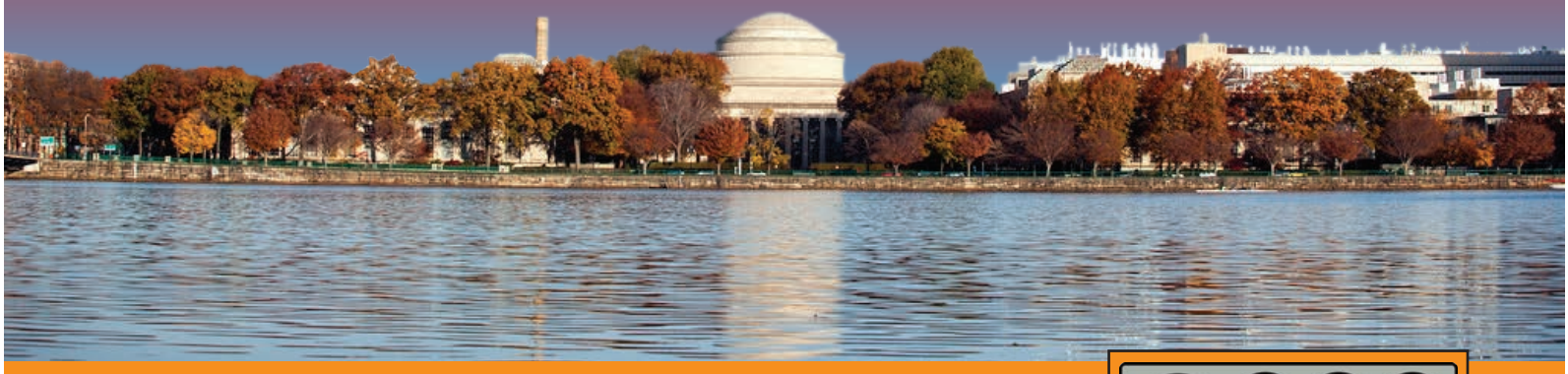
MITOPENCOURSEWARE

Number Theory I

by Andrew Sutherland



1 Absolute values and discrete valuations

1.1 Introduction

At its core, number theory is the study of the integer ring $\mathbb{Z}$. By the fundamental theorem of arithmetic, every element of $\mathbb{Z}$ can be written uniquely as a product of primes (up to multiplication by a unit ± 1), so it is natural to focus on the prime elements of $\mathbb{Z}$. If p is a prime, the ideal $(p) \coloneqq p\mathbb{Z}$ it generates is a maximal ideal ($\mathbb{Z}$ has Krull dimension one), and the residue field $\mathbb{Z}/p\mathbb{Z}$ is the finite field $\mathbb{F}_p$ with p elements (unique up to isomorphism). The fraction field of $\mathbb{Z}$ is the field $\mathbb{Q}$ of rational numbers. The field $\mathbb{Q}$ and the finite fields $\mathbb{F}_p$ together make up the prime fields: every field k contains exactly one of them, according to its characteristic: k has characteristic zero if and only if it contains $\mathbb{Q}$, and k has characteristic p if and only if k contains $\mathbb{F}_p$.

The structure of the ring $\mathbb{Z}$ and the distribution of its primes are both intimately related to properties of the Riemann zeta function

$$\zeta(s) = \sum n^{-s} = \prod_p (1 - p^{-s})^{-1}.$$

As a function of the complex variable s, the Riemann zeta function is holomorphic and nonvanishing on $\mathrm{Re}(s) > 1$ and admits an analytic continuation to the entire complex plane. It has a simple pole at $s = 1$, which implies that there are infinitely many primes (otherwise the product over primes on the RHS would be finite and converge). The distribution of its zeros in the *critical strip* $0 < s < 1$ is directly related to the distribution of primes (via the *explicit formula*, which we will see later in the course). As you are probably aware, Riemann famously conjectured more than 150 years ago that the zeros of $\zeta(s)$ in the critical strip all lie on the *critical line* $\mathrm{Re}(s) = \frac{1}{2}$; this conjecture remains open.

One can also consider finite extensions of $\mathbb{Q}$, such as the field $\mathbb{Q}(i) \coloneqq \mathbb{Q}[x]/(x^2 + 1)$. These are called *number fields*, and each can be constructed as the quotient of the polynomial ring $\mathbb{Q}[x]$ by one of its maximal ideals; the ring $\mathbb{Q}[x]$ is a principal ideal domain and its maximal ideals can all be written as (f) for some monic irreducible $f \in \mathbb{Z}[x]$. Associated to each number field K is a zeta function $\zeta_K(s)$, and each of these has an associated conjecture regarding the location of its zeros (these conjectures all remain open).

Number fields are one of two types of *global fields* that we will spend much of the course studying; the other type are known as *global function fields*. Let $\mathbb{F}_q$ denote the field with q elements, where q is any prime power. The polynomial ring $\mathbb{F}_q[t]$ has much in common with the integer ring $\mathbb{Z}$. Like $\mathbb{Z}$, it is a principal ideal domain of dimension one, and the residue fields $\mathbb{F}_q[t]/(f)$ one obtains by taking the quotient by a maximal ideal (f), where $f \in \mathbb{F}_q[t]$ is any irreducible polynomial, are finite fields $\mathbb{F}_{q^d}$, where d is the degree of f. In contrast to the situation with $\mathbb{Z}$, the residue fields of $\mathbb{F}_q[t]$ all have the same characteristic as its fraction field $\mathbb{F}_q(t)$, which plays a role analogous to $\mathbb{Q}$. Global function fields are finite extensions of $\mathbb{F}_q(t)$ (this includes $\mathbb{F}_q(t)$ itself, an extension of degree 1).

Associated to each global field k is an infinite collection of *local fields* corresponding to the completions of k with respect to its absolute values; when $k = \mathbb{Q}$, these completions are the field of real numbers $\mathbb{R}$ and the p-adic fields $\mathbb{Q}_p$ (as you will prove on Problem Set 1).

The ring $\mathbb{Z}$ is a principal ideal domain (PID), as is $\mathbb{F}_q[t]$, and in such fields every nonzero prime ideal is maximal and thus has an associated *residue field*. For both $\mathbb{Z}$ and $\mathbb{F}_q[t]$ these residue fields are finite, but the characteristics of the residue fields of $\mathbb{Z}$ are all different (and distinct from the characteristic of its fraction field), while those of $\mathbb{F}_q[t]$ are all the same.

We will spend the first part of this course fleshing out this picture, in which we are particularly interested in understanding the integral closure of the rings $\mathbb{Z}$ and $\mathbb{F}_q[t]$ in finite extensions of their fraction fields (such integral closures are known as *rings of integers*), and the prime ideals of these rings. Where possible we will treat number fields and function fields on an equal footing, but we will also note some key differences. Surprisingly, the apparently more complicated function field setting often turns out to be simpler than the number field setting; for example, the analog of the Riemann hypothesis in the function field setting (the Riemann hypothesis for curves), is not an open problem. It was proved by André Weil in the 1940s [5]; a further generalization to varieties of arbitrary dimension was proved by Pierre Deligne in the 1970s [3].

Zeta functions provide the tool we need to understand the distribution of primes, both in general, and within particular residue classes; the proofs of the prime number theorem and Dirichlet's theorem on primes in arithmetic progressions both use zeta functions in an essential way. Dirichlet's theorem states that for each integer $m > 1$ and each integer a coprime to m, there are infinitely many primes $p \equiv a \bmod m$. In fact, more is true: the Chebotarev density theorem tells us that for each modulus m the primes are equidistributed among the residue classes of the integers a coprime to m. We will see this and several other applications of the Chebotarev density theorem in the later part of the course.

Before we begin, let us note the following.

Remark 1.1. Our rings always have a multiplicative identity that is preserved by ring homomorphisms (so the zero ring in which $1 = 0$ is not an initial object in the category of rings, but it is the terminal object in this category). Except where noted otherwise, the rings we shall consider are all commutative.

1.2 Absolute values

We begin with the general notion of an absolute value on a field; a reference for much of this material is [4, Chapter 1].

Definition 1.2. An *absolute value* on a field k is a map $|\ |\colon k \to \mathbb{R}_{\geq 0}$ such that for all $x, y \in k$ the following hold:

1. $|x| = 0$ if and only if $x = 0$;
2. $|xy| = |x||y|$;
3. $|x + y| \leq |x| + |y|$.

If the stronger condition

4. $|x + y| \leq \max(|x|, |y|)$

also holds, then the absolute value is *nonarchimedean*; otherwise it is *archimedean*.

Example 1.3. The map $|\ |\colon k \to \mathbb{R}_{\geq 0}$ defined by

$$|x| = \begin{cases} 1 & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

is the *trivial absolute value* on k. It is nonarchimedean.

Lemma 1.4. *An absolute value $|\ |$ on a field k is nonarchimedean if and only if*

$$|\underbrace{1+\cdots+1}_{n}| \le 1$$

for all $n \ge 1$.

Proof. See Problem Set 1. □

Corollary 1.5. *In a field of positive characteristic every absolute value is nonarchimedean, and the only absolute value on a finite field is the trivial one.*

Definition 1.6. Two absolute values $|\ |$ and $|\ |'$ on the same field k are *equivalent* if there exists an $\alpha \in \mathbb{R}_{>0}$ for which $|x|' = |x|^\alpha$ for all $x \in k$.

1.3 Absolute values on $\mathbb{Q}$

To avoid confusion we will denote the usual absolute value on $\mathbb{Q}$ (inherited from $\mathbb{R}$) by $|\ |_\infty$; it is an archimedean absolute value. But there are are infinitely many others. Recall that any element of $\mathbb{Q}^\times$ may be written as $\pm\prod_q q^{e_q}$, where the product ranges over primes and the exponents $e_q \in \mathbb{Z}$ are uniquely determined (as is the sign).

Definition 1.7. For a prime p the *p-adic valuation* $v_p\colon \mathbb{Q} \to \mathbb{Z}$ is defined by

$$v_p\left(\pm\prod_q q^{e_q}\right) \coloneqq e_p,$$

and we define $v_p(0) \coloneqq \infty$. The *p-adic absolute value* on $\mathbb{Q}$ is defined by

$$|x|_p \coloneqq p^{-v_p(x)},$$

where $|0|_p = p^{-\infty}$ is understood to be 0.

Theorem 1.8 (Ostrowski's Theorem)**.** *Every nontrivial absolute value on $\mathbb{Q}$ is equivalent to $|\ |_p$ for some $p \le \infty$.*

Proof. See Problem Set 1. □

Theorem 1.9 (Product Formula)**.** *For every $x \in \mathbb{Q}^\times$ we have*

$$\prod_{p\le\infty} |x|_p = 1.$$

Proof. See Problem Set 1. □

1.4 Discrete valuations

Definition 1.10. A *valuation* on a field k is a group homomorphism $k^\times \to \mathbb{R}$ such that for all $x, y \in k$ we have

$$v(x+y) \ge \min(v(x), v(y)).$$

We may extend v to a map $k \to \mathbb{R} \cup \{\infty\}$ by defining $v(0) \coloneqq \infty$. For any any $0 < c < 1$, defining $|x|_v \coloneqq c^{v(x)}$ yields a nonarchimedean absolute value. The image of v in $\mathbb{R}$ is the

value group of v. We say that v is a *discrete valuation* if its value group is equal to $\mathbb{Z}$ (every discrete subgroup of $\mathbb{R}$ is isomorphic to $\mathbb{Z}$, so we can always rescale a valuation with a discrete value group so that this holds). Given a field k with valuation v, the set

$$A \coloneqq \{x \in k : v(x) \geq 0\},$$

is the *valuation ring* of k (with respect to v). A *discrete valuation ring* (DVR) is an integral domain that is the valuation ring of its fraction field with respect to a discrete valuation; such a ring A cannot be a field, since $v(\operatorname{Frac} A) = \mathbb{Z} \neq \mathbb{Z}_{\geq 0} = v(A)$.

It is easy to verify that every valuation ring A is a in fact a ring, and even an integral domain (if x and y are nonzero then $v(xy) = v(x) + v(y) \neq \infty$, so $xy \neq 0$), with k as its fraction field. Notice that for any $x \in k^\times$ we have $v(1/x) = v(1) - v(x) = -v(x)$, so at least one of x and $1/x$ has nonnegative valuation and lies in A. It follows that $x \in A$ is invertible (in A) if and only if $v(x) = 0$, hence the unit group of A is

$$A^\times = \{x \in k : v(x) = 0\},$$

We can partition the nonzero elements of k according to the sign of their valuation. Elements with valuation zero are units in A, elements with positive valuation are non-units in A, and elements with negative valuation do not lie in A, but their multiplicative inverses are non-units in A. This leads to a more general notion of a valuation ring.

Definition 1.11. A *valuation ring* is an integral domain A with fraction field k with the property that for every $x \in k$, either $x \in A$ or $x^{-1} \in A$.

Let us now suppose that the integral domain A is the valuation ring of its fraction field with respect to some discrete valuation v (which we shall see is uniquely determined). Any element $\pi \in A$ for which $v(\pi) = 1$ is called a *uniformizer*. Uniformizers exist, since $v(A) = \mathbb{Z}_{\geq 0}$. If we fix a uniformizer π, every $x \in k^\times$ can be written uniquely as

$$x = u\pi^n$$

where $n = v(x)$ and $u = x/\pi^n \in A^\times$ and uniquely determined. It follows that A is a unique factorization domain (UFD), and in fact A is a principal ideal domain (PID). Indeed, every nonzero ideal of A is equal to

$$(\pi^n) = \{a \in A : v(a) \geq n\},$$

for some integer $n \geq 0$. Moreover, the ideal (π^n) depends only on n, not the choice of uniformizer π: if π' is any other uniformizer its unique representation $\pi' = u\pi^1$ differs from π only by a unit. The ideals of A are thus totally ordered, and the ideal

$$\mathfrak{m} = (\pi) = \{a \in A : v(a) > 0\}$$

is the unique maximal ideal of A (and also the only nonzero prime ideal of A).

Definition 1.12. A *local ring* is a commutative ring with a unique maximal ideal.

Definition 1.13. The *residue field* of a local ring A with maximal ideal $\mathfrak{m}$ is the field $A/\mathfrak{m}$.

We can now see how to determine the valuation v corresponding to a discrete valuation ring A. Given a discrete valuation ring A with unique maximal ideal $\mathfrak{m}$, we may define $v\colon A \to \mathbb{Z}$ by letting $v(a)$ be the unique integer n for which $(a) = \mathfrak{m}^n$ and $v(0) \coloneqq \infty$. Extending v to the fraction field k of A via $v(a/b) \coloneqq v(a) - v(b)$ gives a discrete valuation v on k for which $A = \{x \in k : v(x) \geq 0\}$ is the corresponding valuation ring.

Notice that any discrete valuation v on k with A as its valuation ring must satisfy $v(\pi) = 1$ for some $\pi \in \mathfrak{m}$ (otherwise $v(k) \neq \mathbb{Z}$), and we then have $v(\pi) = 1$ if and only if $\mathfrak{m} = (\pi)$. Moreover, v must then coincide with the discrete valuation we just defined: for any DVR A, the discrete valuation on the fraction field of A that yields A as its valuation ring is uniquely determined. It follows that we could have defined a uniformizer to be any generator of the maximal ideal of A without reference to a valuation.

Example 1.14. For the p-adic valuation $v_p\colon \mathbb{Q} \to \mathbb{Z} \cup \{\infty\}$ we have the valuation ring

$$\mathbb{Z}_{(p)} \coloneqq \left\{ \frac{a}{b} : a, b \in \mathbb{Z}, p \nmid b \right\},$$

with maximal ideal $\mathfrak{m} = (p)$; this is the *localization* of the ring $\mathbb{Z}$ at the prime ideal (p). The residue field is $\mathbb{Z}_{(p)}/p\mathbb{Z}_{(p)} \simeq \mathbb{Z}/p\mathbb{Z} \simeq \mathbb{F}_p$.

Example 1.15. For any field k, the valuation $v\colon k((t)) \to \mathbb{Z} \cup \{\infty\}$ on the field of Laurent series over k defined by

$$v\left(\sum_{n \geq n_0} a_n t^n\right) = n_0,$$

where $a_{n_0} \neq 0$, has valuation ring $k[[t]]$, the power series ring over k. For $f \in k((t))^\times$, the valuation $v(f) \in \mathbb{Z}$ is the *order of vanishing* of f at zero. For every $\alpha \in k$ one can similarly define a valuation v_α on k as the order of vanishing of f at α by taking the Laurent series expansion of f about α.

1.5 Discrete Valuation Rings

Discrete valuation rings are in many respects the nicest rings that are not fields. In addition to being an integral domain, every discrete valuation ring A enjoys the following properties:

- *noetherian*: Every increasing sequence $I_1 \subseteq I_2 \subseteq \cdots$ of ideals eventually stabilizes; equivalently, every ideal is finitely generated.
- *principal ideal domain*: Every ideal is principal (generated by a single element).
- *local*: There is a unique maximal ideal $\mathfrak{m}$.
- *dimension one*: The (Krull) *dimension* of a ring R is the supremum of the lengths n of all chains of prime ideals $\mathfrak{p}_0 \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_n$ (which need not be finite, in general). For DVRs, $(0) \subseteq \mathfrak{m}$ is the longest chain of prime ideals, with length 1.
- *regular*: The dimension of the $A/\mathfrak{m}$-vector space $\mathfrak{m}/\mathfrak{m}^2$ is equal to the dimension of A. Non-local rings are regular if this holds for every localization at a prime ideal.
- *integrally closed* (or *normal*): Every element of the fraction field of A that is the root of a monic polynomial in $A[x]$ lies in A.
- *maximal*: There are no intermediate rings strictly between A and its fraction field.

Various combinations of these properties can be used to uniquely characterize discrete valuation rings (and hence give alternative definitions).

Theorem 1.16. *For an integral domain A, the following are equivalent:*

- *A is a DVR.*
- *A is a noetherian valuation ring that is not a field.*
- *A is a local PID that is not a field.*
- *A is an integrally closed noetherian local ring of dimension one.*
- *A is a regular noetherian local ring of dimension one.*
- *A is a noetherian local ring whose maximal ideal is nonzero and principal.*
- *A is a maximal noetherian ring of dimension one.*

Proof. See [1, §23] or [2, §9]. □

1.6 Integral extensions

Integrality plays a key role in number theory, so it is worth discussing it in more detail.

Definition 1.17. Given a ring extension $A \subseteq B$, an element $b \in B$ is *integral over A* if is a root of a monic polynomial in $A[x]$. The ring B is *integral over A* if all its elements are.

Proposition 1.18. *Let $\alpha, \beta \in B$ be integral over $A \subseteq B$. Then $\alpha + \beta$ and $\alpha\beta$ are integral over A.*

Proof. Let $f \in A[x]$ and $g \in A[y]$ be such that $f(a) = g(b) = 0$, where

$$f(x) = a_0 + a_1 x + \cdots + a_{m-1}x^{m-1} + x^m,$$
$$g(y) = b_0 + b_1 y + \cdots + b_{n-1}y^{n-1} + y^n.$$

It suffices to consider the case

$$A = \mathbb{Z}[a_0, \ldots, a_{m-1}, b_0, \ldots, b_{n-1}], \qquad \text{and} \qquad B = \frac{A[x,y]}{(f(x), g(y))},$$

with α and β equal to the images of x and y in B, respectively, since given any $A' \subseteq B'$ we have homomorphisms $A \to A'$ defined by $a_i \to a_i$ and $b_i \to b_i$ and $B \to B'$ defined by $x \mapsto \alpha$ and $y \mapsto \beta$, and if $x + y, xy \in B$ are integral over A then $\alpha + \beta, \alpha\beta \in B'$ must be integral over A'.

Let k be the algebraic closure of the fraction field of A, and let $\alpha_1, \ldots, \alpha_m$ be the roots of f in k and let $\beta_1, \ldots, \beta_n$ be the roots of g in k. The polynomial

$$h(z) = \prod_{i,j} \big(z - (\alpha_i + \beta_j)\big)$$

has coefficients that may be expressed as polynomials in the symmetric functions of the α_i and β_j, equivalently, the coefficients a_i and b_j of f and g, respectively. Thus $h \in A[z]$, and $h(x+y) = 0$, so $x+y$ is integral over A. Applying the same argument to $h(z) = \prod_{i,j}(z - \alpha_i\beta_j)$ shows that xy is also integral over A. □

Definition 1.19. Given a ring extension B/A, the ring $\tilde{A} = \{b \in B : b \text{ is integral over } A\}$ is the *integral closure of A in B*. When $\tilde{A} = A$ we say that A is *integrally closed in B*. For a domain A, its *integral closure* (or *normalization*) is its integral closure in its fraction field, and A is *integrally closed* (or *normal*) if it is integrally closed in its fraction field.

Proposition 1.20. *If $C/B/A$ is a tower of ring extensions in which B is integral over A and C is integral over B then C is integral over A.*

Proof. See [1, Thm. 10.27] or [2, Cor. 5.4]. □

Corollary 1.21. *If B/A is a ring extension, then the integral closure of A in B is integrally closed in B.*

Proposition 1.22. *The ring $\mathbb{Z}$ is integrally closed.*

Proof. We apply the rational root test: suppose $r/s \in \mathbb{Q}$ is integral over $\mathbb{Z}$, where r and s are coprime integers. Then

$$\left(\frac{r}{s}\right)^n + a_{n-1}\left(\frac{r}{s}\right)^{n-1} + \cdots a_1\left(\frac{r}{s}\right) + a_0 = 0$$

for some $a_0, \ldots, a_{n-1} \in \mathbb{Z}$. Clearing denominators yields

$$r^n + a_{n-1}sr^{n-1} + \cdots a_1 s^{n-1}r + a_0 s^n = 0,$$

thus $r^n = -s(a_{n-1}r^{n-1} + \cdots a_1 s^{n-2}r + a_0 s^{n-1})$ is a multiple of s. But r and s are coprime, so $s = \pm 1$ and therefore $r/s \in \mathbb{Z}$. □

Corollary 1.23. *Every unique factorization domain is integrally closed. In particular, every PID is integrally closed.*

Proof. The proof of Proposition 1.22 works for any UFD. □

Example 1.24. The ring $\mathbb{Z}[\sqrt{5}]$ is not a UFD (nor a PID) because it is not integrally closed: consider $\phi = (1+\sqrt{5})/2 \in \operatorname{Frac}\mathbb{Z}[\sqrt{5}]$, which is integral over $\mathbb{Z}$ (and hence over $\mathbb{Z}[\sqrt{5}]$), since $\phi^2 - \phi - 1 = 0$. But $\phi \notin \mathbb{Z}[\sqrt{5}]$, so $\mathbb{Z}[\sqrt{5}]$ is not integrally closed.

The corollary implies that every discrete valuation ring is integrally closed. In fact, more is true.

Proposition 1.25. *Every valuation ring is integrally closed.*

Proof. Let A be a valuation ring with fraction field k and let $\alpha \in k$ be integral over A. Then

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

for some $a_0, a_1, \ldots, a_{n-1} \in A$. Suppose $\alpha \notin A$. Then $\alpha^{-1} \in A$, since A is a valuation ring. Multiplying the equation above by $\alpha^{-(n-1)} \in A$ and moving all but the first term on the LHS to the RHS yields

$$\alpha = -a_{n-1} - a_{n-1}\alpha^{-1} - \cdots - a_1\alpha^{2-n} - a_0\alpha^{1-n} \in A,$$

contradicting our assumption that $\alpha \notin A$. It follows that A is integrally closed. □

Definition 1.26. A *number field* K is a finite extension of $\mathbb{Q}$. The *ring of integers* $\mathcal{O}_K$ is the integral closure of $\mathbb{Z}$ in K.

Remark 1.27. The notation $\mathbb{Z}_K$ is also sometimes used to denote the ring of integers of K. The symbol $\mathcal{O}$ emphasizes the fact that $\mathcal{O}_K$ is an *order* in K; in any $\mathbb{Q}$-algebra K of finite dimension r, an order is a subring of K that is also a free $\mathbb{Z}$-module of rank r, equivalently, a $\mathbb{Z}$-lattice in K that is also a ring. In fact, $\mathcal{O}_K$ is the *maximal order* of K: it contains every order in K.

Proposition 1.28. *Let A be an integrally closed domain with fraction field K. Let α be an element of a finite extension L/K, and let $f \in K[x]$ be its minimal polynomial over K. Then α is integral over A if and only if $f \in A[x]$.*

Proof. The reverse implication is immediate: if $f \in A[x]$ then certainly α is integral over A. For the forward implication, suppose α is integral over A and let $g \in A[x]$ be a monic polynomial for which $g(\alpha) = 0$. In $\overline{K}[x]$ we may factor $f(x)$ as

$$f(x) = \prod_i (x - \alpha_i).$$

For each α_i we have a field embedding $K(\alpha) \to \overline{K}$ that sends α to α_i and fixes K. As elements of $\overline{K}$ we have $g(\alpha_i) = 0$ (since $f(\alpha_i) = 0$ and f must divide g), so each $\alpha_i \in \overline{K}$ is integral over A and lies in the integral closure $\tilde{A}$ of A in $\overline{K}$. Each coefficient of $f \in K[x]$ can be expressed as a sum of products of the α_i, and is therefore an element of the ring $\tilde{A}$ that also lies in K. But $A = \tilde{A} \cap K$, since A is integrally closed in its fraction field K. □

Example 1.29. We saw in Example 1.24 that $(1 + \sqrt{5})/2$ is integral over $\mathbb{Z}$. Now consider $\alpha = (1 + \sqrt{7})/2$. Its minimal polynomial $x^2 - x - 3/2 \notin \mathbb{Z}[x]$, so α is not integral over $\mathbb{Z}$.

References

[1] Allen Altman and Steven Kleiman, *A term of commutative algebra*, Worldwide Center of Mathematics, 2013.

[2] Michael Atiyah and Ian MacDonald, *Introduction to commutative algebra*, Addison–Wesley, 1969.

[3] Pierre Deligne, *La conjecture de Weil: I*, Publications Mathématiques l'I.H.É.S. **43** (1974), 273–307.

[4] Jean-Pierre Serre, *Local fields*, Springer, 1979.

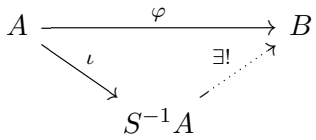
[5] André Weil, *Numbers of solutions of equations in finite fields*, Bulletin of the American Mathematical Society, **55** (1949), 497–508.

2 Localization and Dedekind domains

After a brief review of some commutative algebra background on localizations, in this lecture we begin our study of Dedekind domains, which are commutative rings that play a key role in algebraic number theory and arithmetic geometry (named after Richard Dedekind).

2.1 Localization of rings

Let A be a commutative ring (unital, as always), and let S be a multiplicative subset of A; this means S is closed under finite products (including the empty product, so $1 \in S$), and S does not contain zero. The *localization* of A with respect to S is a ring $S^{-1}A$ equipped with a ring homomorphism $\iota\colon A \to S^{-1}A$ that maps S into $(S^{-1}A)^\times$ and satisfies the following universal property: if $\varphi\colon A \to B$ is a ring homomorphism with $\varphi(S) \subseteq B^\times$ then there is a unique ring homomorphism $S^{-1}A \to B$ that makes the following diagram commute:



and one says that φ factors uniquely through $S^{-1}A$ (via ι). As usual with universal properties, this guarantees that $S^{-1}A$ is unique (hence well-defined), provided that it exists. To prove existence we construct $S^{-1}A$ as the quotient of $A \times S$ modulo the equivalence relation

$$(a, s) \sim (b, t) \Leftrightarrow \exists u \in S \text{ such that } (at - bs)u = 0. \tag{1}$$

We then use a/s to denote the equivalence class of (a, s) and define $\iota(a) := a/1$; one can easily verify that $S^{-1}A$ is a ring with additive identity $0/1$ and multiplicative identity $1/1$, and that $\iota : A \to S^{-1}A$ is a ring homomorphism. If s is invertible in A we can view a/s either as the element as^{-1} of A or the equivalence class of (a, s) in $S^{-1}A$; we have $(a, s) \sim (a/s, 1)$, since $(a \cdot 1 - a/s \cdot s) \cdot 1 = 0$, so this notation should not cause any confusion. For $s \in S$ we have $\iota(s)^{-1} = 1/s$, since $(s/1)(1/s) = s/s = 1/1 = 1$, thus $\iota(S) \subseteq (S^{-1}A)^\times$.

If $\varphi\colon A \to B$ is a ring homomorphism with $\varphi(S) \subseteq B^\times$, then $\varphi = \pi \circ \iota$, where π is defined by $\pi(a/s) := \varphi(a)\varphi(s)^{-1}$. If $\pi\colon S^{-1}A \to B$ is any ring homomorphism that satisfies $\varphi = \pi \circ \iota$, then $\varphi(a)\varphi(s)^{-1} = \pi(\iota(a))\pi(\iota(s))^{-1} = \pi(\iota(a)\iota(s)^{-1}) = \pi((a/1)(1/s)) = \pi(a/s)$, so π is unique.

In the case of interest to us, A is actually an integral domain, in which case $(a, s) \sim (b, t)$ if and only if $at - bs = 0$ (we can always take $u = 1$ in the equivalence relation (1) above), and we can then identify $S^{-1}A$ with a subring of the fraction field of A (which we note is the localization of A with respect to $S = A_{\neq 0}$), and if T is a multiplicative subset A that contains S, then $S^{-1}A \subseteq T^{-1}A$.

When A is an integral domain the map $\iota\colon A \to S^{-1}A$ is injective, allowing us to identify A with its image $\iota(A) \subseteq S^{-1}A$ (in general, ι is injective if and only if S contains no zero divisors). When A is an integral domain we may thus view $S^{-1}A$ as an intermediate ring that lies between A and its fraction field: $A \subseteq S^{-1}A \subseteq \operatorname{Frac} A$.

2.2 Ideals in localizations of rings

If $\varphi\colon A \to B$ is a ring homomorphism and $\mathfrak{b}$ is a B-ideal, then $\varphi^{-1}(\mathfrak{b})$ is an A-ideal called the *contraction* of $\mathfrak{b}$ to A and sometimes denoted $\mathfrak{b}^c$; when A is a subring of B and φ is

the inclusion map we simply have $\mathfrak{b}^c = \mathfrak{b} \cap A$. If $\mathfrak{a}$ is an A-ideal, in general $\varphi(\mathfrak{a})$ is not a B-ideal; but we can instead consider the B-ideal generated by $\varphi(\mathfrak{a})$, the *extension* of $\mathfrak{a}$ to B, sometimes denoted $\mathfrak{a}^e$.

In the case of interest to us, A is an integral domain, $B = S^{-1}A$ is the localization of A with respect to some multiplicative set S, and $\varphi = \iota$ is injective, so we view A as a subring of B. We then have

$$\mathfrak{a}^e = \mathfrak{a}B := (ab : a \in \mathfrak{a}, b \in B). \tag{2}$$

We clearly have $\mathfrak{a} \subseteq \varphi^{-1}((\varphi(\mathfrak{a}))) = \mathfrak{a}^{ec}$ and $\mathfrak{b}^{ce} = (\varphi(\varphi^{-1}(\mathfrak{b}))) \subseteq \mathfrak{b}$; one might ask whether these inclusions are equalities. In general the first is not: if $B = S^{-1}A$ and $\mathfrak{a} \cap S \neq \emptyset$ then $\mathfrak{a}^e = \mathfrak{a}B = B$ and $\mathfrak{a}^{ec} = B \cap A$ are both unit ideals, but we may still have $\mathfrak{a} \subsetneq A$. However when $B = S^{-1}A$ the second inclusion is an equality; see [1, Prop. 11.19] or [2, Prop. 3.11] for a short proof. We also note the following theorem.

Theorem 2.1. *Let S be a multiplicative subset of an integral domain A. There is a one-to-one correspondence between the prime ideals of $S^{-1}A$ and the primes ideals of A that do not intersect S given by the inverse maps $\mathfrak{q} \mapsto \mathfrak{q} \cap A$ and $\mathfrak{p} \mapsto \mathfrak{p}S^{-1}A$.*

Proof. See [1, Cor. 11.20] or [2, Prop. 3.11.iv]. □

Remark 2.2. An immediate consequence of (2) is that if $a_1, \dots, a_n \in A$ generate $\mathfrak{a}$ as an A-ideal, then they also generate $\mathfrak{a}^e = \mathfrak{a}B$ as a B-ideal. As noted above, when $B = S^{-1}A$ we have $\mathfrak{b} = \mathfrak{b}^{ce}$, so every B-ideal is of the form $\mathfrak{a}^e$ (take $\mathfrak{a} = \mathfrak{b}^c$). It follows that if A is noetherian then so are all its localizations, and if A is a PID then so are all of its localizations.

An important special case of localization occurs when $\mathfrak{p}$ is a prime ideal in an integral domain A, and $S = A - \mathfrak{p}$ (the complement of the set $\mathfrak{p}$ in the set A). In this case it is customary to denote $S^{-1}A$ by

$$A_{\mathfrak{p}} := \{a/b : a \in A, b \notin \mathfrak{p}\}/\sim, \tag{3}$$

and call it the *localization of A at $\mathfrak{p}$*. The prime ideals of $A_{\mathfrak{p}}$ are then in bijection with the prime ideals of A that lie in $\mathfrak{p}$. It follows that $\mathfrak{p}A_{\mathfrak{p}}$ is the unique maximal ideal of $A_{\mathfrak{p}}$ and $A_{\mathfrak{p}}$ is therefore a local ring (whence the term *localization*).

Warning 2.3. The notation in (3) makes it tempting to assume that if a/b is an element of $\operatorname{Frac} A$, then $a/b \in A_{\mathfrak{p}}$ if and only if $b \notin \mathfrak{p}$. This is not necessarily true! As an element of $\operatorname{Frac} A$, the notation "a/b" represents an equivalence class; if $a/b = a'/b'$ with $b' \notin A_{\mathfrak{p}}$, then in fact $a/b = a'/b' \in A_{\mathfrak{p}}$. As a trivial example, take $A = \mathbb{Z}$, $\mathfrak{p} = (3)$, $a/b = 9/3$ and $a'/b' = 3/1$. You may object that we should write a/b in lowest terms, but when A is not a unique factorization domain it is not clear what this means.

Example 2.4. For a field k, let $A = k[x]$ and $\mathfrak{p} = (x - 2)$. Then

$$A_{\mathfrak{p}} = \{f \in k(x) : f \text{ is defined at } 2\}.$$

The ring A is a PID, so $A_{\mathfrak{p}}$ is a PID with a unique nonzero maximal ideal (the ideal $\mathfrak{p}A_{\mathfrak{p}}$), hence a DVR. Its maximal ideal is

$$\mathfrak{p}A_{\mathfrak{p}} = \{f \in k(x) : f(2) = 0\}.$$

The valuation on the field $k(x) = \operatorname{Frac} A$ corresponding to the valuation ring $A_{\mathfrak{p}}$ measures the order of vanishing of functions $f \in k(x)$ at 2. The residue field is $A_{\mathfrak{p}}/\mathfrak{p}A_{\mathfrak{p}} \simeq k$, and the quotient map $A_{\mathfrak{p}} \twoheadrightarrow A_{\mathfrak{p}}/\mathfrak{p}A_{\mathfrak{p}}$ sends f to $f(2)$.

Example 2.5. Let $p \in \mathbb{Z}$ be a prime. Then $\mathbb{Z}_{(p)} = \{a/b : a, b \in \mathbb{Z}, p \nmid b\}$. As in the previous example, $\mathbb{Z}$ is a PID and $\mathbb{Z}_{(p)}$ is a DVR; the valuation on $\mathbb{Q}$ is the p-adic valuation. The residue field is $\mathbb{Z}_{(p)}/p\mathbb{Z}_{(p)} \simeq \mathbb{F}_p$ and the quotient map $\mathbb{Z}_{(p)} \twoheadrightarrow \mathbb{F}_p$ is reduction modulo p.

2.3 Localization of modules

The concept of localization generalizes immediately to modules. As above, let A be a commutative ring, let S a multiplicative subset of A, and let M be an A-module. The localization $S^{-1}M$ of M with respect to S is an $S^{-1}A$-module equipped with an A-module homomorphism $\iota\colon M \to S^{-1}M$ with the universal property that if N is an $S^{-1}A$-module and $\varphi\colon M \to N$ is an A-module homomorphism, then φ factors uniquely through $S^{-1}M$ (via ι). Note that in this definition we are viewing $S^{-1}A$-modules as A-modules via the canonical homomorphism $A \to S^{-1}A$ that is part of the definition of $S^{-1}A$. Our definition of $S^{-1}M$ reduces to the definition of $S^{-1}A$ in the case $M = A$.

The explicit construction of $S^{-1}M$ is exactly the same as $S^{-1}A$, one takes the quotient of the $S^{-1}A$-module $M \times S$ modulo the same equivalence relation as in (1):

$$(a, s) \sim (b, t) \Leftrightarrow \exists u \in S \text{ such that } (at - bs)u = 0,$$

where a and b now denote elements of M, and $\iota(a) \coloneqq a/1$ as before. Alternatively, one can define $S^{-1}M \coloneqq M \otimes_A S^{-1}A$ (see [2, Prop. 3.5] for a proof that this is equivalent). In other words, $S^{-1}M$ is the *base change* of M from A to $S^{-1}A$; we will discuss base change more generally in later lectures.

The map $\iota\colon M \to S^{-1}M$ is injective if and only if the map $M \xrightarrow{\times s} M$ is injective for every $s \in S$. This is a strong condition that does not hold in general, even when A is an integral domain (the annihilator of M may be non-trivial), but it applies to all the cases we care about. In particular, if A lies in a field K (in which case A must be an integral domain whose fraction field lies in K) and M is an A-module that is contained in a K-vector space. In this setting multiplication by any nonzero $s \in A$ is injective and we can view M as an A-submodule of any of its localizations $S^{-1}M$.

We will mostly be interested in the case $S = A - \mathfrak{p}$, where $\mathfrak{p}$ is a prime ideal of A, in which case we write $M_\mathfrak{p}$ for $S^{-1}M$, just as we write $A_\mathfrak{p}$ for $S^{-1}A$.

Proposition 2.6. *Let A be a subring of a field K, and let M be an A-module contained in a K-vector space V (equivalently, for which the map $M \to M \otimes_A K$ is injective).*[1] *Then*

$$M = \bigcap_\mathfrak{m} M_\mathfrak{m} = \bigcap_\mathfrak{p} M_\mathfrak{p},$$

where $\mathfrak{m}$ ranges over the maximal ideals of A, $\mathfrak{p}$ ranges over the prime ideals of A, and the intersections take place in V.

Proof. The fact that $M \subseteq \bigcap_\mathfrak{m} M_\mathfrak{m}$ is immediate. Now suppose $x \in \bigcap_\mathfrak{m} M_\mathfrak{m}$ and consider the A-ideal $\mathfrak{a} \coloneqq \{a \in A : ax \in M\}$. For each maximal ideal $\mathfrak{m}$ we can write $x = m/s$ for some $m \in M$ and $s \in A - \mathfrak{m}$; we then have $sx \in M$ and $s \in \mathfrak{a}$, but $s \notin \mathfrak{m}$, so $\mathfrak{a} \not\subseteq \mathfrak{m}$. It follows that $\mathfrak{a}$ and must be the unit ideal, so $1 \in \mathfrak{a}$ and $x = 1 \cdot x \in M$; thus $\cap_\mathfrak{m} M_\mathfrak{m} \subseteq M$.

We now note that each $M_\mathfrak{p}$ contains some $M_\mathfrak{m}$ (since each $\mathfrak{p}$ is contained in some $\mathfrak{m}$), and every maximal ideal is prime, so $\cap_\mathfrak{m} M_\mathfrak{m} = \cap_\mathfrak{p} M_\mathfrak{p}$. □

[1] The image is a tensor product of A-modules that is also a K-vector space. We need the natural map to be injective in order to embed M in it. Note that V necessarily contains a subspace isomorphic to $M \otimes_A K$.

An important special case of this proposition occurs when $K = \operatorname{Frac} A$ and $V = K$, in which case M is an A-submodule of K. Every ideal I of A is an A-submodule of K, and can thus be localized as above. The localization of I (as an A-module) at a prime ideal $\mathfrak{p}$ of A is the same thing as the extension of I (as an A-ideal) to the localization of A at $\mathfrak{p}$. In other words,

$$I_{\mathfrak{p}} = \{i/s : i \in I, s \in A - \mathfrak{p}\} = \{ia/s : i \in I, a \in A, s \in A - \mathfrak{p}\} = IA_{\mathfrak{p}}.$$

We also have the following corollary of Proposition 2.6.

Corollary 2.7. *Let A be an integral domain. Every ideal I of A (including $I = A$) is equal to the intersection of its localizations at the maximal ideals of A, and also to the intersection of its localizations at the prime ideals of A.*

Example 2.8. If $A = \mathbb{Z}$ then $\mathbb{Z} = \bigcap_p \mathbb{Z}_{(p)}$ in $\mathbb{Q}$.

Proposition 2.6 and Corollary 2.7 are powerful tools, because they allow us work in local rings (rings with just one maximal ideal), which often simplifies matters considerably. For example, to prove that an ideal I in an integral domain A satisfies a certain property, it is enough to show that this property holds for all its localizations $I_{\mathfrak{p}}$ at prime ideals $\mathfrak{p}$ and is preserved under intersections. We now want to consider rings A that satisfy some further assumptions that make its localizations become even easier to work with.

2.4 Dedekind domains

Proposition 2.9. *Let A be a noetherian domain. The following are equivalent:*

(i) *For every nonzero prime ideal $\mathfrak{p} \subset A$ the local ring $A_{\mathfrak{p}}$ is a DVR.*

(ii) *The ring A is integrally closed and $\dim A \leq 1$.*

Proof. If A is a field then (i) and (ii) both hold, so let us assume that A is not a field, and put $K := \operatorname{Frac} A$. We first show that (i) implies (ii). Recall that $\dim A$ is the supremum of the length of all chains of prime ideals. It follows from Theorem 2.1 that every chain of prime ideals $(0) \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_n$ extends to a corresponding chain in $A_{\mathfrak{p}_n}$ of the same length; conversely, every chain in $A_{\mathfrak{p}}$ contracts to a chain in A of the same length. Thus

$$\dim A = \sup\{\dim A_{\mathfrak{p}} : \mathfrak{p} \in \operatorname{Spec} A\} = 1,$$

since every $A_{\mathfrak{p}}$ is either a DVR ($\mathfrak{p} \neq (0)$), in which case $\dim A_{\mathfrak{p}} = 1$, or a field ($\mathfrak{p} = (0)$), in which case $\dim A_{\mathfrak{p}} = 0$. Any $x \in K$ that is integral over A is integral over every $A_{\mathfrak{p}}$ (since they all contain A), and the $A_{\mathfrak{p}}$ are integrally closed, since they are DVRs or fields. So $x \in \bigcap_{\mathfrak{p}} A_{\mathfrak{p}} = A$, and therefore A is integrally closed, which shows (ii).

To show that (ii) implies (i), we first show that the following properties are all inherited by localizations of a ring: (1) no zero divisors, (2) noetherian, (3) dimension at most one, (4) integrally closed. (1) is obvious, (2) was noted in Remark 2.2, and (3) follows from Theorem 2.1 since, as argued above, we have $\dim A_{\mathfrak{p}} \leq \dim A$. To show (4), suppose $x \in K$ is integral over $A_{\mathfrak{p}}$. Then

$$x^n + \frac{a_{n-1}}{s_{n-1}}x^{n-1} + \cdots + \frac{a_1}{s_1}x + \frac{a_0}{s_0} = 0$$

for some $a_0, \dots, a_{n-1} \in A$ and $s_0, \dots, s_{n-1} \in A - \mathfrak{p}$. Multiplying both sides by s^n, where $s = s_0 \cdots s_{n-1} \in S$, shows that sx is integral over A, hence an element of A, since A is integrally closed. But then $sx/s = x$ is an element of $A_\mathfrak{p}$, so $A_\mathfrak{p}$ is integrally closed as claimed.

Thus (ii) implies that every $A_\mathfrak{p}$ is an integrally closed noetherian local domain of dimension at most 1, and for $\mathfrak{p} \neq (0)$ we must have $\dim A_\mathfrak{p} = 1$. Thus for every nonzero prime ideal $\mathfrak{p}$, the ring $A_\mathfrak{p}$ is an integrally closed noetherian local domain of dimension 1, and therefore a DVR, by Theorem 1.16. □

Definition 2.10. A noetherian domain satisfying either of the equivalent properties of Proposition 2.9 is called a *Dedekind domain*.

Corollary 2.11. *Every PID is a Dedekind domain. In particular, $\mathbb{Z}$ is a Dedekind domain, as is $k[x]$ for any field k.*

Remark 2.12. Every PID is both a UFD and a Dedekind domain. Not every UFD is a Dedekind domain (consider $k[x, y]$, for any field k), and not every Dedekind domain is a UFD (consider $\mathbb{Z}[\sqrt{-13}]$, in which $(1 + \sqrt{-13})(1 - \sqrt{-13}) = 2 \cdot 7 = 14$). However (as we shall see), every ring that is both a UFD and a Dedekind domain is a PID.

One of our first goals in this course is to prove that ring of integers of number fields and coordinate rings of global function fields are Dedekind domains. More precisely, we will prove that if A is a Dedekind domain and L is a finite separable extension of its fraction field, then the integral closure of A in L is a Dedekind domain. This includes the two main cases of interest to us, in which either $A = \mathbb{Z}$ and L is a number field, or $A = \mathbb{F}_q[t]$ and L is a global function field. Recall from Lecture 1 that number fields and global function fields are the two types of *global fields* (as we will prove in later lectures).

2.5 Fractional ideals

Throughout this subsection, A is a noetherian domain (not necessarily a Dedekind domain) and K is its fraction field.

Definition 2.13. A *fractional ideal* of a noetherian domain A is a finitely generated A-submodule of its fraction field.

Fractional ideals generalize the notion of an ideal: when A is noetherian the ideals of A are precisely the finitely generated A-submodules of A, and when A is also a domain we can extend this notion to its fraction field. Every ideal of A is also a fractional ideal of A, but fractional ideals are typically not ideals because they need not be contained in A. Some authors use the term *integral ideal* to distinguish the fractional ideals that lie in A (and are thus ideals) but we will not use this terminology.

Lemma 2.14. *Let A be a noetherian domain with fraction field K, and let $I \subseteq K$ be an A-module. Then I is finitely generated if and only if $aI \subseteq A$ for some nonzero $a \in A$.*

Proof. For the forward implication, if $r_1/s_1, \dots, r_n/s_n$ generate I as an A-module, then $aI \subseteq A$ for $a = s_1 \cdots s_n$. Conversely, if $aI \subseteq A$, then aI is an ideal, hence finitely generated (since A is noetherian), and if $a_1, \dots, a_n$ generate aI then $a_1/a, \dots, a_n/a$ generate I. □

Remark 2.15. Lemma 2.14 gives an alternative definition of fractional ideals that can be extended to domains that are not necessarily noetherian; they are A-submodules I of K for which there exists a nonzero $r \in A$ such that $rI \subseteq A$. When A is noetherian this coincides with our definition above.

Corollary 2.16. *Every fractional ideal of A can be written in the form $\frac{1}{a}I$, for some nonzero $a \in A$ and ideal I.*

Definition 2.17. A fractional ideal of A is *principal* if it is generated by one element, that is, it has the form xA for some $x \in K$. We will also use the notation $(x) := xA$ to denote the principal fractional ideal generated by $x \in K$.

As with ideals, we can add and multiply fractional ideals:

$$I + J := (i + j : i \in I, j \in J), \qquad IJ := (ij : i \in I, j \in J).$$

Here the notation (S) means the A-module generated by $S \subseteq K$. As with ideals, we actually have $I + J = \{i + j : i \in I, j \in J\}$, but the ideal IJ is typically not the same as set $\{ij : i \in I, j \in J\}$, it consists of all finite sums of elements in this set. We also have a new operation, corresponding to division. For any fractional ideals I, J with J nonzero, the set

$$(I : J) := \{x \in K : xJ \subseteq I\}$$

is called a *colon ideal*. Some texts refer to $(I : J)$ as the *ideal quotient* of I by J, but note that it is **not** a quotient of A-modules (for example, $(\mathbb{Z} : \mathbb{Z}) = \mathbb{Z}$ but $\mathbb{Z}/\mathbb{Z} = \{0\}$).

We do not assume $I \subseteq J$ (or $J \subseteq I$), the definition makes sense for any fractional ideals I and J with J nonzero. [2] If $I = (x)$ and $J = (y)$ are principal fractional ideals then $(I : J) = (x/y)$, so colon ideals can be viewed as a generalization of division in $K^\times$.

Lemma 2.18. *Let I and J be fractional ideals of a noetherian domain A with J nonzero. Then $(I : J)$ is a fractional ideal of A.*

Proof. It is clear from the definition that $(I : J)$ is closed under addition and multiplication by elements of A (since I is), so $(I : J)$ is an A-module of the fraction field of A. To show that $(I : J)$ is finitely generated, we first suppose that $I, J \subseteq A$ are ideals. For any nonzero $j \in J \subseteq A$ we have $j(I : J) \subseteq I \subseteq A$, so $(I : J)$ is finitely generated, by Lemma 2.14. For the general case, choose a and b so that $aI \subseteq A$ and $bJ \subseteq A$ via Lemma 2.14. Then $(I : J) = (abI : abJ)$ with $abI, abJ \subseteq A$, which we have already shown is finitely generated. □

Definition 2.19. A fractional ideal I is *invertible* if $IJ = A$ for some fractional ideal J.

Inverses are unique when they exist: if $IJ = A = IJ'$ then $J = JA = JIJ' = AJ' = J'$. We may use I^{-1} to denote the inverse of a fractional ideal I when it exists.

Lemma 2.20. *A fractional ideal I of A is invertible if and only if $I(A : I) = A$ (in which case $(A : I)$ is its inverse).*

Before proving the lemma, note that $I(A : I) \subseteq A$ always holds, since for $y \in I$ and $x \in (A : I)$ we have $xy \in xI \subseteq A$, by the definition of $(A : I)$. The lemma states that this inclusion is an equality precisely when I is invertible.

[2] The definition still makes sense when J is the zero ideal, but $(I : (0)) = K$ will typically not be finitely generated as an A-module, hence not a fractional ideal.

Proof. Suppose I is invertible, with $IJ = A$. Then $jI \subseteq A$ for all $j \in J$, so $J \subseteq (A : I)$, and $A = IJ \subseteq I(A : I) \subseteq A$, so $I(A : I) = A$. □

In the next lecture we will prove that in a Dedekind domain every nonzero fractional ideal is invertible, but let us first note that this is not true in general.

Example 2.21. Consider the subring $A := \mathbb{Z}+2i\mathbb{Z}$ of the Gaussian integers (with $i^2 = -1$). The set $I := 2\mathbb{Z}[i]$ is a non-invertible A-ideal (even though it is an invertible $\mathbb{Z}[i]$-ideal); indeed, we have $(A : I) = \mathbb{Z}[i]$ and $I(A : I) = 2\mathbb{Z}[i] \subsetneq A$.

References

[1] Allen Altman and Steven Kleiman, *A term of commutative algebra*, Worldwide Center of Mathematics, 2013.

[2] Michael Atiyah and Ian MacDonald, *Introduction to commutative algebra*, Addison–Wesley, 1969.

[3] Anthony W. Knapp, *Advanced Algebra*, Digital Second Edition, 2016.

3 Properties of Dedekind domains

In the previous lecture we defined a Dedekind domain as a noetherian domain A that satisfies either of the following equivalent conditions:

- the localizations of A at its nonzero prime ideals are all discrete valuation rings;
- A is integrally closed and has dimension at most one.

In this lecture we will establish several additional properties enjoyed by Dedekind domains, the most significant of which is unique factorization of ideals. As we noted last time, Dedekind domains are typically not unique factorization domains (this occurs if and only if it is also a principal ideal domain), but ideals can be uniquely factored into prime ideals.

3.1 Invertible fractional ideals and the ideal class group

In this section A is a noetherian domain (not necessarily a Dedekind domain) and K is its fraction field. Recall that a fractional ideal of A is a finitely generated A-submodule of K, and if I and J are fractional ideals, so is the colon ideal

$$(I : J) := \{x \in K : xJ \subseteq I\},$$

and we that a fractional ideal I is invertible if $IJ = A$ for some fractional ideal J. The definition of $(A : I)$ implies $I(A : I) \subseteq A$, and Lemma 2.20 implies that I is invertible precisely when this inclusion is an equality, in which case the inverse of I is $(A : I)$.

Ideal multiplication is commutative and associative, thus the set of nonzero fractional ideals of a noetherian domain form an abelian monoid under multiplication with $A = (1)$ as the identity. It follows that the subset of invertible fractional ideals is an abelian group.

Definition 3.1. The *ideal group* $\mathcal{I}_A$ of a noetherian domain A is the group of invertible fractional ideals. Note that, despite the name, elements of $\mathcal{I}_A$ need not be ideals.

Every nonzero principal fractional ideal (x) is invertible (since $(x)^{-1} = (x^{-1})$), and a product of principal fractional ideals is principal (since $(x)(y) = (xy)$), as is the unit ideal (1), thus the set of nonzero principal fractional ideals $\mathcal{P}_A$ is a subgroup of $\mathcal{I}_A$.

Definition 3.2. Let A be a noetherian domain. The quotient $\mathrm{cl}(A) := \mathcal{I}_A/\mathcal{P}_A$ is the *ideal class group* of A; it is also called the *Picard group* of A and denoted $\mathrm{Pic}(A)$.[1]

Example 3.3. If A is a DVR with uniformizer π then its nonzero fractional ideals are the principal fractional ideals (π^n) with $n \in \mathbb{Z}$ (including $n \le 0$). We have $(\pi^m)(\pi^n) = (\pi^{m+n})$, thus the ideal group of A is isomorphic to $\mathbb{Z}$ (under addition). In this case $\mathcal{P}_A = \mathcal{I}_A$ and the ideal class group $\mathrm{cl}(A)$ is trivial.

Remark 3.4. A Dedekind domain is a UFD if and only if its ideal class group is trivial (see Corollary 3.19 below), thus $\mathrm{cl}(A)$ may be viewed as a measure of how far A is from being a UFD. More generally, the ideal class group of an integrally closed noetherian domain A

[1] In general, the Picard group of a commutative ring A as the group of isomorphism classes of A-modules that are invertible under tensor product (equivalently, projective modules of rank one). When A is a noetherian domain, the Picard group of A is canonically isomorphic to the ideal class group of A and the two notions may be used interchangeably.

is trivial when A is a UFD, and the converse holds if one replaces the ideal class group with the *divisor class group*. One defines a divisor as an equivalence class of fractional ideals modulo the equivalence relation $I \sim J \Leftrightarrow (A : I) = (A : J)$, and in an integrally closed noetherian domain A (or more generally, a Krull domain), the set of divisors forms a group that contains principal divisors as a subgroup; the divisor class group is defined as the quotient, and it is trivial if and only if A is a UFD (this holds more generally for any Krull domain, see [2, Thm. 8.34]). In a Dedekind domain, fractional ideals are always distinct as divisors and every nonzero fractional ideal is invertible, so the ideal class group and divisor class group coincide.[2]

3.2 Invertible ideals in Dedekind domains

In order to prove that every nonzero fractional ideal in a Dedekind domain is invertible, we first note that arithmetic of fractional ideals behaves well under localization.

Lemma 3.5. *Let I and J be fractional ideals of A of a noetherian domain A, and let $\mathfrak{p}$ be a prime ideal of A. Then $I_\mathfrak{p}$ and $J_\mathfrak{p}$ are fractional ideals of $A_\mathfrak{p}$, as are*

$$(I+J)_\mathfrak{p} = I_\mathfrak{p} + J_\mathfrak{p}, \qquad (IJ)_\mathfrak{p} = I_\mathfrak{p} J_\mathfrak{p}, \qquad (I:J)_\mathfrak{p} = (I_\mathfrak{p} : J_\mathfrak{p}).$$

The same applies if we localize with respect to any multiplicative subset S of A.

Proof. $I_\mathfrak{p} = IA_\mathfrak{p}$ is a finitely generated $A_\mathfrak{p}$-module (since I is a finitely generated A-module; see Remark 2.2), hence a fractional ideal of $A_\mathfrak{p}$, and similarly for $J_\mathfrak{p}$. We have

$$(I+J)_\mathfrak{p} = (I+J)A_\mathfrak{p} = IA_\mathfrak{p} + JA_\mathfrak{p} = I_\mathfrak{p} + J_\mathfrak{p},$$

where we use the distributive law in K to get $(I+J)A_\mathfrak{p} = IA_\mathfrak{p} + JA_\mathfrak{p}$. We also have

$$(IJ)_\mathfrak{p} = (IJ)A_\mathfrak{p} = I_\mathfrak{p} J_\mathfrak{p},$$

since $(IJ)A_\mathfrak{p} \subseteq I_\mathfrak{p} J_\mathfrak{p}$ obviously holds and by writing sums of fractions over a common denominator we can see that $I_\mathfrak{p} J_\mathfrak{p} \subseteq (IJ)A_\mathfrak{p}$ also holds. Finally

$$(I:J)_\mathfrak{p} = \{x \in K : xJ \subseteq I\}_\mathfrak{p} = \{x \in K : xJ_\mathfrak{p} \subseteq I_\mathfrak{p}\} = (I_\mathfrak{p} : J_\mathfrak{p}).$$

For the last statement, note that no part of our proof depends on the fact that we localized with respect to a multiplicative set of the from $A - \mathfrak{p}$. □

Theorem 3.6. *Let I be a fractional ideal of a noetherian domain A. Then I is invertible if and only if its localization at every maximal ideal of A is invertible, equivalently, if and only if its localization at every prime ideal of A is invertible.*

Proof. Suppose I is invertible. Then $I(A : I) = A$, and for any maximal ideal $\mathfrak{m}$ we have $I_\mathfrak{m}(A_\mathfrak{m} : I_\mathfrak{m}) = A_\mathfrak{m}$, by Lemma 3.5, so $I_\mathfrak{m}$ is also invertible.

Now suppose $I_\mathfrak{m}$ is invertible for every maximal ideal $\mathfrak{m}$; then $I_\mathfrak{m}(A_\mathfrak{m} : I_\mathfrak{m}) = A_\mathfrak{m}$ for every maximal ideal $\mathfrak{m}$. Applying Lemma 3.5 and Proposition 2.6 yields

$$I(A:I) = \bigcap_\mathfrak{m} (I(A:I))_\mathfrak{m} = \bigcap_\mathfrak{m} I_\mathfrak{m}(A_\mathfrak{m} : I_\mathfrak{m}) = \bigcap_\mathfrak{m} A_\mathfrak{m} = A,$$

so I is invertible. The same proof works for prime ideals. □

[2] In general, the divisor class group and the ideal class group (or Picard group) of an integrally closed noetherian domain A may differ when $\dim A > 1$; see [1, Thm. 19.38] for a dimension 2 an example in which the ideal class group is trivial but the divisor class group is not (implying that A is not a UFD).

Corollary 3.7. *In a Dedekind domain every nonzero fractional ideal is invertible.*

Proof. If A is Dedekind then all of its localizations at maximal ideals are DVRs, hence PIDs, and in a PID every nonzero fractional ideal is invertible. It follows from Theorem 3.6 that every nonzero fractional ideal of A is invertible. □

An integral domain in which every nonzero ideal is invertible is a Dedekind domain (see Problem Set 2), so this gives another way to define Dedekind domains. Let us also note an equivalent condition that will be useful in later lectures.

Lemma 3.8. *A nonzero fractional ideal I in a noetherian local domain A is invertible if and only if it is principal.*

Proof. If I is principal then it is invertible, so we only need to show the converse. Let I be an invertible fractional ideal, and let $\mathfrak{m}$ be the maximal ideal of A. We have $II^{-1} = A$, so $\sum_{i=1}^{n} a_i b_i = 1$ for some $a_i \in I$ and $b_i \in I^{-1}$, and each $a_i b_i$ lies in $II^{-1} = A$. One of the products $a_i b_i$, say $a_1 b_1$, must be a unit, otherwise the sum would not be a unit (note that $A = \mathfrak{m} \sqcup A^\times$, since A is a local ring). For every $x \in I$ we have $a_1 b_1 x \in (a_1)$, since $b_1 x \in A$ (because $x \in I$ and $b_1 \in I^{-1}$). It follows that $x = (a_1 b_1)^{-1} a_1 b_1 x \in (a_1)$, since $(a_1 b_1)^{-1} \in A$, so we have $I \subseteq (a_1) \subseteq I$, which shows that $I = (a_1)$ is principal. □

Corollary 3.9. *A nonzero fractional ideal in a noetherian domain A is invertible if and only if it is locally principal, that is, its localization at every maximal ideal of A is principal.*

3.3 Unique factorization of ideals in Dedekind domains

We are now ready to prove the main result of this lecture, that every nonzero ideal in a Dedekind domain has a unique factorization into prime ideals. As a first step we need to show that every ideal is contained in only finitely many prime ideals.

Lemma 3.10. *Let A be a Dedekind domain and let $a \in A$ be nonzero. The set of prime ideals that contain a is finite.*

Proof. Consider the following subsets S and T of the ideal group $\mathcal{I}_A$:

$$S := \{I \in \mathcal{I}_A : (a) \subseteq I \subseteq A\},$$
$$T := \{I \in \mathcal{I}_A : A \subseteq I \subseteq (a)^{-1}\}.$$

The sets S and T are both non-empty (they contain A) and partially ordered by inclusion. The elements of S are all ideals, and we have bijections

$$\begin{aligned} \varphi_1 \colon S &\to T & \qquad \varphi_2 \colon T &\to S \\ I &\mapsto I^{-1} & \qquad I &\mapsto aI \end{aligned}$$

with φ_1 order-reversing and φ_2 order-preserving. The composition $\varphi := \varphi_2 \circ \varphi_1$ is thus an order-reversing permutation of S. Since A is noetherian, the set S satisfies the ascending chain condition: every chain $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ of ideals in S is eventually constant. By applying our order-reversing permutation φ we see that S also satisfies the descending chain condition: every chain $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$ of ideals in S is eventually constant.

Now if a lies in infinitely many distinct prime ideals $\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{p}_3, \ldots$, then

$$\mathfrak{p}_1 \supseteq \mathfrak{p}_1 \cap \mathfrak{p}_2 \supseteq \mathfrak{p}_1 \cap \mathfrak{p}_2 \cap \mathfrak{p}_3 \supseteq \cdots$$

is a descending chain of ideals in S that must stabilize. Thus for n sufficiently large we have

$$\mathfrak{p}_1 \cdots \mathfrak{p}_{n-1} \subseteq \mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_{n-1} = \mathfrak{p}_1 \cap \cdots \cap \mathfrak{p}_n \subseteq \mathfrak{p}_n.$$

The prime ideal $\mathfrak{p}_n$ contains the product $\mathfrak{p}_1 \cdots \mathfrak{p}_{n-1}$, so it must contain one of the factors $\mathfrak{p}_1, \cdots, \mathfrak{p}_{n-1}$ (this is what it means for an ideal to be prime). But this contradicts $\dim A \leq 1$: we cannot have a chain of prime ideals $(0) \subsetneq \mathfrak{p}_i \subsetneq \mathfrak{p}_n$ of length 2 in A. □

Corollary 3.11. *Let I be a nonzero ideal of a Dedekind domain A. The number of prime ideals of A that contain I is finite.*

Proof. Apply Lemma 3.10 to any nonzero $a \in I$. □

Example 3.12. The Dedekind domain $A = \mathbb{C}[t]$ contains uncountably many nonzero prime ideals $\mathfrak{p}_r = (t - r)$, one for each $r \in \mathbb{C}$. But any nonzero $f \in \mathbb{C}[t]$ lies in only finitely many of them, namely, the $\mathfrak{p}_r$ for which $f(r) = 0$; equivalently, f has finitely many roots.

Let $\mathfrak{p}$ be a nonzero prime ideal in a Dedekind domain A with fraction field K, let π be a uniformizer for the discrete valuation ring $A_\mathfrak{p}$, and let I be a nonzero fractional ideal of A. The localization $I_\mathfrak{p}$ is a nonzero fractional ideal of $A_\mathfrak{p}$, hence of the form (π^n) for some $n \in \mathbb{Z}$ that does not depend on the choice of π (note that n may be negative). We now extend the valuation $v_\mathfrak{p} \colon K \to \mathbb{Z} \cup \{\infty\}$ to fractional ideals by defining $v_\mathfrak{p}(I) := n$ and $v_\mathfrak{p}((0)) := \infty$; for any $x \in K$ we have $v_\mathfrak{p}((x)) = v_\mathfrak{p}(x)$.

The map $v_\mathfrak{p} \colon \mathcal{I}_A \to \mathbb{Z}$ is a group homomorphism: if $I_\mathfrak{p} = (\pi^m)$ and $J_\mathfrak{p} = (\pi^n)$ then

$$(IJ)_\mathfrak{p} = I_\mathfrak{p} J_\mathfrak{p} = (\pi^m)(\pi^n) = (\pi^{m+n}),$$

so $v_p(IJ) = m + n = v_\mathfrak{p}(I) + v_\mathfrak{p}(J)$. It is order-reversing with respect to the partial ordering on $\mathcal{I}_A$ by inclusion and the total order on $\mathbb{Z}$: for any $I, J \in \mathcal{I}_A$, if $I \subseteq J$ then $v_\mathfrak{p}(I) \geq v_\mathfrak{p}(J)$.

Lemma 3.13. *Let $\mathfrak{p}$ be a nonzero prime ideal in a Dedekind domain A. If I is an ideal of A then $v_\mathfrak{p}(I) = 0$ if and only if $\mathfrak{p}$ does not contain I. In particular, if $\mathfrak{q}$ is any nonzero prime ideal different from $\mathfrak{p}$ then $v_\mathfrak{q}(\mathfrak{p}) = v_\mathfrak{p}(\mathfrak{q}) = 0$.*

Proof. If $I \subseteq \mathfrak{p}$ then $v_p(I) \geq v_\mathfrak{p}(\mathfrak{p}) = 1$ is nonzero. If $I \not\subseteq \mathfrak{p}$ then pick $a \in I - \mathfrak{p}$ and note that $0 = v_\mathfrak{p}(a) \geq v_\mathfrak{p}(I) \geq v_\mathfrak{p}(A) = 0$, since $(a) \subseteq I \subseteq A$. The prime ideals $\mathfrak{p}$ and $\mathfrak{q}$ are nonzero, hence maximal (since $\dim A \leq 1$), so neither contains the other and $v_\mathfrak{q}(\mathfrak{p}) = v_\mathfrak{p}(\mathfrak{q}) = 0$. □

Corollary 3.14. *Let A be a Dedekind domain with fraction field K. For each nonzero fractional ideal I we have $v_\mathfrak{p}(I) = 0$ for all but finitely many prime ideals $\mathfrak{p}$. In particular, if $x \in K^\times$ then $v_\mathfrak{p}(x) = 0$ for all but finitely many $\mathfrak{p}$.*

Proof. For $I \subseteq A$ this follows from Corollary 3.11 and Lemma 3.13. For $I \not\subseteq A$ let $I = \frac{1}{a}J$ with $a \in A$ and $J \subseteq A$. Then $v_\mathfrak{p}(I) = v_\mathfrak{p}(J) - v_\mathfrak{p}(a) = 0 - 0 = 0$ for all but finitely many prime ideals $\mathfrak{p}$. This holds in particular for $I = (x)$, for any $x \in K^\times$. □

We are now ready to prove our main theorem.

Theorem 3.15. *Let A be a Dedekind domain. The ideal group $\mathcal{I}_A$ of A is the free abelian group generated by its nonzero prime ideals $\mathfrak{p}$. The isomorphism*

$$\mathcal{I}_A \simeq \bigoplus_\mathfrak{p} \mathbb{Z}$$

is given by the inverse maps

$$I \mapsto (\dots, v_{\mathfrak{p}}(I), \dots)$$
$$\prod_{\mathfrak{p}} \mathfrak{p}^{e_{\mathfrak{p}}} \longleftarrow (\dots, e_{\mathfrak{p}}, \dots)$$

Proof. Corollary 3.14 implies that the first map is well defined (the vector associated to $I \in \mathcal{I}_A$ has only finitely many nonzero entries and is thus an element of the direct sum). For each nonzero prime ideal $\mathfrak{p}$, the maps $I \mapsto v_{\mathfrak{p}}(I)$ and $e_{\mathfrak{p}} \mapsto \mathfrak{p}^{e_{\mathfrak{p}}}$ are group homomorphisms, and it follows that the maps in the theorem are both group homomorphisms. To see that the first map is injective, note that if $v_{\mathfrak{p}}(I) = v_{\mathfrak{p}}(J)$ then $I_{\mathfrak{p}} = J_{\mathfrak{p}}$, and if this holds for every $\mathfrak{p}$ then $I = \cap_{\mathfrak{p}} I_{\mathfrak{p}} = \cap_{\mathfrak{p}} J_{\mathfrak{p}} = J$, by Corollary 2.7. To see that it is surjective, note that Lemma 3.13 implies that for any vector $(\dots, e_{\mathfrak{p}}, \dots)$ in the image we have

$$v_{\mathfrak{q}}\left(\prod_{\mathfrak{p}} \mathfrak{p}^{e_{\mathfrak{p}}}\right) = \sum_{\mathfrak{p}} e_{\mathfrak{p}} v_{\mathfrak{q}}(\mathfrak{p}) = e_{\mathfrak{q}},$$

which implies that $\prod_{\mathfrak{p}} \mathfrak{p}^{e_{\mathfrak{p}}}$ is the pre-image of $(\dots, e_{\mathfrak{p}}, \dots)$; it also shows that the second map is the inverse of the first map. □

Remark 3.16. When A is a DVR, the isomorphism given by Theorem 3.15 is just the discrete valuation map $v_{\mathfrak{p}} \colon \mathcal{I}_A \xrightarrow{\sim} \mathbb{Z}$, where $\mathfrak{p}$ is the unique maximal ideal of A.

Corollary 3.17. *In a Dedekind domain every nonzero fractional ideal I has a unique factorization $I = \prod_{\mathfrak{p}} \mathfrak{p}^{v_{\mathfrak{p}}(I)}$ into nonzero prime ideals $\mathfrak{p}$.*[3]

Remark 3.18. Every integral domain with unique ideal factorization is a Dedekind domain (see Problem Set 2).

The isomorphism of Theorem 3.15 allows us to reinterpret the operations we have defined on fractional ideals. If $I = \prod_{\mathfrak{p}} \mathfrak{p}^{e_{\mathfrak{p}}}$ and $J = \prod_{\mathfrak{p}} \mathfrak{p}^{f_{\mathfrak{p}}}$ are nonzero fractional ideals then

$$\begin{aligned} IJ &= \prod \mathfrak{p}^{e_{\mathfrak{p}} + f_{\mathfrak{p}}}, \\ (I : J) &= \prod \mathfrak{p}^{e_{\mathfrak{p}} - f_{\mathfrak{p}}}, \\ I + J &= \prod \mathfrak{p}^{\min(e_{\mathfrak{p}}, f_{\mathfrak{p}})} = \gcd(I, J), \\ I \cap J &= \prod \mathfrak{p}^{\max(e_{\mathfrak{p}}, f_{\mathfrak{p}})} = \operatorname{lcm}(I, J), \end{aligned}$$

and for all $I, J \in \mathcal{I}_A$ we have

$$IJ = (I \cap J)(I + J).$$

A key consequence of unique factorization is that $I \subseteq J$ if and only if $e_{\mathfrak{p}} \geq f_{\mathfrak{p}}$ for all $\mathfrak{p}$; this implies that J contains I if and only if J divides I. Recall that in any commutative ring, if J divides I (i.e. $JH = I$ for some ideal H) then J contains I (the elements of I are H-linear, hence A-linear, combinations of elements of J and so lie in J), whence the slogan *to divide is to contain*. In a Dedekind domain the converse is also true: *to contain is to divide*. This leads to another characterization of Dedekind domains (see Problem Set 2).

[3] We view $A = \prod_{\mathfrak{p}} \mathfrak{p}^{v_{\mathfrak{p}}(A)} = \prod_{\mathfrak{p}} \mathfrak{p}^0 = (1)$ as an (empty) product of prime ideals.

Given that inclusion and divisibility are equivalent in a Dedekind domain, we may view $I+J$ as the greatest common divisor of I and J (it is the smallest ideal that contains, hence divides, both I and J), and $I \cap J$ as the least common multiple of I and J (it is the largest ideal contained in, hence divisible by, both I and J).[4]

We also note that

$$x \in I \iff (x) \subseteq I \iff v_{\mathfrak{p}}(x) \geq e_{\mathfrak{p}} \text{ for all } \mathfrak{p},$$

(where $I = \prod_{\mathfrak{p}} \mathfrak{p}^{e_{\mathfrak{p}}}$ as above), and therefore

$$I = \{x \in K : v_{\mathfrak{p}}(x) \geq e_{\mathfrak{p}} \text{ for all } \mathfrak{p}\}.$$

We have $I \subseteq A$ if and only if $e_{\mathfrak{p}} \geq 0$ for all $\mathfrak{p}$.

Corollary 3.19. *A Dedekind domain is a UFD if and only if it is a PID, equivalently, if and only if its class group is trivial.*

Proof. Every PID is a UFD, so we only need to prove the reverse implication. The fact that we have unique factorization of ideals implies that it is enough to to show that every prime ideal is principal. Let $\mathfrak{p}$ be a nonzero prime ideal in a Dedekind domain A that is also a UFD, let $a \in \mathfrak{p}$ nonzero, and let $a = p_1 \cdots p_n$ be the unique factorization of a into irreducible elements. Now $\mathfrak{p}$ contains and therefore divides $(a) = (p_1) \cdots (p_n)$, so $\mathfrak{p}$ divides (and therefore contains) some (p_i), which is necessarily a prime ideal (in a UFD, irreducible elements generate prime ideals). But A has dimension one, so we must have $\mathfrak{p} = (p_i)$. □

3.4 Representing ideals in a Dedekind domain

Not all Dedekind domains are PIDs; a typical Dedekind domain will contain ideals that require more than one generator. But it turns out that two generators always suffice, and we can even pick one of them arbitrarily. To prove this we need the following lemma. Recall that two A-ideals I and J are said to be *relatively prime*, or *coprime*, if $I + J = A$; equivalently, $\gcd(I, J) = (1)$.

Lemma 3.20. *Let A be a Dedekind domain and let I and I' be nonzero ideals. There exists an ideal J coprime to I' such that IJ is principal.*

Proof. Let $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ be the nonzero prime ideals dividing I' (a finite list, by Corollary 3.11). For $1 \leq i \leq n$ define the ideal $\mathfrak{a}_i := \mathfrak{p}_1 \cdots \mathfrak{p}_{i-1}\mathfrak{p}_{i+1} \cdots \mathfrak{p}_n$ and choose $a_i \in I$ so that

$$a_i \in \mathfrak{a}_i I \qquad \text{and} \qquad a_i \notin \mathfrak{p}_i I.$$

Note that $\mathfrak{a}_i I \cap \mathfrak{p}_i I \subsetneq \mathfrak{a}_i I$ because $v_{\mathfrak{p}_i}(\mathfrak{a}_i I \cap \mathfrak{p}_i I) = v_{\mathfrak{p}_i}(\mathfrak{p}_i I) > v_{\mathfrak{p}_i}(I) = v_{\mathfrak{p}_i}(\mathfrak{a}_i I)$, so such an a_i exists. Each a_i is necessarily nonzero, and satisfies $v_{\mathfrak{p}_i}(a) = v_{\mathfrak{p}_i}(I)$ since

$$v_{\mathfrak{p}_i}(a_i) \geq v_{\mathfrak{p}_i}(\mathfrak{a}_i I) = v_{\mathfrak{p}_i}(I) \qquad \text{and} \qquad v_{\mathfrak{p}_i}(a_i) < v_{\mathfrak{p}_i}(\mathfrak{p}_i I) = v_{\mathfrak{p}_i}(I) + 1,$$

and for $j \neq i$ we have $v_{\mathfrak{p}_j}(a_i) \geq v_{\mathfrak{p}_j}(\mathfrak{p}_j I) > v_{\mathfrak{p}_j}(I)$. We now define $a := a_1 + \cdots + a_n$, so that $v_{\mathfrak{p}_i}(a) = v_{\mathfrak{p}_i}(a_i) = v_{\mathfrak{p}_i}(I)$ for $1 \leq i \leq n$ (by the nonarchimedean triangle equality; see Problem Set 1). We thus have $v_{\mathfrak{p}}(a) = v_{\mathfrak{p}}(I)$ for all prime ideals $\mathfrak{p}|I'$.

Now (a) is contained in I and therefore divisible by I (since A is a Dedekind domain), so $(a) = IJ$ for some ideal J. For each prime ideal $\mathfrak{p}|I'$ we have $v_{\mathfrak{p}}(J) = v_{\mathfrak{p}}(a) - v_{\mathfrak{p}}(I) = 0$, so J is coprime to I', and $IJ = (a)$ is principal as desired. □

[4] It may seem strange at first glance that the greatest common divisor of I and J is the *smallest* ideal dividing I and J, but note that if $A = \mathbb{Z}$ then $\gcd((a), (b)) = (\gcd(a, b))$ for any $a, b \in \mathbb{Z}$, so the terminology is consistent (note that bigger numbers generate smaller ideals).

One can show that every integral domain satisfying Lemma 3.20 is a Dedekind domain (see Problem Set 2).

Corollary 3.21 (Finite approximation)**.** *Let I be a nonzero fractional ideal in a Dedekind domain A and let $\mathfrak{p}_1, \dots, \mathfrak{p}_n$ be a finite set of nonzero prime ideals of A. Then I contains an element x for which $v_{\mathfrak{p}_i}(x) = v_{\mathfrak{p}_i}(I)$ for $1 \le i \le n$.*

Proof. Let $I = \frac{1}{s}J$ with $s \in A$ and J an ideal. As in the proof of Lemma 3.20, we can pick $a \in J$ so that $v_{\mathfrak{p}_i}(a) = v_{\mathfrak{p}_i}(J)$ for $1 \le i \le n$. If we now let $x = a/s$ then we have $v_{\mathfrak{p}_i}(x) = v_{\mathfrak{p}_i}(a) - v_{\mathfrak{p}_i}(s) = v_{\mathfrak{p}_i}(J) - v_{\mathfrak{p}_i}(s) = v_{\mathfrak{p}_i}(I)$ for $1 \le i \le n$ as desired. □

Corollary 3.22. *Let I be a nonzero ideal in a Dedekind domain A. The quotient ring A/I is a principal ideal ring (every ideal in A/I is principal).*

Proof. Let $\varphi\colon A \to A/I$ be the quotient map, let $\bar{J}$ be an (A/I)-ideal and let $J := \varphi^{-1}(J)$ be its inverse image in A; then $I \subseteq J$, and $\bar{J} \simeq J/I$ as (A/I)-modules. By Corollary 3.21 we may choose $a \in J$ so that $v_{\mathfrak{p}}(a) = v_{\mathfrak{p}}(J)$ for all nonzero prime ideals $\mathfrak{p}|I$. For every nonzero prime ideal $\mathfrak{p}$ we then have $v_{\mathfrak{p}}(J) \le v_{\mathfrak{p}}(I)$ and

$$v_{\mathfrak{p}}((a)+I) = \begin{cases} \min(v_{\mathfrak{p}}(a), v_{\mathfrak{p}}(I)) = v_{\mathfrak{p}}(a) = v_{\mathfrak{p}}(J) & \text{if } \mathfrak{p}|I, \\ \min(v_{\mathfrak{p}}(a), v_{\mathfrak{p}}(I)) = v_{\mathfrak{p}}(I) = 0 = v_{\mathfrak{p}}(J) & \text{if } \mathfrak{p} \nmid I, \end{cases}$$

so $(a) + I = J$ (here we are using unique factorization of ideals; in a Dedekind domain two ideals with the same valuation at every nonzero prime ideal must be equal). If follows that $\bar{J} \simeq J/I = ((a)+I)/I = \varphi((a)) = (\varphi(a))$ is principal. □

The converse of Corollary 3.22 also holds; an integral domain whose quotients by nonzero ideals are principal ideal rings is a Dedekind domain (see Problem Set 2).

Definition 3.23. A ring that has only finitely many maximal ideals is called *semilocal*.

Example 3.24. The ring $\mathbb{Z}_{(3)} \cap \mathbb{Z}_{(5)}$ is semilocal, it has just two maximal ideals.

Corollary 3.25. *Every semilocal Dedekind domain is a principal ideal domain.*

Proof. If we let I' be the product of all the prime ideals in A and apply Lemma 3.20 to any ideal I we will necessarily have $J = A$ and $IJ = I$ principal. □

Theorem 3.26. *Let I be a nonzero ideal in a Dedekind domain A and let $a \in I$ be nonzero. Then $I = (a, b)$ for some $b \in I$.*

Proof. We have $(a) \subseteq I$, so I divides (a) and we have $II' = (\alpha)$ for some nonzero ideal I'. By Lemma 3.20 there is an ideal J coprime to I' such that IJ is principal, so $IJ = (b)$ for some $b \in I$. We have $\gcd((a),(b)) = \gcd(II', IJ) = I$, since $\gcd(I', J) = (1)$, and it follows that $I = (a, b)$. □

Theorem 3.26 gives us a convenient way to represent ideals I in the ring of integers of a global field. We can always pick $a \in \mathbb{Z}$ or $a \in \mathbb{F}_q[t]$; we will see in later lectures that there is a natural choice for a (the absolute norm of I). It also gives us yet another characterization of Dedekind domains: they are precisely the integral domains for which Theorem 3.26 holds.

We end this section with a theorem that summarizes the various equivalent definitions of a Dedekind domain that we have seen.

Theorem 3.27. *Let A be an integral domain. The following are equivalent:*

- *A is an integrally closed noetherian domain of dimension at most one.*
- *A is noetherian and its localizations at nonzero prime ideals are DVRs.*
- *Every nonzero ideal in A is invertible.*
- *Every nonzero ideal in A is a (finite) product of prime ideals.*
- *A is noetherian and "to contain is to divide" holds for ideals in A.*
- *For every ideal I in A there is an ideal J in A such that IJ is principal.*
- *Every quotient of A by a nonzero ideal is a principal ideal ring.*
- *For every nonzero ideal I in A and nonzero $a \in I$ we have $I = (a, b)$ for some $b \in I$.*

Proof. See Problem Set 2. □

References

[1] Pete L. Clark, *Commutative algebra*, 2015.

[2] Max D. Larsen and Paul J. McCarthy, *Multiplicative thoery of ideals*, Academic Press, 1971.

4 Étale algebras, norm and trace

4.1 Separability

In this section we briefly review some standard facts about separable and inseparable field extensions that we will use repeatedly throughout the course. Those familiar with this material should feel free to skim it. In this section K denotes any field, $\overline{K}$ is an algebraic closure that we will typically choose to contain any extensions L/K under consideration, and for any polynomial $f = \sum a_i x^i \in K[x]$ we use $f' := \sum i a_i x^{i-1}$ to denote the formal derivative of f (this definition also applies when K is an arbitrary ring).

Definition 4.1. A polynomial f in $K[x]$ is *separable* if $(f, f') = (1)$, that is, $\gcd(f, f')$ is a unit in $K[x]$. Otherwise f is *inseparable*.

If f is separable then it splits into distinct linear factors over over $\overline{K}$, where it has $\deg f$ distinct roots; this is sometimes used as an alternative definition. Note that the proper of separability is intrinsic to the polynomial f, it does not depend on the field we are working in; in particular, if L/K is any field extension whether or not a polynomial in $f \in K[x] \subseteq L[x]$ does not depend on whether we view f as an element of $K[x]$ or $L[x]$.

Warning 4.2. Older texts (such as Bourbaki) define a polynomial in $K[x]$ to be separable if all of its irreducible factors are separable (under our definition); so $(x-1)^2$ is separable under this older definition, but not under ours. This discrepancy does not change the definition of separable elements or field extensions.

Definition 4.3. Let L/K be an algebraic field extension. An element $\alpha \in L$ is *separable over* K if it is the root of a separable polynomial in $K[x]$ (in which case its minimal polynomial is necessarily separable). The extension L/K is *separable* if every $\alpha \in L$ is separable over K; otherwise it is *inseparable*.

Lemma 4.4. *An irreducible polynomial $f \in K[x]$ is inseparable if and only if $f' = 0$.*

Proof. Let $f \in K[x]$ be irreducible; then f is nonzero and not a unit, so $\deg f > 0$. If $f' = 0$ then $\gcd(f, f') = f \notin K^\times$ and f is inseparable. If f is inseparable then $g := \gcd(f, f')$ is a nontrivial divisor of f and f'. This implies $\deg g = \deg f$, since f is irreducible, but then $\deg f' < \deg f = \deg g$, so g cannot divide f' unless $f' = 0$. □

Corollary 4.5. *Let $f \in K[x]$ be irreducible and let $p \geq 0$ be the characteristic of K. We have $f(x) = g(x^{p^n})$ for some irreducible separable $g \in K[x]$ and integer $n \geq 0$ that are uniquely determined by f.*

Proof. If f is separable the theorem holds with $g = f$ and $n = 0$; for uniqueness, note that if $p = 0$ then $p^n \neq 0$ if and only if $n = 0$, and if $p > 0$ and $g(x^{p^n})$ is inseparable unless $n = 0$ because $g(x^{p^n})' = g'(x^{p^n})p^n x^{p^n - 1} = 0$ (by the previous lemma). Otherwise $f(x) := \sum f_r x^r$ is inseparable and $f'(x) = \sum r f_r x^{r-1} = 0$ (by the lemma), and this can occur only if $p > 0$ and $f_r = 0$ for all $r \geq 0$ not divisible by p. So $f = g(x^p)$ for some (necessarily irreducible) $g \in K[x]$. If g is separable we are done; otherwise we proceed by induction. As above, the uniqueness of g and n is guaranteed by the fact that $g(x^{p^n})' = 0$ for all $n > 0$. □

Corollary 4.6. *If* $\operatorname{char} K = 0$ *then every algebraic extension of K is separable.*

Lemma 4.7. *Let $L = K(\alpha)$ be an algebraic field extension contained in an algebraic closure $\overline{K}$ of K and let $f \in K[x]$ be the minimal polynomial of α over K. Then*

$$\#\operatorname{Hom}_K(L, \overline{K}) = \#\{\beta \in \overline{K} : f(\beta) = 0\} \leq [L : K],$$

with equality if and only if α is separable over K.

Proof. Each element of $\operatorname{Hom}_K(L, \overline{K})$ is uniquely determined by the image of α, which must be a root β of $f(x)$ in $\overline{K}$. The number of these roots is equal to $[L : K] = \deg f$ precisely when f, and therefore α, is separable over K. □

Definition 4.8. Let L/K be a finite extension of fields. The *separable degree* of L/K is

$$[L : K]_s := \#\operatorname{Hom}_K(L, \overline{K}).$$

The *inseparable degree* of f is

$$[L : K]_i := [L : K]/[L : K]_s$$

We will see shortly that $[L : K]_s$ always divides $[L : K]$, so $[L : K]_i$ is an integer (in fact a power of the characteristic of K), but it follows immediately from our definition that

$$[L : K] = [L : K]_s[L : K]_i.$$

holds regardless.

Theorem 4.9. *Let L/K be an algebraic field extension. and let $\phi_K \colon K \to \Omega$ be a homomorphism to an algebraically closed field Ω. Then ϕ_K extends to a homomorphism $\phi_L \colon L \to \Omega$.*

Proof. We use Zorn's lemma. Define a partial ordering on the set $\mathcal{F}$ of pairs (F, ϕ_F) for which F/K is a subextension of L/K and $\phi_F \colon F \to \Omega$ extends ϕ_K by defining

$$(F_1, \phi_{F_1}) \leq (F_2, \phi_{F_2})$$

whenever F_2 contains F_1 and ϕ_{F_2} extends ϕ_{F_1}. Given any totally ordered subset $\mathcal{C}$ of $\mathcal{F}$, let E be the field $\bigcup\{F : (F, \phi_F) \in \mathcal{C}\}$ and define $\phi_E \colon E \to \Omega$ by $\phi_E(x) = \phi_F(x)$ for $x \in F \subseteq E$ (this does not depend on the choice of F because $\mathcal{C}$ is totally ordered). Then (E, ϕ_E) is a maximal element of $\mathcal{C}$, and by Zorn's lemma, $\mathcal{F}$ contains a maximal element (M, ϕ_M).

We claim that $M = L$. If not, then pick $\alpha \in L - M$ and consider the field $F = M(\alpha) \subseteq L$ properly containing M, and extend ϕ_M to $\varphi_F \colon F \to \Omega$ be letting $\phi_F(\alpha)$ be any root of $\alpha_M(f)$ in Ω, where $f \in M[x]$ is the minimal polynomial of α over M and $\alpha_M(f)$ is the image of f in $\Omega[x]$ obtained by applying ϕ_M to each coefficient. Then (M, ϕ_M) is strictly dominated by (F, ϕ_F), contradicting its maximality. □

Lemma 4.10. *Let $L/F/K$ be a tower of finite extensions of fields. Then*

$$\#\operatorname{Hom}_K(L, \overline{K}) = \#\operatorname{Hom}_K(F, \overline{K})\#\operatorname{Hom}_F(L, \overline{K}).$$

Proof. We decompose $L/F/K$ into a tower of simple extensions and proceed by induction. The result is trivial if $L = K$ and otherwise it suffices to consider $K \subseteq F \subseteq F(\alpha) = L$, where $K = F$ in the base case. Theorem 4.9 allows us to define a bijection

$$\operatorname{Hom}_K(F, \overline{K}) \times \operatorname{Hom}_F(F(\alpha), \overline{K}) \to \operatorname{Hom}_K(F(\alpha), \overline{K})$$

that sends (ϕ_1, ϕ_2) to $\phi \colon L \to \overline{K}$ defined by $\phi|_F = \phi_1$ and $\phi(\alpha) = (\hat{\phi}_1\hat{\phi}_2\hat{\phi}_1^{-1})(\alpha)$, where $\hat{\phi}_1, \hat{\phi}_2 \in \operatorname{Aut}_K(\overline{K})$ are arbitrary extensions of ϕ_1, ϕ_2 to $\overline{K}$; note that $\phi(\alpha)$ does not depend on these choices and is a root of $\phi(f)$, where $f \in F[x]$ is the minimal polynomial of α and $\phi(f)$ is its image in $\phi(F)[x]$. The inverse bijection is $\phi_1 = \phi|_F$ and $\phi_2(\alpha) = (\hat{\phi}_1^{-1}\hat{\phi}\hat{\phi}_1)(\alpha)$. □

Corollary 4.11. *Let $L/F/K$ be a tower of finite extensions of fields. Then*

$$[L:K]_s = [L:F]_s[F:K]_s$$
$$[L:K]_i = [L:F]_i[F:K]_i$$

Proof. The first equality follows from the lemma and the second follows from the identities $[L:K] = [L:F][F:K]$ and $[L:K] = [L:K]_s[L:K]_i$. □

Theorem 4.12. *Let L/K be a finite extension of fields. The following are equivalent:*

(a) *L/K is separable;*

(b) *$[L:K]_s = [L:K]$;*

(c) *$L = K(\alpha)$ for some $\alpha \in L$ separable over K;*

(d) *$L \simeq K[x]/(f)$ for some monic irreducible separable polynomial $f \in K[x]$.*

Proof. The equivalence of (c) and (d) is immediate (let f be the minimal polynomial of α and let α be the image of x in $K[x]/(f)$), and the equivalence of (b) and (c) is given by Lemma 4.7. That (a) implies (c) is the PRIMITIVE ELEMENT THEOREM, see [2, §15.8] or [3, §V.7.4] for a proof. It remains only to show that (c) implies (a).

So let $L = K(\alpha)$ with α separable over K. For any $\beta \in L$ we can write $L = K(\beta)(\alpha)$, and we note that α is separable over $K(\beta)$, since its minimal polynomial over $K(\beta)$ divides it minimal polynomial over K, which is separable. Lemma 4.7 implies $[L:K]_s = [L:K]$ and $[L:K(\beta)]_s = [L:K(\beta)]$ (since $L = K(\alpha) = K(\beta)(\alpha)$), and the equalities

$$[L:K] = [L:K(\beta)][K(\beta):K]$$
$$[L:K]_s = [L:K(\beta)]_s[K(\beta):K]_s$$

then imply $[K(\beta):K]_s = [K(\beta):K]$. So β is separable over K (by Lemma 4.7). This applies to every $\beta \in L$, so L/K is separable and (a) holds. □

Corollary 4.13. *Let L/K be a finite extension of fields. Then $[L:K]_s \le [L:K]$ with equality if and only if L/K is separable.*

Proof. We have already established this for simple extensions, and otherwise we my decompose L/K into a finite tower of simple extensions and proceed by induction on the number of extensions, using the previous two corollaries at each step. □

Corollary 4.14. *Let $L/F/K$ is a tower of finite extensions of fields. Then L/K is separable if and only if both L/F and F/K separable.*

Proof. The forward implication is immediate and the reverse implication follows from Corollaries 4.11 and 4.13. □

Corollary 4.15. *Let $L/F/K$ be a tower of algebraic field extensions. Then L/K is separable if and only if both L/F and F/K are separable.*

Proof. As in the previous corollary the forward implication is immediate. To prove the reverse implication, we assume L/F and F/K are separable and show that every $\beta \in L$ is separable over K. If $\beta \in F$ we are done, and if not we at least know that β is separable over F. Let M/K be the subextension of F/K generated by the coefficients of the minimal polynomial $f \in F[x]$ of β over F. This is a finite separable extension of K, and $M(\beta)$ is also a finite separable extension of M, since the minimal polynomial of β over $M(\beta)$ is f, which is separable. By the previous corollary, $M(\beta)$, and therefore β, is separable over K. □

Corollary 4.16. *Let L/K be an algebraic field extension, and let*

$$F = \{\alpha \in L : \alpha \text{ is separable over } K\}.$$

Then F is a separable field extension of K.

Proof. This is clearly a field, since if α and β are both separable over K then $K(\alpha)$ and $K(\alpha, \beta)$ are separable extensions of K (by the previous corollary), thus every element of $K(\alpha, \beta)$, including $\alpha\beta$ and $\alpha + \beta$, is separable over K and lies in F. The field F is then separable by construction. □

Definition 4.17. Let L/K be an algebraic field extension. The field F in Corollary 4.16 is the *separable closure of K in L*. When L is an algebraic closure of K it is simply called a *separable closure* of K and denoted K^{sep}.

When K has characteristic zero the notions of separable closure and algebraic closure necessarily coincide. This holds more generally whenever K is a perfect field.

Definition 4.18. A field K is *perfect* if every algebraic extension of K is separable.

All fields of characteristic zero are perfect. Perfect fields of positive characteristic are characterized by the following property.

Theorem 4.19. *A field K of characteristic $p > 0$ is perfect if and only if $K = K^p$, that is, every element of K is a pth power, equivalently, the map $x \mapsto x^p$ is an automorphism.*

Proof. If $K \neq K^p$ then for any $\alpha \in K - K^p$ the polynomial $x^p - \alpha$ is irreducible and the extension $K[x]/(x^p - \alpha)$ is inseparable, implying that K is not perfect. Now suppose $K = K^p$ and let $f \in K[x]$ be irreducible. By Corollary 4.5, we have $f(x) = g(x^{p^n})$ for some separable $g \in K[x]$ and $n \geq 0$. If $n > 0$ then

$$f(x) = g(x^{p^n}) = \tilde{g}(x^{p^{n-1}})^p,$$

where $\tilde{g}$ is the polynomial obtained from g by replacing each coefficient with its pth root (thus $\tilde{g}(x)^p = g(x^p)$, since we are in characteristic p). But this contradicts the irreducibility of f. So $n = 0$ and $f = g$ is separable. The fact that every irreducible polynomial in $K[x]$ is separable implies that every algebraic extension of K is separable, so K is perfect. □

Corollary 4.20. *Every finite field is a perfect field.*

Proof. If a field K has cardinality p^n then $\#K^\times = p^n - 1$, thus $\alpha = \alpha^{p^n} = (\alpha^{p^{n-1}})^p$ for all $\alpha \in K$ and every element of K is a pth power. □

Definition 4.21. A field K is *separably closed* if K has no nontrivial finite separable extensions. Equivalently, K is equal to its separable closure in any algebraic closure of K.

Definition 4.22. An algebraic extension L/K is *purely inseparable* if $[L : K]_s = 1$.

Remark 4.23. The trivial extension K/K is both separable and purely inseparable (but not inseparable!); conversely, an extension that is separable and purely inseparable is trivial.

Example 4.24. If $K = \mathbb{F}_p(t)$ and $L = K[x]/(x^p - t) = \mathbb{F}_p(t^{1/p})$, then L/K is a purely inseparable extension of degree p.

