{Gal}(K/k)$, and a nonzero prime ideal $\mathfrak{P}$ of $\mathcal{O}_K$. Then

$$Z_{\mathfrak{P}} \stackrel{\text{def}}{=} \{\sigma \in G \mid \sigma\mathfrak{P} = \mathfrak{P}\}$$

is the *decomposition group* of $\mathfrak{P}$ in the extension K/k. Considering the action of G on the set of prime ideals lying above $\mathfrak{P} \cap \mathcal{O}_k$, the orbit-stabilizer theorem shows that $|Z_{\mathfrak{P}}| = ef$, for by the transitivity there is only a single orbit, which has g elements. Moreover

$$\varphi \in Z_{\sigma\mathfrak{P}} \iff \varphi\sigma\mathfrak{P} = \sigma\mathfrak{P} \iff \sigma^{-1}\varphi\sigma\mathfrak{P} = \mathfrak{P} \iff \sigma^{-1}\varphi\sigma \in Z_{\mathfrak{P}},$$

so $Z_{\sigma\mathfrak{P}} = \sigma Z_{\mathfrak{P}} \sigma^{-1}$ for $\sigma \in G$. The ideal $\sigma\mathfrak{P}$ is said to be *conjugate* to the ideal $\mathfrak{P}$. Recall that $\mathcal{O}_K/\mathfrak{P}$ may be regarded as an extension of $\mathcal{O}_k/\mathfrak{p}$. Since each

automorphism σ in $Z_{\mathfrak{P}}$ preserves $\mathfrak{P}$, it induces an automorphism $\widetilde{\sigma}$ of $\mathcal{O}_K/\mathfrak{P}$. This automorphism fixes $\mathcal{O}_k/\mathfrak{p}$ pointwise since σ fixes k pointwise. Thus $\widetilde{\sigma}$ is an element of the Galois group $\mathrm{Gal}((\mathcal{O}_K/\mathfrak{P})/(\mathcal{O}_k/\mathfrak{p}))$.

Theorem of Frobenius. *The mapping*

$$Z_{\mathfrak{P}} \to \mathrm{Gal}((\mathcal{O}_K/\mathfrak{P})/(\mathcal{O}_k/\mathfrak{p}))$$

given by $\sigma \mapsto \widetilde{\sigma}$ is an epimorphism.

Proof. Since

$$\widetilde{(\sigma+\tau)}(\alpha+\mathfrak{P}) = (\sigma+\tau)\alpha+\mathfrak{P} = \sigma\alpha+\mathfrak{P}+\tau\alpha+\mathfrak{P} = \widetilde{\sigma}(\alpha+\mathfrak{P})+\widetilde{\tau}(\alpha+\mathfrak{P})$$

and

$$\widetilde{(\sigma\tau)}(\alpha+\mathfrak{P}) = \sigma\tau\alpha+\mathfrak{P} = \sigma\tau\alpha+\sigma\mathfrak{P}+\mathfrak{P} = \sigma(\tau\alpha+\mathfrak{P})+\mathfrak{P} = \widetilde{\sigma}\widetilde{\tau}(\alpha+\mathfrak{P}),$$

the mapping $\sigma \mapsto \widetilde{\sigma}$ is a homomorphism. It remains to show that it is onto.

Finite extensions of finite fields are simple, so there exists some element $\beta+\mathfrak{P} \in \mathcal{O}_K/\mathfrak{P}$ so that $\mathcal{O}_K/\mathfrak{P} = (\mathcal{O}_k/\mathfrak{p})(\beta+\mathfrak{P})$. If F is the fixed field of $G_{\mathfrak{P}}$, K/F is normal since K/k was assumed normal. Thus all the conjugates of β over F are in K, so the minimal polynomial $m(x)$ of β over F factors as

$$m(x) = \prod_{\sigma \in Z_{\mathfrak{P}}} (x - \sigma\beta).$$

Reducing this polynomial modulo $\mathfrak{P}$ we obtain

$$\widetilde{m}(x) = \prod_{\sigma \in Z_{\mathfrak{P}}} (x - \widetilde{\sigma}(\beta+\mathfrak{P})),$$

and note that the coefficients of $\widetilde{m}(x)$ are in $F/(\mathfrak{P} \cap F)$.

Now $Z_{\mathfrak{P}}$ is transitive on the prime ideals above $\mathfrak{P} \cap F$ while the elements of $Z_{\mathfrak{P}}$ fix $\mathfrak{P}$, so $\mathfrak{P}$ is the only prime ideal of $\mathcal{O}_K$ extending $\mathfrak{P} \cap F$. Thus $n_{K/F} = e(\mathfrak{P}|\mathfrak{P} \cap F)f(\mathfrak{P}|\mathfrak{P} \cap F)$. Moreover g equals the number of distinct prime ideals $\mathfrak{P}_j$ dividing $\mathfrak{p}\mathcal{O}_K$. If $\sigma, \tau \in G$ then

$$\sigma Z_{\mathfrak{P}_\mathrm{j}} = \tau Z_{\mathfrak{P}_\mathrm{j}} \iff \sigma^{-1}\tau Z_{\mathfrak{P}_\mathrm{j}} = Z_{\mathfrak{P}_\mathrm{j}} \iff \sigma^{-1}\tau\mathfrak{P}_\mathrm{j} = \mathfrak{P}_\mathrm{j} \iff \sigma\mathfrak{P}_\mathrm{j} = \tau\mathfrak{P}_\mathrm{j}$$

for each of these prime ideals $\mathfrak{P}_j$. So g equals the index of $G_{\mathfrak{P}}$ in G and thus $g = n_{F/k}$ by Galois theory. But then $n_{K/k} = n_{K/F}n_{F/k}$ and $n_{K/k} = efg$ implies that $n_{K/F} = ef$, so the tower laws show that $e(\mathfrak{P} \cap F|\mathfrak{p})f(\mathfrak{P} \cap F|\mathfrak{p}) = 1$. In particular $F/(\mathfrak{P} \cap F) = k/\mathfrak{p}$, and so $\widetilde{m}(x)$ has coefficients in $k/\mathfrak{p}$.

The minimal polynomial of $\beta+\mathfrak{P}$ over $k/\mathfrak{p}$ divides $\widetilde{m}(x)$, so each conjugate of $\beta+\mathfrak{P}$ over $k/\mathfrak{p}$ is of the form $\widetilde{\sigma}(\beta+\mathfrak{P}) = \sigma\beta+\mathfrak{P}$ for some $\sigma \in Z_{\mathfrak{P}}$. But the elements of $\mathrm{Gal}((\mathcal{O}_K/\mathfrak{P})/(\mathcal{O}_k/\mathfrak{p}))$ are determined by their action on $\beta+\mathfrak{P}$. So each such element is of the form $\widetilde{\sigma}$ for some $\sigma \in G_{\mathfrak{P}}$. □

The kernel

$$T_{\mathfrak{P}} \stackrel{\text{def}}{=} \ker(\sigma \mapsto \widetilde{\sigma})$$

of the epimorphism is called the *inertia group* of $\mathfrak{P}$ in the extension K/k. Reduction modulo $\mathfrak{P}$ gives a canonical isomorphism

$$Z_{\mathfrak{P}}/T_{\mathfrak{P}} \to \mathrm{Gal}((\mathcal{O}_K/\mathfrak{P})/(\mathcal{O}_k/\mathfrak{p})).$$