Proposition 4.25. *Let K be a field of characteristic $p > 0$. If L/K is purely inseparable of degree p then $L = K(a^{1/p}) \simeq K[x]/(x^p - a)$ for some $a \in K - K^p$.*

Proof. Every $\alpha \in L - K$ is inseparable over K, and by Corollary 4.5 its minimal polynomial over K is of the form $f(x) = g(x^p)$ with f monic. We have $1 < \deg f \leq [L : K] = p$, so $g(x)$ must be a monic polynomial of degree 1, which we can write as $g(x) = x - a$. Then $f(x) = x^p - a$, and we must have $a \notin K^p$ since f is irreducible (a difference of pth powers can be factored). We have $[L : K(\alpha)] = 1$, so $L = K(\alpha) \simeq K[x]/(x^p - a)$ as claimed. □

Theorem 4.26. *Let L/K be an algebraic extension and let F be the separable closure of K in L. Then L/F is purely inseparable.*

Proof. If L/K is separable then $L = F$ the theorem holds, so we assume otherwise, in which case the characteristic p of K must be nonzero. Fix an algebraic closure $\overline{K}$ of K that contains L. Let $\alpha \in L - F$ have minimal polynomial f over F. Use Corollary 4.5 to write $f(x) = g(x^{p^n})$ with $g \in F[x]$ irreducible and separable, and $n \geq 0$. We must have $\deg g = 1$, since otherwise the roots of g would be separable over F, and therefore over K, but not lie in in the separable closure F of K in L. Thus $f(x) = x^{p^n} - a$ for some $a \in F$ (since f is monic and $\deg g = 1$). Since we are in characteristic $p > 0$, we can factor f in $F(\alpha)[x]$ as

$$f(x) = x^{p^n} - \alpha^{p^n} = (x - \alpha)^{p^n}.$$

There is thus only one F-homomorphism from $F(\alpha)$ to $\overline{K}$. The same statement applies to any extension of F obtained by adjoining any set of elements of L (even an infinite set). Therefore $\#\operatorname{Hom}_F(L, \overline{K}) = 1$, so $[L : F]_s = 1$ and L/F is purely inseparable. □

Corollary 4.27. *Every algebraic extension L/K can be uniquely decomposed into a tower of algebraic extensions $L/F/K$ with F/K separable and L/F purely inseparable.*

Proof. By Theorem 4.26, we can take F to be the separable closure of K in L, and this is the only possible choice, since we must have $[L : F]_s = 1$. □

Corollary 4.28. *The inseparable degree of any finite extension of fields is a power of the characteristic.*

Proof. This follows from the proof of Theorem 4.26. □

4.2 Étale algebras

We now want to generalize the notion of a separable field extension. By Theorem 4.12, every finite separable extension L/K can be explicitly represented as $L = K[x]/(f)$ for some separable irreducible $f \in K[x]$. If f is not irreducible then we no longer have a field, but we do have a ring $K[x]/(f)$ that is also a K vector space, in which the ring multiplication is compatible with scalar multiplication. In other words, L is a (unital) commutative K-algebra whose elements are all separable over K. The notion of separability extends to elements of a K-algebra (even non-commutative ones): an element is separable over K if and only it is the root of some separable polynomial in $K[x]$ (in which case its minimal polynomial must be separable). Recall that the minimal polynomial of an element α of a K-algebra A is the monic generator of the kernel of the K-algebra homomorphism $K[x] \to A$ defined by $x \mapsto \alpha$; note that if A is not a field, minimal polynomials need not be irreducible.

It follows from the Chinese remainder theorem that if f is separable then the K-algebra $K[x]/(f)$ is isomorphic to a direct product of finite separable extensions of K. Indeed, if $f = f_1 \cdots f_n$ is the factorization of f into irreducibles in $K[x]$ then

$$\frac{K[x]}{(f)} = \frac{K[x]}{(f_1 \cdots f_n)} \simeq \frac{K[x]}{(f_1)} \times \cdots \times \frac{K[x]}{(f_n)},$$

where the isomorphism is both a ring isomorphism and a K-algebra isomorphism. The separability of f implies that the f_i are separable and the ideals (f_i) are pairwise coprime (this justifies our application of the Chinese remainder theorem). We thus obtain a K-algebra that is isomorphic to finite product of separable field extensions $K[x]/(f_i)$ of K. Algebras of this form are called *étale algebras* (or *separable algebras*).

Definition 4.29. Let K be a field. An *étale K-algebra* is a K-algebra L that is isomorphic to a finite product of separable field extensions of K. The dimension of an étale K-algebra is its dimension as a K-vector space. When this dimension is finite we say that L is a *finite étale K-algebra*. A homomorphism of étale K-algebras is a homomorphism of K-algebras (which means a ring homomorphism that commutes with scalar multiplication).

Remark 4.30. One can define the notion of an étale A-algebra for any noetherian domain A (we will consider this in a later lecture).

Example 4.31. If K is a separably closed field then every étale K-algebra A is isomorphic to $K^n = K \times \cdots \times K$ for some positive integer n (and therefore a finite étale K-algebra).

Étale algebras are *semisimple algebras*. Recall that a (not necessarily commutative) ring R is *simple* if it is nonzero and has no nonzero proper (two-sided) ideals, and R is *semisimple* if it is isomorphic to a nonempty finite product of simple rings $\prod R_i$.[1] A commutative ring is simple if and only if it is a field, and semisimple if and only if it is isomorphic to a finite product of fields; this applies in particular to commutative semisimple K-algebras. Every étale K-algebra is thus semisimple (but the converse does not hold).

The ideals of a semisimple commutative ring $R = \prod_{i=1}^n R_i$ are easy to describe; each corresponds to a subproduct. To see this, note that the projection maps $R \to R_i$ are surjective homomorphisms onto a simple ring, thus for any R-ideal I, its image in R_i is either the zero ideal or the whole ring (note that the image of an ideal under a surjective ring homomorphism is an ideal). In particular, for each index i, either every $(r_1, \ldots, r_n) \in I$ has $r_i = 0$ or some $(r_1, \ldots, r_n) \in I$ has $r_i = 1$; it follows that I is isomorphic to the product of the R_i for which I projects onto R_i.

Proposition 4.32. *Let $A = \prod K_i$ be a K-algebra written that is a product of field extensions K_i/K. Every surjective homomorphism $\varphi\colon A \to B$ of K-algebras corresponds to the projection of A on to a subproduct of its factors.*

Proof. The ideal $\ker \varphi$ is a subproduct of $\prod K_i$, thus $A \simeq \ker \varphi \times \operatorname{im} \varphi$ and $B = \operatorname{im} \varphi$ is isomorphic to the complementary subproduct. □

Proposition 4.32 can be viewed as a generalization of the fact that every surjective homomorphism of fields is an isomorphism.

Corollary 4.33. *The decomposition of an étale algebra into field extensions is unique up to permutation and isomorphisms of factors.*

[1] There are many equivalent (and a few inequivalent) definitions, but this is the simplest.

Proof. Let A be an étale K-algebra and suppose A is isomorphic (as a K-algebra) to two products of field extensions of K, say

$$\prod_{i=1}^{m} K_i \simeq A \simeq \prod_{j=1}^{n} L_j.$$

Composing with isomorphisms yields surjective K-algebra homomorphisms $\pi_i\colon \prod L_j \to K_i$ and $\pi_j\colon \prod K_i \to L_j$. Proposition 4.32 then implies that each K_i must be isomorphic to one of the L_j and each L_j must be isomorphic to one of the K_i (and $m = n$). □

Our main interest in étale algebras is that they naturally arise from (and are stable under) *base change*, a notion we now recall.

Definition 4.34. Let $\varphi\colon A \to B$ be a homomorphism of rings (so B is an A-module), and let M be any A-module. The tensor product of A-modules $M \otimes_A B$ is a B-module (with multiplication defined by $b(m \otimes b') := m \otimes bb'$) called the *base change* (or *extension of scalars*) of M from A to B. If M is an A-algebra then its base change to B is a B-algebra.

We have already seen one example of base change: if M is an A-module and $\mathfrak{p}$ is a prime ideal of A then $M_{\mathfrak{p}} = M \otimes_A A_{\mathfrak{p}}$ (this is another way to define the localization of a module).

Remark 4.35. Each $\varphi\colon A \to B$ determines a functor from the category of A-modules to the category of B-modules via base change. It has an adjoint functor called *restriction of scalars* that converts a B-module M into an A-module by the rule $am = \varphi(a)m$ (if φ is inclusion this amounts to restricting the scalar multiplication by B to the subring A).

The ring homomorphism $\varphi\colon A \to B$ will often be an inclusion, in which case we have a ring extension B/A (we may also take this view whenever φ is injective, which is necessarily the case if A is a field). We are specifically interested in the case where B/A is a field extension and M is a finite étale A-algebra.

Proposition 4.36. *Suppose L is a finite étale K-algebra and K'/K is any field extension. Then $L \otimes_K K'$ is a finite étale K'-algebra of the same dimension as L.*

Proof. Without loss of generality we assume that L is actually a field; if not L is a product of fields and we can apply the following argument to each of its factors.

By Theorem 4.12, $L \simeq K[x]/(f)$ for some separable $f \in K[x]$, and if $f = f_1 f_2 \cdots f_m$ is the factorization of f in $K'[x]$, we have isomorphisms of K'-algebras

$$L \otimes_K K' \simeq K'[x]/(f) \simeq \prod_i K'[x]/(f_i),$$

in which each factor $K'[x]/(f_i)$ is a finite separable extension of K' (as discussed above, this follows from the CRT because f is separable). Thus $L \otimes_K K'$ is a finite étale K'-algebra, and $\dim_K L = \deg f = \dim_{K'} K'[x]/(f)$, so the dimension is preserved. □

Example 4.37. Any finite dimensional real vector space V is a finite étale $\mathbb{R}$-algebra (with coordinate-wise multiplication with respect to some basis); the complex vector space $V \otimes_{\mathbb{R}} \mathbb{C}$ is then a finite étale $\mathbb{C}$-algebra of the same dimension.

Note that even when an étale K-algebra L is a field, the base change $L \otimes_K K'$ will often not be a field. For example, if $K = \mathbb{Q}$ and $L \neq \mathbb{Q}$ is a number field, then $L \otimes_K \mathbb{C}$ will never be a field, it will be isomorphic to a $\mathbb{C}$-vector space of dimension $[L : K] > 1$.

Remark 4.38. In the proof of Proposition 4.36 we made essential use of the fact that the elements of an étale K-algebra are separable. Indeed, the proposition does not hold if L is a finite semisimple commutative K-algebra that contains an inseparable element.

Corollary 4.39. *Let $L \simeq K[x]/(f)$ be a finite separable extension of a field K defined by an irreducible separable polynomial $f \in K[x]$. Let K'/K be any field extension, and let $f = f_1 \cdots f_m$ be the factorization of f into distinct irreducible polynomials $f_i \in K'[x]$. We have an isomorphism of finite étale K'-algebras*

$$L \otimes_K K' \simeq \prod_i K'[x]/(f_i)$$

where each $K'[x]/(f_i)$ is a finite separable field extension of K'.

Proof. This follows directly from the proof of Proposition 4.36. □

The following proposition gives several equivalent characterizations of finite étale algebras, including a converse to Corollary 4.39 (provided the field K is not too small). Recall that an element α of a ring is *nilpotent* if $\alpha^n = 0$ for some n, and a ring is *reduced* if it contains no nonzero nilpotents.

Theorem 4.40. *Let L be a commutative K-algebra of finite dimension and assume that the dimension of L is less than the cardinality of K. The following are equivalent:*

(a) *L is a finite étale K-algebra.*

(b) *Every nonzero element of L is separable over K.*

(c) *$L \otimes_K K'$ is reduced for every extension K'/K.*

(d) *$L \otimes_K K'$ is semisimple for every extension K'/K.*

(e) *$L = K[x]/(f)$ for some separable $f \in K[x]$.*

The implications (a) $\Leftrightarrow$ (b) $\Leftrightarrow$ (c) $\Leftrightarrow$ (d) $\Leftarrow$ (e) hold regardless of the dimension of L.

Proof. To show (a) $\Rightarrow$ (b), let $L = \prod_{i=1}^n K_i$ with each K_i/K separable, and consider $\alpha = (\alpha_1, \ldots, \alpha_n) \in L = \prod_{i=1}^n K_i$. Each $\alpha_i \in K_i$ is separable over K with separable minimal polynomial $f_i \in K[x]$, and α is a root of $f := \operatorname{lcm}\{f_1, \ldots, f_n\}$, which is separable (the LCM of a finite set of separable polynomials is separable), thus α is separable.

To show (b) $\Rightarrow$ (c), note that if $\alpha \in L$ is nonzero and separable over K it cannot be nilpotent (the minimal polynomial of a nonzero nilpotent is x^n for some $n > 1$ and is therefore not separable), and separability is preserved under base change.

The equivalence (c) $\Leftrightarrow$ (d) follows from Lemma 4.42 below.

To show (d) $\Rightarrow$ (a), we first note we can assume L is semisimple (take $K' = K$), and it suffices to treat the case where L is a field. By base-changing to the separable closure of K in L, we can further reduce to the case that L/K is a purely inseparable field extension. If $L = K$ we are done. Otherwise we may pick an inseparable $\alpha \in L$, and, as in the proof of Theorem 4.26, the minimal polynomial of α has the form $f(x) = x^{p^n} - a$ for some $a \in K$ and $n \geq 1$. Now consider

$$\gamma := \alpha \otimes 1 - 1 \otimes \alpha \in L \otimes_K L$$

We have $\gamma \neq 0$, since $\gamma \notin K$, but $\gamma^{p^n} = \alpha^{p^n} \otimes 1 - 1 \otimes \alpha^{p^n} = a \otimes 1 - 1 \otimes a = 0$, so γ is a nonzero nilpotent and $L \otimes_K L$ is not reduced, contradicting (c) $\Leftrightarrow$ (d).

We have $(e) \Rightarrow (a)$ form Corollary 4.39. For the converse, suppose $L = \prod_{i=1}^n L_i$ with each L_i/K a finite separable extension of K. Pick a monic irreducible separable polynomial $f_1(x)$ so that $L_1 \simeq K[x]/(f_1(x))$, and then do the same for $i = 2, \dots, n$ ensuring that each polynomial f_j we pick is not equal to f_i for any $i < j$. This can be achieved by replacing $f_j(x)$ with $f_j(x + a)$ for some $a \in K^\times$ if necessary. Here we use the fact that there are at least n distinct choices for a, under our assumption that the dimension of L is less than the cardinality of K (note that if $f(x)$ is irreducible then the polynomials $f(x + a)$ are irreducible and pairwise coprime as a ranges over K). The polynomials $f_1, \dots f_n$ are then coprime and separable, so their product f is separable and $L = K[x]/(f)$, as desired. □

Remark 4.41. K-algebras of the form $L = K[x]/(f(x))$ are *monogenic* (generated by one element). Theorem 4.40 implies that finite étale K-algebras are monogenic whenever the base field K is big enough. This always holds if K is infinite, but if K is a finite field then not every finite étale K-algebra is monogenic. The recent preprint [5] gives exact bounds on the maximal number of generators needed for a finite étale K-algebra over a finite field.

The following lemma is a standard exercise in commutative algebra that we include for the sake of completeness.

Lemma 4.42. *Let K be a field. A commutative K-algebra of finite dimension is semisimple if and only if it is reduced.*

Proof. If A is semisimple it is clearly reduced (otherwise we could project a nonzero nilpotent of A to a nonzero nilpotent in a field); we only need to prove the converse. Every ideal of a commutative K-algebra A is also a K-vector space; this implies that when $\dim_K A$ is finite A satisfies both the ascending and descending chain conditions and is therefore noetherian and artinian. This implies that A has finitely many maximal ideals $M_1, \dots M_n$ and that the intersection of these ideals (the radical of A) is equal to the set of nilpotent elements of A (the nilradical of A); see Exercises 19.12 and 19.13 in [1], for example.

Taking the product of the projection maps $A \twoheadrightarrow A/M_i$ yields a surjective ring homomorphism $\varphi\colon A \twoheadrightarrow \bigoplus_{i=1}^n A/M_i$ from A to a product of fields. If A is reduced then $\ker \varphi = \cap M_i = \{0\}$ and φ is an isomorphism, implying that A is semismiple. □

Proposition 4.43. *Suppose L is a finite étale K-algebra and Ω is a separably closed field extension of K. There is an isomorphism of finite étale Ω-algebras*

$$L \otimes_K \Omega \quad \xrightarrow{\sim} \quad \prod_{\sigma \in \mathrm{Hom}_K(L,\Omega)} \Omega$$

that sends $\beta \otimes 1$ to the vector $(\sigma(\beta))_\sigma$ for each $\beta \in L$.

Proof. We may reduce to the case that $L = K[x]/(f)$ is a separable field extension, and we may then factor $f(x) = (x - \alpha_1) \cdots (x - \alpha_n)$ over Ω, with the α_i are distinct. We have a bijection between $\mathrm{Hom}_K(K[x]/(f), \Omega)$ and the set $\{\alpha_i\}$: each $\sigma \in \mathrm{Hom}_K(K[x]/(f), \Omega)$ is determined by $\sigma(x) \in \{\alpha_i\}$, and for each α_i, the map $x \mapsto \alpha_i$ determines a K-algebra homomorphism $\sigma_i \in \mathrm{Hom}_K(K[x]/(f), \Omega)$. As in the proof of Proposition 4.36 we have Ω-algebra isomorphisms

$$\frac{K[x]}{(f)} \otimes_K \Omega \xrightarrow{\sim} \frac{\Omega[x]}{(f)} \xrightarrow{\sim} \prod_{i=1}^n \frac{\Omega[x]}{(x - \alpha_i)} \xrightarrow{\sim} \prod_{i=1}^n \Omega.$$

which map

$$x \otimes 1 \mapsto x \mapsto (\alpha_1, \dots, \alpha_n) \mapsto (\sigma_1(x), \dots, \sigma_n(x)).$$

The element $x \otimes 1$ generates $L \otimes_K \Omega$ as an Ω-algebra, and it follows that $\beta \otimes 1 \mapsto (\sigma(\beta))_\sigma$ for every $\beta \in L$. □

Remark 4.44. The proof of Proposition 4.43 does not require Ω to be separably closed, we could replace Ω with the compositum of the normal closure of the field extensions L_i/K in the decomposition of $L = \prod L_i$ into separable extensions of K (in the proof above we just needed f to split into linear factors).

Example 4.45. Let $L/K = \mathbb{Q}(i)/\mathbb{Q}$ and $\Omega = \mathbb{C}$. We have $\mathbb{Q}(i) \simeq \mathbb{Q}[x]/(x^2+1)$ and

$$\mathbb{Q}(i) \otimes_{\mathbb{Q}} \mathbb{C} \simeq \frac{\mathbb{Q}[x]}{(x^2+1)} \otimes_{\mathbb{Q}} \mathbb{C} \simeq \frac{\mathbb{C}[x]}{(x^2+1)} \simeq \frac{\mathbb{C}[x]}{(x-i)} \times \frac{\mathbb{C}[x]}{(x+i)} \simeq \mathbb{C} \times \mathbb{C}.$$

As $\mathbb{C}$-algebra isomorphisms, the corresponding maps are determined by

$$i \otimes 1 \mapsto x \otimes 1 \mapsto x \mapsto (x, x) \equiv (i, -i) \mapsto (i, -i).$$

Taking the base change of $\mathbb{Q}(i)$ to $\mathbb{C}$ lets us see the two distinct embeddings of $\mathbb{Q}(i)$ in $\mathbb{C}$, which are determined by the image of i. Note that $\mathbb{Q}(i)$ is canonically embedded in its base change $\mathbb{Q}(i) \otimes_{\mathbb{Q}} \mathbb{C}$ to $\mathbb{C}$ via $\alpha \mapsto \alpha \otimes 1$. We have

$$-1 = i^2 = (i \otimes 1)^2 = i^2 \otimes 1^2 = -1 \otimes 1 = -(1 \otimes 1)$$

Thus as an isomorphism of $\mathbb{C}$-algebras, the basis $(1 \otimes 1, i \otimes 1)$ for $\mathbb{Q}(i) \otimes_{\mathbb{Q}} \mathbb{C}$ is mapped to the basis $\big((1,1),(i,-i)\big)$ for $\mathbb{C} \times \mathbb{C}$. For any $(\alpha, \beta) \in \mathbb{C} \times \mathbb{C}$, the inverse image of

$$(\alpha, \beta) = \frac{\alpha+\beta}{2}(1,1) + \frac{\alpha-\beta}{2i}(i,-i)$$

in $\mathbb{Q}(i) \otimes \mathbb{C}$ under this isomorphism is

$$\frac{\alpha+\beta}{2}(1 \otimes 1) + \frac{\alpha-\beta}{2i}(i \otimes 1) = 1 \otimes \frac{\alpha+\beta}{2} + i \otimes \frac{\alpha-\beta}{2i}.$$

Now $\mathbb{R}/\mathbb{Q}$ is an extension of rings, so we can also consider the base change of the $\mathbb{Q}$-algebra $\mathbb{Q}(i)$ to $\mathbb{R}$. But note that $\mathbb{R}$ is not separably closed and in particular, it does not contain a subfield isomorphic to $\mathbb{Q}(i)$, thus Proposition 4.43 does not apply. Indeed, as an $\mathbb{R}$-module, we have $\mathbb{Q}(i) \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{R}^2$, but as an $\mathbb{R}$-algebra, $\mathbb{Q}(i) \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{C} \not\simeq \mathbb{R}^2$.

4.3 Norms and traces

We now introduce the norm and trace map associated to a finite free ring extension B/A. These are often defined only for field extensions, but in fact the same definition works without modification whenever B is a free A-module of finite rank. One can generalize further to projective modules (with some restrictions), but we will not need this.

Definition 4.46. Let B/A be a (commutative) ring extension in which B is a free A-module of finite rank. The (relative) *norm* $\mathrm{N}_{B/A}(b)$ and *trace* $\mathrm{T}_{B/A}(b)$ of b (down to A) are the determinant and trace of the A-linear multiplication-by-b map $B \to B$ defined by $x \mapsto bx$.

As a special case, note that if A is a field and B is a finite A-algebra (a field extension, for example) then B is an A-vector space of finite dimension, hence a free A-module of finite rank. In practice one computes the norm and trace by picking a basis for B as an A-module and computing the matrix of the multiplication-by-b map with respect to this basis; this is an $n \times n$ matrix with entries in A whose determinant and trace are basis independent.

It follows immediately from the definition that $\mathrm{N}_{B/A}$ is multiplicative, $\mathrm{T}_{B/A}$ is additive, we have group homomorphisms

$$\mathrm{N}_{B/A}\colon B^\times \to A^\times \qquad \text{and} \qquad \mathrm{T}_{B/A}\colon B \to A,$$

and if B_1/A and B_2/A are two ring extensions that are free A-modules of finite rank then

$$\mathrm{N}_{B_1\times B_2/A}(x) = \mathrm{N}_{B_1/A}(x_1)\mathrm{N}_{B_2/A}(x_2) \qquad \text{and} \qquad \mathrm{T}_{B_1\times B_2/A} = \mathrm{T}_{B_1/A}(x_1) + \mathrm{T}_{B_2/A}(x_2)$$

for all $x = (x_1, x_2) \in B_1 \times B_2$.

Example 4.47. Consider $A = \mathbb{R}$ and $B = \mathbb{C}$, which has the A-module basis $(1, i)$. For $b = 2 + 3i$ the matrix of $B \overset{\times b}{\to} B$ with respect to this basis can be written as $\left(\begin{smallmatrix} 2 & -3 \\ 3 & 2 \end{smallmatrix}\right)$, thus

$$\mathrm{N}_{\mathbb{C}/\mathbb{R}}(2+3i) = \det \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} = 13,$$
$$\mathrm{T}_{\mathbb{C}/\mathbb{R}}(2+3i) = \operatorname{tr} \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} = 4.$$

Warning 4.48. In order to write down the matrix of an A-linear transformation $B \to B$ with respect to basis for B as a free A-module of rank n, we not only need to pick a basis, we need to decide whether to represent elements of $B \simeq A^n$ as row vectors with linear transformations acting via matrix multiplication on the right, or as column vectors with linear transformations acting via matrix multiplication on the left. The latter convention is often implicitly assumed in the literature (as in the example above), but the former is often used in computer algebra systems (such as Magma).

We now verify that the norm and trace are well behaved under base change.

Lemma 4.49. *Let B/A be ring extension with B free of rank n over A, and let $\varphi\colon A \to A'$ be a ring homomorphism. The base change $B' = B \otimes_A A'$ of B to A' is a free A'-module of rank n and we for every $b \in B$*

$$\varphi(\mathrm{N}_{B/A}(b)) = \mathrm{N}_{B'/A'}(b \otimes 1) \qquad \textit{and} \qquad \varphi(\mathrm{T}_{B/A}(b)) = \mathrm{T}_{B'/A'}(b \otimes 1).$$

Proof. Let $b \in B$, let $(b_1, \dots, b_n)$ be a basis for B as an A-module, and let $M = (m_{ij}) \in A^{n\times n}$ be the matrix of $B \overset{\times b}{\to} B$ with respect to this basis. Then $(b_1 \otimes 1, \dots, b_n \otimes 1)$ is a basis for B' as an A'-module (thus B' is free of rank n over A') and $M' = (\varphi(m_{ij})) \in A'^{n\times n}$ is the matrix of $B' \overset{\times b\otimes 1}{\to} B'$, and we have

$$\varphi(\mathrm{N}_{B/A}(b)) = \varphi(\det M) = \det M' = \mathrm{N}_{B'/A'}(b \otimes 1)$$
$$\varphi(\mathrm{T}_{B/A}(b)) = \varphi(\operatorname{tr} M) = \operatorname{tr} M' = \mathrm{N}_{B'/A'}(b \otimes 1) \qquad \square$$

Theorem 4.50. *Let K be a field with separable closure Ω and let L be a finite étale K-algebra. For all $\alpha \in L$ we have*

$$\mathrm{N}_{L/K}(\alpha) = \prod_{\sigma\in \mathrm{Hom}_K(L,\Omega)} \sigma(\alpha) \qquad \textit{and} \qquad \mathrm{T}_{L/K}(\alpha) = \sum_{\sigma\in \mathrm{Hom}_K(L,\Omega)} \sigma(\alpha).$$

Proof. Let n be the rank of L as a K-module. By the previous lemma and Proposition 4.43,

$$\mathrm{N}_{L/K}(\alpha) = \mathrm{N}_{(L\otimes_K\Omega)/\Omega}(\alpha\otimes 1) = \mathrm{N}_{\Omega^n/\Omega}(\sigma_1(\alpha),\dots,\sigma_n(\alpha)) = \prod_{i=1}^n \sigma_i(\alpha).$$

The isomorphism $L\otimes_K\Omega \to \prod_\sigma \Omega = \Omega^n$ of Prop. 4.43 sends $\alpha\otimes 1$ to $(\sigma_1(\alpha),\dots,\sigma_n(\alpha))$. Using the standard basis for Ω^n, the matrix of multiplication-by-$(\sigma_1(\alpha),\dots,\sigma_n(\alpha))$ is just the diagonal matrix with $\sigma_i(\alpha)$ in the ith diagonal entry. Similarly,

$$\mathrm{T}_{L/K}(\alpha) = \mathrm{T}_{(L\otimes_K\Omega)/\Omega}(\alpha\otimes 1) = \mathrm{T}_{\Omega^n/\Omega}(\sigma_1(\alpha),\dots,\sigma_n(\alpha)) = \sum_{i=1}^n \sigma_i(\alpha). \qquad \square$$

The proof above demonstrates a useful trick: when working over a field that is not algebraically/separably closed, base change to an algebraic/separable closure. This often turns separable field extensions into étale algebras that are no longer fields.

Proposition 4.51. *Let L/K be a (not necessarily separable) finite extension, let $\overline{K}$ be an algebraic closure of K containing L. Let $\alpha\in L^\times$ have minimal polynomial $f\in K[x]$ with factorization $f(x)=\prod_{i=1}^d(x-\alpha_i)$ in $\overline{K}[x]$, and let $e=[L:K(\alpha)]$. We have*

$$\mathrm{N}_{L/K}(\alpha) = \prod_{i=1}^d \alpha_i^e \qquad \text{and} \qquad \mathrm{T}_{L/K}(\alpha) = e\sum_{i=1}^d \alpha_i.$$

In particular, if $f(x)=\sum_{i=0}^d a_i x^i$, then $\mathrm{N}_{L/K}(\alpha)=(-1)^{de}a_0^e$ and $\mathrm{T}_{L/K}(\alpha)=-ea_{d-1}$.

Proof. See Problem Set 2. $\square$

Corollary 4.52. *Let $M/L/K$ be a tower of finite extensions. Then*

$$\mathrm{N}_{M/K} = \mathrm{N}_{L/K}\circ\mathrm{N}_{M/L} \qquad \text{and} \qquad \mathrm{T}_{M/K} = \mathrm{T}_{L/K}\circ\mathrm{T}_{M/L}.$$

Proof. Fix a separable closure Ω of K that contains M. As in the proof of Lemma 4.10, each $\sigma\in\mathrm{Hom}_K(M,\Omega)$ can be identified with a pair (σ_1,σ_2) with $\sigma_1\in\mathrm{Hom}_L(M,\Omega)$ and $\sigma_2\in\mathrm{Hom}_K(L,\Omega)$. We then note that for any $\alpha\in M^\times$,

$$\mathrm{N}_{M/K}(\alpha) = \prod_{\sigma\in\mathrm{Hom}_K(M,\Omega)}\sigma(\alpha) = \prod_{\sigma_2\in\mathrm{Hom}_K(L,\Omega)}\sigma_2\left(\prod_{\sigma_1\in\mathrm{Hom}_L(M,\Omega)}\sigma_1(\alpha)\right) = N_{L/K}(N_{M/L}(\alpha)),$$

and $\mathrm{T}_{M/K}(\alpha)=\mathrm{T}_{L/K}(\mathrm{T}_{M/L}(\alpha))$ follows similarly by replacing products with sums. $\square$

Corollary 4.53. *Let A be an integrally closed domain with fraction field K and let L/K be a finite extension. if $\alpha\in L$ is integral over A then $\mathrm{N}_{L/K}(\alpha)\in A$ and $\mathrm{T}_{L/K}(\alpha)\in A$.*

Proof. This follows immediately from Propositions 1.28 and 4.51. $\square$

Corollary 4.52 actually holds in much greater generality.

Theorem 4.54 (Transitivity of Norm and Trace)**.** *Let $A\subseteq B\subseteq C$ be rings with C free of finite rank over B and B free of finite rank over A. Then C is free of finite rank over A and*

$$\mathrm{N}_{C/A} = \mathrm{N}_{B/A}\circ\mathrm{N}_{C/B} \qquad \text{and} \qquad \mathrm{T}_{C/A} = \mathrm{T}_{B/A}\circ\mathrm{T}_{C/B}.$$

Proof. See [3, §III.9.4]. $\square$

References

[1] Allen Altman and Steven Kleiman, *A term of commutative algebra*, Worldwide Center of Mathematics, 2013.

[2] Michael Artin, *Algebra*, 2nd edition, Pearson, 2010.

[3] Nicolas Bourbaki, *Algebra I: Chapters 1–3*, Springer, 1989.

[4] Nicolas Bourbaki, *Algebra II: Chapters 4–7*, Springer, 1989.

[5] Uriya First, Zinovy Reichstein, Santiago Salazar, *On the number of generators of a separable algebra over a finite field*, arXiv:1709.06982, 2017.

[6] Anthony W. Knapp, *Advanced Algebra*, Digital Second Edition, 2016.

[7] Joseph J. Rotman, *Advanced Modern Algebra*, 2nd edition, Graduate Studies in Mathematics **114**, AMS, 2010.

5 Dedekind extensions

In this lecture we prove that the integral closure of a Dedekind domain in a finite extension of its fraction field is also a Dedekind domain; this implies, in particular, that the ring of integers of a number field is a Dedekind domain. We then consider the factorization of prime ideals in Dedekind extensions.

5.1 Dual modules, pairings, and lattices

In this section we work in a more general setting, where A is any commutative (unital) ring.

Definition 5.1. Let A be a commutative ring and M an A-module. The *dual module* $M^\vee$ is the A-module $\mathrm{Hom}_A(M, A)$ with scalar multiplication $(af)(m) = af(m)$, where $a \in A$, $f \in \mathrm{Hom}_A(M, A)$, and $m \in M$. If $\varphi\colon M \to N$ is an A-module homomorphism, the dual homomorphism $\varphi^\vee\colon N^\vee \to M^\vee$ is defined by $\varphi^\vee(g)(m) = g(\varphi(m))$, for $g \in N^\vee$ and $m \in M$.

It is easy to check that taking duals preserves identity maps and is compatible with composition: if $\varphi_1 : M \to N$ and $\varphi_2 : N \to P$ are A-module homomorphisms, then $(\varphi_2\varphi_1)^\vee = \varphi_1^\vee\varphi_2^\vee$. We thus have a contravariant functor from the category of A-modules to itself. This functor is compatible with (finite) direct sums, $(M \oplus N)^\vee \simeq M^\vee \oplus N^\vee$.

Lemma 5.2. *Let A be a commutative ring. For all A-modules M and N the A-modules $(M \oplus N)^\vee$ and $M^\vee \oplus N^\vee$ are canonically isomorphic.*

Proof. We have inverse A-module homomorphisms $\varphi \mapsto (m \mapsto \varphi(m, 0), n \mapsto \varphi(0, n))$ and $(\phi, \psi) \mapsto ((m, n) \mapsto \phi(m) + \psi(n))$. □

If A is a field and M is finitely generated, then M is a vector space of finite dimension, $M^\vee$ is its dual space and we have $M^{\vee\vee} \simeq M$. In general not every A-module is isomorphic to its double dual; those that are are said to be *reflexive*.

We have already seen examples of reflexive modules: every invertible fractional ideal is isomorphic to the dual of its inverse, hence to its double dual, and is thus reflexive.

Proposition 5.3. *Let A be an integral domain with fraction field K and let M be a nonzero A-submodule of K. Then $M^\vee \simeq (A : M) := \{x \in K : xM \subseteq A\}$; in particular, if M is an invertible fractional ideal then $M^\vee \simeq M^{-1}$ and $M^{\vee\vee} \simeq M$.*

Proof. For any $x \in (A : M)$ the map $m \mapsto xm$ is an A-linear map from M to A, hence an element of $M^\vee$, and this defines an A-module homomorphism $\varphi\colon (A : M) \to M^\vee$, since the map $x \mapsto (m \mapsto xm)$ is itself A-linear. Since $M \subseteq K$ is a nonzero A-module, it contains some nonzero $a \in A$ (if $a/b \in M$, so is $ba/b = a$). If $f \in M^\vee$ and $m = b/c \in M$ then

$$f(m) = f\left(\frac{b}{c}\right) = \frac{ac}{ac} f\left(\frac{b}{c}\right) = \frac{b}{ac} f\left(\frac{ac}{c}\right) = \frac{b}{ac} f(a) = \frac{f(a)}{a} m,$$

where we have used the fact that $a_1 f(a_2/a_3) = a_2 f(a_1/a_3)$ for any $a_1, a_2, a_3 \in A$ with $a_1/a_3, a_2/a_3 \in M$, by the A-linearity of f. It follows that f corresponds to multiplication by $x = f(a)/a$, which lies in $(A : M)$ since $xm = f(m) \in A$ for all $m \in M$. The map $f \mapsto f(a)/a$ defines an A-module homomorphism $M^\vee \to (A : M)$ inverse to φ, so φ is an isomorphism. When M is an invertible fractional ideal we have $M^\vee \simeq (A : M) = M^{-1}$, by Lemma 2.20, and $M^{\vee\vee} \simeq (M^{-1})^{-1} = M$ follows. □

Example 5.4. As a $\mathbb{Z}$-module, we have $\mathbb{Q}^\vee = \{0\}$ because there are no non-trivial $\mathbb{Z}$-linear homomorphisms from $\mathbb{Q}$ to $\mathbb{Z}$; indeed, $\mathbb{Q}$ is a divisible group and $\mathbb{Z}$ contains no non-trivial divisible subgroups. It follows that $\mathbb{Q}^{\vee\vee} = \{0\}$ (but as $\mathbb{Q}$-modules we have $\mathbb{Q} \simeq \mathbb{Q}^\vee \simeq \mathbb{Q}^{\vee\vee}$). Similarly, the dual of any finite $\mathbb{Z}$-module (any finite abelian group) is the zero module, as is the double dual. More generally, if A is an integral domain every dual (and double dual) A-module must be torsion free, but not all A-modules are torsion free.

One situation where we can recover many of the standard results that hold for vector spaces of finite dimension (with essentially the same proofs), is when M is a free module of finite rank. In particular, not only is M reflexive, we have $M \simeq M^\vee$ (non-canonically) and may explicitly construct a dual basis.

Theorem 5.5. *Let A be a commutative ring and let M be a free A-module of rank n. Then $M^\vee$ is also a free A-module of rank n, and each basis $(e_1, \dots, e_n)$ of M uniquely determines a dual basis $(e_1^\vee, \dots e_n^\vee)$ of $M^\vee$ with the property*

$$e_i^\vee(e_j) = \delta_{ij} := \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Proof. If $n = 0$ then $M = M^\vee = \{0\}$ and the theorem holds. Now assume $n \geq 1$ and fix an A-basis $\mathbf{e} := (e_1, \dots, e_n)$ for M. For each $\mathbf{a} := (a_1, \dots, a_n) \in A^n$, define $f_{\mathbf{a}} \in M^\vee$ by setting $f_{\mathbf{a}}(e_i) = a_i$ and extending A-linearly. The map $\mathbf{a} \mapsto f_{\mathbf{a}}$ gives an A-module homomorphism $A^n \to M^\vee$ with inverse $f \mapsto (f(e_1), \dots, f(e_n))$ and is therefore an isomorphism. It follows that $M^\vee \simeq A^n$ is a free A-module of rank n.

Now let $e_i^\vee := f_{\hat{\imath}}$, where $\hat{\imath} := (0, \dots, 0, 1, 0, \dots, 0) \in A^n$ has a 1 in the ith position. Then $\mathbf{e}^\vee := (e_1^\vee, \dots, e_n^\vee)$ is a basis for $M^\vee$, since $(\hat{1}, \dots, \hat{n})$ is a basis for A^n, and $e_i^\vee(e_j) = \delta_{ij}$. The basis $\mathbf{e}^\vee$ is uniquely determined by $\mathbf{e}$: it must be the image of $(\hat{1}, \dots, \hat{n})$ under the isomorphism $\mathbf{a} \mapsto f_{\mathbf{a}}$ determined by $\mathbf{e}$. □

Definition 5.6. Let A be a commutative ring and M an A-module. A (bilinear) *pairing* on M is an A-linear map $\langle \cdot, \cdot \rangle \colon M \times M \to A$. Explicitly, this means that for all $u, v, w \in M$ and $\lambda \in A$ we have

$$\begin{aligned} \langle u + v, w \rangle &= \langle u, w \rangle + \langle v, w \rangle, \\ \langle u, v + w \rangle &= \langle u, v \rangle + \langle u, w \rangle, \\ \langle \lambda u, v \rangle &= \langle u, \lambda v \rangle = \lambda \langle u, v \rangle. \end{aligned}$$

If $\langle v, w \rangle = \langle w, v \rangle$ then $\langle \cdot, \cdot \rangle$ is *symmetric*, if $\langle v, w \rangle = -\langle w, v \rangle$ then $\langle \cdot, \cdot \rangle$ is *skew-symmetric*, and if $\langle v, v \rangle = 0$ then $\langle \cdot, \cdot \rangle$ is *alternating* (the last two are equivalent provided $\mathrm{char}(A) \neq 2$). The pairing $\langle \cdot, \cdot \rangle$ induces an A-module homomorphism

$$\begin{aligned} \varphi \colon M &\to M^\vee \\ m &\mapsto (n \mapsto \langle m, n \rangle) \end{aligned}$$

If $\ker \varphi = \{0\}$ then $\langle \cdot, \cdot \rangle$ is *nondegenerate*, and if φ is an isomorphism then $\langle \cdot, \cdot \rangle$ is *perfect*.

Every perfect pairing is necessarily nondegenerate. If M is a vector space of finite dimension the converse holds, but this is not true in general, not even for free modules of finite rank: consider the pairing $\langle x, y \rangle := 2xy$ on $\mathbb{Z}$, which is non-degenerate but not perfect.

If M is a free A-module with basis $(e_1,\dots,e_n)$ and $\langle\cdot,\cdot\rangle$ is a perfect pairing, we can apply the inverse of the isomorphism $\varphi\colon M \xrightarrow{\sim} M^\vee$ induced by the pairing to the dual basis $(e_1^\vee,\dots,e_n^\vee)$ given by Theorem 5.5 to obtain a basis $(e'_1,\dots,e'_n)$ for M that satisfies

$$\langle e'_i, e_j\rangle = \delta_{ij}.$$

When $\langle\cdot,\cdot\rangle$ is symmetric we can similarly recover $(e_1,\dots,e_n)$ from $(e'_1,\dots,e'_n)$ in the same way. We record this fact in the following proposition.

Proposition 5.7. *Let A be a commutative ring and let M be a free A-module of rank n with a perfect pairing $\langle\cdot,\cdot\rangle$. For each A-basis $(e_1,\dots,e_n)$ of M there is a unique basis $(e'_1,\dots,e'_n)$ for M such that $\langle e'_i, e_j\rangle = \delta_{ij}$.*

Proof. Existence follows from the discussion above: apply the inverse of the isomorphism $\varphi\colon V \to V^\vee$ induced by $\langle\cdot,\cdot\rangle$ to the dual basis $(e_1^\vee,\dots,e_n^\vee)$ given by Theorem 5.5 to obtain a basis $(e'_1,\dots,e'_n)$ for M with $e'_i = \varphi^{-1}(e_i^\vee)$. We then have $e_i^\vee = \varphi(e'_i) = m \mapsto \langle e'_i, m\rangle$ and

$$\langle e'_i, e_j\rangle = \varphi(e'_i)(e_j) = e_i^\vee(e_j) = \delta_{ij}$$

for $1 \le i,j \le n$. If $(f'_1,\dots,f'_n)$ is another basis for M with the same property then for each i we have $\langle e'_i - f'_i, e_j\rangle = \delta_{ij} - \delta_{ij} = 0$ for every e_j, and therefore $\langle e'_i - f'_i, m\rangle = 0$ for all $m \in M$, but then $e'_i - f'_i \in \ker\varphi = \{0\}$, since the perfect pairing $\langle\cdot,\cdot\rangle$ is nondegenerate, and therefore $f'_i = e'_i$ for each i; uniqueness follows. □

Remark 5.8. In what follows the commutative ring A in Proposition 5.7 will typically be a field K and the free A-module M will be a K-vector space that we will denote V. We may then use A to denote a subring of K and M to denote an A-submodule of V. A perfect paring $\langle\cdot,\cdot\rangle$ on the K-vector space V will typically not restrict to a perfect pairing on the A-module M. For example, the perfect pairing $\langle x, y\rangle = xy$ on $\mathbb{Q}$ does not restrict to a perfect pairing on the $\mathbb{Z}$-module $2\mathbb{Z}$ because the induced map $\varphi\colon 2\mathbb{Z} \to 2\mathbb{Z}^\vee$ defined by $\varphi(m) = (n \mapsto mn)$ is not surjective: the map $x \mapsto x/2$ lies in $2\mathbb{Z}^\vee = \mathrm{Hom}_{\mathbb{Z}}(2\mathbb{Z},\mathbb{Z})$ but it is not in the image of φ.

We now introduce the notion of a lattice in a vector space.

Definition 5.9. Let A be an integral domain with fraction field K and let V be a K-vector space of finite dimension. A (full) *A-lattice in V* is a finitely generated A-submodule M of V that spans V as a K-vector space.

Remark 5.10. Some authors require A-lattices to be free A-modules. When $A = \mathbb{Z}$ (or any PID) this is not a restriction because M is necessarily torsion-free (it lies in a vector space) and any finitely generated torsion-free module over a PID is free (by the structure theorem for finitely generated modules over a PID). But when A is not a PID, finitely generated torsion-free A-modules will typically *not be free*. We do not want to exclude this case! In particular if L/K is an extension of number fields the ring of integers $\mathcal{O}_L$ will typically not be a free $\mathcal{O}_K$-module (even though it is a free $\mathbb{Z}$-module, as we shall shortly prove), but we still want to treat $\mathcal{O}_L$ as an $\mathcal{O}_K$-lattice in L (this will be important in later lectures when we define the *different ideal* $\mathcal{D}_{L/K}$).

Definition 5.11. Let A be a noetherian domain with fraction field K, and let V be a K-vector space of finite dimension with a perfect pairing $\langle\cdot,\cdot\rangle$. If M is an A-lattice in V, its *dual lattice* (with respect to the perfect pairing $\langle\cdot,\cdot\rangle$ on V) is the A-module

$$M^* \coloneqq \{x \in V : \langle x, m\rangle \in A \text{ for all } m \in M\}.$$

It is clear that M^* is an A-submodule of V, but it is not clear that it is an A-lattice in V (it must be finitely generated and span V), nor is it obvious that it is isomorphic to the dual module $M^\vee$. In order to justify the term *dual lattice*, let us now prove both facts. We will need to use the hypothesis that A is noetherian, since in general the dual of a finitely generated A-module need not be finitely generated. Notice that $\langle\cdot,\cdot\rangle$ is a perfect pairing on the K-module V that need not restrict to a perfect pairing on the A-module M.

Theorem 5.12. *Let A be a noetherian domain with fraction field K, let V be a K-vector space with a perfect pairing $\langle\cdot,\cdot\rangle$, and let M be an A-lattice in V. The dual lattice M^* is an A-lattice in V isomorphic to $M^\vee$.*

Proof. Let $\mathbf{e} := (e_1,\dots,e_n)$ be a K-basis for V that lies in M, and let $\mathbf{e}' := (e'_1,\dots,e'_n)$ be the unique K-basis for V given by Proposition 5.7 that satisfies $\langle e'_i, e_j\rangle = \delta_{ij}$.

To show that M^* spans V we write a finite set S of generators for M in terms of the basis $\mathbf{e}$ with coefficients in K and let d be the product of all denominators that appear. We claim that $d\mathbf{e}'$ lies in M^*: for each e'_i and generator $m \in S$, if we put $m = \sum_j m_j e_j$ then

$$\langle de'_i, m\rangle = d\langle e'_i, \textstyle\sum_j m_j e_j\rangle = d\sum_j m_j\langle e'_i, e_j\rangle = d\sum_j m_j\delta_{ij} = dm_i \in A,$$

by our choice of d, and this implies $de'_i \in M^*$. Thus M^* contains a basis $d\mathbf{e}'$ for V.

We now show M^* is finitely generated. Let

$$N := \{a_1e_1 + \cdots + a_ne_n : a_1,\dots,a_n \in A\} \simeq A^n$$

be the free A-submodule of M spanned by $\mathbf{e}$. The A-module N contains a basis for V and is finitely generated, so it is an A-lattice in V. The K-basis $\mathbf{e}'$ for V lies in N^*, since $\langle e'_i, e_j\rangle = \delta_{ij} \in A$, and we claim it is an A-basis for N^*. Given $x \in N^*$, if we write $x = \sum_i x_i e'_i$ then $\langle x, e_i\rangle = x_i\langle e'_i, e_i\rangle = x_i$ lies in A, since $x \in N^*$, so x lies in the A-span of $\mathbf{e}'$. It follows that N^* is a free A-module of rank n, and in particular, a finitely generated module over a noetherian ring and therefore a noetherian module (a module whose submodules are all finitely generated); see [1, Thm. 16.19]. From the definition of the dual lattice we have $N \subseteq M \Rightarrow M^* \subseteq N^*$, so M^* is a submodule of a noetherian module, hence finitely generated.

We now show $M^* \simeq M^\vee$. We have an obvious A-module homomorphism $\varphi\colon M^* \to M^\vee$ given by $x \mapsto (m \mapsto \langle x, m\rangle)$, and the A-module homomorphism $\psi\colon M^\vee \to M^*$ defined by $f \mapsto \sum_i f(e_i)e'_i$ is the inverse of φ. Indeed, for any $x = \sum_i x_i e'_i \in M^*$ we have

$$\psi(\varphi(x)) = \sum_i \varphi(x)(e_i)e'_i = \sum_i \langle x, e_i\rangle e'_i = \sum_i\sum_j x_j\langle e'_j, e_i\rangle e'_i = \sum_i x_i e'_i = x,$$

and for any $f \in M^\vee$ and each generator $m_j = \sum m_j e_j$ for M we have

$$\varphi(\psi(f))(m) = \varphi(\textstyle\sum_i f(e_i)e'_i)(m) = \sum_i \varphi(f(e_i)e'_i)(m) = \sum_i\langle f(e_i)e'_i, \sum_j m_j e_j\rangle = f(m),$$

which implies $\varphi(\psi(f)) = f$ and $\varphi^{-1} = \psi$; thus φ is an isomorphism from M^* to $M^\vee$. □

Corollary 5.13. *Let A be a noetherian domain with fraction field K. If M_1, M_2 are A-lattices in K-vector spaces V_1, V_2 with perfect pairings $\langle\cdot,\cdot\rangle_1$, $\langle\cdot,\cdot\rangle_2$ (resp.), then $\langle\cdot,\cdot\rangle_1 + \langle\cdot,\cdot\rangle_2$ defines a perfect pairing on $V_1 \oplus V_2$ and $(M \oplus N)^* \simeq M^* \oplus N^*$.*

Proof. This follows from Lemma 5.2 and Theorem 5.12. □

Corollary 5.14. *Let A be a noetherian domain with fraction field K, let V be a K-vector space with a perfect pairing $\langle\cdot,\cdot\rangle$, and let M be a free A-lattice in V with A-basis $(e_1,\dots,e_n)$. The dual lattice M^* is a free A-lattice in V that has a unique A-basis $(e_1^*,\dots,e_n^*)$ that satisfies $\langle e_i^*, e_j\rangle = \delta_{ij}$.*

Proof. This follows from the proof of Theorem 5.12 with $N = M$ and $e_i^* = e_i'$. □

You might wonder whether $M^{**} = M$ for an A-lattice M in a vector space V. This is false in general, but it is true when A is a Dedekind domain and we have a symmetric perfect pairing on V. To prove this we first show that the dual lattice respects localization.

Lemma 5.15. *Let A be a noetherian domain with fraction field K, let V be a K-vector space of finite dimension with a perfect pairing $\langle\cdot,\cdot\rangle$, let M be an A-lattice in V, and let S be a multiplicative subset of A. Then $S^{-1}M$ and $S^{-1}M^*$ are $(S^{-1}A)$-lattices in V satisfying $(S^{-1}M)^* = S^{-1}M^*$.*

Proof. It is clear that $S^{-1}M$ are $S^{-1}M^*$ are both $S^{-1}A$-lattices: each contains a basis for V (since M and M^* do), and both are finitely generated as $S^{-1}A$-modules (since M and M^* are finitely generated as A-modules).

Let $m_1,\dots m_n$ be A-module generators for M (and therefore $S^{-1}A$-module generators for $S^{-1}M$). If x is an element of $(S^{-1}M)^*$ then for each m_i we have $\langle x, m_i\rangle = a_i/s_i$ for some $a_i \in A$ and $s_i \in S$, and if we put $s = s_1\cdots s_n$ then $\langle sx, m_i\rangle \in A$ for every m_i, hence for all $m \in M$; thus $sx \in M^*$ and $x \in S^{-1}M^*$. Conversely, if $x = y/s$ is an element of $S^{-1}M^*$ with $y \in M^*$ and $s \in S$, then $\langle y, m_i\rangle \in A$ for every m_i and $\langle x, m_i\rangle = \langle y, m_i\rangle/s \in S^{-1}A$ for every m_i, hence for all $m \in S^{-1}M$, and it follows that $x \in (S^{-1}M)^*$. □

Proposition 5.16. *Let A be a Dedekind domain with fraction field K, let V be a K-vector space of finite dimension with a symmetric perfect pairing $\langle\cdot,\cdot\rangle$, and let M be an A-lattice in V. Then $M^{**} = M$.*

Proof. By Proposition 2.6, it suffices to show $(M^{**})_{\mathfrak{p}} = M_{\mathfrak{p}}$ for each maximal ideal $\mathfrak{p}$ of A. By Lemma 5.15 we have $(M^{**})_{\mathfrak{p}} = M_{\mathfrak{p}}^{**}$, so it is enough to show that the proposition holds when A is replaced by one of its localizations $A_{\mathfrak{p}}$ (a DVR, since A is a Dedekind domain).

So let us assume A that is a DVR. Then A is a PID and M and M^* are both torsion-free modules over a PID, hence free A-modules. So let us choose an A-basis $(e_1,\dots,e_n)$ for M, and let $(e_1^*,\dots,e_n^*)$ be the unique dual A-basis for M^* that satisfies $\langle e_i^*, e_j\rangle = \delta_{ij}$ (given by Corollary 5.14). If we now let $(e_1^{**},\dots,e_n^{**})$ be the unique A-basis for M^{**} that satisfies $\langle e_i^{**}, e_j^*\rangle = \delta_{ij}$ and note that $\langle e_i, e_j^*\rangle = \delta_{ij}$ (since $\langle\cdot,\cdot\rangle$ is symmetric), by uniqueness, we must have $e_i^{**} = e_i$ for all i, and therefore $M^{**} = M$. □

5.2 Extensions of Dedekind domains

Let A be a Dedekind domain with fraction field K, let L/K be a finite extension, and let B be the integral closure of A in L. We wish to prove that B is a Dedekind domain, which we will do by showing that it is an A-lattice in L; this will imply, in particular, that B is finitely generated, which is really the only difficult thing to show. Let us first show that B spans L as a vector space (and in fact L is its fraction field).

Proposition 5.17. *Let A be a Dedekind domain with fraction field K, let L/K be a finite extension, and let B be the integral closure of A in L. Every element of L can be written as b/a with $a \in A$ and $b \in B$. In particular, B spans L as a K-vector space and L is the fraction field of B.*

Proof. Let $\alpha \in L$. By multiplying the minimal polynomial of α in $K[x]$ by the product of the denominators of its coefficients, we obtain a polynomial in $A[x]$:

$$g(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

with $a_n \neq 0$, that has α as a root. We can make this polynomial monic by replacing x with x/a_n and multiplying through by a_n^{n-1} to obtain

$$a_n^{n-1} g(x/a_n) = x^n + a_{n-1}x^{n-1} + a_n a_{n-2} x^{n-2} \cdots + a_n^{n-2} a_1 x + a_n^{n-1} a_0.$$

This is a monic polynomial with coefficients in A that has $a_n\alpha \in L$ as a root. Therefore $a_n\alpha \in B$, since B is the integral closure of A in L, and $\alpha = b/a_n$ for some $b \in B$ and $a_n \in A$ as claimed. It follows that B generates L as a K-vector space (we have $\alpha = b \cdot \frac{1}{a_n}$ with $\frac{1}{a_n} \in K$), and $B \subseteq L \subseteq \operatorname{Frac} B$ implies $L = \operatorname{Frac} B$ (no smaller field can contain B). □

Proposition 5.18. *Let A be a Dedekind domain with fraction field K, let L/K be a finite extension of fields, and let B be the integral closure of A in L. Then $\mathrm{N}_{L/K}(b) \in A$ and $\mathrm{T}_{L/K}(b) \in A$ for all $b \in B$.*

Proof. The minimal polynomial $f = \sum_{i=0}^{d} a_i x^i \in K[x]$ of b has coefficients in A, by Proposition 1.28, and it then follows from Proposition 4.51 that $N_{L/K}(b) = (-1)^{de} a_0^e \in A$ and $T_{L/K}(b) = -e a_{d-1} \in A$ (where $e = [L : K(b)] \in \mathbb{Z}$). □

Definition 5.19. Let B/A be a ring extension with B a free A-module of finite rank. The *trace pairing* on B is the map $B \times B \to A$ defined by

$$\langle x, y \rangle_{B/A} := \mathrm{T}_{B/A}(xy).$$

Theorem 5.20. *Let L be a commutative K-algebra of finite dimension. The trace pairing $\langle \cdot, \cdot \rangle_{L/K}$ is a symmetric bilinear pairing. It is a perfect pairing if and only if L is a finite étale K-algebra.*

Proof. Bilinearity follows from the K-linearity of the trace map $\mathrm{T}_{L/K}$, and symmetry is immediate. The fact that L is a K-vector space implies that the trace pairing is perfect if and only if it is nondegenerate.

If L is not reduced then the proposition holds, since it is not étale (by Theorem 4.40), and the trace pairing is degenerate: for any nonzero nilpotent x the map $y \mapsto T_{L/K}(xy)$ must be the zero map, since every xy is also nilpotent and the trace of any nilpotent element z is zero (the matrix of the multiplication-by-z map is nilpotent, so its trace is zero).

We now assume L is reduced, hence semisimple (by Lemma 4.42) and thus a product of fields. It suffices to consider the case that L is a field, since the trace pairing on a product of field extensions is nondegenerate if and only if the trace pairing on each factor is nondegenerate, and a product of field extensions is ètale if and only if each factor is ètale.

As proved on Problem Set 2, $T_{L/K}$ is the zero map if and only if the field extension L/K is inseparable. If $T_{L/K}$ is the zero map then the trace pairing is clearly degenerate, and otherwise we may pick $z \in L$ for which $\mathrm{T}_{L/K}(z) \neq 0$. Then for every $x \in L^\times$ we have $\langle x, z/x \rangle_{L/K} = \mathrm{T}_{L/K}(z) \neq 0$, so $x \mapsto \langle x, y \rangle_{L/K}$ is not the zero map, and it follows that the trace pairing is nondegenerate. □

Remark 5.21. Theorem 5.20 gives another equivalent definition of a finite étale K-algebra in addition to the six listed in Theorem 4.40: a finite étale K-algebra is a commutative K-algebra of finite dimension for which the trace pairing is a perfect pairing.

We now assume that L/K is separable. For the next several lectures we will be working in the following setting: A is a Dedekind domain with fraction field K, the extension L/K is finite separable, and B is the integral closure of A in L (which we will shortly prove is a Dedekind domain). As a convenient shorthand, we will write "assume $AKLB$" to indicate that we are using this setup.

Proposition 5.22. *Assume $AKLB$. Then B is an A-lattice in L, and in particular, B is finitely generated as an A-module.*

Proof. By Proposition 5.17, B spans L as a K-vector space, so it contains a basis $(e_1, \dots, e_n)$ for L as a K-vector space. Let $M \subseteq B$ be the A-span of $(e_1, \dots, e_n)$. Then M is an A-lattice in L contained in B, and it has a dual lattice M^* that contains the A-module

$$B^* := \{x \in L : \langle x, b\rangle_{L/K} \in A \text{ for all } b \in B\}.$$

Proposition 5.18 implies that $B \subseteq B^*$, and we thus have inclusions

$$M \subseteq B \subseteq B^* \subseteq M^*.$$

By Theorem 5.12, M^* is an A-lattice in L, hence finitely generated, hence noetherian. It follows that its A-submodule B is finitely generated and thus an A-lattice in L. □

Remark 5.23. When L/K is inseparable, B need not be finitely generated as an A-module, not even when A is a PID; see [2, Ex. 11, p. 205]. We used the separability hypothesis in order to get a perfect pairing, which plays a crucial role in the proof of Theorem 5.12.

Lemma 5.24. *Let B/A be an extension of domains with B integral over A, and let $\mathfrak{q}_0 \subsetneq \mathfrak{q}_1$ be primes of B. Then $\mathfrak{q}_0 \cap A \subsetneq \mathfrak{q}_1 \cap A$ and $\dim A \geq \dim B$.*

Proof. We first replace B with $B/\mathfrak{q}_0$ and replace A, $\mathfrak{q}_0$, and $\mathfrak{q}_1$ with their images in $B/\mathfrak{q}_0$ (the new B is integral over the new A, since the image of a monic polynomial in $A[x]$ is a monic polynomial in $(A/(\mathfrak{q}_0 \cap A))[x]$). Then $\mathfrak{q}_0 = (0)$ and $\mathfrak{q}_1$ is a nonzero prime ideal. Let $\alpha \in \mathfrak{q}_1$ be nonzero. Its minimal polynomial $x^n + a_{n-1}x^{n-1} + \cdots + a_0$ over K has coefficients in A (since $\alpha \in \mathfrak{q}_1 \subseteq B$ is integral over A), with $a_0 \neq 0$ (otherwise divide by x). We have $a_0 = -a_1\alpha - \cdots - \alpha^n \in \mathfrak{q}_1$, thus $0 \neq a_0 \in \mathfrak{q}_1 \cap A$. So $\mathfrak{q}_1 \cap A$ is not the zero ideal and therefore properly contains $\mathfrak{q}_0 \cap A = \{0\}$. We can apply this result repeatedly to any chain of distinct prime ideals in B to get a corresponding chain of distinct prime ideals in A. It follows that $\dim A \geq \dim B$. □

Theorem 5.25. *Let A be a Dedekind domain with fraction field K, let L/K be a finite separable extension, and let B be the integral closure of A in L. Then B is a Dedekind domain.*

Proof. Recall that we defined a Dedekind domain as an integrally closed noetherian domain of dimension at most one. Let us verify that each of these conditions holds:

- B is an integrally closed domain (by definition);
- B is finitely generated over the noetherian ring A (by Prop. 5.22), hence noetherian;
- B has dimension at most 1, since $\dim B \leq \dim A \leq 1$, by Lemma 5.24.

Thus B is a Dedekind domain. □

Remark 5.26. Theorem 5.25 holds without the assumption that L/K is separable. This follows from the Krull-Akizuki Theorem, see [4, Thm. 11.7] or [3, §VII.2.5], which is used to prove that B is noetherian even when it is not finitely generated as an A-module.

Corollary 5.27. *The ring of integers of a number field is a Dedekind domain.*

5.3 Splitting primes in Dedekind extensions

We continue in the $AKLB$ setup, in which A is a Dedekind domain, K is its fraction field, L/K is a finite separable[1] extension, and B is the integral closure of A, which we now know is a Dedekind domain with fraction field L. As we proved in earlier lectures, every nonzero ideal in a Dedekind domain can be uniquely factored into prime ideals. Understanding the ideal structure of a Dedekind domain thus boils down to understanding its prime ideals. In order to simplify the language, whenever we have a Dedekind domain A, by a *prime* of A (or of its fraction field K), we always mean a **nonzero prime ideal** of A.

If A has dimension zero then so does B, in which case there are no primes to consider, so we may as well assume $\dim A = 1$, in which case $\dim B = 1$ as well (if B is a field then so is $B \cap K = A$). Henceforth our $AKLB$ setup will include the assumption that $A \neq K$.

Given a prime $\mathfrak{p}$ of A, we can consider the ideal $\mathfrak{p}B$ it generates in B (its extension to B under the inclusion map). The ideal $\mathfrak{p}B$ need not be prime, but it can be uniquely factored into nonzero prime ideals in the Dedekind domain B. We thus have

$$\mathfrak{p}B = \prod_{\mathfrak{q}} \mathfrak{q}^{e_{\mathfrak{q}}},$$

where $\mathfrak{q}$ ranges over primes of B and the exponents $e_{\mathfrak{q}} \geq 0$ are zero for all but finitely many primes $\mathfrak{q}$. The primes $\mathfrak{q}$ for which $e_{\mathfrak{q}} > 0$ are said to *lie over* or *above* the prime ideal $\mathfrak{p}$. As an abuse of notation, we will often write $\mathfrak{q}|\mathfrak{p}$ to indicate this relationship (there is little risk of confusion, the prime ideal $\mathfrak{p}$ is maximal hence not divisible by any prime ideals of A other than itself).

Lemma 5.28. *Let A be a ring of dimension one contained in a Dedekind domain B. Let $\mathfrak{p}$ be a prime of A and let $\mathfrak{q}$ be a prime of B. Then $\mathfrak{q}|\mathfrak{p}$ if and only if $\mathfrak{q} \cap A = \mathfrak{p}$.*

Proof. If $\mathfrak{q}$ divides $\mathfrak{p}B$ then it contains $\mathfrak{p}B$ (to divide is to contain), and therefore $\mathfrak{q} \cap A$ contains $\mathfrak{p}B \cap A$ which contains $\mathfrak{p}$; the ideal $\mathfrak{p}$ is maximal and $\mathfrak{q} \cap A \neq A$ (since $1 \notin \mathfrak{q}$), so $\mathfrak{q} \cap A = \mathfrak{p}$. Conversely, if $\mathfrak{q} \cap A = \mathfrak{p}$ then $\mathfrak{q} = \mathfrak{q}B$ certainly contains $(\mathfrak{q} \cap A)B = \mathfrak{p}B$, and B is a Dedekind domain, so $\mathfrak{q}$ divides $\mathfrak{p}B$ (in a Dedekind domain to contain is to divide). □

Lemma 5.28 implies that contraction gives us a surjective map $\operatorname{Spec} B \to \operatorname{Spec} A$ defined by $\mathfrak{q} \mapsto \mathfrak{q} \cap A$; to see why it is surjective, note that $(0) \cap A = (0)$, and if $\mathfrak{p}$ is a nonzero element of $\operatorname{Spec} A$ then $\mathfrak{p}B$ is nonzero and not the unit ideal, and therefore divisible by at least one $\mathfrak{q} \in \operatorname{Spec} B$. The fibers of this map are finite; we use $\{\mathfrak{q}|\mathfrak{p}\}$ to denote the fiber above a prime $\mathfrak{p}$ of A.

The primes $\mathfrak{p}$ of A are all maximal ideals (since $\dim A = 1$), so each has an associated residue field $A/\mathfrak{p}$, and similarly for primes $\mathfrak{q}$ of B. If $\mathfrak{q}$ lies above $\mathfrak{p}$ then we may regard the residue field $B/\mathfrak{q}$ as a field extension of $A/\mathfrak{p}$: the kernel of the map $A \hookrightarrow B \to B/\mathfrak{q}$ is $\mathfrak{p} = A \cap \mathfrak{q}$, and the induced map $A/\mathfrak{p} = A/(\mathfrak{q} \cap A) \to B/\mathfrak{q}$ is a ring homomorphism of fields, hence injective.

Definition 5.29. Assume $AKLB$, and let $\mathfrak{p}$ be a prime of A. The exponent $e_{\mathfrak{q}}$ in the factorization $\mathfrak{p}B = \prod_{\mathfrak{q}|\mathfrak{p}} \mathfrak{q}^{e_{\mathfrak{q}}}$ is the *ramification index of* $\mathfrak{q}$, and the degree $f_{\mathfrak{q}} = [B/\mathfrak{q} : A/\mathfrak{p}]$

[1] Most of our proofs will not actually use the separability hypothesis (and even when they do, there may be another way to prove the same result, as with Theorem 5.25). In order to simplify the presentation we will use the separability assumption whenever it would be awkward not to. The cases we are most interested in (extensions of local and global fields) are going to be separable in any event.

of the corresponding residue field extension is the *residue degree* (or *inertia degree*) of $\mathfrak{q}$. In situations where more than one extension of Dedekind domains is under consideration, we may write $e_{\mathfrak{q}/\mathfrak{p}}$ for $e_{\mathfrak{q}}$ and $f_{\mathfrak{q}/\mathfrak{p}}$ for $f_{\mathfrak{q}}$.

Lemma 5.30. *Let A be a Dedekind domain with fraction field K, let $M/L/K$ be a tower of finite separable extension, and let B and C be the integral closures of A in L and M respectively. Then C is the integral closure of B in M, and if $\mathfrak{r}$ is a prime of M lying above a prime $\mathfrak{q}$ of L lying above a prime $\mathfrak{p}$ of K then $e_{\mathfrak{r}/\mathfrak{p}} = e_{\mathfrak{r}/\mathfrak{q}} e_{\mathfrak{q}/\mathfrak{p}}$ and $f_{\mathfrak{r}/\mathfrak{p}} = f_{\mathfrak{r}/\mathfrak{q}} f_{\mathfrak{q}/\mathfrak{p}}$.*

Proof. It follows from Proposition 1.20 that the integral closure of B in M lies in C, and it contains C, since $A \subseteq B$. We thus have a tower of Dedekind extensions $C/B/A$. If $\mathfrak{r}|\mathfrak{q}|\mathfrak{p}$ then the factorization of $\mathfrak{p}C$ in C refines the factorization of $\mathfrak{p}B$ in B, so $e_{\mathfrak{r}/\mathfrak{p}} = e_{\mathfrak{r}/\mathfrak{q}} e_{\mathfrak{q}/\mathfrak{p}}$, and the residue field embedding $A/\mathfrak{p} \hookrightarrow C/\mathfrak{r}$ factors as $A/\mathfrak{p} \hookrightarrow B/\mathfrak{q} \hookrightarrow C/\mathfrak{r}$, so $f_{\mathfrak{r}/\mathfrak{p}} = f_{\mathfrak{r}/\mathfrak{q}} f_{\mathfrak{q}/\mathfrak{p}}$. □

Example 5.31. Let $A \coloneqq \mathbb{Z}$, with $K \coloneqq \operatorname{Frac} A = \mathbb{Q}$, and let $L \coloneqq \mathbb{Q}(i)$ with $[L : K] = 2$.

The prime (5) factors in $B = \mathbb{Z}[i]$ into two distinct prime ideals:

$$5\mathbb{Z}[i] = (2+i)(2-i).$$

The prime $(2+i)$ has ramification index $e_{(2+i)} = 1$, and $e_{(2-i)} = 1$ as well. The residue field $\mathbb{Z}/(5)$ is isomorphic to the finite field $\mathbb{F}_5$, and we also have $\mathbb{Z}[i]/(2+i) \simeq \mathbb{F}_5$ (this can be determined by counting the $\mathbb{Z}[i]$-lattice points in a fundamental parallelogram of the sublattice $(2+i)$ in $\mathbb{Z}[i]$), so $f_{(2+i)} = 1$; we similarly have $f_{(2-i)} = 1$.

The prime (7) remains prime in $B = \mathbb{Z}[i]$; its prime factorization is simply

$$7\mathbb{Z}[i] = (7),$$

where the (7) on the RHS denotes a principal ideal in B (this is clear from context). The ramification index of (7) is thus $e_{(7)} = 1$, but its residue field degree is $f_{(7)} = 2$, because $\mathbb{Z}/(7) \simeq \mathbb{F}_7$, but $\mathbb{Z}[i]/(7) \simeq \mathbb{F}_{49}$ has dimension 2 has an $\mathbb{F}_7$-vector space.

The prime (2) factors as

$$(2) = (1+i)^2,$$

since $(1+i)^2 = (1+2i-1) = (2i) = (2)$ (note that i is a unit). You might be thinking that $(2) = (1+i)(1-i)$ factors into distinct primes, but note that $(1+i) = -i(1+i) = (1-i)$. Thus $e_{(1+i)} = 2$, and $f_{(1+i)} = 1$ because $\mathbb{Z}/(2) \simeq \mathbb{F}_2 \simeq \mathbb{Z}[i]/(1+i)$.

Let us now compute the sum $\sum_{\mathfrak{q}|\mathfrak{p}} e_{\mathfrak{q}} f_{\mathfrak{q}}$ for each of the primes $\mathfrak{p}$ we factored above:

$$\sum_{q|(2)} e_{\mathfrak{q}} f_{\mathfrak{q}} = e_{(1+i)} f_{(1+i)} = 2 \cdot 1 = 2,$$
$$\sum_{q|(5)} e_{\mathfrak{q}} f_{\mathfrak{q}} = e_{(2+i)} f_{(2+i)} + e_{(2-i)} f_{(2-i)} = 1 \cdot 1 + 1 \cdot 1 = 2,$$
$$\sum_{q|(7)} e_{\mathfrak{q}} f_{\mathfrak{q}} = e_{(7)} f_{(7)} = 2 \cdot 1 = 2.$$

In all three cases we obtain $2 = [\mathbb{Q}(i) : \mathbb{Q}]$; as we shall shortly prove, this is not an accident.

Example 5.32. Let $A \coloneqq \mathbb{R}[x]$, with $K \coloneqq \operatorname{Frac} A = \mathbb{R}(x)$, and let $L \coloneqq \mathbb{R}(\sqrt{x^3+3x})$. The integral closure of A in L is the Dedekind domain $B = \mathbb{R}[x,y]/(y^2 - x^3 - 3x)$. Then $[L : K] = 2$.

The prime $(x-1)$ factors in B into two distinct prime ideals:

$$(x-1) = (x-1, y-2)(x-1, y+2) \qquad (\text{since } y^2 - 4 = x^3 + 3x - 4 \in (x-1)).$$

We thus have $e_{(x-1,y-2)} = 1$, and $f_{(x-1,y-2)} = [B/(x-1, y-2) : A/(x-1)] = [\mathbb{R} : \mathbb{R}] = 1$. Similarly, $e_{(x-1,y+2)} = 1$ and $f_{(x-1,y+2)} = 1$.

The prime $(x+1)$ remains prime in B (because $y^2 = -1$ has no solutions in $\mathbb{R}$), thus $e_{(x+1)} = 1$, and $f_{(x+1)} = [B/(x+1) : A/(x+1)] \simeq [\mathbb{C} : \mathbb{R}] = 2$.

The prime (x) factors in B as

$$(x) = (x, y)^2,$$

and we have $e_{(x,y)} = 2$ and $f_{(x,y)} = 1$.

As in the previous example, $\sum_{\mathfrak{q}|\mathfrak{p}} e_{\mathfrak{q}} f_{\mathfrak{q}} = [L : K]$ in every case:

$$\sum_{\mathfrak{q}|(x-1)} e_{\mathfrak{q}} f_{\mathfrak{q}} = e_{(x-1,y-2)} f_{(x-1,y+2)} + e_{(x-1,y+2)} f_{(x-1,y+2)} = 1 \cdot 1 + 1 \cdot 1 = 2,$$

$$\sum_{\mathfrak{q}|(x+1)} e_{\mathfrak{q}} f_{\mathfrak{q}} = e_{(x+1)} f_{(x+1)} = 1 \cdot 2 = 2.$$

$$\sum_{\mathfrak{q}|(x)} e_{\mathfrak{q}} f_{\mathfrak{q}} = e_{(x,y)} f_{(x,y)} = 2 \cdot 1 = 2,$$

Before proving that $\sum_{\mathfrak{q}|\mathfrak{p}} e_{\mathfrak{q}} f_{\mathfrak{q}} = [L : K]$ always holds, let us consider the quotient ring $B/\mathfrak{p}B$. The ring $B/\mathfrak{p}B$ is typically not a field, so it is not a field extension of $A/\mathfrak{p}$, but it is an $A/\mathfrak{p}$-algebra. This follows from the fact that B contains A and $\mathfrak{p}B$ contains $\mathfrak{p}$: given $\bar{a} \in A/\mathfrak{p}$ and $\bar{x} \in B/\mathfrak{p}B$, if we choose lifts $a \in A$ of $\bar{a}$ and $x \in B$ of $\bar{x}$ then $\bar{a}\bar{x} = \overline{ax} \in B/\mathfrak{p}B$ is the reduction of $ax \in b$ and does not depend on the choice of a and x since any other choices would be congruent modulo $\mathfrak{p}B$.

Lemma 5.33. *Assume AKLB and let $\mathfrak{p}$ be a prime of A. The dimension of $B/\mathfrak{p}B$ as an $A/\mathfrak{p}$-vector space is equal to the dimension of L as a K-vector space.*

Proof. Let $A_{\mathfrak{p}} := S^{-1}A$ and $B_{\mathfrak{p}} := S^{-1}B$ be localizations of A and B (as A-modules), where $S = A - \mathfrak{p}$. Then $A_{\mathfrak{p}}/\mathfrak{p}A_{\mathfrak{p}} = S^{-1}A/(\mathfrak{p}S^{-1}A) \simeq A/\mathfrak{p}$ and $B_{\mathfrak{p}}/\mathfrak{p}B_{\mathfrak{p}} \simeq S^{-1}B/(\mathfrak{p}S^{-1}B) \simeq B/\mathfrak{p}B$. It follows that if the lemma is true when A is a DVR then it is true in general, so we may assume that A is a DVR, and in particular, a PID.

By Proposition 5.22, B is finitely generated as an A module, and as an integral domain containing A, it must be torsion free. It follows from the structure theorem for finitely generated modules over a PID that B is free of finite rank over A. By Proposition 5.17, B spans L as a K-vector space, so any A-basis for B is a K-basis for L. It follows that B has rank $n := [L : K]$ as a free A-module, that is, $B \simeq A^n$. We then have $\mathfrak{p}B \simeq \mathfrak{p}A^n = (\mathfrak{p}A)^n$, so $B/\mathfrak{p}B \simeq A^n/(\mathfrak{p}A)^n \simeq (A/\mathfrak{p})^n$ is a free $A/\mathfrak{p}$-module of dimension n. □

Example 5.34. Let $A = \mathbb{Z}$, $B = \mathbb{Z}[i]$, and consider $\mathfrak{p} = (2)$. We have $\mathfrak{p}B = 2\mathbb{Z}[i] = (1+i)^2$, and $B/\mathfrak{p}B = \mathbb{Z}[i]/2\mathbb{Z}[i] = \mathbb{Z}[i]/(1+i)^2$ is an $\mathbb{F}_2$-algebra of dimension $2 = [\mathbb{Q}(i) : \mathbb{Q}]$. It contains a nonzero nilpotent (the image of $i+1$), so it is not a finite étale $\mathbb{F}_2$-algebra. It is a ring of cardinality 4 and characteristic 2 isomorphic to $\mathbb{F}_2[x]/(x^2)$.

Theorem 5.35. *Assume AKLB. For each prime $\mathfrak{p}$ of A we have*

$$\sum_{\mathfrak{q}|\mathfrak{p}} e_{\mathfrak{q}} f_{\mathfrak{q}} = [L : K].$$

Proof. We have

$$B/\mathfrak{p}B \simeq \prod_{\mathfrak{q}|\mathfrak{p}} B/\mathfrak{q}^{e_\mathfrak{q}}$$

Applying the previous proposition gives

$$\begin{aligned}[L:K] &= [B/\mathfrak{p}B : A/\mathfrak{p}] \\ &= \sum_{\mathfrak{q}|\mathfrak{p}} [B/\mathfrak{q}^{e_\mathfrak{q}} : A/\mathfrak{p}] \\ &= \sum_{\mathfrak{q}|\mathfrak{p}} e_\mathfrak{q}[B/\mathfrak{q} : A/\mathfrak{p}] \\ &= \sum_{\mathfrak{q}|\mathfrak{p}} e_\mathfrak{q} f_\mathfrak{q}.\end{aligned}$$

The second equality comes from the Chinese Remainder Theorem, and the third uses the fact that $B/\mathfrak{q}^{e_\mathfrak{q}}$ has dimension $e_\mathfrak{q}$ as a $B/\mathfrak{q}$-vector space. Indeed, we have

$$\mathfrak{q}^{e_\mathfrak{q}} = \{x \in B : v_\mathfrak{q}(x) \geq e_\mathfrak{q}\},$$

and if $\pi \in \mathfrak{q}$ is a uniformizer for $B_\mathfrak{q}$ (a generator $\mathfrak{q}B_\mathfrak{q}$ that we can force to lie in $\mathfrak{q}$ by clearing denominators), the images of $(\pi^0, \pi^1, \ldots, \pi^{e_\mathfrak{q}-1})$ in $B/\mathfrak{q}^{e_\mathfrak{q}}$ are a $B/\mathfrak{q}$-basis for $B/\mathfrak{q}^{e_\mathfrak{q}}$. □

For each prime $\mathfrak{p}$ of A, let $g_\mathfrak{p} := \#\{\mathfrak{q}|\mathfrak{p}\}$ denote the cardinality of the fiber above $\mathfrak{p}$.

Corollary 5.36. *Assume $AKLB$ and let $\mathfrak{p}$ be a prime of A. Then $g_\mathfrak{p}$ is an integer in the interval $[1, n]$, where $n = [L : K]$, as are $e_\mathfrak{q}$ and $f_\mathfrak{q}$ for each $\mathfrak{q}|\mathfrak{p}$.*

We now define some standard terminology that we may use in the $AKLB$ setting to describe how a prime $\mathfrak{p}$ of K splits in L (that is, for a nonzero prime ideal $\mathfrak{p}$ of A, how the ideal $\mathfrak{p}B$ factors into nonzero prime ideals $\mathfrak{q}$ of B).

Definition 5.37. Assume $AKLB$, let $\mathfrak{p}$ be a prime of A.

- L/K is *totally ramified at* $\mathfrak{q}$ if $e_\mathfrak{q} = [L : K]$ (equivalently, $f_\mathfrak{q} = 1 = g_\mathfrak{p} = 1$).
- L/K is *unramified at* $\mathfrak{q}$ if $e_\mathfrak{q} = 1$ **and** $B/\mathfrak{q}$ is a separable extension of $A/\mathfrak{p}$.
- L/K is *unramified above* $\mathfrak{p}$ if it is unramified at all $\mathfrak{q}|\mathfrak{p}$, equivalently, if $B/\mathfrak{p}B$ is a finite étale algebra over $A/\mathfrak{p}$.

When L/K is unramified above $\mathfrak{p}$ we say that

- $\mathfrak{p}$ *remains inert in* L if $\mathfrak{q} = \mathfrak{p}B$ is prime (equivalently, $e_\mathfrak{q} = g_\mathfrak{p} = 1$, and $f_\mathfrak{q} = [L : K]$).
- $\mathfrak{p}$ *splits completely in* L if $g_\mathfrak{p} = [L : K]$ (equivalently, $e_\mathfrak{q} = f_\mathfrak{q} = 1$ for all $\mathfrak{q}|\mathfrak{p}$).

In Example 5.34 above for the extension $\mathbb{Q}(i)/\mathbb{Q}$, the prime $\mathfrak{p} = (2)$ is ramified and the quotient ring $B/\mathfrak{p}B$ is not an étale $A/\mathfrak{p}$ algebra, even though the residue field $A/\mathfrak{p} \simeq \mathbb{F}_2$ is a perfect field (note that $B/\mathfrak{p}B$ is not a field). But when $A/\mathfrak{p}$ is a finite field (or any perfect field), for any prime $\mathfrak{q}|\mathfrak{p}$ the residue field $B/\mathfrak{q}$ is necessarily a finite étale $(A/\mathfrak{p})$-algebra, since it must be a separable field extension, and in this case $\mathfrak{q}$ is unramified whenever $e_\mathfrak{q} = 1$. This applies to our primary case of interest, where L/K is an extension of global fields. However, we will occasionally want to consider Dedekind domains A whose residue fields need not be perfect, in which case $e_\mathfrak{q} = 1$ does not imply that $\mathfrak{q}$ is unramified.

References

[1] Allen Altman and Steven Kleiman, *A term of commutative algebra*, Worldwide Center of Mathematics, 2013.

[2] Z.I. Borevich and I.R. Shafarevich, *Number theory*, Academic Press, 1966.

[3] N. Bourbaki, *Commutative Algebra: Chapters 1–7*, Springer, 1989.

[4] H. Matsumura, *Commutative ring theory*, Cambridge University Press, 1986.

6 Ideal norms and the Dedekind-Kummer theorem

In order to better understand how ideals split in Dedekind extensions we want to extend our definition of the norm map to ideals. Recall that for a ring extension B/A in which B is a free A-module of finite rank, we defined the norm map $\mathrm{N}_{B/A}\colon B \to A$ as

$$\mathrm{N}_{B/A}(b) := \det(B \xrightarrow{\times b} B),$$

the determinant of the multiplication-by-b map with respect to an A-basis for B. If B is a free A-module we could define the norm of a B-ideal to be the A-ideal generated by the norms of its elements, but in the case we are most interested in (our "$AKLB$" setup) B is typically *not* a free A-module (even though it is finitely generated as an A-module).

To get around this limitation, we introduce the notion of the *module index*, which we will use to define the norm of an ideal. In the special case where B is a free A-module, the norm of a B-ideal will be equal to the A-ideal generated by the norms of elements.

6.1 The module index

Our strategy is to define the norm of a B-ideal as the intersection of the norms of its localizations at maximal ideals of A (note that B is an A-module, so we can view any ideal of B as an A-module). Recall that by Proposition 2.6 any A-module M in a K-vector space is equal to the intersection of its localizations at primes of A; this applies, in particular, to ideals (and fractional ideals) of A and B. In order to do this we first define the *module index* of two A-lattices, as originally introduced by Fröhlich [3].

Recall that an A-lattice M in a K-vector space V is a finitely generated A-submodule of V that spans V as a K-vector space (Definition 5.9). If M is a free A-module, then any A-basis for M is also a K-basis for V, and we must have $M \simeq A^n$, where $n = \dim_K V$. If A is a Dedekind domain, even when M is not free, its localization $M_\mathfrak{p}$ at any prime $\mathfrak{p}$ of A will be a free $A_\mathfrak{p}$-module. This follows from the following facts: (a) $A_\mathfrak{p}$ is a DVR and therefore a PID, (b) $M_\mathfrak{p}$ is a torsion-free $A_\mathfrak{p}$-module, since it lies in a K-vector space and $A_\mathfrak{p} \subseteq K$, and (c) any finitely generated torsion-free module over a PID is free.

Definition 6.1. Let A be a Dedekind domain with fraction field K, let V be an n-dimensional K-vector space, let M and N be A-lattices in V, and let $\mathfrak{p}$ be a prime of A. Then $A_\mathfrak{p}$ is a PID and we must have $M_\mathfrak{p} \simeq A_\mathfrak{p}^n \simeq N_\mathfrak{p}$, as explained above. Choose an $A_\mathfrak{p}$-module isomorphism $\phi_\mathfrak{p}\colon M_\mathfrak{p} \xrightarrow{\sim} N_\mathfrak{p}$ and let $\hat{\phi}_\mathfrak{p}$ denote the unique K-linear map $V \to V$ extending $\phi_\mathfrak{p}$. The linear map $\hat{\phi}_\mathfrak{p}$ is an isomorphism and therefore has nonzero determinant. The *module index* $[M_\mathfrak{p} : N_\mathfrak{p}]_{A_\mathfrak{p}}$ is the principal fractional $A_\mathfrak{p}$-ideal generated by $\det \hat{\phi}_\mathfrak{p}$:

$$[M_\mathfrak{p} : N_\mathfrak{p}]_{A_\mathfrak{p}} := \big(\det \hat{\phi}_\mathfrak{p}\big).$$

This ideal does not depend on our choice of $\phi_\mathfrak{p}$ because any other choice can be written as $\phi_1\phi_\mathfrak{p}\phi_2$ for some $A_\mathfrak{p}$-module automorphisms $\phi_1\colon M_\mathfrak{p} \xrightarrow{\sim} M_\mathfrak{p}$ and $\phi_2\colon N_\mathfrak{p} \xrightarrow{\sim} N_\mathfrak{p}$ that necessarily have unit determinants. The *module index* $[M : N]_A$ is the A-module

$$[M : N]_A := \bigcap_\mathfrak{p} [M_\mathfrak{p} : N_\mathfrak{p}]_{A_\mathfrak{p}},$$

where $\mathfrak{p}$ ranges over primes of A and the intersection takes place in K. Each $[M_\mathfrak{p}\colon N_\mathfrak{p}]_{A_\mathfrak{p}}$ is an A-submodule of K (which need not be finitely generated), so their intersection is clearly an A-submodule of K, but it is not immediately clear that it finitely generated (or nonzero).

We claim that in fact $[M : N]_A$ is a nonzero fractional ideal of A whose localizations agree with all the local module indexes, that is for every prime $\mathfrak{p}$ of A we have

$$\big([M : N]_A\big)_{\mathfrak{p}} = [M_{\mathfrak{p}} : N_{\mathfrak{p}}]_{A_{\mathfrak{p}}}.$$

This is obvious when M and N are free A-modules: fix a global A-module isomorphism $\phi\colon M \xrightarrow{\sim} N$ so that $(\det \hat{\phi})_{\mathfrak{p}} = (\det \hat{\phi}_{\mathfrak{p}})$ for all primes $\mathfrak{p}$ (where $\phi_{\mathfrak{p}}$ is just the $A_{\mathfrak{p}}$-module isomorphism induced by ϕ). To prove the general case we apply a standard "gluing" argument that will be familiar to those who have studied algebraic geometry.

Proposition 6.2. *Let A be a Dedekind domain with fraction field K and let M and N be A-lattices in a K-vector space of finite dimension. The module index $[M : N]_A$ is a nonzero fractional ideal of A whose localization at each prime $\mathfrak{p}$ of A is equal to the local module index $[M_{\mathfrak{p}} : N_{\mathfrak{p}}]_{A_{\mathfrak{p}}}$.*

Proof. The finitely generated A-module M is locally free in the sense that the module $M_{\mathfrak{p}}$ is a free $A_{\mathfrak{p}}$-module for every prime $\mathfrak{p}$. It follows from [2, Thm. 19.2] that there exist nonzero $a_1, \dots, a_r \in A$ generating the unit ideal such that each $M[1/a_i]$ is a free $A[1/a_i]$-module (here $M[1/a_i]$ denotes the localization of M with respect to the multiplicative set $\{a_i^n : n \in \mathbb{Z}_{\geq 0}\}$). We similarly have nonzero $b_1, \dots, b_s \in A$ generating the unit ideal such that each $N[1/b_j]$ is a free $A[1/b_j]$-module. For any pair a_i and b_j, if we localize at the multiplicative set $S_{ij} := \{a_i^m b_j^n : m, n \in \mathbb{Z}_{\geq 0}\}$ then $S_{ij}^{-1}M$ and $S_{ij}^{-1}N$ will both be free $S_{ij}^{-1}A$-modules and we will have

$$\big([S_{ij}^{-1}M : S_{ij}^{-1}N]_{S_{ij}^{-1}A}\big)_{\mathfrak{p}} = [M_{\mathfrak{p}} : N_{\mathfrak{p}}]_{A_{\mathfrak{p}}},$$

for all primes $\mathfrak{p}$ of A that do not contain either a_i or b_j, since we can fix a global $S_{ij}^{-1}A$-module isomorphism $\phi\colon S_{ij}^{-1}M \to S_{ij}^{-1}N$ that induces $A_{\mathfrak{p}}$-module isomorphisms $\phi_{\mathfrak{p}}\colon M_{\mathfrak{p}} \to N_{\mathfrak{p}}$ with $(\det \hat{\phi})_{\mathfrak{p}} = (\det \hat{\phi}_{\mathfrak{p}})$; note that if $\mathfrak{p}$ contains either a_i or b_j then $\mathfrak{p}S_{ij}^{-1}A$ is the unit ideal (not a prime ideal of $S_{ij}^{-1}A$), thus $[S_{ij}^{-1}M : S_{ij}^{-1}N]_{S_{ij}^{-1}A}$ is equal to the intersection $\cap_{\mathfrak{p}}[M_{\mathfrak{p}} : N_{\mathfrak{p}}]_{A_{\mathfrak{p}}}$ over primes $\mathfrak{p}$ that do not contain a_i or b_j.

We now observe that since the sets $\{a_i\}$ and $\{b_j\}$ both generate the unit ideal, for every prime $\mathfrak{p}$ there is a choice of a_i and b_j that do not lie in $\mathfrak{p}$. It follows that

$$[M : N]_A = \bigcap_{\mathfrak{p}} [M_{\mathfrak{p}} : N_{\mathfrak{p}}]_{A_p} = \bigcap_{ij} [S_{ij}^{-1}M : S_{ij}^{-1}N]_{S_{ij}^{-1}A}.$$

Moreover, $[M : N]_A$ is a nonzero fractional ideal. To see this, let $I_{ij} := [S_{ij}^{-1}M : S_{ij}^{-1}N]_{S_{ij}^{-1}A}$. Each I_{ij} is a nonzero principal fractional $S_{ij}^{-1}A$-ideal, and we can choose a single $\alpha \in K^\times$ so that each αI_{ij} is an $S_{ij}^{-1}A$-ideal. The intersection of the αI_{ij} lies in $\cap_{ij} S_{ij}^{-1}A = A$ and is thus an A-submodule of A, hence an ideal, and finitely generated because A is noetherian. It follows that $[M : N]_A$ is a fractional ideal of A, and it is nonzero, since it contains the product of the the generators of the I_{ij}, for example. The localization of the intersection of a finite set of A-modules is equal to the intersection of their localizations, thus

$$([M : N]_A)_{\mathfrak{p}} = \big(\cap_{ij} [S_{ij}^{-1}M : S_{ij}^{-1}N]_{S_{ij}^{-1}A}\big)_{\mathfrak{p}} = \cap_{ij} \big([S_{ij}^{-1}M : S_{ij}^{-1}N]_{S_{ij}^{-1}A}\big)_{\mathfrak{p}} = [M_{\mathfrak{p}} : N_{\mathfrak{p}}]_{A_{\mathfrak{p}}}$$

as claimed. □

Proposition 6.2 implies that the module index $[M : N]_A$ is an element of the ideal group $\mathcal{I}_A$. If M, N, P are A-lattices in V then

$$[M : N]_A[N : P]_A = [M : P]_A, \tag{1}$$

since for each prime $\mathfrak{p}$ we can write any isomorphism $M_\mathfrak{p} \xrightarrow{\sim} P_\mathfrak{p}$ as a composition of isomorphisms $M_\mathfrak{p} \xrightarrow{\sim} N_\mathfrak{p} \xrightarrow{\sim} P_\mathfrak{p}$; we then note that the determinant map is multiplicative with respect to composition and multiplication of fractional ideals is compatible with localization. Taking $P = M$ yields the identity

$$[M : N]_A[N : M]_A = [M : M]_A = A, \tag{2}$$

thus $[M : N]_A$ and $[N : M]_A$ are inverses in the ideal group $\mathcal{I}_A$. We note that when $N \subseteq M$ the module index $[M : N]_A \subseteq A$ is actually an ideal (not just a fractional ideal), since in this case we can express a basis for $N_\mathfrak{p}$ as $A_\mathfrak{p}$-linear combinations of a basis for $M_\mathfrak{p}$, and the matrix for $\hat{\phi}_\mathfrak{p}$ will then have entries (and determinant) in $A_\mathfrak{p}$.

Remark 6.3. In the special case $V = K$, an A-lattice in V is simply a fractional ideal of A. In this setting each module index $[M : N]_A$ corresponds to a colon ideal

$$[M : N]_A = (N : M). \tag{3}$$

Note that the order of M and N is **reversed**. This unfortunate conflict of notation arises from the fact that the module index is generalizing the notion of an index (for example, $[\mathbb{Z} : 2\mathbb{Z}]_\mathbb{Z} = ([\mathbb{Z} : 2\mathbb{Z}]) = (2)$), whereas colon ideals are generalizing the notion of a ratio (for example, $(\mathbb{Z} : 2\mathbb{Z}) = (1 : 2) = (1/2)$). To see why (3) holds, let π be a uniformizer for $A_\mathfrak{p}$. Then $M_\mathfrak{p} = (\pi^m)$ and $N_\mathfrak{p} = (\pi^n)$ for some $m, n \in \mathbb{Z}$, and we may take $\phi_\mathfrak{p}$ to be the multiplication-by-π^{n-m} map. We then have

$$[M_\mathfrak{p} : N_\mathfrak{p}]_{A_\mathfrak{p}} = (\det \hat{\phi}_\mathfrak{p}) = (\pi^{n-m}) = (\pi^n/\pi^m) = (N_\mathfrak{p} : M_\mathfrak{p}).$$

It follows from the remark that if M and N are nonzero fractional ideals of A then

$$M[M : N]_A = M(N : M) = N.$$

(note we are using the fact that A is a Dedekind domain; we always have $M(N : M) \subseteq N$ but equality does not hold in general), and if $N \subseteq M$ then $I := [M : N]_A \subseteq A$ is an ideal and we have $MI = N = NA$ and therefore $M/N \simeq A/I$ as quotients of A-modules. It follows that $I = \{a \in A : aM \subseteq N\}$ is the *annihilator* of M/N, which is a *cyclic* A-module (has a single generator), since A/I is clearly cyclic (generated by the image of 1). Conversely, if we know that $M/N \simeq A/I$ for nonzero fractional ideals $N \subseteq M$, then we necessarily have $I = [M : N]_A$. The following theorem generalizes this observation.

Theorem 6.4. *Let A be a Dedekind domain with fraction field K, and let $N \subseteq M$ be A-lattices in a K-vector space V of dimension r for which the quotient module M/N is a direct sum of cyclic A-modules:*

$$M/N \simeq A/I_1 \oplus \cdots \oplus A/I_n,$$

where $I_1, \ldots, I_n$ are ideals of A. Then

$$[M : N]_A = I_1 \cdots I_n.$$

Proof. Let $\mathfrak{p}$ be a prime of A, let π be a uniformizer for $A_\mathfrak{p}$, and let $e_j = v_\mathfrak{p}(I_j)$ for $1 \le j \le n$. Pick a basis for $M_\mathfrak{p}$ and an isomorphism $\phi_\mathfrak{p}\colon M_\mathfrak{p} \to N_\mathfrak{p}$ so that $M_\mathfrak{p}/N_\mathfrak{p} = \operatorname{coker}\phi_\mathfrak{p}$. The matrix of $\phi_\mathfrak{p}$ is an $r \times r$ matrix over the PID $A_\mathfrak{p}$ with nonzero determinant. It therefore has Smith normal form UDV, with $U, V \in \mathrm{GL}_r(A_\mathfrak{p})$ and $D = \operatorname{diag}(\pi^{d_1}, \dots, \pi^{d_r})$ for some uniquely determined nonnegative integers $d_1 \le \cdots \le d_r$. We then have

$$A_\mathfrak{p}/(\pi^{e_1}) \oplus \cdots \oplus A_\mathfrak{p}/(\pi^{e_n}) \simeq M_\mathfrak{p}/N_\mathfrak{p} = \operatorname{coker}\phi \simeq A_\mathfrak{p}/(\pi^{d_1}) \oplus \cdots \oplus A_\mathfrak{p}/(\pi^{d_r}).$$

It follows from the structure theorem for modules over a PID that the non-trivial summands on each side are precisely the invariant factors of $M_\mathfrak{p}/N_\mathfrak{p}$, possibly in different orders. We therefore have $\sum_{j=1}^n e_j = \sum_{i=1}^r d_i$, and applying the definition of the module index yields

$$[M_\mathfrak{p} : N_\mathfrak{p}]_{A_\mathfrak{p}} = (\det\phi_\mathfrak{p}) = (\det D) = (\pi^{\sum d_i}) = (\pi^{\sum e_j}) = (\pi_\mathfrak{p}^{e_1}) \cdots (\pi_\mathfrak{p}^{e_n}) = (I_1 \cdots I_n)_\mathfrak{p}.$$

It follows that $[M : N]_A = I_1 \cdots I_n$, since the localizations $([M : N]_A)_\mathfrak{p} = [M_\mathfrak{p} : N_\mathfrak{p}]_{A_\mathfrak{p}}$ and $(I_1 \cdots I_n)_\mathfrak{p}$ coincide for every prime $\mathfrak{p}$. □

6.2 The ideal norm

In the *AKLB* setup the inclusion $A \subseteq B$ induces a homomorphism of ideal groups:

$$\begin{aligned} \mathcal{I}_A &\to \mathcal{I}_B \\ I &\mapsto IB. \end{aligned}$$

We wish define a homomorphism $N_{B/A}\colon \mathcal{I}_B \to \mathcal{I}_A$ in the reverse direction. As we proved in the previous lecture, every fractional B-ideal I is an A-lattice in L, so let us consider

$$\begin{aligned} \mathcal{I}_B &\to \mathcal{I}_A \\ I &\mapsto [B : I]_A. \end{aligned}$$

Definition 6.5. Assume *AKLB*. The *ideal norm* $N_{B/A}\colon \mathcal{I}_B \to \mathcal{I}_A$ is the map $I \mapsto [B : I]_A$. We extend $N_{B/A}$ to the zero ideal by defining $N_{B/A}((0)) = (0)$.

We now show that the ideal norm $N_{B/A}$ is compatible with the field norm $\mathrm{N}_{L/K}$.

Proposition 6.6. *Assume AKLB and let $\alpha \in L$. Then $N_{B/A}((\alpha)) = \big(\mathrm{N}_{L/K}(\alpha)\big)$.*

Proof. The case $\alpha = 0$ is immediate, so assume $\alpha \in L^\times$. We have

$$N_{B/A}((\alpha)) = [B : \alpha B]_A = \bigcap_\mathfrak{p} [B_\mathfrak{p} : \alpha B_\mathfrak{p}]_{A_\mathfrak{p}} = \Big(\det(L \xrightarrow{\times\alpha} L)\Big) = \big(\mathrm{N}_{L/K}(\alpha)\big),$$

since each $B_\mathfrak{p} \xrightarrow{\times\alpha} \alpha B_\mathfrak{p}$ is an isomorphism of free $A_\mathfrak{p}$-modules that are $A_\mathfrak{p}$-lattices in L. □

Proposition 6.7. *Assume AKLB. The map $N_{B/A}\colon \mathcal{I}_B \to \mathcal{I}_A$ is a group homomorphism.*

Proof. Let $\mathfrak{p}$ be a maximal ideal of A. Then $A_\mathfrak{p}$ is a DVR and $B_\mathfrak{p}$ is a semilocal Dedekind domain, hence a PID. Thus every element of $\mathcal{I}_{B_\mathfrak{p}}$ is a principal ideal (α) for some $\alpha \in L^\times$, and the previous proposition implies that $N_{B_\mathfrak{p}/A_\mathfrak{p}}\colon \mathcal{I}_{B_\mathfrak{p}} \to \mathcal{I}_{A_\mathfrak{p}}$ is a group homomorphism, since $N_{L/K}$ is. For any $I, J \in \mathcal{I}_B$ we then have

$$N_{B/A}(IJ) = \bigcap_\mathfrak{p} N_{B_\mathfrak{p}/A_\mathfrak{p}}(I_\mathfrak{p} J_\mathfrak{p}) = \bigcap_\mathfrak{p} N_{B_\mathfrak{p}/A_\mathfrak{p}}(I_\mathfrak{p}) N_{B_\mathfrak{p}/A_\mathfrak{p}}(J_\mathfrak{p}) = N_{B/A}(I) N_{B/A}(J). \qquad \square$$

Corollary 6.8. *Assume AKLB. For all $I, J \in \mathcal{I}_B$ we have*

$$[I : J]_A = N_{B/A}(I^{-1}J) = N_{B/A}((J : I))$$

Proof. The second equality is immediate: $(J : I) = I^{-1}J$ (because B is a Dedekind domain). The first follows from (1), (2), and the previous proposition. Indeed, we have

$$[I : J]_A = [I : B]_A[B : J]_A = [B : I]_A^{-1}[B : J]_A = N_{B/A}(I^{-1})N_{B/A}(J) = N_{B/A}(I^{-1}J). \quad \square$$

Corollary 6.9. *Assume AKLB and let I be a fractional ideal of B. The ideal norm of I is the fractional ideal of A generated by the image of I under the field norm $N_{L/K}$, that is,*

$$N_{B/A}(I) = \big(\mathrm{N}_{L/K}(\alpha) : \alpha \in I\big).$$

Proof. Let J denote the RHS. For any nonzero prime $\mathfrak{p}$ of A, the localization of the ideal $N_{B/A}(I) = [B : I]_A$ at $\mathfrak{p}$ is $[B_\mathfrak{p} : I_\mathfrak{p}]_{A_\mathfrak{p}} = N_{B_\mathfrak{p}/A_\mathfrak{p}}(I_\mathfrak{p})$. The fractional ideal $N_{B_\mathfrak{p}/A_\mathfrak{p}}(I_\mathfrak{p})$ of $A_\mathfrak{p}$ is principal, so $N_{B_\mathfrak{p}/A_\mathfrak{p}}(I_\mathfrak{p}) = J_\mathfrak{p}$ follows from the proposition, and

$$N_{B/A}(I) = \bigcap_\mathfrak{p} N_{B_\mathfrak{p}/A_\mathfrak{p}}(I_\mathfrak{p}) = \bigcap_\mathfrak{p} J_\mathfrak{p} = J. \quad \square$$

The corollary gives us an alternative definition of the ideal norm in terms of the field norm. In view of this we extend our definition of the field norm $N_{L/K}$ to fractional ideals of B, and we may write $N_{L/K}(I)$ instead of $N_{B/A}(I)$. We have the following pair of commutative diagrams, in which the downward arrows map nonzero field elements to the principal fractional ideals they generate. We know that composing the maps $K^\times \to L^\times \to K^\times$ along the top corresponds to exponentiation by $n = [L : K]$ (see Problem Set 2); we now show that this is also true for the composition of the bottom maps.

$$\begin{array}{ccc} K^\times & \hookrightarrow & L^\times \\ \downarrow{\scriptstyle (x)} & & \downarrow{\scriptstyle (y)} \\ \mathcal{I}_A & \xrightarrow{I \mapsto IB} & \mathcal{I}_B \end{array} \qquad \begin{array}{ccc} L^\times & \xrightarrow{N_{L/K}} & K^\times \\ \downarrow{\scriptstyle (y)} & & \downarrow{\scriptstyle (x)} \\ \mathcal{I}_B & \xrightarrow{N_{B/A}} & \mathcal{I}_A \end{array}$$

Theorem 6.10. *Assume AKLB and let $\mathfrak{q}$ be a prime lying above $\mathfrak{p}$. Then $N_{B/A}(\mathfrak{q}) = \mathfrak{p}^{f_\mathfrak{q}}$, where $f_\mathfrak{q} = [B/\mathfrak{q} : A/\mathfrak{p}]$ is the residue field degree of $\mathfrak{q}$.*

Proof. The $(A/\mathfrak{p})$-vector space $B/\mathfrak{q}$ has dimension $f_\mathfrak{q}$ (by definition); as a quotient of A-modules, we have $B/\mathfrak{q} \simeq A/\mathfrak{p} \oplus \cdots \oplus A/\mathfrak{p}$, an $f_\mathfrak{q}$-fold direct sum of cyclic A-modules $A/\mathfrak{p}$, and we may apply Theorem 6.4. Thus $N_{B/A}(\mathfrak{q}) = [B : \mathfrak{q}]_A = \mathfrak{p} \cdots \mathfrak{p} = \mathfrak{p}^{f_\mathfrak{q}}$. $\square$

Corollary 6.11. *Assume AKLB. For $I \in \mathcal{I}_A$ we have $N_{B/A}(IB) = I^n$, where $n = [L : K]$.*

Proof. Since $N_{B/A}$ and $I \mapsto IB$ are group homomorphisms, it suffices to consider the case were $I = \mathfrak{p}$ is a nonzero prime ideal. We then have

$$N_{B/A}(\mathfrak{p}B) = N_{B/A}\left(\prod_{\mathfrak{q}|\mathfrak{p}} \mathfrak{q}^{e_\mathfrak{q}}\right) = \prod_{\mathfrak{q}|\mathfrak{p}} N_{B/A}(\mathfrak{q})^{e_\mathfrak{q}} = \prod_{\mathfrak{q}|\mathfrak{p}} \mathfrak{p}^{e_\mathfrak{q} f_\mathfrak{q}} = \mathfrak{p}^{\sum_{\mathfrak{q}|\mathfrak{p}} e_\mathfrak{q} f_\mathfrak{q}} = \mathfrak{p}^n. \quad \square$$

6.3 The ideal norm in algebraic geometry

The maps $i\colon \mathcal{I}_A \to \mathcal{I}_B$ and $N_{B/A}\colon \mathcal{I}_B \to \mathcal{I}_A$ have a geometric interpretation that will be familiar to those who have studied algebraic geometry: they are the pushforward and pullback maps on divisors associated to the morphism of curves $Y \to X$ induced by the inclusion $A \subseteq B$, where $X = \operatorname{Spec} A$ and $Y = \operatorname{Spec} B$. For the benefit of those who have not seen this before, let us briefly explain the connection (while glossing over some details).

Dedekind domains naturally arise in algebraic geometry as coordinate rings of smooth curves (which for the sake of this discussion one can take to mean geometrically irreducible algebraic varieties of dimension one with no singularities). In order to make this explicit, let us fix a perfect field k and a polynomial $f \in k[x,y]$ that we will assume is irreducible in $\bar{k}[x,y]$. The ring $A = k[x,y]/(f)$ is a noetherian domain of dimension 1, and if we further assume that the algebraic variety X defined by $f(x,y) = 0$ has no singularities, then A is also integrally closed and therefore a Dedekind domain.[1] We call A the *coordinate ring* of X, denoted $k[X]$, and its fraction field is the *function field* of X, denoted $k(X)$.

Conversely, given a Dedekind domain A, we can regard $X = \operatorname{Spec} A$ as a smooth curve whose *closed points* are the maximal ideals of A (all of $\operatorname{Spec} A$ except the zero ideal, which is called the *generic point*). When the field of constants k is algebraically closed, Hilbert's Nullstellensatz gives a one-to-one correspondence between maximal ideals $(x-x_0, y-y_0)$ and points (x_0, y_0) in the affine plane, but in general closed points correspond to $\operatorname{Gal}(\bar{k}/k)$-orbits of $\bar{k}$-points.

Recall that the ideal group $\mathcal{I}_A$ is isomorphic to the free abelian group generated by the nonzero prime ideals $\mathfrak{p}$ of A. The corresponding object in algebraic geometry is the *divisor group* $\operatorname{Div} X$, the free abelian group generated by the closed points P of X. The group $\operatorname{Div} X$ is written additively, so its elements have the form $D = \sum n_P P$ with all but finitely many of the integers n_P equal to 0.

A finite extension of Dedekind domains B/A induces a surjective morphism $\phi\colon Y \to X$ of the corresponding curves $X = \operatorname{Spec} A$ and $Y = \operatorname{Spec} B$. Primes $\mathfrak{q}$ of B in the fiber above a prime $\mathfrak{p}$ of A correspond to closed points Q of Y in the fiber of ϕ above a closed point P of X. The map $\mathcal{I}_A \to \mathcal{I}_B$ defined by $\mathfrak{p} \mapsto \mathfrak{p}B = \prod_{\mathfrak{q}|\mathfrak{p}} \mathfrak{q}^{e_\mathfrak{q}}$ corresponds to the *pullback* map $\phi^*\colon \operatorname{Div} X \to \operatorname{Div} Y$ induced by ϕ, which is defined by

$$\phi^*(P) := \sum_{\phi(Q)=P} e_Q Q$$

where e_Q is the ramification index (one then extends $\mathbb{Z}$-linearly: $\phi^*(\sum n_P P) = \sum n_P \phi^*(P)$). Geometrically we think of e_Q as the "multiplicity" of Q in the fiber above P, although e_Q is typically defined algebraically as the ramification index of the prime Q in the Dedekind extension B/A as we have done (alternatively, as we shall see in later lectures, it can be defined in terms of valuations on $k(X)$ and $k(Y)$ associated to P and Q).

In the other direction, the norm map $N_{B/A}\colon \mathcal{I}_B \to \mathcal{I}_A$, which sends $\mathfrak{q}$ to $N_{B/A}(\mathfrak{q}) = \mathfrak{p}^{f_\mathfrak{q}}$, corresponds to *pushforward* map $\phi_*\colon \operatorname{Div} Y \to \operatorname{Div} X$ induced by ϕ, which is defined by

$$\phi_*(Q) := f_Q \phi(Q) = f_Q P,$$

[1] If A is not integrally closed, we can replace it by its integral closure, thereby obtaining the *normalization* of the curve X. One typically also takes the projective closure of X in order to obtain a *complete* curve; this corresponds to considering all absolute values (*places*) of the function field of X, not just those arising from primes. This distinction does not affect our discussion here but will become relevant in later lectures.

where f_Q counts the number of $\bar{k}$-points in the $\mathrm{Gal}(\bar{k}/k)$-orbit corresponding to the closed point Q, equivalently, the degree of the field extension of k needed to split Q into f_Q distinct closed points after base extension (here we are using our assumption that k is perfect). This is precisely the residue field degree of Q as a prime in the Dedekind extension B/A. Note that when $k = \bar{k}$ we always have $f_Q = 1$ (so over algebraically closed fields one typically omits f_Q from the pushforward map and the degree formula below).

If we compose the pushforward and pullback maps we obtain

$$\phi_*\phi^*(P) = \sum_{\phi(Q)=P} e_Q f_Q P = \deg(\phi)P.$$

Here $\deg(\phi)$ is the *degree* of the morphism $\phi\colon Y \to X$, which is typically defined as the degree of the function field extension $[k(Y) : k(X)]$, but one can take the above formula as an alternative definition (by Theorem 5.35). It is a weighted measure of the cardinality of the fibers of ϕ that reflects both the ramification and degree of each closed point in the fiber (and as a consequence, it is the same for every fiber and is an invariant of ϕ).

6.4 The ideal norm in number fields

We now consider the special case $A = \mathbb{Z}$, $K = \mathbb{Q}$, where $B = \mathcal{O}_L$ is the ring of integers of the number field L. In this situation we may simply write N in place of $N_{B/A}$ and call it the *absolute norm*. If $\mathfrak{q}$ is a nonzero prime ideal of $\mathcal{O}_L$ then Theorem 6.10 implies

$$\mathrm{N}(\mathfrak{q}) = (p^{f_\mathfrak{q}}),$$

where $p \in \mathbb{Z}$ is the unique prime in $\mathfrak{q} \cap \mathbb{Z}$, and f is the degree of the finite field $B/\mathfrak{q}$ as an extension of $\mathbb{F}_p \simeq \mathbb{Z}/p\mathbb{Z}$. The absolute norm

$$\mathrm{N}(\mathfrak{q}) = [\mathcal{O}_L : \mathfrak{q}]_\mathbb{Z} = ([\mathcal{O}_L : \mathfrak{q}])$$

is the principal ideal generated by the (necessarily finite) index $[\mathcal{O}_L : \mathfrak{q}] \in \mathbb{Z}$ of $\mathfrak{q}$ in $\mathcal{O}_L$ as free $\mathbb{Z}$-modules of equal rank; this is just the index of $\mathfrak{q}$ in $\mathcal{O}_L$ as additive groups. More generally, we have the following.

Proposition 6.12. *Let L be a number field with ring of integers $\mathcal{O}_L$. For any nonzero $\mathcal{O}_L$-ideal $\mathfrak{a}$ we have $\mathrm{N}(\mathfrak{a}) = \big([\mathcal{O}_L : \mathfrak{a}]\big)$. If $\mathfrak{b} \subseteq \mathfrak{a}$ are nonzero fractional ideals of $\mathcal{O}_L$, then*

$$[\mathfrak{a} : \mathfrak{b}]_\mathbb{Z} = ([\mathfrak{a} : \mathfrak{b}]).$$

Proof. The ring $\mathcal{O}_L$ is a free $\mathbb{Z}$ module of rank $n := [L : \mathbb{Q}]$. It is free because it is torsion-free and $\mathbb{Z}$ is a PID, and it has rank n because it contains a $\mathbb{Q}$-basis for L, by Proposition 5.17. The same is true of any nonzero fractional ideal of $\mathcal{O}_L$: it is a torsion-free $\mathbb{Z}$-module, hence free, and it has the same rank n as $\mathcal{O}_L$ because it contains some nonzero principal fractional ideal $\alpha\mathcal{O}_L$: the fact that $\mathcal{O}_L$ spans L implies that $\alpha\mathcal{O}_L$ spans L, because the multiplication-by-α map $L \overset{\times\alpha}{\to} L$ is an invertible $\mathbb{Q}$-linear transformation.

Let us now fix $\mathbb{Z}$-bases for $\mathcal{O}_L$ and the nonzero $\mathcal{O}_L$-ideal $\mathfrak{a}$. Let $\Phi \in \mathbb{Z}^{n\times n}$ be the matrix whose columns express each basis element for $\mathfrak{a}$ in terms of our basis for $\mathcal{O}_L$. Multiplication by Φ defines a $\mathbb{Z}$-module isomorphism from $\mathcal{O}_L$ to $\mathfrak{a}$, since it maps our basis for $\mathcal{O}_L$ to our basis for $\mathfrak{a}$. It follows that $[\mathcal{O}_L : \mathfrak{a}]_\mathbb{Z} = (\det \Phi)$: for every prime $p \in \mathbb{Z}$ we can use the matrix Φ to define a $\mathbb{Z}_{(p)}$-module isomorphism $\phi_{(p)}\colon (\mathcal{O}_L)_{(p)} \to \mathfrak{a}_{(p)}$ with $\det \hat{\phi}_{(p)} = \det \Phi$ (any $\mathbb{Z}$-basis for a free $\mathbb{Z}$-module M is also a $\mathbb{Z}_{(p)}$-basis for the free $\mathbb{Z}_{(p)}$-module $M_{(p)}$).

We now observe that the absolute value of the determinant of Φ is equal to the index of $\mathfrak{a}$ in $\mathcal{O}_L$: indeed, if we identify $\mathcal{O}_L$ with $\mathbb{Z}^n$ then $|\det \Phi|$ is the volume of a fundamental parallelepiped for $\mathfrak{a}$, viewed as a sublattice of $\mathbb{Z}^n$. We thus have

$$([\mathcal{O}_L : \mathfrak{a}]) = (\det \Phi) = [\mathcal{O}_L : \mathfrak{a}]_{\mathbb{Z}} = \mathrm{N}(\mathfrak{a}),$$

which proves the first claim.

For any $\alpha \in L^\times$ we have $[\mathfrak{a} : \mathfrak{b}] = [\alpha\mathfrak{a} : \alpha\mathfrak{b}]$ and $[\mathfrak{a} : \mathfrak{b}]_{\mathbb{Z}} = [\alpha\mathfrak{a} : \alpha\mathfrak{b}]_{\mathbb{Z}}$, so we can assume without loss of generality that $\mathfrak{a}$ and $\mathfrak{b}$ are ideals in $\mathcal{O}_L$. We then have a tower of free $\mathbb{Z}$-modules $\mathfrak{b} \subseteq \mathfrak{a} \subseteq \mathcal{O}_L$, and therefore

$$[\mathcal{O}_L : \mathfrak{a}][\mathfrak{a} : \mathfrak{b}] = [\mathcal{O}_L : \mathfrak{b}].$$

Replacing both sides with the $\mathbb{Z}$-ideals they generate, we have

$$\mathrm{N}(\mathfrak{a})\big([\mathfrak{a} : \mathfrak{b}]\big) = \mathrm{N}(\mathfrak{b}),$$

and therefore $([\mathfrak{a} : \mathfrak{b}]) = \mathrm{N}(\mathfrak{a}^{-1}\mathfrak{b}) = [\mathfrak{a} : \mathfrak{b}]_{\mathbb{Z}}$, by Corollary 6.8, proving the second claim. □

Remark 6.13. Since $\mathbb{Z}$ is a principal ideal domain whose only units are ± 1, we can unambiguously identify each fractional ideal with a positive rational number and view the absolute norm $\mathrm{N} \colon \mathcal{I}_{\mathcal{O}_L} \to \mathcal{I}_{\mathbb{Z}}$ as a homomorphism $\mathrm{N} \colon \mathcal{I}_{\mathcal{O}_L} \to \mathbb{Q}^\times_{>0}$ from ideal group of $\mathcal{O}_L$ to the multiplicative group of positive rational numbers. If we write $\mathrm{N}(\mathfrak{a})$ in contexts where an element of $\mathbb{Z}$ or $\mathbb{Q}$ (or $\mathbb{R}$) is expected, it is always with this understanding. When $\mathfrak{a} = (a)$ is a nonzero principal fractional ideal we may also write $\mathrm{N}(a) \coloneqq \mathrm{N}((a)) = |\mathrm{N}_{L/\mathbb{Q}}(a)|$; this is a positive rational number, and for $a \in \mathcal{O}_L$, a positive integer.

6.5 The Dedekind-Kummer theorem

We now give a theorem that provides a practical method for factoring primes in Dedekind extensions. This result was proved by Dedekind for number fields, building on earlier work of Kummer, but we will give a version that works for arbitrary extensions of Dedekind domains B/A whose fraction fields are a finite separable extensions L/K (the *AKLB* setup).

The primitive element theorem implies when L/K is a finite separable extension we can always write $L = K(\alpha)$ for some $\alpha \in L$, and in the *AKLB* setup we can assume $\alpha \in B$, by Proposition 5.17. This does **not** imply that $B = A[\alpha]$; indeed, it may very will happen that there is no $\alpha \in B$ for which $B = A[\alpha]$. Extensions L/K for which $B = A[\alpha]$ for some $\alpha \in B$ are said to be *monogenic*. This necessarily implies that B is a free A-module, hence it has an *integral basis* $\{\beta_1, \dots, \beta_n\}$ that is both an A-basis for B and a K-basis for L. But monogenicity is a much stronger condition: it implies that B has an *integral power basis*, one of the form $\{1, \alpha, \dots, \alpha^{n-1}\}$. When $A = \mathbb{Z}$ every B has an integral basis, but very few have an integral power basis. Examples of monogenic extensions include quadratic and cyclotomic number fields (as extensions of $\mathbb{Q}$); see Problem Set 3 for proofs of these facts and some examples of non-monogenic number fields.

We will first prove the Dedekind-Kummer theorem assuming we have a monogenic extension; in the next section we will address the general case.

Theorem 6.14 (Dedekind-Kummer)**.** *Assume AKLB with $L = K(\alpha)$ and $\alpha \in B$. Let $f \in A[x]$ be the minimal polynomial of α, let $\mathfrak{p}$ be a prime of A, and let*

$$\bar{f} = \bar{g}_1^{e_1} \cdots \bar{g}_r^{e_r}$$

be its factorization into monic irreducibles in $(A/\mathfrak{p})[x]$*. Let* $\mathfrak{q}_i := (\mathfrak{p}, g_i(\alpha))$*, where* $g_i \in A[x]$ *is any lift of* $\bar{g}_i$ *in* $(A/\mathfrak{p})[x]$ *under the reduction map* $A[x] \to (A/\mathfrak{p})[x]$*. If* $B = A[\alpha]$ *then*

$$\mathfrak{p}B = \mathfrak{q}_1^{e_1} \cdots \mathfrak{q}_r^{e_r},$$

is the prime factorization of $\mathfrak{p}B$ *in* B *and the residue field degree of* $\mathfrak{q}_i$ *is* $\deg \bar{g}_i$*.*

Before proving the theorem, last us give an example to illustrate its utility.

Example 6.15. Let $A = \mathbb{Z}$, $K = \mathbb{Q}$, and $L = \mathbb{Q}(\zeta_5)$, where $\alpha = \zeta_5$ is a primitive 5th root of unity with minimal polynomial $f(x) = x^4 + x^3 + x^2 + x + 1$. Then $B = \mathcal{O}_L = \mathbb{Z}[\zeta_5]$ and we can use the theorem to factor any prime of $\mathbb{Z}$ in $\mathcal{O}_L$:

- (2): $f(x)$ is irreducible modulo 2, so $2\mathbb{Z}[\zeta_5]$ is prime and (2) is inert in $\mathbb{Q}(\zeta_5)$.
- (5): $f(x) \equiv (x-1)^4 \bmod 5$, so $5\mathbb{Z}[\zeta_5] = (5, \zeta_5 - 1)^4$ and (5) is totally ramified in $\mathbb{Q}(\zeta_5)$.
- (11): $f(x) \equiv (x-4)(x-9)(x-5)(x-3) \bmod 11$, so
$$11\mathbb{Z}[\zeta_5] = (11, \zeta_5 - 4)(11, \zeta_5 - 9)(11, \zeta_5 - 5)(11, \zeta_5 - 3),$$
and (11) splits completely in $\mathbb{Q}(\zeta_5)$.
- (19): $f(x) \equiv (x^2 + 5x + 1)(x^2 - 4x + 1) \bmod 19$, so
$$19\mathbb{Z}[\zeta_5] = (19, \zeta_5^2 + 5\zeta_5 + 1)(19, \zeta_5^2 - 4\zeta_5 + 1).$$

The four cases above cover every possible prime factorization pattern in the cyclotomic extension $\mathbb{Q}(\zeta_5)/\mathbb{Q}$ (see Problem Set 3 for a proof).

Proof of the Dedekind-Kummer theorem. We have $B = A[\alpha] \simeq A[x]/(f(x))$ and therefore

$$\frac{B}{\mathfrak{q}_i} = \frac{A[\alpha]}{(\mathfrak{p}, g_i(\alpha))} \simeq \frac{A[x]}{(f(x), \mathfrak{p}, g_i(x))} \simeq \frac{(A/\mathfrak{p})[x]}{(\bar{f}(x), \bar{g}_i(x))} \simeq \frac{(A/\mathfrak{p})[x]}{(\bar{g}_i(x))}.$$

The polynomial $\bar{g}_i(x)$ is by assumption irreducible, thus $(\bar{g}_i(x))$ is a maximal ideal (because $(A/\mathfrak{p})[x]$ is a UFD of dimension 1), so the quotient $(A/\mathfrak{p})[x]/(\bar{g}_i(x))$ is a field; indeed, it is an extension of the residue field $A/\mathfrak{p}$ of degree $\deg g_i$. It follows that $\mathfrak{q}_i$ is a prime above $\mathfrak{p}$ with residue field degree $f_{\mathfrak{q}_i} = \deg \bar{g}_i$ as claimed.

The ideal $\prod_i \mathfrak{q}_i^{e_i} = \prod_i (\mathfrak{p}, g_i(\alpha))^{e_i} = \prod_i (\mathfrak{p}B + (g_i(\alpha)))^{e_i}$ is divisible by $\mathfrak{p}B$, since if we expand the ideal product every term is clearly divisible by $\mathfrak{p}B$, including

$$\prod_i (g_i(\alpha)^{e_i}) \equiv (f(\alpha)) \equiv (0) \bmod \mathfrak{p}B.$$

The $\bar{g}_i(x)$ are distinct as elements of $(A/\mathfrak{p})[x]/(f(x)) \simeq A[x]/(\mathfrak{p}, f(x)) \simeq A[\alpha]/\mathfrak{p}A[\alpha]$, and it follows that the $g_i(\alpha)$ are distinct modulo $\mathfrak{p}B$. Therefore the prime ideals $\mathfrak{q}_i$ are distinct, and we must then have $e_i \geq e_{\mathfrak{q}_i}$ and $\{\mathfrak{q}|\mathfrak{p}\} \subseteq \{\mathfrak{q}_i\}$ in order for $\prod_i \mathfrak{q}_i^{e_i}$ to be divisible by $\mathfrak{p}B$; we already showed that each $\mathfrak{q}_i$ is a prime above $\mathfrak{p}$, so we must have $\{\mathfrak{q}_i\} = \{\mathfrak{q}|\mathfrak{p}\}$. Now

$$N_{B/A}\left(\prod_i \mathfrak{q}_i^{e_i}\right) = \prod_i N_{B/A}(\mathfrak{q}_i)^{e_i} = \prod_i (\mathfrak{p}^{f_{\mathfrak{q}_i}})^{e_i} = \mathfrak{p}^{e_i \deg \bar{g}_i} = \mathfrak{p}^{\deg f} = \mathfrak{p}^{[L:K]},$$

so $\sum_i e_i f_{\mathfrak{q}_i} = [L:K] = \sum_{\mathfrak{q}|\mathfrak{p}} e_{\mathfrak{q}} f_{\mathfrak{q}}$. We must have $e_i = e_{\mathfrak{q}_i}$ and the theorem follows. □

We now want to remove the monogenic hypothesis from Theorem 6.14 We can always write $L = K(\alpha)$ for some $\alpha \in B$ (since L/K is separable), but in general the ring $A[\alpha]$ may be a proper subring of B. The relationship between $A[\alpha]$ and B is characterized by the *conductor* of the extension $B/A[\alpha]$.

6.6 The conductor of a ring

We first give the general definition then specialize to subrings of Dedekind domains.

Definition 6.16. Let S/R be an extension of commutative rings. The *conductor of R in S* is the largest S-ideal that is also an R-ideal; equivalently, it is the largest ideal of S contained in R. It can be written as

$$\mathfrak{c} := \{\alpha \in S : \alpha S \subseteq R\} = \{\alpha \in R \colon \alpha S \subseteq R\}.$$

If R is an integral domain, the *conductor of R* is the conductor of R in its integral closure.

Example 6.17. The conductor of $\mathbb{Z}$ in $\mathbb{Z}[i]$ is (0). The conductor of $\mathbb{Z}[\sqrt{-3}]$ in $\mathbb{Z}[\zeta_3]$ is $(2, 1+\sqrt{-3})$ (these may be viewed as generators over $\mathbb{Z}[\sqrt{-3}]$ or $\mathbb{Z}[\zeta_3]$, or even just $\mathbb{Z}$; note that $(2, 1+\sqrt{-3}) = 2\mathbb{Z}[\zeta_3]$ is principal in $\mathbb{Z}[\zeta_3]$ but not in $\mathbb{Z}[\sqrt{-3}]$).

We are interested in the case where R is a noetherian domain.

Lemma 6.18. *Let R be a noetherian domain. The conductor of R in its integral closure S is nonzero if and only if S is finitely generated as an R-module.*

Proof. This is a special case of Lemma 2.14. □

Recall that we defined a fractional ideal of a noetherian domain R as a finitely generated R-submodule of its fraction field. If R has nonzero conductor then its integral closure S is a fractional ideal of R that is also a ring. This means we can write S as $\frac{1}{r}I$ for some $r \in R$ and R-ideal I, and the conductor $\mathfrak{c}$ is precisely the set of denominators $r \in R$ for which $S = \frac{1}{r}I$ for some R-ideal I (note that the representation $\frac{1}{r}I$ is far from unique).

6.7 Orders in Dedekind domains

We now introduce the notion of an *order* (in a Dedekind domain). This should not be confused with the notion of a reflexive, transitive, antisymmetric relation on a set, rather it is a literal translation of the German *Ordnung*, which refers to a ring of algebraic integers.

Definition 6.19. An *order* $\mathcal{O}$ is a noetherian domain of dimension one whose conductor is nonzero, equivalently, whose integral closure is finitely generated as an $\mathcal{O}$-module.[2]

Every Dedekind domain that is not a field is also an order. The integral closure of an order is always a Dedekind domain, but not every ring whose integral closure is a Dedekind domain is an order: as shown by Nagata [5, p. 212], one can construct noetherian domains of dimension one with zero conductor. But in the case of interest to us the conductor is automatically nonzero: in the $AKLB$ setup B is finitely generated over A (by Proposition 5.22), hence over every intermediate ring between A and B, including all those whose integral closure is B. In particular, if $A[\alpha]$ and B have the same fraction field (so $L = K(\alpha)$), then $A[\alpha]$ is an order in B (assuming $B \neq L$).

There is an alternative definition of an order that coincides with our definition in the case of interest to us. Recall that an A-lattice in a K-vector space L is a finitely generated A-submodule of L that spans L as a K-vector space.

[2]Not all authors require an order to have nonzero conductor (e.g. Neukirch [6, §I.12]), but nearly all of the interesting theorems about orders require this assumption, so we include it in the definition.

Definition 6.20. Let A be a noetherian domain with fraction field K, and let L be a (not necessarily commutative) K-algebra of finite dimension. An *A-order* in L is an A-lattice that is also a ring.

Remark 6.21. In general the K-algebra L (and the order $\mathcal{O}$) in Definition 6.20 need not be commutative (even though A necessarily is). For example, the endomorphism ring of an elliptic curve is isomorphic to a $\mathbb{Z}$-order in a $\mathbb{Q}$-algebra L of dimension 1, 2, or 4. This $\mathbb{Z}$-order is necessarily commutative in dimensions 1 and 2, where L is either $\mathbb{Q}$ or an imaginary quadratic field, but it is non-commutative in dimension 4, where L is a quaternion algebra.

Proposition 6.22. *Assume $AKLB$ and let $\mathcal{O}$ be a subring of L. Then $\mathcal{O}$ is an A-order in L if and only if it is an order with integral closure B.*

Proof. We first recall that under our $AKLB$ assumption, $\dim A = 1$, hence $\dim B = 1$, since $A = B \cap K$, and $\mathcal{O} \subseteq L$ is an A-module containing 1, so it contains A.

Suppose $\mathcal{O}$ is an A-order in L. Then $\mathcal{O}$ is an A-lattice, hence finitely generated as an A-module, and therefore integral over A (see [1, Thm. 10.8], for example). Thus $\mathcal{O}$ lies in the integral closure B of A in L. The fraction field of $\mathcal{O}$ is a K-vector space spanning L, hence equal to L, so $\mathcal{O}$ and B have the same fraction field and B is the integral closure of $\mathcal{O}$. Thus $\mathcal{O}$ is a domain of dimension 1 (since B is), and it is noetherian because it is a finitely generated over the noetherian ring A. The integral closure B of $\mathcal{O}$ is finitely generated over A, hence over $\mathcal{O}$; therefore $\mathcal{O}$ is an order.

Now suppose $\mathcal{O}$ is an order with integral closure B. It is an A-submodule of the noetherian A-module B, hence finitely generated over A. It contains a K-basis for L because L is its fraction field (take any K-basis for L written as fractions over $\mathcal{O}$ and clear denominators). Thus $\mathcal{O}$ is an A-lattice in L that is also a ring, hence it is an A-order in L. □

Remark 6.23. There may be subrings $\mathcal{O}$ of L that are orders but not A-orders in L, but these do not have B as their integral closure. Consider $A = B = \mathbb{Z}$, $K = L = \mathbb{Q}$, and $\mathcal{O} = \mathbb{Z}_{(2)}$, for example. In this case $\mathcal{O}$ is a DVR, hence a Dedekind domain, hence an order, but it is not an A-order in L, because it is is not finitely generated over A. But its integral closure is not B (indeed, $\mathcal{O} \not\subseteq B$).

Remark 6.24. An A-order in L is a *maximal order* if it is not properly contained in any other A-order in L. When A is a Dedekind domain one can show that every A-order in L lies in a maximal order. Maximal orders are not unique in general, but in the $AKLB$ setup B is the unique maximal order.

As with Dedekind domains, we call a nonzero prime ideal $\mathfrak{p}$ in an order $\mathcal{O}$ a *prime* of $\mathcal{O}$, and if $\mathfrak{q}$ is a prime of the integral closure B of $\mathcal{O}$ lying above $\mathfrak{p}$ (dividing $\mathfrak{p}B$) then we may write $\mathfrak{q}|\mathfrak{p}$ to indicate this. As in the $AKLB$ setup, we have $\mathfrak{q}|\mathfrak{p}$ if and only if $\mathfrak{q} \cap \mathcal{O} = \mathfrak{p}$, by Lemma 5.28. The fact that B is integrally closed ensures that every prime $\mathfrak{p}$ of $\mathcal{O}$ has at least one prime $\mathfrak{q}$ lying above it (this is a standard fact of commutative algebra). We thus have a surjective map

$$\begin{aligned} \operatorname{Spec} B &\twoheadrightarrow \operatorname{Spec} \mathcal{O} \\ \mathfrak{q} &\mapsto \mathfrak{q} \cap \mathcal{O} \end{aligned}$$

If a prime $\mathfrak{q}$ of B contains the conductor $\mathfrak{c}$, then so does $\mathfrak{p} = \mathfrak{q} \cap \mathcal{O}$ (since $\mathfrak{c} \subseteq \mathcal{O}$), and conversely. It follows that the map is $\operatorname{Spec} B \to \operatorname{Spec} \mathcal{O}$ is still well-defined if we restrict to primes that do not contain $\mathfrak{c}$. In B we can factor $\mathfrak{c}$ into a product of powers of finitely many primes $\mathfrak{q}$; it follows that only finitely many primes $\mathfrak{p}$ of $\mathcal{O}$ contain $\mathfrak{c}$.

Proposition 6.25. *In any order $\mathcal{O}$, only finitely many primes contain the conductor.*

We now show that when we restrict to primes that do not contain the conductor the map $\operatorname{Spec} B \to \operatorname{Spec} \mathcal{O}$ becomes a bijection.

Lemma 6.26. *Let $\mathcal{O}$ be an order with integral closure B and conductor $\mathfrak{c}$ and let $\mathfrak{p}$ be a prime of $\mathcal{O}$ not containing $\mathfrak{c}$. Then $\mathfrak{p}B$ is prime of B.*

Proof. Let $\mathfrak{q}$ be a prime of B lying above $\mathfrak{p}$, so that $\mathfrak{p} = \mathfrak{q} \cap \mathcal{O}$, and pick an element $s \in \mathfrak{c}$ not in $\mathfrak{p}$ (and hence not in $\mathfrak{q}$). Claim: $\mathcal{O}_\mathfrak{p} = B_\mathfrak{q}$. To see that $\mathcal{O}_\mathfrak{p} \subseteq B_\mathfrak{q}$, note that if $a/b \in \mathcal{O}_\mathfrak{p}$ with $a \in \mathcal{O}$ and $b \in \mathcal{O} - \mathfrak{p}$, then $b \in B - \mathfrak{q}$, so $a/b \in B_\mathfrak{q}$. Conversely, if $a/b \in B_\mathfrak{q}$ with $a \in B$ and $b \in B - \mathfrak{q}$ then $sa \in \mathcal{O}$ and $sb \in \mathcal{O} - \mathfrak{p}$, so $(sa)/(sb) = a/b \in \mathcal{O}_p$; here we have used that $sB \subseteq \mathcal{O}$ (since $s \in \mathfrak{c}$) and $sb \notin \mathfrak{q}$ (since $s, b \notin \mathfrak{q}$), so $sb \notin \mathfrak{p}$.

We now note that $\mathfrak{q}' | \mathfrak{p} \Rightarrow B_{\mathfrak{q}'} = \mathcal{O}_\mathfrak{p} = B_\mathfrak{q} \Rightarrow \mathfrak{q}' = \mathfrak{q}$, so there is only one prime $\mathfrak{q}$ lying above $\mathfrak{p}$. It follows that $\mathfrak{p}B = \mathfrak{q}^e$ for some $e \geq 1$, and we claim that $e = 1$. Indeed, we must have $\mathfrak{p}\mathcal{O}_\mathfrak{p} = \mathfrak{q}B_\mathfrak{q}$ (this is the unique maximal ideal of the local ring $\mathcal{O}_\mathfrak{p} = B_\mathfrak{q}$ written in two different ways), so $\mathfrak{q}^e B_\mathfrak{q} = \mathfrak{q}B_\mathfrak{q}$ and therefore $e = 1$. □

Corollary 6.27. *Let $\mathcal{O}$ be an order with integral closure B and conductor $\mathfrak{c}$. The restriction of the map $\operatorname{Spec} B \to \operatorname{Spec} \mathcal{O}$ defined by $\mathfrak{q} \mapsto \mathfrak{q} \cap \mathcal{O}$ to prime ideals not containing $\mathfrak{c}$ is a bijection with inverse $\mathfrak{p} \mapsto \mathfrak{p}B$.*

We now note several conditions on primes of $\mathcal{O}$ that are equivalent to not containing the conductor; these notably include the property of being invertible.

Theorem 6.28. *Let $\mathcal{O}$ be an order with integral closure B and conductor $\mathfrak{c}$, and let $\mathfrak{p}$ be a prime of $\mathcal{O}$. The following are equivalent:*

(a) *$\mathfrak{p}$ does not contain $\mathfrak{c}$;*

(b) *$\mathcal{O} = \{x \in B : x\mathfrak{p} \subseteq \mathfrak{p}\}$;*

(c) *$\mathfrak{p}$ is invertible;*

(d) *$\mathcal{O}_\mathfrak{p}$ is a DVR;*

(e) *$\mathfrak{p}\mathcal{O}_\mathfrak{p}$ is principal.*

If any of these equivalent properties hold, then $\mathfrak{p}B$ is a prime of B.

Proof. See Problem Set 3. □

Remark 6.29. Orders in Dedekind domains also have a geometric interpretation. If $\mathcal{O}$ is an order, the curve $X = \operatorname{Spec} \mathcal{O}$ will have a singularity at each closed point P corresponding to a maximal ideal of $\mathcal{O}$ that contains the conductor. Taking the integral closure B of $\mathcal{O}$ yields a smooth curve $Y = \operatorname{Spec} B$ with the same function field as X and a morphism $Y \to X$ that looks like a bijection above non-singular points (a dominant morphism of degree 1). The curve Y is called the *normalization* of X.

Recall that two ideals I and J in a ring A are said to be *relatively prime* or *coprime* if $I + J = A$; we may also say that I is *prime to* J. When A is a noetherian domain this is equivalent to requiring that $I_\mathfrak{p} + J_\mathfrak{p} = A_\mathfrak{p}$ for every prime ideal $\mathfrak{p}$ of A; this follows from Proposition 2.6 and Lemma 3.5. For prime ideals $\mathfrak{p}$ that do not contain J, we have $J_\mathfrak{p} = A_\mathfrak{p}$, in which case $I_\mathfrak{p} + J_\mathfrak{p} = A_\mathfrak{p}$ certainly holds, so we only need to consider the case where $\mathfrak{p}$ contains J. In this case $J_\mathfrak{p}$ is contained in $\mathfrak{p}A_\mathfrak{p}$ and $I_\mathfrak{p} + J_\mathfrak{p} = A_\mathfrak{p}$ if and only if $I_\mathfrak{p} \not\subseteq \mathfrak{p}A_\mathfrak{p}$, in which case $I_\mathfrak{p} = A_\mathfrak{p}$, equivalently, $IA_\mathfrak{p} = A_\mathfrak{p}$. This leads to the following definition.

Definition 6.30. Let A be a noetherian domain and let J be an ideal of A. A fractional ideal I of A is *prime to* J if $IA_{\mathfrak{p}} = A_{\mathfrak{p}}$ for all prime ideals $\mathfrak{p}$ that contain J. The set of invertible fractional ideals prime to J is denoted $\mathcal{I}_A^J$; it is a subgroup of the ideal group $\mathcal{I}_A$.

To check that $\mathcal{I}_A^J$ is in fact a subgroup, we note that if $\mathfrak{p}$ is any prime containing J then (a) $(1)A_{\mathfrak{p}} = A_{\mathfrak{p}}$, (b) if $IA_{\mathfrak{p}} = A_{\mathfrak{p}}$ then $I^{-1}A_{\mathfrak{p}} = I^{-1}IA_{\mathfrak{p}} = A_{\mathfrak{p}}$ (c) if $I_1A_{\mathfrak{p}} = A_{\mathfrak{p}}$ and $I_2A_{\mathfrak{p}} = A_{\mathfrak{p}}$ then $I_1I_2A_{\mathfrak{p}} = I_2A_{\mathfrak{p}} = A_{\mathfrak{p}}$.

Theorem 6.31. *Let $\mathcal{O}$ be an order with integral closure B. Let $\mathfrak{c}$ be any ideal of B contained in the conductor of $\mathcal{O}$. The map $\mathfrak{q} \mapsto \mathfrak{q} \cap \mathcal{O}$ induces a group isomorphism from $\mathcal{I}_B^{\mathfrak{c}}$ to $\mathcal{I}_{\mathcal{O}}^{\mathfrak{c}}$ and both groups are isomorphic to the free abelian group generated by their prime ideals. In particular, every fractional ideal of $\mathcal{O}$ prime to the conductor has a unique factorization into prime ideals $\prod \mathfrak{p}_i^{e_i}$ which matches the factorization $IB = \prod \mathfrak{q}_i^{e_i}$ with $\mathfrak{p}_i = \mathfrak{q}_i \cap \mathcal{O}$.*

Proof. The B-ideal $\mathfrak{c}$ lies in the conductor of $\mathcal{O}$ and is therefore also an $\mathcal{O}$-ideal, so the subgroups $\mathcal{I}_B^{\mathfrak{c}}$ and $\mathcal{I}_{\mathcal{O}}^{\mathfrak{c}}$ are well defined and the map $\mathfrak{q} \to \mathfrak{q} \cap \mathcal{O}$ gives a bijection between the sets of prime ideals contained in these subgroups, by Corollary 6.27; the theorem follows. □

We now return to the $AKLB$ setup. Let $\mathcal{O}$ be an order in B with conductor $\mathfrak{c}$. For example, we could take $\mathcal{O} = A[\alpha]$, where $L = K(\alpha)$ with $\alpha \in B$, as in the Dedekind-Kummer Theorem. Theorem 6.31 implies that we can determine how primes of A split in B by looking at their factorizations in $\mathcal{O}$, provided we restrict to primes $\mathfrak{p}$ that do not contain $\mathfrak{c} \cap A$. This restriction ensures that the primes $\mathfrak{q}$ of B and $\mathfrak{q}' = \mathfrak{q} \cap \mathcal{O}$ lying above $\mathfrak{p}$ are all prime to $\mathfrak{c}$ and hence to the conductor, so the factorizations of $\mathfrak{p}B$ and $\mathfrak{p}\mathcal{O}$ will match up. In order to complete the picture, we now show that the residue field degrees of the primes in these factorizations also match.

Proposition 6.32. *Assume $AKLB$ and let $\mathcal{O}$ be an order with integral closure B. Let $\mathfrak{c} = (\mathfrak{c}' \cap A)B$, where $\mathfrak{c}'$ is the conductor of $\mathcal{O}$. Then $\mathcal{O}$ is an A-lattice in L and the restrictions of the norm maps $N_{B/A}$ and $N_{\mathcal{O}/A}$ to $I_B^{\mathfrak{c}}$ and $\mathcal{I}_{\mathcal{O}}^{\mathfrak{c}}$ commute with the isomorphism $\mathcal{I}_B^{\mathfrak{c}} \to \mathcal{I}_{\mathcal{O}}^{\mathfrak{c}}$ defined by $\mathfrak{q} \mapsto \mathfrak{q} \cap \mathcal{O}$. If $\mathfrak{q}$ is a prime of B that does not contain $\mathfrak{c}$ and $\mathfrak{q}' = \mathfrak{q} \cap \mathcal{O}$ and $\mathfrak{p} = \mathfrak{q} \cap A$, then $N_{B/A}(\mathfrak{q}) = N_{\mathcal{O}/A}(\mathfrak{q}') = \mathfrak{p}^{f_{\mathfrak{q}}}$ and $[B/\mathfrak{q} : A/\mathfrak{p}] = [\mathcal{O}/\mathfrak{q}' : A/\mathfrak{p}]$.*

Proof. We first note that $(\mathfrak{c}' \cap A)\mathcal{O} \subseteq \mathfrak{c}'$, so $\mathfrak{c} = (\mathfrak{c}' \cap A)B \subseteq \mathfrak{c}'B = \mathfrak{c}'$, thus $\mathfrak{c}$ is contained in the conductor of $\mathcal{O}$. That $\mathcal{O}$ is an A-lattice in L follows from Proposition 6.22. Let $\mathfrak{q}$ be a prime of B that does not contain $\mathfrak{c}$, and define $\mathfrak{q}' := \mathfrak{q} \cap \mathcal{O}$ and $\mathfrak{p} := \mathfrak{q} \cap A$. If $\mathfrak{p}'$ is any prime of A other than $\mathfrak{p}$, then the localization of $\mathfrak{q}$ at $\mathfrak{p}'$ contains B and the localization of $\mathfrak{q}'$ at $\mathfrak{p}'$ contains $\mathcal{O}$ (pick $a \in \mathfrak{p} - \mathfrak{p}'$ and note that $a/a = 1$ lies in both $\mathfrak{q}$ and $\mathfrak{q}'$); we thus have

$$N_{B/A}(\mathfrak{q})_{\mathfrak{p}'} = [B_{\mathfrak{p}'} : \mathfrak{q}_{\mathfrak{p}'}]_{A_{\mathfrak{p}'}} = [B_{\mathfrak{p}'} : B_{\mathfrak{p}'}]_{A_{\mathfrak{p}'}} = A_{\mathfrak{p}'} = [\mathcal{O}_{\mathfrak{p}'} : \mathcal{O}_{\mathfrak{p}'}]_{A_{\mathfrak{p}'}} = [\mathcal{O}_{\mathfrak{p}'} : \mathfrak{q}'_{\mathfrak{p}'}]_{A_{\mathfrak{p}'}} = N_{\mathcal{O}/A}(\mathfrak{q}')_{\mathfrak{p}'}$$

For the prime $\mathfrak{p}$ we proceed as in the proof of Lemma 6.26 and pick $s \in (\mathfrak{c} \cap A) - \mathfrak{p}$. We then find that $B_{\mathfrak{p}} = \mathcal{O}_{\mathfrak{p}}$ and $\mathfrak{q}_{\mathfrak{p}} = \mathfrak{q}'_{\mathfrak{p}}$, and therefore

$$N_{B/A}(\mathfrak{q})_{\mathfrak{p}} = [B_{\mathfrak{p}} : \mathfrak{q}_{\mathfrak{p}}]_{A_{\mathfrak{p}}} = [\mathcal{O}_{\mathfrak{p}} : \mathfrak{q}'_{\mathfrak{p}}]_{A_{\mathfrak{p}}} = N_{\mathcal{O}/A}(\mathfrak{q}')_{\mathfrak{p}}.$$

Thus $N_{B/A}(\mathfrak{q})_{\mathfrak{p}} = N_{B/A}(\mathfrak{q}')_{\mathfrak{p}}$ for all primes $\mathfrak{p}$ of A, and

$$N_{B/A}(\mathfrak{q}) = \cap_{\mathfrak{p}} N_{B/A}(\mathfrak{q})_{\mathfrak{p}} = \cap_{\mathfrak{p}} N_{\mathcal{O}/A}(\mathfrak{q}')_{\mathfrak{p}} = N_{\mathcal{O}/A}(\mathfrak{q}').$$

The proof that $N_{B/A}(\mathfrak{q}) = \mathfrak{p}^{f_{\mathfrak{q}}}$ in Theorem 6.10 does not depend on the fact that B is a Dedekind domain and applies equally to the order $\mathcal{O}$. Thus $N_{\mathcal{O}/A}(\mathfrak{q}') = \mathfrak{p}^{f_{\mathfrak{q}'}}$, where $f_{\mathfrak{q}'} := [\mathcal{O}/\mathfrak{q}' : A/\mathfrak{p}]$. We therefore have $f_{\mathfrak{q}'} = f_{\mathfrak{q}}$ and $[B/\mathfrak{q} : A/\mathfrak{p}] = [\mathcal{O}/q' : A/\mathfrak{p}]$ as claimed. □

Corollary 6.33. *The assumption $B = A[\alpha]$ in the Dedekind-Kummer theorem can be replaced with the assumption that $\mathfrak{p}B$ is prime to the conductor of $A[\alpha]$ in B.*

Remark 6.34. In the special case where $A = \mathbb{Z}$ and $L = \mathbb{Q}(\alpha)$ is a number field generated by an algebraic integer α, for any prime number p, the ideal $p\mathcal{O}_L$ is prime to the conductor of $A[\alpha]$ if and only if p does not divide the index n of $A[\alpha]$ in $\mathcal{O}_L$, as we now explain. The conductor $\mathfrak{c}$ is an $\mathcal{O}_L$-ideal with absolute norm $[\mathcal{O}_L : \mathfrak{c}]$, and it is also an $A[\alpha]$-ideal, hence contained in $A[\alpha]$, so $[\mathcal{O}_L : \mathfrak{c}] = [\mathcal{O}_L : A[\alpha]][A[\alpha] : \mathfrak{c}]$ is divisible by $n = [\mathcal{O}_L : A[\alpha]]$. If $p|n$ then $p|[\mathcal{O}_L : \mathfrak{c}]$ and $p\mathcal{O}_L$ must have a prime of $\mathcal{O}_L$ above p that divides $\mathfrak{c}$. Conversely if $p\mathcal{O}_L$ is not prime to $\mathfrak{c}$ then there is a prime $\mathfrak{q}$ of $\mathcal{O}_L$ above p that divides $\mathfrak{c}$, and it follows that $p = [\mathcal{O}_L : \mathfrak{q}]$ divides $[\mathcal{O}_L : \mathfrak{c}]$, hence p divides either $\mathcal{O}_L : A[\alpha]]$ or $[A[\alpha] : \mathfrak{c}]$. The latter cannot hold because it would imply that $\mathfrak{q}$ is an $A[\alpha]$-ideal, hence divisible by the conductor $\mathfrak{c}$ (and therefore equal to $\mathfrak{c}$), but then $[\mathcal{O}_L : \mathfrak{c}] = [\mathcal{O}_L : \mathfrak{q}]$ and $[\mathcal{O}_L : A[\alpha]] = 1$ which is impossible when $A[\alpha]$ has nontrivial conductor $\mathfrak{c} = \mathfrak{q}$.

For number fields $L = \mathbb{Q}[x]/(x^n + ax^m + b)$ with $m|n$, the article [4] gives a precise characterization of the primes p dividing $[\mathcal{O}_L : A[\alpha]]$ (equivalently, dividing the conductor of $A[\alpha]$, as argued above), including necessary and sufficient criteria for L to be monogenic.

References

[1] Allen Altman and Steven Kleiman, *A term of commutative algebra*, Worldwide Center of Mathematics, 2013.

[2] David Eisenbud, *Commutative algebra with a view toward algebraic geometry*, Springer, 1995.

[3] Albrecht Fröhlich, *Ideals in an extension field as modules over the algebraic integers in a finite number field*, Math. Z. **74** (1960), 29–38.

[4] Anuj Jakhar, Sudesh K. Khanduja, Neraj Sangwan, *On prime divisors of the index of an algebraic integer*, J. Number Theory **2016** (166), 47–61.

[5] M. Nagata, *Local rings*, John Wiley & Sons, 1962.

[6] J. Neukirch, *Algebraic number theory*, Springer, 1999.

7 Galois extensions, Frobenius elements, and the Artin map

In our standard $AKLB$ setup, A is a Dedekind domain with fraction field K, and L/K is a finite separable extension of its fraction field (and B is the integral closure of A in L, also a Dedekind domain). We now consider the case where L/K is also normal, hence Galois, and let $G := \mathrm{Gal}(L/K)$ to denote the Galois group; we will use $AKLBG$ to denote this setup.

7.1 Splitting primes in Galois extensions

We begin by showing that the Galois group G acts on the ideal group $\mathcal{I}_B$ (the invertible, equivalently, nonzero, fractional ideals of B) and that this action is compatible with the group structure of $\mathcal{I}_B$. More precisely, $\mathcal{I}_B$ is a left G-module.

Definition 7.1. Let G be a group. A *left G-module* is an abelian group M equipped with a left G-action that commutes with its group operation; in additive notation we have $\sigma(a+b) = \sigma(a) + \sigma(b)$ for all $\sigma \in G$ and $a, b \in M$. One similarly defines a *right G-module* as an abelian group with a right G-action that commutes with the group operation.

Theorem 7.2. *Assume $AKLBG$. For each fractional ideal I of B and $\sigma \in G$ define*

$$\sigma(I) \coloneqq \{\sigma(x) : x \in I\}.$$

The set $\sigma(I)$ is a fractional ideal of B, and this defines a group action on $\mathcal{I}_B$ that makes it a left G-module. Moreover, the restriction of this action to $\operatorname{Spec} B$ makes it a G-set.

Proof. We first show that $\sigma(B) = B$ for all $\sigma \in G$. Each $b \in B$ is integral over A, hence $f(b) = 0$ for some monic polynomial $f \in A[x]$, and we have

$$0 = \sigma(0) = \sigma(f(b)) = f(\sigma(b)),$$

so $\sigma(b)$ is also integral over A, hence an element of B, since B is the integral closure of A in L. This proves $\sigma(B) \subseteq B$, and the same argument shows $\sigma^{-1}(B) \subseteq B$, hence $B \subseteq \sigma(B)$ and therefore $\sigma(B) = B$ as claimed.

Each $\sigma \in G = \mathrm{Gal}(L/K)$ is a field automorphism of L and thus commutes with addition and multiplication. It follows that if $I \subseteq L$ is a finitely generated B-module (a fractional ideal) then $\sigma(I)$ is a finitely generated $\sigma(B)$-module, and $\sigma(B) = B$, so $\sigma(I)$ is a finitely generated B-module, hence a fraction ideal as claimed. We clearly have $\sigma((0)) = (0)$ for all $\sigma \in G$, so G permutes $\mathcal{I}_B$, the group of nonzero fractional ideals. We also have

$$(\sigma\tau)(I) = \{(\sigma\tau)(x) : x \in I\} = \{\sigma(\tau(x)) : x \in I\} = \{\sigma(y) : y \in \tau(I)\} = \sigma(\tau(I)),$$

and the identity clearly acts trivially, so we have a left G-action on $\mathcal{I}_B$.

Now let $I, J \in \mathcal{I}_B$ and $\sigma \in G$. Each $x \in IJ$ has the form $x = a_1b_1 + \cdots + a_nb_n$ with $a_i \in I$ and $b_i \in J$, and $\sigma(x) = \sigma(a_1)\sigma(b_1) + \cdots + \sigma(a_n)\sigma(b_n) \in \sigma(I)\sigma(J)$. Thus $\sigma(IJ) \subseteq \sigma(I)\sigma(J)$, and applying the same argument to $\sigma(I), \sigma(J)$, and σ^{-1} implies $\sigma^{-1}(\sigma(I)\sigma(J)) \subseteq IJ$ and therefore $\sigma(I)\sigma(J) \subseteq \sigma(IJ)$. Thus $\sigma(IJ) = \sigma(I)\sigma(J)$ for all $I, J \in \mathcal{I}_B$, implying that $\mathcal{I}_B$ is a left G-module.

Let $\mathfrak{p}$ be a prime of B and let $\sigma(\mathfrak{p}) = \mathfrak{q}_1^{e_1} \cdots \mathfrak{q}_n^{e_n}$ be the unique factorization of $\sigma(\mathfrak{p})$ in B. Applying σ^{-1} to both sides yields $\mathfrak{p} = \sigma^{-1}(\mathfrak{q}_1)^{e_1} \cdots \sigma^{-1}(\mathfrak{q}_n)^{e_n}$, and therefore $n = 1$ and $e_1 = 1$, since $\mathfrak{p}$ is prime, thus $\sigma(\mathfrak{p}) = \mathfrak{q}_1$ is prime and the G-action on $\mathcal{I}_B$ restricts to a G-action on $\operatorname{MaxSpec} B$, and on $\operatorname{Spec} B$, since G fixes $\{(0)\} = \operatorname{Spec} B - \operatorname{MaxSpec} B$. □

Recall that by a prime of A (or K) we mean a nonzero prime ideal of A, and similarly for B (and L), and for any prime $\mathfrak{p}$ of A we use $\{\mathfrak{q}|\mathfrak{p}\}$ to denote the set of primes $\mathfrak{q}$ that lie above $\mathfrak{p}$ (equivalently, for which $\mathfrak{q} = A \cap \mathfrak{p}$); in other words, $\{\mathfrak{q}|\mathfrak{p}\}$ is the fiber of the contraction map $\operatorname{MaxSpec} B \to \operatorname{MaxSpec} A$ above $\mathfrak{p}$.

Corollary 7.3. *Assume AKLBG. For each prime $\mathfrak{p}$ of A the group G acts transitively on the set $\{\mathfrak{q}|\mathfrak{p}\}$; in other words, the orbits of the G-action on $\operatorname{Spec} B$ are the fibers of the contraction map $\operatorname{Spec} B \to \operatorname{Spec} A$.*

Proof. Consider any $\sigma \in G$. For $\mathfrak{q}|\mathfrak{p}$ we have $\mathfrak{p}B \subseteq \mathfrak{q}$ and $\sigma(\mathfrak{p}B) \subseteq \sigma(\mathfrak{q})$, so $\sigma(\mathfrak{q})|\mathfrak{p}$ (in a Dedekind domain, to contain is to divide). Thus $\{\mathfrak{q}|\mathfrak{p}\}$ is closed under the action of G, we just need to show that it consists of a single orbit.

Let $\{\mathfrak{q}|\mathfrak{p}\} = \{\mathfrak{q}_1, \dots, \mathfrak{q}_n\}$ and suppose that $\mathfrak{q}_1$ and $\mathfrak{q}_2$ lie in distinct G-orbits. The primes $\mathfrak{q}_1, \dots, \mathfrak{q}_n$ are maximal ideals, hence pairwise coprime, so by the CRT we have a ring isomorphism

$$\frac{B}{\mathfrak{q}_1 \cdots \mathfrak{q}_n} \simeq \frac{B}{\mathfrak{q}_1} \times \cdots \times \frac{B}{\mathfrak{q}_n},$$

and we may choose $b \in B$ such that $b \equiv 0 \bmod \mathfrak{q}_2$ and $b \equiv 1 \bmod \sigma^{-1}(\mathfrak{q}_1)$ for all $\sigma \in G$ (by hypothesis, $\sigma(\mathfrak{q}_2) \neq \mathfrak{q}_1$ for all $\sigma \in G$, since $\mathfrak{q}_1, \mathfrak{q}_2$ lie in different G-orbits). Then $b \in \mathfrak{q}_2$ and

$$N_{L/K}(b) = \prod_{\sigma \in G} \sigma(b) \equiv 1 \bmod \mathfrak{q}_1,$$

hence $N_{L/K}(b) \notin A \cap \mathfrak{q}_1 = \mathfrak{p}$. But $N_{L/K}(b) \in N_{L/K}(\mathfrak{q}_2) = \mathfrak{p}^{f_{\mathfrak{q}_2}} \subseteq \mathfrak{p}$, a contradiction. □

As shown in the proof of Theorem 7.2, we have $\sigma(B) = B$ for all $\sigma \in G = \operatorname{Gal}(L/K)$, thus each $\sigma \in G$ restricts to a ring automorphism of B that fixes every element of the subring $A = B \cap K$, and thus every element of any prime $\mathfrak{p}$ of A. It follows that σ induces an isomorphism of residue field extensions $\bar{\sigma} \in \operatorname{Hom}_{A/\mathfrak{p}}(B/\mathfrak{q}, B/\sigma(\mathfrak{q}))$ defined by $\bar{\sigma}(x + \mathfrak{q}) \coloneqq \sigma(x) + \sigma(\mathfrak{q})$ for $x \in B$, which we may more compactly write as $\bar{\sigma}(\bar{x}) \coloneqq \overline{\sigma(x)}$ (but note that the $\bar{x}$ and $\overline{\sigma(x)}$ are elements of different residue fields).

Corollary 7.4. *Assume AKLBG and let $\mathfrak{p}$ be a prime of A. The residue field degrees $f_{\mathfrak{q}} \coloneqq [B/\mathfrak{q} : A/\mathfrak{p}]$ are the same for every $\mathfrak{q}|\mathfrak{p}$, as are the ramification indices $e_{\mathfrak{q}} \coloneqq v_{\mathfrak{q}}(\mathfrak{p}B)$.*

Proof. For each $\sigma \in G$ we have an isomorphism of the residue fields $B/\mathfrak{q}$ and $B/\sigma(\mathfrak{q})$ that fixes $A/\mathfrak{p}$, so they clearly have the same degree $f_{\mathfrak{q}} = f_{\sigma(\mathfrak{q})}$, and G acts transitively on $\{\mathfrak{q}|\mathfrak{p}\}$, by Corollary 7.3, so the function $\mathfrak{q} \mapsto f_{\mathfrak{q}}$ must be constant on $\{\mathfrak{q}|\mathfrak{p}\}$.

For each $\sigma \in G$ we also have $\sigma(\mathfrak{p}) = \mathfrak{p}$ and $\sigma(B) = B$, so $\sigma(\mathfrak{p}B) = \mathfrak{p}B$, and for each $\mathfrak{q}|\mathfrak{p}$,

$$e_{\mathfrak{q}} = v_{\mathfrak{q}}(\mathfrak{p}B) = v_{\mathfrak{q}}(\sigma(\mathfrak{p}B)) = v_{\mathfrak{q}}\big(\sigma\big(\prod_{\mathfrak{r}|\mathfrak{p}} \mathfrak{r}^{e_{\mathfrak{r}}}\big)\big) = v_{\mathfrak{q}}\big(\prod_{\mathfrak{r}|\mathfrak{p}} \sigma(\mathfrak{r})^{e_{\mathfrak{r}}}\big) = v_{\mathfrak{q}}\big(\prod_{\mathfrak{r}|\mathfrak{p}} \mathfrak{r}^{e_{\sigma^{-1}(\mathfrak{r})}}\big) = e_{\sigma^{-1}(\mathfrak{q})}.$$

The transitivity of the G-action on $\{\mathfrak{q}|\mathfrak{p}\}$ again implies that $\mathfrak{q} \mapsto e_{\mathfrak{q}}$ is constant on $\{\mathfrak{q}|\mathfrak{p}\}$. □

Corollary 7.4 implies that whenever L/K is Galois, we may unambiguously write $e_{\mathfrak{p}}$ and $f_{\mathfrak{p}}$ instead of $e_{\mathfrak{q}}$ and $f_{\mathfrak{q}}$; recall that we previously defined $g_{\mathfrak{p}} \coloneqq \#\{\mathfrak{q}|\mathfrak{p}\}$.

Corollary 7.5. *Assume AKLBG. For each prime $\mathfrak{p}$ of A we have $e_{\mathfrak{p}} f_{\mathfrak{p}} g_{\mathfrak{p}} = [L : K]$.*

Proof. This follows immediately from Theorem 5.35 and Corollary 7.4. □

Example 7.6. Assume $AKLBG$. When $n := [L:K]$ is prime there are just three ways a prime $\mathfrak{p}$ of A can split in B:

- $e_\mathfrak{p} = n$ and $f_\mathfrak{p} = g_\mathfrak{p} = 1$, in which case $\mathfrak{p}$ is totally ramified in L;
- $f_\mathfrak{p} = n$ and $e_\mathfrak{p} = g_\mathfrak{p} = 1$, in which case $\mathfrak{p}$ remains inert in L if $B/\mathfrak{p}B$ is finite étale;
- $g_\mathfrak{p} = n$ and $e_\mathfrak{p} = f_\mathfrak{p} = 1$, in which case $\mathfrak{p}$ splits completely in L if $B/\mathfrak{p}B$ is finite étale.

Recall from Definition 5.37 that we only defined the terms "remains inert" and "splits completely" for unramified primes, which includes the condition that all the residue field extensions $B/\mathfrak{q}$ of $A/\mathfrak{p}$ are separable, equivalently, that $B/\mathfrak{p}B$ is finite ètale over $A/\mathfrak{p}$. This will automatically hold in the primary case of interest to us, where the residue field $A/\mathfrak{p}$ is finite, hence perfect, and all residue field extensions are separable.

7.2 Decomposition and inertia groups

Definition 7.7. Assume $AKLBG$. For each prime $\mathfrak{q}$ of B the *decomposition group* $D_\mathfrak{q}$ (also denoted $D_\mathfrak{q}(L/K)$) is the stabilizer of $\mathfrak{q}$ in G.

Lemma 7.8. *Assume $AKLBG$ and let $\mathfrak{p}$ be a prime of A. The decomposition groups $D_\mathfrak{q}$ for $\mathfrak{q}|\mathfrak{p}$ are all conjugate in G, with $\#D_\mathfrak{q} = e_\mathfrak{p} f_\mathfrak{p}$ and $[G : D_\mathfrak{q}] = g_\mathfrak{p}$.*

Proof. Points in an orbit of group action have conjugate stabilizers, so the $D_\mathfrak{q}$ for $\mathfrak{q}|\mathfrak{p}$ are all conjugate, by Corollary 7.3. The orbit-stabilizer theorem implies $[G:D_\mathfrak{q}] = \#\{\mathfrak{q}|\mathfrak{p}\} = g_\mathfrak{p}$. We have $\#G = [L:K] = e_\mathfrak{p} f_\mathfrak{p} g_\mathfrak{p}$, by Corollary 7.5, so $\#D_\mathfrak{q} = \#G/[G:D_\mathfrak{q}] = e_\mathfrak{p} f_\mathfrak{p}$. □

Let us now consider a particular prime $\mathfrak{q}|\mathfrak{p}$ of B (by writing $\mathfrak{q}|\mathfrak{p}$ we define $\mathfrak{p}$ as $\mathfrak{q} \cap A$). As noted above, each $\sigma \in G$ induces a residue field isomorphism $\bar\sigma \in \mathrm{Hom}_{A/\mathfrak{p}}(B/\mathfrak{q}, B/\sigma(\mathfrak{q}))$. For $\sigma \in D_\mathfrak{q}$, we have $\sigma(\mathfrak{q}) = \mathfrak{q}$, in which case $\bar\sigma \in \mathrm{Aut}_{A/\mathfrak{p}}(B/\mathfrak{q})$. Moreover, the map $\sigma \mapsto \bar\sigma$ defines a group homomorphism $\pi_\mathfrak{q} \colon D_\mathfrak{q} \to \mathrm{Aut}_{A/\mathfrak{p}}(B/\mathfrak{q})$, since for any $x \in B$ we have

$$\overline{\sigma\tau}(\bar x) = \overline{\sigma\tau(x)} = \overline{\sigma(\tau(x))} = \bar\sigma(\overline{\tau(x)}) = \bar\sigma(\bar\tau(\bar x)).$$

Note that $B/\mathfrak{q}$ need not be a Galois extension of $A/\mathfrak{p}$ even when L is a Galois extension of K, which is why we write $\mathrm{Aut}_{A/\mathfrak{p}}(B/\mathfrak{q})$ and not $\mathrm{Gal}((B/\mathfrak{q})/(A/\mathfrak{p}))$.

Proposition 7.9. *Assume $AKLBG$ and let $\mathfrak{q}|\mathfrak{p}$ be a prime of B. The group homomorphism $\pi_\mathfrak{q} \colon D_\mathfrak{q} \to \mathrm{Aut}_{A/\mathfrak{p}}(B/\mathfrak{q})$ defined by $\sigma \mapsto \bar\sigma$ is surjective and $B/\mathfrak{q}$ is normal over $A/\mathfrak{p}$.*

Proof. Let F be the separable closure of $A/\mathfrak{p}$ in $B/\mathfrak{q}$ and for $\bar b \in F$, pick $b \in B$ such that $b \equiv \bar b \bmod \mathfrak{q}$ and $b \equiv 0 \bmod \sigma^{-1}(\mathfrak{q})$ (so $\sigma(b) \equiv 0 \bmod \mathfrak{q}$) for all $\sigma \in G - D_\mathfrak{q}$; the CRT implies that such an b exists, since for $\sigma \in G - D_\mathfrak{q}$ the ideals $\mathfrak{q}$ and $\sigma(\mathfrak{q})$ are distinct and therefore coprime (since they are maximal ideals). Now define

$$g(x) := \prod_{\sigma \in G} \big(x - \sigma(b)\big) \in A[x],$$

and let $\bar g$ denote the image of g in $(A/\mathfrak{p})[x]$. Observe that $\bar b$ is the root of a polynomial $\bar g \in (A/\mathfrak{p})[x]$ that splits completely in $(B/\mathfrak{q})[x]$, and our choice of $\bar b$ was arbitrary, so this applies to every $\bar b \in F^\times$. It follows that F is a normal (hence Galois) extension of $A/\mathfrak{p}$, and we have $\mathrm{Gal}(F/(A/\mathfrak{p})) \simeq \mathrm{Aut}_{A/\mathfrak{p}}(B/\mathfrak{q})$, since F is the separable closure of $A/\mathfrak{p}$ in $B/\mathfrak{q}$.

For each $\sigma \in G - D_{\mathfrak{q}}$ we have $\overline{\sigma}(\bar{b}) = 0$, so 0 is a root of $\overline{g}(x)$ with multiplicity at least $m = \#(G - D_{\mathfrak{q}})$, and the remaining roots are $\overline{\sigma}(\bar{b})$ for $\sigma \in D_{\mathfrak{q}}$, all of which are $\mathrm{Gal}(F/(A/\mathfrak{p}))$-conjugates of $\bar{b}$. It follows that $\overline{g}(x)/x^m$ divides a power of the minimal polynomial $f(x)$ of $\bar{b}$, but $f(x)$ is irreducible in $(A/\mathfrak{p})[x]$, so $\overline{g}(x)/x^m$ is a power of $f(x)$ and every $\mathrm{Gal}(F/(A/\mathfrak{p}))$-conjugate of $\bar{b}$ has the form $\overline{\sigma}(\bar{b})$ for some $\sigma \in D_{\mathfrak{q}}$. Applying this to $\bar{b}$ chosen so that $F = (A/\mathfrak{p})(\bar{b})$ (by the primitive element theorem) shows that the map $\pi_{\mathfrak{q}} \colon D_{\mathfrak{q}} \to \mathrm{Aut}_{A/\mathfrak{p}}(B/\mathfrak{q}) \simeq \mathrm{Gal}(F/(A/\mathfrak{p}))$ is surjective.

To show that $B/\mathfrak{q}$ is a normal extension of $A/\mathfrak{p}$ we proceed as we did for F: for each $b \in B$ define $g \in A[x]$ and $\overline{g} \in (A/\mathfrak{p})[x]$ as above to show that every $b \in B/\mathfrak{q}$ is the root of a polynomial in $(A/\mathfrak{p})[x]$ that splits completely in $(B/\mathfrak{q})[x]$. □

Definition 7.10. Assume *AKLBG*, and let $\mathfrak{q}|\mathfrak{p}$ be a prime of B. The kernel of the surjective homomorphism $\pi_{\mathfrak{q}} \colon D_{\mathfrak{q}} \to \mathrm{Aut}_{A/\mathfrak{p}}(B/\mathfrak{q})$ is the *inertia group* $I_{\mathfrak{q}}$ of $\mathfrak{q}$.

Corollary 7.11. *Assume AKLBG and let $\mathfrak{q}|\mathfrak{p}$ be a prime of B. We have an exact sequence*

$$1 \longrightarrow I_{\mathfrak{q}} \longrightarrow D_{\mathfrak{q}} \longrightarrow \mathrm{Aut}_{A/\mathfrak{p}}(B/\mathfrak{q}) \longrightarrow 1,$$

and $\#I_{\mathfrak{q}} = e_{\mathfrak{p}}[B/\mathfrak{q} : A/\mathfrak{p}]_i$.

We have shown that the residue field $B/\mathfrak{q}$ is always a normal extension of the residue field $A/\mathfrak{p}$. Let us now suppose that it is also separable, hence Galois; this holds, for example, if $A/\mathfrak{p}$ is a perfect field, and in particular, whenever $A/\mathfrak{p}$ is a finite field. We then have

$$D_{\mathfrak{q}}/I_{\mathfrak{q}} \simeq \mathrm{Aut}_{A/\mathfrak{p}}(B/\mathfrak{q}) = \mathrm{Gal}((B/\mathfrak{q})/(A/\mathfrak{p})).$$

Proposition 7.12. *Assume AKLBG, let $\mathfrak{q}|\mathfrak{p}$ be a prime of B, and suppose $B/\mathfrak{q}$ is a separable extension of $A/\mathfrak{p}$. We have a tower of field extensions $K \subseteq L^{D_{\mathfrak{q}}} \subseteq L^{I_{\mathfrak{q}}} \subseteq L$ with*

$$\begin{aligned} e_{\mathfrak{p}} &= [L : L^{I_{\mathfrak{q}}}] = \#I_{\mathfrak{q}}; \\ f_{\mathfrak{p}} &= [L^{I_{\mathfrak{q}}} : L^{D_{\mathfrak{q}}}] = \#D_{\mathfrak{q}}/\#I_{\mathfrak{q}}; \\ g_{\mathfrak{p}} &= [L^{D_{\mathfrak{q}}} : K] = \#\{\mathfrak{q}|\mathfrak{p}\}. \end{aligned}$$

The fields $L^{D_{\mathfrak{q}}}$ and $L^{I_{\mathfrak{q}}}$ are the *decomposition field* and *inertia field* associated to $\mathfrak{q}$.

Proof. The third equality follows immediately from Lemma 7.8. The second follows from Proposition 7.9 and the separability of $(B/\mathfrak{q})/(A/\mathfrak{p})$, since $D_{\mathfrak{q}}/I_{\mathfrak{q}} \simeq \mathrm{Gal}((B/\mathfrak{q})/(A/\mathfrak{p}))$ has cardinality $f_{\mathfrak{p}} = [B/\mathfrak{q} : A/\mathfrak{p}]$. We then have $[L : L^{D_{\mathfrak{q}}}] = \#D_{\mathfrak{q}} = e_{\mathfrak{p}} f_{\mathfrak{p}}$ and $\#D_{\mathfrak{q}}/\#I_{\mathfrak{q}} = f_{\mathfrak{p}}$, so $\#I_q = e_{\mathfrak{p}}$, so the first equality also holds. □

We now consider an intermediate field E lying between K and L. Let us fix a prime $\mathfrak{q}|\mathfrak{p}$ of B, and let $\mathfrak{q}_E \coloneqq \mathfrak{q} \cap E$, so that $\mathfrak{q}|\mathfrak{q}_E$ and $\mathfrak{q}_E|\mathfrak{p}$, and let us use $\overline{G}_{\mathfrak{q}}(L/K) \coloneqq \mathrm{Aut}_{A/\mathfrak{p}}(B/\mathfrak{q})$, $\overline{G}_{\mathfrak{q}}(L/E) \coloneqq \mathrm{Aut}_{(B\cap E)/\mathfrak{q}_E}(B/\mathfrak{q})$, $\overline{G}_{\mathfrak{q}_E}(E/K) \coloneqq \mathrm{Aut}_{A/\mathfrak{p}}((B \cap E)/\mathfrak{q}_E)$ to denote the automorphism groups of the residue field extensions associated to the tower $K \subseteq E \subseteq L$. We use the notation $D_{\mathfrak{q}}(L/E)$ to denote the decomposition group of $\mathfrak{q}$ relative to the extension L/E (note that L/E is Galois since L/K is), and similarly define $D_{\mathfrak{q}}(L/K)$, as well as $I_{\mathfrak{q}}(L/E)$ and $I_{\mathfrak{q}}(L/K)$. In the case that E/K is also Galois, we similarly use $D_{\mathfrak{q}_E}(E/K)$ and $I_{\mathfrak{q}_E}(E/K)$ to denote the decomposition and inertia group of $\mathfrak{q}_E$ (subgroups of $\mathrm{Gal}(E/K)$).

Proposition 7.13. *Assume AKLBG, let E be an intermediate field between K and L. Let $\mathfrak{q}$ be a prime of B and let $\mathfrak{q}_E = \mathfrak{q} \cap E$ and $\mathfrak{p} = \mathfrak{q} \cap K$. Then*

$$D_{\mathfrak{q}}(L/E) = D_{\mathfrak{q}}(L/K) \cap \mathrm{Gal}(L/E) \qquad \text{and} \qquad I_{\mathfrak{q}}(L/E) = I_{\mathfrak{q}}(L/K) \cap \mathrm{Gal}(L/E).$$

If E/K is Galois, then we have the following commutative diagram of exact sequences:

$$\begin{array}{ccccccccc}
 & & 1 & & 1 & & 1 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \longrightarrow & I_{\mathfrak{q}}(L/E) & \longrightarrow & I_{\mathfrak{q}}(L/K) & \longrightarrow & I_{\mathfrak{q}_E}(E/K) & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \longrightarrow & D_{\mathfrak{q}}(L/E) & \longrightarrow & D_{\mathfrak{q}}(L/K) & \longrightarrow & D_{\mathfrak{q}_E}(E/K) & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \longrightarrow & \overline{G}_{\mathfrak{q}}(L/E) & \longrightarrow & \overline{G}_{\mathfrak{q}}(L/K) & \longrightarrow & \overline{G}_{\mathfrak{q}_E}(E/K) & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 1 & & 1 & & 1 & &
\end{array}$$

Proof. Note that $D_{\mathfrak{q}}(L/E) \subseteq \mathrm{Gal}(L/E) \subseteq \mathrm{Gal}(L/K)$. An element σ of $\mathrm{Gal}(L/K)$ lies in $D_{\mathfrak{q}}(L/E)$ if and only if it fixes E (hence lies in $\mathrm{Gal}(L/E)$) and satisfies $\sigma(\mathfrak{q}) = \mathfrak{q}$ (hence lies in $D_{\mathfrak{q}}(L/K)$), which proves the first claim. For the second claim, the restriction of $\pi_{\mathfrak{q}}(L/K)\colon D_{\mathfrak{q}}(L/K) \to \overline{G}_{\mathfrak{q}}(L/K)$ to $D_{\mathfrak{q}}(L/E)$ is the map $\pi_{\mathfrak{q}}(L/E)\colon D_{\mathfrak{q}}(L/E) \to \overline{G}_{\mathfrak{q}}(L/E)$, hence the kernels agree after intersecting with $\mathrm{Gal}(L/E)$.

The exactness of the columns follows from Corollary 7.11; we now argue exactness of the rows. Each row corresponds to an inclusion followed by a restriction in which the inclusion is precisely the kernel of the restriction (for the first two rows this follows from the two claims proved above and for the third row it follows from the main theorem of Galois theory); exactness at the first two groups in each row follows. Surjectivity of the restriction maps follows from the bijection used in the proof of Lemma 4.10. We have a bijection $\mathrm{Hom}_K(L,\Omega) \to \mathrm{Hom}_E(L,\Omega) \times \mathrm{Hom}_K(E,\Omega)$ whose second factor is restriction, and we may view this as a bijection $\phi\colon \mathrm{Gal}(L/K) \to \mathrm{Gal}(L/E) \times \mathrm{Gal}(E/K)$. If $\sigma \in \mathrm{Gal}(E/K)$ stabilizes $\mathfrak{q}_E$ then $\phi^{-1}(1,\sigma) \in \mathrm{Gal}(L/K)$ stabilizes $\mathfrak{q}$ and restricts to σ; this implies surjectivity of the restriction maps in the first two rows, and for the third we replace $L/E/K$ with the corresponding tower of residue field extensions (and forget about stabilizing $\mathfrak{q}_E$).

We now argue commutativity of the four corner squares which suffices to prove the commutativity of the enitre diagram. The upper left square commutes because all the maps are inclusions. The upper right square commutes because inclusion and restriction commute. The lower left square commutes because the horizontal maps are inclusions and the vertical maps coincide on $D_{\mathfrak{q}}(L/E)$. The lower right square commutes because the horizontal maps are restrictions and the vertical maps agree after restriction to E. □

7.3 Frobenius elements

We now add the further assumption that the residue fields $A/\mathfrak{p}$ (and therefore $B/\mathfrak{q}$) are finite for all primes $\mathfrak{p}$ of K.[1] This holds, for example, whenever K is a global field (a finite

[1]There exist Dedekind domains A (PIDs even) with a mixture of finite and infinite residue fields; see [1].

extension of $\mathbb{Q}$ or $\mathbb{F}_q(t)$). In this situation $B/\mathfrak{q}$ is necessarily a Galois extension of $A/\mathfrak{p}$ (we don't need Proposition 7.9 for this, finite extensions of finite fields are always Galois). Indeed, recall that every finite extension of a finite field $\mathbb{F}$ has a cyclic Galois group generated by the $\#\mathbb{F}$-power Frobenius automorphism $x \mapsto x^{\#\mathbb{F}}$.

In order to simplify the notation, when working with finite residue fields we may write $\mathbb{F}_\mathfrak{q} := B/\mathfrak{q}$ and $\mathbb{F}_\mathfrak{p} := A/\mathfrak{p}$; these are finite fields of p-power order, where p is the characteristic of $\mathbb{F}_\mathfrak{p}$ (and of $\mathbb{F}_\mathfrak{q}$). Note that the field K (and L) need not have characteristic p (consider the case of number fields), but if the characteristic of K is positive then it must be p (consider the homomorphism $A \to A/\mathfrak{p}$ from the integral domain A to the field $A/\mathfrak{p}$).

Let $\mathfrak{q}|\mathfrak{p}$ be a prime of B. Corollary 7.11 gives us an exact sequence

$$1 \longrightarrow I_\mathfrak{q} \longrightarrow D_\mathfrak{q} \xrightarrow{\pi_\mathfrak{q}} \mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p}) \longrightarrow 1.$$

If $\mathfrak{p}$ (equivalently, $\mathfrak{q}$) is unramified, then $e_\mathfrak{p} = e_\mathfrak{q} = 1$ and $I_\mathfrak{q}$ is trivial. In this case we have an isomorphism

$$\pi_\mathfrak{q} \colon D_\mathfrak{q} \xrightarrow{\sim} \mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p}).$$

The Galois group $\mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p})$ is the cyclic group of order $f_\mathfrak{p} = [\mathbb{F}_\mathfrak{q} : \mathbb{F}_\mathfrak{p}]$ generated by the Frobenius automorphism

$$x \mapsto x^{\#\mathbb{F}_\mathfrak{p}}.$$

Note that the cardinality of the finite field $\mathbb{F}_\mathfrak{p}$ is necessarily a power of its characteristic p. If $K = \mathbb{Q}$ and $\mathfrak{p} = (p)$ is a prime of $\mathbb{Z}$, then $\mathbb{F}_\mathfrak{p} = \mathbb{Z}/p\mathbb{Z}$ is the field with p elements, but in general the field $\mathbb{F}_\mathfrak{p}$ need not be a prime field (consider $K = \mathbb{Q}(i)$ and $\mathfrak{p} = (7)$).

Definition 7.14. Assume *AKLBG* with finite residue fields and $\mathfrak{q}|\mathfrak{p}$ unramified. The inverse image of the Frobenius automorphism of $\mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p})$ under $\pi_\mathfrak{q} : D_\mathfrak{q} \xrightarrow{\sim} \mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p})$ is the *Frobenius element* $\sigma_\mathfrak{q} \in D_\mathfrak{q} \subseteq G$ (also called the *Frobenius substitution* [2, §8]).

Proposition 7.15. *Assume AKLBG with finite residue fields and $\mathfrak{q}|\mathfrak{p}$ unramified. The Frobenius element $\sigma_\mathfrak{q}$ is the unique $\sigma \in G$ such that for all $x \in B$ we have*

$$\sigma(x) \equiv x^{\#\mathbb{F}_\mathfrak{p}} \bmod \mathfrak{q}.$$

Proof. Clearly $\sigma_\mathfrak{q}$ has this property, we just need to show uniqueness. Suppose $\sigma \in G$ has the desired property. For any $x \in \mathfrak{q}$ we have $x \equiv 0 \bmod \mathfrak{q}$, and $\sigma(x) \equiv x^{\#\mathbb{F}_\mathfrak{p}} \bmod \mathfrak{q}$ implies $\sigma(x) \equiv 0 \bmod \mathfrak{q}$, so $\sigma(x) \in \mathfrak{q}$; it follows that $\sigma(\mathfrak{q}) = \mathfrak{q}$, and therefore $\sigma \in D_\mathfrak{q}$. The isomorphism $\pi_\mathfrak{q} \colon D_\mathfrak{q} \to \mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p})$ maps both σ and $\sigma_\mathfrak{q}$ to the Frobenius automorphism $x \mapsto x^{\#\mathbb{F}_\mathfrak{p}}$, so we must have $\sigma = \sigma_\mathfrak{q}$. □

Proposition 7.16. *Assume AKLBG with finite residue fields and $\mathfrak{q}|\mathfrak{p}$ unramified. For all $\mathfrak{q}'|\mathfrak{p}$ the Frobenius elements $\sigma_\mathfrak{q}$ and $\sigma_{\mathfrak{q}'}$ are conjugate in G.*

Proof. By Corollary 7.3, G acts transitively on $\{\mathfrak{q}|\mathfrak{p}\}$, so let $\tau \in G$ be such that $\mathfrak{q}' = \tau(\mathfrak{q})$. For any $x \in B$ we have

$$\begin{aligned}
\sigma_\mathfrak{q}(x) &\equiv x^{\#\mathbb{F}_\mathfrak{p}} \bmod \mathfrak{q}.\\
\tau(\sigma_\mathfrak{q}(x)) &\equiv \tau\left(x^{\#\mathbb{F}_\mathfrak{p}}\right) \bmod \tau(\mathfrak{q})\\
(\tau\sigma_\mathfrak{q})(x) &\equiv \tau(x)^{\#\mathbb{F}_\mathfrak{p}} \bmod \mathfrak{q}'\\
(\tau\sigma_\mathfrak{q})(\tau^{-1}(x)) &\equiv \tau(\tau^{-1}(x))^{\#\mathbb{F}_\mathfrak{p}} \bmod \mathfrak{q}'\\
(\tau\sigma_\mathfrak{q}\tau^{-1})(x) &\equiv x^{\#\mathbb{F}_\mathfrak{p}} \bmod \mathfrak{q}',
\end{aligned}$$

where we applied τ to both sides in the second line and replaced x by $\tau^{-1}(x)$ in the fourth line. The uniqueness of $\sigma_{\mathfrak{q}'}$ given by Proposition 7.15 implies $\sigma_{\mathfrak{q}'} = \tau\sigma_{\mathfrak{q}}\tau^{-1}$. □

Definition 7.17. Assume *AKLBG* with finite residue fields and $\mathfrak{q}|\mathfrak{p}$ unramified. The conjugacy class of the Frobenius element $\sigma_{\mathfrak{q}} \in G$ is the *Frobenius class* of $\mathfrak{p}$, denoted $\mathrm{Frob}_{\mathfrak{p}}$.

It is common to abuse terminology and refer to $\mathrm{Frob}_{\mathfrak{p}}$ as a Frobenius element $\sigma_{\mathfrak{p}} \in G$ representing its conjugacy class (so $\sigma_{\mathfrak{p}} = \sigma_{\mathfrak{q}}$ for some $\mathfrak{q}|\mathfrak{p}$); there is little risk of confusion so long as we remember that $\sigma_{\mathfrak{p}}$ is only determined up to conjugacy (which usually governs all the properties we care about). There is, however, one situation where this terminology is entirely correct. If G is abelian then each conjugacy classes consists of a single element, in which we case $\mathrm{Frob}_{\mathfrak{p}} = \{\sigma_{\mathfrak{q}} : \mathfrak{q}|\mathfrak{p}\}$ is a singleton set and there is a unique choice for $\sigma_{\mathfrak{p}}$ (note that $\#\{\sigma_{\mathfrak{q}} : \mathfrak{q}|\mathfrak{p}\} = 1$ does not imply $\#\{\mathfrak{q}|\mathfrak{p}\} = 1$; the map $\mathfrak{q} \to \sigma_{\mathfrak{q}}$ is need not be injective).

7.4 Artin symbols

There is another notation commonly used to denote Frobenius elements that includes the field extension in the notation.

Definition 7.18. Assume *AKLBG* with finite residue fields. For each unramified prime $\mathfrak{q}$ of L we define the *Artin symbol*

$$\left(\frac{L/K}{\mathfrak{q}}\right) := \sigma_{\mathfrak{q}}.$$

Proposition 7.19. *Assume AKLBG with finite residue fields and $\mathfrak{q}|\mathfrak{p}$ unramified. Then $\mathfrak{p}$ splits completely if and only if $\left(\frac{L/K}{\mathfrak{q}}\right) = 1$.*

Proof. This follows directly from the definitions: if $\mathfrak{p}$ splits completely then $e_{\mathfrak{p}}f_{\mathfrak{p}} = 1$ and $D_{\mathfrak{q}} = \langle\sigma_{\mathfrak{q}}\rangle = \{1\}$. Conversely, if $D_{\mathfrak{q}} = \langle\sigma_{\mathfrak{q}}\rangle = \{1\}$ then $e_{\mathfrak{p}}f_{\mathfrak{p}} = 1$ and $\mathfrak{p}$ splits completely. □

We will see later in the course that the extension L/K is completely determined by the set of primes $\mathfrak{p}$ that split completely in L. Thus in some sense the Artin symbol captures the essential structure of L/K.

Proposition 7.20. *Assume AKLBG with finite residue fields and let $\mathfrak{q}|\mathfrak{p}$ be unramified. Let E be an intermediate field between K and L, and define $\mathfrak{q}_E := \mathfrak{q} \cap E$. Then*

$$\left(\frac{L/E}{\mathfrak{q}}\right) = \left(\frac{L/K}{\mathfrak{q}}\right)^{[\mathbb{F}_{\mathfrak{q}_E}:\mathbb{F}_{\mathfrak{p}}]},$$

and if E/K is Galois then $\left(\frac{E/K}{\mathfrak{q}_E}\right)$ is the restriction of $\left(\frac{L/K}{\mathfrak{q}}\right)$ to E.

Proof. For the first claim, note that $\#\mathbb{F}_{\mathfrak{q}_E} = (\#\mathbb{F}_{\mathfrak{p}})^{[\mathbb{F}_{\mathfrak{q}_E}:\mathbb{F}_{\mathfrak{p}}]}$. The second claim follows from the commutativity of the lower right square in the commutative diagram of Proposition 7.13: the Frobenius automorphism $x \mapsto x^{\#\mathbb{F}_{\mathfrak{p}}}$ of $\mathrm{Gal}(\mathbb{F}_{\mathfrak{q}_E}/\mathbb{F}_{\mathfrak{p}})$ is the restriction of the Frobenius automorphism $x \mapsto x^{\#\mathbb{F}_{\mathfrak{p}}}$ of $\mathrm{Gal}(\mathbb{F}_{\mathfrak{q}}/\mathbb{F}_{\mathfrak{p}})$ to $\mathbb{F}_{\mathfrak{q}_E}$. □

When L/K is abelian, the Artin symbol takes the same value for all $\mathfrak{q}|\mathfrak{p}$ and we may instead write

$$\left(\frac{L/K}{\mathfrak{p}}\right) := \sigma_{\mathfrak{p}}.$$

In this setting we now view the Artin symbol as a function mapping unramified primes $\mathfrak{p}$ to Frobenius elements $\sigma_{\mathfrak{p}} \in G$. We wish to extend this map to a multiplicative homomorphism from the ideal group $\mathcal{I}_A$ to the Galois group $G = \mathrm{Gal}(L/K)$, but ramified primes $\mathfrak{q}|\mathfrak{p}$ cause problems: the homomorphism $\pi_{\mathfrak{q}} \colon D_{\mathfrak{q}} \to \mathrm{Gal}(\mathbb{F}_{\mathfrak{q}}/\mathbb{F}_{\mathfrak{p}})$ is not a bijection when $\mathfrak{p}$ is ramified (it has nontrivial kernel I_q of order $e_{\mathfrak{q}} = e_{\mathfrak{p}} > 1$).

For any set S of primes of A, let $\mathcal{I}_A^S$ denote the subgroup of $\mathcal{I}_A$ generated by the primes of A that do not lie in S (a free abelian group).

Definition 7.21. Let A be a Dedekind domain with finite residue fields. Let L be a finite abelian extension of $K = \mathrm{Frac}\, A$, and let S be the set of primes of A that ramify in L. The *Artin map* is the homomorphism

$$\begin{aligned}\left(\frac{L/K}{\cdot}\right) : \mathcal{I}_A^S &\to \mathrm{Gal}(L/K)\\ \prod_{i=1}^m \mathfrak{p}_i^{e_i} &\mapsto \prod_{i=1}^m \left(\frac{L/K}{\mathfrak{p}_i}\right)^{e_i}.\end{aligned}$$

Remark 7.22. We will prove in later lectures that the set S of ramified primes is finite, but the definition makes sense in any case.

One of the main results of class field theory is that the Artin map is surjective (this is part of what is known as *Artin reciprocity*). This is a deep theorem that we are not yet ready to prove, but we can verify that it holds in some simple examples.

Example 7.23 (Quadratic fields)**.** Let $K = \mathbb{Q}$ and $L = \mathbb{Q}(\sqrt{d})$ for some square-free integer $d \neq 1$. Then $\mathrm{Gal}(L/K)$ has order 2 and is certainly abelian. As you proved on Problem Set 2, the only ramified primes $\mathfrak{p} = (p)$ of $A = \mathbb{Z}$ are those that divide the *discriminant*

$$D := \mathrm{disc}(L/K) = \begin{cases} d & \text{if } d \equiv 1 \bmod 4,\\ 4d & \text{if } d \not\equiv 1 \bmod 4.\end{cases}$$

If we identify $\mathrm{Gal}(L/K)$ with the multiplicative group $\{\pm 1\}$, then

$$\left(\frac{L/K}{\mathfrak{p}}\right) = \left(\frac{\mathbb{Q}(\sqrt{d})/\mathbb{Q}}{(p)}\right) = \left(\frac{D}{p}\right) = \pm 1,$$

where $(\frac{D}{p})$ is the *Kronecker symbol*. For odd primes $p \nmid D$ we have

$$\left(\frac{D}{p}\right) = \begin{cases} +1 & \text{if } D \text{ is a nonzero square modulo } p,\\ -1 & \text{if } D \text{ is not a square modulo } p,\end{cases}$$

and for $p = 2$ not dividing D (in which case $D = d \equiv 1 \bmod 4$) we have

$$\left(\frac{D}{2}\right) = \begin{cases} +1 & \text{if } D \equiv 1 \bmod 8,\\ -1 & \text{if } D \equiv 5 \bmod 8.\end{cases}$$

The cyclotomic extensions $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ provide another interesting example that you will have an opportunity to explore on Problem Set 4.

References

[1] R. C. Heitmann, *PID's with specified residue fields*, Duke Math. J. **41** (1974), 565–582.

[2] J.-P. Serre, *Local fields*, Springer, 1979.

8 Complete fields and valuation rings

In order to make further progress in our investigation of how primes split in our $AKLB$ setup, and in particular, to determine the primes of K that ramify in L, we introduce a new tool that allows us to "localize" fields. We have seen how useful it can be to localize the Dedekind domain A at a prime ideal $\mathfrak{p}$: this yields a discrete valuation ring $A_{\mathfrak{p}}$, a principal ideal domain with exactly one nonzero prime ideal, which is much easier to study than A, and from Proposition 2.6 we know that the localizations of A at prime ideals collectively determine the structure of A.