Consider a tower $L/K/k$ of normal extensions of number fields. Any element $\sigma \in \mathrm{Gal}(L/k)$ restricts to an element of $\mathrm{Gal}(K/k)$. For $\sigma(K) = K$ since K/k is normal. Suppose that $\sigma \in Z_{\mathfrak{Q}}$ with $\mathfrak{Q}$ a prime ideal of $\mathcal{O}_L$ lying above a prime ideal $\mathfrak{P}$ of $\mathcal{O}_K$. Then $\mathfrak{P}\mathcal{O}_L \subseteq \mathfrak{Q}$, so $\sigma\mathfrak{P}\mathcal{O}_L \subseteq \sigma\mathfrak{Q} = \mathfrak{Q}$. But $\sigma\mathfrak{P}\mathcal{O}_L = \mathfrak{P}\mathcal{O}_L$ since $\mathfrak{P}$ is the only prime ideal lying under $\mathfrak{Q}$, so the restriction of σ to K fixes $\mathfrak{P}$. The restriction homomorphism $Z_{\mathfrak{Q}} \to Z_{\mathfrak{P}}$ for $\mathfrak{Q}$ a prime ideal lying above a prime ideal $\mathfrak{P}$ is an epimorphism. For if $\tau \in \mathrm{Gal}(K/k)$ with $\tau\mathfrak{P} = \mathfrak{P}$, choose $\sigma \in \mathrm{Gal}(L/k)$ as one of the automorphisms of L over k that extend τ. Then $\sigma\mathfrak{Q} = \mathfrak{Q}'$ where $\mathfrak{Q}'$ is one of the prime ideals lying above $\tau\mathfrak{P} = \mathfrak{P}$. Furthermore $\mathrm{Gal}(L/K)$ is transitive on the prime ideals lying above $\mathfrak{P}$, so there is some $\omega \in \mathrm{Gal}(L/K)$ with $\omega\mathfrak{Q}' = \mathfrak{Q}$. Then $\omega\sigma\mathfrak{Q} = \omega\mathfrak{Q}' = \mathfrak{Q}$ so $\omega\sigma \in Z_{\mathfrak{Q}}$, while the restriction of $\omega\sigma$ to K equals the restriction τ of σ to K since ω equals the identity on K.

Suppose that $\sigma \in T_{\mathfrak{Q}}$ with $\mathfrak{Q}$ a prime ideal of $\mathcal{O}_L$ lying above a prime ideal $\mathfrak{P}$ of $\mathcal{O}_K$. Then $\widetilde{\sigma}$ is the identity on $\mathcal{O}_L/\mathfrak{Q}$, so $\sigma\alpha - \alpha \in \mathfrak{Q}$ for all $\alpha \in \mathcal{O}_L$. Now if $\beta \in \mathcal{O}_K$ then $\sigma\beta \in \mathcal{O}_K$, so $\sigma\beta - \beta \in \mathfrak{Q} \cap \mathcal{O}_K = \mathfrak{P}$. Thus the element σ in $T_{\mathfrak{Q}}$ restricts to an element in $T_{\mathfrak{P}}$. The restriction homomorphism $T_{\mathfrak{Q}} \to T_{\mathfrak{P}}$ is moreover an epimorphism. For if $\tau \in T_{\mathfrak{P}}$, choose an $\omega\sigma \in Z_{\mathfrak{Q}}$ that extends τ as above. Then

$$I = \{\tau\alpha - \alpha \in \mathcal{O}_K \mid \alpha \in \mathcal{O}_K\} \subseteq \mathfrak{P}$$

and

$$I \subseteq J = \{\omega\sigma\beta - \beta \in \mathcal{O}_L \mid \beta \in \mathcal{O}_L\}.$$

Furthermore $I\mathcal{O}_L = J$, so $I = J\mathcal{O}_L \subseteq \mathfrak{P}\mathcal{O}_L \subseteq \mathfrak{Q}$. Then $\omega\sigma\beta - \beta \in \mathfrak{Q}$ for $\beta \in \mathcal{O}_L$, so $\widetilde{\omega\sigma}$ is the identity on $\mathcal{O}_L/\mathfrak{Q}$. This implies that the restriction homomorphism is an epimorphism, for τ is the restriction of $\omega\sigma$.

This concludes our scanty account of Hilbert theory - the theory of the decomposition of prime ideals in finite normal extensions of number fields.

The different

The *Dedekind complement* of a fractional ideal $\mathfrak{b}$ of K relative to $\mathcal{O}_k$ is

$$\mathfrak{b}' = \{\alpha \in K \mid \mathrm{tr}_{K/k}(\alpha\beta) \in \mathcal{O}_k \text{ for } \beta \in \mathfrak{b}\}.$$

It is also often called the *codifferent*. We will show that it is again a fractional ideal. The fractional ideal $\mathfrak{b}$ has an ideal basis $\beta_1, \ldots, \beta_n$ over $\mathcal{O}_k$. Express each $\alpha \in \mathfrak{b}'$ as a linear combination

$$\alpha = b_1\beta_1 + \cdots + b_n\beta_n$$

in terms of the ideal basis of $\mathfrak{b}$, with coefficients b_j in k. Then

$$\mathrm{tr}_{K/k}(\alpha\beta_j) = b_1\mathrm{tr}_{K/k}(\beta_1\beta_j) + \cdots + b_n\mathrm{tr}_{K/k}(\beta_n\beta_j),$$

and considered as a linear system of equations for $b_1, \ldots, b_n$ this has coefficient determinant

$$\delta = \det([\mathrm{tr}_{K/k}(\beta_i\beta_j)]_{1\leq i,j\leq n}) = \Delta_{K/k}(\beta_1, \ldots, \beta_n) \neq 0.$$

Clearly the determinant of the system is an element of $\mathcal{O}_k$, and the $\mathrm{tr}_{K/k}(\alpha\beta_j)$ likewise, thus

$$\delta\alpha = \delta b_1\beta_1 + \cdots + \delta b_n\beta_n \in \mathcal{O}_K.$$

The Dedekind complement is closed under addition and under multiplication with elements of $\mathcal{O}_K$, so $\delta\mathfrak{b}'$ is an ideal of $\mathcal{O}_K$, which means that $\mathfrak{b}'$ is a fractional ideal.

Next we show that $\mathfrak{b}\mathfrak{b}' = \mathcal{O}'_K$. Suppose $\alpha \in \mathcal{O}_K$, $\beta \in \mathfrak{b}$ and $\gamma \in \mathfrak{b}'$ to be arbitrary, then we have $\alpha\beta \in \mathcal{O}_K$ and so $\mathrm{tr}_{K/k}(\alpha(\beta\gamma)) = \mathrm{tr}_{K/k}((\alpha\beta)\gamma) \in \mathcal{O}_k$. This implies that $\beta\gamma \in \mathcal{O}'_K$ and thus $\mathfrak{b}\mathfrak{b}' \subseteq \mathcal{O}'_K$. In the other direction, suppose $\alpha \in \mathcal{O}'_K$, $\beta \in \mathfrak{b}$ and $\gamma \in \mathfrak{b}^{-1}$ to be arbitrary, then we have $\beta\gamma \in \mathfrak{b}\mathfrak{b}^{-1} = \mathcal{O}_K$ and so $\mathrm{tr}_{K/k}(\beta(\gamma\alpha)) = \mathrm{tr}_{K/k}((\beta\gamma)\alpha) \in \mathcal{O}_k$. This implies that $\gamma\alpha \in \mathfrak{b}'$ and thus $\mathfrak{b}^{-1}\mathcal{O}'_K \subseteq \mathfrak{b}'$.

Since $\mathfrak{b}\mathfrak{b}' = \mathcal{O}'_K = \mathfrak{b}'\mathfrak{b}''$ we may cancel the common factor to conclude that $\mathfrak{b}'' = \mathfrak{b}$. This will come in useful when proving the functional equation of the Dedekind zeta function.