Localizing A does not change its fraction field K. But there is an operation we can perform on K that is analogous to localizing A: we can construct the *completion* of K with respect to one of its absolute values. When K is a global field, this yields a *local field*, a term that we will define in the next lecture. At first glance taking completions might seem to make things more complicated, but as with localization, it simplifies matters by allowing us to focus on a single prime, and moreover, work in a complete field.

We begin by briefly reviewing some standard background material on completions, topological rings, and inverse limits.

8.1 Completions

Recall that an absolute value on a field K is a function $|\ |\colon K \to \mathbb{R}_{\ge 0}$ for which

1. $|x| = 0$ if and only if $x = 0$;
2. $|xy| = |x||y|$;
3. $|x + y| \le |x| + |y|$.

If in addition the stronger condition

4. $|x + y| \le \max(|x|, |y|)$

holds, then $|\ |$ is nonarchimedean. This definition does not depend on the fact that K is a field, K could be any ring, but absolute values can exist only when K is an integral domains, since $a, b \ne 0 \Rightarrow |a|, |b| \ne 0 \Rightarrow |ab| = |a||b| \ne 0 \Rightarrow ab \ne 0$; of course an absolute value on an integral domain extends to an absolute value on its fraction field, but it will be convenient to consider absolute values on integral domains as well as fields.

For a more general notion, we can instead consider a metric on a set X, which we recall is a function $d\colon X \times X \to \mathbb{R}_{\ge 0}$ that satisfies

1. $d(x, y) = 0$ if and only if $x = y$;
2. $d(x, y) = d(y, x)$;
3. $d(x, z) \le d(x, y) + d(y, z)$.

A metric that also satisfies

4. $d(x, z) \le \max(d(x, y), d(y, z))$

is an *ultrametric* and is said to be *nonarchimedean*. Every absolute value on a ring induces a metric $d(x, y) := |x - y|$, but not every metric on a ring is induced by an absolute value. The metric d defines a topology on X generated by *open balls*

$$B_{<r}(x) := \{y \in X : d(x, y) < r\}.$$

with $r \in \mathbb{R}_{>0}$ and $x \in X$, and we call X a *metric space*. It is a Hausdorff space, since distinct $x, y \in X$ have disjoint open neighborhoods $B_{<r}(x)$ and $B_{<r}(y)$ (take $r = d(x,y)/2$), and we note that each *closed ball*

$$B_{\le r}(x) := \{y \in X : d(x,y) \le r\}$$

is a closed set, since its complement is the union of $B_{<(d(x,y)-r)}(y)$ over $y \in X - B_{\le r}(x)$.

Definition 8.1. Let X be a metric space. A sequence (x_n) of elements of X *converges* (*to* x) if there is an $x \in X$ such that for every $\epsilon > 0$ there is an $N \in \mathbb{Z}_{>0}$ such that $d(x_n, x) < \epsilon$ for all $n \ge N$; the limit x is necessarily unique. The sequence (x_n) is *Cauchy* if for every $\epsilon > 0$ there is an $N \in \mathbb{Z}_{>0}$ such that $d(x_m, x_n) < \epsilon$ for all $m, n \ge N$. Every convergent sequence is Cauchy, but the converse need not hold. A metric space in which every Cauchy sequence converges is said to be *complete*.

When X is an integral domain with an absolute value $|\ |$ that makes it a complete metric space we say that X is complete with respect to $|\ |$. Which sequences converge and which sequences are Cauchy depends very much on the absolute value $|\ |$ that we use; for example, every integral domain is complete with respect to its trivial absolute value, since then every Cauchy sequence must be eventually constant and obviously converges. Equivalent absolute values necessarily agree on which sequences are convergent and which are Cauchy, so if an integral domain is complete with respect to an absolute value it is complete with respect to all equivalent absolute values.

Definition 8.2. Let X be a metric space. Cauchy sequences (x_n) and (y_n) are *equivalent* if $d(x_n, y_n) \to 0$ as $n \to \infty$; this defines an equivalence relation on the set of Cauchy sequences in X and we use $[(x_n)]$ to denote the equivalence class of (x_n). The *completion* of X is the metric space $\widehat{X}$ whose elements are equivalence classes of Cauchy sequences with the metric

$$d([(x_n)],[(y_n)]) := \lim_{n\to\infty} d(x_n, y_n)$$

(this limit exists and depends only on the equivalence classes of (x_n) and (y_n)). We may canonically embed X in its completion $\widehat{X}$ via the map $x \mapsto \hat{x} = [(x, x, \dots)]$.

When X is a ring we extend the ring operations to $\widehat{X}$ a ring by defining

$$[(x_n)] + [(y_n)] := [(x_n + y_n)] \qquad \text{and} \qquad [(x_n)][(y_n)] := [(x_n y_n)];$$

the additive and multiplicative identities $0 := [(0, 0, \cdots)]$ and $1 := [(1, 1, \dots)]$. When the metric on X is induced by an absolute value $|\ |$, we extend $|\ |$ to an absolute value on $\widehat{X}$ via

$$\big|[(x_n)]\big| := \lim_{n\to\infty} |x_n|.$$

This limit exists and depends only on the equivalence of (x_n), as one can show using the triangle inequality and the fact that (x_n) is Cauchy and $\mathbb{R}$ is complete. When X is a field with a metric induced by an absolute value, the completion $\widehat{X}$ is also a field (this is false in general, see Problem Set 4 for a counter example). Indeed, given $[(x_n)] \ne 0$, we can choose (x_n) with $x_n \ne 0$ for all n, and use the multiplicative property of the absolute value (combined with the triangle inequality), to show that $(1/x_n)$ is Cauchy. We then have $1/[(x_n)] = [(1/x_n)]$, since $[(x_n)][(1/x_n)] = [(1, 1, \dots)] = 1$.

If $|\ |$ arises from a discrete valuation v on K (meaning $|x| \coloneqq c^{v(x)}$ for some $c \in (0,1)$), we extend v to a discrete valuation on $\widehat{X}$ by defining

$$v([(x_n)]) \coloneqq \lim_{n\to\infty} v(x_n) \in \mathbb{Z},$$

for $[(x_n)] \neq \hat{0}$ and $v(\hat{0}) \coloneqq \infty$. Note that for $[(x_n)] \neq \hat{0}$ the sequence $(v(x_n))$ is eventually constant (so the limit is an integer), and we have $|[(x_n)]| = c^{v([(x_n)])}$.

8.1.1 Topological fields with an absolute value

Let K be a field with an absolute value $|\ |$. Then K is also a topological space under the metric $d(x,y) = |x-y|$ induced by the absolute value, and moreover it is a *topological field*.

Definition 8.3. An abelian group G is a *topological group* if it is a topological space in which the map $G \times G \to G$ defined by $(g,h) \mapsto g+h$ and the map $G \to G$ defined by $g \mapsto -g$ are both continuous (here $G \times G$ has the product topology). A commutative ring R is a *topological ring* if it is a topological space in which the maps $R \times R \to R$ defined by $(r,s) \mapsto r+s$ and $(r,s) \mapsto rs$ are both continuous; the additive group of R is then a topological group, since $(-1,s) \mapsto -s$ is continuous, but the unit group $R^\times$ need not be a topological group, in general. A field K is a *topological field* if it is a topological ring whose unit group is a topological group.

If R is a ring with an absolute value then it is a topological ring under the induced topology, and its unit group is also a topological group ; in particular, if R is a field with an absolute value, then it is a topological field under the induced topology. These facts follow from the the triangle inequality and the multiplicative property of an absolute value.

Proposition 8.4. *Let K be a field with an absolute value $|\ |$ viewed as a topological field under the induced topology, and let $\widehat{K}$ be the completion K. The field $\widehat{K}$ is complete, and has the following universal property: every embedding of K as a topological field into a complete field L can be uniquely extended to an embedding of $\widehat{K}$ into L which is an isomorphism whenever K is dense in L. Up to a canonical isomorphism, $\widehat{K}$ is the unique topological field with this property.*

Proof. See Problem Set 4. □

The proposition implies that the completion of $\widehat{K}$ is (isomorphic to) itself, since we can apply the universal property of the completion of $\widehat{K}$ to the trivial embedding $\widehat{K} \to \widehat{K}$. Completing a field that is already complete has no effect. In particular, the completion of K with respect to the trivial absolute value is K, since every field is complete with respect to the trivial absolute value.

Two absolute values on the same field induce the same topology if and only if they are equivalent; this follows from the Weak Approximation Theorem.

Theorem 8.5 (Weak Approximation)**.** *Let K be a field and let $|\cdot|_1, \dots, |\cdot|_n$ be pairwise inequivalent nontrivial absolute values on K. Let $a_1, \dots, a_n \in K$ and let $\epsilon_1, \dots, \epsilon_n$ be positive real numbers. Then there exists an $x \in K$ such that $|x - a_i|_i < \epsilon_i$ for $1 \le i \le n$.*

Proof. See Problem Set 4. □

Corollary 8.6. *Let K be a field with absolute values $|\ |_1$ and $|\ |_2$. The induced topologies on K coincide if and only if $|\ |_1$ and $|\ |_2$ are equivalent.*

Proof. See Problem Set 4. □

The topology induced by a nonarchimedean absolute value has some features that may be counterintuitive to the uninitiated. In particular, every open ball is also closed, so the closure of $B_{<r}(x)$ is not $B_{\le r}(x)$ unless these two sets are already equal, which need not be the case since the map $|\ |\colon K \to \mathbb{R}_{\ge 0}$ need not be surjective; indeed, it will have discrete image if $|\ |$ arises from a discrete valuation. This means that is entirely possible to have $B_{<r}(x) = B_{<s}(x)$ for $r \neq s$; indeed occurs uncountably often when $|\ |$ arises from a discrete valuation. The reader may wish to verify that the following hold in any nonarchimedean metric space X:

1. Every point in an open ball is a center: $B_{<r}(y) = B_{<r}(x)$ for all $y \in B_{<r}(x)$.
2. Any pair of open balls are either disjoint or concentric (have a common center).
3. Every open ball is closed and every closed ball is open.
4. X is *totally disconnected*: every pair of distinct points have disjoint open neighborhoods whose union is the whole space (every connected component is a point).

For any topological space X, the continuity of a map $f\colon X \times X \to X$ implies that for every fixed $x \in X$ the maps $X \to X$ defined by $y \mapsto f(x,y)$ and $y \mapsto f(y,x)$ are continuous, since each is the composition $f \circ \phi$ of f with the continuous map $\phi\colon X \to X \times X$ defined by $y \mapsto (x,y)$ and $y \mapsto (y,x)$, respectively. For an additive topological group G this means that every translation-by-h map $g \mapsto g + h$ is a homeomorphism, since it is continuous and has a continuous inverse (translate by $-h$); in particular, translates of open sets are open and translates of closed sets are closed. Thus in order to understand the topology of a topological group, we can focus on neighborhoods of the identity; a base of open neighborhoods about the identity determines the entire topology. It also means that any topological property of a subgroup (such as being open, closed, or compact) applies to all of its cosets.

If $\widehat{K}$ is the completion of a field K with respect to an absolute value $|\ |$, then $\widehat{K}$ is a topological field with the topology induced by $|\ |$, and the subspace topology on $K \subseteq \widehat{K}$ is the same as the topology on K induced by $|\ |$. By construction, K is *dense* in $\widehat{K}$; indeed, $\widehat{K}$ is precisely the set of limit points of K. More generally, every open ball $B_{<r}(x)$ in K is dense in the corresponding open ball $B_{<r}(\hat{x})$ in $\widehat{K}$.

8.1.2 Inverse limits

Inverse limits are a general construction that can be applied in any category with products, but we will only be concerned with inverse limits in familiar concrete categories such as groups, rings, and topological spaces, all of which are *concrete categories* whose objects can be defined as sets (more formally, concrete categories admit a faithful functor to the category of sets), which allows many concepts to be defined more concretely.

Definition 8.7. A *directed set* is a set I with a relation "$\le$" that is reflexive ($i \le i$), anti-symmetric ($i \le j \le i \Rightarrow i = j$), and transitive ($i \le j \le k \Rightarrow i \le k$), in which every finite subset has an upper bound (in particular, I is non-empty).

Definition 8.8. An *inverse system* (*projective system*) in a category is a family of objects $\{X_i : i \in I\}$ indexed by a directed set I and a family of morphisms $\{f_{ij}\colon X_i \leftarrow X_j : i \le j\}$ (note the direction) such that each f_{ii} is the identity and $f_{ik} = f_{ij} \circ f_{jk}$ for all $i \le j \le k$.[1]

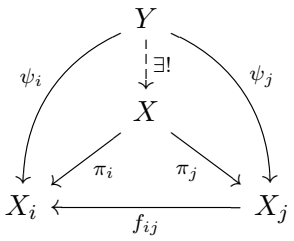
[1] Some (but not all) authors reserve the term *projective system* for cases where the f_{ij} are epimorphisms. This distinction is not relevant to us, as our inverse systems will all use epimorphisms (surjections, in fact).

Definition 8.9. Let (X_i, f_{ij}) be an inverse system in a concrete category with products. The *inverse limit* (or *projective limit*) of (X_i, f_{ij}) is the object

$$X \coloneqq \varprojlim X_i \coloneqq \left\{ x \in \prod_{i\in I} X_i : x_i = f_{ij}(x_j) \text{ for all } i \le j \right\} \subseteq \prod_{i\in I} X_i$$

(whenever such an object X exists in the category). The restrictions $\pi_i \colon X \to X_i$ of the projections $\prod X_i \to X_i$ satisfy $\pi_i = f_{ij} \circ \pi_j$ for $i \le j$.

The object $X = \varprojlim X_i$ has the universal property that if Y is another object with morphisms $\psi_i \colon Y \to X_i$ that satisfy $\psi_i = f_{ij} \circ \psi_j$ for $i \le j$, then there is a unique morphism $Y \to X$ for which all of the diagrams



commute (this universal property defines an inverse limit in any category with products).

As with other categorical constructions satisfying (or defined by) universal properties, uniqueness is guaranteed, but existence is not. However, in any concrete category for which the faithful functor to the category of sets has a left adjoint, inverse limits necessarily exist; this applies to all the categories we shall consider, including the categories of groups, rings, and topological spaces, all of which admit a "free object functor" from the category of sets.

Proposition 8.10. *Let (X_i, f_{ij}) be an inverse system of Hausdorff topological spaces. Then $X \coloneqq \varprojlim X_i$ is a closed subset of $\prod X_i$, and if the X_i are compact then X is compact.*

Proof. The set X is the intersection of the sets $Y_{ij} \coloneqq \{x \in \prod X_i : x_i = f_{ij}(x_j)\}$ with $i \le j$, each of which can be written as $Y_{ij} = \prod_{k\ne i,j} X_k \times Z_{ij}$, where Z_{ij} is the preimage of the diagonal $\Delta_i \coloneqq \{(x_i, x_i) : x_i \in X_i\} \subseteq X_i \times X_i$ under the continuous map $X_i \times X_j \to X_i \times X_i$ defined by $(x_i, x_j) \mapsto (x_i, f_{ij}(x_j))$. Each Δ_i is closed in $X_i \times X_i$ (because X_i is Hausdorff), so each Z_{ij} is closed in $X_i \times X_j$, and each Y_{ij} is closed in $\prod X_i$; it follows that X is a closed subset of $\prod X_i$. By Tychonoff's theorem [1, Thm. I.9.5.3], if the X_i are compact then so is their product $\prod X_i$, in which case the closed subset X is also compact. □

8.2 Valuation rings in complete fields

We now want to specialize to absolute values induced by a discrete valuation $v \colon K^\times \twoheadrightarrow \mathbb{Z}$. If we pick a positive real number $c < 1$ and define $|x|_v \coloneqq c^{v(x)}$ for $x \in K^\times$ and $|0|_v \coloneqq 0$ then we obtain a nontrivial nonarchimedean absolute value $|\ |_v$. Different choices of c yield equivalent absolute values and thus do not change the topology induced by $|\ |_v$ or the completion $K_v \coloneqq \widehat{K}$ of K with respect to $|\ |_v$. We will see later that there is a canonical choice for c when the residue field k of the valuation ring of K is finite (one takes $c = 1/\#k$).

It follows from our discussion that the valuation ring

$$A_v \coloneqq \{x \in K_v : v(x) \ge 0\} = \{x \in K_v : |x|_v \le 1\}$$

is a closed (and therefore open) ball in K_v; indeed, it is the closure of the valuation ring A of K inside K_v. Note that K_v is the fraction field of A_v, since we have $x \in K_v - A_v$ if and only if $1/x \in \hat{A}$; so rather than defining A_v as the valuation ring of K_v we could equivalently define A_v as the completion of A (with respect to $|\ |_v$) and then define K_v as its fraction field.

We now give another characterization of A_v as an inverse limit.

Proposition 8.11. *Let K be a field with absolute value $|\ |_v$ induced by a discrete valuation v, let A be the valuation ring of K, and let π be a uniformizer. The valuation ring of the completion K_v of K with respect to $|\ |_v$ is a complete discrete valuation ring A_v with uniformizer π, and we have an isomorphism of topological rings*

$$A_v \simeq \varprojlim_{n\to\infty} \frac{A}{\pi^n A}.$$

It is immediately clear that A_v is a complete DVR with uniformizer π: it is complete because it is a closed subset of the complete field K_v, it is a DVR with uniformizer π because v extends to a discrete valuation on A_v with $v(\pi) = 1$.

Before proving the non-trivial part of the proposition, let us check that we understand the topology of the inverse limit $X \coloneqq \varprojlim_n A/\pi^n A$. The valuation ring A is a closed ball $B_{\le 1}(0)$ (hence an open set) in the nonarchimedean metric space K, and this also applies to each of the sets $\pi^n A$ (they are closed balls of radius c^n about 0). Each quotient $A/\pi^n A$ therefore has the discrete topology, since the inverse image of any point under the quotient map is a coset of the open subgroup $\pi^n A$. The inverse limit X is a subspace of the product space $\prod_n A/\pi^n A$, whose basic open sets project onto $A/\pi^n A$ for all but finitely many factors (by definition of the product topology). It follows that every basic open subset U of X is the full inverse image (under the canonical projection maps given by the inverse limit construction) of a subset of $A/\pi^m A$ for some $m \ge 1$; all open sets are unions of these basic open sets. When this set is a point we can describe U as a coset $a + \pi^m A$, for some $a \in A$; as a subset $U = \prod_n U_n$ of $\prod_n A/\pi^n A$ each U_n is the image of $a + \pi^m A$ under the quotient map $A \to A/\pi^n A$. In general, U is a union of such sets (all with the same m).

We can alternatively describe the topology on X in terms of an absolute value: for nonzero $x = (x_n) \in X = \varprojlim A/\pi^n A$, let $v(x)$ be the least $n \ge 0$ for which $x_{n+1} \ne 0$, and define $|x|_v \coloneqq c^{v(x)}$. If we embed A in X in the obvious way, $a \mapsto (\bar{a}, \bar{a}, \bar{a}, \ldots)$, the absolute value on X restricts to the absolute value $|\ |_v$ on A, and the subspace topology A inherits from X is the same as that induced by $|\ |_v$. The open sets of X are unions of open balls $B_{<r}(a)$, where we can always choose $a \in A$ (because A is dense in X). If we let $m \ge 0$ be the least integer for which $c^m < r$, where $c \in (0,1)$ is the constant for which $|x| = c^{v(x)}$ for all $x \in A$, then $B_{<r}(a)$ corresponds to a coset $a + \pi^m A$ as above.

Let us now prove the proposition.

Proof. The ring A_v is complete and contains A. For each $n \ge 1$ we define a ring homomorphism $\phi_n \colon A_v \to A/\pi^n A$ as follows: for each $\hat{a} = [(a_i)]$ let $\phi_n(\hat{a})$ be the limit of the eventually constant sequence $(\overline{a}_i)$ of images of a_i in $A/(\pi^n)$. We thus obtain an infinite sequence of surjective maps $\phi_n \colon A_v \to A/\pi^n A$ that are compatible in that for all $n \ge m > 0$ and all $a \in A_v$ the image of $\phi_n(a)$ in $A/\pi^m A$ is $\phi_m(a)$. This defines a surjective ring homomorphism $\phi \colon A_v \to \varprojlim A/\pi^n A$. Now note that

$$\ker \phi = \bigcap_{n\ge 1} \pi^n A_v = \{0\}, \tag{1}$$

so ϕ is injective and therefore an isomorphism. To show that ϕ is also a homeomorphism, it suffices to note that if $a+\pi^m A$ is a coset of $\pi^m A$ in A and U is the corresponding open set in $\varprojlim A/\pi^n A$, then $\phi^{-1}(U)$ is the closure of $a+\pi^m A$ in A_v, which is the coset $a+\pi^m A_v$, an open subset in A_v (as explained in the discussion above, every open set in the inverse limit corresponds to a finite union of cosets $a+\pi^m A$ for some m). Conversely ϕ maps open sets $a+\pi^m A_v$ to open sets in $\varprojlim A/\pi^n A$. □

Remark 8.12. Given any ring R with an ideal I, one can define the *I-adic completion* of R as the inverse limit of topological rings $\varprojlim_n R/I^n$, where each R/I^n is given the discrete topology. Proposition 8.11 shows that when R is a DVR with maximal ideal $\mathfrak{m}$, taking the completion of R with respect to the absolute value $|\ |_{\mathfrak{m}}$ is the same thing as taking the $\mathfrak{m}$-adic completion. This is not true in general. In particular, the $\mathfrak{m}$-adic completion of a (not necessarily discrete) valuation ring R with respect to its maximal ideal $\mathfrak{m}$ need not be complete (either in the sense of Definition 8.1 or in the sense of being isomorphic to its $\mathfrak{m}$-adic completion). The key issue that arises is that the kernel in (1) need not be trivial; indeed, if $\mathfrak{m}^2=\mathfrak{m}$ (which can happen) it certainly won't be. This problem does not occur for valuation rings that are noetherian, but these are necessarily DVRs.

Example 8.13. Let $K=\mathbb{Q}$ and let v_p be the p-adic valuation for some prime p and let $|x|_p := p^{-v_p(x)}$ denote the corresponding absolute value. The completion of $\mathbb{Q}$ with respect to $|\ |_p$ is the field $\mathbb{Q}_p$ of p-adic numbers. The valuation ring of $\mathbb{Q}$ corresponding to v_p is the local ring $\mathbb{Z}_{(p)}$. Taking $\pi=p$ as our uniformizer, we get

$$\widehat{\mathbb{Z}_{(p)}} \simeq \varprojlim_{n\to\infty} \frac{\mathbb{Z}_{(p)}}{p^n\mathbb{Z}_{(p)}} \simeq \varprojlim_{n\to\infty} \frac{\mathbb{Z}}{p^n\mathbb{Z}} \simeq \mathbb{Z}_p,$$

the ring of p-adic integers (note that this example gives two equivalent definitions of $\mathbb{Z}_p$).

Example 8.14. Let $K=\mathbb{F}_q(t)$ be the rational function field over a finite field $\mathbb{F}_q$ and let v_t be the t-adic valuation and let $|x|_t := q^{-v_t(x)}$ be the corresponding absolute value. with uniformizer $\pi=t$. The completion of $\mathbb{F}_q(t)$ with respect to $|\ |_t$ is isomorphic to the field $\mathbb{F}_q((t))$ of Laurent series over $\mathbb{F}_q$. The valuation ring of $\mathbb{F}_q(t)$ with respect to v_t is the local ring $\mathbb{F}_q[t]_{(t)}$ consisting of rational functions whose denominators have nonzero constant term. Taking $\pi=t$ as our uniformizer, we get

$$\widehat{\mathbb{F}_q[t]_{(t)}} \simeq \varprojlim_{n\to\infty} \frac{\mathbb{F}_q[t]_{(t)}}{t^n\mathbb{F}_q[t]_{(t)}} \simeq \varprojlim_{n\to\infty} \frac{\mathbb{F}_q[t]}{t^n\mathbb{F}_q[t]} \simeq \mathbb{F}_q[[t]],$$

where $\mathbb{F}_q[[t]]$ denotes the power series ring over $\mathbb{F}_q$.

Example 8.15. The isomorphism $\mathbb{Z}_p \simeq \varprojlim \mathbb{Z}/p^n\mathbb{Z}$ gives us a canonical way to represent elements of $\mathbb{Z}_p$: we can write $a\in\mathbb{Z}_p$ as a sequence (a_n) with $a_{n+1}\equiv a_n \bmod p^n$, where each

$a_n \in \mathbb{Z}/p^n\mathbb{Z}$ is uniquely represented by an integer in $[0, p^n - 1]$. In $\mathbb{Z}_7$, for example:

$$\begin{aligned}
2 &= (2, 2, 2, 2, 2, \ldots) \\
2002 &= (0, 42, 287, 2002, 2002, \ldots) \\
-2 &= (5, 47, 341, 2399, 16805, \ldots) \\
2^{-1} &= (4, 25, 172, 1201, 8404, \ldots) \\
\sqrt{2} &= \begin{cases} (3, 10, 108, 2166, 4567 \ldots) \\ (4, 39, 235, 235, 12240 \ldots) \end{cases} \\
\sqrt[5]{2} &= (4, 46, 95, 1124, 15530, \ldots)
\end{aligned}$$

While this representation is canonical, it is also redundant. The value of a_n constrains the value of a_{n+1} to just p possible values among the p^{n+1} elements of $\mathbb{Z}/p^{n+1}\mathbb{Z}$, namely, those that are congruent to a_n modulo p^n. We can always write $a_{n+1} = a_n + p^n b_n$ for some $b_n \in [0, p-1]$, namely, $b_n = (a_{n+1} - a_n)/p^n$.

Definition 8.16. Let $a = (a_n)$ be a p-adic integer with each a_n uniquely represented by an integer in $\in [0, p^n - 1]$. The sequence $(b_0, b_1, b_2, \ldots)$ with $b_0 = a_1$ and $b_n = (a_{n+1} - a_n)/p^n$ is called the *p-adic expansion of a*.

Proposition 8.17. *Every element of $\mathbb{Z}_p$ has a unique p-adic expansion and every sequence $(b_0, b_1, b_2, \ldots)$ of integers in $[0, p-1]$ is the p-adic expansion of an element of $\mathbb{Z}_p$.*

Proof. This follows immediately from the definition: we can recover (a_n) from its p-adic expansion $(b_0, b_1, b_2, \ldots)$ via $a_1 = b_0$ and $a_{n+1} = a_n + pb_n$ for all $n \geq 1$. □

Thus we have a bijection between $\mathbb{Z}_p$ and the set of all sequences of integers in $[0, p-1]$ indexed by the nonnegative integers.

Example 8.18. We have the following p-adic expansion in $\mathbb{Z}_7$:

$$\begin{aligned}
2 &= (2, 0, 0, 0, 0, 0, 0, 0, 0, 0, \ldots) \\
2002 &= (0, 6, 5, 5, 0, 0, 0, 0, 0, 0, \ldots) \\
-2 &= (5, 6, 6, 6, 6, 6, 6, 6, 6, 6, \ldots) \\
2^{-1} &= (4, 3, 3, 3, 3, 3, 3, 3, 3, 3, \ldots) \\
5^{-1} &= (3, 1, 4, 5, 2, 1, 4, 5, 2, 1, \ldots) \\
\sqrt{2} &= \begin{cases} (3, 1, 2, 6, 1, 2, 1, 2, 4, 6 \ldots) \\ (4, 5, 4, 0, 5, 4, 5, 4, 2, 0 \ldots) \end{cases} \\
\sqrt[5]{2} &= (4, 6, 1, 3, 6, 4, 3, 5, 4, 6 \ldots)
\end{aligned}$$

You can easily recreate these examples (and many more) in Sage. To create the ring of 7-adic integers, use `Zp(7)`. By default Sage uses 20 digits of p-adic precision, but you can change this to n digits using `Zp(p,n)`.

Performing arithmetic in $\mathbb{Z}_p$ using p-adic expansions is straight-forward. One computes a sum of p-adic expansions $(b_0, b_1, \ldots) + (c_0, c_1, \ldots)$ by adding digits mod p and carrying to the right (don't forget to carry!). Multiplication corresponds to computing products of formal power series in p, e.g. $(\sum b_n p^n)(\sum c_n p^n)$, and can be performed by hand (or in Sage) using the standard schoolbook algorithm for multiplying integers represented in base 10, except now one works in base p. For more background on p-adic numbers, see [2, 3, 4, 5].

8.3 Extending valuations

Recall from Lecture 3 that each prime $\mathfrak{p}$ of a Dedekind domain A determines a discrete valuation (a surjective homomorphism) $v_\mathfrak{p}\colon \mathcal{I}_A \to \mathbb{Z}$ that assigns to a nonzero fractional ideal I the exponent $n_\mathfrak{p}$ appearing in the unique factorization of $I = \prod \mathfrak{p}^{n_\mathfrak{p}}$ into prime ideals; equivalently, $v_\mathfrak{p}(I)$ is the unique integer n for which $IA_\mathfrak{p} = \mathfrak{p}^n A_\mathfrak{p}$. This induces a discrete valuation $v_\mathfrak{p}(x) := v_\mathfrak{p}(xA)$ on the fraction field K, and a corresponding absolute value $|x|_\mathfrak{p} := c^{v_\mathfrak{p}(x)}$ (with $0 < c < 1$). In the $AKLB$ setup, where L/K is a finite separable extension and B is the integral closure of A in L, the primes $\mathfrak{q}|\mathfrak{p}$ of B similarly give rise to discrete valuations $v_\mathfrak{q}$ on L, each of which restricts to a valuation on K, but this valuation need not be equal to $v_\mathfrak{p}$. We want to understand how the discrete valuations $v_\mathfrak{q}$ relate to $v_\mathfrak{p}$.

Definition 8.19. Let L/K be a finite separable extension, and let v and w be discrete valuations on K and L respectively. If $w_{|_K} = ev$ for some $e \in \mathbb{Z}_{>0}$, then we say that w *extends* v *with index* e.

Theorem 8.20. *Assume $AKLB$ and let $\mathfrak{p}$ be a prime of A. For each prime $\mathfrak{q}|\mathfrak{p}$, the discrete valuation $v_\mathfrak{q}$ extends $v_\mathfrak{p}$ with index $e_\mathfrak{q}$, and every discrete valuation on L that extends $v_\mathfrak{p}$ arises in this way. In other words, the map $\mathfrak{q} \mapsto v_\mathfrak{q}$ gives a bijection from $\{\mathfrak{q}|\mathfrak{p}\}$ to the set of discrete valuations of L that extend $v_\mathfrak{p}$.*

Proof. For each prime $\mathfrak{q}|\mathfrak{p}$ we have $v_\mathfrak{q}(\mathfrak{p}B) = e_\mathfrak{q}$ (by definition of the ramification index $e_\mathfrak{q}$), while $v_\mathfrak{q}(\mathfrak{p}'B) = 0$ for all primes $\mathfrak{p}' \neq \mathfrak{p}$ of A (since $\mathfrak{q}$ lies above only the prime $\mathfrak{p} = \mathfrak{q} \cap A$). If $I = \prod_{\mathfrak{p}'} (\mathfrak{p}')^{n_{\mathfrak{p}'}}$ is any nonzero fractional ideal of A then

$$v_\mathfrak{q}(IB) = v_\mathfrak{q}\left(\prod_{\mathfrak{p}'} (\mathfrak{p}')^{n_{\mathfrak{p}'}} B\right) = v_\mathfrak{q}(\mathfrak{p}^{n_\mathfrak{p}} B) = v_\mathfrak{q}(\mathfrak{p}B) n_\mathfrak{p} = e_\mathfrak{q} n_\mathfrak{p} = e_\mathfrak{q} v_\mathfrak{p}(I),$$

so $v_\mathfrak{q}(x) = v_\mathfrak{q}(xB) = e_\mathfrak{q} v_\mathfrak{p}(xA) = e_\mathfrak{q} v_\mathfrak{p}(x)$ for all $x \in K^\times$; thus $v_\mathfrak{q}$ extends $v_\mathfrak{p}$ with index $e_\mathfrak{q}$.

If $\mathfrak{q}$ and $\mathfrak{q}'$ are two distinct primes above $\mathfrak{p}$, then neither contains the other and for any $x \in \mathfrak{q} - \mathfrak{q}'$ we have $v_\mathfrak{q}(x) > 0 \geq v_{\mathfrak{q}'}(x)$, thus $v_\mathfrak{q} \neq v_{\mathfrak{q}'}$ and the map $\mathfrak{q} \mapsto v_\mathfrak{q}$ is injective.

Let w be a discrete valuation on L that extends $v_\mathfrak{p}$, let $W = \{x \in L : w(x) \geq 0\}$ be the associated DVR, and let $\mathfrak{m} = \{x \in L : w(x) > 0\}$ be its maximal ideal. Since $w_{|_K} = ev_\mathfrak{p}$, the discrete valuation w is nonnegative on $A = \{x \in K : w(x) \geq 0\}$ therefore $A \subseteq W$, and elements of A with nonzero valuation are precisely the elements of $\mathfrak{p}$, thus $\mathfrak{p} = \mathfrak{m} \cap A$. The discrete valuation ring W is integrally closed in its fraction field L, so $B \subseteq W$. Let $\mathfrak{q} = \mathfrak{m} \cap B$. Then $\mathfrak{q}$ is prime (since $\mathfrak{m}$ is), and $\mathfrak{p} = \mathfrak{m} \cap A = \mathfrak{q} \cap A$, so $\mathfrak{q}$ lies over $\mathfrak{p}$. The ring W contains $B_\mathfrak{q}$ and is contained in $\operatorname{Frac} B_\mathfrak{q} = L$. But there are no intermediate rings between a DVR and its fraction field, so $W = B_\mathfrak{q}$ and $w = v_\mathfrak{q}$ (and $e = e_\mathfrak{q}$). □

References

[1] N. Bourbaki, *General Topology: Chapters 1-4*, Springer, 1985.

[2] F.Q. Gouvea, *p-adic numbers*, Springer, 1993.

[3] N. Koblitz, *p-adic numbers, p-adic analysis, and zeta functions*, Springer, 1984.

[4] A.M. Robert, *A course in p-adic analysis*, Springer, 2000.

[5] J.-P. Serre, *A course in arithmetic*, Springer, 1973.

9 Local fields and Hensel's lemmas

In this lecture we introduce the notion of a *local field*; these are precisely the fields that arise as completions of a *global field* (finite extensions of $\mathbb{Q}$ or $\mathbb{F}_q(t)$), but they can be defined in a more intrinsic way. In later lectures we will see that global fields can also be defined in a more intrinsic way, as fields whose completions are local fields and which admit a suitable product formula.

9.1 Local fields

Definition 9.1. A *local field* is a field with a nontrivial absolute value $|\ |$ that is locally compact under the topology induced by $|\ |$.

Recall that a topological space is *locally compact* if every point has a compact neighborhood.[1] The topology induced by $|\ |$ is given by the metric $d(x,y) := |x-y|$. A metric space is locally compact if and only if every point lies in a compact closed ball.

Example 9.2. Under the standard archimedean absolute value both $\mathbb{R}$ and $\mathbb{C}$ are local fields but $\mathbb{Q}$ is not. Indeed no closed ball in $\mathbb{Q}$ is compact, since it is missing limit points (all irrational real numbers), and in a metric space a compact set must contains all its limit points. Finite fields are not local fields because they have no nontrivial absolute values.

Our first goal is to classify local fields by showing that they are precisely the fields we get by completing a global field. As in the previous lecture, we use $B_{<r}(x) := \{y : |y-x| < r\}$ to denote the open ball of radius $r \in \mathbb{R}_{>0}$ about x, and $B_{\le r}(x) := \{y : |y-x| \le r\}$ to denote a closed ball. Open balls are always open sets and closed balls are always closed sets, but in a nonarchimedean metric space, open balls are both open and closed, as are closed balls.

Remark 9.3. For nonarchimedean metric spaces whose metric is induced by a discrete valuation, every open ball of radius r is also a closed ball of some radius $s \le r$, but we need not have $s = r$; in particular, the closure of $B_{<r}(x)$ (which is already closed) need not be equal to $B_{\le r}(x)$, it could be strictly contained in $B_{\le r}(x)$. The key point is that not every $r \in \mathbb{R}_{\ge 0}$ actually arises as a distance, only countably many do.

Lemma 9.4. *Let K be a field with a nontrivial absolute value $|\ |$. Then K is a local field if and only if every (equivalently, any) closed ball in K is compact.*

Proof. Suppose K is a local field. Then $0 \in K$ lies in a compact closed ball $B_{\le s}(0)$. Let us fix $\alpha \in K^\times$ with $|\alpha| > 1$ (such an α exists because $|\ |$ is nontrivial). The map $x \mapsto \alpha x$ is continuous and $|\ |$ is multiplicative, so $B_{\le |\alpha|^n s}(0)$ is compact for every $n \in \mathbb{Z}_{>0}$ (recall that the continuous image of a compact set is compact). We thus have compact balls about 0 of arbitrarily large radii, implying that every closed ball $B_{\le r}(0)$ is a closed subset of a compact set, hence compact. For every $z \in K$ the translation map $x \mapsto x + z$ is continuous, so every closed ball $B_{\le r}(z)$ is compact. This proves the forward implication, and the reverse implication follows immediately from the definition of local compactness. For the parenthetical, replace $B_{\le s}(0)$ in the argument above by any closed ball. □

Corollary 9.5. *Let K be a local field with absolute value $|\ |$. Then K is complete.*

[1] Weaker definitions of locally compact are sometimes used, but they all imply this one, and for Hausdorff spaces these weaker definitions are all equivalent to the one given here.

Proof. Suppose not. Then there is a Cauchy sequence (x_n) in K that converges to a limit $x \in \widehat{K} - K$. Pick $N \in \mathbb{Z}_{>0}$ so that $|x_n - x| < 1/2$ for all $n \geq N$ (here we are using the extension of $|\ |$ to $\widehat{K}$), and consider the closed ball $S := B_{\leq 1}(x_N)$ in K, which is compact by Lemma 9.4. The Cauchy sequence $(x_n)_{n \geq N}$ in S has a convergent subsequence whose limit lies in $S \subseteq K$, since S is compact and therefore sequentially compact (because K is a metric space). But this limit must be equal to $x \notin K$, a contradiction. □

Proposition 9.6. *Let K be a field with absolute value $|\ |_v$ induced by a discrete valuation v with valuation ring A and uniformizer π. Then K is a local field if and only if K is complete and the residue field $A/\pi A$ is finite.*

Proof. If K is a local field then K is complete, by Corollary 9.5, and the valuation ring

$$A = \{x \in K : v(x) \geq 0\} = \{x \in K : |x|_v \leq 1\} = B_{\leq 1}(0)$$

is a closed ball, hence compact, by Lemma 9.4. The cosets $x + \pi A$ of the subgroup $\pi A \subseteq A$ are open balls $B_{<1}(x)$, since $y \in x + \pi A$ if and only if $|x - y|_v \leq |\pi|_v < 1$. The collection $\{x + \pi A : x \in A\}$ of cosets of πA is an open cover of A by disjoint sets which must be finite, since A is compact; thus $A/\pi A$ is finite.

Now suppose that K is complete and $A/\pi A$ is finite. The valuation ring $A \subseteq K$ is also complete, and Proposition 8.11 gives an isomorphism of topological rings

$$A = \hat{A} \simeq \varprojlim_n \frac{A}{\pi^n A}.$$

Each quotient $A/\pi^n A$ is finite, since $A/\pi A$ is finite, and therefore compact; it follows that the inverse limit, and therefore A, is compact, by Proposition 8.10. Lemma 9.4 implies that K is a local field, since it contains a compact closed ball $B_{\leq 1}(0) = A$ and $|\ |_v$ is nontrivial (recall that discrete valuations surject onto $\mathbb{Z}$ and are thus non-trivial by definition). □

Corollary 9.7. *Let L be a global field with a nontrivial absolute value $|\ |_v$. Then the completion L_v of L with respect to $|\ |_v$ is a local field.*

Proof. Let L/K be a finite extension with $K = \mathbb{Q}$ or $K = \mathbb{F}_q(t)$ and $A = \mathbb{Z}$ or $A = \mathbb{F}_q[t]$, so that $K = \operatorname{Frac} A$. Then A is a Dedekind domain, as is its integral closure B in L, by Theorem 5.25 (and Remark 5.26 in the case that L/K is inseparable).[2]

If $|\ |_v$ is archimedean, then $K = \mathbb{Q}$ and the completion of L with respect to $|\ |_v$ must contain the completion of $\mathbb{Q}$ with respect to the restriction of $|\ |_v$ to $\mathbb{Q}$, which must be isomorphic to $\mathbb{R}$ (as shown on Problem Set 1, every archimedean absolute value on $\mathbb{Q}$ is equivalent to the usual Euclidean absolute value). Thus L_v is a finite extension of $\mathbb{R}$ and must be isomorphic to either $\mathbb{R}$ or $\mathbb{C}$ (as a topological field), both of which are local fields.

We now assume that $|\ |_v$ is nonarchimedean. We claim that in this case $|\ |_v$ is induced by a discrete valuation. Let $C := \{x \in L : |x|_v \leq 1\}$ be the valuation ring of L with respect to $|\ |_v$, and let $\mathfrak{m} := \{x \in L : |x|_v < 1\}$ be its maximal ideal, which is nonzero because $|\ |_v$ is nontrivial. The restriction of $|\ |_v$ to K is a nonarchimedean absolute value, and from the classification of absolute values on $\mathbb{Q}$ and $\mathbb{F}_q(t)$ proved on Problem Set 1, we can assume it is induced by a discrete valuation on A; in particular, $|x|_v \leq 1$ for all $x \in A$, and therefore

[2] In fact, we can always choose K so that L/K is separable: if L has positive characteristic p, let $\mathbb{F}_q$ be the algebraic closure of $\mathbb{F}_p$ in L, choose a separating transcendental element t, and put $K := \mathbb{F}_q(t)$. Such a t exists because $\mathbb{F}_q$ is perfect and $L/\mathbb{F}_q$ is finitely generated, see [3, Thm. 7.20].

$A \subseteq C$. Like all valuation rings (discrete or not), C is integrally closed in its fraction field L, and C contains A, so C contains B, since B is the integral closure of A in L. The ideal $\mathfrak{q} = \mathfrak{m} \cap B$ is maximal, and the DVR $B_{\mathfrak{q}}$ lies in C and must equal C, since there are no intermediate rings between a DVR and its fraction field (we cannot have $C = L$ because C is not a field). It follows that the absolute value induced by $v_{\mathfrak{q}}$ is equivalent to $|\ |_v$, since they have the same valuation rings. By choosing $0 < c < 1$ appropriately, we can assume that $|\cdot|_v = c^{v_{\mathfrak{q}}(\cdot)}$ is induced by $v_{\mathfrak{q}}$, which proves the claim.

The residue field $B_{\mathfrak{q}}/\mathfrak{q}B_{\mathfrak{q}} \simeq B/\mathfrak{q}$ is finite, since $B/\mathfrak{q}$ is a finite extension of the finite field $A/\mathfrak{p}$, where $\mathfrak{p} = \mathfrak{q} \cap A$. If we now consider the completion L_v with valuation ring B_v, we can take any uniformizer π for $\mathfrak{q} \subseteq B \subseteq B_v$ as a uniformizer for B_v, and we have

$$\frac{B}{\mathfrak{q}} \simeq \frac{B_{\mathfrak{q}}}{\mathfrak{q}B_{\mathfrak{q}}} = \frac{B_{\mathfrak{q}}}{\pi B_{\mathfrak{q}}} \simeq \frac{B_v}{\pi B_v},$$

so $B_v/\pi B_v$ is finite. Thus L_v is a complete field with an absolute value induced by a discrete valuation and finite residue field, and therefore a local field, by Proposition 9.6. □

In order to classify all local fields we require the following result from topology (here nondiscrete simply means that not every set is open).

Proposition 9.8. *A locally compact topological vector space over a nondiscrete locally compact field has finite dimension.*

Proof. See [4, Prop. 4-13.iv]. □

Theorem 9.9. *Let L be a local field. If L is archimedean then it is isomorphic to $\mathbb{R}$ or $\mathbb{C}$; otherwise, L is isomorphic to a finite extension of $\mathbb{Q}_p$ or $\mathbb{F}_q((t))$.*

Proof. Let L be a local field with nontrivial absolute value $|\ |$; then L is complete, by Corollary 9.5. If L has characteristic zero then the prime field of L is $\mathbb{Q}$, and L contains the completion of $\mathbb{Q}$ with respect to the restriction of $|\ |$ to $\mathbb{Q}$. By Ostrowski's theorem, the restriction of $|\ |$ to $\mathbb{Q}$ is equivalent to either the standard archimedean absolute value, in which case the completion is $\mathbb{R}$, or it is equivalent to a p-adic absolute value, in which case the completion is $\mathbb{Q}_p$ (which, by definition, is the completion of $\mathbb{Q}$ with respect to the p-adic absolute value). Thus L contains a subfield K isomorphic to $\mathbb{R}$ or to $\mathbb{Q}_p$ for some prime p.

If L has positive characteristic p then the prime field of L is $\mathbb{F}_p$, and L must contain a transcendental element s, since no algebraic extension of $\mathbb{F}_p$ has a nontrivial absolute value (if $|\alpha| > 1$ for some algebraic $\alpha \in L$, then the restriction of $|\ |$ to the finite field $\mathbb{F}_p(\alpha)$ is nontrivial, but this is impossible). It follows that L contains $\mathbb{F}_p(s)$ and therefore contains the completion of $\mathbb{F}_p(s)$ with respect to $|\ |$. Every completion of $\mathbb{F}_p(s)$ is isomorphic to $\mathbb{F}_q((t))$ for some q a power of p and t transcendental over $\mathbb{F}_q$ (see Problem Set 5). Thus L contains a subfield K isomorphic to $\mathbb{F}_q((t))$.

If K is archimedean then $K = \mathbb{R}$ is a local field, and if K is nonarchimedean then $K = \mathbb{Q}_p$ or $K = \mathbb{F}_q((t))$ is a complete field with a discrete valuation and finite residue field, hence a local field by Proposition 9.6. The field K is therefore locally compact, and it is nondiscrete because its absolute value is nontrivial. Proposition 9.8 implies that L has finite degree over K. If K is archimedean then $K = \mathbb{R}$, and L must be $\mathbb{R}$ or $\mathbb{C}$; otherwise, L is a finite extension of $\mathbb{Q}_p$ or $\mathbb{F}_q((t))$ as claimed. □

9.2 Hensel's lemmas

Definition 9.10. Let R be a (commutative) ring, and let $f(x) = \sum f_i x^i \in R[x]$ be a polynomial. The (*formal*) *derivative* f' of f is the polynomial $f'(x) := \sum i f_i x^{i-1} \in R[x]$.

Note that the canonical ring homomorphism $\mathbb{Z} \to R$ defined by $1 \mapsto 1$ allows us to view the integers $i = 1+1+\cdots+1$ as elements of R (the map $\mathbb{Z} \to R$ will be injective only when R has characteristic zero, but it is well defined in any case). It is easy to verify that for all $a, b \in R$ and $f, g \in R[x]$ the formal derivative satisfies the usual identities:

$$\begin{aligned}(af+bg)' &= af'+bg', && \text{(linearity)}\\ (fg)' &= f'g+fg', && \text{(Leibniz rule)}\\ (f\circ g)' &= (f'\circ g)g', && \text{(chain rule)}\end{aligned}$$

When the characteristic of R is positive, we may have $\deg f' < \deg f - 1$. Indeed, if R has characteristic $p > 0$ and $g(x) = f(x^p)$ for some $f \in R[x]$, then $g' = f'(x^p)px^{p-1} = 0$.

Lemma 9.11. *Let R be a ring, let $f = \sum f_i x^i \in R[x]$ be a polynomial, and let $a \in R$. Then $f(x) = f(a) + f'(a)(x-a) + g(x)(x-a)^2$ for a unique $g \in R[x]$.*

Proof. We have

$$\begin{aligned}f(x) = f(a+(x-a)) &= \sum_{i\ge 0} f_i (a+(x-a))^i = \sum_{i\ge 0} f_i \sum_{0\le j\le i} \binom{i}{j} a^j (x-a)^{i-j}\\ &= f(a) + \sum_{i\ge 1} f_i \sum_{0\le j< i} \binom{i}{j} a^j (x-a)^{i-j}\\ &= f(a) + f'(a)(x-a) + \sum_{i\ge 2} f_i \sum_{0\le j\le i-2} \binom{i}{j} a^j (x-a)^{i-j}\\ &= f(a) + f'(a)(x-a) + g(x)(x-a)^2,\end{aligned}$$

where $g(x) = \sum_{i\ge 2} f_i \sum_{0\le j\le i-2} \binom{i}{j} a^j (x-a)^{i-2-j} \in R[x]$. □

Remark 9.12. The lemma can be viewed as giving the first two terms of a formal Taylor expansion of $f(x)$ about a. Note that the binomial coefficients $\binom{i}{j}$ are integers, hence well defined elements of R under the canonical homomorphism $\mathbb{Z} \to R$, even when $j!$ is divisible by the characteristic of R. In the usual Taylor expansion

$$f(x) = \sum_{i=0}^{\infty} \frac{f^{(i)}(a)}{i!}(x-a)^i$$

used in characteristic zero, if f is a polynomial then $f^{(i)}(a)$ is necessarily a multiple of $i!$, so $f^{(i)}(a)/i!$ is always a well defined element of R, even in positive characteristic.

Corollary 9.13. *Let R be a ring, $f \in R[x]$, and $a \in R$. Then $f(a) = f'(a) = 0$ if and only if a is (at least) a double root of f, that is, $f(x) = (x-a)^2 g(x)$ for some $g \in R[x]$.*

Definition 9.14. Let $f \in R[x]$ be a polynomial over a ring R and let $a \in R$. If $f(a) = 0$ and $f'(a) \neq 0$ then a is a *simple root* of f.

If R is a ring and I is an R-ideal, by a *lift* of an element $\bar{r}$ of the quotient R/I, we mean a preimage of $\bar{r}$ under the quotient map $R \twoheadrightarrow R/I$.

Lemma 9.15 (Hensel's Lemma I)**.** *Let A be a complete DVR with maximal ideal $\mathfrak{p}$ and residue field $k := A/\mathfrak{p}$. Suppose $f \in A[x]$ is a monic polynomial whose reduction to $k[x]$ has a simple root $\bar{a} \in k$. Then $\bar{a}$ can be lifted to a root of f in A.*

Proof. We work in the fraction field K of A. Let a_0 be any lift of $\bar{a}$ to A; the element a_0 is not necessarily a root of f, but it is a root modulo $\mathfrak{p}$. We will show that a_0 is the first term of a Cauchy sequence (a_n) in which each a_n is a root of f modulo $\mathfrak{p}^{2^n}$. To simplify the notation we fix $0 < c < 1$ and define the absolute value $|\cdot| := c^{v_{\mathfrak{p}}(\cdot)}$. The fact that $\bar{a}$ is a simple root implies that $f(a_0) \in \mathfrak{p}$ but $f'(a_0) \notin \mathfrak{p}$, so $|f(a_0)| \le c < 1$ and $|f'(a_0)| = 1$. We now define

$$\epsilon := \frac{|f(a_0)|}{|f'(a_0)|^2} < 1.$$

In what follows we will only use $\epsilon < 1$, which will allow our proof to work in cases where $\bar{a}$ is not necessarily a simple root (in particular, we won't assume $|f'(a_0)| = 1$).

For each $n \ge 0$ we define

$$a_{n+1} := a_n - f(a_n)/f'(a_n).$$

We will prove by induction on n that

- **(a)** $|a_n| \le 1$ ($a_n \in A$);
- **(b)** $|a_n - a_0| \le \epsilon < 1$ ($a_n \equiv a_0 \bmod \mathfrak{p}$, so a_n is a lift of $\bar{a}$);
- **(c)** $|f'(a_n)| = |f'(a_0)|$ (with (d) this ensures $f'(a_n) | f(a_n)$, so a_{n+1} is well defined);
- **(d)** $|f(a_n)| \le \epsilon^{2^n} |f'(a_0)|^2$ ($|f(a_n)|$ and therefore $f(a_n)$ converges rapidly to 0).

The case $n = 0$ is clear. We now assume (a), (b), (c), (d) for n and prove them for $n+1$:

(a) $|a_{n+1} - a_n| = |f(a_n)/f'(a_n)| \le \epsilon^{2^n} |f'(a_0)|^2 / |f'(a_0)| = \epsilon^{2^n} |f'(a_0)| \le \epsilon^{2^n}$, by (c) and (d), therefore $|a_{n+1}| = |a_{n+1} - a_n + a_n| \le \max(|a_{n+1} - a_n|, |a_n|) \le 1$, by (a).

(b) $|a_{n+1} - a_0| \le \max(|a_{n+1} - a_n|, |a_n - a_0|) \le \max(\epsilon^{2^n}, \epsilon) = \epsilon$ (as above and using (b)).

(c) Applying Lemma 9.11 to $f'(x)$ at a_n and substituting a_{n+1} for x yields

$$f'(a_{n+1}) = f'(a_n) - f''(a_n)\frac{f(a_n)}{f'(a_n)} + g(a_{n+1})\left(\frac{f(a_n)}{f'(a_n)}\right)^2,$$

where we have used $a_{n+1} - a_n = -f(a_n)/f'(a_n)$. We have $f''(a_n), g(a_{n+1}) \in A$, so $|f''(a_n)|, |g(a_{n+1})| \le 1$, and $|f(a_n)/f'(a_n)| = |f(a_n)|/|f'(a_0)| \le \epsilon^{2^n} |f'(a_0)|$, by (d), so the absolute values of the last two terms on the RHS are strictly smaller than the first term $|f'(a_n)| = |f'(a_0)|$. Therefore $|f'(a_{n+1})| = |f'(a_n)| = |f'(a_0)|$.

(d) Applying Lemma 9.11 to $f(x)$ and substituting a_{n+1} for x yields

$$f(a_{n+1}) = f(a_n) - f'(a_n)\frac{f(a_n)}{f'(a_n)} + h(a_{n+1})\left(\frac{f(a_n)}{f'(a_n)}\right)^2 = h(a_{n+1})\left(\frac{f(a_n)}{f'(a_n)}\right)^2,$$

for some $h \in A[x]$. We have $|h(a_{n+1})| \le 1$, so (c) and (d) imply

$$|f(a_{n+1})| \le |f(a_n)|^2 / |f'(a_n)|^2 = |f(a_n)|^2 / |f'(a_0)|^2 \le \epsilon^{2^{n+1}} |f'(a_0)|^2.$$

which completes our inductive proof.

We have $|a_{n+1} - a_n| \le \epsilon^{2^n} \to 0$ as $n \to \infty$, and for a nonarchimedean absolute value this implies that (a_n) is Cauchy. Thus $a := \lim_{n\to\infty} a_n \in A$, since A is complete. We have $f(a) = \lim_{n\to\infty} f(a_n) = 0$, so a is a root of f, and $|a - a_0| = \lim_{n\to\infty} |a_n - a_0| < 1$, so $a \equiv a_0 \bmod \mathfrak{p}$ is a lift of $\bar{a}$. □

Our proof of Lemma 9.15 did not actually use the assumption that f is monic, nor did it actually require $\bar{a}$ to be a simple root. Let us record the (apparently stronger) form of Hensel's lemma that what we actually proved.

Lemma 9.16 (Hensel's Lemma II)**.** *Let A be a complete DVR. Let $f \in A[x]$, and suppose $a_0 \in A$ satisfies*

$$|f(a_0)| < |f'(a_0)|^2$$

(so in particular, $f'(a_0)$ divides $f(a_0)$), and for $n \ge 0$ define

$$a_{n+1} := a_n - f(a_n)/f'(a_n).$$

The sequence (a_n) is well-defined and converges to the unique root $a \in A$ of f for which

$$|a - a_0| \le \epsilon := |f(a_0)|/|f'(a_0)|^2.$$

Moreover, $|f(a_n)| \le \epsilon^{2n}|f'(a_0)|^2$ for all $n \ge 0$.

Lemma 9.16 can be viewed as a nonarchimedean version of Newton's method for finding (or more closely approximating) a root of a polynomial given an initial approximation. Like Newton's method, the recurrence in Lemma 9.16 converges quadratically, meaning that we double the precision of our approximation with each iteration. But Lemma 9.16 is better than Newton's method, for two reasons: (1) in the most common scenario the residue field is finite, which makes finding an initial approximation very easy, and (2) once we have an initial approximation with $\epsilon < 1$, convergence is guaranteed.

Remark 9.17. In Lemmas 9.15 and 9.16 it is not actually necessary for A to be complete (or an integral domain). A local ring A for which Lemma 9.15 holds is called a *henselian ring* (this is a definition). One can show that Lemma 9.16 necessarily also holds in any henselian ring, as do many other forms of "Hensel's Lemma", including Lemma 9.19 below. In general, any condition that holds for a local ring if and only if it is a henselian ring may be called "Hensel's Lemma"; see [5, Lemma 10.148.3] for more than a dozen candidates. One can define the *henselization* of a noetherian local ring R as the minimal extension of R that is henselian (as usual, it is minimal in the sense of satisfying a universal property, and this forces it to be unique up to isomorphism). When R is a DVR its henselization is simply $\widehat{R} \cap K^{\text{sep}}$, where K is the fraction field of R. Loosely speaking, in henselian rings, Cauchy sequences that converge (in the completion) to the root of a polynomial are required to converge, but not every Cauchy sequence needs to converge.

Example 9.18. Let $A = \mathbb{Z}_5$ and $f(x) = x^2 - 6 \in \mathbb{Z}_5[x]$. Then $\bar{f}(x) = x^2 - 1 \in \mathbb{F}_5[x]$ has $\bar{a} = 1$ as a simple root. By Lemma 9.15 there is a unique $a \in \mathbb{Z}_5$ such that $a^2 - 6 = 0$ and $a \equiv 1 \bmod 5$. We could also have chosen $\bar{a} = -1$, which would give another distinct root of $f(x)$, which must be $-a$. Thus $\mathbb{Z}_5$ contains two distinct square roots of 6.

Now let $A = \mathbb{Z}_2$ and $f(x) = x^2 - 17$. Then $\bar{f}(x) = x^2 - 1 = (x-1)^2$ has no simple roots (note $\bar{f}' = 0$). But if we let $a_0 = 1$, then $f(a_0) = -16$ and $|f(a_0)| = 1/16$, while $f'(a_0) = 2$ and $|f'(a_0)| = 1/2$. We thus have $|f(a_0)| < |f'(a_0)|^2$ and can apply Lemma 9.16 to get a square root of 17 in $\mathbb{Z}_2$.

There is a another version of Hensel's Lemma that we need, which is arguably the most powerful (of course they are all equivalent by definition, but this version is most easily seen to imply all the others).

Lemma 9.19 (Hensel's lemma III)**.** *Let A be a complete DVR with maximal ideal $\mathfrak{p}$ and residue field k, let $f \in A[x]$ have image $\bar{f}$ in $k[x]$, and suppose $\bar{f} = \bar{g}\bar{h}$ for some coprime $\bar{g}, \bar{h} \in k[x]$. Then there exist polynomials $g, h \in A[x]$ for which $f = gh$ with $g \equiv \bar{g} \bmod \mathfrak{p}$ and $h \equiv \bar{h} \bmod \mathfrak{p}$ such that $\deg g = \deg \bar{g}$.*

Proof. See [2, Theorem II.4.6] or [5, Lemma 10.148.3]. □

This form of Hensel's lemma has the following useful corollary, which is itself another form Hensel's lemma in the sense that it characterizes henselian fields (see Remark 9.17).[3]

Lemma 9.20 (Hensel-Kürschák lemma)**.** *Let A be a complete DVR with fraction field K, and let $f \in K[x]$ be an irreducible polynomial whose leading and constant coefficients lie in A. Then $f \in A[x]$.*

Proof. Let $\mathfrak{p} = (\pi)$ be the maximal ideal of A, let $k := A/\mathfrak{p}$, and write $f = \sum_{i=0}^{n} f_i x^i$ with $f_n \neq 0$. We must have $n > 0$ and $f_0 \neq 0$, since f is irreducible. Let $m := \min\{v_{\mathfrak{p}}(f_i)\}$. Suppose for the sake of contradiction that $m < 0$, and let $g := \pi^{-m} f = \sum_{i=0}^{n} g_i x^i \in A[x]$. Then g is an irreducible polynomial in $A[x]$ with $g_0, g_n \in \mathfrak{p}$, since $m < 0$ and $f_0, f_n \in A$, and g_i is a unit for some $0 < i < n$, by the minimality of m. The reduction $\bar{g}$ of g to $k[x]$ has positive degree and constant term 0, and is therefore divisible by x. If we let $\bar{u} := x^d$ be the largest power of x dividing $\bar{g}$, then $0 < d \leq \deg \bar{g} < n$ and $\bar{v} := \bar{g}/x^d \in k[x]$ is coprime to $\bar{u}$ (possibly $\deg \bar{v} = 0$). Lemma 9.19 implies that $g = uv$ for some $u, v \in A[x]$ with $0 < \deg u = \deg \bar{u} < n$. But this means g is not irreducible, a contradiction. □

Corollary 9.21. *Let A be a complete DVR with fraction field K, and let L/K be a finite extension of degree n. Then $\alpha \in L$ is integral over A if and only if $\mathrm{N}_{L/K}(\alpha) \in A$.*

Proof. Let $f = \sum_{i=0}^{d} f_i x^i \in K[x]$ be the minimal polynomial of α. If α is integral over A then $f \in A[x]$, by Proposition 1.28, and $N_{L/K}(\alpha) = (-1)^n f(0)^e \in A$, where $e = [L : K(\alpha)]$, by Proposition 4.51. Conversely, if $N_{L/K}(\alpha) = (-1)^n f(0)^e \in A$, then $f(0) \in A$, since $f(0) \in K$ is a root of $x^e - (-1)^n N_{L/K}(\alpha) \in A[x]$ and A is integrally closed. The constant coefficient of f thus lies in A, as does its leading coefficient (it is monic), so $f \in A[x]$, by Lemma 9.20, and α is therefore integral over A. □

References

[1] N. Bourbaki, *General Topology: Chapters 1-4*, Springer, 1985.

[2] J. Neukirch, *Algebraic number theory*, Springer, 1999.

[3] A. W. Knapp, *Advanced algebra*, Birkhauser, 2008.

[4] D. Ramakrishnan and R.J. Valenza, *Fourier analysis on number fields*, Springer, 1999.

[5] Stacks Project Authors, *Stacks Project*, http://stacks.math.columbia.edu.

[3] See [2, §II.6] for a proof of this.

10 Extensions of complete DVRs

Recall that in our *AKLB* setup, A is a Dedekind domain with fraction field K, the field L is a finite separable extension of K, and B is the integral closure of A in L; as we proved in Theorem 5.25, this implies that B is also a Dedekind domain (with L as its fraction field). We now want to consider the special case where A is a complete DVR; in this case B is also a complete DVR, but this will take a little bit of work to prove. We first show that B is a DVR.

Theorem 10.1. *Assume AKLB and that A is a complete DVR with maximal ideal $\mathfrak{p}$. Then B is a DVR whose maximal ideal $\mathfrak{q}$ is necessarily the unique prime above $\mathfrak{p}$.*

Proof. We first show that $\#\{\mathfrak{q}|\mathfrak{p}\} = 1$. At least one prime $\mathfrak{q}$ of B lies above $\mathfrak{p}$, since the factorization of $\mathfrak{p}B \subsetneq B$ is non-trivial. Now suppose for the sake of contradiction that $\mathfrak{q}_1, \mathfrak{q}_2 \in \{\mathfrak{q}|\mathfrak{p}\}$ with $\mathfrak{q}_1 \neq \mathfrak{q}_2$. Choose $b \in \mathfrak{q}_1 - \mathfrak{q}_2$ and consider the ring $A[b] \subseteq B$. The ideals $\mathfrak{q}_1 \cap A[b]$ and $\mathfrak{q}_2 \cap A[b]$ are distinct prime ideals of $A[b]$ containing $\mathfrak{p}A[b]$, and both are maximal, since they are nonzero and $\dim A[b] = \dim A = 1$ (note that $A[b]$ is integral over A and therefore has the same dimension). The quotient ring $A[b]/\mathfrak{p}A[b]$ thus has at least two maximal ideals. Let $f \in A[x]$ be the minimal polynomial of b over K, and let $\bar{f} \in (A/\mathfrak{p})[x]$ be its reduction to the residue field $A/\mathfrak{p}$.

$$\frac{(A/\mathfrak{p})[x]}{(\bar{f})} \simeq \frac{A[x]}{(\mathfrak{p}, f)} \simeq \frac{A[b]}{\mathfrak{p}A[b]},$$

thus the ring $(A/\mathfrak{p})[x]/(\bar{f})$ has at least two maximal ideals, which implies that $\bar{f}$ is divisible by two distinct irreducible polynomials (because $(A/\mathfrak{p})[x]$ is a PID). We can thus factor $\bar{f} = \bar{g}\bar{h}$ with $\bar{g}$ and $\bar{h}$ coprime. By Hensel's Lemma 9.19, we can lift this to a non-trivial factorization $f = gh$ of f in $A[x]$, contradicting the irreducibility of f.

Every maximal ideal of B lies above a maximal ideal of A, but A has only the maximal ideal $\mathfrak{p}$ and $\#\{\mathfrak{q}|\mathfrak{p}\} = 1$, so B has a unique (nonzero) maximal ideal $\mathfrak{q}$. Thus B is a local Dedekind domain, hence a local PID, and not a field, so B is a DVR, by Theorem 1.16. □

Remark 10.2. The assumption that A is complete is necessary. For example, if A is the DVR $\mathbb{Z}_{(5)}$ with fraction field $K = \mathbb{Q}$ and we take $L = \mathbb{Q}(i)$, then the integral closure of A in L is $B = \mathbb{Z}_{(5)}[i]$, which is a PID but not a DVR: the ideals $(1+2i)$ and $(1-2i)$ are both maximal (and not equal). But if we take completions we get $A = \mathbb{Z}_5$ and $K = \mathbb{Q}_5$, and now $L = \mathbb{Q}_5(i) = \mathbb{Q}_5 = K$, since $x^2 + 1$ has a root in $\mathbb{F}_5 \simeq \mathbb{Z}_5/5\mathbb{Z}_5$ that we can lift to $\mathbb{Z}_5$ via Hensel's lemma; thus if we complete A then $B = A$ is a DVR as required.

Definition 10.3. Let K be a field with absolute value $|\ |$ and let V be a K-vector space. A *norm* on V is a function $\|\ \| : V \to \mathbb{R}_{\geq 0}$ such that

- $\|v\| = 0$ if and only if $v = 0$.
- $\|\lambda v\| = |\lambda|\,\|v\|$ for all $\lambda \in K$ and $v \in V$.
- $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$.

Each norm $\|\ \|$ induces a topology on V via the distance metric $d(v, w) := \|v - w\|$.

Example 10.4. Let V be a K-vector space with basis (e_i), and for $v \in V$ let $v_i \in K$ denote the coefficient of e_i in $v = \sum_i v_i e_i$. The *sup-norm* $\|v\|_\infty := \sup\{|v_i|\}$ is a norm on V (thus

every vector space has at least one norm). If V is also a K-algebra, an absolute value $\|\ \|$ on V (as a ring) is a norm on V (as a K-vector space) if and only if it extends the absolute value on K (fix $v \neq 0$ and note that $\|\lambda\|\,\|v\| = \|\lambda v\| = |\lambda|\,\|v\| \Leftrightarrow \|\lambda\| = |\lambda|$).

Proposition 10.5. *Let V be a vector space of finite dimension over a complete field K. Every norm on V induces the same topology, in which V is a complete metric space.*

Proof. See Problem Set 5. □

Theorem 10.6. *Let A be a complete DVR with fraction field K, maximal ideal $\mathfrak{p}$, discrete valuation $v_\mathfrak{p}$, and absolute value $|x|_\mathfrak{p} := c^{v_\mathfrak{p}(x)}$, with $0 < c < 1$. Let L/K be a finite extension of degree n. The following hold.*

(i) *There is a unique absolute value $|x| := |\mathrm{N}_{L/K}(x)|_\mathfrak{p}^{1/n}$ on L that extends $|\ |_\mathfrak{p}$;*

(ii) *The field L is complete with respect to $|\ |$, and its valuation ring $\{x \in L : |x| \leq 1\}$ is equal to the integral closure B of A in L;*

(iii) *If L/K is separable then B is a complete DVR whose maximal ideal $\mathfrak{q}$ induces*

$$|x| = |x|_\mathfrak{q} := c^{\frac{1}{e_\mathfrak{q}} v_\mathfrak{q}(x)},$$

where $e_\mathfrak{q}$ is the ramification index of $\mathfrak{q}$, that is, $\mathfrak{p}B = \mathfrak{q}^{e_\mathfrak{q}}$.

Proof. Assuming for the moment that $|\ |$ is actually an absolute value (which is not obvious!), for any $x \in K$ we have

$$|x| = |\mathrm{N}_{L/K}(x)|_\mathfrak{p}^{1/n} = |x^n|_\mathfrak{p}^{1/n} = |x|_\mathfrak{p},$$

so $|\ |$ extends $|\ |_\mathfrak{p}$ and is therefore a norm on L. The fact that $|\ |_\mathfrak{p}$ is nontrivial means that $|x|_\mathfrak{p} \neq 1$ for some $x \in K^\times$, and $|x|^a = |x|_\mathfrak{p} = |x|$ only for $a = 1$, which implies that $|\ |$ is the unique absolute value in its equivalence class extending $|\ |_\mathfrak{p}$. Every norm on L induces the same topology (by Proposition 10.5), so $|\ |$ is the only absolute value on L that extends $|\ |_\mathfrak{p}$.

We now show $|\ |$ is an absolute value. Clearly $|x| = 0 \Leftrightarrow x = 0$ and $|\ |$ is multiplicative; we only need to check the triangle inequality. It suffices to show $|x| \leq 1 \Rightarrow |x+1| \leq |x| + 1$, since we always have $|y+z| = |z||y/z+1|$ and $|y| + |z| = |z|(|y/z| + 1)$, and without loss of generality we assume $|y| \leq |z|$. In fact the stronger implication $|x| \leq 1 \Rightarrow |x+1| \leq 1$ holds:

$$|x| \leq 1 \iff |\mathrm{N}_{L/K}(x)|_\mathfrak{p} \leq 1 \iff N_{L/K}(x) \in A \iff x \in B \iff x{+}1 \in B \iff |x{+}1| \leq 1.$$

The first biconditional follows from the definition of $|\ |$, the second follows from the definition of $|\ |_\mathfrak{p}$, the third is Corollary 9.21, the fourth is obvious, and the fifth follows from the first three after replacing x with $x + 1$. This completes the proof of (i), and also proves (ii).

We now assume L/K is separable. Then B is a DVR, by Theorem 10.1, and it is complete because it is the valuation ring of L. Let $\mathfrak{q}$ be the unique maximal ideal of B. The valuation $v_\mathfrak{q}$ extends $v_\mathfrak{p}$ with index $e_\mathfrak{q}$, by Theorem 8.20, so $v_\mathfrak{q}(x) = e_\mathfrak{q} v_\mathfrak{p}(x)$ for $x \in K^\times$. We have $0 < c^{1/e_\mathfrak{q}} < 1$, so $|x|_\mathfrak{q} := (c^{1/e_\mathfrak{q}})^{v_\mathfrak{q}(x)}$ is an absolute value on L induced by $v_\mathfrak{q}$. To show it is equal to $|\ |$, it suffices to show that it extends $|\ |_\mathfrak{p}$, since we already know that $|\ |$ is the unique absolute value on L with this property. For $x \in K^\times$ we have

$$|x|_\mathfrak{q} = c^{\frac{1}{e_\mathfrak{q}} v_\mathfrak{q}(x)} = c^{\frac{1}{e_\mathfrak{q}} e_\mathfrak{q} v_\mathfrak{p}(x)} = c^{v_\mathfrak{p}(x)} = |x|_\mathfrak{p},$$

and the theorem follows. □

Remark 10.7. The transitivity of $\mathrm{N}_{L/K}$ in towers (Corollary 4.52) implies that we can uniquely extend the absolute value on the fraction field K of a complete DVR to an algebraic closure $\overline{K}$. In fact, this is another form of Hensel's lemma in the following sense: one can show that a (not necessarily discrete) valuation ring A is Henselian if and only if the absolute value of its fraction field K can be uniquely extended to $\overline{K}$; see [4, Theorem 6.6].

Corollary 10.8. *Assume AKLB and that A is a complete DVR with maximal ideal $\mathfrak{p}$ and let $\mathfrak{q}|\mathfrak{p}$. Then $v_{\mathfrak{q}}(x) = \frac{1}{f_{\mathfrak{q}}} v_{\mathfrak{p}}(\mathrm{N}_{L/K}(x))$ for all $x \in L$.*

Proof. $v_{\mathfrak{p}}(\mathrm{N}_{L/K}(x)) = v_{\mathfrak{p}}(\mathrm{N}_{L/K}((x))) = v_{\mathfrak{p}}(\mathrm{N}_{L/K}(\mathfrak{q}^{v_{\mathfrak{q}}(x)})) = v_{\mathfrak{p}}(\mathfrak{p}^{f_{\mathfrak{q}} v_{\mathfrak{q}}(x)}) = f_{\mathfrak{q}} v_{\mathfrak{q}}(x)$. □

Remark 10.9. One can generalize the notion of a discrete valuation to a *valuation*, a surjective homomorphism $v\colon K^\times \to \Gamma$, in which Γ is a (totally) ordered abelian group and $v(x+y) \geq \min(v(x), v(y))$; we extend v to K by defining $v(0) = \infty$ to be strictly greater than any element of Γ. In the $AKLB$ setup with A a complete DVR, one can then define a valuation $v(x) = \frac{1}{e_{\mathfrak{q}}} v_{\mathfrak{q}}(x)$ with image $\frac{1}{e_{\mathfrak{q}}}\mathbb{Z}$ that restricts to the discrete valuation $v_{\mathfrak{p}}$ on K. The valuation v then extends to a valuation on $\overline{K}$ with $\Gamma = \mathbb{Q}$. Some texts take this approach, but we will generally stick with discrete valuations (so our absolute value on L restricts to K, but our discrete valuations on L do not restrict to discrete valuations on K, they extend them with index $e_{\mathfrak{q}}$).

Remark 10.10. Recall that a *valuation ring* is an integral domain A with fraction field K such that for every $x \in K^\times$ either $x \in A$ or $x^{-1} \in A$ (possibly both). As you will show on Problem Set 6, if A is a valuation ring, then there exists a valuation $v\colon K \to \Gamma \cup \{\infty\}$ for some totally ordered abelian group Γ such that $A = \{x \in K : v(x) \geq 0\}$ is the valuation ring of K with respect to this valuation.

10.1 The Dedekind-Kummer theorem in a local setting

Recall that the Dedekind-Kummer theorem (Theorem 6.14) allows us to factor primes in our $AKLB$ setting by factoring polynomials over the residue field, provided that B is monogenic (of the form $A[\alpha]$ for some $\alpha \in B$), or the prime of interest does not contain the conductor. We now show that in the special case where A and B are DVRs and the residue field extension is separable, B is always monogenic; this holds, for example, whenever K is a local field. To prove this, we first recall a form of Nakayama's lemma.

Lemma 10.11 (Nakayama's Lemma)**.** *Let A be a local ring with maximal ideal $\mathfrak{p}$, and let M be a finitely generated A-module. If the images of $x_1, \dots, x_n \in M$ generate $M/\mathfrak{p}M$ as an $(A/\mathfrak{p})$-vector space then $x_1, \dots, x_n$ generate M as an A-module.*

Proof. See [1, Corollary 4.8b]. □

Before proving our theorem on local monogenicity, we record a few corollaries of Nakayama's Lemma that will be useful later.

Corollary 10.12. *Let A be a local noetherian ring with maximal ideal $\mathfrak{p}$, let $g \in A[x]$, and let $B \coloneqq A[x]/(g(x))$. Every maximal ideal $\mathfrak{m}$ of B contains the ideal $\mathfrak{p}B$.*

Proof. Suppose not. Then $\mathfrak{m} + \mathfrak{p}B = B$ for some maximal ideal $\mathfrak{m}$ of B. The ring B is finitely generated over the noetherian ring A, hence a noetherian A-module, so its A-submodules are all finitely generated. Let $z_1, \dots, z_n$ be A-module generators for $\mathfrak{m}$. Every coset of $\mathfrak{p}B$

in B can be written as $z + \mathfrak{p}B$ for some A-linear combination z of $z_1, \dots, z_n$, so the images of $z_1, \dots, z_n$ generate $B/\mathfrak{p}B$ as an $(A/\mathfrak{p})$-vector space. By Nakayama's lemma, $z_1, \dots, z_n$ generate B, in which case $\mathfrak{m} = B$, a contradiction. □

As a corollary, we immediately obtain a local version of the Dedekind-Kummer theorem that does not even require A and B to be Dedekind domains.

Corollary 10.13. *Let A be a local noetherian ring with maximal ideal $\mathfrak{p}$, let $g \in A[x]$ be a polynomial with reduction $\bar{g} \in (A/\mathfrak{p})[x]$, and let α be the image of x in the ring $B \coloneqq A[x]/(g(x)) = A[\alpha]$. The maximal ideals of B are $(\mathfrak{p}, g_i(\alpha))$, where $g_1, \dots, g_m \in A[x]$ are lifts of the distinct irreducible polynomials $\bar{g}_i \in (A/\mathfrak{p})[x]$ that divide $\bar{g}$.*

Proof. By Corollary 10.12, the quotient map $B \to B/\mathfrak{p}B$ gives a one-to-one correspondence between maximal ideals of B and maximal ideals of $B/\mathfrak{p}B$, and we have

$$\frac{B}{\mathfrak{p}B} \simeq \frac{A[x]}{(\mathfrak{p}, g(x))} \simeq \frac{(A/\mathfrak{p})[x]}{(\bar{g}(x))}.$$

Each maximal ideal of $(A/\mathfrak{p})[x]/(\bar{g}(x))$ is the reduction of an irreducible divisor of $\bar{g}$, hence one of the $\bar{g}_i$ (because $(A/\mathfrak{p})[x]$ is a PID). The corollary follows. □

Theorem 10.14. *Assume AKLB, with A and B DVRs with residue fields $k \coloneqq A/\mathfrak{p}$ and $l \coloneqq B/\mathfrak{q}$. If l/k is separable then $B = A[\alpha]$ for some $\alpha \in B$; if L/K is unramified this holds for every lift α of any generator $\bar{\alpha}$ for $l = k(\bar{\alpha})$.*

Proof. Let $\mathfrak{p}B = \mathfrak{q}^e$ be the factorization of $\mathfrak{p}B$ and let $f = [l : k]$ be the residue field degree, so that $ef = n \coloneqq [L : K]$. The extension l/k is separable, so we may apply the primitive element theorem to write $l = k(\alpha_0)$ for some $\alpha_0 \in l$ whose minimal polynomial $\bar{g}$ is separable of degree equal to f. Let $g \in A[x]$ be a monic lift of $\bar{g}$, and let α_0 be any lift of $\bar{\alpha}_0$ to B. If $v_{\mathfrak{q}}(g(\alpha_0)) = 1$ then let $\alpha \coloneqq \alpha_0$. Otherwise, let π_0 be any uniformizer for B and let $\alpha \coloneqq \alpha_0 + \pi_0 \in B$ (so $\alpha \equiv \bar{\alpha}_0 \bmod \mathfrak{q}$) Writing $g(x + \pi_0) = g(x) + \pi_0 g'(x) + \pi_0^2 h(x)$ for some $h \in A[x]$ via Lemma 9.11, we have

$$v_{\mathfrak{q}}(g(\alpha)) = v_{\mathfrak{q}}(g(\alpha_0 + \pi_0)) = v_{\mathfrak{q}}(g(\alpha_0) + \pi_0 g'(\alpha_0) + \pi_0^2 h(\alpha_0)) = 1,$$

so $\pi \coloneqq g(\alpha)$ is also a uniformizer for B.

We now claim $B = A[\alpha]$, equivalently, that $1, \alpha, \dots, \alpha^{n-1}$ generate B as an A-module. By Nakayama's lemma, it suffices to show that the reductions of $1, \alpha, \dots, \alpha^{n-1}$ span $B/\mathfrak{p}B$ as an k-vector space. We have $\mathfrak{p} = \mathfrak{q}^e$, so $\mathfrak{p}B = (\pi^e)$. We can represent each element of $B/\mathfrak{p}B$ as a coset

$$b + \mathfrak{p}B = b_0 + b_1\pi + b_2\pi \cdots + b_{e-1}\pi^{e-1} + \mathfrak{p}B,$$

where $b_0, \dots, b_{e-1}$ are determined up to equivalence modulo πB. Now $1, \bar{\alpha}, \dots, \bar{\alpha}^{f-1}$ are a basis for $B/\pi B = B/\mathfrak{q}$ as a k-vector space, and $\pi = g(\alpha)$, so we can rewrite this as

$$\begin{aligned} b + \mathfrak{p}B = {} & (a_0 + a_1\alpha + \cdots a_{f-1}\alpha^{f-1}) \\ & + (a_f + a_{f+1}\alpha + \cdots a_{2f-1}\alpha^{f-1})g(\alpha) \\ & + \cdots \\ & + (a_{ef-f+1} + a_{ef-f+2}\alpha + \cdots a_{ef-1}\alpha^{f-1})g(\alpha)^{e-1} + \mathfrak{p}B. \end{aligned}$$

Since $\deg g = f$, and $n = ef$, this expresses $b + \mathfrak{p}B$ in the form $b' + \mathfrak{p}B$ with b' in the A-span of $1, \dots, \alpha^{n-1}$. Thus $B = A[\alpha]$.

We now note that if L/K is unramified then l/k is separable (this is part of the definition of unramified), and $e = 1$, $f = n$, in which case there is no need to require $g(\alpha)$ to be a uniformizer and we can just take $\alpha = \alpha_0$ to be any lift of any $\bar{\alpha}_0$ that generates l over k. □

In our *AKLB* setup, if A is a complete DVR with maximal ideal $\mathfrak{p}$ then B is a complete DVR with maximal ideal $\mathfrak{q}|\mathfrak{p}$ and the formula $[L : K] = \sum_{\mathfrak{q}|\mathfrak{p}} e_{\mathfrak{q}} f_{\mathfrak{q}}$ given by Theorem 5.35 has only one term $e_{\mathfrak{q}} f_{\mathfrak{q}}$. We now simplify matters even further by reducing to the two extreme cases $f_{\mathfrak{q}} = 1$ (a totally ramified extension) and $e_{\mathfrak{q}} = 1$ (an unramified extension, provided that the residue field extension is separable).[1]

10.2 Unramified extensions of a complete DVR

Let A be a complete DVR with fraction field K and residue field k. Associated to any finite unramified extension of L/K of degree n is a corresponding finite separable extension of residue fields l/k of the same degree n. Given that the extensions L/K and l/k are finite separable extensions of the same degree, we might wonder how they are related. More precisely, if we fix K with residue field k, what is the relationship between finite unramified extensions L/K of degree n and finite separable extensions l/k of degree n? Each L/K uniquely determines a corresponding l/k, but what about the converse?

This question has a surprisingly nice answer. The finite unramified extensions L of K form a category $\mathcal{C}_K^{\mathrm{unr}}$ whose morphisms are K-algebra homomorphisms, and the finite separable extensions l of k form a category $\mathcal{C}_k^{\mathrm{sep}}$ whose morphisms are k-algebra homomorphisms. These two categories are equivalent.

Theorem 10.15. *Let A be a complete DVR with fraction field K and residue field $k := A/\mathfrak{p}$. The categories $\mathcal{C}_K^{\mathrm{unr}}$ and $\mathcal{C}_k^{\mathrm{sep}}$ are equivalent via the functor $\mathcal{F}\colon \mathcal{C}_K^{\mathrm{unr}} \to \mathcal{C}_k^{\mathrm{sep}}$ that sends each unramified extension L of K to its residue field l, and each K-algebra homomorphism $\varphi\colon L_1 \to L_2$ to the k-algebra homomorphism $\bar{\varphi}\colon l_1 \to l_2$ defined by $\bar{\varphi}(\bar{\alpha}) := \overline{\varphi(\alpha)}$, where α is any lift of $\bar{\alpha} \in l_1 := B_1/\mathfrak{q}_1$ to B_1 and $\overline{\varphi(\alpha)}$ is the reduction of $\varphi(\alpha) \in B_2$ to $l_2 := B_2/\mathfrak{q}_2$; here $\mathfrak{q}_1, \mathfrak{q}_2$ are the maximal ideals of the valuation rings B_1, B_2 of L_1, L_2, respectively.*

In particular, $\mathcal{F}$ gives a bijection between the isomorphism classes in $\mathcal{C}_K^{\mathrm{unr}}$ and $\mathcal{C}_k^{\mathrm{sep}}$, and if L_1, L_2 and have residue fields l_1, l_2 then $\mathcal{F}$ induces a bijection of finite sets

$$\mathrm{Hom}_K(L_1, L_2) \xrightarrow{\sim} \mathrm{Hom}_k(l_1, l_2).$$

Proof. Let us first verify that $\mathcal{F}$ is well-defined. It is clear that it maps finite unramified extensions L/K to finite separable extensions l/k, but we should check that the map on morphisms does not depend on the lift α of $\bar{\alpha}$ we pick. So let $\varphi\colon L_1 \to L_2$ be a K-algebra homomorphism, and for $\bar{\alpha} \in l_1$, let α and α' be two lifts of $\bar{\alpha}$ to B_1. Then $\alpha - \alpha' \in \mathfrak{q}_1$, and this implies that $\varphi(\alpha - \alpha') \in \varphi(\mathfrak{q}_1) = \varphi(B_1) \cap \mathfrak{q}_2 \subseteq \mathfrak{q}_2$, and therefore $\overline{\varphi(\alpha)} = \overline{\varphi(\alpha')}$. The identity $\varphi(\mathfrak{q}_1) = \varphi(B_1) \cap \mathfrak{q}_2 \subseteq \mathfrak{q}_2$ follows from the fact that φ restricts to an injective ring homomorphism $B_1 \to B_2$ and $B_2/\varphi(B_1)$ is a finite extension of DVRs in which $\mathfrak{q}_2$ lies over the prime $\varphi(\mathfrak{q}_1)$ of $\varphi(B_1)$. It's easy to see that $\mathcal{F}$ sends identity morphisms to identity morphisms and that it is compatible with composition, so we have a well-defined functor.

To show that $\mathcal{F}$ is an equivalence of categories we need to prove two things:

[1] Recall from Definition 5.37 that separability of the residue field extension is part of the *definition* of an unramified extension. If the residue field is perfect (as when K is a local field, for example), the residue field extension is automatically separable, but in general it need not be, even when L/K is unramified.

- $\mathcal{F}$ is essentially surjective: each separable l/k is isomorphic to the residue field of some unramified L/K
- $\mathcal{F}$ is full and faithful: the induced map $\mathrm{Hom}_K(L_1, L_2) \to \mathrm{Hom}_k(l_1, l_2)$ is a bijection.

We first show that $\mathcal{F}$ is essentially surjective. Given a finite separable extension l/k, we may apply the primitive element theorem to write

$$l \simeq k(\bar{\alpha}) = \frac{k[x]}{(\bar{g}(x))},$$

for some $\bar{\alpha} \in l$ whose minimal polynomial $\bar{g} \in k[x]$ is necessarily monic, irreducible, separable, and of degree $n := [l : k]$. Let $g \in A[x]$ be any monic lift of $\bar{g}$; then g is also irreducible, separable, and of degree n. Now let

$$L := \frac{K[x]}{(g(x))} = K(\alpha),$$

where α is the image of x in $K[x]/g(x)$. Then L/K is a finite separable extension, and by Corollary 10.13, $(\mathfrak{p}, g(\alpha))$ is the unique maximal ideal of $A[\alpha]$ (since $\bar{g}$ is irreducible) and

$$\frac{B}{\mathfrak{q}} \simeq \frac{A[\alpha]}{(\mathfrak{p}, g(\alpha))} \simeq \frac{A[x]}{(\mathfrak{p}, g(x))} \simeq \frac{(A/\mathfrak{p})[x]}{(\bar{g}(x))} \simeq l.$$

We thus have $[L : K] = \deg g = [l : k] = n$, and it follows that L/K is an unramified extension of degree $n = f := [l : k]$: the ramification index of $\mathfrak{q}$ is necessarily $e = n/f = 1$, and the extension l/k is separable by assumption (so in fact $B = A[\alpha]$, by Theorem 10.14).

We now show that the functor $\mathcal{F}$ is full and faithful. Given finite unramified extensions L_1, L_2 with valuation rings B_1, B_2 and residue fields l_1, l_2, we have induced maps

$$\mathrm{Hom}_K(L_1, L_2) \xrightarrow{\sim} \mathrm{Hom}_A(B_1, B_2) \longrightarrow \mathrm{Hom}_k(l_1, l_2).$$

The first map is given by restriction from L_1 to B_1, and since tensoring with K gives an inverse map in the other direction, it is a bijection. We need to show that the same is true of the second map, which sends $\varphi\colon B_1 \to B_2$ to the k-homomorphism $\overline{\varphi}$ that sends $\overline{\alpha} \in l_1 = B_1/\mathfrak{q}_1$ to the reduction of $\varphi(\alpha)$ modulo $\mathfrak{q}_2$, where α is any lift of $\bar{\alpha}$.

As above, use the primitive element theorem to write $l_1 = k(\bar{\alpha}) = k[x]/(\bar{g}(x))$ for some $\bar{\alpha} \in l_1$. If we now lift $\bar{\alpha}$ to $\alpha \in B_1$, we must have $L_1 = K(\alpha)$, since $[L_1 : K] = [l_1 : k]$ is equal to the degree of the minimal polynomial $\bar{g}$ of $\bar{\alpha}$ which cannot be less than the degree of the minimal polynomial g of α (both are monic). Moreover, we also have $B_1 = A[\alpha]$, since this is true of the valuation ring of every finite unramified extension in our category.

Each A-module homomorphism in

$$\mathrm{Hom}_A(B_1, B_2) = \mathrm{Hom}_A\left(\frac{A[x]}{(g(x))}, B_2\right)$$

is uniquely determined by the image of x in B_2. Thus gives us a bijection between $\mathrm{Hom}_A(B_1, B_2)$ and the roots of g in B_2. Similarly, each k-algebra homomorphism in

$$\mathrm{Hom}_k(l_1, l_2) = \mathrm{Hom}_k\left(\frac{k[x]}{(\bar{g}(x))}, l_2\right)$$

is uniquely determined by the image of x in l_2, and there is a bijection between $\mathrm{Hom}_k(l_1, l_2)$ and the roots of $\bar{g}$ in l_2. Now $\bar{g}$ is separable, so every root of $\bar{g}$ in $l_2 = B_2/\mathfrak{q}_2$ lifts to a unique root of g in B_2, by Hensel's Lemma 9.15. Thus the map $\mathrm{Hom}_A(B_1, B_2) \longrightarrow \mathrm{Hom}_k(l_1, l_2)$ induced by $\mathcal{F}$ is a bijection. □

Remark 10.16. In the proof above we actually only used the fact that L_1/K is unramified. The map $\mathrm{Hom}_K(L_1, L_2) \to \mathrm{Hom}_k(l_1, l_2)$ is a bijection even if L_2/K is not unramified.

Let us note the following corollary, which follows from our proof of Theorem 10.15.

Corollary 10.17. *Assume AKLB with A a complete DVR with residue field k. Then L/K is unramified if and only if $B = A[\alpha]$ for some $\alpha \in L$ whose minimal polynomial $g \in A[x]$ has separable image $\bar{g}$ in $k[x]$.*

Proof. The forward direction was proved in the proof of the theorem, and for the reverse direction note that $\bar{g}$ must be irreducible, since otherwise we could use Hensel's lemma to lift a non-trivial factorization of $\bar{g}$ to a non-trivial factorization of g, so the residue field extension is separable and has the same degree as L/K, so L/K is unramified. □

Corollary 10.18. *Let A be a complete DVR with fraction field K and residue field k, and let ζ_n be a primitive nth root of unity in some algebraic closure of K, with n prime to the characteristic of k. The extension $K(\zeta_n)/K$ is unramified.*

Proof. The field $K(\zeta_n)$ is the splitting field of $f(x) = x^n - 1$ over K. The image $\bar{f}$ of f in $k[x]$ is separable when $p \nmid n$, since $\gcd(\bar{f}, \bar{f}') \neq 1$ only when $\bar{f}' = nx^{n-1}$ is zero, equivalently, only when $p|n$. When $\bar{f}$ is separable, so are all of its divisors, including the reduction of the minimal polynomial of ζ_n, which must be irreducible since otherwise we could obtain a contradiction by lifting a non-trivial factorization via Hensel's lemma. It follows that the residue field of $K(\zeta_n)$ is a separable extension of k, thus $K(\zeta_n)/K$ is unramified. □

When the residue field k is finite (always the case if K is a local field), we can give a precise description of the finite unramified extensions L/K.

Corollary 10.19. *Let A be a complete DVR with fraction field K and finite residue field $\mathbb{F}_q$, and let L be a degree n extension of K. Then L/K is unramified if and only if $L \simeq K(\zeta_{q^n-1})$. When this holds, $A[\zeta_{q^n-1}]$ is the integral closure of A in L and L/K is a Galois extension with $\mathrm{Gal}(L/K) \simeq \mathbb{Z}/n\mathbb{Z}$.*

Proof. The reverse implication is implied by Corollary 10.18; note that $K(\zeta_{q^n-1})$ has degree n over K because its residue field is the splitting field of $x^{q^n-1} - 1$ over $\mathbb{F}_q$, which is an extension of degree n (indeed, one can take this as the definition of $\mathbb{F}_{q^n}$).

Now suppose L/K is unramified. The residue field has degree n and is thus isomorphic to $\mathbb{F}_{q^n}$, so its multiplicative group is a cyclic of order $q^n - 1$ generated by some $\bar{\alpha}$. The minimal polynomial $\bar{g} \in \mathbb{F}_q[x]$ of $\bar{\alpha}$ divides $x^{q^n-1} - 1$, and since $\bar{g}$ is irreducible, it is coprime to the quotient $(x^{q^n-1} - 1)/\bar{g}$. By Hensel's Lemma 9.19, we can lift $\bar{g}$ to a polynomial $g \in A[x]$ that divides $x^{q^n-1} - 1 \in A[x]$, and by Hensel's Lemma 9.15 we can lift $\bar{\alpha}$ to a root α of g, in which case α is also a root of $x^{q^n-1} - 1$; it must be a primitive $(q^n - 1)$-root of unity because its reduction $\bar{\alpha}$ is.

Let B be the integral closure of A in L. We have $B \simeq A[\zeta_{q^n-1}]$ by Theorem 10.14, and L is the splitting field of $x^{q^n-1} - 1$, since its residue field $\mathbb{F}_{q^n}$ is (we can lift the factorization of $x^{q^n-1} - 1$ from $\mathbb{F}_{q^n}$ to L via Hensel's lemma). It follows that L/K is Galois, and the bijection between $(q^n - 1)$-roots of unity in L and $\mathbb{F}_{q^n}$ induces an isomorphism $\mathrm{Gal}(L/K) \simeq \mathrm{Gal}(l/k) = \mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q) \simeq \mathbb{Z}/n\mathbb{Z}$. □

Corollary 10.20. *Let A be a complete DVR with fraction field K and finite residue field of characteristic p, and suppose that K does not contain a primitive pth root of unity. The extension $K(\zeta_m)/K$ is ramified if and only if p divides m.*

Proof. If p does not divide m then Corollary 10.18 implies that $K(\zeta_m)/K$ is unramified. If p divides m then $K(\zeta_m)$ contains $K(\zeta_p)$, which by Corollary 10.19 is unramified if and only if $K(\zeta_p) \simeq K(\zeta_{p^n-1})$ with $n := [K(\zeta_p) : K]$, which occurs if and only if p divides $p^n - 1$ (since $\zeta_p \notin K$), which it does not; thus $K(\zeta_p)$ and therefore $K(\zeta_m)$ is ramified when $p|m$. □

Example 10.21. Consider $A = \mathbb{Z}_p$, $K = \mathbb{Q}_p$, $k = \mathbb{F}_p$, and fix $\overline{\mathbb{F}}_p$ and $\overline{\mathbb{Q}}_p$. For each positive integer n, the finite field $\mathbb{F}_p$ has a unique extension of degree n in $\overline{\mathbb{F}}_p$, namely, $\mathbb{F}_{p^n}$. Thus for each positive integer n, the local field $\mathbb{Q}_p$ has a unique unramified extension of degree n; it can be explicitly constructed by adjoining a primitive root of unity ζ_{p^n-1} to $\mathbb{Q}_p$. The element ζ_{p^n-1} will necessarily have minimal polynomial of degree n dividing $x^{p^n-1} - 1$.

Another useful consequence of Theorem 10.15 that applies when the residue field is finite is that the norm map $\mathrm{N}_{L/K}$ restricts to a surjective map $B^\times \to A^\times$ on unit groups; in fact, this property characterizes unramified extensions.

Theorem 10.22. *Assume AKLB with A a complete DVR with finite residue field. Then L/K is unramified if and only if $\mathrm{N}_{L/K}(B^\times) = A^\times$.*

Proof. See Problem Set 6. □

Definition 10.23. Let L/K be a separable extension. The *maximal unramified extension of K in L* is the subfield

$$\bigcup_{\substack{K \subseteq E \subseteq L \\ E/K \text{ fin. unram.}}} E \subseteq L$$

where the union is over finite unramified subextensions E/K. When $L = K^{\mathrm{sep}}$ is the separable closure of K, this is the *maximal unramified extension of K*, denoted K^{unr}.

Example 10.24. The field $\mathbb{Q}_p^{\mathrm{unr}}$ is an infinite extension of $\mathbb{Q}_p$ with Galois group

$$\mathrm{Gal}(\mathbb{Q}_p^{\mathrm{unr}}/\mathbb{Q}_p) \simeq \mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p) = \varprojlim_n \mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p) \simeq \varprojlim_n \mathbb{Z}/n\mathbb{Z} =: \hat{\mathbb{Z}},$$

where the inverse limit is taken over positive integers n ordered by divisibility. The ring $\hat{\mathbb{Z}}$ is the *profinite completion* of $\mathbb{Z}$. The field $\mathbb{Q}_p^{\mathrm{unr}}$ has value group $\mathbb{Z}$ and residue field $\overline{\mathbb{F}}_p$.

Theorem 10.25. *Assume AKLB with A a complete DVR and separable residue field extension l/k. Let $e_{L/K}$ and $f_{L/K}$ be the ramification index and residue field degrees, respectively. The following hold:*

(i) *There is a unique intermediate field extension E/K that contains every unramified extension of K in L and it has degree $[E : K] = f_{L/K}$.*

(ii) *The extension L/E is totally ramified and has degree $[L : E] = e_{L/K}$.*

(iii) *If L/K is Galois then $\mathrm{Gal}(L/E) = I_{L/K}$, where $I_{L/K} = I_{\mathfrak{q}}$ is the inertia subgroup of $\mathrm{Gal}(L/K)$ for the unique prime $\mathfrak{q}$ of B.*

Proof. (i) Let E/K be the finite unramified extension of K in L corresponding to the finite separable extension l/k given by Theorem 10.15; then $[E : K] = [l : k] = f_{L/K}$ as desired. The maximal unramified extension E' of K in L has the same residue field l as L, which is also the residue field of E, and equivalence of categories given by Theorem 10.15 implies that the trivial isomorphism $\ell \simeq \ell$ corresponds to an isomorphism $E \simeq E'$ that allows us to

view E as a subfield of L; the same applies to any unramified extension of K with residue field l, so E is unique up to isomorphism.

(ii) We have $f_{L/E} = [l : l] = 1$, so $e_{L/E} = [L : E] = [L : K]/[E : K] = e_{L/K}$.

(iii) By Proposition 7.13, we have $I_{L/E} = \mathrm{Gal}(L/E) \cap I_{L/K}$, and these three groups all have the same order $e_{L/K}$ so they must coincide. □

References

[1] David Eisenbud, *Commutative algebra with a view toward algebraic geometry*, Springer, 1995.

[2] N. Koblitz, *p-adic numbers, p-adic analysis, and zeta functions*, Springer, 1984.

[3] S. Lang, *Algebraic number theory*, second edition, Springer, 1994.

[4] J. Neukirch, *Algebraic number theory*, Springer, 1999.

11 Totally ramified extensions and Krasner's lemma

In the previous lecture we showed that in the *AKLB* setup, if A is a complete DVR with maximal ideal $\mathfrak{p}$ then B is a complete DVR with maximal ideal $\mathfrak{q}$ and $[L:K] = n = e_{\mathfrak{q}} f_{\mathfrak{q}}$; see Theorem 10.6 (note that the *AKLB* setup includes the assumption that L/K is separable). In this setting we may unambiguously write $e_{L/K}$ for $e_{\mathfrak{q}}$ and $f_{L/K}$ for $f_{\mathfrak{q}}$, since $\mathfrak{q}$ is the unique prime of L. Provided the residue field extension is separable (always the case if K is a local field), we can decompose the extension L/K as a tower of field extensions $L/E/K$ in which E/K is unramified (so $e_{E/K} = 1$ and $f_{E/K} = f_{L/K}$) and L/E is totally ramified (so $e_{L/E} = e_{L/K}$ and $f_{L/E} = 1$), by Theorem 10.25.

In the previous lecture we classified unramified extensions of (fraction fields of) complete DVRs, and showed that when the residue field is finite (always true for local fields), unramified extensions are all cyclotomic extensions of the form $K(\zeta_n)/K$ for some n coprime to the residue field characteristic; see Corollary 10.19. In this lecture we will classify totally ramified extensions of complete DVRs.

11.1 Totally ramified extensions of a complete DVR

Definition 11.1. Let A be a DVR with maximal ideal $\mathfrak{p}$. A monic polynomial

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in A[x]$$

is *Eisenstein* (or an *Eisenstein polynomial*) if $a_i \in \mathfrak{p}$ for $0 \le i < n$ and $a_0 \notin \mathfrak{p}^2$; equivalently, $v_{\mathfrak{p}}(a_i) \ge 1$ for $0 \le i < n$ and $v_{\mathfrak{p}}(a_0) = 1$. Note that this means a_0 is a uniformizer.

Lemma 11.2 (Eisenstein irreducibility)**.** *Let A be a DVR with fraction field K and let $f \in A[x]$ be Eisenstein. Then f is irreducible in both $A[x]$ and $K[x]$.*

Proof. Suppose not. Then $f = gh$ has degree $n \ge 2$ for some non-constant monic $g, h \in A[x]$. Put $f = \sum_i f_i x^i$, $g = \sum_i g_i x^i$, $h = \sum_i h_i x^i$. Then $v_{\mathfrak{p}}(f_0) = v_p(g_0 h_0) = v_{\mathfrak{p}}(g_0) + v_{\mathfrak{p}}(h_0) = 1$, where $\mathfrak{p}$ is the maximal ideal of A, and without loss of generality we may assume $v_{\mathfrak{p}}(g_0) = 0$ and $v_{\mathfrak{p}}(h_0) = 1$. Let $i > 0$ be the least i for which $v_{\mathfrak{p}}(h_i) = 0$; such an $i < n$ exists since h is monic and $\deg h < n$. We have

$$f_i = g_0 h_i + g_1 h_{i-1} + \cdots + g_{i-1} h_1 + g_i h_0,$$

with $v_{\mathfrak{p}}(f_i) \ge 1$ since f is Eisenstein and $i < n$, but the valuation of the RHS is zero, since $v_{\mathfrak{p}}(g_0 h_i) = 0$ and $v_{\mathfrak{p}}(g_j h_{i-j}) \ge 1$ for $0 \le j < i$, by the minimality of i, which is a contradiction. Thus f is irreducible in $A[x]$, and since A is a DVR, and therefore a PID and thus a UFD, f is irreducible in $K[x]$, by Gauss's Lemma [1]. □

Remark 11.3. We can apply Lemma 11.2 to any polynomial $f(x)$ over a Dedekind domain A that is Eisenstein over a localization $A_{\mathfrak{p}}$; the rings $A_{\mathfrak{p}}$ and A have the same fraction field K and f is then irreducible in $K[x]$, hence in $A[x]$; this gives the *Eisenstein criterion* for irreducibility.

Lemma 11.4. *Let A be a DVR and let $f \in A[x]$ be an Eisenstein polynomial. Then $B = A[\pi] \coloneqq A[x]/(f)$ is a DVR with uniformizer π, where π is the image of x in $A[x]/(f)$.*

Proof. Let $\mathfrak{p}$ be the maximal ideal of A. We have $f \equiv x^n \bmod \mathfrak{p}$, so by Corollary 10.13 the ideal $\mathfrak{q} = (\mathfrak{p}, x) = (\mathfrak{p}, \pi)$ is the only maximal ideal of B. Let $f = \sum f_i x^i$; then $\mathfrak{p} = (f_0)$ and $\mathfrak{q} = (f_0, \pi)$, and $f_0 = -f_1\pi - f_2\pi^2 - \cdots - \pi^n \in (\pi)$, so $\mathfrak{q} = (\pi)$. The unique maximal ideal (π) of B is nonzero and principal, so B is a DVR with uniformizer π. □

Theorem 11.5. *Assume AKLB with A a complete DVR and π a uniformizer for B. The extension L/K is totally ramified if and only if $B = A[\pi]$ and the minimal polynomial of π is Eisenstein.*

Proof. Let $n = [L : K]$, let $\mathfrak{p}$ be the maximal ideal of A, let $\mathfrak{q}$ be the maximal ideal of B (which we recall is a complete DVR, by Theorem 10.6), and let π be a uniformizer for B with minimal polynomial f. If $B = A[\pi]$ and f is Eisenstein, then as in Lemma 11.4 we have $\mathfrak{p} = \mathfrak{q}^n$, so $v_\mathfrak{q}$ extends $v_\mathfrak{p}$ with index $e_\mathfrak{q} = n$ and L/K is totally ramified.

We now suppose L/K is totally ramified. Then $v_\mathfrak{q}$ extends $v_\mathfrak{p}$ with index n, which implies $v_\mathfrak{q}(K) = n\mathbb{Z}$. The set $\{\pi^0, \pi^1, \pi^2, \ldots, \pi^{n-1}\}$ is linearly independent over K, since the valuations of $\pi^0, \ldots \pi^{n-1}$ are distinct modulo $v_\mathfrak{q}(K) = n\mathbb{Z}$ (if $\sum_{i=0}^{n-1} a_i\pi^i = 0$ we must have $v_\mathfrak{q}(a_i\pi^i) = v_\mathfrak{q}(a_j\pi^j)$ for some nonzero a_i and a_j with $i \neq j$, which is impossible). Thus $L = K(\pi)$.

Let $f = \sum_{i=0}^{n} a_i x^i \in A[x]$ be the minimal polynomial of π. We have $v_\mathfrak{q}(f(\pi)) = \infty$ and $v_\mathfrak{q}(a_i\pi^i) \equiv i \bmod n$ for $0 \leq i \leq n$. This is possible only if

$$v_\mathfrak{q}(a_0) = v_\mathfrak{q}(a_0\pi^0) = v_q(a_n\pi^n) = v_q(\pi^n) = n,$$

and $v_\mathfrak{q}(a_i) \geq n$ for $0 \leq i < n$. This implies that $v_\mathfrak{p}(a_0) = 1$, since $v_\mathfrak{q}$ extends $v_\mathfrak{p}$ with index n, and $v_\mathfrak{p}(a_i) \geq 1$ for $0 \leq i < n$. Thus f is Eisenstein and Lemma 11.4 implies that $A[\pi] \subseteq B$ is a DVR, hence maximal, so $B = A[\pi]$. □

Example 11.6. Let $K = \mathbb{Q}_3$. As shown in an earlier problem set, there are just three distinct quadratic extensions of $\mathbb{Q}_3$: $\mathbb{Q}_3(\sqrt{2})$, $\mathbb{Q}_3(\sqrt{3})$, and $\mathbb{Q}_3(\sqrt{6})$. The extension $\mathbb{Q}_3(\sqrt{2})$ is the unique unramified quadratic extension of $\mathbb{Q}_3$, and we note that it can be written as a cyclotomic extension $\mathbb{Q}_3(\zeta_8)$. The other two are both ramified, and can be defined by the Eisenstein polynomials $x^2 - 3$ and $x^2 - 6$.

Definition 11.7. Assume *AKLB* with A a complete DVR and separable residue field extension of characteristic $p \geq 0$. The extension L/K is *tamely ramified* if $p \nmid e_{L/K}$ (always true if $p = 0$); note that unramified extensions are tamely ramified. Otherwise L/K is *wildly ramified* if $p | e_{L/K}$. A totally ramified extension L/K is *totally tamely ramified* if $p \nmid e_{L/K}$, and it is *totally wildly ramified* if $e_{L/K}$ is a power of p (a totally ramified extension that is wildly ramified need not be totally wildly ramified).

Recall that ramification indices multiply in towers (Lemma 5.30), and separability is a transitive in towers (Corollary 4.14). This yields the following proposition, which we note applies to all nonarchimedean local fields.

Proposition 11.8. *The properties of being unramified, tamely ramified, wildly ramified, totally ramified, totally tamely ramified, and totally wildly ramified are all transitive in towers of extensions of fraction fields of complete DVRs with separable residue field extensions.*

Proof. This follows immediately from the transitivity of separability and the multiplicativity of ramification indices and degrees in towers. □

Remark 11.9. A compositum of totally ramified extensions need not be totally ramified. From Example 11.6 we see that the compositum of the totally ramified quadratic extensions $\mathbb{Q}_3(\sqrt{3})$ and $\mathbb{Q}_3(\sqrt{6})$ of $\mathbb{Q}_3$ contains the unramified quadratic extension $\mathbb{Q}_3(\sqrt{2})$ of $\mathbb{Q}_3$.

Theorem 11.10. *Assume AKLB with A a complete DVR and separable residue field extension of characteristic $p \geq 0$ not dividing $n := [L : K]$. The extension L/K is totally tamely ramified if and only if $L = K(\pi_A^{1/n})$ for some uniformizer π_A of A.*

Proof. If $L = K(\pi_A^{1/n})$ then $\pi = \pi_A^{1/n}$ has minimal polynomial $x^n - \pi_A$, which is Eisenstein, so $A[\pi]$ is a DVR by Lemma 11.4. This implies $B = A[\pi]$, since DVRs are maximal, and Theorem 11.5 implies that L/K is totally tamely ramified, since $p \nmid n$.

Now assume L/K is totally tamely ramified and let $\mathfrak{p}$ and $\mathfrak{q}$ be the maximal ideals of A and B with uniformizers π_A and π_B respectively. Then $v_\mathfrak{q}$ extends $v_\mathfrak{p}$ with index $e_\mathfrak{q} = n$ and $v_\mathfrak{q}(\pi_B^n) = n = v_\mathfrak{q}(\pi_A)$. This implies that $\pi_B^n = u\pi_A$ for some unit $u \in B^\times$. We have $f_\mathfrak{q} = 1$, so B and A have the same residue field, and if we lift the image of u in $B/\mathfrak{q} \simeq A/\mathfrak{p}$ to a unit u_A in A and replace π_A with $u_A^{-1}\pi_A$, we can assume that $u \equiv 1 \bmod \mathfrak{q}$. Now define $g(x) := x^n - u \in B[x]$ with reduction $\bar{g} = x^n - 1$ in $(B/\mathfrak{q})[x]$. We have $\bar{g}'(1) = n \neq 0$ (since $p \nmid n$), so by Hensel's Lemma 9.15 we can lift the root 1 of $\bar{g}(x)$ in $B/\mathfrak{q}$ to a root r of $g(x)$ in B. Now let $\pi := \pi_B/r$. Then π is a uniformizer for B and $B = A[\pi]$ by Theorem 11.5, so $L = K(\pi)$, and $\pi^n = \pi_B^n/r^n = \pi_B^n/u = \pi_A$, so $L = K(\pi_A^{1/n})$ as desired. □

Proposition 11.11. *Let L be a totally ramified extension of the fraction field K of a complete DVR. There is a unique intermediate field E such that E/K is totally tamely ramified and L/E is totally wildly ramified.*

Proof. Let $e := e_{L/K}$ be the ramification index and let $p \geq 0$ be the characteristic of the residue field. If $p \nmid e$ then the proposition holds with $E = L$, so we assume $p|e$, and put $e = mp^a$ with $p \nmid m$ (possibly $m = 1$).

Let A be the valuation ring of K with maximal ideal $\mathfrak{p}$, and let B be the valuation ring of L (also a complete DVR) with maximal ideal $\mathfrak{q}$. As in the proof of Theorem 11.10, we can choose uniformizers π_A of A and π_B of B such that $\pi_B^n = u\pi_A$ with $u \in B^\times$ and $u \equiv 1 \bmod \mathfrak{q}$. Let $g(x) = x^m - u \in B[x]$; as in the proof of the theorem we can construct a root $r \in B$ of $g(x)$ by Hensel lifting the root 1 of $\bar{g} \in (B/\mathfrak{q})[x]$. Now consider the field extension $E := K(\pi)$, where $\pi := \pi_B^{p^a}/r$. We have $\pi^m = \pi_B^n/r^m = \pi_B^n/u = \pi_A$, so $E = K(\pi_A^{1/m})$ with $p \nmid m$. The polynomial $x^m - \pi_A$ of π is Eisenstein, hence irreducible, and has π as a root, so E/K has degree m. By Theorem 11.10, the extension E/K is totally tamely ramified (the residue field extension is trivial, so it is certainly separable), and the extension L/E has degree p^a and is thus totally wildly ramified.

To see that E is unique, suppose $E' \subseteq L$ is another totally tamely ramified extensions of K such that L/E' is totally wildly ramified. Then E' must also be of the form $E' = K(\pi_A^{1/m})$, by Theorem 11.10 and its proof (we can use the same $\pi_A = u^{-1}\pi_B^n$ for intermediate field), in other words, E and E' are both generated by (possibly different) roots of $x^m - \pi_A$. The ratio of these roots is a (not necessarily primitive) mth root of unity $\zeta \in L$ that must lie in K because L/K is totally ramified and the extension $K(\zeta)/K$ is necessarily unramified, by Corollary 10.18, since $p \nmid m$. It follows that $E' = E$. □

Corollary 11.12. *Let L be a finite separable extension of the fraction field field K of a complete DVR with separable residue field extension. There is a unique intermediate field E such that E/K is tamely ramified and L/E is totally wildly ramified.*

Proof. Let F be the maximal unramified extension of K in L. By Corollary 10.17 we can assume $K = F(\alpha)$ where α is an integral element whose minimal polynomial g has separable image in $k[x]$, where k is the residue field of K. Applying the previous proposition to the totally ramified extension L/F yields a tamely ramified extension E/F with L/E totally wildly ramified. Unramified extensions are tamely ramified, so $E/F/K$ is a tower of tamely ramified extensions, hence tamely ramified.

Any field E' with L/E' totally wildly ramified must contain α, otherwise $E'(\alpha)$ would be a non-trivial unramified subextension L/E' (here we are again applying Corollary 10.17 and the fact that the image of the minimal polynomial of α over E' must divide g and thus has separable image in $k[x]$ and in $k'[x]$, where k' is the residue field of E', since k' is an extension of k). Proposition 11.11 then implies $E' = E$. □

11.2 Krasner's lemma

Let K be the fraction field of a complete DVR with absolute value $|\ |$. By Theorem 10.6 we can uniquely extend $|\ |$ to any finite extension L/K by defining $|x| := |N_{L/K}(x)|^{1/n}$, where $n = [L : K]$; as noted in Remark 10.7, this induces a unique absolute value on $\overline{K}$ that restricts to the absolute value of K.

Lemma 11.13. *Let K be the fraction field of a complete DVR with algebraic closure $\overline{K}$ and absolute value $|\ |$ extended to $\overline{K}$. For all $\alpha \in \overline{K}$ and $\sigma \in \mathrm{Aut}_K(\overline{K})$ we have $|\sigma(\alpha)| = |\alpha|$.*

Proof. The elements α and $\sigma(\alpha)$ must have the same minimal polynomial $f \in K[x]$, since $f(\sigma(\alpha)) = \sigma(f(\alpha)) = 0$, so $N_{K(\alpha)/K}(\alpha) = f(0) = N_{K(\sigma(\alpha))/K}(\sigma(\alpha))$, by Proposition 4.51. It follows that $|\sigma(\alpha)| = |N_{K(\sigma(\alpha))/K}(\alpha)|^{1/n} = |N_{K(\alpha)/K}(\alpha)|^{1/n} = |\alpha|$, where $n = \deg f$. □

Definition 11.14. Let K be the fraction field of a complete DVR with absolute value $|\ |$ extended to an algebraic closure $\overline{K}$. For $\alpha, \beta \in \overline{K}$, we say β *belongs to* α if $|\beta - \alpha| < |\beta - \sigma(\alpha)|$ for all $\sigma \in \mathrm{Aut}_K(\overline{K})$ with $\sigma(\alpha) \neq \alpha$, that is, β is strictly closer to α than it is to any of its conjugates. This is equivalent to requiring that $|\beta - \alpha| < |\alpha - \sigma(\alpha)|$ for all $\sigma(\alpha) \neq \alpha$, since every nonarchimedean triangle is isosceles (if one side is shorter than another, it is the shortest of all three sides).

Lemma 11.15 (Krasner's lemma)**.** *Let K be the fraction field of a complete DVR and let $\alpha, \beta \in \overline{K}$, with α separable over K. If β belongs to α then $K(\alpha) \subseteq K(\beta)$.*

Proof. Suppose not. Then β belongs to α but $\alpha \notin K(\beta)$. The extension $K(\alpha, \beta)/K(\beta)$ is separable and non-trivial, so there is an automorphism $\sigma \in \mathrm{Aut}_{K(\beta)}(\overline{K}/K(\beta))$ for which $\sigma(\alpha) \neq \alpha$ (let σ send α to a different root of the minimal polynomial of α over $K(\beta)$). Applying Lemma 11.13 to $\beta - \alpha \in \overline{K}$, for any $\sigma \in \mathrm{Aut}_{K(\beta)}(\overline{K}/K(\beta))$ we have

$$|\beta - \alpha| = |\sigma(\beta - \alpha)| = |\sigma(\beta) - \sigma(\alpha)| = |\beta - \sigma(\alpha)|,$$

since σ fixes β. But this contradicts the hypothesis that β belongs to α, since $\sigma(\alpha) \neq \alpha$. □

Remark 11.16. Krasner's lemma is another "Hensel's lemma" in the sense that it characterizes Henselian fields (fraction fields of Henselian rings); although the lemma is named after Krasner [2], it was proved earlier by Ostrowski in [3].

Definition 11.17. For a field K with absolute value $|\ |$ the *L^1-norm* of $f \in K[x]$ is defined by.

$$\|f\|_1 := \sum_i |f_i|,$$

where $f = \sum_i f_i x^i \in K[x]$; it is easily verified that $\|\ \|_1$ satisfies all the properties of Definition 10.3 and is thus a norm on the K-vector space $K[x]$.

Lemma 11.18. *Let K be a field with absolute value $|\ |$ and let $f := \prod_{i=1}^n (x-\alpha_i) \in K[x]$ be a monic polynomial with roots $\alpha_1, \dots, \alpha_n \in L$, where L/K is a field with an absolute value that extends $|\ |$. Then $|\alpha| < \|f\|_1$ for every root α of f.*

Proof. The lemma is clear for $n \le 1$, so assume $n \ge 2$. If $\|f\|_1 = 1$ then we must have $f = x^n$ and $\alpha = 0$, in which case $|\alpha| = 0 < 1 = \|f\|_1$ and the lemma holds. Otherwise $\|f\|_1 > 1$, and if $|\alpha| \le 1$ the lemma holds, so let α is a root of f with $|\alpha| > 1$. We have

$$0 = |f(\alpha)| = \left|\alpha^n + \sum_{i=0}^{n-1} f_i \alpha^i\right| \ge |\alpha|^n - \sum_{i=0}^{n-1} |f_i||\alpha|^i \ge |\alpha|^n - |\alpha|^{n-1} \sum_{i=0}^{n-1} |f_i| \ge |\alpha| - (\|f\|_1 - 1),$$

where we have used $|a| = |a+b-b| \le |a+b| + |-b| = |a+b| + |b|$ to get the general inequality $|a+b| \ge |a| - |b|$ which we applied repeatedly to get the first inequality above, we used $|\alpha| > 1$ to get the second (replacing $|\alpha|^i$ with $|\alpha|^{n-1}$ in each term) and the third (dividing by $|\alpha|^{n-1} \ge 1$). Thus $\|f\|_1 - 1 \ge |\alpha|$, and therefore $\|f\|_1 \ge |\alpha| + 1 > |\alpha|$. □

Theorem 11.19 (Continuity of roots)**.** *Let K be the fraction field of a complete DVR and $f \in K[x]$ a monic irreducible separable polynomial. There exists $\delta = \delta(f) \in \mathbb{R}_{>0}$ such that for every monic polynomial $g \in K[x]$ with $\|f-g\|_1 < \delta$ the following holds:*

Every root β of g belongs to a root α of f for which $K(\beta) = K(\alpha)$.

In particular, every such g is separable, irreducible, and has the same splitting field as f.

Proof. We first note that we can always pick $\delta < 1$, in which case any monic $g \in K[x]$ with $\|f-g\|_1 < \delta$ must have the same degree as f, so we can assume $\deg g = \deg f$. Let us fix an algebraic closure $\overline{K}$ of K with absolute value $|\ |$ extending the absolute value on K. Let $\alpha_1, \dots, \alpha_n$ be the roots of f in $\overline{K}$, and write

$$f(x) = \prod_i (x - \alpha_i) = \sum_{i=0}^n f_i x^i.$$

Let ϵ be the lesser of 1 and the minimum distance $|\alpha_i - \alpha_j|$ between any two distinct roots of f. We now define

$$\delta := \delta(f) := \left(\frac{\epsilon}{2(\|f\|_1 + 1)}\right)^n > 0,$$

and note that $\delta < 1$, since $\|f\|_1 \ge 1$ and $\epsilon \le 1$. Let $g(x) = \sum_i g_i x^i$ be a monic polynomial of degree n with $\|f-g\|_1 < \delta$. We then have

$$\|g\|_1 \le \|f\|_1 + \|g-f\|_1 = \|f\|_1 + \|f-g\|_1 < \|f\|_1 + \delta,$$

and for any root $\beta \in \overline{K}$ of g we have

$$|f(\beta)| = |f(\beta) - g(\beta)| = |(f-g)(\beta)| = \left|\sum_{i=0}^n (f_i - g_i)\beta^i\right| \le \sum_{i=0}^n |f_i - g_i||\beta|^i.$$

We have $|\beta| < \|g\|_1$ by Lemma 11.18, and $\|g\|_1 \geq 1$, so $|\beta|^i < \|g\|_1^i \leq \|g\|_1^n$. Thus

$$|f(\beta)| < \|f-g\|_1 \cdot \|g\|_1^n < \delta(\|f\|_1+\delta)^n < \delta(\|f\|_1+1)^n \leq (\epsilon/2)^n,$$

and therefore

$$\prod_{i=1}^{n} |\beta - \alpha_i| = |f(\beta)| < (\epsilon/2)^n.$$

It follows that $|\beta - \alpha_i| < \epsilon/2$ for at least one α_i, and the triangle inequality implies that this α_i must be unique since $|\alpha_i - \alpha_j| \geq \epsilon$ for $i \neq j$. Therefore β belongs to $\alpha := \alpha_i$.

By Krasner's lemma, $K(\alpha) \subseteq K(\beta)$, and we have $n = [K(\alpha):K] \leq [K(\beta):K] \leq n$, so $K(\alpha) = K(\beta)$. It follows that g is the minimal polynomial of β, since $\deg(g) = [K(\beta):K]$. Thus g is irreducible, and it is also separable, since $\beta \in K(\beta) = K(\alpha)$ lies in a separable extension of K. We now observe that if a root β of g belongs to a root α of f, then for any $\tau \in \mathrm{Aut}_K(\overline{K})$ and all $\sigma \in \mathrm{Aut}_K(\overline{K})$ such that $\sigma(\alpha) \neq \alpha$ we have

$$|\tau(\beta) - \tau(\alpha)| = |\tau(\beta-\alpha)| = |\beta - \alpha| < |\alpha - \sigma(\alpha)| = |\tau(\alpha - \sigma(\alpha))| = |\tau(\alpha) - \tau(\sigma(\alpha))|.$$

Noting that $\sigma(\alpha) \neq \alpha \iff \tau(\sigma(\alpha)) \neq \tau(\alpha)$, this implies that $\tau(\beta)$ belongs to $\tau(\alpha)$. Now $\mathrm{Aut}_K(\overline{K})$ acts transitively on the roots of f and g, so every root β of g belongs to a distinct root α of f for which $K(\beta) = K(\alpha)$. Therefore g has the same splitting field as f. □

11.3 Local extensions come from global extensions

Let $\hat{L}$ be a local field. From our classification of local fields (Theorem 9.9), we know that $\hat{L}$ is (isomorphic to) a finite extension of $\hat{K} = \mathbb{Q}_p$ (some $p \leq \infty$) or $\hat{K} = \mathbb{F}_q((t))$ (some q). We also know that the completion of a global field at any of its nontrivial absolute values is a local field (Corollary 9.7). It thus reasonable to ask whether $\hat{L}$ is the completion of a corresponding global field L that is a finite extension of $K = \mathbb{Q}$ or $K = \mathbb{F}_q(t)$.

More generally, for any fixed global field K and local field $\hat{K}$ that is the completion of K with respect to one of its nontrivial absolute values $|\ |$, we may ask whether every finite extension of local fields $\hat{L}/\hat{K}$ necessarily corresponds to an extension of global fields L/K, where $\hat{L}$ is the completion of L with respect to one of its absolute values (whose restriction to K must be equivalent to $|\ |$). The answer is yes. In order to simplify matters we restrict our attention to the case where $\hat{L}/\hat{K}$ is separable, but this is true in general.

Theorem 11.20. *Let K be a global field with a nontrivial absolute value $|\ |$, and let $\hat{K}$ be the completion of K with respect to $|\ |$. Every finite separable extension $\hat{L}$ of $\hat{K}$ is the completion of a finite separable extension L of K with respect to an absolute value that restricts to $|\ |$. One can choose L so that $[L:K] = [\hat{L}:\hat{K}]$, in which case $\hat{L} = \hat{K} \cdot L$.*

Proof. Let $\hat{L}/\hat{K}$ be a separable extension of degree n. If $|\ |$ is archimedean then K is a number field and $\hat{K}$ is either $\mathbb{R}$ or $\mathbb{C}$; the only nontrivial case is $\hat{K} \simeq \mathbb{R}$ and $n = 2$, and we may then assume that $\hat{L} = \hat{K}(\sqrt{d}) \simeq \mathbb{C}$ where $d \in \mathbb{Z}_{<0}$ is any nonsquare in K (such a d exists because $K/\mathbb{Q}$ is finite). We may assume without loss of generality that $|\ |$ is the Euclidean absolute value on $\hat{K} \simeq \mathbb{R}$ (it must be equivalent to it), and uniquely extend $|\ |$ to $L := K(\sqrt{d})$ by requiring $|\sqrt{d}| = \sqrt{-d}$. Then $\hat{L}$ is the completion of L with respect to $|\ |$, and clearly $[L:K] = [\hat{L}:\hat{K}] = 2$, and $\hat{L}$ is the compositum of $\hat{K}$ and L.

We now suppose that $|\ |$ is nonarchimedean, in which case the valuation ring of $\hat{K}$ is a complete DVR and $|\ |$ is induced by its discrete valuation. By the primitive element theorem

(Theorem 4.12), we may assume $\hat{L} = \hat{K}[x]/(f)$ where $f \in \hat{K}[x]$ is monic, irreducible, and separable. The field K is dense in its completion $\hat{K}$, so we can find a monic $g \in K[x] \subseteq \hat{K}[x]$ such that $\|g - f\|_1 < \delta$ for any $\delta > 0$. It then follows from Theorem 11.19 that $\hat{L} = \hat{K}[x]/(g)$ (and that g is separable). The field $\hat{L}$ is a finite separable extension of the fraction field of a complete DVR, so by Theorem 10.6 it is itself the fraction field of a complete DVR and has a unique absolute value that extends the absolute value $|\ |$ on $\hat{K}$.

Now let $L := K[x]/(g)$. The polynomial g is irreducible in $\hat{K}[x]$, hence in $K[x]$, so $[L : K] = \deg g = [\hat{L} : \hat{K}]$. The field $\hat{L}$ contains both $\hat{K}$ and L, and it is clearly the smallest field that does (since g is irreducible in $\hat{K}[x]$), so $\hat{L}$ is the compositum of $\hat{K}$ and L. The absolute value on $\hat{L}$ restricts to an absolute value on L extending the absolute value $|\ |$ on K, and $\hat{L}$ is complete, so $\hat{L}$ contains the completion of L with respect to $|\ |$. On the other hand, the completion of L with respect $|\ |$ contains L and $\hat{K}$, so it must be $\hat{L}$. □

In the preceding theorem, when the local extension $\hat{L}/\hat{K}$ is Galois one might ask whether the corresponding global extension L/K is also Galois, and whether $\mathrm{Gal}(\hat{L}/\hat{K}) \simeq \mathrm{Gal}(L/K)$. As shown by the following example, this need not be the case.

Example 11.21. Let $K = \mathbb{Q}$, $\hat{K} = \mathbb{Q}_7$ and $\hat{L} = \hat{K}[x]/(x^3 - 2)$. The extension $\hat{L}/\hat{K}$ is Galois because $\hat{K} = \mathbb{Q}_7$ contains ζ_3 (we can lift the root 2 of $x^2 + x + 1 \in \mathbb{F}_7[x]$ to a root of $x^2 + x + 1 \in \mathbb{Q}_7[x]$ via Hensel's lemma), and this implies that $x^3 - 2$ splits completely in $\hat{L}$. But $L = K[x]/(x^3 - 2)$ is not a Galois extension of K because it contains only one root of $x^3 - 2$. However, we can replace K with $\mathbb{Q}(\zeta_3)$ without changing $\hat{K}$ (take the completion of K with respect to the absolute value induced by a prime above 7) or $\hat{L}$, but now $L = K[x]/(x^3 - 2)$ is a Galois extension of K.

In the example we were able to adjust our choice of the global field K without changing the local fields extension $\hat{L}/\hat{K}$ in a way that ensures that $\hat{L}/\hat{K}$ and L/K have the same automorphism group. Indeed, this is always possible.

Corollary 11.22. *For every finite Galois extension $\hat{L}/\hat{K}$ of local fields there is a finite Galois extension of global fields L/K and an absolute value $|\ |$ on L such that $\hat{L}$ is the completion of L with respect to $|\ |$, $\hat{K}$ is the completion of K with respect to the restriction of $|\ |$ to K, and* $\mathrm{Gal}(L/K) \simeq \mathrm{Gal}(\hat{L}/\hat{K})$.

Proof. The archimedean case is already covered by Theorem 11.20 (take $K = \mathbb{Q}$), so we assume $\hat{L}$ is nonarchimedean and note that we may take $|\ |$ to be the absolute value on both $\hat{K}$ and on $\hat{L}$, by Theorem 10.6. The field $\hat{K}$ is an extension of either $\mathbb{Q}_p$ or $\mathbb{F}_q((t))$, and by applying Theorem 11.20 to this extension we may assume $\hat{K}$ is the completion of a global field K with respect to the restriction of $|\ |$. As in the proof of the theorem, let $g \in K[x]$ be a monic separable polynomial irreducible in $\hat{K}[x]$ such that $\hat{L} = \hat{K}[x]/(g)$ and define $L := K[x]/(g)$ so that $\hat{L}$ is the compositum of $\hat{K}$ and L.

Now let M be the splitting field of g over K, the minimal extension of K that contains all the roots of g (which are distinct because g is separable). The field $\hat{L}$ also contains these roots (since $\hat{L}/\hat{K}$ is Galois) and $\hat{L}$ contains K, so $\hat{L}$ contains a subextension of K isomorphic to M (by the universal property of a splitting field), which we now identify with M; note that $\hat{L}$ is also the completion of M with respect to the restriction of $|\ |$ to M.

We have a group homomorphism $\varphi\colon \mathrm{Gal}(\hat{L}/\hat{K}) \to \mathrm{Gal}(M/K)$ induced by restriction, and φ is injective (each $\sigma \in \mathrm{Gal}(\hat{L}/\hat{K})$ is determined by its action on any root of g in M). If we now replace K by the fixed field of the image of φ and replace L with M, the completion of K with respect to the restriction of $|\ |$ is still equal to $\hat{K}$, and similarly for L and $\hat{L}$, and now $\mathrm{Gal}(L/K) \simeq \mathrm{Gal}(\hat{L}/\hat{K})$ as desired. □

11.4 Completing a separable extension of Dedekind domains

We now return to our general *AKLB* setup: A is a Dedekind domain with fraction field K with a finite separable extension L/K, and B is the integral closure of A in L, which is also a Dedekind domain. Recall from Theorem 8.20 that if $\mathfrak{p}$ is a prime of K (a nonzero prime ideal of A), each prime $\mathfrak{q}|\mathfrak{p}$ induces a valuation $v_\mathfrak{q}$ of L that extends the valuation $v_\mathfrak{p}$ of K with index $e_\mathfrak{q}$, meaning that $v_\mathfrak{q}|_K = e_\mathfrak{q} v_\mathfrak{p}$ (and every valuation of L that extends $v_\mathfrak{p}$ arises in this way). We now want to look at what happens when we complete K with respect to the absolute value $|\ |_\mathfrak{p}$ induced by $v_\mathfrak{p}$ to obtain a complete field $K_\mathfrak{p}$, and similarly complete L with respect to $|\ |_\mathfrak{q}$ for some $\mathfrak{q}|\mathfrak{p}$ to obtain $L_\mathfrak{q}$. This includes the case where L/K is an extension of global fields, in which case we get a corresponding extension $L_\mathfrak{q}/K_\mathfrak{p}$ of local fields for each $\mathfrak{q}|\mathfrak{p}$; as proved below, the embedding $K \hookrightarrow L$ induces an embedding $K_\mathfrak{p} \hookrightarrow L_\mathfrak{q}$ of topological fields in which the absolute value $|\ |_\mathfrak{p}$ on $K_\mathfrak{p}$ is equivalent to the restriction of $|\ |_\mathfrak{q}$ to $K_\mathfrak{p}$ (if we define $|\ |_\mathfrak{q}$ as in Theorem 10.6 then $|\ |_\mathfrak{p}$ will be the restriction of $|\ |_\mathfrak{q}$).

In general the extension $L_\mathfrak{q}/K_\mathfrak{p}$ may have smaller degree than L/K. If $L \simeq K[x]/(f)$, the irreducible polynomial $f \in K[x]$ need not be irreducible over $K_\mathfrak{p}$. Indeed, this will necessarily be the case if there is more than one prime $\mathfrak{q}$ lying above $\mathfrak{p}$; the Dedekind-Kummer theorem gives a one-to-one correspondence between irreducible factors of f in $K_\mathfrak{p}[x]$ and primes $\mathfrak{q}|\mathfrak{p}$ (via Hensel's Lemma). The following theorem gives a complete description of the situation.

Theorem 11.23. *Assume AKLB, let $\mathfrak{p}$ be a prime of A, and let $\mathfrak{p}B = \prod_{\mathfrak{q}|\mathfrak{p}} \mathfrak{q}^{e_\mathfrak{q}}$ be the factorization of $\mathfrak{p}B$ in B. Let $K_\mathfrak{p}$ be the completion of K with respect to $|\ |_\mathfrak{p}$, and let $\hat{\mathfrak{p}}$ be the maximal ideal of its valuation ring. For each $\mathfrak{q}|\mathfrak{p}$, let $L_\mathfrak{q}$ denote the completion of L with respect to $|\ |_\mathfrak{q}$, and $\hat{\mathfrak{q}}$ the maximal ideal of its valuation ring. The following hold:*

(1) *Each $L_\mathfrak{q}$ is a finite separable extension of $K_\mathfrak{p}$ with $[L_\mathfrak{q} : K_\mathfrak{p}] \le [L : K]$.*

(2) *Each $\hat{\mathfrak{q}}$ is the unique prime of $L_\mathfrak{q}$ lying over $\hat{\mathfrak{p}}$.*

(3) *Each $\hat{\mathfrak{q}}$ has ramification index $e_{\hat{\mathfrak{q}}} = e_\mathfrak{q}$ and residue field degree $f_{\hat{q}} = f_\mathfrak{q}$.*

(4) $[L_\mathfrak{q} : K_\mathfrak{p}] = e_\mathfrak{q} f_\mathfrak{q}$;

(5) *The map $L \otimes_K K_\mathfrak{p} \to \prod_{\mathfrak{q}|\mathfrak{p}} L_\mathfrak{q}$ defined by $\ell \otimes x \mapsto (\ell x, \dots, \ell x)$ is an isomorphism of finite étale $K_\mathfrak{p}$-algebras.*

(6) *If L/K is Galois then each $L_\mathfrak{q}/K_\mathfrak{p}$ is Galois and we have isomorphisms of decomposition groups $D_\mathfrak{q} \simeq D_{\hat{\mathfrak{q}}} = \mathrm{Gal}(L_\mathfrak{q}/K_\mathfrak{p})$ and inertia groups $I_\mathfrak{q} \simeq I_{\hat{\mathfrak{q}}}$.*

Proof. We first note that the $K_\mathfrak{p}$ and the $L_\mathfrak{q}$ are all fraction fields of complete DVRs; this follows from Proposition 8.11 (note that we are not assuming they are local fields).

(1) For each $\mathfrak{q}|\mathfrak{p}$ the embedding $K \hookrightarrow L$ induces an embedding $K_\mathfrak{p} \hookrightarrow L_\mathfrak{q}$ via the map $[(x_n)] \mapsto [(x_n)]$ on equivalence classes of Cauchy sequences; a sequence (x_n) that is Cauchy in K with respect to $|\ |_\mathfrak{p}$, is also Cauchy in L with respect to $|\ |_\mathfrak{q}$ because $v_\mathfrak{q}$ extends $v_\mathfrak{p}$. We may thus view $K_\mathfrak{p}$ as a topological subfield of $L_\mathfrak{q}$, and it is clear that $[L_\mathfrak{q} : K_\mathfrak{p}] \le [L : K]$, since any K-basis $b_1, \dots, b_m$ for $L \subseteq L_\mathfrak{q}$ spans $L_\mathfrak{q}$ as a $K_\mathfrak{p}$-vector space: given a Cauchy sequence $y := (y_n)$ of elements in L, if we write each y_n as $x_{1,n}b_1 + \cdots + x_{m,n}b_m$ with $x_{i,n} \in K$ we obtain Cauchy sequences $x_1 := (x_{1,n}), \cdots, x_m := (x_{m,n})$ of elements in K (linear maps of finite dimensional normed spaces are uniformly continuous and thus preserves Cauchy sequences), and we can write $[y] = [x_1]b_1 + \cdots [x_m]b_m$ as a $K_\mathfrak{p}$-linear combination of $b_1, \dots, b_m$.

The field L is a finite étale K-algebra, since L/K is a separable extension, so its base change $L \otimes_K K_\mathfrak{p}$ to $K_\mathfrak{p}$ is a finite étale $K_\mathfrak{p}$-algebra, by Proposition 4.36. Let us now consider the $K_\mathfrak{p}$-algebra homomorphism $\phi_\mathfrak{q}\colon L \otimes_K K_\mathfrak{p} \to L_\mathfrak{q}$ defined by $\ell \otimes x \mapsto \ell x$. We have $\phi_\mathfrak{q}(b_i \otimes 1) = b_i$ for each of our K-basis elements $b_i \in L$, and as noted above, $b_1, \dots b_m$ span $L_\mathfrak{q}$ as $K_\mathfrak{p}$-vector space, thus $\phi_\mathfrak{q}$ is surjective. As a finite étale $K_\mathfrak{p}$-algebra, $L \otimes_K K_\mathfrak{p}$ is by definition isomorphic to a finite product of finite separable extensions of $K_\mathfrak{p}$; by Proposition 4.32, $L_\mathfrak{q}$ is isomorphic to a subproduct and thus also a finite étale $K_\mathfrak{p}$-algebra; in particular, $L_\mathfrak{q}/K_\mathfrak{p}$ is separable.

(2) As noted above, the valuation rings of $K_\mathfrak{p}$ and the $L_\mathfrak{q}$ are complete DVRs, so this follows immediately from Theorem 10.1.

(3) The valuation $v_{\hat{\mathfrak{q}}}$ extends $v_\mathfrak{q}$ with index 1, which in turn extends $v_\mathfrak{p}$ with index $e_\mathfrak{q}$. The valuation $v_{\hat{\mathfrak{p}}}$ extends $v_\mathfrak{p}$ with index 1, and it follows that $v_{\hat{q}}$ extends $v_{\hat{p}}$ with index $e_\mathfrak{q}$ and therefore $e_{\hat{\mathfrak{q}}} = e_\mathfrak{q}$. The residue field of $\hat{\mathfrak{p}}$ is the same as that of $\mathfrak{p}$: for any Cauchy sequence (a_n) over K the a_n will eventually all have the same image in the residue field at $\mathfrak{p}$ (since $v_\mathfrak{p}(a_n - a_m) > 0$ for all sufficiently large m and n). Similar comments apply to each $\hat{\mathfrak{q}}$ and $\mathfrak{q}$, and it follows that $f_{\hat{\mathfrak{q}}} = f_\mathfrak{q}$.

(4) It follows from (2) that $[L_\mathfrak{q} : K_\mathfrak{p}] = e_{\hat{\mathfrak{q}}} f_{\hat{\mathfrak{q}}}$, since $\hat{\mathfrak{q}}$ is the only prime above $\hat{\mathfrak{p}}$, and (3) then implies $[L_\mathfrak{q} : K_\mathfrak{p}] = e_\mathfrak{q} f_\mathfrak{q}$, by Theorem 5.35.

(5) Let $\phi := \prod_{\mathfrak{q}|\mathfrak{p}} \phi_\mathfrak{q}$, where $\phi_\mathfrak{q}\colon L \otimes_K K_\mathfrak{p} \to L_\mathfrak{q}$ is the surjective $K_\mathfrak{p}$-algebra homomorphisms defined in the proof of (1). Then $\phi\colon L \otimes_K K_\mathfrak{p} \to \prod_{\mathfrak{q}|\mathfrak{p}} L_\mathfrak{q}$ is a $K_\mathfrak{p}$-algebra homomorphism. Applying (4) and the fact that taking the base change of a finite étale algebra does not change its dimension (see Proposition 4.36), we have

$$\dim_{K_\mathfrak{p}} (L \otimes_K K_\mathfrak{p}) = \dim_K L = [L : K] = \sum_{\mathfrak{q}|\mathfrak{p}} e_\mathfrak{q} f_\mathfrak{q} = \sum_{\mathfrak{q}|\mathfrak{p}} [L_\mathfrak{q} : K_\mathfrak{p}] = \dim_{K_\mathfrak{p}} \prod_{\mathfrak{q}|\mathfrak{p}} L_\mathfrak{q}.$$

Pick a $K_\mathfrak{p}$-basis $\{\beta_i\}$ for $\prod_{\mathfrak{q}|\mathfrak{p}} L_\mathfrak{q}$, fix $\epsilon > 0$, and for each basis element $\beta_i = (\beta_{i,\mathfrak{q}})_{\mathfrak{q}|\mathfrak{p}}$ use the weak approximation theorem proved in Problem Set 4 to construct $\alpha_i \in L$ such that $|\alpha_i - \beta_{i,\mathfrak{q}}|_\mathfrak{q} < \epsilon$ for all $\mathfrak{q}|\mathfrak{p}$. In the metric space $\prod_{\mathfrak{q}|\mathfrak{p}} L_\mathfrak{q}$ (with the sup norm), each $\phi(\alpha_i \otimes 1)$ is close to β_i. The $K_\mathfrak{p}$-matrix whose jth column expresses $\phi(\alpha_j \otimes 1)$ in terms of the basis $\{\beta_i\}$ is then close to the identity matrix (with respect to $|\ |_\mathfrak{p}$), and the determinant D of this matrix is close to 1 (the determinant is continuous). For sufficiently small ϵ we must have $D \neq 0$, and then $\{\phi(\alpha_i \otimes 1)\}$ is a basis for $\prod_{\mathfrak{q}|\mathfrak{p}} L_\mathfrak{q}$. It follows that ϕ is surjective and therefore an isomorphism, since its domain and codomain have the same dimension.

(6) We now assume L/K is Galois. Each $\sigma \in D_\mathfrak{q}$ acts on L and respects the valuation $v_\mathfrak{q}$, since it fixes $\mathfrak{q}$ (if $x \in \mathfrak{q}^n$ then $\sigma(x) \in \sigma(\mathfrak{q}^n) = \sigma(\mathfrak{q})^n = \mathfrak{q}^n$). It follows that if (x_n) is a Cauchy sequence in L, then so is $(\sigma(x_n))$, thus σ is an automorphism of $L_\mathfrak{q}$, and it fixes $K_\mathfrak{p}$. We thus have a group homomorphism $\varphi\colon D_\mathfrak{q} \to \mathrm{Aut}_{K_\mathfrak{p}}(L_\mathfrak{q})$.

If $\sigma \in D_\mathfrak{q}$ acts trivially on $L_\mathfrak{q}$ then it acts trivially on $L \subseteq L_\mathfrak{q}$, so $\ker \varphi$ is trivial. Also,

$$e_\mathfrak{q} f_\mathfrak{q} = |D_\mathfrak{q}| \leq \#\mathrm{Aut}_{K_\mathfrak{p}}(L_\mathfrak{q}) \leq [L_\mathfrak{q} : K_\mathfrak{p}] = e_\mathfrak{q} f_\mathfrak{q},$$

by Theorem 11.23, so $\#\mathrm{Aut}_{K_\mathfrak{p}}(L_\mathfrak{q}) = [L_\mathfrak{q} : K_\mathfrak{p}]$ and $L_\mathfrak{q}/K_\mathfrak{p}$ is Galois, and this also shows that φ is surjective and therefore an isomorphism. There is only one prime $\hat{q}$ of $L_\mathfrak{q}$, and it is necessarily fixed by every $\sigma \in \mathrm{Gal}(L_\mathfrak{q}/K_\mathfrak{p})$, so $\mathrm{Gal}(L_\mathfrak{q}/K_\mathfrak{p}) \simeq D_{\hat{\mathfrak{q}}}$. The inertia groups $I_\mathfrak{q}$ and $I_{\hat{\mathfrak{q}}}$ both have order $e_\mathfrak{q} = e_{\hat{\mathfrak{q}}}$, and φ restricts to a homomorphism $I_\mathfrak{q} \to I_{\hat{\mathfrak{q}}}$, so the inertia groups are also isomorphic. □

Corollary 11.24. *Assume AKLB and let $\mathfrak{p}$ be a prime of A. For every $\alpha \in L$ we have*

$$\mathrm{N}_{L/K}(\alpha) = \prod_{\mathfrak{q}|\mathfrak{p}} \mathrm{N}_{L_\mathfrak{q}/K_\mathfrak{p}}(\alpha) \qquad \text{and} \qquad \mathrm{T}_{L/K}(\alpha) = \sum_{\mathfrak{q}|\mathfrak{p}} \mathrm{T}_{L_\mathfrak{q}/K_\mathfrak{q}}(\alpha).$$

where we view α as an element of $L_\mathfrak{q}$ and $\mathrm{N}_{L/K}(\alpha)$ as an element of $K_\mathfrak{p}$ via the canonical embeddings $L \hookrightarrow L_\mathfrak{q}$ and $K \hookrightarrow K_\mathfrak{p}$.

Proof. The norm and trace are defined as the determinant and trace of K-linear maps $L \xrightarrow{\times\alpha} L$ that are unchanged upon tensoring with $K_\mathfrak{p}$; the corollary then follows from the isomorphism in part (5) of Theorem 11.23, which commutes with the norm and trace. □

Remark 11.25. Theorem 11.23 can be stated more generally in terms of equivalence classes of absolute values, or *places*. Rather than working with a prime $\mathfrak{p}$ of K and primes $\mathfrak{q}|\mathfrak{p}$ of L, one works with an absolute value $|\ |_v$ of K (for example, $|\ |_\mathfrak{p}$) and inequivalent absolute values $|\ |_w$ of L that extend $|\ |_v$. Places will be discussed further in the next lecture.

Corollary 11.26. *Assume AKLB and let $\mathfrak{p}$ be a prime of A. Let $\mathfrak{p}B = \prod \mathfrak{q}^{e_\mathfrak{q}}$ be the factorization of $\mathfrak{p}B$ in B. Let $\hat{A}_\mathfrak{p}$ denote the completion of A with respect to $|\ |_\mathfrak{p}$, and for each $\mathfrak{q}|\mathfrak{p}$, let $\hat{B}_\mathfrak{q}$ denote the completion of B with respect to $|\ |_\mathfrak{q}$. Then $B \otimes_A \hat{A}_\mathfrak{p} \simeq \prod_{\mathfrak{q}|\mathfrak{p}} \hat{B}_\mathfrak{q}$, as $\hat{A}_\mathfrak{p}$-algebras*

Proof. After replacing A with $A_\mathfrak{p}$ and B with $B_\mathfrak{p}$ (localizing B as an A-module), we may assume that A is a DVR and B/A is a free A module of rank $n := [L : K] = \sum_{\mathfrak{q}|\mathfrak{p}} e_\mathfrak{q} f_\mathfrak{q}$. Then $B \otimes_A \hat{A}_\mathfrak{p}$ is a free $\hat{A}_\mathfrak{p}$-module of rank n. Viewing $\hat{A}_\mathfrak{p}$ and the $\hat{B}_\mathfrak{q}$ as valuation rings of $K_\mathfrak{p}$ and $L_\mathfrak{q}$, it follows from part (4) of Theorem 11.23 that $\prod \hat{B}_\mathfrak{q}$ is a free $\hat{A}_\mathfrak{p}$-module of rank $\sum_{\mathfrak{q}|\mathfrak{p}} [L_\mathfrak{q} : K_\mathfrak{p}] = \sum_{\mathfrak{q}|\mathfrak{p}} e_\mathfrak{q} f_\mathfrak{q} = n$. These isomorphic $\hat{A}_\mathfrak{p}$-modules lie in isomorphic finite étale $K_\mathfrak{p}$-algebras $L \otimes_K K_\mathfrak{p} \simeq \prod L_\mathfrak{q}$, by part (5) of Theorem 11.23, and this $K_\mathfrak{p}$-algebra isomorphism restricts to an $\hat{A}_\mathfrak{p}$-algebra isomorphism. □

Remark 11.27. Let A be a Dedekind domain with fraction field K. If we localize A at a prime $\mathfrak{p}$ we obtain a DVR $A_\mathfrak{p}$ with the same fraction field K. We can then complete $A_\mathfrak{p}$ with respect to $|\ |_\mathfrak{p}$ to obtain a complete DVR $\hat{A}_\mathfrak{p}$ whose fraction field $K_\mathfrak{p}$ is the completion of K with respect to $|\ |_\mathfrak{p}$, and $\hat{A}_\mathfrak{p}$ is then the valuation ring of $K_\mathfrak{p}$. Alternatively, we could first complete A with respect to the absolute value $|\ |_\mathfrak{p}$ induced by $\mathfrak{p}$ and then localize. But as explained in Lecture 8, completing A with respect to $|\ |_\mathfrak{p}$ is the same thing as taking the valuation ring of $K_\mathfrak{p}$, so the completion of A is already the complete DVR $\hat{A}_\mathfrak{p}$ we obtained by localizing and completing; there is no need to localize and nothing would change if we did. Completion not only commutes with localization, it makes localization unnecessary.

Henceforth if A is a Dedekind domain and $\mathfrak{p}$ is a prime of A (a nonzero prime ideal), by the *completion of A at $\mathfrak{p}$* we mean the ring $\hat{A}_\mathfrak{p}$.

References

[1] Michael Artin, *Algebra*, 2nd edition, Pearson, 2010.

[2] Marc Krasner, *Théorie non abélienne des corps de classes pour les extensions finies et séparables des corps valués complets: principes fondamentaux; espaces de polynomes et transformation T ; lois d'unicité, d'ordination et d'existence*, C. R. Acad. Sci. Paris **222** (1946), 626–628.

[3] Alexander Ostrowski, *Über sogenannte perfekte Körper*, J. Reine Angew. Math. **147** (1917), 191–204

12 The different and the discriminant

12.1 The different

We continue in our usual *AKLB* setup: A is a Dedekind domain, K is its fraction field, L/K is a finite separable extension, and B is the integral closure of A in L (a Dedekind domain with fraction field L). We would like to understand the primes that ramify in L/K. Recall that a prime $\mathfrak{q}|\mathfrak{p}$ of L is unramified if and only if $e_{\mathfrak{q}} = 1$ and $B/\mathfrak{q}$ is a separable extension of $A/\mathfrak{p}$, equivalently, if and only if $B/\mathfrak{q}^{e_{\mathfrak{q}}}$ is a finite étale $A/\mathfrak{p}$ algebra (by Theorem 4.40).[1] A prime $\mathfrak{p}$ of K is unramified if and only if all the primes $\mathfrak{q}|\mathfrak{p}$ lying above it are unramified, equivalently, if and only if the ring $B/\mathfrak{p}B$ is a finite étale $A/\mathfrak{p}$ algebra.[2]

Our main tools for studying ramification are the *different* $\mathcal{D}_{B/A}$ and *discriminant* $D_{B/A}$. The different is a B-ideal that is divisible by precisely the ramified primes $\mathfrak{q}$ of L, and the discriminant is an A-ideal divisible by precisely the ramified primes $\mathfrak{p}$ of K. Moreover, the valuation $v_{\mathfrak{q}}(\mathcal{D}_{B/A})$ will give us information about the ramification index $e_{\mathfrak{q}}$ (its exact value when $\mathfrak{q}$ is tamely ramified).

Recall from Lecture 5 the trace pairing $L \times L \to K$ defined by $(x, y) \mapsto \mathrm{T}_{L/K}(xy)$; under our assumption that L/K is separable, it is a perfect pairing. An A-lattice M in L is a finitely generated A-module that spans L as a K-vector space (see Definition 5.9). Every A-lattice M in L has a *dual lattice* (see Definition 5.11)

$$M^* := \{x \in L : \mathrm{T}_{L/K}(xm) \in A \ \forall m \in M\},$$

which is an A-lattice in L isomorphic to the dual A-module $M^\vee := \mathrm{Hom}_A(M, A)$ (see Theorem 5.12). In our *AKLB* setting we have $M^{**} = M$, by Proposition 5.16.

Every fractional ideal I of B is finitely generated as a B-module, and therefore finitely generated as an A module (since B is finite over A). If I is nonzero, it necessarily spans L, since B does. It follows that every element of the group $\mathcal{I}_B$ of nonzero fractional ideals of B is an A-lattice in L. We now show that $\mathcal{I}_B$ is closed under the operation of taking duals.

Lemma 12.1. *Assume AKLB. If $I \in \mathcal{I}_B$ then $I^* \in \mathcal{I}_B$.*

Proof. The dual lattice I^* is a finitely generated A-module, thus to show that it is a finitely generated B-module it is enough to show it is closed under multiplication by elements of B. So consider any $b \in B$ and $x \in I^*$. For all $m \in I$ we have $\mathrm{T}_{L/K}((bx)m) = \mathrm{T}_{L/K}(x(bm)) \in A$, since $x \in I^*$ and $bm \in I$, so $bx \in I^*$ as desired. □

Definition 12.2. Assume *AKLB*. The *different* $\mathcal{D}_{L/K}$ of L/K (and the *different* $\mathcal{D}_{B/A}$ of B/A), is the inverse of B^* in $\mathcal{I}_B$. Explicitly, we have

$$B^* := \{x \in L : \mathrm{T}_{L/K}(xb) \in A \text{ for all } b \in B\},$$

and we define

$$\mathcal{D}_{L/K} := \mathcal{D}_{B/A} := (B^*)^{-1} = (B : B^*) = \{x \in L : xB^* \subseteq B\}.$$

Note that $B \subseteq B^*$, since $\mathrm{T}_{L/K}(ab) \in A$ for $a, b \in B$ (by Corollary 4.53), and this implies $\mathcal{D}_{B/A} = (B^*)^{-1} \subseteq B^{-1} = B$. Thus the different is an ideal, not just a fractional ideal.

[1] Note that $B/\mathfrak{q}^{e_{\mathfrak{q}}}$ is reduced if and only if $e_{\mathfrak{q}} = 1$; consider the image of a uniformizer in $B/\mathfrak{q}^{e_{\mathfrak{q}}}$.

[2] As usual, by a *prime* of A or K we mean a nonzero prime ideal of A, and similarly for B and L. The notation $\mathfrak{q}|\mathfrak{p}$ means that $\mathfrak{q}$ is a prime of B lying above $\mathfrak{p}$ (so $\mathfrak{p} = \mathfrak{q} \cap A$ and $\mathfrak{q}$ divides $\mathfrak{p}B$).

The different respects localization and completion.

Proposition 12.3. *Assume AKLB and let S be a multiplicative subset of A. Then*

$$S^{-1}\mathcal{D}_{B/A} = \mathcal{D}_{S^{-1}B/S^{-1}A}.$$

Proof. This follows from the fact that inverses and duals are both compatible with localization, by Lemmas 3.5 and 5.15. □

Proposition 12.4. *Assume AKLB and let $\mathfrak{q}|\mathfrak{p}$ be a prime of B. Then*

$$\mathcal{D}_{\hat{B}_\mathfrak{q}/\hat{A}_\mathfrak{p}} = \mathcal{D}_{B/A}\hat{B}_\mathfrak{q},$$

where $\hat{A}_\mathfrak{p}$ and $\hat{B}_\mathfrak{q}$ are the completions of A and B at $\mathfrak{p}$ and $\mathfrak{q}$, respectively.

Proof. Let $\hat{L} := L \otimes K_\mathfrak{p}$ be the base change of the finite étale K-algebra L to $K_\mathfrak{p}$. By (5) of Theorem 11.23, we have $\hat{L} \simeq \prod_{\mathfrak{q}|\mathfrak{p}} L_\mathfrak{q}$. Note that even though $\hat{L}$ need not be a field, in general, is is a free $K_\mathfrak{p}$-module of finite rank, and is thus equipped with a trace map that necessarily satisfies $\mathrm{T}_{\hat{L}/K_\mathfrak{p}}(x) = \sum_{\mathfrak{q}|\mathfrak{p}} \mathrm{T}_{\hat{L}/K_\mathfrak{p}}(x)$ that defines a trace pairing on $\hat{L}$.

Now let $\hat{B} := B \otimes \hat{A}_\mathfrak{p}$; it is an $A_\mathfrak{p}$-lattice in the $K_\mathfrak{p}$-vector space $\hat{L}$. By Corollary 11.26, $\hat{B} \simeq \prod_{\mathfrak{q}|\mathfrak{p}} \hat{B}_\mathfrak{q} \simeq \bigoplus_{\mathfrak{q}|\mathfrak{p}} \hat{B}_\mathfrak{q}$, and therefore $\hat{B}^* \simeq \bigoplus_{\mathfrak{q}|\mathfrak{p}} \hat{B}_\mathfrak{q}^*$, by Corollary 5.13. It follows that $\hat{B}^* \simeq B^* \otimes_A \hat{A}_\mathfrak{p}$. In particular, B^* generates each fractional ideal $\hat{B}_\mathfrak{q}^* \in \mathcal{I}_{\hat{B}_\mathfrak{q}}$. Taking inverses, $\mathcal{D}_{B/A} = (B^*)^{-1}$ generates the $\hat{B}_\mathfrak{q}$-ideal $(\hat{B}_\mathfrak{q}^*)^{-1} = \mathcal{D}_{\hat{B}_\mathfrak{q}/\hat{A}_\mathfrak{p}}$. □

12.2 The discriminant

Definition 12.5. Let S/R be a ring extension in which S is a free R-module of rank n. For any $x_1, \dots, x_n \in S$ we define the *discriminant*

$$\mathrm{disc}(x_1, \dots, x_n) := \mathrm{disc}_{S/R}(x_1, \dots, x_n) := \det[\mathrm{T}_{S/R}(x_i x_j)]_{i,j} \in R.$$

Note that we do not require $x_1, \dots, x_n$ to be an R-basis for S, but if they satisfy a non-trivial R-linear relation then the discriminant will be zero (by linearity of the trace).

In our *AKLB* setup, we have in mind the case where $e_1, \dots, e_n \in B$ is a basis for L as a K-vector space, in which case $\mathrm{disc}(e_1, \dots, e_n) = \det[\mathrm{T}_{L/K}(e_i e_j)]_{ij} \in A$. Note that we do not need to assume that B is a free A-module; L is certainly a free K-module. The fact that the discriminant lies in A when $e_1, \dots, e_n \in B$ follows immediately from Corollary 4.53.

Proposition 12.6. *Let L/K be a finite separable extension of degree n, and let Ω/K be a field extension for which there are distinct $\sigma_1, \dots, \sigma_n \in \mathrm{Hom}_K(L, \Omega)$. For any $e_1, \dots, e_n \in L$ we have*

$$\mathrm{disc}(e_1, \dots, e_n) = \det[\sigma_i(e_j)]_{ij}^2,$$

and for any $x \in L$ we have

$$\mathrm{disc}(1, x, x^2, \dots, x^{n-1}) = \prod_{i<j} (\sigma_i(x) - \sigma_j(x))^2.$$

Such a field extension Ω/K always exists, since L/K is separable ($\Omega = K^{\mathrm{sep}}$ works).

Proof. For $1 \le i, j \le n$ we have $\mathrm{T}_{L/K}(e_i e_j) = \sum_{k=1}^{n} \sigma_k(e_i e_j)$, by Theorem 4.50. Therefore

$$\begin{aligned}\operatorname{disc}(e_1, \dots, e_n) &= \det[\mathrm{T}_{L/K}(e_i e_j)]_{ij} \\ &= \det\left([\sigma_k(e_i)]_{ik}[\sigma_k(e_j)]_{kj}\right) \\ &= \det\left([\sigma_k(e_i)]_{ik}[\sigma_k(e_j)]_{jk}^{\mathrm{t}}\right) \\ &= \det[\sigma_i(e_j)]_{ij}^2\end{aligned}$$

since the determinant is multiplicative and $\det M = \det M^{\mathrm{t}}$ for any matrix M.

Now let $x \in L$ and put $e_i := x^{i-1}$ for $1 \le i \le n$. Then

$$\operatorname{disc}(1, x, x^2, \dots, x^{n-1}) = \det[\sigma_i(x^{j-1})]_{ij}^2 = \det[\sigma_i(x)^{j-1}]_{ij}^2 = \prod_{i<j} (\sigma_i(x) - \sigma_j(x))^2,$$

since $[\sigma_i(x)^{j-1}]_{ij}$ is a Vandermonde matrix (rows of the form $z^0, \dots, z^{n-1}$ for some z); see [2, p. 258] for a proof of this standard fact. □

Definition 12.7. For a polynomial $f(x) = \prod_i (x - \alpha_i)$, the *discriminant* of f is

$$\operatorname{disc}(f) := \prod_{i<j} (\alpha_i - \alpha_j)^2.$$

Equivalently, if A is a Dedekind domain, $f \in A[x]$ is a monic separable polynomial, and α is the image of x in $A[x]/(f(x))$, then

$$\operatorname{disc}(f) = \operatorname{disc}(1, \alpha, \alpha^2, \dots, \alpha^{n-1}) \in A.$$

Example 12.8. $\operatorname{disc}(x^2 + bx + c) = b^2 - 4c$ and $\operatorname{disc}(x^3 + ax + b) = -4a^3 - 27b^2$.

Now assume $AKLB$ and let M be an A-lattice in L. Then M is a finitely generated A-module that contains a K-basis for L. We want to define the discriminant of M in a way that does not require us to choose a basis.

Let us first consider the case where M is a free A-lattice. If $e_1, \dots, e_n \in M \subseteq L$ and $e'_1, \dots, e'_n \in M \subseteq L$ are two A-bases for M, then

$$\operatorname{disc}(e'_1, \dots, e'_n) = u^2 \operatorname{disc}(e_1, \dots, e_n)$$

for some unit $u \in A^\times$; this follows from the fact that the change of basis matrix $P \in A^{n \times n}$ is invertible and its determinant is therefore a unit u. This unit gets squared because we need to apply the change of basis matrix twice in order to change $\mathrm{T}(e_i e_j)$ to $\mathrm{T}(e'_i e'_j)$. Explicitly, writing bases as row-vectors, let $e = (e_1, \dots, e_n)$ and $e' = (e'_1, \dots, e'_n)$ satisfy $e' = eP$. Then

$$\begin{aligned}\operatorname{disc}(e') &= \det[\mathrm{T}_{L/K}(e'_i e'_j)]_{ij} \\ &= \det[\mathrm{T}_{L/K}((eP)_i (eP)_j)]_{ij} \\ &= \det[P^{\mathrm{t}}[\mathrm{T}_{L/K}(e_i e_j)]_{ij} P] \\ &= (\det P^{\mathrm{t}}) \operatorname{disc}(e) (\det P) \\ &= (\det P)^2 \operatorname{disc}(e),\end{aligned}$$

where we have used the linearity of $\mathrm{T}_{L/K}$ to go from the second equality to the third.

This actually gives us a basis independent definition when $A = \mathbb{Z}$. In this case B is always a free $\mathbb{Z}$-lattice, and the only units in $\mathbb{Z}$ are $u = \pm 1$, so $u^2 = 1$.

Definition 12.9. Assume *AKLB*, let M be an A-lattice in L, and let $n \coloneqq [L:K]$. The *discriminant* $D(M)$ of M is the A-module generated by $\{\operatorname{disc}(x_1,\dots,x_n) : x_1,\dots,x_n \in M\}$.

Lemma 12.10. *Assume AKLB and let $M' \subseteq M$ be free A-lattices in L. The discriminants $D(M') \subseteq D(M)$ are nonzero principal fractional ideals. If $D(M') = D(M)$ then $M' = M$.*

Proof. Let $e \coloneqq (e_1,\dots,e_n)$ be an A-basis for M. Then $\operatorname{disc}(e) \in D(M)$, and for any row vector $x \coloneqq (x_1,\dots,x_n)$ with entries in M there is a matrix $P \in A^{n\times n}$ for which $x = eP$, and we then have $\operatorname{disc}(x) = (\det P)^2 \operatorname{disc}(e)$ as above. It follows that

$$D(M) = (\operatorname{disc}(e))$$

is principal, and it is nonzero because e is a basis for L and the trace pairing is nondegenerate. If we now let $e' \coloneqq (e'_1,\dots,e'_n)$ be an A-basis for M' then $D(M') = (\operatorname{disc}(e'))$ is also a nonzero and principal. Our assumption that $M' \subseteq M$ implies that $e' = eP$ for some matrix $P \in A^{n\times n}$, and we have $\operatorname{disc}(e') = (\det P)^2 \operatorname{disc}(e)$. If $D(M') = D(M)$ then $\det P$ must be a unit, in which case P is invertible and $e = e'P^{-1}$. This implies $M \subseteq M'$, so $M' = M$. □

Proposition 12.11. *Assume AKLB and let M be an A-lattice in L. Then $D(M) \in \mathcal{I}_A$.*

Proof. The A-module $D(M) \subseteq K$ is nonzero because M contains a K-basis $e = (e_1,\dots,e_n)$ for L and $\operatorname{disc}(e) \neq 0$ because the trace pairing is nondegenerate. To show that $D(M)$ is a finitely generated A-module (and thus a fractional ideal), we use the usual trick: make it a submodule of a noetherian module. So let N be the free A-lattice in L generated by e and then pick a nonzero $a \in A$ such that $M \subseteq a^{-1}N$ (write each generator for M in terms of the K-basis e and let a be the product of all the denominators that appear; note that M is finitely generated). We then have $D(M) \subseteq D(a^{-1}N)$, and $D(a^{-1}N)$ is a principal fractional ideal of A, hence a noetherian A-module (since A is noetherian), so its submodule $D(M)$ must be finitely generated. □

Definition 12.12. Assume *AKLB*. The *discriminant* $D_{L/K}$ of L/K (and the *discriminant* $D_{B/A}$ of B/A) is the discriminant of B as an A-module:

$$D_{L/K} \coloneqq D_{B/A} \coloneqq D(B) \in \mathcal{I}_A,$$

which is an A-ideal, since $\operatorname{disc}(x_1,\dots,x_n) = \det[T_{B/A}(x_i x_j)]_{i,j} \in A$ for all $x_1,\dots,x_n \in B$.

Example 12.13. Consider the case $A = \mathbb{Z}$, $K = \mathbb{Q}$, $L = \mathbb{Q}(i)$, $B = \mathbb{Z}[i]$. Then B is a free A-lattice with basis $(1, i)$ and we can compute $D_{L/K}$ in three ways:

- $\operatorname{disc}(1,i) = \det \begin{bmatrix} \mathrm{T}_{L/K}(1\cdot 1) & \mathrm{T}_{L/K}(1\cdot i) \\ \mathrm{T}_{L/K}(i\cdot 1) & \mathrm{T}_{L/K}(i\cdot i) \end{bmatrix} = \det \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} = -4.$
- The non-trivial automorphism of L/K fixes 1 and sends i to $-i$, so we could instead compute $\operatorname{disc}(1,i) = \left(\det \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}\right)^2 = (-2i)^2 = -4.$
- We have $B = \mathbb{Z}[i] = \mathbb{Z}[x]/(x^2+1)$ and can compute $\operatorname{disc}(x^2+1) = -4$.

In every case the discriminant $D_{L/K}$ is the ideal $(-4) = (4)$.

Remark 12.14. If $A = \mathbb{Z}$ then B is the ring of integers of the number field L, and B is a free A-lattice, because it is a torsion-free module over a PID and therefore a free module. In this situation it is customary to define the *absolute discriminant* D_L of the number field L to be the *integer* $\operatorname{disc}(e_1, \dots, e_n) \in \mathbb{Z}$, for any basis $(e_1, \dots, e_n)$ of B, rather than the ideal it generates. As noted above, this integer is independent of the choice of basis because $u^2 = 1$ for all $u \in \mathbb{Z}^\times$; in particular, the sign of D_L is well defined (as we shall see, the sign of D_L carries information about L). In the example above, the absolute discriminant is $D_L = -4$.

Like the different, the discriminant respects localization.

Proposition 12.15. *Assume AKLB and let S be a multiplicative subset of A. Then*

$$S^{-1} D_{B/A} = D_{S^{-1}B/S^{-1}A}.$$

Proof. Let $x = s^{-1} \operatorname{disc}(e_1, \dots, e_n) \in S^{-1} D_{B/A}$ for some $s \in S$ and $e_1, \dots, e_n \in B$. Then $x = s^{2n-1} \operatorname{disc}(s^{-1}e_1, \dots, s^{-1}e_n)$ lies in $D_{S^{-1}B/S^{-1}A}$. This proves the forward inclusion.

Conversely, for any $e_1, \dots, e_n \in S^{-1}B$ we can choose a single $s \in S \subseteq A$ so that each se_i lies in B. We then have $\operatorname{disc}(e_1, \dots, e_n) = s^{-2n} \operatorname{disc}(se_1, \dots, se_n) \in S^{-1} D_{B/A}$, which proves the reverse inclusion. □

Proposition 12.16. *Assume AKLB and let $\mathfrak{p}$ be a prime of A. Then*

$$D_{B/A} \hat{A}_\mathfrak{p} = \prod_{\mathfrak{q}|\mathfrak{p}} D_{\hat{B}_\mathfrak{q}/\hat{A}_\mathfrak{p}}$$

where $\hat{A}_\mathfrak{p}$ and $\hat{B}_\mathfrak{q}$ are the completions of A and B at $\mathfrak{p}$ and $\mathfrak{q}$, respectively.

Proof. After localizing at $\mathfrak{p}$ we can assume A is a DVR and B is a free A-module of rank n. As in the proof of Proposition 12.4, we have a trace pairing on the finite étale $K_\mathfrak{p}$-algebra $\hat{L} \coloneqq L \otimes K_\mathfrak{p}$ and $\hat{B} \coloneqq B \otimes \hat{A}_\mathfrak{p} \simeq \bigoplus_{\mathfrak{q}|\mathfrak{p}} \hat{B}_\mathfrak{q}$ is an $\hat{A}_\mathfrak{p}$-lattice in the $K_\mathfrak{p}$-vector space $\hat{L}$ that is a direct sum of free $\hat{A}_\mathfrak{p}$-modules, and thus a free $\hat{A}_\mathfrak{p}$-module of rank $n = \sum e_\mathfrak{q} f_\mathfrak{q}$; see Corollary 11.26.

We can choose $\hat{A}_\mathfrak{p}$ bases for each $\hat{B}_\mathfrak{q}$ using elements in B; this follows from weak approximation (Theorem 8.5) and the fact that B is dense in $\hat{B}_\mathfrak{q}$ (or see [1, Thm. 2.3]). From these bases we can construct an $\hat{A}_\mathfrak{p}$-basis $\hat{e}$ for the direct sum $\bigoplus_{\mathfrak{q}|\mathfrak{p}} \hat{B}_\mathfrak{q} \simeq \hat{B}$ whose elements each have nonzero projections to exactly one of the $\hat{B}_\mathfrak{q}$, along with a corresponding A-basis e for B obtained from $\hat{e}$ as the union of these projections.

The matrix $[T_{\hat{L}/K_\mathfrak{p}}(\hat{e}_i \hat{e}_j)]$ is block diagonal; each block corresponds to a matrix whose determinant is the discriminant of the $\hat{A}_\mathfrak{p}$-basis we chose for one of the $\hat{B}_\mathfrak{q}$. It follows that $D_{\hat{B}/\hat{A}_\mathfrak{p}} = \prod_{\mathfrak{q}|\mathfrak{p}} D_{\hat{B}_\mathfrak{q}/\hat{A}_\mathfrak{p}}$ (here we are using the fact that $\hat{B} \simeq \bigoplus_{\mathfrak{q}|\mathfrak{p}} \hat{B}_\mathfrak{q}$ is both an isomorphism of rings and an isomorphism of $A_\mathfrak{p}$-modules, hence it preserves traces to $\hat{A}_\mathfrak{p}$). We now observe that

$$\operatorname{disc}_{B/A}(e_1, \dots, e_n) = \operatorname{disc}_{(B \otimes A_\mathfrak{p})/\hat{A}_\mathfrak{p}}(e_1 \otimes 1, \dots, e_n \otimes 1)$$

generates $D_{B/A}$ as an A-ideal, and also generates $D_{\hat{B}/\hat{A}_\mathfrak{p}}$ as an $\hat{A}_\mathfrak{p}$-ideal (note that $\hat{B}$ is a free $\hat{A}_\mathfrak{p}$-module, so $D_{B/\hat{A}_\mathfrak{p}}$ is the principal ideal generated by the discriminant of any $A_\mathfrak{p}$-basis for $\hat{B}$). It follows that $D_{B/A}\hat{A}_\mathfrak{p} = D_{\hat{B}/\hat{A}_\mathfrak{p}} = \prod_{\mathfrak{q}|\mathfrak{p}} D_{\hat{B}_\mathfrak{q}/\hat{A}_\mathfrak{p}}$. □

We have defined two different ideals associated to a finite separable extension of Dedekind domains B/A in the $AKLB$ setup. We have the different $\mathcal{D}_{B/A}$, which is a fractional ideal of B, and the discriminant $D_{B/A}$, which is a fractional ideal of A. We now relate these two ideals in terms of the ideal norm $N_{B/A}\colon \mathcal{I}_B \to \mathcal{I}_A$, which for $I \in \mathcal{I}_B$ is defined as $\mathrm{N}_{B/A}(I) := [B : I]_A$, where $[B : I]_A$ is the module index (see Definitions 6.1 and 6.5).

Theorem 12.17. *Assume $AKLB$. Then $D_{B/A} = N_{B/A}(\mathcal{D}_{B/A})$.*

Proof. The different and discriminant are both compatible with localization, by Propositions 12.3 and 12.15, and the A-modules $D_{B/A}$ and $N_{B/A}(\mathcal{D}_{B/A})$ of A are both determined by the intersections of their localizations at maximal ideals (Proposition 2.6), so it suffices to prove that the theorem holds when we replace A by its localization A at a prime of A. Then A is a DVR and B is a free A-lattice in L; let us fix an A-basis $(e_1, \dots, e_n)$ for B.

The dual A-lattice

$$B^* = \{x \in L : \mathrm{T}_{L/K}(xb) \in A \ \forall b \in B\} \in \mathcal{I}_B$$

is also a free A-lattice in L, with basis $(e_1^*, \dots, e_n^*)$ uniquely determined by $\mathrm{T}_{L/K}(e_i^* e_j) = \delta_{ij}$, where δ_{ij} is the Kronecker delta function; see Corollary 5.14. If we write $e_i = \sum a_{ij} e_j^*$ in terms of the K-basis $(e_1^*, \dots, e_n^*)$ for L then

$$T_{L/K}(e_i e_j) = T_{L/K}\left(\sum_k a_{ik} e_k^* e_j\right) = \sum_k a_{ik} T_{L/K}(e_k^* e_j) = \sum_k a_{ik}\delta_{kj} = a_{ij}.$$

It follows that $P := [\mathrm{T}_{L/K}(e_i e_j)]_{ij}$ is the change-of-basis matrix from $e^* := (e_1^*, \dots, e_n^*)$ to $e := (e_1, \dots, e_n)$ (as row vectors we have $e = e^* P$). If we let ϕ denote the K-linear transformation with matrix P (or its transpose, if you prefer to work with column vectors), then ϕ is an isomorphism of free A-modules and

$$D_{B/A} = \big(\det[\mathrm{T}_{L/K}(e_i e_j)]_{ij}\big) = (\det \phi) = [B^* : B]_A,$$

where $[B^* : B]_A$ is the module index (see Definition 6.1). Applying Corollary 6.8 yields

$$D_{B/A} = [B^* : B]_A = N_{B/A}((B : B^*)) = N_{B/A}((B^*)^{-1}) = N_{B/A}(\mathcal{D}_{B/A}).$$

(the last three equalities each hold by definition). □

12.3 Ramification

Having defined the different and discriminant ideals we now want to understand how they relate to ramification. Recall that in our $AKLB$ setup, if $\mathfrak{p}$ is a prime of A then we can factor the B-ideal $\mathfrak{p}B$ as

$$\mathfrak{p}B = \mathfrak{q}_1^{e_1} \cdots \mathfrak{q}_r^{e_r}.$$

The Chinese remainder theorem implies

$$B/\mathfrak{p}B \simeq B/\mathfrak{q}_1^{e_1} \times \cdots \times B/\mathfrak{q}_r^{e_r}.$$

This is a commutative $A/\mathfrak{p}$-algebra of dimension $\sum e_i f_i$, where $f_i = [B/\mathfrak{q}_i : A/\mathfrak{p}]$ is the residue degree (see Theorem 5.35). It is a product of fields if and only if we have $e_i = 1$ for all i, and it is a finite étale-algebra if and only if it is a product of fields that are separable extensions of $A/\mathfrak{p}$. The following lemma relates the discriminant to the property of being a finite étale algebra.

Lemma 12.18. *Let k be a field and let R be a commutative k-algebra with k-basis $r_1, \dots, r_n$. Then R is a finite étale k-algebra if and only if* $\operatorname{disc}(r_1, \dots, r_n) \neq 0$.

Proof. By Theorem 5.20, R is a finite étale k-algebra if and only if the trace pairing on R is a perfect pairing, which is equivalent to being nondegenerate, since k is a field.

If the trace pairing is degenerate then for some nonzero $x \in R$ we have $\mathrm{T}_{R/k}(xy) = 0$ for all $y \in R$. If we write $x = \sum_i x_i r_i$ with $x_i \in k$ then $\mathrm{T}_{R/k}(xr_j) = \sum_i x_i \mathrm{T}_{R/k}(r_i r_j) = 0$ for all r_j (take $y = r_j$), and this implies that the columns of the matrix $[\mathrm{T}_{R/k}(r_i r_j)]_{ij}$ are linearly dependent and $\operatorname{disc}(r_1, \dots, r_n) = \det[\mathrm{T}_{R/k}(r_i r_j)]_{ij} = 0$.

Conversely, if $\operatorname{disc}(r_1, \dots, r_n) = 0$ then the columns of $\det[\mathrm{T}_{R/k}(r_i r_j)]_{ij}$ are linearly dependent and for some $x_i \in k$ not identically zero we must have $\sum_i x_i \mathrm{T}_{R/k}(r_i r_j) = 0$ for all j. For $x \coloneqq \sum_i x_i r_i$ and any $y = \sum_j y_j r_j \in R$ we have $\mathrm{T}_{R/k}(xy) = \sum_j y_j \sum_i x_i \mathrm{T}_{R/k}(r_i r_j) = 0$, which shows that the trace pairing is degenerate. □

Theorem 12.19. *Assume AKLB, let $\mathfrak{q}$ be a prime of B lying above a prime $\mathfrak{p}$ of A such that $B/\mathfrak{q}$ is a separable extension of $A/\mathfrak{p}$. The extension L/K is unramified at $\mathfrak{q}$ if and only if $\mathfrak{q}$ does not divide $\mathcal{D}_{B/A}$, and it is unramified at $\mathfrak{p}$ if and only if $\mathfrak{p}$ does not divide $D_{B/A}$.*

Proof. We first consider the different $\mathcal{D}_{B/A}$. By Proposition 12.4, the different is compatible with completion, so it suffices to consider the case that A and B are complete DVRs (complete K at $\mathfrak{p}$ and L at $\mathfrak{q}$ and apply Theorem 11.23). We then have $[L : K] = e_{\mathfrak{q}} f_{\mathfrak{q}}$, where $e_{\mathfrak{q}}$ is the ramification index and $f_{\mathfrak{q}}$ is the residue field degree, and $\mathfrak{p}B = \mathfrak{q}^{e_{\mathfrak{q}}}$.

Since B is a DVR with maximal ideal $\mathfrak{q}$, we must have $\mathcal{D}_{B/A} = \mathfrak{q}^m$ for some $m \geq 0$. By Theorem 12.17 we have

$$D_{B/A} = N_{B/A}(\mathcal{D}_{B/A}) = N_{B/A}(\mathfrak{q}^m) = \mathfrak{p}^{f_{\mathfrak{q}} m}.$$

Thus $\mathfrak{q} | \mathcal{D}_{B/A}$ if and only if $\mathfrak{p} | D_{B/A}$. Since A is a PID, B is a free A-module and we may choose an A-module basis $e_1, \dots, e_n$ for B that is also a K-basis for L. Let $k \coloneqq A/\mathfrak{p}$, and let $\overline{e}_i$ be the reduction of e_i to the k-algebra $R \coloneqq B/\mathfrak{p}B$. Then $(\overline{e}_1, \dots, \overline{e}_n)$ is a k-basis for R: it clearly spans, and we have $[R : k] = [B/\mathfrak{q}^{e_{\mathfrak{q}}} : A/\mathfrak{p}] = e_{\mathfrak{q}} f_{\mathfrak{q}} = [L : K] = n$.

Since B has an A-module basis, we may compute its discriminant as

$$D_{B/A} = (\operatorname{disc}(e_1, \dots, e_n)).$$

Thus $\mathfrak{p} | D_{B/A}$ if and only if $\operatorname{disc}(e_1, \dots, e_n) \in \mathfrak{p}$, equivalently, $\operatorname{disc}(\overline{e}_1, \dots, \overline{e}_n) = 0$ (note that $\operatorname{disc}(e_1, \dots, e_n)$ is a polynomial in the $\mathrm{T}_{L/K}(e_i e_j)$ and $T_{R/k}(\overline{e}_i \overline{e}_j)$ is the trace of the multiplication-by-$\overline{e}_i \overline{e}_j$ map, which is the same as the reduction to $k = A/\mathfrak{p}$ of the trace of the multiplication-by-$e_i e_j$ map $T_{L/K}(e_i e_j) \in A$). By Lemma 12.18, $\operatorname{disc}(\overline{e}_1, \dots, \overline{e}_n) = 0$ if and only if the k-algebra $B/\mathfrak{p}B$ is not finite étale, equivalently, if and only if $\mathfrak{p}$ is ramified. Thus $\mathfrak{p} | D_{B/A}$ if and only if $\mathfrak{p}$ is ramified. There is only one prime $\mathfrak{q}$ above $\mathfrak{p}$, so we also have $\mathfrak{q} | \mathcal{D}_{B/A}$ if and only if $\mathfrak{q}$ is ramified. □

We now note an important corollary of Theorem 12.19.

Corollary 12.20. *Assume AKLB. Only finitely many primes of A (or B) ramify.*

Proof. A and B are Dedekind domains, so the ideals $D_{B/A}$ and $\mathcal{D}_{B/A}$ both have unique factorizations into prime ideals in which only finitely many primes appear. □

12.4 The discriminant of an order

Recall from Lecture 6 that an order $\mathcal{O}$ is a noetherian domain of dimension one whose conductor is nonzero (see Definitions 6.16 and 6.19), and the integral closure of an order is always a Dedekind domain. In our *AKLB* setup, the orders with integral closure B are precisely the A-lattices in L that are rings (see Proposition 6.22); if $L = K(\alpha)$ with $\alpha \in B$, then $A[\alpha]$ is an example. The discriminant $D_{\mathcal{O}/A}$ of such an order $\mathcal{O}$ is its discriminant $D(\mathcal{O})$ as an A-module. The fact that $\mathcal{O} \subseteq B$ implies that $D(\mathcal{O}) \subseteq D_{B/A}$ is an A-ideal.

If $\mathcal{O}$ is an order of the form $A[\alpha]$, where $\alpha \in B$ generates $L = K(\alpha)$ with minimal polynomial $f \in A[x]$, then $\mathcal{O}$ is a free A-lattice with basis $1, \alpha, \dots, \alpha^{n-1}$, where $n = \deg f$, and we may compute its discriminant as

$$D_{\mathcal{O}/A} = (\operatorname{disc}(1, \alpha, \dots, \alpha^{n-1})) = (\operatorname{disc}(f)),$$

which is a principal A-ideal contained in $D_{B/A}$. If B is also a free A-lattice, then as in the proof of Lemma 12.10 we have

$$D_{\mathcal{O}/A} = (\det P)^2 D_{B/A} = [B:\mathcal{O}]_A^2 D_{B/A},$$

where P is the matrix of the A-linear map $\phi\colon B \to \mathcal{O}$ that sends an A-basis for B to an A-basis for $\mathcal{O}$ and $[B:\mathcal{O}]_A$ is the module index (a principal A-ideal).

In the important special case where $A = \mathbb{Z}$ and L is a number field, the integer $(\det P)^2$ is uniquely determined and it necessarily divides $\operatorname{disc}(f)$, the generator of the principal ideal $D(\mathcal{O}) = D(A[\alpha])$. It follows that if $\operatorname{disc}(f)$ is squarefree then we must have $B = \mathcal{O} = A[\alpha]$. More generally, any prime p for which $v_p(\operatorname{disc}(f))$ is odd must be ramified, and any prime that does not divide $\operatorname{disc}(f)$ must be unramified. Another useful observation that applies when $A = \mathbb{Z}$: the module index $[B:\mathcal{O}]_{\mathbb{Z}} = ([B:\mathcal{O}])$ is the principal ideal generated by the index of $\mathcal{O}$ in B (as $\mathbb{Z}$-lattices), and we have the relation

$$D_{\mathcal{O}} = [B:\mathcal{O}]^2 D_B$$

between the absolute discriminant of the order $\mathcal{O}$ and its integral closure B.

Example 12.21. Consider $A = \mathbb{Z}$, $K = \mathbb{Q}$ with $L = \mathbb{Q}(\alpha)$, where $\alpha^3 - \alpha - 1 = 0$. We can compute the absolute discriminant of $\mathbb{Z}[\alpha]$ as

$$\operatorname{disc}(1, \alpha, \alpha^2) = \operatorname{disc}(x^3 - x - 1) = -4(-1)^3 - 27(-1)^2 = -23.$$

The fact that -23 is squarefree immediately implies that 23 is the only prime of A that ramifies, and we have $D_{\mathbb{Z}[\alpha]} = -23 = [\mathcal{O}_L : \mathbb{Z}[\alpha]]^2 D_L$, which forces $[\mathcal{O}_L : \mathbb{Z}[\alpha]] = 1$, so $D_L = -23$ and $\mathcal{O}_L = \mathbb{Z}[\alpha]$.

More generally, we have the following theorem.

Theorem 12.22. *Assume AKLB and let $\mathcal{O}$ be an order with integral closure B and conductor $\mathfrak{c}$. Then $D_{\mathcal{O}/A} = \mathrm{N}_{B/A}(\mathfrak{c}) D_{B/A}$.*

Proof. See Problem Set 6. □

12.5 Computing the discriminant and different

We conclude with a number of results that allow one to explicitly compute the discriminant and different in many cases.

Proposition 12.23. *Assume AKLB. If $B = A[\alpha]$ for some $\alpha \in L$ and $f \in A[x]$ is the minimal polynomial of α, then*

$$\mathcal{D}_{B/A} = (f'(\alpha))$$

is the B-ideal generated by $f'(\alpha)$.

Proof. See Problem Set 6. □

The assumption $B = A[\alpha]$ in Proposition 12.23 does not always hold, but if we want to compute the power of $\mathfrak{q}$ that divides $\mathcal{D}_{B/A}$ we can complete L at $\mathfrak{q}$ and K at $\mathfrak{p} = \mathfrak{q} \cap A$ so that A and B become complete DVRs, in which case $B = A[\alpha]$ does hold (by Lemma 10.14), so long as the residue field extension is separable (always true if K and L are global fields, since the residue fields are then finite, hence perfect). The following definition and proposition give an alternative approach.

Definition 12.24. Assume $AKLB$ and let $\alpha \in B$ have minimal polynomial $f \in A[x]$. The *different of α* is defined by

$$\delta_{B/A}(\alpha) := \begin{cases} f'(\alpha) & \text{if } L = K(\alpha), \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 12.25. *Assume AKLB. Then $\mathcal{D}_{B/A} = \big(\delta_{B/A}(\alpha) : \alpha \in B\big)$.*

Proof. See [3, Thm. III.2.5]. □

We can now more precisely characterize the ramification information given by the different ideal.

Theorem 12.26. *Assume AKLB and let $\mathfrak{q}$ be a prime of L lying above $\mathfrak{p} = \mathfrak{q} \cap A$ for which the residue field extension $(B/\mathfrak{q})/(A/\mathfrak{p})$ is separable. Then*

$$e_{\mathfrak{q}} - 1 \le v_{\mathfrak{q}}(\mathcal{D}_{B/A}) \le e_{\mathfrak{q}} - 1 + v_{\mathfrak{q}}(e_{\mathfrak{q}}),$$

and the lower bound is an equality if and only if $\mathfrak{q}$ is tamely ramified.

Proof. See Problem Set 6. □

We also note the following proposition, which shows how the discriminant and different behave in a tower of extensions.

Proposition 12.27. *Assume AKLB and let M/L be a finite separable extension and let C be the integral closure of A in M. Then*

$$\mathcal{D}_{C/A} = \mathcal{D}_{C/B} \cdot \mathcal{D}_{B/A}$$

(where the product on the right is taken in C), and

$$D_{C/A} = (D_{B/A})^{[M:L]} N_{B/A}(D_{C/B}).$$

Proof. See [4, Prop. III.8]. □

If $M/L/K$ is a tower of finite separable extensions, we note that the primes $\mathfrak{p}$ of K that ramify are precisely those that divide either $D_{L/K}$ or $N_{L/K}(D_{M/L})$.

References

[1] A. Jakhar, B. Jhorar, S. K. Khanduja, N. Sangwan, *Discriminant as a product of local discriminant*, J. Algebra App. **16** (2017), 1750198 (7 pages).

[2] S. Lang, *Algebra*, third edition, Springer, 2002.

[3] J. Neukirch, *Algebraic number theory*, Springer, 1999.

[4] J.-P. Serre, *Local fields*, Springer, 1979.

[5] Stacks Project Authors, *Stacks Project*, `http://stacks.math.columbia.edu`.

13 Global fields and the product formula

Up to this point we have defined global fields as finite extensions of $\mathbb{Q}$ (number fields) or of $\mathbb{F}_q(t)$ (global function fields). Our goal in this lecture is to prove a generalization of the product formula that you proved on Problem Set 1 for $K = \mathbb{Q}$ and $K = \mathbb{F}_q(t)$, which will then allow us to give a more natural definition of global fields: they are fields whose completions are local fields and which satisfy a suitable product formula.

13.1 Places of a field

Definition 13.1. Let K be a field. A *place* of K is an equivalence class of nontrivial absolute values on K. Recall that the completion of K at an absolute value depends only on its equivalence class, so there is a one-to-one correspondence between places of K and completions of K. We may use M_K to denote the set of places of K, and for each place v we use $|\ |_v$ to denote any representative absolute and K_v to denote the completion of K with respect to $|\ |_v$ (this does not depend on the choice of $|\ |_v$). We call a place v *archimedean* when K_v is archimedean and *nonarchimedean* otherwise.

Now let K be a global field. By Corollary 9.7, for any place v of K the completion K_v is a local field. From our classification of local fields (Theorem 9.9), if K_v is archimedean then $K_v \simeq \mathbb{R}$ or $K_v \simeq \mathbb{C}$, and otherwise the absolute value of K_v is induced by a discrete valuation that we also denote v; note that while the absolute value $|x|_v := c^{-v(x)}$ depends on a choice of $c \in [0,1]$, the discrete valuation $v\colon K_v \to \mathbb{Z}$ is uniquely determined. We now introduce the following terminology:

- if $K_v \simeq \mathbb{R}$ then v is a *real place*;
- if $K_v \simeq \mathbb{C}$ then v is a *complex place*;
- if $|\ |_v$ is induced by a discrete valuation $v_\mathfrak{p}$ corresponding to a prime $\mathfrak{p}$ of K then v is a *finite place*; otherwise v is an *infinite place*.

Every finite place is nonarchimedean. Infinite places are archimedean in characteristic zero and nonarchimedean otherwise. Every archimedean place is an infinite place, but nonarchimedean places may be finite or infinite (the latter only in positive characteristic).

Example 13.2. As you proved on Problem Set 1, the set $M_\mathbb{Q}$ consists of finite places p corresponding to p-adic absolute values $|\ |_p$, and a single archimedean infinite place ∞ corresponding to the Euclidean absolute value $|\ |_\infty$. The set $M_{\mathbb{F}_q(t)}$ consist of finite places corresponding to irreducible polynomials in $\mathbb{F}_q[t]$ and a single nonarchimedean infinite place ∞ corresponding to the absolute value $|\cdot|_\infty := q^{\deg(\cdot)}$.

Remark 13.3. There is nothing special about the infinite place of $\mathbb{F}_q(t)$, it is an artifact of our choice of the separating element t, which we could change by applying an automorphism $t \mapsto (at+b)/(ct+d)$ of $\mathbb{F}_q(t)$. If we put $z := 1/t$ and consider $\mathbb{F}_q(z) \simeq \mathbb{F}_q(t)$, the absolute value $|\ |_\infty$ on $\mathbb{F}_q(t)$ is the same as the absolute value $|\ |_z$ on $\mathbb{F}_q(z)$ corresponding to the irreducible polynomial $z \in \mathbb{F}_q[z]$. This is analogous to the situation with the projective line $\mathbb{P}^1$, where we may distinguish the projective point $(1:0)$ as the "point at infinity", but this distinction is not invariant under automorphisms of $\mathbb{P}^1$.