The ideal

$$\mathfrak{d}_{K/k} \stackrel{\text{def}}{=} (\mathcal{O}'_K)^{-1},$$

where the Dedekind complement is taken with respect to $\mathcal{O}_k$, is called the *different* of K/k. Though by its definition it is merely a fractional ideal, it actually is an ideal of $\mathcal{O}_K$. For $1 \in \mathcal{O}'_K$ by the definition of the Dedekind complement, but the inverse of a fractional ideal containing 1 is an ideal, by the definition of the inverse. We may express the Dedekind complement in terms of the different by the formula

$$\mathfrak{b}' = \mathfrak{b}^{-1}\mathfrak{d}_{K/k}^{-1},$$

where $\mathfrak{b}$ is a fractional ideal of K and $\mathfrak{b}'$ is taken with respect to $\mathrm{tr}_{K/k}$ and $\mathcal{O}_k$. This follows from the relationship $\mathfrak{b}\mathfrak{b}' = \mathcal{O}'_K$.

The equivalences

$$\begin{aligned}\alpha \in \mathfrak{d}_{L/k}^{-1} &\iff \mathrm{tr}_{L/k}(\alpha\mathcal{O}_L) \subseteq \mathcal{O}_k \iff \mathrm{tr}_{K/k}(\mathrm{tr}_{L/K}(\alpha\mathcal{O}_L)) \subseteq \mathcal{O}_k \\ &\iff \mathrm{tr}_{K/k}(\mathrm{tr}_{L/K}(\alpha\mathcal{O}_L)\mathcal{O}_K) \subseteq \mathcal{O}_k \iff \mathrm{tr}_{L/K}(\alpha\mathcal{O}_L) \subseteq \mathcal{O}'_K \\ &\iff \mathrm{tr}_{L/K}(\alpha\mathcal{O}_L)\mathfrak{d}_{K/k} \subseteq \mathcal{O}_K \iff \mathrm{tr}_{L/K}(\alpha\mathfrak{d}_{K/k}\mathcal{O}_L) \subseteq \mathcal{O}_K \\ &\iff \alpha\mathfrak{d}_{K/k} \subseteq \mathcal{O}'_L\end{aligned}$$

hold for $L/K/k$ a tower of extensions of number fields, and they imply that $\mathfrak{d}_{L/k}^{-1}\mathfrak{d}_{K/k} = \mathfrak{d}_{L/K}^{-1}$. Thus $\mathfrak{d}_{L/k} = \mathfrak{d}_{L/K}\mathfrak{d}_{K/k}$, which is known as the *tower different theorem.* This may be expressed in terms of a product of ideals in $\mathcal{O}_K$ by interpreting $\mathfrak{d}_{K/k}$ as the extension $\mathfrak{d}_{K/k}\mathcal{O}_K$.

The trace form $\mathrm{tr}_{K/k}(\alpha\beta)$ is a nondegenerate bilinear form, so any basis $\omega_1, \ldots, \omega_n$ of K over k has a complementary basis $\omega'_1, \ldots, \omega'_n$, satisfying

$$\mathrm{tr}_{K/k}(\omega_i\omega'_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases}$$

with respect to $\mathrm{tr}_{K/k}$. Considering an extension $K/\mathbb{Q}$, if $\omega_1, \ldots, \omega_n$ is an integral basis for K, then

$$\begin{aligned}&\Delta_{K/\mathbb{Q}}(\omega_1, \ldots, \omega_n)\Delta_{K/\mathbb{Q}}(\omega'_1, \ldots, \omega'_n) \\ &= (\det([\sigma_i(\omega_j)]_{1\le i,j\le n}) \det([\sigma_l(\omega'_m)]_{1\le l,m\le n}))^2 \\ &= (\det([\sigma_j(\omega_i)]_{1\le i,j\le n}) \det([\sigma_j(\omega'_m)]_{1\le j,m\le n}))^2 \\ &= \det([\sigma_1(\omega_i)\sigma_1(\omega'_m) + \cdots + \sigma_n(\omega_i)\sigma_n(\omega'_m)]_{1\le i,m\le n})^2 \\ &= \det([\mathrm{tr}(\omega_i\omega'_m)]_{1\le i,m\le n})^2 = \det(I_{n\times n})^2 = 1.\end{aligned}$$

We note that $\omega'_1, \ldots, \omega'_n$ is an ideal basis over $\mathcal{O}_k$ for the fractional ideal $\mathcal{O}'_K$.

Supposing that $\mathfrak{a}$ is an ideal of $\mathcal{O}_K$ with an ideal basis $\alpha_1, \ldots, \alpha_n$, and $\omega_1, \ldots, \omega_n$ an integral basis for K, there is a rational integer matrix A such that

$$[\alpha_1 \ldots \alpha_n]^t = A[\omega_1 \ldots \omega_n]^t$$

Now repeat the same argument that was used to show that $N((\alpha)) = |N_{K/\mathbb{Q}}(\alpha)|$, to conclude that $N(\mathfrak{a}) = |\det(A)|$. Then

$$\Delta_{K/\mathbb{Q}}(\alpha_1, \ldots, \alpha_n) = \det(A)^2\Delta_{K/\mathbb{Q}}(\omega_1, \ldots, \omega_n) = N(\mathfrak{a})^2 d_K,$$

and this formula extends to ideal bases of fractional ideals by homogeneity. Applying it to the fractional ideal $\mathcal{O}'_K$ yields

$$\begin{aligned}d_K^{-1} &= \Delta_{K/\mathbb{Q}}(\omega_1, \ldots, \omega_n)^{-1} = \Delta_{K/\mathbb{Q}}(\omega_1^*, \ldots, \omega_n^*) \\ &= N(\mathcal{O}'_K)^2\Delta_{K/\mathbb{Q}}(\omega_1, \ldots, \omega_n) = N(\mathfrak{d}_{K/\mathbb{Q}}^{-1})^2 d_K = N(\mathfrak{d}_{K/\mathbb{Q}})^{-2} d_K\end{aligned}$$

and so $|d_K| = N(\mathfrak{d}_{K/\mathbb{Q}})$. This formula is interesting because it expresses the discriminant, or at least its absolute value, in terms of something that does

not depend on a choice of integral basis for K. At the moment we only have a definition of the discriminant in the case $K/\mathbb{Q}$ and not in the case K/k of relative extensions. For the case of relative extensions, the analogue of an integral basis does not exist in general, and there are other problems in extending the earlier definition. So instead we define the *discriminant* $\mathfrak{d}_{K/k}$ *of a relative extension* K/k as an ideal of $\mathcal{O}_k$ by the formula

$$\mathfrak{d}_{K/k} = N_{K/k}(\mathfrak{d}_{K/k}).$$

Note that if $k = \mathbb{Q}$, then $\mathfrak{d}_{K/\mathbb{Q}} = (d_K)$, and more generally we may view the discriminant as an element of $\mathcal{O}_k$ up to associates if $\mathcal{O}_k$ is a principal ideal domain.

The different $\mathfrak{d}_{K/k}$ gives quite precise information about the ramification of prime ideals of $\mathcal{O}_k$ in an extension K/k of number fields: The *Dedekind different theorem* states that if $\mathfrak{P}$ is a prime ideal upstairs, $\mathfrak{p}$ is the prime ideal lying below $\mathfrak{P}$, and $\mathfrak{P}^e$ is the exact power of $\mathfrak{P}$ dividing $\mathfrak{p}\mathcal{O}_K$, then $\mathfrak{P}^{e-1}|\mathfrak{d}_{K/k}$, with $\mathfrak{P}^{e-1}$ the exact power dividing $\mathfrak{d}_{K/k}$ if and only if e and $N(\mathfrak{P})$ are coprime. This very powerful theorem has the *Dedekind discriminant theorem* as an immediate corollary: A prime ideal $\mathfrak{p}$ in $\mathcal{O}_k$ ramifies in K/k if and only if $\mathfrak{p}|\mathfrak{d}_{K/k}$. This result would be convenient to apply when discussing convergence of the Euler products that define the Artin L-functions, because it implies that in any extension only a finite number of prime ideals ramify. The factors in the Euler products are more complicated in the ramified case than in the unramified case. However, we can do without it, and a proof together with the necessary preparations seems excessively long for this summary, so we forgo it. We must however prove the absolute case $K/\mathbb{Q}$ of the Dedekind discriminant theorem, otherwise our proof of Minkowski's theorem would be incomplete.