Definition 13.4. If L/K is an extension of global fields, for every place w of L, any absolute value $|\ |_w$ that represents the equivalence class w restricts to an absolute value on K that represents a place v of K that is independent of the choice of $|\ |_w$. We write $w|v$ to indicate this relationship and say that w *extends* v or that w *lies above* v.

Theorem 13.5. *Let L/K be a finite separable extension of global fields and let v be a place of K. We have an isomorphism of finite étale K_v-algebras*

$$L \otimes_K K_v \xrightarrow{\sim} \prod_{w|v} L_w$$

defined by $\ell \otimes x \mapsto (\ell x, \dots, \ell x)$.

For nonarchimedean places this follows from part (v) of Theorem 11.23, but here we give a different proof that works for any place of K.

Proof. The separable extension L/K is a finite étale K-algebra, so the base change $L \otimes_K K_v$ is a finite étale K_v-algebra, by Proposition 4.36, and is therefore isomorphic to a finite product $\prod_{i\in I} L_i$ of finite separable extensions L_i of K_v, each of which is a local field (any finite extension of a local field is a local field). We just need to show that there is a one-to-one correspondence between the sets of local fields $\{L_i : i \in I\}$ and $\{L_w : w|v\}$.

Let us fix an absolute value $|\ |_v$ on K_v representing the place v. Each L_i is a local field extending K_v, and therefore has a unique absolute value $|\ |_w$ that restricts to $|\ |_v$; this follows from Theorem 10.6 when v is nonarchimedean and is obvious when v is archimedean, since then either $K_v \simeq L_w$ or $K_v \simeq \mathbb{R} \subseteq \mathbb{C} \simeq L_w$ and the Euclidean absolute value on $\mathbb{R}$ is the restriction of the Euclidean absolute value on $\mathbb{C}$. The map $L \hookrightarrow L \otimes_K K_v \simeq \prod_i L_i \twoheadrightarrow L_i$ allows us to view L as a subfield of each L_i, so the absolute value $|\ |_w$ on L_i restricts to an absolute value on L that uniquely determines a place $w|v$. This defines a map $\phi\colon \{L_i : i \in I\} \to \{L_w : w|v\}$ that we will show is a bijection satisfying $\phi(L_i) \simeq L_i$.

We may view $L \otimes_K K_v \simeq \prod_i L_i$ as an isomorphism of topological rings, since both sides are finite dimensional vector spaces over the complete field K_v and thus have a unique topology induced by the sup norm, by Proposition 10.5, and this topology agrees with the product topology on $\prod_i L_i$. The image of the canonical embedding $L \hookrightarrow L \otimes_K K_v$ defined by $\ell \mapsto \ell \otimes 1$ is dense because $K \subseteq L$ is dense in K_v: given any $\ell \otimes x$ in $L \otimes_K K_v$ with $\ell \in L$ and $x \in K_v$, we can choose $y \in K^\times$ arbitrarily close to x so that $\ell y \otimes 1 = \ell \otimes y$ is an element of the image of L arbitrarily close to $\ell \otimes x$ (and similarly for sums of pure tensors). The image of L is therefore also dense in $\prod_i L_i$ and has dense image under the projections $\prod_i L_i \twoheadrightarrow L_i$ and $\prod_i L_i \twoheadrightarrow L_i \times L_j$ $(i \neq j)$.

If $\phi(L_i) = L_w$ then $L_i \simeq L_w$ since L is dense in the complete field L_i, and L_w is the completion of L with respect to the restriction of the absolute value on L_i to L, by the universal property of completions (Proposition 8.4). To show that ϕ is injective, note that if $\phi(L_i) = \phi(L_j) = L_w$ for some $i \neq j$ we obtain a contradiction because the image of the diagonal embedding $L \to L_w \times L_w$ is not dense in $L_w \times L_w$ (its closure is isomorphic to L_w), but the image of L is dense in $L_i \times L_j$.

It remains only to show that ϕ is surjective. For each $w|v$ we may define a continuous homomorphism of finite étale K_v-algebras and topological rings:

$$\begin{aligned} \varphi_w\colon L \otimes_K K_v &\to L_w \\ \ell \otimes x &\mapsto \ell x. \end{aligned}$$

The map φ_w is surjective because its image contains L and is complete, and L_w is the completion of L. It follows from Corollary 4.32 that φ_w factors through the projection of $L \otimes_K K_v \simeq \prod_i L_i$ on to one of its factors L_i and induces a homeomorphism from L_i to L_w. It follows that $L_i \simeq L_w$ as topological fields, so $\phi(L_i) = L_w$ and ϕ is surjective. □

Corollary 13.6. *Let L/K be a finite separable extension of global fields, v be a place of K, and $f \in K[x]$ be an irreducible polynomial such that $L \simeq K[x]/(f(x))$. There is a one-to-one correspondence between the irreducible factors of f in $K_v[x]$ and the places of L lying above v. If $f = f_1 \cdots f_r$ is the factorization of f in $K_v[x]$, then we can order the set $\{w|v\} = \{w_1, \ldots, w_r\}$ so that $L_{w_i} \simeq K_v[x]/(f_i(x))$ for $1 \le i \le r$.*

Proof. Note that the f_i are distinct because f is separable over K and therefore separable over every extension of K, including K_v. The corollary then follows from Proposition 4.33, Corollary 4.39, and Theorem 13.5. □

Given a finite separable extension of global fields L/K and a place v of K, if we fix an algebraic closure $\overline{K}_v$ of K_v and consider the set $\mathrm{Hom}_K(L, \overline{K}_v)$ of K-embeddings of L into $\overline{K}_v$, the Galois group $\mathrm{Gal}(\overline{K}_v/K_v)$ acts on the set $\mathrm{Hom}_K(L, \overline{K}_v)$ via composition: given $\sigma \in \mathrm{Gal}(\overline{K}_v/K_v)$ and $\tau \in \mathrm{Hom}_K(L, \overline{K}_v)$, we have $\sigma \circ \tau \in \mathrm{Hom}_K(L, \overline{K}_v)$, and this clearly defines a group action (composition is associative and the identity acts trivially).

Corollary 13.7. *Let L/K be a finite separable extension of global fields and v a place of K. We have a bijection*

$$\mathrm{Hom}_K(L, \overline{K}_v)/\mathrm{Gal}(\overline{K}_v/K_v) \longleftrightarrow \{w|v\},$$

between $\mathrm{Gal}(\overline{K}_v/K_v)$-orbits of K-embeddings of L into $\overline{K}_v$ and the places of L above v.

Proof. By the primitive element theorem, we may assume $L \simeq K(\alpha) = K[x]/(f)$ for some $\alpha \in L$ with minimal polynomial $f \in K[x]$. We then have a bijection between $\mathrm{Hom}_K(L, \overline{K}_v)$ and the roots α_i of f in $\overline{K}_v$ that is compatible with the action of $\mathrm{Gal}(\overline{K}_v/K_v)$ on both sets. If $f = f_1 \cdots f_r$ is the factorization of f in $K_v[x]$, each f_i corresponds to an orbit of the action of $\mathrm{Gal}(\overline{K}_v/K_v)$ on the roots of f, and by the previous corollary, these are in one-to-one correspondence with the places of L above v. □

For $K = \mathbb{Q}$ and $v = \infty$, Corollary 13.7 implies that $\mathrm{Hom}_{\mathbb{Q}}(L, \mathbb{C})/\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ is in bijection with the set $\{w|\infty\}$ of infinite places of the number field L; note that $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ is the cyclic group of order 2 generated by complex conjugation, so the orbits of $\mathrm{Hom}_{\mathbb{Q}}(L, \mathbb{C})$ all have size 1 or 2, depending on whether the embedding of L into $\mathbb{C}$ is fixed by complex conjugation or not. Each real place w corresponds to a $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$-orbit of size 1; this occurs for the elements of $\mathrm{Hom}_{\mathbb{Q}}(L, \mathbb{C})$ whose image lies in $\mathbb{R}$ and may also be viewed as elements of $\mathrm{Hom}_{\mathbb{Q}}(L, \mathbb{R})$. Each complex place corresponds to a $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$-orbit of size two in $\mathrm{Hom}_{\mathbb{Q}}(L, \mathbb{C})$; these are conjugate pairs whose images do not lie in $\mathbb{R}$.

Definition 13.8. Let K be a number field. Elements of $\mathrm{Hom}_{\mathbb{Q}}(K, \mathbb{R})$ are *real embeddings*, and elements of $\mathrm{Hom}_{\mathbb{Q}}(K, \mathbb{C})$ whose image does not lie in $\mathbb{R}$ are *complex embeddings*.

There is a one-to-one correspondence between real embeddings and real places, but complex embeddings come in conjugate pairs; each pair of complex embeddings corresponds to a single complex place.

Corollary 13.9. *Let K be a number field with r real places and s complex places. Then*

$$[K : \mathbb{Q}] = r + 2s.$$

Proof. We may write $K \simeq \mathbb{Q}[x]/(f)$ for some irreducible separable $f \in \mathbb{Q}[x]$, and we then have $[K:\mathbb{Q}] = \deg f = \#\operatorname{Hom}_{\mathbb{Q}}(K,\mathbb{C})$, since there is a one-to-one correspondence between $\operatorname{Hom}_{\mathbb{Q}}(K,\mathbb{C})$ and the roots of f. The action of $\operatorname{Gal}(\mathbb{C}/\mathbb{R})$ on $\operatorname{Hom}_{\mathbb{Q}}(K,\mathbb{C})$ has r orbits of size 1, and s orbits of size 2, and the corollary follows. □

Example 13.10. Let $K = \mathbb{Q}[x]/(x^3-2)$. There are three embeddings $K \hookrightarrow \mathbb{C}$, one for each root of x^3-2; explicitly:

$$(1)\ x \mapsto \sqrt[3]{2}, \qquad (2)\ x \mapsto e^{2\pi i/3}\cdot\sqrt[3]{2}, \qquad (3)\ x \mapsto e^{4\pi i/3}\cdot\sqrt[3]{2}.$$

The first embedding is real, while the second two are complex and conjugate to each other. Thus K has $r = 1$ real place and $s = 1$ complex place, and we have $[K:\mathbb{Q}] = 1\cdot 1 + 2\cdot 1 = 3$.

We conclude this section with a result originally due to Brill [2] that relates the parity of the number of complex places to the sign of the absolute discriminant of a number field.

Proposition 13.11. *Let K be a number field with s complex places. The sign of the absolute discriminant $D_K \in \mathbb{Z}$ is $(-1)^s$.*

Proof. Let $\alpha_1,\dots,\alpha_n$ be a $\mathbb{Z}$-basis for $\mathcal{O}_K$, let $\operatorname{Hom}_{\mathbb{Q}}(K,\mathbb{C}) = \{\sigma_1,\dots,\sigma_n\}$ and consider the matrix $A := [\sigma_i(\alpha_j)]_{ij}$ with determinant $\det A =: x + yi \in \mathbb{C}$; recall that $D_K = (\det A)^2$, by Proposition 12.6. Each real embedding σ_i corresponds to a row of A fixed by complex conjugation, while each pair of complex conjugate embeddings $\sigma_i, \bar\sigma_i$ corresponds to a pair of rows of A that are interchanged by complex conjugation. Swapping a pair of rows negates the determinant, thus $\det \bar A = (-1)^s \det A$, and we have

$$x + yi = \det A = (-1)^s \det \bar A = (-1)^s(x - yi).$$

Either $(-1)^s = 1$, in which case $y = 0$ and $D_K = x^2$ has sign $+1 = (-1)^s$, or $(-1)^s = -1$, in which case $x = 0$ and $D_K = -y^2$ has sign $-1 = (-1)^s$. □

13.2 Haar measures

Definition 13.12. Let X be a locally compact Hausdorff space. The *σ-algebra* Σ of X is the collection of subsets of X generated by the open and closed sets under countable unions and countable intersections. Its elements are *Borel sets*, or *measurable sets*. A *Borel measure* on X is a countably additive function

$$\mu\colon \Sigma \to \mathbb{R}_{\ge 0} \cup \{\infty\}.$$

A *Radon measure* on X is a Borel measure on X that additionally satisfies

1. $\mu(S) < \infty$ if S is compact,
2. $\mu(S) = \inf\{\mu(U) : S \subseteq U,\ U \text{ open}\}$,
3. $\mu(S) = \sup\{\mu(C) : C \subseteq S,\ C \text{ compact}\}$,

for all Borel sets S.[1]

[1] Some authors additionally require X to be σ-compact (a countable union of compact sets). Local fields are σ-compact so this distinction will not concern us.

Definition 13.13. A *locally compact group* is a topological group that is Hausdorff and locally compact.[2] A (*left*) *Haar measure* μ on a locally compact group G is a nonzero Radon measure that is *translation invariant*, meaning that

$$\mu(S) = \mu(x + S)$$

for all $x \in G$ and measurable $S \subseteq X$ (we write the group operation additively because we have in mind the additive group of a local field). A compact group is a locally compact group that is compact; in compact groups every measurable set has finite measure.

One defines a right Haar measure analogously, but in most cases they coincide and in our situation we are working with an abelian group (the additive group of a field), in which case they necessarily do. The key fact we need about Haar measures is that they exist and are unique up to scaling. For compact groups existence was proved by Haar and uniqueness by von Neumann; the general result for locally compact groups was proved by Weil.

Theorem 13.14 (Weil)**.** *Every locally compact group G has a Haar measure. If μ and μ' are two Haar measure on G, then there is a positive real number λ such that $\mu'(S) = \lambda\mu(S)$ for all measurable sets S.*

Proof. See [3, §7.2]. □

Example 13.15. The standard Lebesgue measure μ on $\mathbb{R}^n$ with $\mu(\prod_i [a_i, b_i]) = \prod_i |b_i - a_i|$ is the unique Haar measure on $\mathbb{R}^n$ for which the unit cube has measure 1.[3]

The additive group of a local field K is a locally compact group (it is a metric space, hence Hausdorff). For compact groups G, it is standard to normalize the Haar measure so that $\mu(G) = 1$, but local fields are never compact, and we will always have $\mu(K) = \infty$. For nonarchimedean local fields the valuation ring $A = B_{\le 1}(0)$ is a compact group, and it is then natural to normalize the Haar measure on K so that $\mu(A) = 1$. The key point is that there is a unique absolute value on K that is compatible with every Haar measure μ on K, no matter how it is normalized.

Proposition 13.16. *Let K be a local field with discrete valuation v, residue field k, and absolute value*

$$|\cdot|_v := (\#k)^{-v(\cdot)},$$

and let μ be a Haar measure on K. For every $x \in K$ and measurable set $S \subseteq K$ we have

$$\mu(xS) = |x|_v \mu(S).$$

Moreover, the absolute value $|\ |_v$ is the unique absolute value compatible with the topology on K for which this is true.

Proof. Let A be the valuation ring of K with maximal ideal $\mathfrak{p}$. The proposition clearly holds for $x = 0$, so let $x \neq 0$. The map $\phi_x \colon y \mapsto xy$ is an automorphism of the additive group of K, and it follows that the composition $\mu_x = \mu \circ \phi_x$ is a Haar measure on K, hence

[2] Note that the Hausdorff assumption is part of the definition. Some authors include it in the definition of locally compact, and some also include it in the definition of compact (Bourbaki, for example); the convention varies by field and with geography. But everyone agrees that locally compact groups are Hausdorff.

[3] Strictly speaking, the Haar measure on $\mathbb{R}^n$ is the restriction of the Lebesgue measure to the σ-algebra.

a multiple of μ, say $\mu_x = \lambda_x \mu$, for some $\lambda_x \in \mathbb{R}_{>0}$. Define the function $\chi\colon K^\times \to \mathbb{R}_{>0}$ by $\chi(x) := \lambda_x = \mu_x(A)/\mu(A)$. Then $\mu_x = \chi(x)\mu$, and for all $x, y \in K^\times$ we have

$$\chi(xy) = \frac{\mu_{xy}(A)}{\mu(A)} = \frac{\mu_x(yA)}{\mu(A)} = \frac{\chi(x)\mu_y(A)}{\mu(A)} = \frac{\chi(x)\chi(y)\mu(A)}{\mu(A)} = \chi(x)\chi(y).$$

Thus χ is multiplicative, and we claim that in fact $\chi(x) = |x|_v$ for all $x \in K^\times$. Since both χ and $|\cdot|_v$ are multiplicative, it suffices to consider $x \in A - \{0\}$. For any such x, the ideal xA is equal to $\mathfrak{p}^{v(x)}$, since A is a DVR. The residue field $k := A/\mathfrak{p}$ is finite, hence A/xA is also finite; indeed it is a k-vector space of dimension $v(x)$ and has cardinality $[A : xA] = (\#k)^{v(x)}$. Writing A as a finite disjoint union of cosets of xA, we have

$$\mu(A) = [A : xA]\mu(xA) = (\#k)^{v(x)}\chi(x)\mu(A),$$

and therefore $\chi(x) = (\#k)^{-v(x)} = |x|_v$ as claimed. It follows that

$$\mu(xS) = \mu_x(S) = \chi(x)\mu(S) = |x|_v\mu(S),$$

for all $x \in K$ and measurable $S \subseteq K$. To prove uniqueness, if $|\ |$ is an absolute value on K that induces the same topology as $|\ |_v$ then for some $0 < c < 1$ we have $|x| = |x|_v^c$ for all $x \in K^\times$. Let us fix $x \in K^\times$ with $|x|_v \neq 1$ (take any x with $v(x) \neq 0$). If $|\ |$ also satisfies $\mu(xS) = |x|\mu(S)$ then

$$\frac{\mu(xA)}{\mu(A)} = |x| = |x|_v^c = \left(\frac{\mu(xA)}{\mu(A)}\right)^c,$$

which implies $c = 1$, meaning that $|\ |$ and $|\ |_v$ are the same absolute value. □

13.3 The product formula for global fields

Definition 13.17. Let K be a global field. For each place v of K the *normalized absolute value* $\|\ \|_v\colon K_v \to \mathbb{R}_{\geq 0}$ on the completion of K at v is defined by

$$\|x\|_v := \frac{\mu(xS)}{\mu(S)},$$

where μ is a Haar measure on K_v and $S \subseteq K_v$ is a measurable set with finite nonzero measure (such as the set $\{x \in K_v : |x|_v \leq 1\}$, for example).

This definition is independent of the choice of μ and S (by Theorem 13.14). If v is nonarchimedean then the normalized absolute value $\|\ \|_v$ is precisely the absolute value $|\ |_v$ defined in Proposition 13.16. If v is a real place then the normalized absolute value $\|\ \|_v$ is just the usual Euclidean absolute value $|\ |_{\mathbb{R}}$ on $\mathbb{R}$, since for the Euclidean Haar measure $\mu_{\mathbb{R}}$ on $\mathbb{R}$ we have $\mu_{\mathbb{R}}(xS) = |x|_{\mathbb{R}}\mu_{\mathbb{R}}(S)$ for every measurable set S. But when v is a complex place the normalized absolute value $\|\ \|_v$ is the **square** of the Euclidean absolute value $|\ |_{\mathbb{C}}$ on $\mathbb{C}$, since in $\mathbb{C}$ we have $\mu_{\mathbb{C}}(xS) = |x|_{\mathbb{C}}^2\mu_{\mathbb{C}}(S)$.

Remark 13.18. When v is a complex place the normalized absolute value $\|\ \|_v$ is **not** an absolute value, because it does not satisfy the triangle inequality. For example, if $K = \mathbb{Q}(i)$ and $v|\infty$ is the complex place of K then $\|1\|_v = |1|_{\mathbb{C}}^2 = 1$ but

$$\|1 + 1\|_v = \|2\|_v = |2|_{\mathbb{C}}^2 = 4 > 2 = \|1\|_v + \|1\|_v.$$

Nevertheless, the normalized absolute value $\|\ \|_v$ is always multiplicative and compatible with the topology on K_v in the sense that the open balls $B_{<r}(x) := \{y \in K_v : \|y - x\|_v < r\}$ are a basis for the topology on K_v; these are the properties that we care about for the product formula (and for the topology on the ring of adéles $\mathbb{A}_K$ that we will see later).

Lemma 13.19. *Let L/K be a finite separable extension of global fields, let v be a place of K and let $w|v$ be a place of L. Then*

$$\|x\|_w = \|\mathrm{N}_{L_w/K_v}(x)\|_v.$$

Proof. The lemma is trivially true if $[L_w : K_v] = 1$ so assume $[L_w : K_v] > 1$. If v is archimedean then $L_w \simeq \mathbb{C}$ and $K_v \simeq \mathbb{R}$, in which case for any $x \in L_w$ we have

$$\|x\|_w = \mu(xS)/\mu(S) = |x|_{\mathbb{C}}^2 = |x\overline{x}|_{\mathbb{R}} = |\mathrm{N}_{\mathbb{C}/\mathbb{R}}(x)|_{\mathbb{R}} = \|\mathrm{N}_{L_w/K_v}(x)\|_v,$$

where $|\ |_{\mathbb{R}}$ and $|\ |_{\mathbb{C}}$ are the Euclidean absolute values on $\mathbb{R}$ and $\mathbb{C}$.

We now assume v is nonarchimedean. Let π_v and π_w be uniformizers for the local fields K_v and L_w, respectively, and let f be the degree of the corresponding residue field extension k_w/k_v. Without loss of generality, we may assume $x = \pi_w^{w(x)}$, since $\|x\|_v = |x|_v$ depends only on $w(x)$. Theorem 6.10 and Proposition 13.16 imply

$$\|\mathrm{N}_{L_w/K_v}(\pi_w)\|_v = \|\pi_v^f\|_v = (\#k_v)^{-f},$$

so $\|\mathrm{N}_{L_w/K_v}(x)\|_v = (\#k_v)^{-fw(x)}$. Proposition 13.16 then implies

$$\|x\|_w = (\#k_w)^{-w(x)} = (\#k_v)^{-fw(x)} = \|\mathrm{N}_{L_w/K_v}(x)\|_v. \qquad \square$$

Remark 13.20. Note that if v is a nonarchimedean place of K extended by a place $w|v$ of L/K, the absolute value $\|\ \|_w$ is **not** the unique absolute value on L_w that extends the absolute value on $\|\ \|_v$ on K_v given by Theorem 10.6, it differs by a power of $n = [L_w : K_v]$, but it is equivalent to it. It might seem strange to use a normalization here that does not agree with the one we used when considering extensions of local fields in Lecture 9. The difference is that here we are thinking about a single global field K that has many different completions K_v, and we want the normalized absolute values on the various K_v to be compatible (so that the product formula will hold). By contrast, in Lecture 9 we considered various extensions L_w of a single local field K_v and wanted to normalize the absolute values on the L_w compatibly so that we could work in K_v and any of its extensions (all the way up to $\overline{K}_v$) using the same absolute value. These two objectives cannot be met simultaneously and it is better to use the "right" normalization in each setting.

Theorem 13.21 (Product Formula)**.** *Let L be a global field. For all $x \in L^\times$ we have*

$$\prod_{v \in M_L} \|x\|_v = 1,$$

where $\|\ \|_v$ denotes the normalized absolute value for each place $v \in M_L$.

Proof. The global field L is a finite separable extension of $K = \mathbb{Q}$ or $K = \mathbb{F}_q(t)$.[4] Let p be a place of K. By Theorem 13.5, any basis for L as a K-vector space is also a basis for

$$L \otimes_K K_p \simeq \prod_{v|p} L_v$$

[4]Here we are using the fact that if $\mathbb{F}_q$ is the field of constants of L (the largest finite field in L), then L is a finite extension of $\mathbb{F}_q(z)$ and we can choose some $t \in \mathbb{F}_q(z) - \mathbb{F}_q$ so that $\mathbb{F}_q(z) \simeq \mathbb{F}_q(t)$ and $L/\mathbb{F}_q(t)$ is separable (such a t is called a *separating element*).

as a K_v-vector space. Thus

$$\mathrm{N}_{L/K}(x) = \mathrm{N}_{(L\otimes_K K_p)/K_p}(x) = \prod_{v|p} \mathrm{N}_{L_v/K_p}(x).$$

Taking normalized absolute values on both sides yields

$$\big\|\mathrm{N}_{L/K}(x)\big\|_p = \prod_{v|p} \|\mathrm{N}_{L_v/K_p}(x)\|_p = \prod_{v|p} \|x\|_v.$$

We now take the product of both sides over all places $p \in M_K$ to obtain

$$\prod_{p\in M_K} \|\mathrm{N}_{L/K}(x)\|_p = \prod_{p\in M_K} \prod_{v|p} \|x\|_v = \prod_{v\in M_L} \|x\|_v.$$

The LHS is equal to 1, by the product formula for K proved on Problem Set 1. □

With the product formula in hand, we can now give an axiomatic definition of a global field, which up to now we have simply defined as a finite extension of $\mathbb{Q}$ or $\mathbb{F}_q(t)$, due to Emil Artin and George Whaples [1].

Definition 13.22. A *global field* is a field K with at least one place whose completion at each of its places $v \in M_K$ is a local field K_v, and which has a product formula of the form

$$\prod_{v\in M_K} \|x\|_v = 1,$$

where each normalized absolute value $\| \ \|_v \colon K_v \to \mathbb{R}_{\ge 0}$ satisfies $\| \ \|_v = | \ |_v^{m_v}$ for some absolute value $| \ |_v$ representing v and some fixed $m_v \in \mathbb{R}_{>0}$.

Theorem 13.23 (Artin-Whaples)**.** *Every global field is a finite extension of* $\mathbb{Q}$ *or* $\mathbb{F}_q(t)$*.*

Proof. See Problem Set 7. □

References

[1] Emil Artin and George Whaples, *Axiomatic characterization of fields by the product formula for valuations*, Bull. Amer. Math. Soc. **51** (1945), 469–492.

[2] Alexander von Brill, *Ueber die Discriminante*, Math. Ann. **12** (1877), 87–89.

[3] Joe Diestel and Angela Spalsbury, *The Joys of Haar Measure*, American Mathematical Society, 2014.

14 The geometry of numbers

14.1 Lattices in real vector spaces

Recall that for an integral domain A with fraction field K, an *A-lattice* in a finite dimensional K-vector space V is a finitely generated A-submodule of V that contains a K-basis for V (see Definition 5.9). We now want to specialize to the case $A = \mathbb{Z}$, but rather than working with the fraction field $K = \mathbb{Q}$ we will instead work with its completion $\mathbb{R}$ at the unique infinite place of $\mathbb{Q}$.

Remark 14.1. In this lecture we shall focus specifically on number fields, but we will make remarks along the way about how one can similarly treat global function fields (where one would take $A = \mathbb{F}_q[t]$ and work with its completion $\mathbb{F}_q(t)_\infty \simeq \mathbb{F}_q((\frac{1}{t}))$ at the unique infinite place of $\mathbb{F}_q(t)$). In Problem Set 7 you will have the opportunity to explore the function field case in more detail.

A finitely generated $\mathbb{Z}$-submodule of a vector space is necessarily a free module, since $\mathbb{Z}$ is a PID and every submodule of a vector space is torsion-free. Now V is an $\mathbb{R}$-vector space of some finite dimension n, and has a canonical structure as a topological metric space isomorphic to $\mathbb{R}^n$ (by Proposition 10.5, there is a unique topology on V compatible with the topology of $\mathbb{R}$, because $\mathbb{R}$ is complete). This topology makes V a locally compact Hausdorff space, thus V is a locally compact group and therefore has a Haar measure μ that is unique up to scaling, by Theorem 13.14.

Definition 14.2. A subgroup H of a topological group G is *discrete* if the subspace topology on H is the discrete topology (every point is open), and *cocompact* if H is a normal subgroup of G and the quotient G/H is compact (here G/H denotes the group G/H with the quotient topology given by identifying elements of G that lie in the same coset of H).

Definition 14.3. Let V be an $\mathbb{R}$-vector space of finite dimension. A (full) *lattice* in V is a $\mathbb{Z}$-submodule generated by an $\mathbb{R}$-basis for V; equivalently, a discrete cocompact subgroup.

See Problem Set 7 for a proof that these two definitions are equivalent.

Remark 14.4. A discrete subgroup of a Hausdorff topological group is always closed; see [1, III.2.1.5] for a proof. This implies that the quotient of a Hausdorff topological group by a normal discrete subgroup is Hausdorff (which is false for topological spaces in general); see [1, III.2.1.18]. It follows that the quotient of a Hausdorff topological group (including all locally compact groups) by a discrete cocompact subgroup is a compact group. These facts are easy to see in the case of lattices: $\mathbb{Z}$ is closed in $\mathbb{R}$ (as the complement of a union of open intervals), so $\mathbb{Z}^n$ is closed in $\mathbb{R}^n$. Given a lattice Λ in V, each $\mathbb{Z}$-basis for Λ determines an isomorphism of topological groups $\Lambda \simeq \mathbb{Z}^n$ and $V \simeq \mathbb{R}^n$, and the quotient $V/\Lambda \simeq \mathbb{R}^n/\mathbb{Z}^n \simeq (\mathbb{R}/\mathbb{Z})^n$ (an *n-torus*), is compact Hausdorff and thus a compact group.

Remark 14.5. You might ask why we are using the archimedean completion $\mathbb{R} = \mathbb{Q}_\infty$ rather than some other completion $\mathbb{Q}_p$. The reason is $\mathbb{Z}$ is not a discrete subgroup of $\mathbb{Q}_p$ for any finite place p (elements of $\mathbb{Z}$ can be arbitrarily close to 0 under the p-adic metric). Similarly, $\mathbb{F}_q[t]$ is a discrete subgroup of $\mathbb{F}_q(t)_\infty$, but not of any other completion of $\mathbb{F}_q(t)$.

Any basis $v_1, \dots, v_n$ for V determines a parallelepiped

$$F(v_1, \dots, v_n) := \{t_1 v_1 + \cdots + t_n v_n : t_1, \dots, t_n \in [0, 1)\}$$

that we may view as the unit cube by fixing an isomorphism $\varphi\colon V \xrightarrow{\sim} \mathbb{R}^n$ that maps $(v_1,\dots,v_n)$ to the standard basis of unit vectors for $\mathbb{R}^n$. It then makes sense to normalize the Haar measure μ so that $\mu(F(v_1,\dots,v_n)) = 1$, and we then have $\mu(S) = \mu_{\mathbb{R}^n}(\varphi(S))$ for every measurable set $S \subseteq V$, where $\mu_{\mathbb{R}^n}$ denotes the standard Lebesgue measure on $\mathbb{R}^n$.

For any other basis $e_1,\dots,e_n$ of V, if we let $E = [e_{ij}]_{ij}$ be the matrix whose jth column expresses $e_j = \sum_i e_{ij} v_i$, in terms of our normalized basis $v_1,\dots,v_n$, then

$$\mu(F(e_1,\dots,e_n)) = |\det E| = \sqrt{\det E^{\mathrm{t}} \det E} = \sqrt{\det(E^{\mathrm{t}}E)} = \sqrt{\det[\langle e_i, e_j\rangle]_{ij}}, \tag{1}$$

where $\langle e_i, e_j\rangle$ is the canonical inner product (the dot product) on $\mathbb{R}^n$. Here we have used the fact that the determinant of a matrix in $\mathbb{R}^{n\times n}$ is the signed volume of the parallelepiped spanned by its columns (or rows). This is a consequence of the following more general result, which is independent of the choice of basis or the normalization of μ.

Proposition 14.6. *Let $T\colon V \to V$ be a linear transformation of $V \simeq \mathbb{R}^n$. For any Haar measure μ on V and every measurable set $S \subseteq V$ we have*

$$\mu(T(S)) = |\det T|\,\mu(S). \tag{2}$$

Proof. See [11, Ex. 1.2.21]. □

If Λ is a lattice $e_1\mathbb{Z} + \cdots + e_n\mathbb{Z}$ in V, the quotient V/Λ is a compact group that we may identify with the parallelepiped $F(e_1,\dots,e_n) \subseteq V$, which forms a set of unique coset representatives. More generally, we make the following definition.

Definition 14.7. Let Λ be a lattice in $V \simeq \mathbb{R}^n$. A *fundamental domain* for Λ is a measurable set $F \subseteq V$ such that

$$V = \bigsqcup_{\lambda\in\Lambda} (F + \lambda).$$

In other words, F is a measurable set of coset representatives for V/Λ. Fundamental domains exist: if $\Lambda = e_1\mathbb{Z} + \cdots + e_n\mathbb{Z}$ we may take the parallelepiped $F(e_1,\dots,e_n)$.

Proposition 14.8. *Let Λ be a lattice in $V \simeq \mathbb{R}^n$ and let μ be a Haar measure on V. Every fundamental domain for Λ has the same measure, and this measure is finite and nonzero.*

Proof. Let F and G be two fundamental domains for Λ. Using the translation invariance and countable additivity of μ (note that $\Lambda \simeq \mathbb{Z}^n$ is a countable set) along with the fact that Λ is closed under negation, we obtain

$$\begin{aligned}\mu(F) = \mu(F\cap V) &= \mu\left(F \cap \bigsqcup_{\lambda\in\Lambda}(G+\lambda)\right) = \mu\left(\bigsqcup_{\lambda\in\Lambda}(F\cap(G+\lambda))\right)\\ &= \sum_{\lambda\in\Lambda}\mu(F\cap(G+\lambda)) = \sum_{\lambda\in\Lambda}\mu((F-\lambda)\cap G) = \sum_{\lambda\in\Lambda}\mu(G\cap(F+\lambda)) = \mu(G),\end{aligned}$$

where the last equality follows from the first four (swap F and G). If we fix a $\mathbb{Z}$-basis $e_1,\dots,e_n$ for Λ, the parallelepiped $F(e_1,\dots,e_n)$ is a fundamental domain for Λ, and its closure is compact, so $\mu(F(e_1,\dots,e_n))$ is finite, and it is nonzero because there is an isomorphism $V \simeq \mathbb{R}^n$ that maps the closure of $F(e_1,\dots,e_n)$ to the unit cube in $\mathbb{R}^n$ whose Lebesgue measure is nonzero (whether a set has zero measure or not does not depend on the normalization of the Haar measure and is therefore preserved by isomorphisms of locally compact groups). □

Definition 14.9. Let Λ be a lattice in $V \simeq \mathbb{R}^n$ and fix a Haar measure μ on V. The *covolume* $\operatorname{covol}(\Lambda) \in \mathbb{R}_{>0}$ of Λ is the measure $\mu(F)$ of any fundamental domain F for Λ.

Note that covolumes depend on the normalization of μ, but ratios of covolumes do not.

Proposition 14.10. *If $\Lambda' \subseteq \Lambda$ are lattices in $V \simeq \mathbb{R}^n$, then $\operatorname{covol}(\Lambda') = [\Lambda : \Lambda'] \operatorname{covol}(\Lambda)$.*

Proof. Fix a fundamental domain F for Λ and a set of coset representatives S for Λ/Λ'. Then

$$F' := \bigsqcup_{\lambda \in S} (F + \lambda)$$

is a fundamental domain for Λ', and $\#S = [\Lambda : \Lambda'] = \mu(F')/\mu(F)$ is finite. We then have

$$\operatorname{covol}(\Lambda') = \mu(F') = \sum_{\lambda \in S} \mu(F + \lambda) = (\#S)\mu(F) = [\Lambda : \Lambda'] \operatorname{covol}(\Lambda),$$

since every translation $F + \lambda$ of F is a fundamental domain for Λ. □

Definition 14.11. Let S be a subset of a real vector space. The set S is *symmetric* if it is closed under negation, and *convex* if for all $x, y \in S$ we have $\{tx + (1-t)y : t \in [0,1]\} \subseteq S$.

Theorem 14.12 (Minkowski's Lattice Point Theorem)**.** *Let Λ be a lattice in $V \simeq \mathbb{R}^n$ and μ a Haar measure on V. If $S \subseteq V$ is a symmetric convex measurable set that satisfies*

$$\mu(S) > 2^n \operatorname{covol}(\Lambda),$$

then S contains a nonzero element of Λ.

Proof. See Problem Set 6. □

Note that the inequality in Theorem 14.12 bounds the ratio of the measures of two sets (S and a fundamental domain for Λ), and is thus independent of the choice of μ.

Remark 14.13. In the function field analog of Theorem 14.12 the convexity assumption is not needed and the factor of 2^n can be removed.

14.2 The canonical inner product

Let $K/\mathbb{Q}$ be a number field of degree n with r real places and s complex places; then $n = r + 2s$, by Corollary 13.9. We now want to consider the base change of K to $\mathbb{R}$ and $\mathbb{C}$:

$$K_{\mathbb{R}} := K \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{R}^r \times \mathbb{C}^s,$$
$$K_{\mathbb{C}} := K \otimes_{\mathbb{Q}} \mathbb{C} \simeq \mathbb{C}^n.$$

The isomorphism $K_{\mathbb{R}} \simeq \mathbb{R}^r \times \mathbb{C}^s$ follows from Theorem 13.5 and the isomorphism $K_{\mathbb{C}} \simeq \mathbb{C}^n$ follows from the fact that $\mathbb{C}$ is separably closed; see Example 4.31. We note that $K_{\mathbb{R}}$ is an $\mathbb{R}$-vector space of dimension n, thus $K_{\mathbb{R}} \simeq \mathbb{R}^n$, but this is an isomorphism of $\mathbb{R}$-vector spaces and is not an $\mathbb{R}$-algebra isomorphism unless $s = 0$.

We have a sequence of injective homomorphisms of topological rings

$$\mathcal{O}_K \hookrightarrow K \hookrightarrow K_{\mathbb{R}} \hookrightarrow K_{\mathbb{C}}, \tag{3}$$

which are defined as follows:

- the map $\mathcal{O}_K \hookrightarrow K$ is inclusion;
- the map $K \hookrightarrow K_{\mathbb{R}} = K \otimes_{\mathbb{Q}} \mathbb{R}$ is the canonical embedding $\alpha \mapsto \alpha \otimes 1$;
- the map $K_{\mathbb{R}} \simeq \mathbb{R}^r \times \mathbb{C}^s \hookrightarrow \mathbb{C}^r \times \mathbb{C}^{2s} \simeq K_{\mathbb{C}}$ embeds each factor of $\mathbb{R}^r$ in a corresponding factor of $\mathbb{C}^r$ via inclusion and each $\mathbb{C}$ in $\mathbb{C}^s$ is mapped to $\mathbb{C} \times \mathbb{C}$ in $\mathbb{C}^{2s}$ via $z \mapsto (z, \bar{z})$.

To better understand the last map, note that each $\mathbb{C}$ in $\mathbb{C}^s$ arises as $\mathbb{R}[\alpha] = \mathbb{R}[x]/(f) \simeq \mathbb{C}$ for some monic irreducible $f \in \mathbb{R}[x]$ of degree 2, but when we base-change to $\mathbb{C}$ the field $\mathbb{R}[\alpha]$ splits into the étale algebra $\mathbb{C}[x]/(x-\alpha) \times \mathbb{C}[x]/(x-\bar{\alpha}) \simeq \mathbb{C} \times \mathbb{C}$. The composition $K \hookrightarrow K_{\mathbb{R}} \hookrightarrow K_{\mathbb{C}}$ is given by the map

$$x \mapsto (\sigma_1(x), \dots, \sigma_n(x)),$$

where $\mathrm{Hom}_{\mathbb{Q}}(K, \mathbb{C}) = \{\sigma_1, \dots, \sigma_n\}$. If we put $K = \mathbb{Q}(\alpha) := K[x]/(f)$ and let $\alpha_1, \dots, \alpha_n \in \mathbb{C}$ be the roots of f in $\mathbb{C}$, each σ_i is the $\mathbb{Q}$-algebra homomorphism $K \to \mathbb{C}$ defined by $\alpha \mapsto \alpha_i$.

If we fix a $\mathbb{Z}$-basis for $\mathcal{O}_K$, its image under the maps in (3) is a $\mathbb{Q}$-basis for K, an $\mathbb{R}$-basis for $K_{\mathbb{R}}$, and a $\mathbb{C}$-basis for $K_{\mathbb{C}}$, all of which are vector spaces of dimension $n = [K : \mathbb{Q}]$. We may thus view the injections in (3) as inclusions of topological groups (but not rings!)

$$\mathbb{Z}^n \hookrightarrow \mathbb{Q}^n \hookrightarrow \mathbb{R}^n \hookrightarrow \mathbb{C}^n.$$

The ring of integers $\mathcal{O}_K$ is a lattice in the real vector space $K_{\mathbb{R}} \simeq \mathbb{R}^n$, which inherits an inner product from the canonical Hermitian inner product on $K_{\mathbb{C}} \simeq \mathbb{C}^n$ defined by

$$\langle z, z' \rangle := \sum_{i=1}^{n} z_i \bar{z}_i' \in \mathbb{C}.$$

For elements $x, y \in K \hookrightarrow K_{\mathbb{R}} \hookrightarrow K_{\mathbb{C}}$ the Hermitian inner product can be computed as

$$\langle x, y \rangle := \sum_{\sigma \in \mathrm{Hom}_{\mathbb{Q}}(K, \mathbb{C})} \sigma(x)\overline{\sigma(y)} \in \mathbb{R}, \tag{4}$$

which is a real number because the non-real embeddings in $\mathrm{Hom}_{\mathbb{Q}}(K, \mathbb{C})$ come in complex conjugate pairs. The inner product defined in (4) agrees with the restriction of the Hermitian inner product on $K_{\mathbb{R}} \hookrightarrow K_{\mathbb{C}}$. The metric space topology it induces on $K_{\mathbb{R}}$ is the same as the Euclidean topology on $K_{\mathbb{R}} \simeq \mathbb{R}^n$ induced by the usual dot product on $\mathbb{R}^n$, but the corresponding norm $\|x\| := \langle x, x \rangle$ has a different normalization, as we now explain.

If we write elements $z \in K_{\mathbb{C}} \simeq \mathbb{C}^n$ as vectors (z_σ) indexed by the set $\sigma \in \mathrm{Hom}_{\mathbb{Q}}(K, \mathbb{C})$ in some fixed order, we may identify $K_{\mathbb{R}}$ with its image in $K_{\mathbb{C}}$ as the set

$$K_{\mathbb{R}} = \{z \in K_{\mathbb{C}} : \bar{z}_\sigma = z_{\bar{\sigma}} \text{ for all } \sigma \in \mathrm{Hom}_{\mathbb{Q}}(K, \mathbb{C})\}.$$

For real embeddings $\sigma = \bar{\sigma}$ we have $z_\sigma \in \mathbb{R} \subseteq \mathbb{C}$, and for pairs of conjugate complex embeddings $(\sigma, \bar{\sigma})$ we get the embedding $z \mapsto (z_\sigma, z_{\bar{\sigma}}) = (z_\sigma, \bar{z}_\sigma)$ of $\mathbb{C}$ into $\mathbb{C} \times \mathbb{C}$ used to defined the map $K_{\mathbb{R}} \hookrightarrow K_{\mathbb{C}}$ above. Each $z \in K_{\mathbb{R}}$ can be uniquely written in the form

$$(w_1, \dots, w_r,\ x_1 + iy_1,\ x_1 - iy_1,\ \dots,\ x_s + iy_s,\ x_s - iy_s), \tag{5}$$

with $w_i, x_j, y_j \in \mathbb{R}$. Each w_i corresponds to a z_σ with $\sigma = \bar{\sigma}$, and each $(x_j + iy_j, x_j - iy_j)$ corresponds to a complex conjugate pair $(z_\sigma, z_{\bar{\sigma}})$ with $\sigma \neq \bar{\sigma}$. The canonical inner product on $K_{\mathbb{R}}$ can then be written as

$$\langle z, z' \rangle = \sum_{i=1}^{r} w_i w_i' + 2 \sum_{j=1}^{s} (x_j x_j' + y_j y_j').$$

Thus if we take $w_1, \dots, w_r, x_1, y_1, \dots, x_x, y_s$ as coordinates for $K_{\mathbb{R}} \simeq \mathbb{R}^n$ (as $\mathbb{R}$-vector spaces), in order to normalize the Haar measure μ on $K_{\mathbb{R}}$ so that it is consistent with the Lebesgue measure $\mu_{\mathbb{R}^n}$ on $\mathbb{R}^n$ we define

$$\mu(S) \coloneqq 2^s \mu_{\mathbb{R}^n}(S) \tag{6}$$

for any measurable set $S \subseteq K_{\mathbb{R}}$ that we may view as a subset of $\mathbb{R}^n$ by expressing it in w_i, x_j, y_j coordinates as above. With this normalization, the identity (1) still holds when we replace $\mu_{\mathbb{R}^n}$ with μ and the dot product on $\mathbb{R}^n$ with the Hermitian inner product on $K_{\mathbb{R}}$, that is, for any $\mathbb{R}$-basis $e_1, \dots, e_n$ of $K_{\mathbb{R}}$ we still have

$$\mu(F(e_1, \dots, e_n)) = \sqrt{|\det[\langle e_i, e_j \rangle]_{ij}|} \tag{7}$$

Using the Hermitian inner product on $K_{\mathbb{R}} \subseteq K_{\mathbb{C}}$ rather than the dot product on $K_{\mathbb{R}} \simeq \mathbb{R}^n$ multiplies $2s$ of the columns in the matrix $[\langle e_i, e_j \rangle]_{ij}$ by 2, and thus multiplies the RHS by $\sqrt{2^{2s}} = 2^s$; our normalization of $\mu = 2^s \mu_{\mathbb{R}^n}$ multiplies the LHS by 2^s so that (7) still holds.

Remark 14.14. In the function field case one replaces the separable closure $\mathbb{C}$ of $\mathbb{R}$ with a separable closure $\mathbb{F}_q(t)_\infty^{\text{sep}}$ of $\mathbb{F}_q(t)_\infty$. The situation is slightly more complicated, since unlike $\mathbb{C}/\mathbb{R}$, the extension $\mathbb{F}_q(t)_\infty^{\text{sep}}/\mathbb{F}_q(t)_\infty$ is not finite, but for any finite separable extension $K/\mathbb{F}_q(t)$ (a finite étale $\mathbb{F}_q(t)$-algebra) one can base change K to $\mathbb{F}_q(t)_\infty$ and $\mathbb{F}_q(t)_\infty^{\text{sep}}$; these play the role of $K_{\mathbb{R}}$ and $K_{\mathbb{C}}$.

14.3 Covolumes of fractional ideals

Having fixed a normalized Haar measure μ for $K_{\mathbb{R}}$, we can now compute covolumes of lattices in $K_{\mathbb{R}} \simeq \mathbb{R}^n$. This includes not only (the image of) the ring of integers $\mathcal{O}_K$, but also any nonzero fractional ideal I of $\mathcal{O}_K$: every such I contains a nonzero principal fraction ideal $a\mathcal{O}_K$, and if $e_1, \dots, e_n$ is a $\mathbb{Z}$-basis for $\mathcal{O}_K$ then $ae_1, \dots, ae_n$ is a $\mathbb{Z}$-basis for $a\mathcal{O}_K$ that is an $\mathbb{R}$-basis for $K_{\mathbb{R}}$ that lies in I.

Recall from Remark 12.14 that the discriminant of a number field K is the integer

$$D_K \coloneqq \operatorname{disc} \mathcal{O}_K \coloneqq \operatorname{disc}(e_1, \dots, e_n) \in \mathbb{Z}.$$

Proposition 14.15. *Let K be a number field. Using the normalized Haar measure on $K_{\mathbb{R}}$ defined in (6),*

$$\operatorname{covol}(\mathcal{O}_K) = \sqrt{|D_K|}.$$

Proof. Let $e_1, \dots, e_n \in \mathcal{O}_K$ be a $\mathbb{Z}$-basis for $\mathcal{O}_K$, let $\operatorname{Hom}_{\mathbb{Q}}(K, \mathbb{C}) = \{\sigma_1, \dots, \sigma_n\}$, and define $A \coloneqq [\sigma_i(e_j)]_{ij} \in \mathbb{C}^{n \times n}$. Then $D_K = \operatorname{disc}(e_1, \dots, e_n) = (\det A)^2$, by Proposition 12.6

Viewing $\mathcal{O}_K \hookrightarrow K_{\mathbb{R}}$ as a lattice in $K_{\mathbb{R}}$ with basis $e_1, \dots, e_n$, we may use (7) to compute $\operatorname{covol}(\mathcal{O}_K) = \mu(F(e_1, \dots, e_n)) = \sqrt{|\det[\langle e_i, e_j \rangle]_{ij}|}$. Applying (4) yields

$$\det[\langle e_i, e_j \rangle]_{ij} = \det\left[\textstyle\sum_k \sigma_k(e_i)\overline{\sigma_k(e_j)}\right]_{ij} = \det(A^{\mathrm{t}}\overline{A}) = (\det A)(\det \overline{A}).$$

Noting that $\det A$ is the square root of an integer (hence either real or purely imaginary), we have $\operatorname{covol}(\mathcal{O}_K)^2 = |(\det A)^2| = |D_K|$, and the proposition follows. □

Recall from Remark 6.13 that for number fields K we view the absolute norm

$$\begin{aligned} \mathrm{N}\colon \mathcal{I}_{\mathcal{O}_K} &\to \mathcal{I}_{\mathbb{Z}} \\ I &\mapsto [\mathcal{O}_K : I]_{\mathbb{Z}} \end{aligned}$$

as having image in $\mathbb{Q}_{>0}$ by identifying $\mathrm{N}(I) \in \mathcal{I}_{\mathbb{Z}}$ with a positive generator for $\mathrm{N}(I)$ (note that $\mathbb{Z}$ is a PID). Recall that $[\mathcal{O}_K : I]_{\mathbb{Z}}$ is a module index of $\mathbb{Z}$-lattices in the $\mathbb{Q}$-vector space K (see Definitions 6.1 and 6.5), and for ideals $I \subseteq \mathcal{O}_K$ this is just the positive integer $[\mathcal{O}_K : I]_{\mathbb{Z}} = [\mathcal{O}_K : I]$. When $I = (a)$ is a principal fractional ideal with $a \in K$, we may simply write $\mathrm{N}(a) \coloneqq \mathrm{N}((a)) = |\mathrm{N}_{K/\mathbb{Q}}(a)|$.

Corollary 14.16. *Let K be a number field and let I be a nonzero fractional ideal of $\mathcal{O}_K$. Then*

$$\operatorname{covol}(I) = \mathrm{N}(I)\sqrt{|D_K|}$$

Proof. Let $n = [K : \mathbb{Q}]$. Since $\operatorname{covol}(bI) = b^n \operatorname{covol}(I)$ and $\mathrm{N}(bI) = b^n \mathrm{N}(I)$ for any $b \in \mathbb{Z}_{>0}$, without loss of generality we may assume $I \subseteq \mathcal{O}_K$ (replace I with a suitable bI if not). Applying Propositions 14.10 and 14.15, we have

$$\operatorname{covol}(I) = [\mathcal{O}_K : I] \operatorname{covol}(\mathcal{O}_K) = \mathrm{N}(I) \operatorname{covol}(\mathcal{O}_K) = \mathrm{N}(I)\sqrt{|D_K|}$$

as claimed. □

14.4 The Minkowski bound

Theorem 14.17. MINKOWSKI BOUND *Let K be a number field of degree n with s complex places. Define the Minkowski constant m_K for K as the positive real number*

$$m_K := \frac{n!}{n^n}\left(\frac{4}{\pi}\right)^s \sqrt{|D_K|}.$$

For every nonzero fractional ideal I of $\mathcal{O}_K$ there is a nonzero $a \in I$ for which

$$\mathrm{N}(a) \le m_K \mathrm{N}(I).$$

To prove this theorem we need the following lemma.

Lemma 14.18. *Let K be a number field of degree n with r real and s complex places. For each $t \in \mathbb{R}_{>0}$, the measure of the convex symmetric set*

$$S_t := \left\{(z_\sigma) \in K_{\mathbb{R}} : \sum |z_\sigma| \le t\right\} \subseteq K_{\mathbb{R}}$$

with respect to the normalized Haar measure μ on $K_{\mathbb{R}}$ is

$$\mu(S_t) = 2^r \pi^s \frac{t^n}{n!}.$$

Proof. As in (5), we may uniquely write each $z = (z_\sigma) \in K_{\mathbb{R}}$ in the form

$$(w_1, \dots, w_r,\ x_1 + iy_1,\ x_1 - iy_1\ \dots,\ x_s + iy_s,\ x_s - iy_s)$$

with $w_i, x_j, y_j \in \mathbb{R}$. We will have $\sum_\sigma |z_\sigma| \le t$ if and only if

$$\sum_{i=1}^{r} |w_i| + \sum_{j=1}^{s} 2\sqrt{|x_j|^2 + |y_j|^2} \le t. \tag{8}$$

We now compute the volume of this region in $\mathbb{R}^n$ by relating it to the volume of the simplex

$$U_t := \{(u_1, \dots, u_n) \in \mathbb{R}^n_{\geq 0} \colon u_1 + \cdots + u_n \leq t\} \subseteq \mathbb{R}^n,$$

which is $\mu_{\mathbb{R}^n}(U_t) = t^n/n!$ (volume of the standard simplex in $\mathbb{R}^n$ scaled by a factor of t).

If we view all the w_i, x_j, y_j as fixed except the last pair (x_s, y_s), then (x_s, y_s) ranges over a disk of some radius $d \in [0, t/2]$ determined by (8). If we replace (x_s, y_s) with (u_{n-1}, u_n) ranging over the triangular region bounded by $u_{n-1} + u_n \leq 2d$ and $u_{n-1}, u_n \geq 0$, we need to incorporate a factor of $\pi/2$ to account for the difference between $(2d)^2/2 = 2d^2$ and πd^2; repeat this s times. Similarly, if we hold everything but w_r fixed and replace w_r ranging over $[-d, d]$ for some $d \in [0, t]$ with u_r ranging over $[0, d]$, we need to incorporate a factor of 2 to account for this change of variable; repeat r times. We then have

$$\mu(S_t) = 2^s \mu_{\mathbb{R}^n}(S_t) = 2^s \left(\frac{\pi}{2}\right)^s 2^r \mu_{\mathbb{R}^n}(U) = 2^r \pi^s \frac{t^n}{n!}. \qquad \square$$

Proof of Theorem 14.17. Let I be a nonzero fractional ideal of $\mathcal{O}_K$. By Theorem 14.12, if we choose t so that $\mu(S_t) > 2^n \operatorname{covol}(I)$, then S_t will contain a nonzero $a \in I$. By Lemma 14.18 and Corollary 14.16, it suffices to choose t so that

$$\left(\frac{t}{n}\right)^n = \frac{n!\mu(S_t)}{n^n 2^r \pi^s} > \frac{n! 2^n}{n^n 2^r \pi^s} \operatorname{covol}(I) = \frac{n!}{n^n} \left(\frac{4}{\pi}\right)^s \sqrt{|D_K|} \mathrm{N}(I) = m_K \mathrm{N}(I).$$

Let us now pick t so that $\left(\frac{t}{n}\right)^n > m_K \mathrm{N}(I)$. Then S_t contains $a \in I$ with $\sum_\sigma |\sigma(a)| \leq t$ Recalling that the geometric mean is bounded above by the arithmetic mean, we then have

$$\mathrm{N}(a) = \left(\mathrm{N}(a)^{1/n}\right)^n = \left(\prod_\sigma |\sigma(a)|^{1/n}\right)^n \leq \left(\frac{1}{n} \sum_\sigma |\sigma(a)|\right)^n \leq \left(\frac{t}{n}\right)^n,$$

Taking the limit as $\left(\frac{t}{n}\right)^n \to m_K \mathrm{N}(I)$ from above yields $\mathrm{N}(a) \leq m_K \mathrm{N}(I)$. $\square$

14.5 Finiteness of the class group

Recall that the ideal class group $\operatorname{cl} \mathcal{O}_K$ is the quotient of the ideal group $\mathcal{I}_K$ of $\mathcal{O}_K$ by its subgroup of principal fractional ideals. We now use the Minkowski bound to prove that every ideal class $[I] \in \operatorname{cl} \mathcal{O}_K$ can be represented by an ideal $I \subseteq \mathcal{O}_K$ of small norm. It will then follow that the ideal class group is finite.

Theorem 14.19. *Let K be a number field. Every ideal class in $\operatorname{cl} \mathcal{O}_K$ contains an ideal $I \subseteq \mathcal{O}_K$ of absolute norm $\mathrm{N}(I) \leq m_K$, where m_K is the Minkowski constant for K.*

Proof. Let $[J]$ be an ideal class of $\mathcal{O}_K$ represented by the nonzero fractional ideal J. By Theorem 14.17, the fractional ideal J^{-1} contains a nonzero element a for which

$$\mathrm{N}(a) \leq m_K \mathrm{N}(J^{-1}) = m_K \mathrm{N}(J)^{-1},$$

and therefore $\mathrm{N}(aJ) = \mathrm{N}(a)\mathrm{N}(J) \leq m_K$. We have $a \in J^{-1}$, thus $aJ \subseteq J^{-1}J = \mathcal{O}_K$, so $I = aJ$ is an $\mathcal{O}_K$-ideal in the ideal class $[J]$ with $\mathrm{N}(I) \leq m_K$ as desired. $\square$

Lemma 14.20. *Let K be a number field and let M be a real number. The set of ideals $I \subseteq \mathcal{O}_K$ with $\mathrm{N}(I) \leq M$ is finite.*

Proof 1. As a lattice in $K_{\mathbb{R}} \simeq \mathbb{R}^n$, the additive group $\mathcal{O}_K \simeq \mathbb{Z}^n$ has only finitely many subgroups I of index m for each positive integer $m \le M$, since $[\mathbb{Z}^n : I] = m$ implies

$$(m\mathbb{Z})^n \subseteq I \subseteq \mathbb{Z}^n,$$

and $(m\mathbb{Z})^n$ has finite index $m^n = [\mathbb{Z}^n : m\mathbb{Z}^n] = [\mathbb{Z} : m\mathbb{Z}]^n$ in $\mathbb{Z}^n$. □

The proof of Lemma 14.20 is effective: the number of ideals $I \subseteq \mathcal{O}_K$ with $\mathrm{N}(I) \le M$ clearly cannot exceed M^{n+1}. But in fact we can give a much better bound than this.

Proof 2. Let I be an ideal of absolute norm $\mathrm{N}(I) \le M$ and let $I = \mathfrak{p}_1 \cdots \mathfrak{p}_k$ be its factorization into (not necessarily distinct) prime ideals. Then $M \ge \mathrm{N}(I) = \mathrm{N}(\mathfrak{p}_1) \cdots \mathrm{N}(\mathfrak{p}_k) \ge 2^k$, since the norm of each $\mathfrak{p}_i$ is a prime power, and in particular, at least 2. It follows that $k \le \log_2 M$ is bounded, independent of I. Each prime ideal $\mathfrak{p}$ lies above some prime $p \le M$, of which there are fewer than M, and for each prime p the number of primes $\mathfrak{p}|p$ is at most n. Thus there are fewer than $(nM)^{\log_2 M}$ ideals of norm at most M in $\mathcal{O}_K$. □

Corollary 14.21. *Let K be a number field. The ideal class group of $\mathcal{O}_K$ is finite.*

Proof. By Theorem 14.19, each ideal class is represented by an ideal of norm at most m_K, and by Lemma 14.20, the number of such ideals is finite. □

Remark 14.22. The geometry of numbers is not a necessary ingredient to Corollary 14.21, there are purely algebraic proofs that apply to any global field; see [10] for an example.

Remark 14.23. For imaginary quadratic fields $K = \mathbb{Q}(\sqrt{-d})$ it is known that the *class number* $h_K := \#\,\mathrm{cl}\,\mathcal{O}_K$ tends to infinity as $d \to \infty$ ranges over square-free integers. This was conjectured by Gauss in his *Disquisitiones Arithmeticae* [3] and proved by Heilbronn [5] in 1934; the first fully explicit lower bound was obtained by Oesterlé in 1988 [7]. This implies that there are only a finite number of imaginary quadratic fields with any particular class number. It was conjectured by Gauss that there are exactly 9 imaginary quadratic fields with class number one, but this was not proved until the 20th century by Stark [9] and Heegner [4].[1] Complete lists of imaginary quadratic fields for each class number $h_K \le 100$ are now available [12]. By contrast, Gauss predicted that infinitely many real quadratic fields should have class number 1, however this question remains completely open.[2]

Corollary 14.24. *Let K be a number field of degree n with s complex places. Then*

$$|D_K| \ge \left(\frac{n^n}{n!}\right)^2 \left(\frac{\pi}{4}\right)^{2s} > \frac{1}{e^2 n} \left(\frac{\pi e^2}{4}\right)^n.$$

Proof. If I is an ideal and $a \in I$ is nonzero, then $\mathrm{N}(a) \ge \mathrm{N}(I)$, so Theorem 14.19 implies

$$m_K = \frac{n!}{n^n} \left(\frac{4}{\pi}\right)^s \sqrt{|D_K|} \ge 1,$$

the first inequality follows. The second uses an explicit form of Stirling's approximation,

$$n! \le e\sqrt{n} \left(\frac{n}{e}\right)^n,$$

and the fact that $2s \le n$. □

[1] Heegner's 1952 result [4] was essentially correct but contained some gaps that prevented it from being generally accepted until 1967 when Stark gave a complete proof in [9].

[2] In fact it is conjectured that $h_K = 1$ for approximately 75.446% of real quadratic fields with prime discriminant; this follows from the Cohen-Lenstra heuristics [2].

We note that $\pi e^2/4 \approx 5.8 > 1$, so the minimum value of $|D_K|$ increases exponentially with $n = [K:\mathbb{Q}]$. The lower bounds for $n \in [2,7]$ given by the corollary are listed below, along with the least value of $|D_K|$ that actually occurs. As can be seen in the table, $|D_K|$ appears to grow much faster than the corollary suggests. Better lower bounds can be proved using more advanced techniques, but a significant gap still remains.

	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$
lower bound from Corollary 14.24	3	13	44	259	986	6267
minimum value of $\lvert D_K\rvert$	3	23	275	4511	92799	2306599

Corollary 14.25. *If K is a number field other than $\mathbb{Q}$ then $|D_K| > 1$; equivalently, there are no nontrivial unramified extensions of $\mathbb{Q}$.*

Theorem 14.26. *For every real M the set of number fields K with $|D_K| < M$ is finite.*

Proof. It follows from Corollary 14.24 that it suffices to prove this for fixed $n := [K:\mathbb{Q}]$, since for all sufficiently large n we will have $|D_K| > M$ for all number fields K of degree n.

Case 1: Let K be a totally real field (so every place $v|\infty$ is real) with $|D_K| < M$. Then $r = n$ and $s = 0$, so $K_{\mathbb{R}} \simeq \mathbb{R}^r \times \mathbb{C}^s = \mathbb{R}^n$. Consider the convex symmetric set

$$S := \{(x_1,\dots,x_n) \in K_{\mathbb{R}} \simeq \mathbb{R}^n : |x_1| \le \sqrt{M} \text{ and } |x_i| < 1 \text{ for } i > 1\}$$

with measure

$$\mu(S) = 2\sqrt{M}2^{n-1} = 2^n\sqrt{M} > 2^n\sqrt{|D_K|} = 2^n \operatorname{covol}(\mathcal{O}_K).$$

By Theorem 14.12, the set S contains a nonzero $a \in \mathcal{O}_K \subseteq K \hookrightarrow K_{\mathbb{R}}$ that we may write as $a = (a_1,\dots,a_n) = (\sigma_1(a),\dots,\sigma_n(a))$, where the σ_i are the n embeddings of K into $\mathbb{C}$, all of which are real embeddings. We have

$$\mathrm{N}(a) = \left|\prod_i \sigma_i(a)\right| \ge 1,$$

since $\mathrm{N}(a)$ must be a positive integer, and $|a_2|,\dots,|a_n| < 1$, so $|a_1| > 1 > |a_i|$ for all $i \ne 1$.

We claim that $K = \mathbb{Q}(a)$. If not, each $a_i = \sigma_i(a)$ would be repeated $[K:\mathbb{Q}(a)] > 1$ times in the vector $(a_1,\dots,a_n)$, since there must be $[K:\mathbb{Q}(a)]$ elements of $\mathrm{Hom}_{\mathbb{Q}}(K,\mathbb{C})$ that fix $\mathbb{Q}(a)$, namely, those lying in the kernel of the map $\mathrm{Hom}_{\mathbb{Q}}(K,\mathbb{C}) \to \mathrm{Hom}_{\mathbb{Q}}(\mathbb{Q}(a),\mathbb{C})$ induced by restriction. But this is impossible since $a_i \ne a_1$ for $i \ne 1$.

The minimal polynomial $f \in \mathbb{Z}[x]$ of a is a monic irreducible polynomial of degree n. The roots of $f(x)$ in $\mathbb{C}$ are precisely the $a_i = \sigma_i(a) \in \mathbb{R}$, all of which are bounded by $|a_i| \le \sqrt{M}$. Each coefficient f_i of $f(x)$ is an elementary symmetric functions of its roots, hence also bounded in absolute value (certainly $|f_i| \le 2^n M^{n/2}$ for all i). The f_i are integers, so there are only finitely many possibilities for $f(x)$, hence only finitely many totally real number fields K of degree n.

Case 2: K has r real and $s > 0$ complex places, and $K_{\mathbb{R}} \simeq \mathbb{R}^r \times \mathbb{C}^s$. Now let

$$S := \{(w_1,\dots,w_r,\ z_1,\ \dots,\ z_s) \in K_{\mathbb{R}} : |z_1|^2 < c\sqrt{M} \text{ and } |w_i|, |z_j| < 1 \ (j > 1)\}$$

with c chosen so that $\mu(S) > 2^n \operatorname{covol}(\mathcal{O}_K)$ (the exact value of c depends on s and n). The argument now proceeds as in case 1: we get a nonzero $a \in \mathcal{O}_K \cap S$ for which $K = \mathbb{Q}(a)$, and only a finite number of possible minimal polynomials $f \in \mathbb{Z}[x]$ for a. □

Lemma 14.27. *Let K be a number field of degree n. For each prime number p we have*

$$v_p(D_K) \le n\lfloor \log_p n \rfloor + n - 1.$$

In particular, $v_p(D_K) \le n\lfloor \log_2 n \rfloor + n - 1$ for all p.

Proof. We have

$$v_p(D_K) = v_p(N_{K/\mathbb{Q}}(\mathcal{D}_{K/\mathbb{Q}})) = \sum_{\mathfrak{q}|p} f_{\mathfrak{q}} v_{\mathfrak{q}}(\mathcal{D}_{K/\mathbb{Q}})$$

where $\mathcal{D}_{K/\mathbb{Q}}$ is the different ideal and $f_{\mathfrak{q}}$ is the residue degree of $\mathfrak{q}|p$. Using Theorem 12.26 to bound $v_{\mathfrak{q}}(\mathcal{D}_{K/\mathbb{Q}})$ yields

$$v_p(D_K) \le \sum_{\mathfrak{q}|p} f_{\mathfrak{q}}(e_{\mathfrak{q}} - 1 + v_{\mathfrak{q}}(e_{\mathfrak{q}})) = n - \sum_{q|p} f_{\mathfrak{q}} + \sum_{\mathfrak{q}|p} f_{\mathfrak{q}} e_{\mathfrak{q}} v_p(e_{\mathfrak{q}}) \le n - 1 + n\lfloor \log_p n \rfloor,$$

where we have used -1 as an upper bound on $-\sum_{\mathfrak{q}|p} f_{\mathfrak{q}}$ and $\lfloor \log_p n \rfloor$ as an upper bound on each $v_p(e_{\mathfrak{q}})$ (since $e_{\mathfrak{q}} \le n$), and the fact that $\sum_{\mathfrak{q}|p} e_{\mathfrak{q}} f_{\mathfrak{q}} = n$ (by Theorem 5.35). □

Remark 14.28. The bound in Lemma 14.27 is tight; it is achieved by $K = \mathbb{Q}[x]/(x^{p^e} - p)$, for example.

Theorem 14.29 (Hermite)**.** *Let S be a finite set of places of $\mathbb{Q}$, and let n be an integer. The number of extensions $K/\mathbb{Q}$ of degree n unramified outside of S is finite.*

Proof. By Lemma 14.27, since n is fixed, the valuation $v_p(D_K)$ is bounded for each $p \in S$ and must be zero for $p \notin S$. Thus $|D_K|$ is bounded, and the theorem then follows from Proposition 14.26. □

Remark 14.30. In the function field analogs of Theorem 14.26 and Theorem 14.29 one requires K to be a separable extension of $\mathbb{F}_q(t)$ with constant field $\mathbb{F}_q$ (so $K \cap \overline{\mathbb{F}}_q = \mathbb{F}_q$). This is not really a restriction in the sense that every global function field K contains a subfield $\mathbb{F}_q(t)$ for which this is true, but one needs to take $q = \#(K \cap \overline{\mathbb{F}}_q)$ and to choose t to be a separating element (such a t exists by [6, Thm. 7.20]). Unlike the number field setting where the embedding of the rational numbers $\mathbb{Q}$ in a number field K is unique, there are many ways to embed the rational function field $\mathbb{F}_q(t)$ in a global function field K. The notion of an absolute discriminant D_K doesn't really make sense in this setting, one can speak of the discriminant $D_{K/\mathbb{F}_q(t)}$ only after fixing a suitable choice of $\mathbb{F}_q(t)$. As you showed on Problem Set 6, the valuation of the discriminant of an extension of global function fields is not bounded as a function of the degree, in general, and this means that the function field analog of Lemma 14.27 only holds when we use the discriminant of a separable extension.

References

[1] Nicolas Bourbaki, *General Topology: Chapters 1-4*, Springer, 1995.

[2] Henri Cohen and Hendrik W. Lenstra Jr., *Heuristics on class groups of number fields*, in *Number Theory (Noordwijkerhout 1983)*, Lecture Notes in Mathematics **1068**, Springer, 1984, 33–62.

[3] Carl F. Gauss, *Disquisitiones Arithmeticae*, Göttingen (1801), English translation by Arthur A. Clark, revised by William C. Waterhouse, Spring-Verlag 1986 reprint of Yale University Press 1966 edition.

[4] Kurt Heegner, *Diophantische Analysis und Modulfunktionen*, Math. Z. **56** (1952), 227–253.

[5] Hans Heilbronn, *On the class number in imaginary quadratic fields*, Quart. J. of Math. Oxford **5** (1934), 150–160.

[6] Anthony W. Knapp, *Advanced Algebra*, Digital Second Edition, 2016.

[7] Joseph Oesterlé, *La probléme de Gauss sur le nombre de classes*, Enseign. Math. **34** (1988), 43–67.

[8] Michael Rosen, *A geometric proof of Hermite's theorem in function fields*, J. Théor. Nombres Bordeaux **29** (2017), 799–813.

[9] Harold Stark, *A complete determination of the complex quadratic fields of class–number one*, Mich. Math. J. **14** (1967), 1–27.

[10] Alexander Stasinski, *Finiteness of the class group of basic arithmetic rings*, arXiv:1909.07121v1.

[11] Terence Tao, *An introduction to measure theory*, Graduate Studies in Mathematics **126**, AMS, 2010.

[12] Mark Watkins, *Class numbers of imaginary quadratic fields*, Math. Comp. **73** (2004), 907–938.

15 Dirichlet's unit theorem

Let K be a number field. The two main theorems of classical algebraic number theory are:

- The class group $\operatorname{cl}\mathcal{O}_K$ is finite.
- The unit group $\mathcal{O}_K^\times$ is finitely generated.

We proved the first result in the previous lecture; in this lecture we will prove the second, due to Dirichlet. Dirichlet (1805–1859) died five years before Minkowski (1864–1909) was born, so he did not have Minkowski's lattice point theorem (Theorem 14.12) to work with. But we do, and this simplifies the proof considerably.

15.1 The group of Arakelov divisors of a global field

Let K be a global field. As in previous lectures, we use M_K to denote the set of places (equivalence classes of absolute values) of K. For each place $v \in M_K$ we use K_v to denote the completion of K with respect to v (a local field), and we have a normalized absolute value $\| \ \|_v \colon K_v \to \mathbb{R}_{\geq 0}$ defined by

$$\|x\|_v := \frac{\mu(xS)}{\mu(S)},$$

where μ is a Haar measure on K_v and S is any measurable set of positive finite measure. This definition does not depend on the particular choice of μ or S; it is determined by the topology of K_v, which is an invariant of the place v (see Definition 13.17).

When K_v is nonarchimedean its topology is induced by a discrete valuation that we also denote v, and we use k_v to denote the residue field (the quotient of the valuation ring by its maximal ideal), which is a finite field (see Proposition 9.6). In Lecture 13 we showed that

$$\|x\|_v = \begin{cases} |x|_v = (\#k_v)^{-v(x)} & \text{if } v \text{ is nonarchimedean,} \\ |x|_{\mathbb{R}} & \text{if } K_v \simeq \mathbb{R}, \\ |x|_{\mathbb{C}}^2 & \text{if } K_v \simeq \mathbb{C}. \end{cases}$$

While $\| \ \|_v$ is not always an absolute value (when $K_v \simeq \mathbb{C}$ it does not satisfy the triangle inequality), it is always multiplicative and defines a continuous homomorphism $K_v^\times \to \mathbb{R}_{>0}^\times$ of locally compact groups that is surjective precisely when v is archimedean.

Definition 15.1. Let K be a global field. A (multiplicative) *Arakelov divisor* (or M_K*-divisor*) is a sequence of positive real numbers $c = (c_v)$ indexed by $v \in M_K$ with all but finitely many $c_v = 1$ and $c_v \in \|K_v^\times\| := \{\|x\|_v : x \in K_v^\times\}$.[1] The set of Arakelov divisors $\operatorname{Div} K$ forms an abelian group under pointwise multiplication $(c_v)(d_v) := (c_v d_v)$. The multiplicative group $K^\times$ is canonically embedded in $\operatorname{Div} K$ via the map $x \mapsto (\|x\|_v)$, where it forms the subgroup of *principal Arakelov divisors*.

Remark 15.2. Many authors define $\operatorname{Div} K$ as an additive group by taking logarithms (for nonarchimedean places v, one replaces $c_v = (\#k_v)^{-v(c)}$ with the integer $v(c)$), as in [4] for example. The multiplicative convention we use here is due to Weil [5] and better suited to our application to the multiplicative group $\mathcal{O}_K^\times$.[2]

[1] When v is archimedean we have $\|K_v^\times\| = \mathbb{R}_{>0}$ and this constraint is automatically satisfied.
[2] Weil calls them K*-divisors* [5, p. 422], while Lang uses M_K*-divisors* [2].

Definition 15.3. Let K be a global field. The *size* of an Arakelov divisor c is the real number

$$\|c\| := \prod_{v \in M_K} c_v \in \mathbb{R}_{>0}.$$

The map $\operatorname{Div} K \to \mathbb{R}_{>0}^{\times}$ defined by $c \mapsto \|c\|$ is a group homomorphism that contains the subgroup of principal Arakelov divisors in its kernel (by the product formula, Theorem 13.21). Corresponding to each Arakelov divisor c is a subset $L(c)$ of K defined by

$$L(c) := \{x \in K : \|x\|_v \le c_v \text{ for all } v \in M_K\}.$$

and a nonzero fractional ideal of $\mathcal{O}_K$ defined by

$$I_c := \prod_{v \nmid \infty} \mathfrak{q}_v^{v(c)},$$

where $\mathfrak{q}_v := \{a \in \mathcal{O}_K : v(a) > 0\}$ is the prime ideal corresponding to the discrete valuation v that induces $\| \ \|_v$, and $v(c) := -\log_{\#k_v}(c_v) \in \mathbb{Z}$ (so $v(x) = v(c)$ if and only if $\|x\|_v = c_v$). We have $L(c) \subseteq I_c \subseteq K$, and the map $c \mapsto I_c$ defines a group homomorphism $\operatorname{Div} K \to \mathcal{I}_K$. Observe that to specify an Arakelov divisor c it suffices to specify the fractional ideal I_c and the real numbers $c_v > 0$ for $v|\infty$ (a finite set).

Remark 15.4. The quotient of $\operatorname{Div} K$ by its subgroup of principal divisors is denoted $\operatorname{Pic} K$. The homomorphism $\operatorname{Div} K \to \mathcal{I}_K$ sends principal Arakelov divisors to principal fractional ideals, and it follows that the ideal class group $\operatorname{cl} \mathcal{O}_K$ is a quotient of $\operatorname{Pic} K$. We have a commutative diagram

$$\begin{array}{ccc} \operatorname{Div} K & \longrightarrow & \mathcal{I}_K \\ \downarrow & & \downarrow \\ \operatorname{Pic} K & \longrightarrow & \operatorname{cl} \mathcal{O}_K. \end{array}$$

The Arakelov divisors of size 1 form a subgroup of $\operatorname{Div} K$ denoted $\operatorname{Div}^0 K$ that contains the subgroup of principal divisors and surjects onto $\mathcal{I}_K$ via the map $\operatorname{Div} K \to \mathcal{I}_K$ (we are free to choose any $I_c \in \mathcal{I}_K$ because we can always choose the c_v at infinite places to ensure $\|c\| = 1$). The quotient of $\operatorname{Div}^0 K$ by the subgroup of principal Arakelov divisors is the *Arakelov class group* $\operatorname{Pic}^0 K$, which admits the ideal class group $\operatorname{cl} \mathcal{O}_K$ as a finite quotient. See [4] for more background on Arakelov class groups and how to compute them.

Remark 15.5. The set $L(c)$ associated to an Arakelov divisor c is directly analogous to the *Riemann-Roch space*

$$L(D) := \{f \in k(X) : v_P(f) \ge -n_P \text{ for all closed points } P \in X\},$$

associated to a divisor $D \in \operatorname{Div} X$ of a smooth projective curve X/k, which is a k-vector space of finite dimension. Recall that a divisor is a formal sum $D = \sum n_P P$ over the closed points ($\operatorname{Gal}(\bar{k}/k)$-orbits) of the curve X with $n_P \in \mathbb{Z}$ and all but finitely many n_P zero.

If k is a finite field then $K = k(X)$ is a global field and there is a one-to-one correspondence between closed points of X and places of K, and a normalized absolute value $\| \ \|_P$ for each closed point P (indeed, one can take this as a definition). The constraint $v_P(f) \ge -n_P$ is equivalent to $\|f\|_P \le (\#k_P)^{n_P}$, where k_P is the residue field corresponding

to P. If we put $c_P := (\#k_P)^{n_P}$ then $c = (c_P)$ is an Arakelov divisor with $L(c) = L(D)$. The Riemann-Roch space $L(D)$ is finite (since k is finite), and we will prove below that $L(c)$ is also finite (but note that when K is a number field the finite set $L(c)$ is not a vector space).

In §6.3 we described the divisor group $\operatorname{Div} X$ as the additive analog of the ideal group of the ring of integers $A = \mathcal{O}_K$, equivalently, the coordinate ring $A = k[X]$, of the global function field $K = k(X)$. When X is a smooth projective curve this is not a perfect analogy because divisors in $\operatorname{Div} X$ may include terms corresponding to "points at infinity" which do not correspond to fractional ideal of A. The group of Arakelov divisors $\operatorname{Div} K$ takes these infinite places into account and is a more exact analog of $\operatorname{Div} X$ when X is a smooth projective curve over a finite field.

We now specialize to the case where K is a number field. Recall that the absolute norm $\mathrm{N}(I)$ of a fractional ideal of $\mathcal{O}_K$ is the unique $t \in \mathbb{Q}_{>0}$ for which $N_{\mathcal{O}_K/\mathbb{Z}}(I) = (t)$. We have

$$\mathrm{N}(I_c) = \prod_{v\nmid\infty} \mathrm{N}(\mathfrak{q}_v)^{v(c)} = \prod_{v\nmid\infty} (\#k_v)^{v(c)} = \prod_{v\nmid\infty} c_v^{-1},$$

and therefore

$$\|c\| = \mathrm{N}(I_c)^{-1} \prod_{v|\infty} c_v, \tag{1}$$

We also define

$$R_c := \{x \in K_{\mathbb{R}} : |x|_v \le c_v \text{ for all } v|\infty\},$$

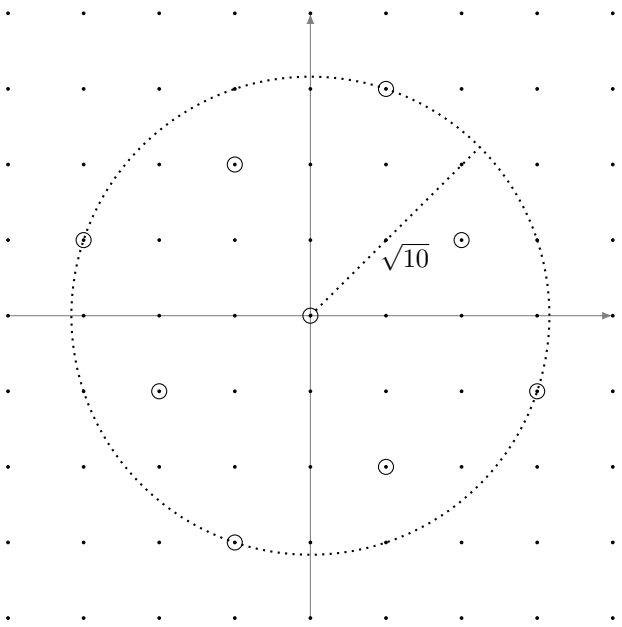
which we note is a compact, convex, symmetric subset of the real vector space

$$K_{\mathbb{R}} := K \otimes_{\mathbb{Q}} \mathbb{R} \simeq \mathbb{R}^r \times \mathbb{C}^s,$$

where r is the number of real places of K, and s is the number of complex places. If we view I_c and $L(c)$ as subsets of $K_{\mathbb{R}}$ via the canonical embedding $K \hookrightarrow K_{\mathbb{R}}$, then

$$L(c) = I_c \cap R_c.$$

Example 15.6. Let $K = \mathbb{Q}(i)$. The ideal $(2+i)$ lying above 5 is prime and corresponds to a finite place v_1, and there is a unique infinite place $v_2|\infty$ which is complex. Let $c_{v_1} = 1/5$, let $c_{v_2} = 10$, and set $c_v = 1$ for all other $v \in M_K$. We then have $I_c = (2+i)$ and the image of $L(c) = \{x \in (2+i) : |x|_\infty \le 10\}$ under the canonical embedding $K \hookrightarrow K_{\mathbb{R}} \simeq \mathbb{C}$ is the set of lattice points in the image of the ideal I_c that lie within the circle $R_c \subseteq K_{\mathbb{R}} \simeq \mathbb{C}$ of radius $\sqrt{10}$. Note that $\| \ \|_{v_2} = | \ |_{\mathbb{C}}^2$ is the square of the usual absolute value on $\mathbb{C}$, which is why the circle has radius $\sqrt{10}$ rather than 10.



The set $L(c)$ is clearly finite; it contains exactly 9 points.

Lemma 15.7. *Let c be an Arakelov divisor of a global field K. The set $L(c)$ is finite.*

Proof. We assume K is a number field; see Problem Set 7 for the function field case. The fractional ideal I_c is a lattice in $K_{\mathbb{R}}$ (under the canonical embedding $K \hookrightarrow K_{\mathbb{R}}$), and is thus a closed discrete subset of $K_{\mathbb{R}}$ (recall from Remark 14.4 that lattices are closed). In $K_{\mathbb{R}}$ we may view $L(c) = I_c \cap R_c$ as the intersection of a discrete closed set with a compact set, which is a compact discrete set and therefore finite. □

Corollary 15.8. *Let K be a global field, and let μ_K denote the torsion subgroup of $K^\times$ (equivalently, the roots of unity in K). The group μ_K is finite and equal to the kernel of the map $K^\times \to \operatorname{Div} K$ defined by $x \mapsto (\|x\|_v)$; it is also the torsion subgroup of $\mathcal{O}_K^\times$.*

Proof. Each $\zeta \in \mu_K$ satisfies $\zeta^n = 1$ for some positive integer n. For every place $v \in M_K$ we have $\|\zeta^n\|_v = \|\zeta\|_v^n = 1$, and therefore $\|\zeta\|_v = 1$. It follows that $\mu_K \subseteq \ker(K^\times \to \operatorname{Div} K)$. Let c be the Arakelov divisor with $c_v = 1$ for all $v \in M_K$. Then $\ker(K^\times \to \operatorname{Div} K) \subseteq L(c)$ is a finite subgroup of $K^\times$ and is therefore contained in the torsion subgroup μ_K. Every element of μ_K is an algebraic integer (in fact a root of $x^n - 1$), so $\mu_K \subseteq \mathcal{O}_K^\times$. □

It follows from Corollary 15.8 that for any global field K we have the following exact sequence of abelian groups

$$1 \longrightarrow \mu_K \longrightarrow K^\times \longrightarrow \operatorname{Div} K \longrightarrow \operatorname{Pic}_K \longrightarrow 1.$$

Proposition 15.9. *Let K be a number field with s complex places, define*

$$B_K := \left(\frac{2}{\pi}\right)^s \sqrt{|D_K|}.$$

If c is an Arakelov divisor of size $\|c\| > B_K$ then $L(c)$ contains an element of $K^\times$.

Proof. Our strategy is to apply Minkowski's lattice point theorem (see Theorem 14.12) to the convex symmetric set R_c and the lattice $I_c \subseteq K \subseteq K_{\mathbb{R}}$; we just need to show that if $\|c\| > B_K$ then the ratio of the Haar measure of R_c to the covolume of I_c exceeds 2^n, where $n = r + 2s$ is the degree of K (which is the real dimension of $K_{\mathbb{R}}$). As defined in §14.2, we normalize the Haar measure μ on the locally compact group $K_{\mathbb{R}} \simeq \mathbb{R}^r \times \mathbb{C}^s \simeq \mathbb{R}^n$ so that $\mu(S) = 2^s \mu_{\mathbb{R}^n}(S)$ for measurable $S \subseteq K_{\mathbb{R}}$. For each real place v, the constraint $\|x\|_v = |x|_{\mathbb{R}} \le c_v$ contributes a factor of $2c_v$ to $\mu(R_c)$, and for each complex place v the constraint $\|x\|_v = |x|_{\mathbb{C}}^2 \le c_v$ contributes a factor of πc_v (the area of a circle of radius $\sqrt{c_v}$). We may then compute

$$\begin{aligned}\frac{\mu(R_c)}{\operatorname{covol}(I_c)} &= \frac{2^s \mu_{\mathbb{R}^n}(R_c)}{\operatorname{covol}(I_c)} = \frac{2^s \left(\prod_{v \text{ real}} 2c_v\right)\left(\prod_{v \text{ complex}} \pi c_v\right)}{\operatorname{covol}(I_c)} \\ &= \frac{2^r (2\pi)^s \prod_{v|\infty} c_v}{\sqrt{|D_K|}\mathrm{N}(I_c)} = \frac{2^r (2\pi)^s}{\sqrt{|D_K|}}\|c\| = \frac{\|c\|}{B_K} 2^n > 2^n\end{aligned}$$

where we have used Corollary 14.16 and (1) in the second line. Theorem 14.12 implies that $L(c) = R_c \cap I_c$ contains a nonzero element (which lies in $K^\times \subseteq K_{\mathbb{R}}$, since $I_c \subseteq K \subseteq K_{\mathbb{R}}$). □

Remark 15.10. The bound in Proposition 15.9 can be turned into an asymptotic, that is, for $c \in \operatorname{Div} K$, as $\|c\| \to \infty$ we have

$$\#L(c) = \left(\frac{2^r (2\pi)^s}{\sqrt{|D_K|}} + o(1)\right)\|c\|. \tag{2}$$

This can be viewed as a multiplicative analog of the Riemann-Roch theorem for function fields, which states that for divisors $D = \sum n_P P$, as $\deg D := \sum n_P \to \infty$ we have

$$\dim L(D) = 1 - g + \deg D. \tag{3}$$

The nonnegative integer g is the *genus*, an important invariant of a function field that is often defined by (3); one could similarly use (2) to define the nonnegative integer $|D_K|$. For all sufficiently large $\|c\|$ the $o(1)$ error term will be small enough so that (2) uniquely determines $|D_K|$. Conversely, with a bit more work one can adapt the proofs of Lemma 15.7 and Proposition 15.9 to give a proof of the Riemann-Roch theorem for global function fields.

15.2 The unit group of a number field

Let K be a number field with ring of integers $\mathcal{O}_K$. The multiplicative group $\mathcal{O}_K^\times$ is the *unit group* of $\mathcal{O}_K$, and may also be called the unit group of K. Of course the unit group of the ring K is $K^\times$, but this is typically referred to as the multiplicative group of K.

As a ring, the finite étale $\mathbb{R}$-algebra $K_{\mathbb{R}} = K \otimes_{\mathbb{Q}} \mathbb{R}$ also has a unit group, and we have an isomorphism of topological groups[3]

$$K_{\mathbb{R}}^\times \simeq \prod_{v|\infty} K_v^\times \simeq \prod_{\text{real } v|\infty} \mathbb{R}^\times \prod_{\text{complex } v|\infty} \mathbb{C}^\times = (\mathbb{R}^\times)^r \times (\mathbb{C}^\times)^s.$$

[3]The additive group of $K_{\mathbb{R}}$ is isomorphic to $\mathbb{R}^n$ as a topological group (and $\mathbb{R}$-vector space), a fact we have used in our study of lattices in $K_{\mathbb{R}}$. But as topological rings $K_{\mathbb{R}} \simeq \mathbb{R}^r \times \mathbb{C}^s \not\simeq \mathbb{R}^n$ unless $s = 0$.

Writing elements of $K_{\mathbb{R}}^\times$ as vectors $x = (x_v)$ indexed by the infinite places v of K, we now define a surjective homomorphism of locally compact groups

$$\begin{aligned} \operatorname{Log}\colon K_{\mathbb{R}}^\times &\to \mathbb{R}^{r+s} \\ (x_v) &\mapsto (\log \|x_v\|_v). \end{aligned}$$

It is surjective and continuous because each of the maps $x_v \mapsto \log \|x_v\|_v$ is, and it is a group homomorphism because

$$\operatorname{Log}(xy) = (\log \|x_v y_v\|_v) = (\log \|x_v\|_v + \log \|y_v\|_v) = (\log \|x_v\|_v) + (\log \|y_v\|_v) = \operatorname{Log} x + \operatorname{Log} y;$$

here we have used the fact that the normalized absolute value $\| \ \|_v$ is multiplicative.

Recall from Corollary 13.7 that there is a one-to-one correspondence between the infinite places of K and the $\operatorname{Gal}(\mathbb{C}/\mathbb{R})$-orbits of $\operatorname{Hom}_{\mathbb{Q}}(K, \mathbb{C})$. For each $v|\infty$ let us now pick a representative σ_v of its corresponding $\operatorname{Gal}(\mathbb{C}/\mathbb{R})$-orbit in $\operatorname{Hom}_{\mathbb{Q}}(K, \mathbb{C})$; for real places v there is a unique choice for σ_v, while for complex places there are two choices, σ_v and its complex conjugate $\bar{\sigma}_v$. Regardless of our choices, we then have

$$\|x\|_v = \begin{cases} |\sigma_v(x)|_{\mathbb{R}} & \text{if } v|\infty \text{ is real} \\ |\sigma_v(x)\bar{\sigma}_v(x)|_{\mathbb{R}} & \text{if } v|\infty \text{ is complex.} \end{cases}$$

The absolute norm $\operatorname{N}\colon K^\times \to \mathbb{Q}_{>0}^\times$ extends naturally to a continuous homomorphism of locally compact groups

$$\begin{aligned} \operatorname{N}\colon K_{\mathbb{R}}^\times &\to \mathbb{R}_{>0}^\times \\ (x_v) &\mapsto \prod_{v|\infty} \|x_v\|_v \end{aligned}$$

which is compatible with the canonical embedding $K^\times \hookrightarrow K_{\mathbb{R}}^\times$. Indeed, we have

$$\operatorname{N}(x) = \left|\operatorname{N}_{K/\mathbb{Q}}(x)\right| = \left|\prod_\sigma \sigma(x)\right|_{\mathbb{R}} = \prod_{v|\infty} \|x\|_v.$$

We thus have a commutative diagram

$$\begin{array}{ccccc} K^\times & \hookrightarrow & K_{\mathbb{R}}^\times & \xrightarrow{\operatorname{Log}} & \mathbb{R}^{r+s} \\ \downarrow{\scriptstyle \operatorname{N}} & & \downarrow{\scriptstyle \operatorname{N}} & & \downarrow{\scriptstyle \operatorname{T}} \\ \mathbb{Q}_{>0}^\times & \hookrightarrow & \mathbb{R}_{>0}^\times & \xrightarrow{\log} & \mathbb{R}, \end{array}$$

where $\operatorname{T}\colon \mathbb{R}^{r+s} \to \mathbb{R}$ is defined by $\operatorname{T}(x) = \sum_i x_i$. We may view Log as a map from $K^\times$ to $\mathbb{R}^{r+s}$ via the embedding $K^\times \hookrightarrow K_{\mathbb{R}}^\times$, and similarly view N as a map from $K^\times$ to $\mathbb{R}_{>0}^\times$.

We can succinctly summarize the commutativity of the above diagram by the identity

$$\operatorname{T}(\operatorname{Log} x) = \log \operatorname{N}(x),$$

which holds for all $x \in K^\times$, and all $x \in K_{\mathbb{R}}^\times$. The norm of a unit in $\mathcal{O}_K$ must be a unit in $\mathbb{Z}$, hence have absolute value 1. Thus $\mathcal{O}_K^\times$ lies in the kernel of the map $x \mapsto \log \operatorname{N}(x)$

and therefore also in the kernel of the map $x \mapsto \mathrm{T}(\mathrm{Log}\, x)$. It follows that $\mathrm{Log}(\mathcal{O}_K^\times)$ is a subgroup of the *trace zero hyperplane*

$$\mathbb{R}_0^{r+s} := \{x \in \mathbb{R}^{r+s} : \mathrm{T}(x) = 0\},$$

which we note is both a subgroup of $\mathbb{R}^{r+s}$, and an $\mathbb{R}$-vector subspace of dimension $r+s-1$. The proof of Dirichlet's unit theorem amounts to showing that $\mathrm{Log}(\mathcal{O}_K^\times)$ is a lattice in $\mathbb{R}_0^{r+s}$.

Proposition 15.11. *Let K be a number field with r real and s complex places, and let Λ_K be the image of the unit group $\mathcal{O}_K^\times$ in $\mathbb{R}_0^{r+s}$ under the* Log *map. The following hold:*

(1) *We have a split exact sequence of finitely generated abelian groups*

$$1 \to \mu_K \to \mathcal{O}_K^\times \xrightarrow{\mathrm{Log}} \Lambda_K \to 0;$$

(2) *Λ_K is a lattice in the trace zero hyperplane $\mathbb{R}_0^{r+s}$.*

Here μ_K is not a Haar measure, it denotes the group of roots of unity in K, all of which are clearly torsion elements of $\mathcal{O}_K^\times$, and any torsion element of $\mathcal{O}_K^\times$ is clearly a root of unity.

Proof. (1) We first show exactness. Let Z be the kernel of $\mathcal{O}_K^\times \xrightarrow{\mathrm{Log}} \Lambda_K$. Clearly $\mu_K \subseteq Z$, since $\Lambda_K \subseteq \mathbb{R}_0^{r+s}$ is torsion free. Let c be the Arakelov divisor with $I_c = \mathcal{O}_K$ and $c_v = 2$ for $v|\infty$, so that

$$L(c) = \{x \in \mathcal{O}_K : \|x\|_v \le 2 \text{ for all } v|\infty\}.$$

For $x \in \mathcal{O}_K^\times$ we have

$$x \in L(c) \iff \mathrm{Log}(x) \in \mathrm{Log}\, R_c = \{z \in \mathbb{R}^{r+s} : z_i \le \log 2\}.$$

The set on the RHS includes the zero vector, thus $Z \subseteq L(c)$, which by Lemma 15.7 is a finite set. As a finite subgroup of $\mathcal{O}_K^\times$, we must have $Z \subseteq \mu_K$, so $Z = \mu_K$ and the sequence is exact (the map from $\mathcal{O}_K^\times$ to Λ_K is surjective by the definition of Λ_K).

We now show the sequence splits. Note that $\Lambda_K \cap \mathrm{Log}(R_c) = \mathrm{Log}\left(\mathcal{O}_K^\times \cap L(c)\right)$ is finite, since $L(c)$ is finite. It follows that 0 is an isolated point of Λ_K in $\mathbb{R}^{r+s}$, and in $\mathbb{R}_0^{r+s}$, so Λ_K is a discrete subgroup of the $\mathbb{R}$-vector space $\mathbb{R}_0^{r+s}$. It is therefore a free $\mathbb{Z}$-module of finite rank at most $r+s-1$, since it spans some subspace of $\mathbb{R}_0^{r+s}$ in which it is both discrete and cocompact, hence a lattice. It follows that $\mathcal{O}_K^\times$ is finitely generated, since it lies in a short exact sequence whose left and right terms are finitely generated (recall that μ_K is finite, by Corollary 15.8). By the structure theorem for finitely generated abelian groups, the sequence must split, since μ_K is the torsion subgroup of $\mathcal{O}_K^\times$.

(2) Having proved (1) it remains only to show that Λ_K spans $\mathbb{R}_0^{r+s}$. Let V be the subspace of $\mathbb{R}_0^{r+s}$ spanned by Λ_K and suppose for the sake of contradiction that $\dim V < \dim \mathbb{R}_0^{r+s}$. The orthogonal subspace $V^\perp$ then contains a unit vector u, and for every $\lambda \in \mathbb{R}_{>0}$ the open ball $B_{<\lambda}(\lambda u)$ does not intersect Λ_K. Thus $\mathbb{R}_0^{r+s}$ contains points arbitrarily far away from every point in Λ_K (with respect to any norm on $\mathbb{R}_0^{r+s} \subseteq \mathbb{R}^{r+s}$). To obtain a contradiction it is enough to show that there is a constant $M \in \mathbb{R}_{>0}$ such that for every $h \in \mathbb{R}_0^{r+s}$ there is an $\ell \in \Lambda_K$ for which $\|h - \ell\| := \max_i |h_i - \ell_i| < M$ (here we are using $\|\ \|$ to denote the sup norm on the $\mathbb{R}$-vector space $\mathbb{R}^{r+s}$).