An *algebra* A is a vector space with an associative and distributive multiplication. The multiplication need not be commutative. An element $a \in A$ is *nilpotent* if $a^m = 0$ for some positive integer m.

Dedekind discriminant theorem - absolute case. *A rational prime ramifies in $\mathcal{O}_K$ if and only if it divides the discriminant d_K.*

Proof. Consider the finite dimensional algebra $A = \mathcal{O}_K/p\mathcal{O}_K$ over $\mathbb{F}_p$. We claim that p ramifies in $\mathcal{O}_K$ if and only if A has a nontrivial nilpotent. For if

$$p\mathcal{O}_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_g^{e_g},$$

then

$$A \cong (\mathcal{O}_K/\mathfrak{P}_1^{e_1}) \times \cdots \times (\mathcal{O}_K/\mathfrak{P}_g^{e_g})$$

by the Chinese Remainder Theorem. If $e_1 = \cdots = e_g = 1$ then as a ring A is isomorphic to a direct product of fields, but fields have no nontrivial nilpotents. While if $e_j > 1$ for some j there is some $\alpha \in \mathfrak{P}_j$ with $\alpha \notin \mathfrak{P}_j^{e_j}$, and then $\alpha + \mathfrak{P}_j^{e_j}$ is a nontrivial nilpotent in $\mathcal{O}_K/\mathfrak{P}_j^{e_j}$. This gives a nontrivial nilpotent in A by setting all other factors in the direct product equal to zero.

Let $(\omega_1, \cdots, \omega_n)$ be some integral basis for $\mathcal{O}_K$ and $(\underline{\omega}_1, \cdots, \underline{\omega}_n)$ its reduction modulo $p\mathcal{O}_K$. Then $(\underline{\omega}_1, \cdots, \underline{\omega}_n)$ is a basis for A over $\mathbb{F}_p$. But the ordered basis $(\underline{\omega}_1, \cdots, \underline{\omega}_n)$ is related to any ordered basis $(a_1, \ldots, a_n)$ by a nonsingular matrix over $\mathbb{F}_p$, and thus the discriminants of these two bases differ by a factor that is a nonzero square in $\mathbb{F}_p$. So

$$d_K = \Delta(\omega_1, \cdots, \omega_n) \equiv \Delta(\underline{\omega}_1, \cdots, \underline{\omega}_n) \equiv D^2 \Delta(a_1, \ldots, a_n) \pmod{p}$$

with $0 \neq D^2 \in \mathbb{F}_p$, since $\omega_j \equiv \underline{\omega}_j \pmod{p}$. Note that we are using the expression for the discriminant as the square of the determinant of a matrix of traces $\mathrm{tr}(a_i a_j)$.

Next we must relate nilpotents to the discriminant. Just as for field extensions, multiplication in an algebra A yields for each fixed $a \in A$ a linear map $M_a : A \to A$ given by $b \mapsto ab$. We consider its trace $\mathrm{tr}(M_a)$, which is an element of $\mathbf{F}_p$. If a is nilpotent, then $\mathrm{tr}(M_a)$ is zero, for when a is nilpotent the characteristic polynomial of M_a is x^m where m is the dimension of A over $\mathbb{F}_p$. We write $\mathrm{tr}(a) = \mathrm{tr}(M_a)$. If a_1 is a nontrivial nilpotent of A it can be extended to a basis $(a_1, \ldots, a_n)$ for A over $\mathbb{F}_p$. Then each product $a_1 a_j$ is nilpotent by commutativity and so $\mathrm{tr}(a_1 a_j) = 0$ for $1 \leq j \leq n$. The first row of the trace matrix $[\mathrm{tr}(a_i a_j)]_{1 \leq i,j \leq n}$ is zero, so $\Delta(a_1, \ldots, a_n) = 0$.

Suppose that p ramifies in $\mathcal{O}_K$. Then we may take a_1 to be a nontrivial nilpotent, which by the congruence for d_K implies that $p | d_K$.

Suppose that p does not ramify in $\mathcal{O}_K$. Then A is a product of finite fields F_j, and we may choose an ordered basis $(a_1, \ldots, a_n)$ for A over $\mathbb{F}_p$ which is a union of bases for these fields over $\mathbb{F}_p$. If a_i and a_j lie in different fields, then

$$\mathrm{tr}(a_i a_j) = \mathrm{tr}(M_{a_i a_j}) = \mathrm{tr}(M_{a_i} M_{a_j}) = \mathrm{tr}(0) = 0$$

because the image of the linear operator M_{a_j} lies in the kernel of the linear operator M_{a_i}. If a_i and a_j lie in the same field F_j, the trace $\mathrm{tr}(a_i a_j)$ is the trace of the element $a_i a_j$ in F_j over $\mathbb{F}_p$. This is a trace in a finite extension of finite fields, with a primitive element α. Then we can calculate the determinant of the trace matrix in $F_j/\mathbb{F}_p$ by passing to the power basis generated by α and compute the trace matrix $[\mathrm{tr}(\alpha^{l-1}\alpha^{m-1})]$ by means of the formula for the Vandermonde determinant to see that it is nonzero in $\mathbb{F}_p$. Up to a nonzero factor in $\mathbb{F}_p$ the discriminant $\Delta(a_1, \ldots, a_n)$ equals the square of the product of these determinants of trace matrices for the various fields F_j, and is thus itself nonzero modulo p. Then $p \nmid d_K$ follows as above. □

This proof was taken from lecture notes on *Algebraic Numbers* by Ehud de Shalit.

The law of decomposition of prime ideals in an extension K/k is a complete description of how the extensions to $\mathcal{O}_K$ of the prime ideals in $\mathcal{O}_k$ factor into prime ideals, including the ramification indices and residue class degrees. We prove a theorem of J. W. R. Dedekind and E. E. Kummer that is very helpful in determining the law of decomposition of rational primes in number fields.

Dedekind-Kummer theorem. *Let K be a number field, $\mathcal{O}_K = \mathbb{Z}[\alpha]$, and $m(x)$ the minimal polynomial of α over $\mathbb{Z}$. Assume that*

$$m(x) \equiv m_1(x)^{e_1} \cdots m_g(x)^{e_g}$$

modulo a prime p, where $m_1(x), \cdots, m_g(x)$ are monic polynomials over $\mathbb{Z}$, irreducible modulo p, and pairwise distinct over $\mathbb{Z}/(p)$. Then

$$p\mathcal{O}_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_g^{e_g}$$

where $\mathfrak{P}_j = (p, m_j(\alpha))$ is a prime ideal of $\mathcal{O}_K$ with residue class degree $f_j = \deg(m_j)$ for $1 \le j \le g$.

Proof. Consider the homomorphisms

$$\mathbb{Z}[\alpha] \leftarrow \mathbb{Z}[x] \to \mathbb{Z}[x]/(m_j(x)) \to (\mathbb{Z}/(p))[x]/(\widetilde{m}_j(x))$$

given by evaluation, quotienting modulo an ideal, and reduction modulo p, respectively. Here $\widetilde{m}_j(x) \in (\mathbb{Z}/(p))[x]$ is the polynomial obtained by reducing the coefficients of $m_j(x)$ modulo p. If $\beta \in \mathcal{O}_K$ is represented as $\beta = f(\alpha) = g(\alpha)$ with $f(x), g(x) \in \mathbb{Z}[x]$, then $f(x) \equiv g(x) \pmod{m(x)}$ and thus $f(x) \equiv g(x) \pmod{m_j(x)}$. This yields a well defined epimorphism $\tilde{h}_j : \mathcal{O}_K \to (\mathbb{Z}/(p))[x]/(\widetilde{m}_j(x))$. Putting $\mathfrak{P}_j = \ker(\tilde{h}_j)$ we obtain an isomorphism

$$\mathcal{O}_K/\mathfrak{P}_j \to (\mathbb{Z}/(p))[x]/(\widetilde{m}_j(x)),$$

so $\mathfrak{P}_j$ is a prime ideal, since the image is a field. Clearly this prime ideal lies above (p), and the residue class degree is $f_j = [\mathcal{O}_K/\mathfrak{P}_j : \mathbb{Z}/(p)] = \deg(\widetilde{m}_j) = \deg(m_j)$.

Next we determine the prime ideals $\mathfrak{P}_j$. Since

$$\tilde{h}_j(p) = p + (p) + (\widetilde{m}_j(x)) = (\widetilde{m}_j(x))$$

and

$$\tilde{h}_j(m_j(\alpha)) = \widetilde{m}_j(x) + (\widetilde{m}_j(x)) = (\widetilde{m}_j(x))$$

we see that $p \in \ker(\tilde{h}_j)$ and $m_j(\alpha) \in \ker(\tilde{h}_j)$, so $(p, m_j(\alpha)) \subseteq \mathfrak{P}_j$. If on the other hand $\beta \in \mathfrak{P}_j$ then $\beta = f(\alpha)$ with some $f(x) \in \mathbb{Z}[x]$. Reducing $f(x)$ modulo p yields a polynomial $\tilde{f}(x) \in (\mathbb{Z}/(p))[x]$ with $\tilde{f}(x) \in (\widetilde{m}_j(x))$ since $f(\alpha) \in \mathfrak{P}_j$. Since $\tilde{f}(x)$ is divisible by $\widetilde{m}_j(x)$ we have $\tilde{f}(x) = \widetilde{m}_j(x)\tilde{g}(x)$ with some polynomial $g(x) \in \mathbb{Z}[x]$. Then the coefficients of $f(x) - m_j(x)g(x)$ are divisible by p, so $f(\alpha) - m_j(\alpha)g(\alpha) \in p\mathbb{Z}[\alpha]$ and thus $\beta \in (p, m_j(\alpha))$ so that $\mathfrak{P}_j \subseteq (p, m_j(\alpha))$.

The prime ideals $\mathfrak{P}_j$ are pairwise distinct, for the quotients $(\mathbb{Z}/(p))[x]/(\widetilde{m}_j(x))$ are pairwise distinct since the irreducible monic polynomials $m_j(x)$ are pairwise distinct modulo p. Supposing that $\mathfrak{P}$ is a prime ideal of $\mathcal{O}_K$ with $\mathfrak{P}|p\mathcal{O}_K$, the finite field $\mathcal{O}_K/\mathfrak{P}$ is an extension of $\mathbb{Z}/(p)$, so there exists some monic polynomial $l(x) \in \mathbb{Z}[x]$ with irreducible reduction $\tilde{l}(x)$ modulo p, for which there is an epimorphism $\tilde{h} : \mathcal{O}_K \to (\mathbb{Z}/(p))[x]/(\tilde{l}(x))$ well defined by

$$\mathbb{Z}[\alpha] \leftarrow \mathbb{Z}[x] \to \mathbb{Z}[x]/(l(x)) \to (\mathbb{Z}/(p))[x]/(\widetilde{l}(x))$$

as above, with $\mathfrak{P} = \ker(\tilde{h})$. Then $\mathfrak{P} = m(\alpha) + \mathfrak{P} \mapsto \tilde{m}(x) + (\tilde{l}(x))$, so $\tilde{m}(x) \in (\tilde{l}(x))$. Due to irreducibility modulo p this implies that $\tilde{l}(x) = \tilde{m}_j(x)$ for some $1 \leq j \leq g$, so $\mathfrak{P} = \mathfrak{P}_j$. We conclude that

$$p\mathcal{O}_K = \mathfrak{P}_1^{d_1} \cdots \mathfrak{P}_g^{d_g}$$

where $d_1, \ldots, d_g$ are nonnegative integers. Substituting $x = \alpha$ into the congruence $m(x) \equiv m_1(x)^{e_1} \cdots m_g(x)^{e_g} \pmod{p}$ yields $m_1(\alpha)^{e_1} \cdots m_g(\alpha)^{e_g} \in p\mathcal{O}_K$. Furthermore $\mathfrak{P}_j = (p, m_j(\alpha))$ implies $\mathfrak{P}_j^{e_j} \subseteq (p, m_j(\alpha)^{e_j})$ so

$$\mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_g^{e_g} \subseteq (p, m_1(\alpha)^{e_1} \cdots m_g(\alpha)^{e_g}) \subseteq p\mathcal{O}_K = \mathfrak{P}_1^{d_1} \cdots \mathfrak{P}_g^{d_g}.$$

But to contain is to divide, so $d_j \leq e_j$ for $1 \leq j \leq g$. Now

$$d_1 f_1 + \cdots + d_g f_g = n_K = \deg(m) = e_1 f_1 + \cdots + e_g f_g$$

by the fundamental formula relating degree, ramification indices, and residue class degrees. Since all the residue class degrees f_j are positive, we conclude that $d_j = e_j$ for $1 \leq j \leq g$. □

This result has the limitation that the ring of integers of a number field is not always of the form $\mathbb{Z}[\alpha]$. The first counterexample was given by Dedekind.

We shall now find the law of decomposition of rational primes in quadratic number fields K. If $p | d_K$ then p ramifies as the square of a prime ideal of $\mathcal{O}_K$, by the Dedekind discriminant theorem and consideration of degree. This corresponds to $(d_K|p) = 0$ in terms of the Legendre symbol, or the Kronecker symbol if $p = 2$. If $p \nmid d_K$ and d_K is even, p must be odd. In this case we may use the Dedekind-Kummer theorem with $\alpha = \sqrt{d_K}/2$ and $m(x) = x^2 - d_K/4$. If $d_K/4$ is a quadratic residue modulo p, $m(x) \equiv m_1(x)m_2(x)$ modulo p with $m_1(x)$ and $m_2(x)$ monic linear factors. Thus p splits completely into two distinct prime ideals if $(d_K|p) = 1$. If $d_K/4$ is a quadratic nonresidue modulo p, $m(x)$ is an irreducible quadratic polynomial modulo p. Thus p remains inert if $(d_K|p) = -1$. If $p \nmid d_K$ and d_k is odd, we may use the theorem of Kummer with $\alpha = (1 + \sqrt{d_K}/2$ and $m(x) = x^2 - x + (1 - d_K)/4$. If p is odd, 2 is invertible modulo p so

$$m(x) \equiv 0 \iff (x + 1/2)^2 \equiv d_K/4$$

is solvable modulo p if and only if d_K is a quadratic residue modulo p. If it is a quadratic residue modulo p, $m(x) \equiv m_1(x)m_2(x)$ modulo p with $m_1(x)$ and $m_2(x)$ monic linear factors. Thus p splits completely into two distinct prime ideals if $(d_K|p) = 1$. If $d_K/4$ is a quadratic nonresidue modulo p, $m(x)$ is an irreducible quadratic polynomial modulo p. So p remains inert if $(d_K|p) = -1$. If $p = 2$ then

$$m(x) \equiv 0 \iff \frac{1 - d_K}{4} \equiv 0$$

modulo 2. Thus $m(x) = x(x+1)$ if $(1-d_K)/4$ is even, so 2 splits completely into two distinct prime ideals if $(d_K|2) = 1$. While $m(x)$ is an irreducible quadratic

polynomial modulo 2 if $(1 - d_k)/4$ is odd, so 2 remains inert if $(d_K|2) = -1$. The criterion for $p = 2$ is formulated in terms of the Kronecker symbol modulo 2.