Let us fix a real number $B > B_K$, where B_K is as in Proposition 15.9, so that for every $c \in \mathrm{Div}\, K$ with $\|c\| \ge B$ the set $L(c)$ contains a nonzero element, and fix a vector $b \in \mathbb{R}^{r+s}$

with nonnegative components b_i such that $\mathrm{T}(b) = \sum_i b_i = \log B$. Let $(\alpha_1), \dots, (\alpha_m)$ be the list of all nonzero principal ideals with $\mathrm{N}(\alpha_j) \le B$ (by Lemma 14.20 this is a finite list). Let M be twice the maximum of $(r+s)B$ and $\max_j \|\operatorname{Log}(\alpha_j)\|$.

Now let $h \in \mathbb{R}_0^{r+s}$, and define $c \in \operatorname{Div} K$ by $I_c := \mathcal{O}_K$ and $c_v := \exp(h_i + b_i)$ for $v|\infty$, where i is the coordinate in $\mathbb{R}^{r+s}$ corresponding to v under the Log map. We have

$$\|c\| = \prod_v c_v = \exp\Big(\sum_i (h_i + b_i)\Big) = \exp \mathrm{T}(h+b) = \exp(\mathrm{T}(h) + \mathrm{T}(b)) = \exp \mathrm{T}(b) = B > B_K,$$

thus $L(c)$ contains a nonzero $\gamma \in I_c \cap K = \mathcal{O}_K$, and $g = \operatorname{Log}(\gamma)$ satisfies $g_i \le \log c_v = h_i + b_i$. We also have $\mathrm{T}(g) = \mathrm{T}(\operatorname{Log}\gamma) = \log \mathrm{N}(\gamma) \ge 0$, since $\mathrm{N}(\gamma) \ge 1$ for all nonzero $\gamma \in \mathcal{O}_K$. The vector $v := g - h \in \mathbb{R}^{r+s}$ satisfies $\sum_i v_i = \mathrm{T}(v) = \mathrm{T}(g) - \mathrm{T}(h) = \mathrm{T}(g) \ge 0$ and $v_i \le b_i \le B$ which together imply $|v_i| \le (r+s)B$, so $\|g - h\| = \|v\| \le M/2$. We also have

$$\log \mathrm{N}(\gamma) = \mathrm{T}(\operatorname{Log}(\gamma)) \le \mathrm{T}(h+b) = \mathrm{T}(b) = \log B,$$

so $\mathrm{N}(\gamma) \le B$ and $(\gamma) = (\alpha_j)$ for one of the α_j fixed above. Thus $\gamma/\alpha_j \in \mathcal{O}_K^\times$ is a unit, and

$$\ell := \operatorname{Log}(\gamma/\alpha_j) = \operatorname{Log}(\gamma) - \operatorname{Log}(\alpha_j) \in \Lambda_K$$

satisfies $\|g - \ell\| = \|\operatorname{Log}(\alpha_j)\| \le M/2$. We then have

$$\|h - \ell\| \le \|h - g\| + \|g - \ell\| \le M$$

as desired (by the triangle inequality for the sup-norm). □

Dirichlet's unit theorem follows immediately from Proposition 15.11.

Theorem 15.12 (Dirichlet's Unit Theorem)**.** *Let K be a number field with r real and s complex places. Then $\mathcal{O}_K^\times \simeq \mu_K \times \mathbb{Z}^{r+s-1}$ is a finitely generated abelian group.*

Proof. The image of the torsion-free part of the unit group $\mathcal{O}_K^\times$ under the Log map is the lattice Λ_K in the trace-zero hyperplane $\mathbb{R}_0^{r+s}$, which has dimension $r + s - 1$. □

We can restate this theorem in a more general form so that it applies to all global fields. As usual, when we consider global function fields we view them as extensions of $\mathbb{F}_q(t)$, with q chosen so that $K \cap \overline{\mathbb{F}}_q = \mathbb{F}_q$ and t chosen so that $K/\mathbb{F}_q(t)$ is separable.

Theorem 15.13 (Unit Theorem for Global Fields)**.** *Let K/F be a finite separable extension, with $F = \mathbb{Q}$ or $F = \mathbb{F}_q(t)$, let $S \subseteq M_K$ be the set of places of K lying above the unique infinite place of F, and define $\mathcal{O}_K^\times := \{x \in K^\times : v(x) = 0 \text{ for all } v \in M_K - S\}$. Then $\mathcal{O}_K^\times \simeq \mu_K \times \mathbb{Z}^{\#S-1}$ is a finitely generated abelian group.*

Proof. For $F = \mathbb{Q}$ we have $\#S = r + s$ and this is simply Dirichlet's unit theorem; for $F = \mathbb{F}_q(t)$, see [3, Prop. 14.1]. □

Remark 15.14. We should be careful in how we interpret 15.13 in the case $F = \mathbb{F}_q(t)$. By applying an automorphism of $\mathbb{F}_q(t)$ (replace t by $t - a$ for some $a \in \mathbb{F}_q$, say) we can move any degree-one place to infinity. This will change the group $\mathcal{O}_K^\times$ and may change the number of places of K above our new point at infinity. In contrast to the number field setting (where the place of $\mathbb{Q}$ at infinity is invariant because it is the only archimedean place) the ring $\mathcal{O}_K$ and the set S are not intrinsic to K in the function field setting; they depend on the choice of the separating element t used to construct the separable extension $K/\mathbb{F}_q(t)$.

Example 15.15. Let $K = \mathbb{Q}(\sqrt{d})$ be a quadratic field with $d \neq 1$ squarefree. If $d < 0$ then $r = 0$ and $s = 1$, in which case the unit group $\mathcal{O}_K^\times$ has rank 0 and $\mathcal{O}_K^\times = \mu_K$ is finite.

If $d > 0$ then $K = \mathbb{Q}(\sqrt{d}) \subseteq \mathbb{R}$ is a real quadratic field with $r = 2$ and $s = 0$, and the unit group $\mathcal{O}_K^\times$ has rank 1. The only torsion elements of $\mathcal{O}_K^\times \subseteq \mathbb{R}$ are ± 1, thus

$$\mathcal{O}_K^\times = \{\pm\epsilon^n : n \in \mathbb{Z}\},$$

for some $\epsilon \in \mathcal{O}_K^\times$ of infinite order. We may assume $\epsilon > 1$: if $\epsilon < 0$ then replace ϵ by $-\epsilon$, and if $\epsilon < 1$ then replace ϵ by ϵ^{-1} (we cannot have $\epsilon = 1 \in \mu_K$).

The assumption $\epsilon > 1$ uniquely determines ϵ. This follows from the fact that for $\epsilon > 1$ we have $|\epsilon^n| > |\epsilon|$ for all $n > 1$ and $|\epsilon^n| \leq 1$ for all $n \leq 0$.

This unique ϵ is the *fundamental unit* of $\mathcal{O}_K$ (and of K). To explicitly determine ϵ, let $D = \operatorname{disc} \mathcal{O}_K$ (so $D = d$ if $d \equiv 1 \bmod 4$ and $D = 4d$ otherwise). Every element of $\mathcal{O}_K$ can be uniquely written as

$$\frac{x + y\sqrt{D}}{2},$$

where x and Dy are integers of the same parity. In the case of a unit we must have $\mathrm{N}(\frac{x+y\sqrt{D}}{2}) = \pm 1$, equivalently,

$$x^2 - Dy^2 = \pm 4. \tag{4}$$

Conversely, any solution $(x, y) \in \mathbb{Z}^2$ to the above equation has x and Dy with the same parity and corresponds to an element of $\mathcal{O}_K^\times$. The constraint $\epsilon = \frac{x+y\sqrt{D}}{2} > 1$ forces $x, y > 0$. This follows from the fact that $\epsilon^{-1} = \frac{|x-y\sqrt{D}|}{2} < 1$, so $-2 < x - y\sqrt{D} < 2$, and adding and subtracting $x + y\sqrt{D} > 2$ shows $x > 0$ and $y > 0$ (respectively).

Thus we need only consider positive integer solutions (x, y) to (4). Among such solutions, $x_1 + y_1\sqrt{D} < x_2 + y_2\sqrt{D}$ implies $x_1 < x_2$, so the solution that minimizes x will give us the fundamental unit ϵ.

Equation (4) is a (generalized) *Pell equation*. Solving the Pell equation is a well-studied problem and there are a number of algorithms for doing so. The most well known uses continued fractions and is explored on Problem Set 7; this is not the most efficient method, but it is dramatically faster than an exhaustive search; see [1] for a comprehensive survey. A remarkable feature of this problem is that even when D is quite small, the smallest solution to (4) may be very large. For example, when $D = d = 889$ the fundamental unit is

$$\epsilon = \frac{26463949435607314430 + 887572376826907008\sqrt{889}}{2}.$$

15.3 The regulator of a number field

Let K be a number field with r real places and s complex places, and let $\mathbb{R}_0^{r+s}$ be the trace-zero hyperplane in $\mathbb{R}^{r+s}$. Choose any coordinate projection $\pi\colon \mathbb{R}^{r+s} \to \mathbb{R}^{r+s-1}$, and use the induced isomorphism $\mathbb{R}_0^{r+s} \xrightarrow{\sim} \mathbb{R}^{r+s-1}$ to endow $\mathbb{R}_0^{r+s}$ with a Euclidean measure. By Proposition 15.11, the image Λ_K of the unit group $\mathcal{O}_K^\times$ is a lattice in $\mathbb{R}_0^{r+s}$, and we can measure its covolume using the Euclidean measure on $\mathbb{R}_0^{r+s}$.

Definition 15.16. The *regulator* of a number field K is

$$R_K := \operatorname{covol}(\pi(\operatorname{Log}(\mathcal{O}_K^\times))) \in \mathbb{R}_{>0},$$

where $\pi\colon \mathbb{R}^{r+s} \to \mathbb{R}^{r+s-1}$ is any coordinate projection; the value of R_K does not depend on the choice of π, since we use π to normalize the Haar measure on $\mathbb{R}_0^{r+s} \simeq \mathbb{R}^{r+s-1}$. If $\epsilon_1, \dots, \epsilon_{r+s-1}$ is a fundamental system of units (a $\mathbb{Z}$-basis for the free part of $\mathcal{O}_K^\times$), then R_K can be computed as the absolute value of the determinant of any $(r+s-1) \times (r+s-1)$ minor of the $(r+s) \times (r+s-1)$ matrix whose columns are the vectors $\mathrm{Log}(\epsilon_i) \in \mathbb{R}^{r+s}$.

Example 15.17. If K is a real quadratic field with absolute discriminant D and fundamental unit $\epsilon = \frac{x+y\sqrt{D}}{2}$, then $r+s = 2$ and the product of the two real embeddings $\sigma_1(\epsilon), \sigma_2(\epsilon) \in \mathbb{R}$ is $\mathrm{N}(\epsilon) = \pm 1$. Thus $\log|\sigma_2(\epsilon)| = -\log|\sigma_1(\epsilon)|$ and

$$\mathrm{Log}(\epsilon) = (\log|\sigma_1(\epsilon)|, \log|\sigma_2(\epsilon)|) = (\log|\sigma_1(\epsilon)|, -\log|\sigma_1(\epsilon)|).$$

The 1×1 minors of the 2×1 transpose of $\mathrm{Log}(\epsilon)$ have determinant $\pm\log|\sigma_1(\epsilon)|$; the absolute value of the determinant is the same in both cases, and since we have require the fundamental unit to satisfy $\epsilon > 1$ (which forces a choice of embedding), the regulator of K is simply $R_K = \log\epsilon$.

References

[1] Michael J. Jacobson and Hugh C. Williams, *Solving the Pell equation*, Springer, 2009.

[2] Serge Lang, *Fundamentals of diophantine geometry*, Springer, 1983.

[3] Michael Rosen, *Number theory in function fields*, Springer, 2002.

[4] R. Schoof, *Computing Arakelov class groups*, in *Algorithmic Number Theory: lattices, number fields, curves, and cryptography*. MSRI Publications **44** (2008), 447–495.

[5] André Weil, *Arithmetic on algebraic varieties*, Annals of Mathematics (2) **53** (1951), 412–444.

16 Riemann's zeta function and the prime number theorem

We now divert our attention from algebraic number theory to talk about zeta functions and L-functions. As we shall see, every global field has a zeta function that is intimately related to the distribution of its primes. We begin with the zeta function of the rational field $\mathbb{Q}$, which we will use to prove the prime number theorem.

We will need some basic results from complex analysis, all of which can be found in any introductory textbook (such as [1, 2, 3, 7, 12]). A short glossary of terms and a list of the basic theorems we will use can be found at the end of these notes.[1]

16.1 The Riemann zeta function

Definition 16.1. The *Riemann zeta function* is the complex function defined by the series

$$\zeta(s) := \sum_{n\geq 1} n^{-s},$$

for $\mathrm{Re}(s) > 1$, where n varies over positive integers. It is easy to verify that this series converges absolutely and locally uniformly on $\mathrm{Re}(s) > 1$ (use the integral test on an open ball strictly to the right of the line $\mathrm{Re}(s) = 1$). By Theorem 16.17, it defines a holomorphic function on $\mathrm{Re}(s) > 1$, since each term $n^{-s} = e^{-s\log n}$ is holomorphic.

Theorem 16.2 (Euler product)**.** *For* $\mathrm{Re}(s) > 1$ *we have*

$$\zeta(s) = \sum_{n\geq 1} n^{-s} = \prod_p (1-p^{-s})^{-1},$$

where the product converges absolutely. In particular, $\zeta(s) \neq 0$ *for* $\mathrm{Re}(s) > 1$.

The product in the theorem above ranges over primes p. This is a standard practice in analytic number theory that we will follow: the symbol p always denotes a prime, and any sum or product over p is understood to be over primes, even if this is not explicitly stated.

Proof. We have

$$\sum_{n\geq 1} n^{-s} = \sum_{n\geq 1} \prod_p p^{-v_p(n)s} = \prod_p \sum_{e\geq 0} p^{-es} = \prod_p (1-p^{-s})^{-1}.$$

To justify the second equality, consider the *partial zeta function* $\zeta_m(s)$, which restricts the summation in $\zeta(s)$ to the set S_m of m-smooth integers (those with no prime factors $p > m$). If $p_1, \ldots, p_k$ are the primes up to m, absolute convergence implies

$$\zeta_m(s) := \sum_{n\in S_m} n^{-s} = \sum_{e_1,\ldots,e_k\geq 0} (p_1^{e_1}\cdots p_k^{e_k})^{-s} = \prod_{1\leq i\leq k} \sum_{e_i\geq 0} (p_i^{-s})^{e_i} = \prod_{p\leq m} (1-p^{-s})^{-1}.$$

For any $\delta > 0$ the sequence of functions $\zeta_m(s)$ converges uniformly on $\mathrm{Re}(s) > 1+\delta$ to $\zeta(s)$; indeed, for any $\epsilon > 0$ and any such s we have

$$|\zeta_m(s) - \zeta(s)| \leq \left|\sum_{n\geq m} n^{-s}\right| \leq \sum_{n\geq m} |n^{-s}| = \sum_{n\geq m} n^{-\mathrm{Re}(s)} \leq \int_m^\infty x^{-1-\delta} dx \leq \frac{1}{\delta} m^{-\delta} < \epsilon,$$

[1] Those familiar with this material should still glance at §16.3.2 which touches on some convergence issues that are particularly relevant to number theoretic applications.

for all sufficiently large m. It follows that the sequence $\zeta_m(s)$ converges locally uniformly to $\zeta(s)$ on $\mathrm{Re}(s) > 1$. The sequence of functions $P_m(s) := \prod_{p \le m}(1-p^{-s})^{-1}$ clearly converges locally uniformly to $\prod(1-p^{-s})^{-1}$ on any region in which the latter function is absolutely convergent (or even just convergent). For any s in $\mathrm{Re}(s) > 1$ we have

$$\sum_p |\log(1-p^{-s})^{-1}| = \sum_p \left| \sum_{e \ge 1} \frac{1}{e} p^{-es} \right| \le \sum_p \sum_{e \ge 1} |p^{-s}|^e = \sum_p (|p^s| - 1)^{-1} < \infty,$$

where we have used the identity $\log(1-z) = -\sum_{n \ge 1} \frac{1}{n} z^n$, valid for $|z| < 1$. It follows that $\prod_p (1-p^{-s})^{-1}$ is absolutely convergent (and in particular, nonzero) on $\mathrm{Re}(s) > 1$. □

Theorem 16.3 (Analytic continuation I). *For* $\mathrm{Re}(s) > 1$ *we have*

$$\zeta(s) = \frac{1}{s-1} + \phi(s),$$

where $\phi(s)$ *is a holomorphic function on* $\mathrm{Re}(s) > 0$. *Thus* $\zeta(s)$ *extends to a meromorphic function on* $\mathrm{Re}(s) > 0$ *that has a simple pole at* $s = 1$ *with residue* 1 *and no other poles.*

Proof. For $\mathrm{Re}(s) > 1$ we have

$$\zeta(s) - \frac{1}{s-1} = \sum_{n \ge 1} n^{-s} - \int_1^\infty x^{-s} dx = \sum_{n \ge 1} \left(n^{-s} - \int_n^{n+1} x^{-s} dx \right) = \sum_{n \ge 1} \int_n^{n+1} \left(n^{-s} - x^{-s} \right) dx.$$

For each $n \ge 1$ the function $\phi_n(s) := \int_n^{n+1} (n^{-s} - x^{-s}) dx$ is holomorphic on $\mathrm{Re}(s) > 0$. For each fixed s in $\mathrm{Re}(s) > 0$ and $x \in [n, n+1]$ we have

$$|n^{-s} - x^{-s}| = \left| \int_n^x s t^{-s-1} dt \right| \le \int_n^x \frac{|s|}{|t^{s+1}|} dt = \int_n^x \frac{|s|}{t^{1+\mathrm{Re}(s)}} dt \le \frac{|s|}{n^{1+\mathrm{Re}(s)}},$$

and therefore

$$|\phi_n(s)| \le \int_n^{n+1} \left| n^{-s} - x^{-s} \right| dx \le \frac{|s|}{n^{1+\mathrm{Re}(s)}}.$$

For any s_0 with $\mathrm{Re}(s_0) > 0$, if we put $\epsilon := \mathrm{Re}(s_0)/2$ and $U := B_{<\epsilon}(s_0)$, then for each $n \ge 1$,

$$\sup_{s \in U} |\phi_n(s)| \le \frac{|s_0| + \epsilon}{n^{1+\epsilon}} =: M_n,$$

and $\sum_n M_n = (|s_0| + \epsilon)\zeta(1+\epsilon)$ converges. The series $\sum_n \phi_n$ thus converges locally normally on $\mathrm{Re}(s) > 0$. By the Weierstrass M-test (Theorem 16.19), $\sum_n \phi_n$ converges to a function $\phi(s) = \zeta(s) - \frac{1}{s-1}$ that is holomorphic on $\mathrm{Re}(s) > 0$. □

We now show that $\zeta(s)$ has no zeros on $\mathrm{Re}(s) = 1$; this fact is crucial to the prime number theorem. For this we use the following ingenious lemma, attributed to Mertens.[2]

Lemma 16.4 (Mertens). *For* $x, y \in \mathbb{R}$ *with* $x > 1$ *we have* $|\zeta(x)^3 \zeta(x+iy)^4 \zeta(x+2iy)| \ge 1$.

[2] If this lemma strikes you as pulling a rabbit out of a hat, well, it is. For a slight variation, see [15, IV], which uses an alternative approach due to Hadamard.

Proof. From the Euler product $\zeta(s) = \prod_p (1-p^{-s})^{-1}$, we see that for $\operatorname{Re}(s) > 1$ we have

$$\log|\zeta(s)| = -\sum_p \log|1-p^{-s}| = -\sum_p \operatorname{Re}\log(1-p^{-s}) = \sum_p \sum_{n\geq 1} \frac{\operatorname{Re}(p^{-ns})}{n},$$

since $\log|z| = \operatorname{Re}\log z$ and $\log(1-z) = -\sum_{n\geq 1}\frac{z^n}{n}$ for $|z|<1$. Plugging in $s = x+iy$ yields

$$\log|\zeta(x+iy)| = \sum_p \sum_{n\geq 1} \frac{\cos(ny\log p)}{np^{nx}},$$

since $\operatorname{Re}(p^{-ns}) = p^{-nx}\operatorname{Re}(e^{-iny\log p}) = p^{-nx}\cos(-ny\log p) = p^{-nx}\cos(ny\log p)$. Thus

$$\log|\zeta(x)^3\zeta(x+iy)^4\zeta(x+2iy)| = \sum_p \sum_{n\geq 1} \frac{3+4\cos(ny\log p)+\cos(2ny\log p)}{np^{nx}}.$$

We now note that the trigonometric identity $\cos(2\theta) = 2\cos^2\theta - 1$ implies

$$3+4\cos\theta+\cos(2\theta) = 2(1+\cos\theta)^2 \geq 0.$$

Taking $\theta = ny\log p$ yields $\log|\zeta(x)^3\zeta(x+iy)^4\zeta(x+2iy)| \geq 0$, which proves the lemma. □

Corollary 16.5. *$\zeta(s)$ has no zeros on $\operatorname{Re}(s) \geq 1$.*

Proof. We know from Theorem 16.2 that $\zeta(s)$ has no zeros on $\operatorname{Re}(s) > 1$, so suppose $\zeta(1+iy) = 0$ for some $y \in \mathbb{R}$. Then $y \neq 0$, since $\zeta(s)$ has a pole at $s = 1$, and we know that $\zeta(s)$ does not have a pole at $1+2iy \neq 1$, by Theorem 16.3. We therefore must have

$$\lim_{x\to 1} |\zeta(x)^3\zeta(x+iy)^4\zeta(x+2iy)| = 0, \tag{1}$$

since $\zeta(s)$ has a simple pole at $s = 1$, a zero at $1+iy$, and no pole at $1+2iy$. But this contradicts Lemma 16.4. □

16.2 The Prime Number Theorem

The prime counting function $\pi\colon \mathbb{R} \to \mathbb{Z}_{\geq 0}$ is defined by

$$\pi(x) := \sum_{p\leq x} 1;$$

it counts the number of primes up to x. The prime number theorem (PNT) states that

$$\pi(x) \sim \frac{x}{\log x}.$$

The notation $f(x) \sim g(x)$ means $\lim_{x\to\infty} f(x)/g(x) = 1$; one says that f is *asymptotic to* g.

This conjectured growth rate for $\pi(x)$ dates back to Gauss and Legendre in the late 18th century. In fact Gauss believed the asymptotically equivalent but more accurate statement[3]

$$\pi(x) \sim \operatorname{Li}(x) := \int_2^x \frac{dt}{\log t}.$$

[3]More accurate in the sense that $|\pi(x) - \operatorname{Li}(x)|$ grows more slowly than $|\pi(x) - \frac{x}{\log x}|$ as $x \to \infty$.

However it was not until a century later that the prime number theorem was independently proved by Hadamard [5] and de la Vallée Poussin [9] in 1896. Their proofs are both based on the work of Riemann [10], who in 1860 showed that there is a precise connection between the zeros of $\zeta(s)$ and the distribution of primes (we shall say more about this later), but was unable to prove the prime number theorem.

The proof we will give is more recent and due to Newman [8], but it relies on the same properties of the Riemann zeta function that were exploited by both Hadamard and de la Vallée, the most essential of which is the fact that $\zeta(s)$ has no zeros on $\mathrm{Re}(s) \geq 1$ (Corollary 16.5). A concise version of Newman's proof by Zagier can be found in [15]; we will follow Zagier's outline but be slightly more expansive in our presentation. We should note that there are also "elementary" proofs of the prime number theorem independently obtained by Erdös [4] and Selberg [11] in the 1940s that do not use the Riemann zeta function, but they are elementary only in the sense that they do not use complex analysis; the details of these proofs are considerably more complicated than the one we will give.

Rather than work directly with $\pi(x)$, it is more convenient to work with the log-weighted prime-counting function defined by Chebyshev[4]

$$\vartheta(x) := \sum_{p \leq x} \log p,$$

whose growth rate differs from that of $\pi(x)$ by a logarithmic factor.

Theorem 16.6 (Chebyshev)**.** *$\pi(x) \sim \frac{x}{\log x}$ if and only if $\vartheta(x) \sim x$.*

Proof. We clearly have $0 \leq \vartheta(x) \leq \pi(x) \log x$, thus

$$\frac{\vartheta(x)}{x} \leq \frac{\pi(x) \log x}{x}.$$

For every $\epsilon \in (0,1)$ we have

$$\begin{aligned}\vartheta(x) \geq \sum_{x^{1-\epsilon} < p \leq x} \log p &\geq (1-\epsilon)(\log x)\big(\pi(x) - \pi(x^{1-\epsilon})\big) \\ &\geq (1-\epsilon)(\log x)(\pi(x) - x^{1-\epsilon}),\end{aligned}$$

and therefore

$$\pi(x) \leq \left(\frac{1}{1-\epsilon}\right)\frac{\vartheta(x)}{\log x} + x^{1-\epsilon}.$$

Thus for all $\epsilon \in (0,1)$ we have

$$\frac{\vartheta(x)}{x} \leq \frac{\pi(x)\log x}{x} \leq \left(\frac{1}{1-\epsilon}\right)\frac{\vartheta(x)}{x} + \frac{\log x}{x^{\epsilon}}.$$

The second term on the RHS tends to 0 as $x \to \infty$, and the lemma follows: by choosing ϵ sufficiently small we can make the ratios of $\vartheta(x)$ to x and $\pi(x)$ to $x/\log x$ arbitrarily close together as $x \to \infty$, so if one of them tends to 1, so must the other. □

[4] As with most Russian names, there is no canonical way to write Chebyshev in the latin alphabet and one finds many variations in the literature; in English, the spelling Chebyshev is now the most widely used.

In view of Chebyshev's result, the prime number theorem is equivalent to $\vartheta(x) \sim x$. We thus want to prove $\lim_{x\to\infty} \vartheta(x)/x = 1$; let us first show that $\lim_{x\to\infty} \vartheta(x)/x$ bounded, which is indicated by the asymptotic notation $\vartheta(x) = O(x)$.[5]

Lemma 16.7 (Chebyshev)**.** *For $x \geq 1$ we have $\vartheta(x) \leq (4\log 2)x$, thus $\vartheta(x) = O(x)$.*

Proof. For any integer $n \geq 1$, the binomial theorem implies

$$2^{2n} = (1+1)^{2n} = \sum_{m=0}^{2n} \binom{2n}{m} \geq \binom{2n}{n} = \frac{(2n)!}{n!n!} \geq \prod_{n<p\leq 2n} p = \exp(\vartheta(2n) - \vartheta(n)),$$

since $(2n)!$ is divisible by every prime $p \in (n, 2n]$ but $n!$ is not divisible by any such p. Taking logarithms on both sides yields

$$\vartheta(2n) - \vartheta(n) \leq 2n \log 2,$$

valid for all integers $n \geq 1$. For any integer $m \geq 1$ we have

$$\vartheta(2^m) = \sum_{n=1}^{m} \big(\vartheta(2^n) - \vartheta(2^{n-1})\big) \leq \sum_{n=1}^{m} 2^n \log 2 \leq 2^{m+1} \log 2.$$

For any real $x \geq 1$ we can choose an integer $m \geq 1$ so that $2^{m-1} \leq x < 2^m$, and then

$$\vartheta(x) \leq \vartheta(2^m) \leq 2^{m+1} \log 2 = (4\log 2)2^{m-1} \leq (4\log 2)x,$$

as claimed. □

In order to prove $\vartheta(x) \sim x$, we will use a general analytic criterion applicable to any non-decreasing real function $f(x)$.

Lemma 16.8. *Let $f\colon \mathbb{R}_{\geq 1} \to \mathbb{R}$ be a nondecreasing function. If the integral $\int_1^\infty \frac{f(t)-t}{t^2} dt$ converges then $f(x) \sim x$.*

Proof. Let $F(x) := \int_1^x \frac{f(t)-t}{t^2} dt$. The hypothesis is that $\lim_{x\to\infty} F(x)$ exists. This implies that for all $\lambda > 1$ and all $\epsilon > 0$ we have $|F(\lambda x) - F(x)| < \epsilon$ for all sufficiently large x.

Fix $\lambda > 1$ and suppose there is an unbounded sequence (x_n) such that $f(x_n) \geq \lambda x_n$ for all $n \geq 1$. For each x_n we have

$$F(\lambda x_n) - F(x_n) = \int_{x_n}^{\lambda x_n} \frac{f(t)-t}{t^2} dt \geq \int_{x_n}^{\lambda x_n} \frac{\lambda x_n - t}{t^2} dt = \int_1^\lambda \frac{\lambda - t}{t^2} dt = c,$$

for some $c > 0$, where we used the fact that f is non-decreasing to get the middle inequality. Taking $\epsilon < c$, we have $|F(\lambda x_n) - F(x_n)| = c > \epsilon$ for arbitrarily large x_n, a contradiction. Thus $f(x) < \lambda x$ for all sufficiently large x. A similar argument shows that $f(x) > \frac{1}{\lambda} x$ for all sufficiently large x. These inequalities hold for all $\lambda > 1$, so $\lim_{x\to\infty} f(x)/x = 1$. Equivalently, $f(x) \sim x$. □

[5]The equality sign in the big-O notation $f(x) = O(g(x))$ is a standard abuse of notation; it simply means $\limsup_{x\to\infty} |f(x)|/|g(x)| < \infty$ (and nothing more). In more complicated equalities a big-O expression should be interpreted as a set of functions, one of which makes the equality true, for example, $\sum_{n\geq 1} \frac{1}{n} = \log n + O(1)$.

In order to show that the hypothesis of Lemma 16.8 is satisfied for $f = \vartheta$, we will work with the function $H(t) = \vartheta(e^t)e^{-t} - 1$; the change of variables $t = e^u$ shows that

$$\int_1^\infty \frac{\vartheta(t) - t}{t^2} dt \text{ converges} \qquad \iff \qquad \int_0^\infty H(u)du \text{ converges .}$$

We now recall the Laplace transform.

Definition 16.9. Let $h\colon \mathbb{R}_{>0} \to \mathbb{R}$ be a piecewise continuous function. The *Laplace transform* $\mathcal{L}h$ of h is the complex function defined by

$$\mathcal{L}h(s) := \int_0^\infty e^{-st} h(t) dt,$$

which is holomorphic on $\mathrm{Re}(s) > c$ for any $c \in \mathbb{R}$ for which $h(t) = O(e^{ct})$.

The following properties of the Laplace transform are easily verified.

- $\mathcal{L}(g + h) = \mathcal{L}g + \mathcal{L}h$, and for any $a \in \mathbb{R}$ we have $\mathcal{L}(ah) = a\mathcal{L}h$.
- If $h(t) = a \in \mathbb{R}$ is constant then $\mathcal{L}h(s) = \frac{a}{s}$.
- $\mathcal{L}(e^{at}h(t))(s) = \mathcal{L}(h)(s - a)$ for all $a \in \mathbb{R}$.

We now define the auxiliary function

$$\Phi(s) := \sum_p p^{-s} \log p,$$

which is related to $\vartheta(x)$ by the following lemma.

Lemma 16.10. $\mathcal{L}(\vartheta(e^t))(s) = \frac{\Phi(s)}{s}$ *is holomorphic on* $\mathrm{Re}(s) > 1$.

Proof. By Lemma 16.7, $\vartheta(e^t) = O(e^t)$, so $\mathcal{L}(\vartheta(e^t))$ is holomorphic on $\mathrm{Re}(s) > 1$. Let p_n be the nth prime, and put $p_0 := 0$. The function $\vartheta(e^t)$ is constant on $t \in (\log p_n, \log p_{n+1})$, so

$$\int_{\log p_n}^{\log p_{n+1}} e^{-st}\vartheta(e^t)dt = \vartheta(p_n) \int_{\log p_n}^{\log p_{n+1}} e^{-st} dt = \frac{1}{s}\vartheta(p_n)\big(p_n^{-s} - p_{n+1}^{-s}\big).$$

We then have

$$\begin{aligned}
(\mathcal{L}\vartheta(e^t))(s) = \int_0^\infty e^{-st}\vartheta(e^t)dt &= \frac{1}{s}\sum_{n=1}^\infty \vartheta(p_n)\big(p_n^{-s} - p_{n+1}^{-s}\big) \\
&= \frac{1}{s}\sum_{n=1}^\infty \vartheta(p_n)p_n^{-s} - \frac{1}{s}\sum_{n=1}^\infty \vartheta(p_{n-1})p_n^{-s} \\
&= \frac{1}{s}\sum_{n=1}^\infty \big(\vartheta(p_n) - \vartheta(p_{n-1})\big)p_n^{-s} \\
&= \frac{1}{s}\sum_{n=1}^\infty p_n^{-s}\log p_n = \frac{\Phi(s)}{s}.
\end{aligned}$$

□

Let us now consider the function $H(t) := \vartheta(e^t)e^{-t} - 1$. It follows from the lemma and standard properties of the Laplace transform that on $\mathrm{Re}(s) > 0$ we have

$$\mathcal{L}H(s) = \mathcal{L}(\vartheta(e^t)e^{-t})(s) - (\mathcal{L}1)(s) = \mathcal{L}(\vartheta(e^t))(s+1) - \frac{1}{s} = \frac{\Phi(s+1)}{s+1} - \frac{1}{s}.$$

Lemma 16.11. *The function* $\Phi(s) - \frac{1}{s-1}$ *extends to a meromorphic function on* $\mathrm{Re}(s) > \frac{1}{2}$ *that is holomorphic on* $\mathrm{Re}(s) \geq 1$.

Proof. By Theorem 16.3, $\zeta(s)$ extends to a meromorphic function on $\mathrm{Re}(s) > 0$, which we also denote $\zeta(s)$, that has only a simple pole at $s = 1$ and no zeros on $\mathrm{Re}(s) \geq 1$, by Corollary 16.5. It follows that the logarithmic derivative $\zeta'(s)/\zeta(s)$ of $\zeta(s)$ is meromorphic on $\mathrm{Re}(s) > 0$, with no zeros on $\mathrm{Re}(s) \geq 1$ and only a simple pole at $s = 1$ with residue -1 (see §16.3.1 for standard facts about the logarithmic derivative of a meromorphic function). In terms of the Euler product, for $\mathrm{Re}(s) > 1$ we have[6]

$$\begin{aligned} -\frac{\zeta'(s)}{\zeta(s)} = (-\log\zeta(s))' &= \left(-\log\prod_p (1-p^{-s})^{-1}\right)' = \left(\sum_p \log(1-p^{-s})\right)' \\ &= \sum_p \frac{p^{-s}\log p}{1-p^{-s}} = \sum_p \frac{\log p}{p^s-1} = \sum_p \left(\frac{1}{p^s} + \frac{1}{p^s(p^s-1)}\right)\log p \\ &= \Phi(s) + \sum_p \frac{\log p}{p^s(p^s-1)}. \end{aligned}$$

The sum on the RHS converges absolutely and locally uniformly to a holomorphic function on $\mathrm{Re}(s) > 1/2$. The LHS is meromorphic on $\mathrm{Re}(s) > 0$, and on $\mathrm{Re}(s) \geq 1$ it has only a simple pole at $s = 1$ with residue 1. It follows that $\Phi(s) - \frac{1}{s-1}$ extends to a meromorphic function on $\mathrm{Re}(s) > \frac{1}{2}$ that is holomorphic on $\mathrm{Re}(s) \geq 1$. □

Corollary 16.12. *The functions* $\Phi(s+1) - \frac{1}{s}$ *and* $(\mathcal{L}H)(s) = \frac{\Phi(s+1)}{s+1} - \frac{1}{s}$ *both extend to meromorphic functions on* $\mathrm{Re}(s) > -\frac{1}{2}$ *that are holomorphic on* $\mathrm{Re}(s) \geq 0$.

Proof. The first statement follows immediately from the lemma. For the second, note that

$$\frac{\Phi(s+1)}{s+1} - \frac{1}{s} = \frac{1}{s+1}\left(\Phi(s+1) - \frac{1}{s}\right) - \frac{1}{s+1}$$

is meromorphic on $\mathrm{Re}(s) > -\frac{1}{2}$ and holomorphic on $\mathrm{Re}(s) \geq 0$, since it is a sum of products of such functions. □

The final step of the proof relies on the following analytic result due to Newman [8].

Theorem 16.13. *Let* $f\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ *be a bounded piecewise continuous function, and suppose its Laplace transform extends to a holomorphic function* $g(s)$ *on* $\mathrm{Re}(s) \geq 0$. *Then the integral* $\int_0^\infty f(t)dt$ *converges and is equal to* $g(0)$.

Proof. Without loss of generality we assume $f(t) \leq 1$ for all $t \geq 0$. For $\tau \in \mathbb{R}_{>0}$, define $g_\tau(s) \coloneqq \int_0^\tau f(t)e^{-st}dt$, By definition $\int_0^\infty f(t)dt = \lim_{\tau\to\infty} g_\tau(0)$, thus it suffices to prove

$$\lim_{\tau\to\infty} g_\tau(0) = g(0).$$

For $r > 0$, let γ_r be the boundary of the region $\{s : |s| \leq r \text{ and } \mathrm{Re}(s) \geq -\delta_r\}$ with $\delta_r > 0$ chosen so that g is holomorphic on γ_r; such a δ_r exists because g is holomorphic on $\mathrm{Re}(s) \geq 0$, hence on some open ball $B_{\leq 2\delta(y)}(iy)$ for each $y \in [-r, r]$, and we may take

[6] As is standard when computing logarithmic derivatives, we are taking the principal branch of the complex logarithm and can safely ignore the negative real axis where it is not defined since we are assuming $\mathrm{Re}(s) > 1$.

$\delta_r := \inf\{\delta(y) : y \in [r, -r]\}$, which is positive because $[-r, r]$ is compact. Each γ_r is a simple closed curve, and for each $\tau > 0$ the function $h(s) := (g(s) - g_\tau(s))e^{s\tau}(1 + \frac{s^2}{r^2})$ is holomorphic on a region containing γ_r. Using Cauchy's integral formula (Theorem 16.26) to evaluate $h(0)$ yields

$$g(0) - g_\tau(0) = h(0) = \frac{1}{2\pi i}\int_{\gamma_r} \big(g(s) - g_\tau(s)\big)e^{s\tau}\left(\frac{1}{s} + \frac{s}{r^2}\right) ds. \tag{2}$$

We will show the LHS tends to 0 as $\tau \to \infty$ by showing that for any $\epsilon > 0$ we can set $r = 3/\epsilon > 0$ so that the absolute value of the RHS is less than ϵ for all sufficiently large τ.

Let γ_r^+ denote the part of γ_r in $\mathrm{Re}(s) > 0$, a semicircle of radius r. The integrand is absolutely bounded by $1/r$ on γ_r^+, since for $|s| = r$ and $\mathrm{Re}(s) > 0$ we have

$$\begin{aligned}
\big|g(s) - g_\tau(s)\big| \cdot \left|e^{s\tau}\left(\frac{1}{s} + \frac{s}{r^2}\right)\right| &= \left|\int_\tau^\infty f(t)e^{-st}dt\right| \cdot \frac{e^{\mathrm{Re}(s)\tau}}{r} \cdot \left|\frac{r}{s} + \frac{s}{r}\right| \\
&\le \int_\tau^\infty e^{-\mathrm{Re}(s)t}dt \cdot \frac{e^{\mathrm{Re}(s)\tau}}{r} \cdot \frac{2\,\mathrm{Re}(s)}{r} \\
&= \frac{e^{-\mathrm{Re}(s)\tau}}{\mathrm{Re}(s)} \cdot \frac{e^{\mathrm{Re}(s)\tau}}{r} \cdot \frac{2\,\mathrm{Re}(s)}{r} \\
&= 2/r^2.
\end{aligned}$$

Therefore

$$\left|\frac{1}{2\pi i}\int_{\gamma_r^+} \big(g(s) - g_\tau(s)\big)e^{s\tau}\left(\frac{1}{s} + \frac{s}{r^2}\right) ds\right| \le \frac{1}{2\pi} \cdot \pi r \cdot \frac{2}{r^2} = \frac{1}{r} \tag{3}$$

Now let γ_r^- be the part of γ_r in $\mathrm{Re}(s) < 0$, a truncated semi-circle. For any fixed r, the first term $g(s)e^{s\tau}(s^{-1} + sr^{-2})$ in the integrand of (2) tends to 0 as $\tau \to \infty$ for $\mathrm{Re}(s) < 0$ and $|s| \le r$. For the second term we note that since $g_\tau(s)$ is holomorphic on $\mathbb{C}$, it makes no difference if we instead integrate over the semicircle of radius r in $\mathrm{Re}(s) < 0$. For $|s| = r$ and $\mathrm{Re}(s) < 0$ we then have

$$\begin{aligned}
\left|g_\tau(s)e^{s\tau}\left(\frac{1}{s} + \frac{s}{r^2}\right)\right| &= \left|\int_0^\tau f(t)e^{-st}dt\right| \cdot \frac{e^{\mathrm{Re}(s)\tau}}{r} \cdot \left|\frac{r}{s} + \frac{s}{r}\right| \\
&\le \int_0^\tau e^{-\mathrm{Re}(s)t}dt \cdot \frac{e^{\mathrm{Re}(s)\tau}}{r} \frac{(-2\,\mathrm{Re}(s))}{r} \\
&= \left(1 - \frac{e^{-\mathrm{Re}(s)\tau}}{\mathrm{Re}(s)}\right) \frac{e^{\mathrm{Re}(s)\tau}}{r} \frac{(-2\,\mathrm{Re}(s))}{r} \\
&= 2/r^2 \cdot (1 - e^{\mathrm{Re}(s)\tau}\,\mathrm{Re}(s)),
\end{aligned}$$

where the factor $(1 - e^{\mathrm{Re}(s)\tau}\,\mathrm{Re}(s))$ on the RHS tends to 1 as $\tau \to \infty$ since $\mathrm{Re}(s) < 0$. We thus obtain the bound $1/r + o(1)$ when we replace γ_r^+ with γ_r^- in (3), and the RHS of (2) is bounded by $2/r + o(1)$ as $\tau \to \infty$. It follows that for any $\epsilon > 0$, for $r = 3/\epsilon > 0$ we have

$$|g(0) - g_\tau(0)| < 3/r = \epsilon$$

for all sufficiently large τ. Therefore $\lim_{\tau\to\infty} g_\tau(0) = g(0)$ as desired. □

Remark 16.14. Theorem 16.13 is an example of what is known as a *Tauberian theorem*. For a piecewise continuous function $f\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$, its Laplace transform

$$\mathcal{L}f(s) := \int_0^\infty e^{-st} f(t)dt,$$

is typically not defined on $\mathrm{Re}(s) \leq c$, where c is the least c for which $f(t) = O(e^{ct})$. Now it may happen that the function $\mathcal{L}f$ has an analytic continuation to a larger domain; for example, if $f(t) = e^t$ then $(\mathcal{L}f)(s) = \frac{1}{s-1}$ extends to a holomorphic function on $\mathbb{C} - \{1\}$. But plugging values of s with $\mathrm{Re}(s) \leq c$ into the integral usually does not work; in our $f(t) = e^t$ example, the integral diverges on $\mathrm{Re}(s) \leq 1$. The theorem says that when $\mathcal{L}f$ extends to a holomorphic function on the entire half-plane $\mathrm{Re}(s) \geq 0$, its value at $s = 0$ is exactly what we would get by simply plugging 0 into the integral defining $\mathcal{L}f$.

More generally, Tauberian theorems refer to results related to transforms $f \to \mathcal{T}(f)$ that allow us to deduce properties of f (such as the convergence of $\int_0^\infty f(t)dt$) from properties of $\mathcal{T}(f)$ (such as analytic continuation to $\mathrm{Re}(s) \geq 0$). The term "Tauberian" was coined by Hardy and Littlewood and refers to Alfred Tauber, who proved a theorem of this type as a partial converse to a theorem of Abel.

Theorem 16.15 (Prime Number Theorem)**.** $\pi(x) \sim \frac{x}{\log x}$.

Proof. $H(t) = \vartheta(e^t)e^{-t} - 1$ is piecewise continuous and bounded, by Lemma 16.7, and its Laplace transform extends to a holomorphic function on $\mathrm{Re}(s) \geq 0$, by Corollary 16.12. Theorem 16.13 then implies that the integral

$$\int_0^\infty H(t)dt = \int_0^\infty \big(\vartheta(e^t)e^{-t} - 1\big)dt$$

converges. Replacing t with $\log x$, we see that

$$\int_1^\infty \left(\vartheta(x)\frac{1}{x} - 1\right)\frac{dx}{x} = \int_1^\infty \frac{\vartheta(x) - x}{x^2}dx$$

converges. Lemma 16.8 implies $\vartheta(x) \sim x$, equivalently, $\pi(x) \sim \frac{x}{\log x}$, by Theorem 16.6. □

One disadvantage of our proof is that it does not give us an error term. Using more sophisticated methods, Korobov [6] and Vinogradov [14] independently obtained the bound

$$\pi(x) = \mathrm{Li}(x) + O\left(\frac{x}{\exp\big((\log x)^{3/5+o(1)}\big)}\right),$$

in which we note that the error term is bounded by $O(x/(\log x)^n)$ for all n but not by $O(x^{1-\epsilon})$ for any $\epsilon > 0$. Assuming the Riemann Hypothesis, which states that the zeros of $\zeta(s)$ in the critical strip $0 < \mathrm{Re}(s) < 1$ all lie on the line $\mathrm{Re}(s) = \frac{1}{2}$, one can prove

$$\pi(x) = \mathrm{Li}(x) + O(x^{1/2+o(1)}).$$

More generally, if we knew that $\zeta(s)$ has no zeros in the critical strip with real part greater than c, for some $c \geq 1/2$ strictly less than 1, we could prove $\pi(x) = \mathrm{Li}(x) + O(x^{c+o(1)})$.

There thus remains a large gap between what we can prove about the distribution of prime numbers and what we believe to be true. Remarkably, other than refinements to the $o(1)$ term appearing in the Korobov-Vinogradov bound, essentially no progress has been made on this problem in the last 60 years.

16.3 A quick recap of some basic complex analysis

The complex numbers $\mathbb{C}$ are a topological field under the distance metric $d(x,y) = |x-y|$ induced by the standard absolute value $|z| := \sqrt{z\bar{z}}$, which is also a norm on $\mathbb{C}$ as an $\mathbb{R}$-vector space; all references to the topology on $\mathbb{C}$ (open, compact, convergence, limits, etc.) are made with this understanding.

16.3.1 Glossary of terms and standard theorems

Let f and g denote complex functions defined on an open subset of $\mathbb{C}$.

- f is *differentiable* at z_0 if $\lim_{z\to z_0} \frac{f(z)-f(z_0)}{z-z_0}$ exists.
- f is *holomorphic* at z_0 if it is differentiable on an open neighborhood of z_0.
- f is *analytic* at z_0 if there is an open neighborhood of z_0 in which f can be defined by a power series $f(z) = \sum_{n=0} a_n(z-z_0)^n$; equivalently, f is infinitely differentiable and has a convergent Taylor series on an open neighborhood of z_0.
- **Theorem**: f is holomorphic at z_0 if and only if it is analytic at z_0.
- **Theorem**: If C is a connected set containing a nonempty open set U and f and g are holomorphic on C with $f_{|U} = g_{|U}$, then $f_{|C} = g_{|C}$.
- With U and C as above, if f is holomorphic on U and g is holomorphic on C with $f_{|U} = g_{|U}$, then g is the (unique) *analytic continuation* of f to C and f *extends* to g.
- If f is holomorphic on a punctured open neighborhood of z_0 and $|f(z)| \to \infty$ as $z \to z_0$ then z_0 is a *pole* of f; note that the set of poles of f is necessarily a discrete set.
- f is *meromorphic* at z_0 if it is holomorphic at z_0 or has z_0 as a pole.
- **Theorem**: If f is meromorphic at z_0 then it can be defined by a Laurent series $f(z) = \sum_{n\geq n_0} a_n(z-z_0)^n$ that converges on an open punctured neighborhood of z_0.
- The *order of vanishing* $\operatorname{ord}_{z_0}(f)$ of a nonzero function f that is meromorphic at z_0 is the least index n of the nonzero coefficients a_n in its Laurent series expansion at z_0. Thus z_0 is a pole of f iff $\operatorname{ord}_{z_0}(f) < 0$ and z_0 is a zero of f iff $\operatorname{ord}_{z_0}(f) > 0$.
- If $\operatorname{ord}_{z_0}(f) = 1$ then z_0 is a *simple zero* of f, and if $\operatorname{ord}_{z_0}(f) = -1$ it is a *simple pole*.
- The *residue* $\operatorname{res}_{z_0}(f)$ of a function f meromorphic at z_0 is the coefficient a_{-1} in its Laurent series expansion $f(z) = \sum_{n\geq n_0} a_n(z-z_0)^n$ at z_0.
- **Theorem**: If z_0 is a simple pole of f then $\operatorname{res}_{z_0}(f) = \lim_{z\to z_0}(z-z_0)f(z)$.
- **Theorem**: If f is meromorphic on a set S then so is its *logarithmic derivative* f'/f, and f'/f has only simple poles in S and $\operatorname{res}_{z_0}(f'/f) = \operatorname{ord}_{z_0}(f)$ for all $z_0 \in S$. In particular the poles of f'/f are precisely the zeros and poles of f.

16.3.2 Convergence

Recall that a series $\sum_{n=1}^{\infty} a_n$ of complex numbers *converges absolutely* if the series $\sum_n |a_n|$ of nonnegative real numbers converges. An equivalent definition is that the function $a(n) := a_n$ is integrable with respect to the counting measure μ on the set of positive integers $\mathbb{N}$. Indeed, if the series is absolutely convergent then

$$\sum_{n=1}^{\infty} a_n = \int_{\mathbb{N}} a(n)\mu,$$

and if the series is not absolutely convergent, the integral is not defined. Absolute convergence is effectively built-in to the definition of the Lebesgue integral, which requires that in order for the function $a(n) = x(n) + iy(n)$ to be integrable, the positive real functions $|x(n)|$ and $|y(n)|$ must both be integrable (summable), and separately computes sums of the positive and negative subsequences of $(x(n))$ and $(y(n))$ as suprema over finite subsets.

The measure-theoretic perspective has some distinct advantages. It makes it immediately clear that we may replace the index set $\mathbb{N}$ with any set of the same cardinality, since the counting measure depends only on the cardinality of $\mathbb{N}$, not its ordering. We are thus free to sum over any countable index set, including $\mathbb{Z}$, $\mathbb{Q}$, any finite product of countable sets, and any countable coproduct of countable sets (such as countable direct sums of $\mathbb{Z}$); such sums are ubiquitous in number theory and many cannot be meaningfully interpreted as limits of partial sums in the usual sense, since this assumes that the index set is well ordered (not the case with $\mathbb{Q}$, for example). The measure-theoretic view makes also makes it clear that we may convert any absolutely convergent sum of the form $\sum_{X\times Y}$ into an iterated sum $\sum_X \sum_Y$ (or vice versa), via Fubini's theorem.

We say that an infinite product $\prod_n a_n$ of nonzero complex numbers is *absolutely convergent* when the sum $\sum_n \log a_n$ is, in which case $\prod_n a_n := \exp(\sum_n \log a_n)$.[7] This implies that an absolutely convergent product cannot converge to zero, and the sequence (a_n) must converge to 1 (no matter how we order the a_n). All of our remarks above about absolutely convergent series apply to absolutely convergent products as well.

A series or product of complex functions $f_n(z)$ is *absolutely convergent on S* if the series or product of complex numbers $f_n(z_0)$ is absolutely convergent for all $z_0 \in S$.

Definition 16.16. A sequence of complex functions (f_n) *converges uniformly on S* if there is a function f such that for every $\epsilon > 0$ there is an integer N for which $\sup_{z\in S} |f_n(z) - f(z)| < \epsilon$ for all $n \geq N$. The sequence (f_n) *converges locally uniformly on S* if every $z_0 \in S$ has an open neighborhood U for which (f_n) converges uniformly on $U \cap S$. When applied to a series of functions these terms refer to the sequence of partial sums.

Because $\mathbb{C}$ is locally compact, locally uniform convergence is the same thing as compact convergence: a sequence of functions converges locally uniformly on S if and only if it converges uniformly on every compact subset of S.

Theorem 16.17. *A sequence or series of holomorphic functions f_n that converges locally uniformly on an open set U converges to a holomorphic function f on U, and the sequence or series of derivatives f_n' then converges locally uniformly to f' (and if none of the f_n has a zero in U and $f \neq 0$, then f has no zeros in U).*

Proof. See [3, Thm. III.1.3] and [3, Thm. III.7.2]. □

Definition 16.18. A series of complex functions $\sum_n f_n(z)$ converges *normally* on a set S if $\sum_n \|f_n\| := \sum_n \sup_{z\in S} |f_n(z)|$ converges. The series $\sum_n f_n(z)$ converges *locally normally* on S if every $z_0 \in S$ has an open neighborhood U on which $\sum_n f_n(z)$ converges normally.

Theorem 16.19 (Weierstrass M-test)**.** *Every locally normally convergent series of functions converges absolutely and locally uniformly. Moreover, a series $\sum_n f_n$ of holomorphic functions on S that converges locally normally converges to a holomorphic function f on S, and then $\sum_n f_n'$ converges locally normally to f'.*

[7] In this definition we use the principal branch of $\log z := \log|z| + i \operatorname{Arg} z$ with $\operatorname{Arg} z \in (-\pi, \pi)$.

Proof. See [3, Thm. III.1.6]. □

Remark 16.20. To show a series $\sum_n f_n$ is locally normally convergent on a set S amounts to proving that for every $z_0 \in S$ there is an open neighborhood U of z_0 and a sequence of real numbers (M_n) such that $|f_n(z)| \le M_n$ for $z \in U \cap S$ and $\sum_n M_n < \infty$, whence the term "M-test".

16.3.3 Contour integration

We shall restrict our attention to integrals along contours defined by piecewise-smooth parameterized curves; this covers all the cases we shall need.

Definition 16.21. A *parameterized curve* is a continuous function $\gamma\colon [a,b] \to \mathbb{C}$ whose domain is a compact interval $[a,b] \subseteq \mathbb{R}$. We say that γ is *smooth* if it has a continuous nonzero derivative on $[a,b]$, and *piecewise-smooth* if $[a,b]$ can be partitioned into finitely many subintervals on which the restriction of γ is smooth. We say that γ is *closed* if $\gamma(a) = \gamma(b)$, and *simple* if it is injective on $[a,b)$ and $(a,b]$. Henceforth we will use the term *curve* to refer to any piecewise-smooth parameterized curve γ, or to its oriented image of in the complex plane (directed from $\gamma(a)$ to $\gamma(b)$), which we may also denote γ.

Definition 16.22. Let $f\colon \Omega \to \mathbb{C}$ be a continuous function and let γ be a curve in Ω. We define the *contour integral*

$$\int_\gamma f(z)dz := \int_a^b f(\gamma(t))\gamma'(t)dt,$$

whenever the integral on the RHS (which is defined as a Riemann sum in the usual way) converges. Whether $\int_\gamma f(z)dz$ converges, and if so, to what value, does not depend on the parameterization of γ: if γ' is another parameterized curve with the same (oriented) image as γ, then $\int_{\gamma'} f(z)dz = \int_\gamma f(z)dz$.

We have the following analog of the fundamental theorem of calculus.

Theorem 16.23. *Let $\gamma\colon [a,b] \to \mathbb{C}$ be a curve in an open set Ω and let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function Then*

$$\int_\gamma f'(z)dz = f(\gamma(b)) - f(\gamma(a)).$$

Proof. See [2, Prop. 4.12]. □

Recall that the Jordan curve theorem implies that every simple closed curve γ partitions $\mathbb{C}$ into two components, one of which we may unambiguously designate as the *interior* (the one on the left as we travel along our oriented curve). We say that γ is *contained* in an open set U if both γ and its interior lie in U. The interior of γ is a simply connected set, and if an open set U contains γ then it contains a simply connected open set that contains γ.

Theorem 16.24 (Cauchy's Theorem)**.** *Let U be an open set containing a simple closed curve γ. For any function f that is holomorphic on U we have*

$$\int_\gamma f(z)dz = 0.$$

Proof. See [2, Thm. 8.6] (we can restrict U to a simply connected set). □

Cauchy's theorem generalizes to meromorphic functions.

Theorem 16.25 (Cauchy Residue Formula)**.** *Let U be an open set containing a simple closed curve γ. Let f be a function that is meromorphic on U, let $z_1, \dots, z_n$ be the poles of f that lie in the interior of γ, and suppose that no pole of f lies on γ. Then*

$$\int_\gamma f(z)dz = 2\pi i \sum_{i=1}^{n} \operatorname{res}_{z_i}(f).$$

Proof. See [2, Thm. 10.5] (we can restrict U to a simply connected set). □

To see where the $2\pi i$ comes from, consider $\int_\gamma \frac{dz}{z}$ with $\gamma(t) = e^{it}$ for $t \in [0, 2\pi]$. In general one weights residues by a corresponding *winding number*, but the winding number of a simple closed curve about a point in its interior is always 1.

Theorem 16.26 (Cauchy's Integral Formula)**.** *Let U be an open set containing a simple closed curve γ. For any function f holomorphic on U and a in the interior of γ,*

$$f(a) = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{z-a} dz.$$

Proof. Apply Cauchy's residue formula to $g(z) = f(z)/(z-a)$; the only poles of g in the interior of γ are a simple pole at $z = a$ with $\operatorname{res}_a(g) = f(a)$. □

Cauchy's residue formula can also be used to recover the coefficients $f^{(n)}(a)/n!$ appearing in the Laurent series expansion of a meromorphic function at a (apply it to $f(z)/(z-a)^{n+1}$). One of many useful consequences of this is Liouville's theorem, which can be proved by showing that the Laurent series expansion of a bounded holomorphic function on $\mathbb{C}$ about any point has only one nonzero coefficient (the constant coefficient).

Theorem 16.27 (Liouville's theorem)**.** *Bounded entire functions are constant.*

Proof. See [2, Thm. 5.10]. □

We also have the following converse of Cauchy's theorem.

Theorem 16.28 (Morera's Theorem)**.** *Let f be a continuous function and on an open set U, and suppose that for every simple closed curve γ contained in U we have*

$$\int_\gamma f(z)dz = 0.$$

Then f is holomorphic on U.

Proof. See [3, Thm. II.3.5]. □

References

[1] Lars V. Ahlfors, *Complex analysis: an introduction to the theory of analytic functions of one complex variable*, 3rd edition, McGraw-Hill, 1979.

[2] Joseph Bak and Donald J. Newman, *Complex analysis*, Springer, 2010.

[3] Rolf Busam and Eberhard Freitag, *Complex analysis*, 2nd edition, Springer 2009.

[4] Paul Erdös, *On a new method in elementary number theory which leads to an elementary proof of the prime number theorem*, Proc. Nat. Acad. Scis. U.S.A. **35** (1949), 373–384.

[5] Jacques Hadamard, *Sur la distribution des zéros de la function $\zeta(s)$ et ses conséquences arithmétique*, Bull. Soc. Math. France **24** (1896), 199–220.

[6] Nikolai M. Korobov, *Estimates for trigonometric sums and their applications*, Uspechi Mat. Nauk **13** (1958), 185–192.

[7] Serge Lange, *Complex analysis*, 4th edition, Springer, 1985.

[8] David J. Newman, *Simple analytic proof of the Prime Number Theorem*, Amer. Math. Monthly **87** (1980), 693–696.

[9] Charles Jean de la Vallée Poussin, *Reserches analytiques sur la théorie des nombres premiers*, Ann. Soc. Sci. Bruxelles **20** (1896), 183–256.

[10] Bernhard Riemann, *Über die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsberichte der Berliner Akademie, 1859.

[11] Alte Selberg, *An elementary proof of the Prime-Number Theorem*, Ann. Math. **50** (1949), 305–313.

[12] Elias M. Stein and Rami Shakarchi, *Complex analysis*, Princeton University Press, 2003.

[13] Alfred Tauber, *Ein Satz aus der Theorie der unendlichen Reihen*, Monatsh f. Mathematik und Physik **8** (1897), 273–277.

[14] Ivan M. Vinogradov, *A new estimate of the function $\zeta(1+it)$*, Izv. Akad. Nauk SSSR. Ser. Mat. **22** (1958), 161–164.

[15] Don Zagier, *Newman's short proof of the Prime Number Theorem*, Amer. Math. Monthly **104** (1997), 705–708.

17 The functional equation

In the previous lecture we proved that the Riemann zeta function $\zeta(s)$ has an Euler product and an analytic continuation to the right half-plane $\mathrm{Re}(s) > 0$. In this lecture we complete the picture by deriving a *functional equation* that relates the values of $\zeta(s)$ to those of $\zeta(1-s)$. This will then also allow us to extend $\zeta(s)$ to a meromorphic function on $\mathbb{C}$ that is holomorphic except for a simple pole at $s = 1$.

17.1 Fourier transforms and Poisson summation

A key tool we will use to derive the functional equation is the *Poisson summation formula*, a result from harmonic analysis that we now recall.

Definition 17.1. A *Schwartz function* on $\mathbb{R}$ is a complex-valued C^∞ function $f\colon \mathbb{R} \to \mathbb{C}$ that decays rapidly to zero: for all $m, n \in \mathbb{Z}_{\geq 0}$ we have

$$\sup_{x\in\mathbb{R}} \left| x^m f^{(n)}(x) \right| < \infty,$$

where $f^{(n)}$ denotes the nth derivative of f. The *Schwartz space* $\mathcal{S}(\mathbb{R})$ of all Schwartz functions on $\mathbb{R}$ is a (non-unital) $\mathbb{C}$-algebra of infinite dimension.

Example 17.2. All compactly supported C^∞ functions are Schwartz functions, as is the *Gaussian* function $g(x) := e^{-\pi x^2}$. Non-examples include functions that do not tend to zero as $x \to \pm\infty$ (such as polynomials), and functions like $(1 + x^{2n})^{-1}$ and $e^{-x^2}\sin(e^{x^2})$ that either do not tend to zero quickly enough, or have derivatives that do not tend to zero as $x \to \pm\infty$.

Remark 17.3. For any $p \in \mathbb{R}_{\geq 1}$, the Schwartz space $\mathcal{S}(\mathbb{R})$ is contained in the space $L^p(\mathbb{R})$ of functions on $f\colon \mathbb{R} \to \mathbb{C}$ for which the Lebesgue integral $\int_{\mathbb{R}} |f(x)|^p dx$ exists. The space $L^p(\mathbb{R})$ is a complete normed $\mathbb{C}$-vector space under the L^p-norm $\|f\|_p := (\int_{\mathbb{R}} |f(x)|^p dx)^{1/p}$, and is thus a Banach space. The Schwartz space $\mathcal{S}(\mathbb{R})$ is not complete under the L^p-norm, but it is dense in $L^p(\mathbb{R})$ (in the subspace topology). One can equip the Schwartz space with a translation-invariant metric of its own under which it is a complete metric space (and thus a Fréchet space, since it is also locally convex), but the topology of $\mathcal{S}(\mathbb{R})$ will not concern us here. Similar comments apply to $\mathcal{S}(\mathbb{R}^n)$.

It follows immediately from the definition and standard properties of the derivative that the Schwartz space $\mathcal{S}(\mathbb{R})$ is closed under differentiation, multiplication by polynomials, and linear change of variable. It is also closed under *convolution*: for any $f, g \in \mathcal{S}(\mathbb{R})$ the function

$$(f * g)(x) := \int_{\mathbb{R}} f(y)g(x-y)dy$$

is also an element of $\mathcal{S}(\mathbb{R})$. Convolution is commutative, associative, and bilinear.

Definition 17.4. The *Fourier transform* of a Schwartz function $f \in \mathcal{S}(\mathbb{R})$ is the function

$$\hat{f}(y) := \int_{\mathbb{R}} f(x)e^{-2\pi i x y}dx,$$

which is also a Schwartz function [1, Thm. 5.1.3]. We can recover $f(x)$ from $\hat{f}(y)$ via the inverse transform

$$f(x) = \int_{\mathbb{R}} \hat{f}(y) e^{+2\pi i x y} dy;$$

see [1, Thm. 5.1.9] for a proof of this fact. The maps $f \mapsto \hat{f}$ and $\hat{f} \mapsto f$ are thus inverse linear operators on $\mathcal{S}(\mathbb{R})$ (they are also continuous in the metric topology of $\mathcal{S}(\mathbb{R})$ and thus homeomorphisms).

Remark 17.5. The invertibility of the Fourier transform on the Schwartz space $\mathcal{S}(\mathbb{R})$ is a key motivation for its definition. For functions in $L^1(\mathbb{R})$ (the largest space of functions for which our definition of the Fourier transform makes sense), the Fourier transform of a smooth function decays rapidly to zero, and the Fourier transform of a function that decays rapidly to zero is smooth; this leads one to consider the subspace $\mathcal{S}(\mathbb{R})$ of smooth functions that decay rapidly to zero. One can show that $\mathcal{S}(\mathbb{R})$ is the largest subspace of $L^1(\mathbb{R})$ closed under multiplication by polynomials on which the Fourier transform is invertible.[1]

The Fourier transform changes convolutions into products, and vice versa. We have

$$\widehat{f * g} = \hat{f}\hat{g} \qquad \text{and} \qquad \widehat{fg} = \hat{f} * \hat{g},$$

for all $f, g \in \mathcal{S}(\mathbb{R})$ (see Problem Set 8). One can thus view the Fourier transform as an isomorphism of (non-unital) $\mathbb{C}$-algebras that sends $(\mathcal{S}(\mathbb{R}), +, \times)$ to $(\mathcal{S}(\mathbb{R}), +, *)$.

Lemma 17.6. *For all $a \in \mathbb{R}_{>0}$ and $f \in \mathcal{S}(\mathbb{R})$, we have $\widehat{f(ax)}(y) = \frac{1}{a}\hat{f}(\frac{y}{a})$.*

Proof. Applying the substitution $t = ax$ yields

$$\widehat{f(ax)}(y) = \int_{\mathbb{R}} f(ax) e^{-2\pi i x y} dx = \frac{1}{a} \int_{\mathbb{R}} f(t) e^{-2\pi i t y/a} dt = \frac{1}{a} \hat{f}\left(\frac{y}{a}\right). \qquad \square$$

Lemma 17.7. *For $f \in \mathcal{S}(\mathbb{R})$ we have $\frac{d}{dy}\hat{f}(y) = -2\pi i \widehat{x f(x)}(y)$ and $\widehat{\frac{d}{dx} f(x)}(y) = 2\pi i y \hat{f}(y)$.*

Proof. Noting that $xf \in \mathcal{S}(\mathbb{R})$, the first identity follows from

$$\tfrac{d}{dy}\hat{f}(y) = \tfrac{d}{dy}\left(\int_{\mathbb{R}} f(x) e^{-2\pi i x y} dx\right) = \int_{\mathbb{R}} f(x)(-2\pi i x) e^{-2\pi i x y} dx = -2\pi i \widehat{x f(x)}(y),$$

since we may differentiate under the integral via dominated convergence. For the second, we note that $\lim_{x \to \pm\infty} f(x) = 0$, so integration by parts yields

$$\widehat{\tfrac{d}{dx} f(x)}(y) = \int_{\mathbb{R}} f'(x) e^{-2\pi i x y} dx = 0 - \int_{\mathbb{R}} f(x)(-2\pi i y) e^{-2\pi i x y} dx = 2\pi i y \hat{f}(y). \qquad \square$$

The Fourier transform is compatible with the inner product $\langle f, g \rangle := \int_{\mathbb{R}} f(x)\overline{g(x)} dx$ on $L^2(\mathbb{R})$ (which contains $\mathcal{S}(\mathbb{R})$). Indeed, we can easily derive Parseval's identity:

$$\langle f, g \rangle = \int_{\mathbb{R}} f(x)\overline{g(x)} dx = \int_{\mathbb{R}} \int_{\mathbb{R}} \hat{f}(y)\overline{g(x)} e^{+2\pi i x y} dx dy = \int_{\mathbb{R}} \hat{f}(y)\overline{\hat{g}(y)} dy = \langle \hat{f}, \hat{g} \rangle,$$

which when applied to $g = f$ yields Plancherel's identity:

$$\|f\|_2^2 = \langle f, f \rangle = \langle \hat{f}, \hat{f} \rangle = \|\hat{f}\|_2^2,$$

where $\|f\|_2 = (\int_{\mathbb{R}} |f(x)|^2 dx)^{1/2}$ is the L^2-norm. For number-theoretic applications there is an analogous result due to Poisson.

[1] I thank Keith Conrad and Terry Tao for clarifying this point.

Theorem 17.8 (Poisson Summation Formula)**.** *For all $f \in \mathcal{S}(\mathbb{R})$ we have the identity*

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \hat{f}(n).$$

Proof. We first note that both sums are well defined; the rapid decay property of Schwartz functions guarantees absolute convergence. Let $F(x) := \sum_{n\in\mathbb{Z}} f(x+n)$. Then F is a periodic C^∞-function, so it has a Fourier series expansion

$$F(x) = \sum_{n\in\mathbb{Z}} c_n e^{2\pi i n x},$$

with Fourier coefficients

$$c_n = \int_0^1 F(t)e^{-2\pi i n t}dt = \int_0^1 \sum_{m\in\mathbb{Z}} f(t+m)e^{-2\pi i n t}dt = \int_{\mathbb{R}} f(t)e^{-2\pi i n t}dt = \hat{f}(n).$$

We then note that

$$\sum_{n\in\mathbb{Z}} f(n) = F(0) = \sum_{n\in\mathbb{Z}} c_n = \sum_{n\in\mathbb{Z}} \hat{f}(n). \qquad \square$$

Finally, we note that the Gaussian function $e^{-\pi x^2}$ is its own Fourier transform.

Lemma 17.9. *Let $g(x) := e^{-\pi x^2}$. Then $\hat{g}(y) = g(y)$.*

Proof. The function $g(x)$ satisfies the first order ordinary differential equation

$$g' + 2\pi x g = 0, \tag{1}$$

with initial value $g(0) = 1$. Multiplying both sides by $-i$ and taking Fourier transforms yields

$$-i(\widehat{g'} + 2\pi\widehat{xg}) = -i(2\pi i x\hat{g} + i\hat{g}') = \hat{g}' + 2\pi x\hat{g} = 0,$$

via Lemma 17.7. So $\hat{g}$ also satisfies (1), and $\hat{g}(0) = \int_{\mathbb{R}} e^{-\pi x^2}dx = 1$, so $\hat{g} = g$. $\square$

17.1.1 Jacobi's theta function

We now define the *theta function*[2]

$$\Theta(\tau) := \sum_{n\in\mathbb{Z}} e^{\pi i n^2 \tau}.$$

The sum is absolutely convergent for $\operatorname{im}\tau > 0$ and thus defines a holomorphic function on the upper half plane. It is easy to see that $\Theta(\tau)$ is periodic modulo 2, that is,

$$\Theta(\tau+2) = \Theta(\tau),$$

but it it also satisfies another functional equation.

Lemma 17.10. *For all $a \in \mathbb{R}_{>0}$ we have $\Theta(ia) = \Theta(i/a)/\sqrt{a}$.*

Proof. Put $g(x) := e^{-\pi x^2}$ and $h(x) := g(\sqrt{a}x) = e^{-\pi x^2 a}$. Lemmas 17.6 and 17.9 imply

$$\hat{h}(y) = \widehat{g(\sqrt{a}x)}(y) = \hat{g}\big(y/\sqrt{a}\big)/\sqrt{a} = g\big(y/\sqrt{a}\big)/\sqrt{a}.$$

Plugging $\tau = ia$ into $\Theta(\tau)$ and applying Poisson summation (Theorem 17.8) yields

$$\Theta(ia) = \sum_{n\in\mathbb{Z}} e^{-\pi n^2 a} = \sum_{n\in\mathbb{Z}} h(n) = \sum_{n\in\mathbb{Z}} \hat{h}(n) = \sum_{n\in\mathbb{Z}} g\big(n/\sqrt{a}\big)/\sqrt{a} = \Theta(i/a)/\sqrt{a}. \qquad \square$$

[2]The function $\Theta(\tau)$ we define here is a special case of one of four parameterized families of theta functions $\Theta_i(z:\tau)$ originally defined by Jacobi for $i = 0, 1, 2, 3$, which play an important role in the theory of elliptic functions and modular forms; in terms of Jacobi's notation, $\Theta(\tau) = \Theta_3(0;\tau)$.

17.1.2 Euler's gamma function

You are probably familiar with the gamma function $\Gamma(s)$, which plays a key role in the functional equation of not only the Riemann zeta function but many of the more general zeta functions and L-series we wish to consider. Here we recall some of its analytic properties. We begin with the definition of $\Gamma(s)$ as a Mellin transform.

Definition 17.11. The *Mellin transform* of a function $f\colon \mathbb{R}_{>0} \to \mathbb{C}$ is the complex function defined by

$$\mathcal{M}(f)(s) := \int_0^\infty f(t)t^{s-1}dt,$$

whenever this integral converges. It is holomorphic on $\operatorname{Re} s \in (a,b)$ for any interval (a,b) in which the integral $\int_0^\infty |f(t)|t^{\sigma-1}dt$ converges for all $\sigma \in (a,b)$.

Definition 17.12. The *Gamma function*

$$\Gamma(s) := \mathcal{M}(e^{-t})(s) = \int_0^\infty e^{-t}t^{s-1}dt,$$

is the Mellin transform of e^{-t}. Since $\int_0^\infty |e^{-t}|t^{\sigma-1}dt$ converges for all $\sigma > 0$, the integral defines a holomorphic function on $\operatorname{Re}(s) > 0$.

Integration by parts yields

$$\Gamma(s) = \frac{t^s e^{-t}}{s}\Big|_0^\infty + \frac{1}{s}\int_0^\infty e^{-t}t^s dt = \frac{\Gamma(s+1)}{s},$$

thus $\Gamma(s)$ has a simple pole at $s = 0$ with residue 1 (since $\Gamma(1) = \int_0^\infty e^{-t}dt = 1$), and

$$\Gamma(s+1) = s\Gamma(s) \tag{2}$$

for $\operatorname{Re}(s) > 0$. Equation (2) allows us to extend $\Gamma(s)$ to a meromorphic function on $\mathbb{C}$ with simple poles at $s = 0, -1, -2, \ldots$, and no other poles.

An immediate consequence of (2) is that for integers $n > 0$ we have

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\Gamma(n-1) = n(n-1)(n-2)\cdots 2\cdot 1\cdot \Gamma(1) = n!,$$

thus the gamma function can be viewed as an extension of the factorial function. The gamma function satisfies many useful identities in addition to (2), including the following.

Theorem 17.13 (Euler's Reflection Formula)**.** *We have*

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin(\pi s)}.$$

as meromorphic functions on $\mathbb{C}$ with simple poles at each integer $s \in \mathbb{Z}$.

Proof. Let $f(s) := \Gamma(s)\Gamma(1-s)\sin(\pi s)$. The function $\Gamma(s)\Gamma(1-s)$ has a simple pole at each $s \in \mathbb{Z}$ and no other poles, while $\sin(\pi s)$ has a zero at each $s \in \mathbb{Z}$ and no poles, so $f(s)$ is holomorphic on $\mathbb{C}$. We now note that

$$f(s+1) = \Gamma(s+1)\Gamma(-s)\sin(\pi s+\pi) = -s\Gamma(s)\Gamma(-s)\sin(\pi s) = \Gamma(s)\Gamma(1-s)\sin(\pi s) = f(s),$$

so f is periodic (with period 1). Using the substitution $u = e^t$ we obtain

$$|\Gamma(s)| \le \int_0^\infty |e^{-t}t^{s-1}|dt = \int_{-\infty}^\infty |e^{-e^u}e^{u(s-1)}|e^u du = \int_{-\infty}^\infty e^{u\,\mathrm{Re}(s)-e^u}du.$$

This implies $|\Gamma(s)|$ is bounded on $\mathrm{Re}(s) \in [1,2]$, hence on $\mathrm{Re}(s) \in [0,1] \cap \mathrm{Im}(s) \ge 1$, via (2). It follows that in the strip $\mathrm{Re}(s) \in [0,1]$ we have

$$|f(s)| = |\Gamma(s)||\Gamma(1-s)||\sin(\pi s)| = O(e^{\mathrm{Im}(s)}),$$

as $\mathrm{Im}(s) \to \infty$, since $|\sin(\pi s)| = \frac{1}{2}|e^{is} - e^{i\bar{s}}| = O(e^{\mathrm{Im}(s)})$. By Lemma 17.14 below, $f(s)$ is constant. To determine the constant, as $s \to 0$ we have $\Gamma(s) \sim \frac{1}{s}$ and $\sin(\pi s) \sim \pi s$, thus

$$f(0) = \lim_{s\to 0} f(s) = \lim_{s\to 0} \Gamma(s)\Gamma(1-s)\sin(\pi s) = \lim_{s\to 0} \frac{1}{s}\cdot 1 \cdot \pi s = \pi,$$

and the theorem follows. □

Lemma 17.14. *Let $f(s)$ be a holomorphic function on $\mathbb{C}$ such that $f(s+1) = f(s)$ and $|f(s)| = O(e^{\mathrm{Im}(s)})$ as $\mathrm{Im}(s) \to \infty$ in the vertical strip $\mathrm{Re}(s) \in [0,1]$. Then f is constant.*

Proof. The function

$$g(s) = \frac{f(s) - f(a)}{\sin(\pi(s-a))}$$

is holomorphic on $\mathbb{C}$, since $f(s) - f(a)$ is holomorphic and vanishes at the zeros $a + \mathbb{Z}$ of $\sin(\pi(s-a))$ (all of which are simple). We also have $g(s+1) = g(s)$, and $|g(s)|$ is bounded on $\mathrm{Re}(s) \in [0,1]$, since as $\mathrm{Im}(s) \to \infty$ we have $|f(s) - f(a)| = O(e^{\mathrm{Im}(s)})$ and $|\sin(\pi(s-a))| \sim e^{\pi\,\mathrm{Im}(s)}$. It follows that $g(s)$ is bounded on $\mathbb{C}$, hence constant, by Liouville's theorem. We must have $g = 0$, since $|g(s)| = O(e^{(1-\pi)\,\mathrm{Im}(s)}) = o(1)$ as $\mathrm{Im}(s) \to \infty$, and this implies $f(s) = f(a)$ for all $s \in \mathbb{C}$. □

Example 17.15. Putting $s = \frac{1}{2}$ in the reflection formula yields $\Gamma(\frac{1}{2})^2 = \pi$, so $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Corollary 17.16. *The function $\Gamma(s)$ has no zeros on $\mathbb{C}$.*

Proof. Suppose $\Gamma(s_0) = 0$. The RHS of the reflection formula $\Gamma(s)\Gamma(1-s) = \pi/\sin(\pi s)$ is never zero, since $\sin(\pi s)$ has no poles, so $\Gamma(1-s)$ must have a pole at s_0. Therefore $1 - s_0 \in \mathbb{Z}$, equivalently, $s_0 \in \mathbb{Z}$, but for $s_0 \in \mathbb{Z}_{>0}$ we have $\Gamma(s_0) = (s_0 - 1)! \ne 0$, and for $s_0 \in \mathbb{Z}_{\le 0}$ we cannot have $\Gamma(s_0) = 0$ because $\Gamma(s)$ has a pole at all non-positive integers. □

17.1.3 Completing the zeta function

Let us now consider the function

$$F(s) := \pi^{-s}\Gamma(s)\zeta(2s),$$

which is holomorphic on $\mathrm{Re}(s) > 1/2$. In the region $\mathrm{Re}(s) > 1/2$ we have an absolutely convergent sum

$$F(s) = \pi^{-s}\Gamma(s)\sum_{n\ge 1} n^{-2s} = \sum_{n\ge 1}(\pi n^2)^{-s}\Gamma(s) = \sum_{n\ge 1}\int_0^\infty (\pi n^2)^{-s}t^{s-1}e^{-t}dt,$$

and the substitution $t = \pi n^2 y$ with $dt = \pi n^2 dy$ yields

$$F(s) = \sum_{n\geq 1} \int_0^\infty (\pi n^2)^{-s} (\pi n^2 y)^{s-1} e^{-\pi n^2 y} \pi n^2 dy = \sum_{n\geq 1} \int_0^\infty y^{s-1} e^{-\pi n^2 y} dy.$$

By the Fubini-Tonelli theorem, we can swap the sum and the integral to obtain

$$F(s) = \int_0^\infty y^{s-1} \sum_{n\geq 1} e^{-\pi n^2 y} dy.$$

We have $\Theta(iy) = \sum_{n\in\mathbb{Z}} e^{-\pi n^2 y} = 1 + 2\sum_{n\geq 1} e^{-\pi n^2 y}$, thus

$$\begin{aligned} F(s) &= \frac{1}{2} \int_0^\infty y^{s-1} (\Theta(iy) - 1) dy \\ &= \frac{1}{2} \left(\int_0^1 y^{s-1} \Theta(iy) dy - \frac{1}{s} + \int_1^\infty y^{s-1} (\Theta(iy) - 1) dy \right) \end{aligned}$$

We now focus on the first integral on the RHS. The change of variable $t = \frac{1}{y}$ yields

$$\int_0^1 y^{s-1} \Theta(iy) dy = \int_\infty^1 t^{1-s} \Theta(i/t)(-t^{-2}) dt = \int_1^\infty t^{-s-1} \Theta(i/t) dt.$$

By Lemma 17.10, $\Theta(i/t) = \sqrt{t}\Theta(it)$, and adding $-\int_1^\infty t^{-s-1/2} dt + \int_1^\infty t^{-s-1/2} dt = 0$ yields

$$\begin{aligned} &= \int_1^\infty t^{-s-1/2} \big(\Theta(it) - 1\big) dt + \int_1^\infty t^{-s-1/2} dt \\ &= \int_1^\infty t^{-s-1/2} \big(\Theta(it) - 1\big) dt - \frac{1}{1/2 - s}. \end{aligned}$$

Plugging this back into our equation for $F(s)$ we obtain the identity

$$F(s) = \frac{1}{2} \int_1^\infty \big(y^{s-1} + y^{-s-1/2}\big)\big(\Theta(iy) - 1\big) dy - \frac{1}{2s} - \frac{1}{1-2s},$$

valid on $\mathrm{Re}(s) > 1/2$. We now observe that $F(s) = F(\frac{1}{2} - s)$ for $s \neq 0, \frac{1}{2}$, which allows us to analytically extend $F(s)$ to a meromorphic function on $\mathbb{C}$ with poles only at $s = 0, 1/2$. Replacing s with $s/2$ leads us to define the *completed zeta function*

$$Z(s) := \pi^{-s/2} \Gamma(\tfrac{s}{2}) \zeta(s), \tag{3}$$

which is meromorphic on $\mathbb{C}$ and satisfies the *functional equation*

$$Z(s) = Z(1 - s). \tag{4}$$

It has simple poles at 0 and 1 (and no other poles). The only zeros of $Z(s)$ on $\mathrm{Re}(s) > 0$ are the zeros of $\zeta(s)$, since by Corollary 17.16, the gamma function $\Gamma(s)$ has no zeros (and neither does $\pi^{-s/2}$). Thus the zeros of $Z(s)$ on $\mathbb{C}$ all lie in the critical strip $0 < \mathrm{Re}(s) < 1$.

The functional equation also allows us to analytically extend $\zeta(s)$ to a meromorphic function on $\mathbb{C}$ whose only pole is a simple pole at $s = 1$; the pole of $Z(s)$ at $s = 0$ comes from the pole of $\Gamma(s/2)$ at $s = 0$. The function $\Gamma(s/2)$ also has poles at $-2, -4, \ldots$ where $Z(s)$ does not, so our extended $\zeta(s)$ must have zeros at $-2, -4, \ldots$. These are *trivial zeros*;

all the interesting zeros of $\zeta(s)$ lie in the critical strip and are conjectured to lie only on the critical line $\mathrm{Re}(s) = 1/2$ (this is the Riemann hypothesis).

We can compute $\zeta(0)$ using the functional equation. From (3) and (4) we have

$$\zeta(s) = \frac{Z(s)}{\pi^{-s/2}\Gamma(\frac{s}{2})} = \frac{Z(1-s)}{\pi^{-s/2}\Gamma(\frac{s}{2})} = \frac{\pi^{\frac{(s-1)}{2}}\Gamma(\frac{1-s}{2})}{\pi^{-s/2}\Gamma(\frac{s}{2})}\zeta(1-s) = \frac{\pi^{s-1/2}\Gamma(\frac{1-s}{2})}{\Gamma(\frac{s}{2})}\zeta(1-s). \tag{5}$$

We know that $\zeta(s)$ has a simple pole with residue 1 at $s = 1$, so

$$1 = \lim_{s\to 1^+}(s-1)\zeta(s) = \lim_{s\to 1^+}\frac{(s-1)\pi^{s-1/2}\Gamma(\frac{1-s}{2})}{\Gamma(\frac{s}{2})}\zeta(1-s).$$

When $s = 1$, the denominator on the RHS is $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, which cancels $\pi^{1-1/2} = \sqrt{\pi}$ in the numerator. Using $\Gamma(z) = \frac{1}{z}\Gamma(z+1)$ to shift $\Gamma(\frac{1-s}{2})$ in the numerator yields

$$1 = \lim_{s\to 1^+}(s-1)\tfrac{2}{1-s}\Gamma\left(\tfrac{3-s}{2}\right)\zeta(1-s) = -2\Gamma(1)\zeta(0) = -2\zeta(0).$$

Thus $\zeta(0) = -1/2$.

Using the reflection formula to replace $\Gamma(\frac{s}{2}) = \pi/(\Gamma(\frac{2-s}{2})\sin(\frac{\pi s}{2}))$ in (5), we have

$$\zeta(s) = \pi^{s-3/2}\Gamma(\tfrac{1-s}{2})\Gamma(\tfrac{2-s}{2})\sin(\tfrac{\pi s}{2})\zeta(1-s).$$

Applying the duplication formula $\Gamma(2z) = \pi^{-1/2}2^{2z-1}\Gamma(z)\Gamma(z+\frac{1}{2})$ with $z = \frac{1-s}{2}$ then yields

$$\zeta(s) = 2^s\pi^{s-1}\sin(\tfrac{\pi s}{2})\Gamma(1-s)\zeta(1-s), \tag{6}$$

which is how one often sees the functional equation for $\zeta(s)$ written.

17.2 Gamma factors and a holomorphic zeta function

If we write out the Euler product for the completed zeta function, we have

$$Z(s) = \pi^{-s/2}\Gamma(\tfrac{s}{2}) \cdot \prod_p (1-p^{-s})^{-1}.$$

One should think of this as a product over the places of the field $\mathbb{Q}$; the leading factor

$$\Gamma_{\mathbb{R}}(s) := \pi^{-s/2}\Gamma(\tfrac{s}{2})$$

that distinguishes the completed zeta function $Z(s)$ from $\zeta(s)$ corresponds to the real archimedean place of $\mathbb{Q}$. When we discuss Dedekind zeta functions in a later lecture we will see that there are gamma factors $\Gamma_{\mathbb{R}}$ and $\Gamma_{\mathbb{C}}$ associated to each of the real and complex places of a number field.

If we insert an additional factor of $\binom{s}{2} := \frac{s(s-1)}{2}$ in $Z(s)$ we can remove the poles at 0 and 1, yielding a function $\xi(s)$ holomorphic on $\mathbb{C}$. This yields Riemann's seminal result.

Theorem 17.17 (Analytic Continuation II)**.** *The function*

$$\xi(s) := \tbinom{s}{2}\Gamma_{\mathbb{R}}(s)\zeta(s)$$

is holomorphic on $\mathbb{C}$ and satisfies the functional equation

$$\xi(s) = \xi(1-s).$$

The zeros of $\xi(s)$ all lie in the critical strip $0 < \mathrm{Re}(s) < 1$.

Remark 17.18. We will usually work with $Z(s)$ and deal with the poles rather than making it holomorphic by introducing additional factors; some authors use $\xi(s)$ to denote our $Z(s)$.

17.3 Zeros in the critical strip

The zeros of $\xi(s)$ in the critical strip are precisely the zeros of $\zeta(s)$ in the critical strip. Let $N(T)$ denote the number of zeros of $\xi(s)$ in the critical strip that satisfy $0 < \operatorname{im} s < T$. If we fix $\epsilon > 0$ and let R be the rectangle $\{-\epsilon \le \operatorname{Re}(s) \le 1+\epsilon,\ 0 \le \operatorname{im} s \le T\}$, we can compute $N(T)$ using Cauchy's argument principle via

$$N(T) = \frac{1}{2\pi i}\int_{\partial R} \frac{\xi'(s)}{\xi(s)} ds,$$

provided that there are no zeros on the lines $\operatorname{im} s = 0$ and $\operatorname{im} s = T$. From this formula and the functional equation one derive the asymptotic formula

$$N(T) \sim \frac{1}{2\pi} T \log\left(\frac{T}{2\pi e}\right),$$

along with an explicit error term that allows one to compute the integer $N(T)$ exactly. Note that this formula implies that there are infinitely many zeros in the critical strip. The Riemann hypothesis states that all of these zeros lie on the critical line $\operatorname{im} s = 1/2$.

One can count zeros on the critical line by counting zeros of the *Hardy Z-function*

$$e^{i\theta(t)}\zeta(1/2 + it)$$

in a region $0 \le t \le T$, where $\theta(t)$ is the *Riemann-Siegel function*

$$\theta(t) := \arg\left(\Gamma\left(\frac{2it+1}{4}\right)\right) - \frac{\log \pi}{2} t.$$

There are asymptotic expansions of the Hardy Z-function that allow one to do this efficiently (one counts sign changes and checks for multiple roots). By comparing the result to $N(T)$ one can determine whether all the zeros in the critical strip with $0 < \operatorname{im} s < T$ lie on the critical line or not. This has been done for values of T exceeding 10^{13}; more precisely, it has been verified that when ordered by their imaginary parts, the first 10^{13} zeros above the real axis all lie on the critical line; see [2] for details.

References

[1] Elias M. Stein and Rami Shakarchi, *Fourier analysis: an introduction*, Princeton University Press, 2007.

[2] Xavier Gourdon, *The 10^{13} first zeros of the Riemann zeta function and zeros computation at very large height*, preprint, 2004.

18 Dirichlet L-functions, primes in arithmetic progressions

Having proved the prime number theorem, we would like to prove an analogous result for primes in arithmetic progressions. We begin with Dirichlet's theorem on primes in arithmetic progressions, a result that predates the prime number theorem by sixty years.

Theorem 18.1 (Dirichlet 1837)**.** *For all coprime integers a and m there are infinitely many primes $p \equiv a \bmod m$.*

In fact Dirichlet proved more than this. In a sense that we will make precise below, he proved that for every fixed modulus m the primes are equidistributed among the residue classes in $(\mathbb{Z}/m\mathbb{Z})^\times$. The equidistribution statement that Dirichlet was able to prove is a bit weaker than one might like, but it is more than enough to establish Theorem 18.1.

Remark 18.2. Many of the standard tools of complex analysis we take for granted were not available to Dirichlet in 1837. Riemann was the first to seriously study $\zeta(s)$ as a function of a complex variable, some twenty years after Dirichlet proved Theorem 18.1. We will work in a more modern setting, but our approach still follows the spirit of Dirichlet's proof.

18.1 Infinitely many primes

To motivate Dirichlet's method of proof, let us consider the following (admittedly clumsy) proof that there are infinitely many primes. It is sufficient to show that the Euler product

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1}$$

diverges as $s \to 1^+$. Of course we know $\zeta(s)$ has a pole at $s = 1$ (by Theorem 16.3), but let us suppose for the moment that we did not already know this. Taking logarithms yields

$$\log \zeta(s) = -\sum_p \log(1 - p^{-s}) = \sum_p p^{-s} + O(1), \tag{1}$$

as $s \to 1^+$, where we have used the asymptotic bounds

$$-\log(1-x) = x + O(x^2) \quad (\text{as } x \to 0) \qquad \text{and} \qquad \sum_p O(p^{-2s}) = O(1) \quad (\mathrm{Re}(s) > 1/2 + \epsilon).$$

We can estimate $\sum_{p \le x} \frac{1}{p}$ via Mertens' second theorem, one of three he proved in [4].

Theorem 18.3 (Mertens 1874)**.** *As $x \to \infty$ we have*

(1) $\displaystyle\sum_{p \le x} \frac{\log p}{p} = \log x + R(x)$, *where* $|R(x)| < 2$.[1]

(2) $\displaystyle\sum_{p \le x} \frac{1}{p} = \log\log x + B + O\big(\tfrac{1}{\log x}\big)$, *where* $B = 0.261497\ldots$ *is Mertens' constant;*

(3) $\displaystyle\sum_{p \le x} \log\big(1 - \tfrac{1}{p}\big) = -\log\log x - \gamma + O\big(\tfrac{1}{\log x}\big)$, *where* $\gamma = 0.577216\ldots$ *is Euler's constant.*

[1] In fact, $R(x) = -B_3 + o(1)$ where $B_3 = 1.332582\ldots$ is an explicit constant.

Proof. See Problem Set 9. □

Thus not only does $\sum p^{-s}$ diverge as $s \to 1^+$, we can say with a fair degree of precision how quickly this happens. We should note, however, that Mertens' estimate is not as strong as the prime number theorem. Indeed, as you will prove on Problem Set 9, the Prime Number Theorem is equivalent to the statement

$$\sum_{p \le x} \frac{1}{p} = \log\log x + B + o\left(\tfrac{1}{\log x}\right),$$

which is (ever so slightly) sharper than Mertens' estimate.[2]

18.1.1 Infinitely many primes congruent to 1 modulo 4

To demonstrate how the argument above generalizes to primes in arithmetic progressions, let us prove there are infinitely many primes congruent to 1 mod 4. We might initially consider

$$\prod_{p \equiv 1 \bmod 4} (1 - p^{-s})^{-1} = \sum_{\substack{n \ge 1 \\ p \mid n \Rightarrow p \equiv 1 \bmod 4}} n^{-s},$$

but the sum on the RHS is a bit awkward. Let us instead define a *Dirichlet character*

$$\chi(n) := \begin{cases} 1 & \text{if } n \equiv 1 \bmod 4, \\ -1 & \text{if } n \equiv -1 \bmod 4, \\ 0 & \text{otherwise,} \end{cases}$$

and consider the *Dirichlet L-function*

$$L(s,\chi) := \prod_p (1 - \chi(p)p^{-s})^{-1} = \sum_{n \ge 1} \chi(n) n^{-s} = 1 - 3^{-s} + 5^{-s} - 7^{-s} + 9^{-s} + \cdots,$$

which converges absolutely on $\mathrm{Re}(s) > 1$. As $s \to 1^+$ we have

$$\begin{aligned} \log L(s,\chi) &= -\sum_p \log(1 - \chi(p)p^{-s}) = \sum_p \chi(p) p^{-s} + O(1) \\ &= \sum_{p \equiv 1 \bmod 4} p^{-s} - \sum_{p \equiv 3 \bmod 4} p^{-s} + O(1), \end{aligned}$$

and

$$\log \zeta(s) = \sum_{p \equiv 1 \bmod 4} p^{-s} + \sum_{p \equiv 3 \bmod 4} p^{-s} + O(1),$$

thus

$$\frac{\log \zeta(s) + \log L(s,\chi)}{2} = \sum_{p \equiv 1 \bmod 4} p^{-s} + O(1).$$

Provided $\log L(s,\chi) = O(1)$ as $s \to 1^+$, the LHS (and hence the RHS) must tend to infinity as $s \to 1^+$, since $\zeta(s) \to \infty$ as $s \to 1^+$. It thus suffices to show that $L(s,\chi)$ has an analytic

[2] The error term in the PNT actually implies $\sum_{p \le x} \frac{1}{p} = \log\log x + B + O\left(\frac{1}{x}\right)$, but an $o(\frac{1}{\log x})$ bound is already enough to show $\pi(x) \sim x/\log x$. That the difference between a little-o and a big-O is the difference between proving the PNT and not proving it demonstrates how critical it is to understand error terms.

continuation to a neighborhood of $s = 1$ with $L(1, \chi) \neq 0$ (in which case there is a branch of the complex logarithm holomorphic on a neighborhood of $L(1, \chi)$). We will prove this in the next lecture. Assuming this for the moment, we then have

$$\sum_{\substack{p \le x \\ p \equiv 1 \bmod 4}} \frac{1}{p} = \frac{1}{2} \log\log x + O(1).$$

Mertens' second theorem implies that the same holds if we instead sum over $p \equiv 3 \mod 4$. The primes are thus equidistributed modulo $m = 4$ in the sense that for all integers a coprime to m we have

$$\sum_{\substack{p \le x \\ p \equiv a \bmod m}} \frac{1}{p} \sim \frac{1}{\phi(m)} \sum_{p \le x} \frac{1}{p} \sim \frac{1}{\phi(m)} \log\log x.$$

We should note that this statement is weaker than the prime number theorem for arithmetic progressions, which states that

$$\pi(x; m, a) \sim \frac{1}{\phi(m)} \pi(x),$$

where $\pi(x; m, a)$ counts the primes $p \le x$ for which $p \equiv a \bmod m$ (see Problem Set 9).

Dirichlet did not have Mertens' asymptotic bounds so he stated his results in a different way, using what is now called the *Dirichlet density* of a set of primes S,

$$d(S) := \lim_{s \to 1^+} \frac{\sum_{p \in S} p^{-s}}{\sum_p p^{-s}},$$

defined whenever this limit exists (one can also define notions of lower and upper Dirichlet density using lim inf and lim sup that are always defined and agree whenever $d(S)$ is defined). This definition differs from the more common notion of *natural density*

$$\delta(S) := \lim_{x \to \infty} \frac{\#\{p \le x : p \in S\}}{\#\{p \le x\}}.$$

Dirichlet proved that for all coprime integers a and m the set of primes $p \equiv a \bmod m$ has Dirichlet density $1/\phi(m)$, whereas the prime number theorem for arithmetic progressions states that this set has natural density $1/\phi(m)$. If a set of primes S has a natural density then it has a Dirichlet density and the two are equal, but the converse need not hold: there are sets of primes that have a Dirichlet density but no natural density (see Problem Set 9).