The conclusion is that p ramifies, splits completely, or remains inert in the ring of integers of a quadratic number field K according as to whether $(d_K|p) = 0$, $(d_K|p) = 1$ or $(d_K|p) = -1$ respectively.

Frobenius coset and Artin map

Assume that K/k is a normal extension of number fields. Recall that the Galois group of any extension of finite fields is cyclic and has a canonical generator σ_p^m where σ_p is the Frobenius endomorphism and m is the degree of the ground field of the extension over $\mathbb{F}_p$. Pulling this generator in $\mathrm{Gal}((\mathcal{O}_K/\mathfrak{P})/(\mathcal{O}_k/\mathfrak{p}))$ back to $Z_\mathfrak{P}/T_\mathfrak{P}$ by the canonical isomorphism

$$Z_\mathfrak{P}/T_\mathfrak{P} \to \mathrm{Gal}((\mathcal{O}_K/\mathfrak{P})/(\mathcal{O}_k/\mathfrak{p}))$$

yields a uniquely determined coset of $T_\mathfrak{P}$ in $Z_\mathfrak{P}$. We fix some notation and terminology for this coset. It will be denoted $\mathrm{Frob}_\mathfrak{P}$ and called the *Frobenius coset.* Any element of this coset will be denoted by $\sigma_\mathfrak{P}$ and called a *Frobenius element.* When $T_\mathfrak{P}$ is trivial there is only a single Frobenius element, and in that situation we will denote by $\mathrm{Frob}_\mathfrak{P}$ or $\sigma_\mathfrak{P}$ the unique Frobenius element. When $T_\mathfrak{P}$ is trivial, the notation

$$\left[\frac{K/k}{\mathfrak{P}}\right]$$

is commonly used for the Frobenius element, but we shall not use it.

The Frobenius coset is a generator of the cyclic group $Z_\mathfrak{P}/T_\mathfrak{P}$. We would prefer $\mathrm{Frob}_\mathfrak{P}$ to be a unique element in $Z_\mathfrak{P}$ (and thus in the Galois group), rather than a coset. So when is $T_\mathfrak{P}$ trivial? Since the order of $\mathrm{Gal}((\mathcal{O}_K/\mathfrak{P})/(\mathcal{O}_k/\mathfrak{p}))$ equals the residue class degree f and the order of $Z_\mathfrak{P}$ equals ef, it is clear that the order of $T_\mathfrak{P}$ equals the ramification index e of $\mathfrak{P}$ over $\mathfrak{p}$. Recalling that K/k was assumed normal, so the ramification indices are equal, we see that $\mathrm{Frob}_\mathfrak{P}$ is an element rather than a coset in $Z_\mathfrak{P}$ if and only if $\mathfrak{p}$ does not ramify in the extension. But because we need the Frobenius coset when defining the factors corresponding to ramified prime ideals in the Euler products of Artin L-functions, we cannot limit ourselves to the case when $T_\mathfrak{P}$ is trivial.

Since $N(\mathfrak{p}) = |\mathcal{O}_k/\mathfrak{p}| = p^m$ any Frobenius element $\sigma_\mathfrak{P}$ satisfies the congruence

$$\sigma_\mathfrak{P}(\alpha) \equiv \alpha^{N(\mathfrak{p})} \pmod{\mathfrak{P}}$$

for all $\alpha \in \mathcal{O}_K$. But if ω is an element of $\mathrm{Gal}(K/k)$ with

$$\omega(\alpha) \equiv \alpha^{N(\mathfrak{p})} \pmod{\mathfrak{P}}$$

for all $\alpha \in \mathcal{O}_K$, then

$$\omega\mathfrak{P} \equiv \mathfrak{P}^{N(\mathfrak{p})} \equiv 0 \pmod{\mathfrak{P}},$$

and so $\omega\mathfrak{P} = \mathfrak{P}$ since ω takes prime ideals to prime ideals. Thus $\omega \in Z_{\mathfrak{P}}$, and so

$$\widetilde{\omega}(\alpha + \mathfrak{P}) = \omega(\alpha) + \mathfrak{P} = \alpha^{N(\mathfrak{p})} + \mathfrak{P} = \sigma_{\mathfrak{P}}(\alpha) + \mathfrak{P} = \widetilde{\sigma_{\mathfrak{P}}}(\alpha + \mathfrak{P})$$

for all $\alpha \in \mathcal{O}_K$, for some Frobenius element $\sigma_{\mathfrak{P}}$. Thus $\widetilde{\omega} = \widetilde{\sigma_{\mathfrak{P}}}$ and so ω is also a Frobenius element. Thus the congruence

$$\sigma_{\mathfrak{P}}(\alpha) \equiv \alpha^{N(\mathfrak{p})} \pmod{\mathfrak{P}}$$

uniquely determines the Frobenius coset in the Galois group.

Again recalling that K/k was supposed normal, we know that if $\mathfrak{P}_0$ and $\mathfrak{P}_1$ lie above $\mathfrak{p}$, there is some $\sigma \in \mathrm{Gal}(K/k)$ such that $\mathfrak{P}_0 = \sigma\mathfrak{P}_1$. Then

$$\sigma_{\mathfrak{P}_0}(\alpha) \equiv \alpha^{N(\mathfrak{p})} \pmod{\mathfrak{P}_0}$$

implies

$$\sigma^{-1}\sigma_{\mathfrak{P}_0}\sigma(\beta) \equiv \beta^{N(\mathfrak{p})} \pmod{\mathfrak{P}_1}$$

if $\alpha = \sigma\beta$. Thus $\mathrm{Frob}_{\mathfrak{P}_1} = \sigma^{-1}\mathrm{Frob}_{\mathfrak{P}_0}\sigma$ by the uniqueness of the Frobenius coset. So the two Frobenius cosets are conjugate in the Galois group.

Suppose that $L/K/k$ is a tower of extensions of number fields, with L/k a normal extension, and $\mathfrak{Q}$ a prime ideal of $\mathcal{O}_L$ lying over a prime ideal $\mathfrak{P}$ of $\mathcal{O}_K$ lying above a prime ideal $\mathfrak{p}$ of $\mathcal{O}_k$. If $\sigma_{\mathfrak{Q}}$ is a Frobenius element relative to the extension L/k, then

$$\sigma_{\mathfrak{Q}}^{f(\mathfrak{P}/\mathfrak{p})}$$

is a Frobenius element of $\mathfrak{Q}$ relative to the extension L/K. For

$$\sigma_{\mathfrak{Q}}^{f(\mathfrak{P}/\mathfrak{p})}(\alpha) \equiv \alpha^{N(\mathfrak{p})f(\mathfrak{P}/\mathfrak{p})} \equiv \alpha^{N(\mathfrak{P})} \pmod{\mathfrak{Q}}$$

so the statement follows by the uniqueness of the Frobenius coset.

Let $L/K/k$ and $\mathfrak{Q}, \mathfrak{P}, \mathfrak{p}$ be as above, and assume in addition that K/k is normal. Then the restriction to K of any Frobenius element $\sigma_{\mathfrak{Q}}$ relative to L/k is equal to some Frobenius element $\sigma_{\mathfrak{P}}$ relative to K/k. Denote the restriction by σ, which is an element of $\mathrm{Gal}(K/k)$ by the assumed normality. Now

$$\sigma(\alpha) \equiv \alpha^{N(\mathfrak{p})} \pmod{\mathfrak{Q}}$$

for all $\alpha \in \mathcal{O}_K \subseteq \mathcal{O}_L$. But this implies that

$$\sigma(\alpha) \equiv \alpha^{N(\mathfrak{p})} \pmod{\mathfrak{P}}$$

for all $\alpha \in \mathcal{O}_K$, so the statement follows by the uniqueness of the Frobenius coset.

The Frobenius coset carries information about the splitting of prime ideals in extensions. Suppose that K/k is a normal extension of number fields and $\mathfrak{P}$ a prime ideal of $\mathcal{O}_K$ lying above a prime ideal $\mathfrak{p}$ of $\mathcal{O}_k$. The ramification index is $e = |\mathrm{Frob}(\mathfrak{P})|$ and the residue class degree f is equal to the order of $\mathrm{Frob}_{\mathfrak{P}}$. Moreover $efg = n$ shows that $\mathrm{Frob}_{\mathfrak{P}}$ also determines g. In particular $\mathfrak{p}$ splits

completely if and only if $\mathrm{Frob}_{\mathfrak{P}} = \{1\}$, while $\mathfrak{p}$ remains inert if and only if $\mathrm{Frob}_{\mathfrak{P}}$ has order equal to $n_{K/k}$.

From here on we limit ourselves to the unramified case. Consider two normal extensions K/k and L/k and their compositum KL/k. The extension KL/k is normal, for any embedding $\sigma(KL)$ of the compositum KL is the compositum $\sigma(K)\sigma(L)$ of the embeddings of K and L induced by restriction. The Galois group $\mathrm{Gal}(KL/k)$ injects into the direct product $\mathrm{Gal}(K/k) \times \mathrm{Gal}(L/k)$ by restriction, and we may identify $\mathrm{Gal}(KL/k)$ with its image in the direct product. Let $\mathfrak{R}$ be a prime ideal in $\mathcal{O}_{KL}$ lying above the prime ideal $\mathfrak{P}$ of $\mathcal{O}_K$ and the prime ideal $\mathfrak{Q}$ of $\mathcal{O}_L$. Then $\mathrm{Frob}_{\mathfrak{R}} = \mathrm{Frob}_{\mathfrak{P}} \times \mathrm{Frob}_{\mathfrak{Q}}$, for the restriction of $\mathrm{Frob}_{\mathfrak{R}}$ to K equals $\mathrm{Frob}_{\mathfrak{P}}$ and to L equals $\mathrm{Frob}_{\mathfrak{Q}}$, as above.

A prime ideal $\mathfrak{p}$ splits completely in a normal extension if and only if the Frobenius element of any prime ideal lying above $\mathfrak{p}$ equals the identity. Thus the above result implies that if a prime ideal splits completely in two finite normal extensions of the same number field, then it splits completely in their compositum. This result proves useful for a rather remarkable application of the Dedekind zeta function.

We now consider how the Frobenius element relates to the prime ideal $\mathfrak{p}$ downstairs when the latter is unramified in K/k. We have already seen that in this case the Frobenius elements of prime ideals lying above $\mathfrak{p}$ form a conjugacy class in the Galois group of K/k. Since the conjugacy classes of an abelian group are singletons, the Frobenius element depends only on the prime ideal downstairs if the prime ideal does not ramify and the Galois group is abelian. In this case we denote the Frobenius element by

$$\left(\frac{K/k}{\mathfrak{p}}\right)$$

and call it the *Artin symbol.* The congruence

$$\left(\frac{K/k}{\mathfrak{p}}\right)(\alpha) \equiv \alpha^{N(\mathfrak{p})} \pmod{\mathfrak{P}}$$

holds for $\alpha \in \mathcal{O}_K$ and each prime ideal $\mathfrak{P}$ lying above $\mathfrak{p}$. Thus

$$\left(\frac{K/k}{\mathfrak{p}}\right)(\alpha) \equiv \alpha^{N(\mathfrak{p})} \pmod{\mathfrak{p}\mathcal{O}_K}$$

if $\mathfrak{p}$ is a prime ideal in $\mathcal{O}_k$ that does not ramify in the abelian extension K/k.

A fractional ideal

$$\mathfrak{a} = \mathfrak{p}_1^{a_1} \cdots \mathfrak{p}_s^{a_s}$$

of $\mathcal{O}_k$ factored into powers of distinct prime ideals is said to be *coprime* with a prime ideal $\mathfrak{p}$ of $\mathcal{O}_k$ if $\mathfrak{p}$ does not occur in the factorization. Note that $\mathfrak{p}^{-1}$ is not coprime with $\mathfrak{p}$. This concept extends to $\mathfrak{a}$ being coprime with a set S of prime ideals in the obvious way. A set S of prime ideals in $\mathcal{O}_k$ is called an *exceptional set* for the extension K/k if all prime ideals of $\mathcal{O}_k$ that ramify in the extension K/k are contained in S. The Dedekind discriminant theorem implies that only

a finite number of prime ideals ramify in an extension of number fields, so S may be assumed finite. The set of fractional ideals of $\mathcal{O}_k$ coprime with an exceptional set S for the extension K/k constitutes a subgroup J_k^S of the group of fractional ideals J_k of $\mathcal{O}_k$. If a normal extension K/k of number fields is abelian, i. e. the Galois group of the extension is commutative, the Artin symbol

$$\left(\frac{K/k}{\mathfrak{a}}\right) = \left(\frac{K/k}{\mathfrak{p}_1^{a_1}\cdots\mathfrak{p}_s^{a_s}}\right) = \left(\frac{K/k}{\mathfrak{p}_1}\right)^{a_1}\cdots\left(\frac{K/k}{\mathfrak{p}_s}\right)^{a_s}$$

on fractional ideals $\mathfrak{a}$ in J_k^S is well-defined, with $\mathfrak{a}$ factored as above. This yields a homomorphism

$$\left(\frac{K/k}{}\right) : J_k^S \to \mathrm{Gal}(K/k),$$

which is called the *Artin map*. The determination of the kernel and image of the Artin map is the content of a central theorem of algebraic number theory called the Artin Reciprocity Law. Unfortunately a complete proof requires intricate arguments with cyclotomic extensions, and is a good deal more demanding than anything we do in this Summary. The theorem plays an indispensable part for some results in the analytic theory of number fields.

For simple examples of Artin maps, consider quadratic extensions $K/\mathbb{Q}$. We know that if p is a rational prime not contained in an exceptional set for the extension, there are only two cases: Either (p) splits completely, if $(d_K|p) = 1$, or remains inert if $(d_k|p) = -1$. The Artin map is given by

$$(n) \mapsto \left(\frac{d_K}{|n|}\right)$$

on integral ideals in J_K^S. Here the Galois group of $K/\mathbb{Q}$ is identified with the multiplicative group ± 1.

Recall that the Galois group of the extension $\mathbb{Q}(\sqrt{2},\sqrt{-3})/\mathbb{Q}$ is isomorphic to the direct product $(\mathbb{Z}/2\mathbb{Z})\times(\mathbb{Z}/2\mathbb{Z})$. That $M = \mathbb{Q}(\sqrt{2},\sqrt{-3})$ is the compositum of $K = \mathbb{Q}(\sqrt{2})$ and $L = \mathbb{Q}(\sqrt{-3})$ implies that

$$\left(\frac{M/\mathbb{Q}}{p}\right) = \left(\frac{d_K}{p}\right)\times\left(\frac{d_L}{p}\right) = \left(\frac{8}{p}\right)\times\left(\frac{-3}{p}\right)$$

for rational primes $p \neq 2, 3$.

The cyclotomic extension $\mathbb{Q}(\zeta_5)/\mathbb{Q}$ affords another example. It has a cyclic Galois group of order four, and the discriminant equals 125. The value of the Artin map

$$\left(\frac{\mathbb{Q}(\zeta_5)/\mathbb{Q}}{n}\right)$$

for rational integers $n \not\equiv 0 \pmod 5$ is determined by the assignment $\zeta_5 \mapsto \zeta_5^n$. For the Artin symbol is determined by the congruence

$$\left(\frac{\mathbb{Q}(\zeta_5)/\mathbb{Q}}{p}\right)(\alpha) \equiv \alpha^p \pmod{p\mathcal{O}_{\mathbb{Q}(\zeta_5)}}$$

for rational primes $p \neq 5$, and ζ_5 generates the extension.

This concludes the summary of elementary and algebraic number theory. The reader may well have found the outline of elementary number theory too desiccated, as well as rather hurried and sparse.

The eighth edition of *The Higher Arithmetic* by H. Davenport, edited and with additional material by J. H. Davenport, is highly recommended as an introduction to the central topics of classical number theory. The standard comprehensive treatment is the sixth edition of *An Introduction to the Theory of Numbers* by G. H. Hardy and Edward M. Wright, revised by D. R. Heath-Brown and J. H. Silverman, and with a preface by Andrew Wiles and a new chapter by Silverman. *Introduction to Number Theory* by Hua Loo Keng, translated into English by Peter Shiu and supplemented with notes by Wang Yuan, is also very good and contains a wealth of material. The second edition of *A Classical Introduction to Modern Number Theory* by Kenneth Ireland and Michael Rosen is an excellent exposition of elementary number theory with a selection of further topics from algebraic number theory and Diophantine analysis. Unlike the other treatments of elementary number theory mentioned, this one requires some knowledge of abstract algebra.

The outline of algebraic number theory is very sketchy. It only covers background material for the chapter on the analytic theory of number fields in my *A Course in Analytic Number Theory,* and has scarcely any examples.

Algebraic number theory by A. Fröhlich and M. J. Taylor is a graduate-level introduction with a generous selection of examples. *Number Fields* by Daniel Marcus has many exercises. *Algebraic Number Theory* by Jürgen Neukirch, translated into English by Norbert Schappacher, is both more comprehensive and more demanding, and serves as a modern reference. It covers the ideal-theoretic and valuation-theoretic approaches to algebraic number theory in depth, and has extensive treatments of class field theory and of analytic methods. *Algebraic Number Theory* by S. Lang is another excellent reference.

Lectures on the Theory of Algebraic Numbers by Erich Hecke, translated into English by George U. Brauer and Jay R. Goldman with the assistance of R. Kotzen, is a classic treatise on algebraic number theory along ideal-theoretic lines that is well worth a few comments. Up until analytic methods appear in the last three chapters, only knowledge of polynomial algebra is required to read this book. Even the theory of finitely generated abelian groups is developed *ab initio.* Hecke's treatise has been commended by Atle Selberg, Harold M. Stark and André Weil as an excellent exposition. Even today, ninety years after it first appeared, it is still of interest, especially the last three chapters dealing with analytic methods. But it is not suitable as a reference, partly because there is no index.

The proceedings *Algebraic Number Theory* of the 1965 instructional conference at Brighton, edited by J. W. S. Cassels and A. Fröhlich, is a more modern classic. It is a collection of notes of lectures by distinguished algebraists and number theorists, mainly on topics related to the valuation-theoretic approach to algebraic number theory and class field theory. Also included is the thesis of Tate applying valuation theory and abstract harmonic analysis to the functional equations of L-functions.

List of Notations

$\mathfrak{a}$	conventional notation for an ideal, page 40
$\mathfrak{b}'$	Dedekind complement of a fractional ideal $\mathfrak{b}$, page 55
$\mathrm{char}(F)$	the characteristic of a field F, page 46
$\deg(\mathfrak{p})$	residue class degree of $\mathfrak{p}$ over $\mathbb{Q}$, page 49
$\mathfrak{d}_{K/k}$	the different ideal, page 55
d	conventional notation for a divisor, page 4
e_j	ramification index, page 49
$\mathbb{F}_p$	finite field with p elements up to isomorphism, page 46
$\mathrm{Frob}_{\mathfrak{P}}$	Frobenius coset or element, page 61
f_j	residue class degree, page 49
$\mathrm{Gal}(K/k)$	Galois group of K over k, page 29
$\gcd(a, b)$	greatest common divisor of a and b, page 6
g	conventional notation for a primitive root, page 12
$\mathrm{ind}_g(a)$	index of a to base g, page 14
J_K	the group of fractional ideals of $\mathcal{O}_K$, page 43
J_K^S	the group of fractional ideals coprime with S, page 64
$N(\mathfrak{a})$	the norm of an ideal $\mathfrak{a}$, page 43
$k(\alpha)$	simple extension of k by α, page 25
KL	compositum of K and L, page 25
$\mathrm{lcm}[a, b]$	least common multiple of a and b, page 6
$L/K/k$	tower of field extensions, page 27
K/k	conventional notation for a field extension, page 25
m	conventional notation for a modulus in a congruence, page 9
$m(x)$	minimal polynomial of an algebraic element, page 25
$\mathbb{N}$	the set of positive integers $1, 2, \ldots$, page 1
$\mathbb{N}_0$	the set of nonnegative integers $0, 1, 2, \ldots$, page 1

$n_{K/k}$	degree of K over k, page 25
$N_{K/k}(\alpha)$	the norm of α in K over k, page 30
n_K	degree of K over $\mathbb{Q}$, page 25
$N_K(\alpha)$	the norm of α in K over $\mathbb{Q}$, page 30
$\mathcal{O}_K$	the ring of algebraic integers in K, page 32
$\mathfrak{P}$	conventional notation for a prime ideal, page 41
$\mathfrak{p}$	conventional notation for a prime ideal, page 41
p	conventional notation for a prime, page 2
P_K	the group of principal fractional ideals of $\mathcal{O}_K$, page 43
q	conventional notation for a modulus in a congruence, page 9
r_1	the number of real embeddings, page 27
r_2	the number of conjugate pairs of complex embeddings, page 27
S	an exceptional set of prime ideals, page 63
s	conventional notation for a squarefree integer, page 5
$\mathrm{tr}_{K/k}(\alpha)$	the trace of α in K over k, page 31
w_K	the number of units of $\mathcal{O}_K$ of finite order, page 37
$\mathbb{Z}$	the ring of integers $\ldots, -2, -1, 0, 1, 2, \ldots$, page 1
$T_{\mathfrak{P}}$	inertia group, page 54
$Z_{\mathfrak{P}}$	decomposition group, page 52
$\phi(m)$	number of residue classes modulo m coprime with m, page 10
$\Phi_q(x)$	the q-th cyclotomic polynomial, page 36
σ_p	the Frobenius endomorphism, page 46
$\sigma_{\mathfrak{P}}$	Frobenius element, page 61
ε_d	fundamental unit of the ring of integers of $\mathbb{Q}(\sqrt{d})$, page 38
(a, b)	abbreviated form for greatest common divisor of a and b, page 6
$[a, b]$	abbreviated form for least common multiple of a and b, page 6
$[K : k]$	degree of K over k, page 25
$\equiv$	congruence relation $a \equiv b \pmod{q}$ meaning q divides $a - b$, page 9
$\underline{a}$	residue class of a modulo m, page 9

Index