In order to complete our proof that there are infinitely many primes $p \equiv 1 \bmod 4$, we still need to show $L(1, \chi) \neq 0$. We will achieve this in the next lecture, but for now let us show that this reduces to understanding the behavior of the *Dedekind zeta function*[3] $\zeta_{\mathbb{Q}(i)}(s)$ at $s = 1$. In general the Dedekind zeta function of a number field K is defined by

$$\zeta_K(s) := \sum_{\mathfrak{a}} \mathrm{N}(\mathfrak{a})^{-s} = \prod_{\mathfrak{p}} (1 - \mathrm{N}(\mathfrak{p})^{-s})^{-1},$$

[3]The Dedekind zeta function is named after Richard Dedekind, the last doctoral student of Gauss. He received his Ph.D. in 1854, the same year as Riemann, another student of Gauss. Dedekind and Riemann both studied under Dirichlet as well.

where the sum ranges over nonzero ideals of the ring of integers $\mathcal{O}_K$, the product ranges over nonzero prime ideals of $\mathcal{O}_K$ (primes of K), and $\mathrm{N}(\mathfrak{a}) := [\mathcal{O}_K : \mathfrak{a}]$ is the absolute norm. Note that $\zeta_{\mathbb{Q}}(s) = \zeta(s)$, so this is a natural generalization of the Riemann zeta function.

That the Euler product for $\zeta_K(s)$ converges for $\mathrm{Re}(s) > 1$ follows easily from the case $\zeta_{\mathbb{Q}}(s) = \zeta(s)$ proved in Theorem 16.2. We use unique factorization of ideals in the Dedekind domain $\mathcal{O}_K$ to convert the sum over ideals $\mathfrak{a}$ into a product over prime ideals $\mathfrak{p}$. We then note that for each rational prime p we have $\#\{\mathfrak{p}|p\} \le [K : \mathbb{Q}] = n$ and $\mathrm{N}(\mathfrak{p}) = p^{f_{\mathfrak{p}}} \ge p$ (by Theorems 5.35 and 6.10), and it follows that

$$\sum_{\mathfrak{p}} \left|\log(1 - \mathrm{N}(\mathfrak{p})^{-s})\right| \;\le\; n \sum_{p} \left|\log(1 - p^{-s})\right|.$$

The sum on the RHS converges on $\mathrm{Re}(s) > 1$, so the sum on the LHS must as well.

For $K = \mathbb{Q}(i)$ we can rewrite the Euler product for $\zeta_K(s)$ as

$$\begin{aligned}
\zeta_K(s) &= \prod_{\mathfrak{p}} (1 - N(\mathfrak{p})^{-s})^{-1} = \prod_{p} \prod_{\mathfrak{p}|p} (1 - \mathrm{N}(\mathfrak{p})^{-s})^{-1} \\
&= (1 - 2^{-s})^{-1} \prod_{p \equiv 1 \bmod 4} (1 - p^{-s})^{-1} (1 - p^{-s})^{-1} \prod_{p \equiv 3 \bmod 4} (1 - p^{-2s})^{-1} \\
&= (1 - 2^{-s})^{-1} \prod_{p \equiv 1 \bmod 4} (1 - p^{-s})^{-1} (1 - p^{-s})^{-1} \prod_{p \equiv 3 \bmod 4} (1 - p^{-s})^{-1} (1 + p^{-s})^{-1} \\
&= \prod_{p} (1 - p^{-s})^{-1} \prod_{p} (1 - \chi(p) p^{-s})^{-1} \\
&= \zeta(s) L(s, \chi),
\end{aligned}$$

where we have used the fact that we have

- one prime $\mathfrak{p}$ of norm $\mathrm{N}(\mathfrak{p}) = 2$ above the single prime $p = 2$ that ramifies in $\mathbb{Q}(i)$;
- two primes $\mathfrak{p}, \bar{\mathfrak{p}}$ of norm $\mathrm{N}(\mathfrak{p}) = \mathrm{N}(\bar{\mathfrak{p}}) = p$ above each prime p that splits in $\mathbb{Q}(i)$, equivalently, the primes $p \equiv 1 \bmod 4$;
- one prime $\mathfrak{p}$ of norm $\mathrm{N}(\mathfrak{p}) = p^2$ above each prime p that remains inert in $\mathbb{Q}(i)$, equivalently, the primes $p \equiv 3 \bmod 4$.

We know $\zeta(s)$ has a simple pole at $s = 1$. If we can show $\zeta_K(s)$ extends to a meromorphic function with a simple pole at $s = 1$, then $L(s, \chi)$ must extend to a function that is holomorphic and nonvanishing at $s = 1$, since

$$\mathrm{ord}_{s=1} L(s, \chi) = \mathrm{ord}_{s=1} \zeta_K(s) - \mathrm{ord}_{s=1} \zeta(s) = -1 - (-1) = 0.$$

In fact, $\zeta_K(s)$ extends to a meromorphic function on $\mathrm{Re}(s) > \frac{1}{2}$ with a simple pole at $s = 1$; this can be proved directly, but it follows from a much more general and striking result, the *analytic class number formula*, which was also proved by Dirichlet (at least for quadratic fields). We will prove the analytic class number formula in the next lecture. For the remainder of this lecture we will focus on generalizing our approach to handle arbitrary moduli m.

18.2 Dirichlet characters

We now define the notion of a Dirichlet character. Historically, these preceded the notion of a group character; they were introduced by Dirichlet in 1831, well before the notion of an abstract group was in common use.[4] In order to simplify the exposition we will occasionally invoke some standard facts about characters of finite abelian groups that we recall in §18.6.

Definition 18.4. A function $f\colon \mathbb{Z} \to \mathbb{C}$ is called an *arithmetic function*.[5] The function f is *multiplicative* if $f(1) = 1$ and $f(mn) = f(m)f(n)$ for all coprime $m, n \in \mathbb{Z}$; it is *totally multiplicative* (or *completely multiplicative*) if $f(1) = 1$ and $f(mn) = f(m)f(n)$ for all $m, n \in \mathbb{Z}$. For $m \in \mathbb{Z}_{>0}$ we say that f is *m-periodic* if $f(n+m) = f(n)$ for all $n \in \mathbb{Z}$, and we call m the *period* of f it is the least $m > 0$ for which this holds.

Definition 18.5. A *Dirichlet character* is a periodic totally multiplicative arithmetic function $\chi\colon \mathbb{Z} \to \mathbb{C}$.

The image of a Dirichlet character is a finite multiplicatively closed subset of $\mathbb{C}$, hence the union of a finite subgroup of $\mathrm{U}(1)$ and a subset of $\{0\}$. The constant function $\mathbb{1}(n) := 1$ is the *trivial Dirichlet character*; it is the unique Dirichlet character of period 1. Each m-periodic Dirichlet character χ restricts to a group character χ on $(\mathbb{Z}/m\mathbb{Z})^\times$. Conversely, every group character χ of $(\mathbb{Z}/m\mathbb{Z})^\times$ can be extended to a Dirichlet character χ by defining $\chi(n) = 0$ for $n \notin (\mathbb{Z}/m\mathbb{Z})^\times$; this is called *extension by zero*.[6]

Definition 18.6. A *Dirichlet character of modulus m* is an m-periodic Dirichlet character χ that is the extension by zero of a group character on $(\mathbb{Z}/m\mathbb{Z})^\times$; equivalently, an m-periodic Dirichlet character for which $n \in (\mathbb{Z}/m\mathbb{Z})^\times \iff \chi(n) \neq 0$.

Remark 18.7. Some authors only define Dirichlet characters of modulus m, thereby baking m into the definition of a Dirichlet character; we simply view Dirichlet characters as functions $\mathbb{Z} \to \mathbb{C}$ that satisfy certain properties. Note that a single Dirichlet character may be a Dirichlet character of modulus m for infinitely many m (for example, the unique Dirichlet character of modulus 2 is also a Dirichlet character of modulus 2^k for all $k \geq 1$).

The Dirichlet characters of modulus m form a group under pointwise multiplication that is canonically isomorphic to the character group of $(\mathbb{Z}/m\mathbb{Z})^\times$. Not every m-periodic Dirichlet character χ is a Dirichlet character of modulus m, since an m-periodic Dirichlet character need not vanish on $n \notin (\mathbb{Z}/m\mathbb{Z})^\times$. More generally, we have the following lemma.

Lemma 18.8. *Let χ be a Dirichlet character of period m. Then χ is a Dirichlet character of modulus m' if and only if $m | m' | m^k$ for some k (which holds in particular for $m' = m$).*

Proof. Suppose for the sake of contradiction that $\chi(n) \neq 0$ for some $n \in \mathbb{Z}$ that has a prime factor p in common with m. Then $\chi(p) \neq 0$, since $\chi(p)\chi(n/p) = \chi(n) \neq 0$, and for $r \in \mathbb{Z}$,

$$\chi(r)\chi(p) = \chi(rp) = \chi(rp + m) = \chi(r + m/p)\chi(p),$$

which implies $\chi(r) = \chi(r + m/p)$, since $\chi(p) \neq 0$. Thus χ is (m/p)-periodic, but this contradicts the minimality of the period m. Therefore $\chi(n) = 0$ for all $n \notin (\mathbb{Z}/m\mathbb{Z})^\times$.

[4] Galois' seminal paper was rejected that same year; it wasn't published until 12 years after his death.

[5] Many authors restrict the domain of an arithmetic function to $\mathbb{Z}_{\geq 1}$; for the periodic arithmetic functions we are interested in here, this distinction is irrelevant, and it is slightly more natural to work with $\mathbb{Z}$.

[6] When we write $n \notin (\mathbb{Z}/m\mathbb{Z})^\times$ we of course refer to the image of n under the quotient map $\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$.

Conversely, for any $n \in (\mathbb{Z}/m\mathbb{Z})^\times$ we can pick an integer $a = n^e \equiv 1 \bmod m$ so that $\chi(1) = \chi(a) = \chi(n^e) = \chi(n)^e \neq 0$ and $\chi(n) \neq 0$. So χ is a Dirichlet character of modulus m.

If $m|m'|m^k$, then the prime divisors of m' coincide with those of m. It follows that

$$n \in (\mathbb{Z}/m'\mathbb{Z})^\times \iff n \in (\mathbb{Z}/m\mathbb{Z})^\times \iff \chi(n) \neq 0,$$

and χ is clearly m'-periodic (since $m|m'$), so χ is a Dirichlet character of modulus m'.

Conversely, if χ is a Dirichlet character of modulus m', then χ is m'-periodic, and therefore $m|m'$, since m is the period of χ. And since χ is a Dirichlet character of modulus m and of modulus m', for each prime p we have

$$p \notin (\mathbb{Z}/m\mathbb{Z})^\times \iff \chi(p) = 0 \iff p \notin (\mathbb{Z}/m'\mathbb{Z})^\times,$$

thus the prime divisors of m and m' coincide and m' must divide some power m^k of m. □

18.2.1 Primitive Dirichlet characters

Given a Dirichlet character χ_1 of modulus m_1 dividing m_2, we can always create a Dirichlet character χ_2 of modulus m_2 by taking the extension by zero of the restriction of χ_1 to $(\mathbb{Z}/m_2\mathbb{Z})^\times$; in other words, let $\chi_2(n) := \chi_1(n)$ for $n \in (\mathbb{Z}/m_2\mathbb{Z})^\times$ and $\chi_2(n) := 0$ otherwise. If m_2 is divisible by a prime p that does not divide m_1, the Dirichlet characters χ_1 and χ_2 will not be the same ($\chi_2(p) = 0 \neq \chi_1(p)$, for example), they will agree on $n \in (\mathbb{Z}/m_2\mathbb{Z})^\times$ but not on $n \in (\mathbb{Z}/m_1\mathbb{Z})^\times$.[7] We can create infinitely many new Dirichlet characters from χ_1 in this way, but they will differ from χ_1 only in a rather trivial sense. We would like to distinguish the Dirichlet characters that arise in this way from those that do not.

Definition 18.9. Let χ_1 and χ_2 be Dirichlet characters of modulus m_1 and m_2, respectively, with $m_1|m_2$. If $\chi_2(n) = \chi_1(n)$ for $n \in (\mathbb{Z}/m_2\mathbb{Z})^\times$ then χ_2 is *induced* by χ_1. A Dirichlet character that is not induced by any character other than itself is *primitive*.

Lemma 18.10. *A Dirichlet character χ_2 of modulus m_2 is induced by a Dirichlet character of modulus $m_1|m_2$ if and only if χ_2 is constant on residue classes in $(\mathbb{Z}/m_2\mathbb{Z})^\times$ that are congruent modulo m_1. When this holds, the Dirichlet character χ_1 of modulus m_1 that induces χ_2 is uniquely determined.*

Proof. If χ_2 is induced by χ_1 then it must be constant on residue classes in $(\mathbb{Z}/m_2\mathbb{Z})^\times$ that are congruent modulo m_1, since χ_1 is. To prove the converse we first show that the surjective ring homomorphism $\mathbb{Z}/m_2\mathbb{Z} \to \mathbb{Z}/m_1\mathbb{Z}$ given by reduction modulo m_1 induces a surjective homomorphism $\pi\colon (\mathbb{Z}/m_2\mathbb{Z})^\times \to (\mathbb{Z}/m_1\mathbb{Z})^\times$ of unit groups.[8]

Suppose $u_1 \in \mathbb{Z}$ is a unit modulo m_1. Let a be the product of all primes dividing m_2/m_1 but not u_1. Then $u_2 = u_1 + m_1 a$ is not divisible by any prime $p|m_1$ (since u_1 isn't), nor is it divisible by any prime $p|(m_2/m_1)$: by construction, such a p divides exactly one of u_1 and $m_1 a$. Thus u_2 is a unit modulo m_2 that reduces to u_1 modulo m_1 and π is surjective.

If χ_2 is a Dirichlet character of modulus m_2 constant on fibers of π we can define a Dirichlet character χ_1 of modulus m_1 via $\chi_1(n_1) := \chi_2(n_2)$ for $n_1 \in (\mathbb{Z}/m_1\mathbb{Z})^\times$ with $n_2 \in \pi^{-1}(n_1)$ (any such n_2 will do). Thus χ_1 induces χ_2, and if χ_1' also induces χ_2 it must satisfy the same condition $\chi_1(n_1) = \chi_2(n_2)$ that uniquely determines χ_1. □

[7] Note that while $\#(\mathbb{Z}/m_1\mathbb{Z})^\times \leq \#(\mathbb{Z}/m_2\mathbb{Z})^\times$, the set of integers $n \in (\mathbb{Z}/m_1\mathbb{Z})^\times$ (the n coprime to m_1) contains the set of integers $n \in (\mathbb{Z}/m_2\mathbb{Z})^\times$ (the n coprime to m_2) and is usually larger.

[8] In fact, one can show that every surjective homomorphism of finite rings induces a surjective homomorphism of unit groups, but this does not hold in general (consider $\mathbb{Z} \to \mathbb{Z}/5\mathbb{Z}$, for example).

Definition 18.11. A Dirichlet character χ induced by $\mathbb{1}$ is called *principal* (and is primitive if only if $\chi = \mathbb{1}$). For $m \in \mathbb{Z}_{>0}$ we use $\mathbb{1}_m$ to denote the principal Dirichlet character of modulus m; it corresponds to the trivial character of $(\mathbb{Z}/m\mathbb{Z})^\times$.

Lemma 18.12. *Let χ be a Dirichlet character of modulus m. Then*

$$\sum_{n \in \mathbb{Z}/m\mathbb{Z}} \chi(n) \neq 0 \quad \iff \quad \chi = \mathbb{1}_m.$$

Proof. We have $\chi(n) = 0$ for $n \notin (\mathbb{Z}/m\mathbb{Z})^\times$, and the sum over $(\mathbb{Z}/m\mathbb{Z})^\times$ is nonzero if and only if χ restricts to the trivial character on $(\mathbb{Z}/m\mathbb{Z})^\times$, by the orthogonality of characters; see Corollary 18.37. □

Note that the principal Dirichlet characters $\mathbb{1}_m$ and $\mathbb{1}_{m'}$ necessarily coincide when $m|m'|m^k$; for example the principal Dirichlet character of modulus 2 (the parity function) is the same as the principal Dirichlet character of modulus 4 (and every power of 2).

Theorem 18.13. *Every Dirichlet character χ is induced by a primitive Dirichlet character $\widetilde{\chi}$ that is uniquely determined by χ.*

Proof. Let us define a partial ordering $\preceq$ on the set of all Dirichlet characters by defining $\chi_1 \preceq \chi_2$ if χ_1 induces χ_2. The relation $\preceq$ is clearly reflexive, and it follows from Lemma 18.10 that it is transitive.

Let χ be a Dirichlet character of period m and consider the set $X = \{\chi' : \chi' \preceq \chi\}$. Each $\chi' \in X$ necessarily has period m' dividing m and there is at most one χ' of period m' for each divisor m' of m, by Lemma 18.10. Thus X is finite, and nonempty (since $\chi \in X$).

Suppose $\chi_1, \chi_2 \in X$ have periods m_1 and m_2, respectively. Then m_1 and m_2 both divide m, as does $m_3 = \gcd(m_1, m_2)$. We have a commutative square of surjective unit group homomorphisms induced by reduction maps:

$$\begin{array}{ccc} (\mathbb{Z}/m\mathbb{Z})^\times & \twoheadrightarrow & (\mathbb{Z}/m_1\mathbb{Z})^\times \\ \downarrow & & \downarrow \\ (\mathbb{Z}/m_2\mathbb{Z})^\times & \twoheadrightarrow & (\mathbb{Z}/m_3\mathbb{Z})^\times. \end{array}$$

From Lemma 18.10 we know that χ is constant on residue classes in $(\mathbb{Z}/m\mathbb{Z})^\times$ that are congruent modulo either m_1 or m_2, and therefore χ is constant on residue classes in $(\mathbb{Z}/m\mathbb{Z})^\times$ that are congruent modulo m_3, as are χ_1 and χ_2 (which are determined by χ). It follows that there is a unique Dirichlet character χ_3 of modulus m_3 that induces χ, χ_1, and χ_2.

Thus every pair $\chi_1, \chi_2 \in X$ has a lower bound χ_3 under the partial ordering $\preceq$ that is compatible with the total ordering of X by period. This implies that X contains a unique element $\widetilde{\chi}$ that is minimal, both with respect to the partial ordering $\preceq$ and with respect to the total ordering by period; it must be primitive, by the transitivity of $\preceq$. □

Definition 18.14. The *conductor* of a Dirichlet character χ is the period of the unique primitive Dirichlet character $\widetilde{\chi}$ that induces χ.

Corollary 18.15. *For a Dirichlet character χ of modulus m we have $\sum_{n \in \mathbb{Z}/m\mathbb{Z}} \chi(n) \neq 0$ if and only if χ has conductor 1.*

Proof. This follows immediately from Lemma 18.12. □

Corollary 18.16. *Let $M(m)$ denote the set of Dirichlet characters of modulus m, let $X(m)$ denote the set of primitive Dirichlet characters of conductor dividing m, and let $\widehat{G}(m)$ denote the character group of $(\mathbb{Z}/m\mathbb{Z})^\times$. We have canonical bijections*

$$\begin{aligned} M(m) &\xrightarrow{\sim} X(m) \xrightarrow{\sim} \widehat{G}(m) \\ \chi &\longmapsto \widetilde{\chi} \qquad \longmapsto (n \mapsto \widetilde{\chi}(n)). \end{aligned}$$

Proof. By Theorem 18.13, the map $\chi \to \widetilde{\chi}$ is injective, and it is also surjective: each $\widetilde{\chi} \in X(m)$ induces the character $\chi \in M(m)$ by setting $\chi(n) := \widetilde{\chi}(n)$ for $n \in (\mathbb{Z}/m\mathbb{Z})^\times$ and extending by zero. As previously noted, the map $\chi \to (m \mapsto \chi(m))$ defines a bijection $M \to \widehat{G}(m)$ (a group isomorphism, in fact), and this bijection factors through the map $\chi \mapsto \widetilde{\chi}$, since $\widetilde{\chi}(n) = \chi(n)$ for $n \in (\mathbb{Z}/m\mathbb{Z})^\times$. □

Remark 18.17. Corollary 18.16 implies that we can make $X(m)$ a group by defining $\widetilde{\chi_1}\widetilde{\chi_2} := \widetilde{\chi_1\chi_2}$. Note that $\widetilde{\chi_1\chi_2}$ is **not** the pointwise product of $\widetilde{\chi}_1$ and $\widetilde{\chi}_2$ (which is typically not primitive), it is the unique primitive character that induces the pointwise product.

Example 18.18. 12-periodic Dirichlet characters, ordered by period m and conductor c.

m	c	0	1	2	3	4	5	6	7	8	9	10	11	mod-12	principal	primitive
1	1	1	1	1	1	1	1	1	1	1	1	1	1	no	yes	yes
2	1	0	1	0	1	0	1	0	1	0	1	0	1	no	yes	no
3	1	0	1	1	0	1	1	0	1	1	0	1	1	no	yes	no
3	3	0	1	-1	0	1	-1	0	1	-1	0	1	-1	no	no	yes
4	4	0	1	0	-1	0	1	0	-1	0	1	0	-1	no	no	yes
6	1	0	1	0	0	0	1	0	1	0	0	0	1	yes	yes	no
6	3	0	1	0	0	0	-1	0	1	0	0	0	-1	yes	no	no
12	4	0	1	0	0	0	1	0	-1	0	0	0	-1	yes	no	no
12	12	0	1	0	0	0	-1	0	-1	0	0	0	1	yes	no	yes

The fact that $\chi(n) \in \{0, \pm 1\}$ for all 12-periodic Dirichlet characters χ follows from the fact that the exponent of $(\mathbb{Z}/m\mathbb{Z})^\times$ is 2; thus $(\operatorname{im}\chi) \cap \mathrm{U}(1) \subseteq \mu_2 = \{\pm 1\}$.

18.3 Dirichlet L-functions

Definition 18.19. The *Dirichlet L-function* associated to a Dirichlet character χ is

$$L(s,\chi) := \prod_p (1 - \chi(p)p^{-s})^{-1} = \sum_{n \geq 1} \chi(n) n^{-s}.$$

The sum and product converge absolutely for $\operatorname{Re} s > 1$, since $|\chi(n)| \leq 1$, thus $L(s,\chi)$ is holomorphic on $\operatorname{Re}(s) > 1$.

For the trivial Dirichlet character $\mathbb{1}$ have $L(s,\mathbb{1}) = \zeta(s)$. For the principal character $\mathbb{1}_m$ of modulus m induced by $\mathbb{1}$ we have

$$\zeta(s) = L(s,\mathbb{1}_m) \prod_{p \mid m} (1 - p^{-s})^{-1}.$$

The product on the RHS is finite, hence bounded and nonzero as $s \to 1^+$, so the L-function $L(s,\mathbb{1}_m)$ has a simple pole at $s = 1$ with residue

$$\operatorname{res}_{s=1} L(s,\mathbb{1}_m) = \lim_{s \to 1} (s-1)\zeta(s) \prod_{p \mid m} (1 - p^{-s}) = \prod_{p \mid m} (1 - p^{-1}) = \frac{\phi(m)}{m}.$$

The L-functions of non-principal Dirichlet characters do not have a pole at $s = 1$.

Proposition 18.20. *Let χ be a non-principal Dirichlet character of modulus m. Then $L(s,\chi)$ extends to a holomorphic function on* $\operatorname{Re} s > 0$.

Proof. Define the function $T\colon \mathbb{R}_{\geq 0} \to \mathbb{C}$ by

$$T(x) := \sum_{0<n\leq x} \chi(n).$$

For any $x \in \mathbb{R}_{\geq 0}$ Lemma 18.15 implies

$$T(x+m) - T(x) = \sum_{x<n\leq x+m} \chi(n) = \sum_{n\in\mathbb{Z}/m\mathbb{Z}} \chi(n) = 0,$$

since χ is non-principal. Thus $T(x)$ is periodic modulo m and therefore bounded.

Writing $L(s,\chi)$ as a Stieltjes integral (see §18.5) and integrating by parts yields

$$\begin{aligned} L(s,\chi) &= \sum_{n\geq 1} \chi(n) n^{-s} \\ &= \int_0^\infty x^{-s}\, dT(x) \\ &= x^{-s}T(x)\,\Big|_0^\infty - \int_0^\infty T(x)\, d(x^{-s}) \\ &= 0 - \int_0^\infty T(x)(-sx^{-s-1})dx \\ &= s\int_0^\infty T(x)x^{-s-1}dx. \end{aligned}$$

The RHS is holomorphic on $\operatorname{Re} s > 0$, since it is the limit of the uniformly converging sequence of functions $\phi_n(s) := s\int_0^n T(x)x^{-s-1}dx$ (here we use the fact that $T(x)$ is bounded), and is thus the analytic continuation of $L(x,\chi)$ to $\operatorname{Re}(s) > 0$. □

Remark 18.21. In fact, $L(s,\chi)$ extends to a holomorphic function on $\mathbb{C}$ whenever χ is non-principal.

18.4 Primes in arithmetic progressions

We now return to our goal of proving Dirichlet's theorem on primes in arithmetic progressions. We want to show that for coprime integers $a \perp m$ the set $S := \{p \equiv a \bmod m\}$ is infinite, and it suffices to show the sum

$$\sum_{p\equiv a \bmod m} p^{-s}$$

is unbounded as $s \to 1^+$. To convert this to a sum over all primes we use Proposition 18.36 to construct the indicator function

$$\frac{1}{\phi(m)} \sum_{\chi\in X(m)} \chi(p/a) = \begin{cases} 1 & \text{if } p \equiv a \bmod m, \\ 0 & \text{otherwise} \end{cases}$$

where p/a is computed modulo m and χ ranges over primitive Dirichlet characters of conductor dividing m (which we identify with the character group of $(\mathbb{Z}/m\mathbb{Z})^\times$ via Corollary 18.16).

As $s \to 1^+$ we have

$$\begin{aligned}\sum_{p \equiv a \bmod m} p^{-s} &= \sum_p p^{-s} \frac{1}{\phi(m)} \sum_{\chi \in X(m)} \chi(p/a) \\ &= \sum_{\chi \in X(m)} \frac{\chi(1/a)}{\phi(m)} \sum_p \chi(p) p^{-s} \\ &= \sum_{\chi \in X(m)} \frac{\chi(1/a)}{\phi(m)} \big(\log L(s,\chi) + O(1)\big) \\ &= \frac{\log \zeta(s)}{\phi(m)} + \sum_{\substack{\chi \in X(m) \\ \chi \neq \mathbb{1}}} \frac{\chi(1/a)}{\phi(m)} \log L(s,\chi) + O(1).\end{aligned}$$

We now make the key claim that so long as χ is not principal, we have

$$L(1,\chi) \neq 0.$$

This implies that $\log L(s,\chi) = O(1)$ as $s \to 1^+$ and therefore

$$\sum_{p \equiv a \bmod m} p^{-s} = \frac{\log \zeta(s)}{\phi(m)} + O(1)$$

is unbounded as $s \to 1^+$, since $\zeta(s)$ is. Moreover, Mertens' second theorem implies

$$\sum_{\substack{p \le x \\ p \equiv a \bmod m}} \frac{1}{p} \sim \frac{\log\log x}{\phi(m)}.$$

This proves that the set $S = \{p \equiv a \bmod m\}$ is infinite. Moreover, we have

$$d(S) = \lim_{s \to 1^+} \frac{\sum_{p \in S} p^{-s}}{\sum_p p^{-s}} = \lim_{s \to 1^+} \frac{\log \zeta(s)/\phi(m)}{\log \zeta(s)} = \frac{1}{\phi(m)}.$$

We will prove the key claim that $L(1,\chi) \neq 0$ for non-principal χ in the next lecture.

18.5 Stieltjes integrals

For the benefit of those who have not seen them before, we recall a few facts about Stieltjes integrals (also called Riemann-Stieltjes integrals), taken from [1, Ch. 7]. These generalize the Riemann integral but are less general than the Lebesgue integral; they provide a handy way for converting sums to integrals that is often used in analytic number theory.

Definition 18.22. Let f and g be (real or complex valued) functions defined on a nonempty real interval $[a,b]$. For any partition $P = (x_0, \dots, x_n)$ of $[a,b]$ and sequence $T = (t_1, \dots, t_n)$ with $t_k \in [x_{k-1}, x_k]$, we define the *Riemann-Stieltjes sum*

$$S(P,T,f,g) := \sum_{k=1}^n f(t_k)\big(g(x_k) - g(x_{k-1})\big)$$

We say that f is *Riemann-Stieltjes integrable with respect to* g and write $f \in S(g)$ if there is a (real or complex) number S such that for every $\epsilon > 0$ there is a partition P_ϵ of $[a,b]$

such that for every refinement $P = (x_0, \dots, x_n)$ of P_ϵ and every sequence $T = (t_1, \dots, t_n)$ with $t_k \in [x_{k-1}, x_k]$ we have $|S(P, T, f, g) - S| < \epsilon$.[9]

When such an S exists it is necessarily unique and we denote it by $\int_a^b f\,dg$, the *Riemann-Stieltjes integral of f with respect to g*. Improper Riemann-Stieltjes integrals are then defined as limits

$$\int_a^\infty f\,dg := \lim_{b\to\infty} \int_a^b f\,dg$$

(and similarly for the lower limit), and we define $\int_b^a f\,dg = -\int_a^b f\,dg$ and $\int_a^a f\,dg = 0$.

Taking $g(x) = x$ yields the Riemann integral. The Riemann-Stieltjes integral satisfies the usual properties of linearity, summability, and integration by parts.

Proposition 18.23. *Let f, g, and h be functions on $[a, b]$ and let c_1 and c_2 be constants. The following hold:*

- *If $f, g \in S(h)$ then $\int_a^b (c_1 f + c_2 g)\,dh = c_1 \int_a^b f\,dh + c_2 \int_a^b g\,dh$.*
- *If $f \in S(g), S(h)$ then $\int_a^b f\,d(c_1 g + c_2 h) = c_i \int_a^b f\,dg + c_2 \int_a^b f\,dh$.*
- *If $f \in S(g)$ then for any $c \in [a, b]$ we have $\int_a^b f\,dg = \int_a^c f\,dg + \int_c^b f\,dg$.*
- *If $f \in S(g)$ then $g \in S(f)$ and $\int_a^b f\,dg + \int_a^b g\,df = f(b)g(b) - f(a)g(a)$.*
- *If $f = f_1 + if_2$ and $g = g_1 + ig_2$ with $f_1, f_2 \in S(g_1), S(g_2)$ then*

$$\int_a^b f\,dg = \left(\int_a^b f_1\,dg_1 - \int_a^b f_2\,dg_2\right) + i\left(\int_a^b f_2\,dg_1 + \int_a^b f_1\,dg_2\right).$$

Proof. See [1, Thm. 7.2-7,7.50]. □

The last identity allows us to reduce complex-valued integrals to real-valued integrals. The following proposition allows us to reduce Stieltjes integrals to Riemann integrals.

Proposition 18.24. *Let f and g be real-valued functions on $[a, b]$ and suppose g has a continuous derivative g' on $[a, b]$. Then*

$$\int_a^b f\,dg = \int_a^b f(x)g'(x)dx.$$

Proof. See [1, Thm. 7.8]. □

A key advantage of the Stieltjes integral $\int_a^b f\,dg$ is that neither the integrand f nor the integrator g is required to be continuous. It suffices for f and g to be of bounded variation and not share any discontinuities (and they can even share certain discontinuities, see Theorem 18.26).

Definition 18.25. Let f be a (real or complex valued) function defined on a nonempty real interval $[a, b]$. Then f is of *bounded variation* if there exists a real number M such that

$$\sum_{i=0}^{n-1} |f(x_{i+1}) - f(x_i)| < M$$

[9] This definition (due to Pollard) is more general than that originally given by Stieltjes but is now standard.

for every partition $P = (x_0, \dots, x_n)$ of $[a,b]$. If f has a continuous derivative f' on $[a,b]$ this is equivalent to requiring $\int_a^b |f'(x)|dx < \infty$. Every piecewise monotone function is of bounded variation. In particular, any step function with finitely many discontinuities on $[a,b]$ is of bounded variation.

Theorem 18.26. *Let f and g be functions on $[a,b]$ of bounded variation such that for every $c \in [a,b]$ the function f is continuous from the left at c and the function g is continuous from the right at c. Then $\int_a^b f\,dg$ and $\int_a^b g\,df$ both exist.*

Proof. See [2, Thm. 3.7]. □

Corollary 18.27. *Let f and g be functions on $[a,b]$ such that f and g are not both discontinuous from the left or from the right at integers $n \in [a,b]$, and let $G(x) = \sum_{a<n\le x} g(n)$. Then*

$$\sum_{a<n\le b} f(n)g(n) = \int_a^b f(x)\,dG(x).$$

In particular, the integral on the RHS always exists.

Proof. See [1, Thm. 7.11]. □

As an example of using Stieltjes integrals, let us derive an asymptotic estimate for the the harmonic sum

$$H(x) := \sum_{1\le n\le x} \frac{1}{n}.$$

Theorem 18.28. *For $x \in \mathbb{R}_{\ge 1}$, as $x \to \infty$ we have*

$$H(x) = \log x + \gamma + O\left(\tfrac{1}{x}\right)$$

where $\gamma = \lim_{x\to\infty}(H(x) - \log x) = 0.577216\dots$ is Euler's constant.

Proof. Let $[t]$ denote the greatest integer function. Applying Corollary 18.27 with $g(t) = 1$ and $G(t) = \sum_{1\le n\le t} 1 = [t]$, we have

$$\begin{aligned}
H(x) = \sum_{1\le n\le x} \frac{1}{n} &= \int_{1^-}^{x} \frac{1}{t}\,d[t] \\
&= \frac{[t]}{t}\Big|_{1^-}^{x} - \int_{1^-}^{x} [t]\,d\frac{1}{t} \\
&= \frac{[x]}{x} + \int_{1^-}^{x} \frac{[t]}{t^2}\,dt \\
&= \frac{[x]}{x} + \int_{1^-}^{x} \frac{1}{t}\,dt - \int_{1^-}^{x} \frac{t-[t]}{t^2}\,dt \\
&= \frac{[x]}{x} + \log x - \int_{1^-}^{x} \frac{t-[t]}{t^2}\,dt,
\end{aligned}$$

where we used integration by parts in the second line and applied Proposition 18.24 to get the third line. Now let $\gamma = 1 - \int_{1^-}^{\infty}(t-[t])/t^2\,dt$. Then

$$\begin{aligned}
H(x) &= \frac{[x]}{x} + \log x - 1 + \gamma + \int_x^{\infty} \frac{t-[t]}{t^2}\,dt \\
&= \log x + \gamma + \left(\frac{[x]-x}{x} + \int_x^{\infty} \frac{t-[t]}{t^2}\,dt\right). \qquad (2)
\end{aligned}$$

Both summands in the parenthesized quantity in (2) are clearly $O(\frac{1}{x})$; thus

$$\gamma = \lim_{x\to\infty} (H(x) - \log x),$$

and the theorem follows. □

Remark 18.29. We can refine this estimate by applying a similar analysis to the parenthesized quantity in (2); the key point is that the error term is an exact expression, not an asymptotic estimate, and we can continue this process until we obtain an asymptotic expansion to whatever precision we require. For example, one finds that

$$H(x) = \log x + \gamma + \frac{1}{2x} - \frac{1}{2x^2} + \frac{1}{120x^4} + O\left(\frac{1}{x^6}\right).$$

18.6 A quick recap of the character theory of finite abelian groups

In this section we recall some standard results on characters of finite abelian groups.

Definition 18.30. A *character* of a group G is a homomorphism $\chi\colon G \to \mathrm{U}(1)$.[10] The *character group* (or *dual group*) of G is the abelian group

$$\widehat{G} := \mathrm{Hom}(G, \mathrm{U}(1))$$

under pointwise multiplication: $(\chi_1\chi_2)(g) := \chi_1(g)\chi_2(g)$. The inverse of χ is given by complex conjugation: $\chi^{-1}(g) = \overline{\chi}(g) := \overline{\chi(g)}$. The identity element of $\widehat{G}$ is the *trivial character* $g \mapsto 1$.

Remark 18.31. This definition generalizes to locally compact abelian groups G, in which case each character $\chi\colon G \to \mathrm{U}(1)$ is a homomorphism of topological groups and the dual group $\widehat{G}$ is locally compact under the *compact-open topology* which has a basis of neighborhoods of the identity the sets $U(C,V) := \{\chi \in \widehat{G} : \chi(C) \subseteq V\}$, where C ranges over compact subsets of G and V ranges over open neighborhoods of the identity in $\mathrm{U}(1)$. The locally compact group $\widehat{G}$ is called the *Pontryagin dual* of G.[11] When G is finite it necessarily has the discrete topology (since it must be Hausdorff), every homomorphism $G \to \mathrm{U}(1)$ is automatically continuous, and the compact-open topology on $\widehat{G}$ is also discrete.

Proposition 18.32. *Let G be a finite abelian group with character group $\widehat{G}$. Then $G \simeq \widehat{G}$.*

Proof. As a finite abelian group we can write G as a direct product of cyclic groups

$$G = \langle g_1\rangle \times \cdots \times \langle g_n\rangle \simeq \mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_r\mathbb{Z},$$

with $n_i = |g_i|$, and each $g \in G$ can be uniquely written as $g = \prod_i g_i^{e_i}$ with $0 \le e_i < n_1$. Now fix (not necessarily distinct) primitive n_i-th roots of unity $\alpha_i \in \mathrm{U}(1)$ and for $1 \le i \le r$ define $\chi_i \in \widehat{G}$ via

$$\chi_i(g_j) := \begin{cases} \alpha_i & \text{if } i = j, \\ 1 & \text{if } i \ne j. \end{cases}$$

[10] Some authors call these *unitary characters*, allowing characters to have image in $\mathbb{C}^\times$. When G is finite every character is a unitary character, so this distinction won't concern us.

[11] Some authors define the topology on the Pontryagin duality using uniform convergence on compact sets; for topological groups this is equivalent to the compact-open topology. The unitary group $\mathrm{U}(1) \simeq \mathbb{R}/\mathbb{Z}$ is also referred to as the 1-*torus* or *circle group* and may be denoted $\mathbb{T}$ or S^1 and viewed as an additive group.

Then $|\chi_i| = |\alpha_i| = n_i$, and each $\chi \in \widehat{G}$ can be written uniquely as $\prod_i \chi_i^{e_i}$ with $0 \le e_i < n_i$, where $\chi(g_i) = \alpha_i^{e_i}$ (because a character is completely determined by its values on generators and the χ_i are clearly orthogonal). Therefore

$$\widehat{G} = \langle \chi_1 \rangle \times \cdots \times \langle \chi_n \rangle \simeq \mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_r\mathbb{Z}. \qquad \square$$

Corollary 18.33. *Let G be a finite abelian group. Then $g \in G$ is the identity if and only if $\chi(g) = 1$ for all $\chi \in \widehat{G}$ and $\chi \in \widehat{G}$ is the identity if and only if $\chi(g) = 1$ for all $g \in G$.*

The isomorphism in Proposition 18.32 is not canonical. Indeed, there are $\#\mathrm{Aut}(G)$ distinct ways to choose the α_i used to construct the isomorphism $G \simeq \widehat{G}$. But there is a canonical isomorphism from G to the character group of $\widehat{G}$, the *double dual* of G.

Corollary 18.34. *Let G be a finite abelian group. The evaluation map*

$$g \mapsto (\chi \mapsto \chi(g))$$

is a canonical isomorphism from G to its double dual.

Proof. It is clear that the map above is a homomorphism, and Proposition 18.32 implies that G is isomorphic to its dual group $\widehat{G}$, which is in turn isomorphic to its dual group, the double dual of G). So it suffices to show the map is injective, which follows from Corollary 18.33: if g lies in the kernel then $\chi(g) = 1$ for all $\chi \in \widehat{G}$ and $g = 1_G$, by Corollary 18.33, $\square$

Corollary 18.34 allows us to view G as the character group of $\widehat{G}$ by defining $g(\chi) := \chi(g)$.

Remark 18.35. Corollary 18.34 is a special case of *Pontryagin duality*, which applies to any locally compact abelian group G. For infinite groups, G and $\widehat{G}$ need not be isomorphic; for example, the character group of $\mathbb{Z}$ is isomorphic to $\mathrm{U}(1)$ (but in some cases they are, as when G is $\mathbb{R}$ or $\mathbb{Q}_p$, or any local field, see [3, XV, Lemma 2.2.1]). But the canonical isomorphism between G and its double dual always holds.

This is analogous to the situation with vector spaces: a finite dimensional vector space (which may be an infinite abelian group) is non-canonically isomorphic to its dual space but canonically isomorphic to its double dual via the evaluation map. We should note that for a locally compact topological vector space V over a field k, the Pontryagin dual is not the same thing as the vector space dual: the Pontryagin dual corresponds to $\mathrm{Hom}(V, \mathrm{U}(1))$ (morphisms of locally compact groups) while the vector space dual corresponds to $\mathrm{Hom}_k(V, k)$ (morphisms of topological k-vector spaces). For example, the vector space dual of $\mathbb{Q}$ is isomorphic to $\mathbb{Q}$, as is its double dual, but the Pontryagin dual of $\mathbb{Q}$ is uncountable (thus not isomorphic to $\mathbb{Q}$), even though the Pontryagin double dual is isomorphic to $\mathbb{Q}$.

Proposition 18.36. *Let G be a finite abelian group. For all $g_1, g_2 \in G$ we have*

$$\langle g_1, g_2 \rangle := \frac{1}{\#G} \sum_{\chi \in \widehat{G}} \chi(g_1)\overline{\chi(g_2)} = \begin{cases} 1 & \text{if } g_1 = g_2, \\ 0 & \text{if } g_1 \ne g_2, \end{cases}$$

and for all $\chi_1, \chi_2 \in \widehat{G}$ we have

$$\langle \chi_1, \chi_2 \rangle := \frac{1}{\#G} \sum_{g \in G} \chi_1(g)\overline{\chi_2(g)} = \begin{cases} 1 & \text{if } \chi_1 = \chi_2, \\ 0 & \text{if } \chi_1 \ne \chi_2. \end{cases}$$

Proof. By duality it suffices to consider $\langle g_1, g_2\rangle$. If $g_1 = g_2$ then $\chi(g_1)\overline{\chi(g_2)} = 1$ for all $\chi \in \widehat{G}$ and $\langle g_1, g_2\rangle = \#\widehat{G}/\#G = 1$. If $g_1 \neq g_2$ then by Corollary 18.33 there exists $\lambda \in \widehat{G}$ for which $\alpha := \lambda(g_1)\overline{\lambda(g_2)} = \lambda(g_1 g_2^{-1}) \neq 1$. We then have

$$\alpha\langle g_1, g_2\rangle = \frac{1}{\#G}\sum_{\chi\in\widehat{G}}(\lambda\chi)(g_1)\overline{(\lambda\chi)(g_2)} = \frac{1}{\#G}\sum_{\chi\in\lambda\widehat{G}}\chi(g_1)\overline{\chi(g_2)} = \langle g_1, g_2\rangle,$$

which implies $\langle g_1, g_2\rangle = 0$, since $\alpha \neq 1$. □

Corollary 18.37. *For $\chi \in \widehat{G}$ we have $\sum_{g\in G}\chi(g) \neq 0$ if and only χ is the trivial character. For $g \in G$ we have $\sum_{\chi\in\widehat{G}}\chi(g) \neq 0$ if and only if g is the trivial element.*

Remark 18.38. The *orthogonality of characters* given by Proposition 18.36 is a special case of the orthogonality of characters one encounters in Fourier analysis on compact groups; since G is finite, the weighted sum over G amounts to integrating against its Haar measure (the counting measure μ normalized so that $\mu(G) = 1$).

We conclude our discussion of character groups with a theorem analogous to the fundamental theorem of Galois theory.

Proposition 18.39. *Let G be a finite abelian group. There is an inclusion reversing bijection φ between subgroups H of G and subgroups K of $\widehat{G}$ defined by*

$$\varphi(H) := \{\chi \in \widehat{G} : \chi(h) = 1 \text{ for all } h \in H\}.$$

The inverse bijection ϕ is given by

$$\phi(K) := \{g \in G : \chi(g) = 1 \text{ for all } \chi \in K\},$$

and $\widehat{H} \simeq \widehat{G}/\varphi(H)$ and $K \simeq G/\phi(K)$; in particular, $\#H = [\widehat{G}:\varphi(H)]$ and $\#K = [G:\phi(K)]$.

Proof. Its clear from the definitions that φ and ϕ are inclusion reversing. Let H be a subgroup of G. The group $K = \varphi(H)$ consists of the characters of G whose kernel contains H. It is clear that $H' := \phi(K)$ contains H, since it is equal to the intersection of these kernels, and by duality it is similarly clear that $K' := \varphi(H')$ contains K. We then have $H \subseteq H'$ and $\varphi(H) \subseteq \varphi(H')$, but φ is inclusion reversing so $H = H'$; thus $\phi \circ \varphi$ is the identity map, and by duality, so is $\varphi \circ \phi$.

The restriction map $\widehat{G} \to \widehat{H}$ defined by $\chi \mapsto \chi_{|H}$ is a group homomorphism with kernel $K = \varphi(H)$. It is surjective because if we let $\chi_1 := 1_{\widehat{G}}$ then we have

$$\#H\#K = \sum_{h\in H}\sum_{\chi\in K}\chi(h) = \sum_{h\in H}\sum_{\chi\in K}\chi(h)\overline{\chi_1(h)} = \sum_{g\in G}\sum_{\chi\in K}\chi(g)\overline{\chi_1(g)} = \#G,$$

by Proposition 18.36, and therefore $\#\widehat{H}\#K = \#\widehat{G}$ (by Proposition 18.32). It follows that $\widehat{H} \simeq \widehat{G}/\varphi(H)$, and by duality, $K \simeq G/\phi(K)$. □

References

[1] Tom Apostol, *Mathematical analysis*, 2nd edition, Addison-Wesley, 1974.

[2] Paul Bateman and Harold Diamond, *Analytic number theory: An introductory course*, World Scientific, 2004.

[3] J.W.S. Cassels and A. Fröhlich, *Algebraic number theory*, 2nd edition, London Mathematical Society, 2010.

[4] Franz Mertenz, *Ein Beitrag zur analytischen Zahlentheorie*, J. reine agnew. Math., **78** (1874), 46–62.

19 The analytic class number formula

In the previous lecture we proved Dirichlet's theorem on primes in arithmetic progressions modulo the claim that the L-function $L(s,\chi)$ is holomorphic and nonvanishing at $s=1$ for all non-principal Dirichlet characters χ. To establish this claim we will prove a more general result that has many other applications.

Recall that the Dedekind zeta function of a number field K is defined by

$$\zeta_K(s) := \sum_{\mathfrak{a}} \mathrm{N}(\mathfrak{a})^{-s} = \prod_{\mathfrak{p}} (1-\mathrm{N}(\mathfrak{p})^{-s})^{-1},$$

where $\mathfrak{a}$ ranges over nonzero ideals of $\mathcal{O}_K$ and $\mathfrak{p}$ ranges over nonzero prime ideals of $\mathcal{O}_K$; as we showed in the previous lecture the sum and product converge absolutely on $\mathrm{Re}(s)>1$.

The following theorem is often attributed to Dirichlet, although he originally proved it only for quadratic fields (this is all he needed to prove his theorem on primes in arithmetic progressions, but we will use it in a stronger form). The formula for the limit in the theorem was proved by Dedekind [2, Supplement XI] (as a limit from the right, without an analytic continuation to a punctured neighborhood of $z=1$), and analytic continuation was proved by Landau [3]. Hecke later showed that, like the Riemann zeta function, the Dedekind zeta function has an analytic continuation to all of $\mathbb{C}$ and satisfies a functional equation [1], but we won't take the time to prove this; see Remark §19.13 for details.

Theorem (Analytic Class Number Formula)**.** *Let K be a number field of degree n. The Dedekind zeta function $\zeta_K(z)$ extends to a meromorphic function on $\mathrm{Re}(z) > 1-\frac{1}{n}$ that is holomorphic except for a simple pole at $z=1$ with residue*

$$\lim_{z\to 1^+}(z-1)\zeta_K(z) = \frac{2^r(2\pi)^s h_K R_K}{w_K |D_K|^{1/2}},$$

where r and s are the number of real and complex places of K, respectively, $h_K := \#\operatorname{cl}\mathcal{O}_K$ is the class number, R_K is the regulator, $w_K := \#\mu_K$ is the number of roots of unity, and $D_K := \operatorname{disc}\mathcal{O}_K$ is the absolute discriminant.

Recall that $|D_K|^{1/2}$ is the covolume of $\mathcal{O}_K$ as a lattice in $K_{\mathbb{R}} := K\otimes_{\mathbb{Q}}\mathbb{R} \simeq \mathbb{R}^r\times\mathbb{C}^s$ (Proposition 14.15), and R_K is the covolume of $\Lambda_K := \mathrm{Log}(\mathcal{O}_K^\times)$ as a lattice in the trace-zero hyperplane $\mathbb{R}_0^{r+s}$ (see Definition 15.16). The residue of $\zeta_K(z)$ at $z=1$ thus reflects both the additive and multiplicative structure of the ring of integers $\mathcal{O}_K$.

Remark 19.1. In practice the class number h_K is usually the most difficult quantity in the analytic class number formula to compute. We can approximate the limit on the LHS to any desired precision using a finite truncation of either the sum or product defining $\zeta_K(s)$. Provided we can compute the other quantities to similar precision, this provides a method for computing (or at least bounding) the class number h_K; this explains the origin of the term "analytic class number formula". You will have an opportunity to explore a computational application of this formula on Problem Set 9.

Example 19.2. For $K=\mathbb{Q}$ we have $n=1$, $r=1$, $s=0$, $h=1$, $w=\#\{\pm1\}=2$, $D=1$, and the regulator R is the covolume of a lattice in a zero-dimensional vector space, equivalently, the determinant of a 0×0 matrix, which is 1. In this case the theorem states that $\zeta_{\mathbb{Q}}(z)=\zeta(z)$ is holomorphic on $\mathrm{Re}\,z > 1-\frac{1}{1}=0$ except for a simple pole at $z=1$ with residue

$$\lim_{z\to 1^+}(z-1)\zeta_{\mathbb{Q}}(z) = \frac{2^1(2\pi)^0\cdot 1\cdot 1}{2\cdot |1|^{1/2}} = 1.$$

19.1 Lipschitz parametrizability

In order to prove the analytic class number formula we need an asymptotic estimate for the number of nonzero $\mathcal{O}_K$-ideals $\mathfrak{a}$ with absolute norm $\mathrm{N}(\mathfrak{a})$ bounded by a parameter $t \in \mathbb{R}_{>0}$ that we will let tend to infinity; this is necessary for us to understand the behavior of $\zeta_K(z) = \sum_{\mathfrak{a}} \mathrm{N}(\mathfrak{a})^{-z}$ as $z \to 1^+$. Our strategy is to count points in $\mathrm{Log}(\mathcal{O}_K \cap K^\times)$ that lie inside a suitably chosen region S of $\mathbb{R}^{r+s}$ that we will than scale by t. In order to bound this count as a function of t we need a condition on S that ensures that the count grows smoothly with t; this requires S to have a "reasonable" shape. A sufficient condition for this is *Lipschitz parametrizability*.

Definition 19.3. Let X and Y be metric spaces. A function $f : X \to Y$ is *Lipschitz continuous* if there exists $c > 0$ such that for all distinct $x_1, x_2 \in X$

$$d(f(x_1), f(x_2)) \le c \cdot d(x_1, x_2).$$

Every Lipschitz continuous function is uniformly continuous, but the converse need not hold. For example, the function $f(x) = \sqrt{x}$ on $[0, 1]$ is uniformly continuous but not Lipschitz continuous, since $|\sqrt{1/n} - 0|/|1/n - 0| = \sqrt{n}$ is unbounded as $1/n \to 0$.

Definition 19.4. A set B in a metric space X is *d-Lipschitz parametrizable* if it is the union of the images of a finite number of Lipschitz continuous functions $f_i \colon [0, 1]^d \to X$.

Before stating our next result, we recall the asymptotic notation

$$f(t) = g(t) + O(h(t)) \qquad (\text{as } t \to a),$$

for real or complex valued functions f, g, h of a real variable t, which means

$$\limsup_{t \to a} \left| \frac{f(t) - g(t)}{h(t)} \right| < \infty.$$

Typically $a = \infty$, and this is assumed if a is not specified.

Lemma 19.5. *Let $S \subseteq \mathbb{R}^n$ be a measurable set whose boundary $\partial S := \overline{S} - S^0$ is $(n-1)$-Lipschitz parametrizable. Then*

$$\#(tS \cap \mathbb{Z}^n) = \mu(S)t^n + O(t^{n-1}),$$

as $t \to \infty$, where μ is the standard Lebesgue measure on $\mathbb{R}^n$.

Proof. It suffices to prove the lemma for positive integers, since $\#(tS \cap \mathbb{Z}^n)$ and $\mu(S)t^n$ are both monotonically increasing functions of t and $\mu(S)(t+1)^n - \mu(S)t^n = O(t^{n-1})$. We can partition $\mathbb{R}^n$ as the disjoint union of half-open cubes of the form

$$C(a_1, \dots, a_n) = \{(x_1, \dots, x_n) \in \mathbb{R}_n : x_i \in [a_i, a_i + 1)\},$$

with $a_1, \dots, a_n \in \mathbb{Z}$. Let $\mathcal{C}$ be the set of all such half-open cubes C. For each $t > 0$ define

$$B_0(t) := \#\{C \in \mathcal{C} : C \subseteq tS\},$$
$$B_1(t) := \#\{C \in \mathcal{C} : C \cap tS \neq \emptyset\}.$$

For every $t > 0$ we have

$$B_0(t) \le \#(tS \cap \mathbb{Z}^n) \le B_1(t).$$

We can bound $B_1(t) - B_0(t)$ by noting that each $C(a_1, \dots, a_n)$ counted by this difference contains a point $(a_1, \dots, a_n) \in \mathbb{Z}^n$ within a distance $\sqrt{n} = O(1)$ of a point in $\partial tS = t\partial S$.

Let $f_1, \dots, f_m$ be Lipschitz functions $[0,1]^{n-1} \to \partial S$ whose images cover ∂S, and let $c_1, \dots c_m$ be constants such that $d(f_i(x_1), f_i(x_2)) \le c_i d(x_1, x_2)$ for all $x_1, x_2 \in [0,1]^{n-1}$. For any $y \in \partial S$, we have $y = f_i(x_1, \dots, x_{n-1})$ for some i, and if we put $r_j = \lfloor t x_j \rfloor \in \mathbb{Z}$ so that $0 \le x_j - r_j/t \le 1/t$, then

$$d(y, f_i(\tfrac{r_1}{t}, \dots, \tfrac{r_{n-1}}{t})) \;\le\; c_i \cdot d\big((x_1, \dots, x_{n-1}), (\tfrac{r_1}{t}, \dots, \tfrac{r_{n-1}}{t})\big) \;<\; c_i\sqrt{n}/t \;\le\; c/t,$$

where $c \coloneqq \sqrt{n} \max_i c_i$. Thus every $y \in \partial S$ lies within a distance c/t of a point in the set

$$\mathcal{P} = \left\{ f_i\left(\tfrac{r_1}{t}, \dots, \tfrac{r_{n-1}}{t}\right) : 1 \le i \le m,\ 0 \le r_1, \dots, r_{n-1} \le t \right\},$$

which has cardinality $m(t+1)^{n-1} = O(t^{n-1})$. It follows that every point of ∂tS is within a distance c of one of the $O(t^{n-1})$ points in $t\mathcal{P}$. The number of integer lattice points within a distance $\sqrt{n}$ of a point in $t\partial S$ is thus also $O(t^{n-1})$, and therefore

$$B_1(t) - B_0(t) = O(t^{n-1}).$$

We now note that $B_0(t) \le \mu(tS) \le B_1(t)$ and $\mu(tS) = t^n \mu(S)$; the lemma follows. □

Corollary 19.6. *Let Λ be a lattice in an $\mathbb{R}$-vector space $V \simeq \mathbb{R}^n$ and let $S \subseteq V$ be a measurable set whose boundary is $(n-1)$-Lipschitz parametrizable. Then*

$$\#(tS \cap \Lambda) = \frac{\mu(S)}{\operatorname{covol}(\Lambda)} t^n + O(t^{n-1}).$$

Proof. The case $\Lambda \subseteq \mathbb{Z}^n$ is given by the lemma; note that the normalization of the Haar measure μ is irrelevant, since we are taking a ratio of volumes which is necessarily preserved under the isomorphism of topological vector spaces $V \simeq \mathbb{R}^n$. We now note that if the corollary holds for $s\Lambda$, for some $s > 0$, then it also holds for Λ, since $tS \cap s\Lambda = (t/s)S \cap \Lambda$. For any lattice Λ, we can choose $s > 0$ so that $s\Lambda$ is arbitrarily close to an integer lattice (for example, take s to be the LCM of all denominators appearing in rational approximations of the coordinates of a basis for Λ), which is necessarily a finite index subgroup of $\mathbb{Z}^n$. The corollary follows. □

Remark 19.7. Recall that $\operatorname{covol}(\Lambda) = \mu(F)$ for any fundamental region F for Λ, so the ratio $\mu(S)/\operatorname{covol}(\Lambda) = \mu(S)/\mu(F)$ in Corollary 19.6 does not depend on the normalization of the Haar measure μ. However, we plan to apply the corollary to $\Lambda = \mathcal{O}_K$ and want to replace $\operatorname{covol}(\mathcal{O}_K)$ with $\sqrt{|\operatorname{disc}(\mathcal{O}_K)|} = |D_K|^{1/2}$ via Proposition 14.15, which requires us to use the normalized Haar measure on $K_{\mathbb{R}}$ defined in §14.2.

19.1.1 Counting algebraic integers of bounded norm

Recall from §15.2 that the unit group $K_{\mathbb{R}}^\times$ of $K_{\mathbb{R}} \coloneqq K \otimes_{\mathbb{Q}} \mathbb{R}$ is the locally compact group

$$K_{\mathbb{R}}^\times \simeq \prod_{v|\infty} K_v^\times \simeq \prod_{\text{real } v|\infty} \mathbb{R}^\times \times \prod_{\text{complex } v|\infty} \mathbb{C}^\times.$$

We have a natural embedding

$$\begin{aligned} K^\times &\hookrightarrow K_{\mathbb{R}}^\times \\ x &\mapsto (x_v), \end{aligned}$$

where v ranges over the $r+s$ archimedean places of K; this allows us to view $K^\times$ as a subgroup of $K_{\mathbb{R}}^\times$ that contains the nonzero elements of $\mathcal{O}_K$. In Lecture 15 we defined the continuous homomorphism

$$\begin{aligned}\operatorname{Log} : K_{\mathbb{R}}^\times &\to \mathbb{R}^{r+s}\\ (x_v) &\mapsto (\log \|x_v\|_v),\end{aligned}$$

and proved that we have an exact sequence of abelian groups

$$1 \longrightarrow \mu_K \longrightarrow \mathcal{O}_K^\times \xrightarrow{\operatorname{Log}} \Lambda_K \to 0,$$

in which Λ_K is a lattice in the trace-zero hyperplane $\mathbb{R}_0^{r+s} := \{x \in \mathbb{R}^{r+s} : \mathrm{T}(x) = 0\}$ (where $\mathrm{T}(x)$ is the sum of the coordinates of x). The regulator R_K is the covolume of Λ_K in $\mathbb{R}_0^{r+s}$ (see Definition 15.16), where we endow $\mathbb{R}_0^{r+s}$ with the Euclidean measure induced by any coordinate projection $\mathbb{R}^{r+s} \to \mathbb{R}^{r+s-1}$. By Dirichlet's unit theorem (Theorem 15.12), we can write

$$\mathcal{O}_K^\times = U \times \mu_K,$$

where $U \subseteq \mathcal{O}_K^\times$ is free of rank $r+s-1$ (the subgroup U is not uniquely determined, but let us fix a choice).

We want to estimate the quantity

$$\#\{\mathfrak{a} : \mathrm{N}(\mathfrak{a}) \le t\},$$

where $\mathfrak{a}$ ranges over the nonzero ideals of $\mathcal{O}_K$, as $t \to \infty$. As a first step, let us restrict our attention to nonzero principal ideals $(\alpha) \subseteq \mathcal{O}_K$. We then want to estimate the cardinality of $\{(\alpha) : \mathrm{N}(\alpha) \le t\}$. We have $(\alpha) = (\alpha')$ if and only if $\alpha/\alpha' \in \mathcal{O}_K^\times$, so this is equivalent to

$$\{\alpha \in K^\times \cap \mathcal{O}_K : \mathrm{N}(\alpha) \le t\}/\mathcal{O}_K^\times,$$

where for any set $S \subseteq K_{\mathbb{R}}^\times$, the notation $S/\mathcal{O}_K^\times$ denotes the set of equivalence classes of S under the equivalence relation $\alpha \sim \alpha' \Leftrightarrow \alpha = u\alpha'$ for some $u \in \mathcal{O}_K^\times$. If we now define

$$K_{\mathbb{R},\le t}^\times := \{x \in K_{\mathbb{R}}^\times : \mathrm{N}(x) \le t\} \subseteq K_{\mathbb{R}}^\times \subseteq K_{\mathbb{R}},$$

then we want to estimate the cardinality of the finite set

$$\left(K_{\mathbb{R},\le t}^\times \cap \mathcal{O}_K\right)/\mathcal{O}_K^\times,$$

where the intersection takes place in $K_{\mathbb{R}}$ and produces a subset of $K_{\mathbb{R}}^\times$ that we partition into equivalence classes modulo $\mathcal{O}_K^\times$. To simplify matters, let us replace $\mathcal{O}_K^\times$ with the free group $U \subseteq \mathcal{O}_K^\times$; we then have a w_K–to–1 map

$$(K_{\mathbb{R},\le t}^\times \cap \mathcal{O}_K)/U \longrightarrow \left(K_{\mathbb{R},\le t}^\times \cap \mathcal{O}_K\right)/\mathcal{O}_K^\times.$$

It suffices to estimate the cardinality of $(K_{\mathbb{R},\le t}^\times \cap \mathcal{O}_K)/U$ and divide the result by w_K.

Recall that for $x = (x_v) \in K_{\mathbb{R}}^\times$, the norm map $\mathrm{N}\colon K_{\mathbb{R}}^\times \to \mathbb{R}_{>0}^\times$ is defined by

$$\mathrm{N}(x) := \prod_{v|\infty} \|x_v\|_v = \prod_{v \text{ real}} |x_v|_{\mathbb{R}} \prod_{v \text{ complex}} |x_v|_{\mathbb{C}}^2,$$

and satisfies $\mathrm{T}(\mathrm{Log}\, x) = \log \mathrm{N}(x)$ for all $x \in K_{\mathbb{R}}^{\times}$. We now define a surjective homomorphism

$$\begin{aligned} \nu\colon K_{\mathbb{R}}^{\times} &\twoheadrightarrow K_{\mathbb{R},1}^{\times} \\ x &\mapsto x\mathrm{N}(x)^{-1/n}. \end{aligned}$$

The image of $K_{\mathbb{R},1}^{\times}$ under the Log map is precisely the trace zero hyperplane $\mathbb{R}_0^{r+s}$ in $\mathbb{R}^{r+s}$ in which $\mathrm{Log}(U) = \mathrm{Log}(\mathcal{O}_K^{\times}) = \Lambda_K$ is a lattice. Let us fix a fundamental domain F for the lattice Λ_K in $\mathbb{R}_0^{r+s}$ so that

$$S := \nu^{-1}\left(\mathrm{Log}^{-1}(F)\right)$$

is a set of unique coset representatives for the quotient $K_{\mathbb{R}}^{\times}/U$. If we now define

$$S_{\le t} \coloneqq \{x \in S : \mathrm{N}(x) \le t\} \subseteq K_{\mathbb{R}},$$

we want to estimate the cardinality of the finite set

$$S_{\le t} \cap \mathcal{O}_K.$$

The set $\mathcal{O}_K$ is a lattice in the $\mathbb{R}$-vector space $K_{\mathbb{R}}$ of dimension n. We have $tS_{\le 1} = S_{\le t^n}$, so we can estimate the cardinality of $S_{\le t} = t^{1/n}S_{\le 1}$ via Corollary 19.6 with $S = S_{\le 1}$ and $\Lambda = \mathcal{O}_K$ by replacing t with $t^{1/n}$, provided that the boundary of $S_{\le 1}$ is $(n-1)$-Lipschitz parametrizable, which we now argue.

The kernel of the Log map is $\{\pm 1\}^r \times \mathrm{U}(1)^s$, where $\mathrm{U}(1) = \{z \in \mathbb{C} : z\bar{z} = 1\}$ is the unit circle in $\mathbb{C}$. We thus have a continuous isomorphism of locally compact groups

$$\begin{aligned} K_{\mathbb{R}}^{\times} = (\mathbb{R}^{\times})^r \times (\mathbb{C}^{\times})^s &\xrightarrow{\sim} \mathbb{R}^{r+s} \times \{\pm 1\}^r \times [0, 2\pi)^s \qquad (1) \\ x = (x_1, \dots, x_r, z_1, \dots, z_s) &\longmapsto (\mathrm{Log}\, x) \times (\operatorname{sgn} x_1, \dots, \operatorname{sgn} x_r) \times (\arg z_1, \dots, \arg z_s), \end{aligned}$$

where the map to $\mathbb{R}^{r+s}$ is the Log map, the map to $\{\pm 1\}^r$ is the vector of signs of the r real components, and the map to $[0, 2\pi)^s$ is the vector of angles $\arg z$ such that $z/|z| = e^{i \arg z}$ of the s complex components.

The set $S_{\le 1}$ consists of 2^r connected components, one for each element of $\{\pm 1\}^r$. We can parametrize each of these component using n real parameters as follows:

- $r + s - 1$ parameters in $[0, 1)$ that encode a point in F as an $\mathbb{R}$-linear combination of $\mathrm{Log}(\epsilon_1), \dots, \mathrm{Log}(\epsilon_{r+s-1})$, where $\epsilon_1, \dots, \epsilon_{r+s-1}$ are a basis for U;
- s parameters in $[0, 1)$ that encode an element of $\mathrm{U}(1)^s$;
- a parameter in $(0, 1]$ that encodes the nth-root of the norm.

These parameterizations define a continuously differentiable bijection from the set

$$C = [0, 1)^{n-1} \times (0, 1] \subseteq [0, 1]^n$$

to each of the 2^r disjoint components of $S_{\le 1}$; it can be written out explicitly in terms of exponentials and the identity function. The boundary ∂C is the boundary of the unit n-cube, which is clearly $(n-1)$-Lipschitz parametrizable; thus each component of $S_{\le 1}$, and therefore $S_{\le 1}$ itself, is $(n-1)$-Lipschitz parametrizable.

We now apply Corollary 19.6 to the lattice $\mathcal{O}_K$ and the set $S_{\le 1}$ in the n-dimensional $\mathbb{R}$-vector space $K_{\mathbb{R}}$ with t replaced by $t^{1/n}$, since $S_{\le t} = t^{1/n}S_{\le 1}$. This yields

$$\#(S_{\le t} \cap \mathcal{O}_K) = \frac{\mu(S_{\le 1})}{\mathrm{covol}(\mathcal{O}_K)}(t^{1/n})^n + O\big((t^{1/n})^{n-1}\big) = \left(\frac{\mu(S_{\le 1})}{|D_K|^{1/2}}\right) t + O\big(t^{1-1/n}\big). \qquad (2)$$

Our next task is compute $\mu(S_{\le 1})$; as noted in Remark 19.7, we must use the normalized Haar measure μ on $K_{\mathbb{R}}$ defined in §14.2 when doing so. We will use the isomorphism in (1) to make a change of coordinates, we just need to understand how this affects the Haar measure μ on $K_{\mathbb{R}} = \prod_{v|\infty} K_v \simeq \mathbb{R}^r \times \mathbb{C}^s$. In terms of the standard Lebesgue measures dx and dA on $\mathbb{R}$ and $\mathbb{C}$, we have $\mu = (dx)^r(2dA)^s$, where the $2dA$ reflects the fact that the normalized absolute value $\| \ \|_v$ for each complex place v is the square of the Euclidean absolute value on $\mathbb{C}$. For each factor of $K_{\mathbb{R}}^\times = \prod_{v|\infty} K_v \simeq (\mathbb{R}^\times)^r \times (\mathbb{C}^\times)^s \subseteq \mathbb{R}^r \times \mathbb{C}^s$ we define the maps

$$\begin{aligned} \mathbb{R}^\times &\to \mathbb{R} \times \{\pm 1\} & \mathbb{C}^\times &\to \mathbb{C} \times [0, 2\pi) \\ x &\mapsto (\log|x|, \operatorname{sgn} x) & z &\mapsto (2\log|z|, \arg z) \\ \pm e^\ell &\leftarrow\!\shortmid (\ell, \pm 1) & e^{\ell/2+i\theta} &\leftarrow\!\shortmid (\ell, \theta) \\ dx &\mapsto e^\ell d\ell \mu_{\{\pm 1\}} & 2dA &\mapsto 2e^{\ell/2} d(e^{\ell/2}) d\theta = e^\ell d\ell d\theta, \end{aligned}$$

where $d\ell$ is the Lebesgue measure on $\mathbb{R}$, $\mu_{\{\pm 1\}}$ is the counting measure on $\{\pm 1\}$, and $d\theta$ is the Lebesgue measure on $[0, 2\pi)$. We thus have

$$\begin{aligned} K_{\mathbb{R}}^\times &\xrightarrow{\sim} \mathbb{R}^{r+s} \times \{\pm 1\}^r \times [0, 2\pi)^s \\ \mu &\mapsto e^{\mathrm{T}(\cdot)} \mu_{\mathbb{R}^{r+s}} \mu_{\{\pm 1\}}^r \mu_{[0,2\pi)}^s, \end{aligned}$$

where the trace function $\mathrm{T}(\cdot)$ sums the coordinates of a vector in $\mathbb{R}^{r+s}$.

We now make one further change of coordinates:

$$\begin{aligned} \mathbb{R}^{r+s} &\to \mathbb{R}^{r+s-1} \times \mathbb{R} \\ x = (x_1, \dots, x_{r+s}) &\mapsto (x_1, \dots, x_{r+s-1}, y := \mathrm{T}(x)) \\ e^{\mathrm{T}(x)} \mu_{\mathbb{R}^{r+s}} &\mapsto e^y \mu_{\mathbb{R}^{r+s-1}} dy. \end{aligned}$$

If we let $\pi\colon \mathbb{R}^{r+s} \to \mathbb{R}^{r+s-1}$ denote the coordinate projection, then the measure of $\pi(F)$ in $\mathbb{R}^{r+s-1}$ is, by definition, the regulator R_K (see Definition 15.16).

The Log map gives us a bijection

$$\begin{aligned} S_{\le 1} &\xrightarrow{\sim} F + (-\infty, 0] \left(\frac{1}{n}, \dots, \frac{1}{n}, \frac{2}{n}, \dots, \frac{2}{n}\right), \\ x = \mathrm{N}(x)^{1/n} \nu(x) &\mapsto \mathrm{Log}\, \nu(x) + \log \mathrm{N}(x) \left(\frac{1}{n}, \dots, \frac{1}{n}, \frac{2}{n}, \dots, \frac{2}{n}\right). \end{aligned}$$

The coordinate $y \in (-\infty, 0]$ is given by $y = \mathrm{T}(\mathrm{Log}\, x) = \log \mathrm{N}(x)$, so we can view $S_{\le 1}$ as an infinite union of cosets of $\mathrm{Log}^{-1}(F)$ parameterized by $e^y = \mathrm{N}(x) \in (0, 1]$.

Under our change of coordinates we thus have

$$\begin{aligned} K_{\mathbb{R}}^\times &\xrightarrow{\sim} \mathbb{R}^{r+s-1} \times \mathbb{R} \times \{\pm 1\}^r \times [0, 2\pi)^s \\ S_{\le 1} &\to \pi(F) \times (-\infty, 0] \times \{\pm 1\}^r \times [0, 2\pi)^s. \end{aligned}$$

Since $R_K = \mu_{\mathbb{R}^{r+s-1}}(\pi(F))$, we have

$$\begin{aligned} \mu(S_{\le 1}) &= \int_{-\infty}^0 e^y R_K 2^r (2\pi)^s dy \\ &= 2^r (2\pi)^s R_K. \end{aligned}$$

Plugging this into (2) yields

$$\#(S_{\le t} \cap \mathcal{O}_K) = \left(\frac{2^r (2\pi)^s R_K}{|D_K|^{1/2}}\right) t + O\left(t^{1-1/n}\right). \tag{3}$$

19.2 Proof of the analytic class number formula

We are now ready to prove the analytic class number formula. Our main tool is the following theorem, which uses our analysis in the previous section to give a precise asymptotic estimate on the number of ideals of bounded norm.

Theorem 19.8. *Let K be a number field of degree n As $t \to \infty$, the number of nonzero $\mathcal{O}_K$-ideals $\mathfrak{a}$ of absolute norm $\mathrm{N}(\mathfrak{a}) \le t$ is*

$$\left(\frac{2^r(2\pi)^s h_K R_K}{w_k |D_K|^{1/2}}\right) t + O\left(t^{1-1/n}\right),$$

where r and s are the number of real and complex places of K, respectively, $h_K = \#\operatorname{cl}\mathcal{O}_K$ is the class number, R_K is the regulator, $w_K := \#\mu_K$ is the number of roots of unity, and $D_K := \operatorname{disc}\mathcal{O}_K$ is the absolute discriminant.

Proof. In order to count the nonzero $\mathcal{O}_K$-ideals $\mathfrak{a}$ of absolute norm $\mathrm{N}(\mathfrak{a}) \le t$ we group them by ideal class. For the trivial class, we just need to count nonzero principal ideals (α), equivalently, the number of nonzero $\alpha \in \mathcal{O}_K$ with $\mathrm{N}(\alpha) \le t$, modulo the unit group $\mathcal{O}_K^\times$. Dividing (3) by w_K to account for the w_K-to-1 map

$$S_{\le t} \cap \mathcal{O}_K \longrightarrow (K^\times_{\mathbb{R},\le t} \cap \mathcal{O}_K)/\mathcal{O}_K^\times,$$

we obtain

$$\#\{(\alpha) \subseteq \mathcal{O}_K : \mathrm{N}(\alpha) \le t\} = \left(\frac{2^r(2\pi)^s R_K}{w_K |D_K|^{1/2}}\right) t + O\left(t^{1-1/n}\right). \tag{4}$$

To complete the proof we now show that we get the same answer for every ideal class; the nonzero ideals $\mathfrak{a}$ of norm $\mathrm{N}(\mathfrak{a}) \le t$ are asymptotically equidistributed among ideal classes.

Fix an ideal class $[\mathfrak{a}]$, with $\mathfrak{a} \subseteq \mathcal{O}_K$ nonzero (every ideal class contains an integral ideal, by Theorem 14.19). Multiplication by $\mathfrak{a}$ gives a bijection

$$\begin{aligned}\{\text{ideals } \mathfrak{b} \in [\mathfrak{a}^{-1}] : \mathrm{N}(\mathfrak{b}) \le t\} &\xrightarrow{\times\mathfrak{a}} \{\text{nonzero principal ideals } (\alpha) \subseteq \mathfrak{a} : \mathrm{N}(\alpha) \le t\mathrm{N}(\mathfrak{a})\} \\ &\longrightarrow \{\text{nonzero } \alpha \in \mathfrak{a} : \mathrm{N}(\alpha) \le t\mathrm{N}(\mathfrak{a})\}/\mathcal{O}_K^\times.\end{aligned}$$

Let $S_{[\mathfrak{a}],\le t}$ denote the set on the RHS. The estimate in (4) derived from Corollary 19.6 applies to any lattice in $K_\mathbb{R}$, not just $\mathcal{O}_K$. Replacing $\mathcal{O}_K$ with $\mathfrak{a}$ in (4) we obtain

$$\begin{aligned}\#S_{[\mathfrak{a}],\le t} &= \left(\frac{2^r(2\pi)^s R_K}{w_k \operatorname{covol}(\mathfrak{a})}\right) t\,\mathrm{N}(\mathfrak{a}) + O\big(t^{1-1/n}\big) \\ &= \left(\frac{2^r(2\pi)^s R_K}{w_k \operatorname{covol}(\mathcal{O}_K)\mathrm{N}(\mathfrak{a})}\right) t\,\mathrm{N}(\mathfrak{a}) + O\big(t^{1-1/n}\big) \\ &= \left(\frac{2^r(2\pi)^s R_K}{w_k |D_K|^{1/2}}\right) t + O\big(t^{1-1/n}\big),\end{aligned}$$

since $\operatorname{covol}(\mathfrak{a}) = N(\mathfrak{a})\operatorname{covol}(\mathcal{O}_K)$, by Corollary 14.16. Note that the RHS does not depend on the ideal class $[\mathfrak{a}]$. Summing over ideal classes yields

$$\#\{\text{nonzero ideals } \mathfrak{b} \subseteq \mathcal{O}_K : \mathrm{N}(\mathfrak{b}) \le t\} = \sum_{[\mathfrak{a}] \in \mathrm{cl}(\mathcal{O}_K)} \#S_{[\mathfrak{a}],\le t} = \left(\frac{2^r(2\pi)^s h_K R_K}{w_K |D_K|^{1/2}}\right) t + O\left(t^{1-1/n}\right),$$

as claimed. □

Lemma 19.9. *Let $a_1, a_2, \dots$ be a sequence of complex numbers and let σ be a real number. Suppose that*

$$a_1 + \cdots + a_t = O(t^\sigma) \qquad (\text{as } t \to \infty).$$

Then the Dirichlet series $\sum a_n n^{-s}$ defines a holomorphic function on $\operatorname{Re} s > \sigma$.

Proof. Let $A(x) := \sum_{0<n\le x} a_n$. Writing the Dirichlet sum as a Stieltjes integral (apply Corollary 18.27 with $f(n) = n^{-s}$ and $g(n) = a_n$), for $\operatorname{Re}(s) > \sigma$ we have

$$\begin{aligned}\sum_{n=1}^{\infty} a_n n^{-s} &= \int_{1^-}^{\infty} x^{-s}\, dA(x)\\ &= \frac{A(x)}{x^s}\Big|_{1^-}^{\infty} - \int_{1^-}^{\infty} A(x)\, dx^{-s}\\ &= (0-0) - \int_{1^-}^{\infty} A(x)(-sx^{-s-1})\, dx\\ &= s\int_{1^-}^{\infty} \frac{A(x)}{x^{s+1}}\, dx.\end{aligned}$$

Note that we used $|A(x)| = O(x^\sigma)$ and $\operatorname{Re}(s) > \sigma$ to conclude $\lim_{x\to\infty} A(x)/x^s = 0$. The integral on the RHS converges locally uniformly on $\operatorname{Re}(s) > \sigma$ and the lemma follows. □

Remark 19.10. Lemma 19.9 gives us an *abscissa of convergence* σ for the Dirichlet series $\sum a_n n^{-s}$; this is analogous to the radius of convergence of a power series.

Lemma 19.11. *Let $a_1, a_2, \dots$ be a sequence of complex numbers that satisfies*

$$a_1 + \cdots + a_t = \rho t + O(t^\sigma) \qquad (\text{as } t \to \infty)$$

for some $\sigma \in [0,1)$ and $\rho \in \mathbb{C}^\times$. The Dirichlet series $\sum a_n n^{-s}$ converges on $\operatorname{Re}(s) > 1$ and has a meromorphic continuation to $\operatorname{Re}(s) > \sigma$ that is holomorphic except for a simple pole at $s = 1$ with residue ρ.

Proof. Define $b_n := a_n - \rho$. Then $b_1 + \cdots + b_t = O(t^\sigma)$ and

$$\sum a_n n^{-s} = \rho \sum n^{-s} + \sum b_n n^{-s} = \rho\,\zeta(s) + \sum b_n n^{-s}.$$

We have already proved that the Riemann zeta function $\zeta(s)$ is holomorphic on $\operatorname{Re}(s) > 1$ and has a meromorphic continuation to $\operatorname{Re}(s) > 0$ that is holomorphic except for a simple pole at 1 with residue 1. By the previous lemma, $\sum b_n n^{-s}$ is holomorphic on $\operatorname{Re}(s) > \sigma$, and since $\sigma < 1$, it is holomorphic at $s = 1$. So the entire RHS has a meromorphic continuation to $\operatorname{Re}(s) > \sigma$ that is holomorphic except for the simple pole at 1 coming from $\zeta(s)$, and the residue at $s = 1$ is $\rho \cdot 1 + 0 = \rho$. □

We are now ready to prove the analytic class number formula.

Theorem 19.12 (Analytic Class Number Formula)**.** *Let K be a number field of degree n. The Dedekind zeta function $\zeta_K(z)$ extends to a meromorphic function on $\operatorname{Re}(z) > 1 - \frac{1}{n}$ that is holomorphic except for a simple pole at $z = 1$ with residue*

$$\lim_{z\to 1^+} (z-1)\zeta_K(z) = \rho_K := \frac{2^r (2\pi)^s h_K R_K}{w_K |D_K|^{1/2}},$$

where r and s are the number of real and complex places of K, respectively, $h_K := \#\operatorname{cl}\mathcal{O}_K$ is the class number, R_K is the regulator, $w_K := \mu_K$ is the number of roots of unity, and $D_K := \operatorname{disc}\mathcal{O}_K$ is the absolute discriminant.

Proof. We have

$$\zeta_K(z) = \sum_{\mathfrak{a}} \mathrm{N}(\mathfrak{a})^{-z} = \sum_{t \geq 1} a_t t^{-z},$$

where $\mathfrak{a}$ ranges over nonzero ideals of $\mathcal{O}_K$, and $a_t := \#\{\mathfrak{a} : \mathrm{N}(\mathfrak{a}) = t\}$ with $t \in \mathbb{Z}_{\geq 1}$. If we now define

$$\rho_K := \frac{2^r (2\pi)^s h_K R_K}{w_K |D_K|^{1/2}},$$

then by Theorem 19.8 we have

$$a_1 + \cdots + a_t = \#\{\mathfrak{a} : \mathrm{N}(\mathfrak{a}) \leq t\} = \rho_K t + O(t^{1-1/n}) \qquad (\text{as } t \to \infty).$$

Applying Lemma 19.11 with $\sigma = 1 - 1/n$, we see that $\zeta_K(z) = \sum a_t t^{-z}$ extends to a meromorphic function on $\mathrm{Re}(z) > 1 - 1/n$ that is holomorphic except for a simple pole at $z = 1$ with residue ρ_K. □

Remark 19.13. As previously noted, Hecke proved that $\zeta_K(z)$ extends to a meromorphic function on $\mathbb{C}$ with no poles other than the simple pole at $z = 1$, and it satisfies a functional equation. If we define the *gamma factors*[1]

$$\Gamma_{\mathbb{R}}(z) := \pi^{-z/2}\Gamma\left(\tfrac{z}{2}\right), \qquad \text{and} \qquad \Gamma_{\mathbb{C}}(z) := \Gamma_{\mathbb{R}}(z)\Gamma_{\mathbb{R}}(z+1) = 2(2\pi)^{-z}\Gamma(z),$$

and the *completed zeta function*

$$\xi_K(z) := |D_K|^{z/2}\Gamma_{\mathbb{R}}(z)^r \Gamma_{\mathbb{C}}(z)^s \zeta_K(z),$$

where r and s are the number of real and complex places of K, respectively, then $\xi_K(z)$ is holomorphic except for simple poles at $z = 0, 1$ and satisfies the *functional equation*

$$\xi_K(z) = \xi_K(1 - z).$$

In the case $K = \mathbb{Q}$, we have $r = 1$ and $s = 0$, so

$$\xi_{\mathbb{Q}}(z) = \Gamma_{\mathbb{R}}(z)\zeta(z) = \pi^{-z/2}\Gamma(\tfrac{z}{2})\zeta_{\mathbb{Q}}(z),$$

which is precisely the completed zeta function $Z(z)$ we defined for the Riemann zeta function $\zeta(z) = \zeta_{\mathbb{Q}}(z)$ in Lecture 17 (without any extra factors to remove the zeros at $z = 0, 1$).

19.3 Cyclotomic zeta functions and Dirichlet L-functions

Having proved the analytic class number formula, we now want to complete the proof of Dirichlet's theorem on primes in arithmetic progressions that we began in the previous lecture. To do this we need to establish a connection between Dirichlet L-functions and Dedekind zeta functions of cyclotomic fields.

Recall from Problem Set 4 that we have an isomorphism $\varphi\colon \mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q}) \xrightarrow{\sim} (\mathbb{Z}/m\mathbb{Z})^\times$ canonically defined by $\sigma(\zeta_m) = \zeta_m^{\varphi(\sigma)}$ (independent of the choice of ζ_m). The canonical bijection given by Corollary 18.16 allows us to identify the set $X(m)$ of primitive Dirichlet characters of conductor dividing m with the character group of $(\mathbb{Z}/m\mathbb{Z})^\times \simeq \mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q})$.[2]

[1] The rightmost equality follows from the duplication formula for $\Gamma(s)$. In older texts one may find $\Gamma_{\mathbb{C}}(s)$ defined as $(2\pi)^{-z}\Gamma(z)$, which yields the same functional equation.

[2] As noted in Remark 18.17, the group operation on $X(m)$ is not pointwise multiplication, one multiplies elements of $X(m)$ by taking the unique primitive character that induces the pointwise product.

More generally, given any finite set of primitive Dirichlet characters, if we let m be the LCM of their conductors and consider the subgroup H of $X(m)$ they generate, we may associate to H the subfield $K := \mathbb{Q}(\zeta_m)^{\phi(H)}$, where

$$\phi(H) := \{\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q}) : \chi(\sigma) = 1 \text{ for all } \chi \in H\};$$

we may then regard H as the character group of $\mathrm{Gal}(K/\mathbb{Q})$ via Proposition 18.39. The same applies if we replace m with any multiple m', since $H \subseteq X(m) \subseteq X(m')$ for all $m|m'$ and we will get the same field $K \subseteq \mathbb{Q}(\zeta_m) \subseteq \mathbb{Q}(\zeta_{m'})$.

Conversely, for each subfield K of a cyclotomic field $\mathbb{Q}(\zeta_m)$ there is a corresponding subgroup

$$H := \{\chi \in X(m) : \chi(\sigma) = 1 \text{ for all } \sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_m)/K)\},$$

for which $K = \mathbb{Q}(\zeta_m)^{\phi(H)}$. Note that $K/\mathbb{Q}$ is Galois, since $\mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q})$ is abelian (every subgroup is normal), and we may view H as the character group of $\mathrm{Gal}(K/\mathbb{Q})$. We thus have a one-to-one correspondence between subgroups $H \subseteq X(m)$ and subfields of $K \subseteq \mathbb{Q}(\zeta_m)$ in which H corresponds to the character group of $\mathrm{Gal}(K/\mathbb{Q})$ and $K = \mathbb{Q}(\zeta_m)^{\phi(H)}$.

We will prove that under this correspondence, the Dedekind zeta function of $\zeta_K(s)$ is the product of the Dirichlet L-functions $L(s,\chi)$ for $\chi \in H$. We first note the following.

Proposition 19.14. *Let p be a prime, let m be a positive integer, and let $m' = m/p^{v_p(m)}$. Then $\mathbb{Q}(\zeta_{m'})$ is the maximal extension of $\mathbb{Q}$ in $\mathbb{Q}(\zeta_m)$ unramified at p. In particular, if p does not divide m then $\mathbb{Q}(\zeta_m)$ is unramified at p.*

Proof. By Corollary 10.20, the extension $\mathbb{Q}_p(\zeta_{m'})/\mathbb{Q}_p$ is unramified. It follows from Proposition 12.4 that $\mathbb{Q}(\zeta_{m'})/\mathbb{Q}$ is unramified at p. Applying the same argument to all primes $q \neq p$ dividing m shows that the extension $\mathbb{Q}(\zeta_{p^{v_p(m)}})$ is ramified only at p. By Corollary 14.25, there are no nontrivial unramified extensions of $\mathbb{Q}$, so every subfield of $\mathbb{Q}(\zeta_{p^{v_p(m)}})$ that properly contains $\mathbb{Q}$ is ramified at p. Now $\mathbb{Q}(\zeta_m)$ is the compositum of $\mathbb{Q}(\zeta_{p^{v_p(m)}})$ and $\mathbb{Q}(\zeta_{m'})$, which intersect in $\mathbb{Q}$, so any nontrivial extension of $\mathbb{Q}(\zeta'_m)$ in $\mathbb{Q}(\zeta_m)$ contains a subfield of $\mathbb{Q}(\zeta_{p^{v_p(m)}})$ properly containing $\mathbb{Q}$ which must be ramified at p; the proposition follows. □

Theorem 19.15. *Let $H \subseteq X(m)$ be a group of primitive Dirichlet characters and let $K = \mathbb{Q}(\zeta_m)^{\phi(H)}$ be the corresponding subfield of $\mathbb{Q}(\zeta_m)$, with $\phi(H)$ defined as above. Then*

$$\zeta_K(s) = \prod_{\chi \in H} L(s,\chi).$$

Proof. On the LHS we have

$$\zeta_K(s) = \prod_{\mathfrak{p}} \left(1 - \mathrm{N}(\mathfrak{p})^{-s}\right)^{-1} = \prod_{p} \prod_{\mathfrak{p}|p} \left(1 - \mathrm{N}(\mathfrak{p})^{-s}\right)^{-1},$$

and on the RHS we have

$$\prod_{\chi \in H} L(s,\chi) = \prod_{\chi \in H} \prod_{p} \left(1 - \chi(p)p^{-s}\right)^{-1} = \prod_{p} \prod_{\chi \in H} \left(1 - \chi(p)p^{-s}\right)^{-1}.$$

It thus suffices to prove

$$\prod_{\mathfrak{p}|p} \left(1 - \mathrm{N}(\mathfrak{p})^{-s}\right) \stackrel{?}{=} \prod_{\chi \in H} \left(1 - \chi(p)p^{-s}\right) \tag{5}$$

for each prime p.

Since $K/\mathbb{Q}$ is Galois, we have $[K:\mathbb{Q}] = e_p f_p g_p$, where e_p is the ramification index, f_p is the residue field degree, and $g_p = \#\{\mathfrak{p}|p\}$. On the LHS of (5) we have

$$\prod_{\mathfrak{p}|p} \left(1 - \mathrm{N}(\mathfrak{p})^{-s}\right) = \left(1 - (p^{f_p})^{-s}\right)^{g_p} = \left(1 - (p^{-s})^{f_p}\right)^{g_p},$$

which we note does not change if we replace K with the maximal subfield K' of K in which p is unramified (since K/K' is totally ramified at every prime of K' above p, only e_p changes, not f_p or g_p). On the RHS of (5), we have $\chi(p) = 0$ for all $\chi \in H$ with conductor divisible by p, so we can replace H with the subgroup H' of Dirichlet characters with conductors prime to p. It follows from Proposition 19.14 that $K' = \mathbb{Q}(\zeta_m)^{\phi(H')}$ (to see this, note that if we put $m' = m/p^{v_p(m)}$ then $K' = K \cap \mathbb{Q}(\zeta_{m'})$ and $H' = H \cap X(m')$). Thus without loss of generality we assume $p \nmid m$, so K is unramified at p and we have $\#H = [K:\mathbb{Q}] = f_p g_p$.

Since $K/\mathbb{Q}$ is abelian and unramified at p, the Artin map gives us a Frobenius element σ_p corresponding to the Frobenius automorphism $x \mapsto x^p$ of the residue field, which by definition has order f_p, so σ_p has order f_p in $\mathrm{Gal}(K/\mathbb{Q})$. Viewing H as the character group of $\mathrm{Gal}(K/\mathbb{Q})$, the map $\chi \mapsto \chi(\sigma_p)$ defines a surjective homomorphism from H to the group of f_p-th roots of unity $\alpha \in \mathrm{U}(1)$, and the kernel of this map has cardinality $\#H/f_p = g_p$. Therefore

$$\prod_{\chi \in H} \left(1 - \chi(p)p^{-s}\right) = \prod_{\alpha^{f_p}=1} \left(1 - \alpha p^{-s}\right)^{g_p} = \left(1 - (p^{-s})^{f_p}\right)^{g_p},$$

where the second equality follows from the identity $\prod_{\alpha^{f_p}=1}(1 - \alpha T) = 1 - T^{f_p} \in \mathbb{C}[T]$. □

19.4 Non-vanishing of Dirichlet L-functions with non-principal character

We are now ready to prove the key claim needed to complete our proof of Dirichlet's theorem on primes in arithmetic progressions.

Theorem 19.16. *Let ψ be any non-principal Dirichlet character. Then $L(1,\psi) \neq 0$.*

Proof. Let ψ be a non-principal Dirichlet character, say of modulus m. Then ψ is induced by a non-trivial primitive Dirichlet character $\widetilde{\psi}$ of conductor $\widetilde{m}$ dividing m. The L-functions of ψ and $\widetilde{\psi}$ differ at only finitely many Euler factors $(1 - \psi(p)p^{-s})^{-1}$ (corresponding to primes p dividing $m/\widetilde{m}$), and these factors are clearly nonzero at $s = 1$, since $p > 1$. We thus assume without loss of generality that $\psi = \widetilde{\psi}$ is primitive.

Let K be the mth cyclotomic field $\mathbb{Q}(\zeta_m)$. By Theorem 19.15 we have

$$\zeta_K(s) = \prod_{\chi} L(s,\chi),$$

where χ ranges over the primitive Dirichlet characters of conductor dividing m, including ψ. By the analytic class number formula (Theorem 19.12), the LHS has a simple pole at $s = 1$,

and the same must be true of the RHS. Thus

$$\begin{aligned}\operatorname{ord}_{s=1}\zeta_K(s) &= \operatorname{ord}_{s=1}\prod_{\chi} L(s,\chi)\\ -1 &= \operatorname{ord}_{s=1}L(s,\mathbb{1})\prod_{\chi\neq\mathbb{1}} L(s,\chi)\\ -1 &= \operatorname{ord}_{s=1}\zeta(s)\prod_{\chi\neq\mathbb{1}} L(s,\chi)\\ -1 &= -1+\sum_{\chi\neq\mathbb{1}}\operatorname{ord}_{s=1}L(s,\chi).\end{aligned}$$

Each $\chi \neq \mathbb{1}$ in the sum is necessarily non-principal (since it is primitive). We proved in Proposition 18.20 that for non-principal χ the Dirichlet L-series $L(s,\chi)$ is holomorphic on $\operatorname{Re}(s) > 0$, thus $\operatorname{ord}_{s=1}L(s,\chi) \geq 0$ for all χ appearing in the sum, which can therefore be zero if and only if every term $\operatorname{ord}_{s=1}L(s,\chi)$ is zero. So $L(1,\chi) \neq 0$ for every non-trivial primitive Dirichlet character χ of conductor dividing m, including ψ. □

References

[1] Erich Hecke, *Über die Zetafunktion beliebiger algebraischer Zahlkörper*, Nachr. Ges. Wiss. Göttingen (1917), 77–89.

[2] P.G. Lejeune Dirichlet and Richard Dedekind, *Vorlesungun über Zahlentheorie*, Braunschweig F. Veiweg, 1894.

[3] Edmund Landau, *Neuer Beweis des Primzahlsatzes und Beweis des Primidealsatzes*, Math. Ann. **56**, 645–670.

20 The Kronecker-Weber theorem

In the previous lecture we established a relationship between finite groups of Dirichlet characters and subfields of cyclotomic fields. Specifically, we showed that there is a one-to-one-correspondence between finite groups H of primitive Dirichlet characters of conductor dividing m and subfields K of $\mathbb{Q}(\zeta_m)$ under which H can be viewed as the character group of the finite abelian group $\mathrm{Gal}(K/\mathbb{Q})$ and the Dedekind zeta function of K factors as

$$\zeta_K(s) = \prod_{\chi \in H} L(s, \chi).$$

Now suppose we are given an arbitrary finite abelian extension $K/\mathbb{Q}$. Does the character group of $\mathrm{Gal}(K/\mathbb{Q})$ correspond to a group of Dirichlet characters, and can we then factor the Dedekind zeta function $\zeta_K(s)$ as a product of Dirichlet L-functions?

The answer is yes! This is a consequence of the *Kronecker-Weber theorem*, which states that every finite abelian extension of $\mathbb{Q}$ lies in a cyclotomic field. This theorem was first stated in 1853 by Kronecker [2], who provided a partial proof for extensions of odd degree. Weber [7] published a proof 1886 that was believed to address the remaining cases; in fact Weber's proof contains some gaps (as noted in [5]), but in any case an alternative proof was given a few years later by Hilbert [1]. The proof we present here is adapted from [6, Ch. 14]

20.1 Local and global Kronecker-Weber theorems

We now state the (global) Kronecker-Weber theorem.

Theorem 20.1. *Every finite abelian extension of $\mathbb{Q}$ lies in a cyclotomic field $\mathbb{Q}(\zeta_m)$.*

There is also a local version.

Theorem 20.2. *Every finite abelian extension of $\mathbb{Q}_p$ lies in a cyclotomic field $\mathbb{Q}_p(\zeta_m)$.*

We first show that the local version implies the global one.

Proposition 20.3. *The local Kronecker-Weber theorem implies the global Kronecker-Weber theorem.*

Proof. Let $K/\mathbb{Q}$ be a finite abelian extension. For each ramified prime p of $\mathbb{Q}$, pick a prime $\mathfrak{p}|p$ and let $K_\mathfrak{p}$ be the completion of K at $\mathfrak{p}$ (the fact that $K/\mathbb{Q}$ is Galois means that every $\mathfrak{p}|p$ is ramified with the same ramification index; it makes no difference which $\mathfrak{p}$ we pick). We have $\mathrm{Gal}(K_\mathfrak{p}/\mathbb{Q}_p) \simeq D_\mathfrak{p} \subseteq \mathrm{Gal}(K/\mathbb{Q})$, by Theorem 11.23, so $K_\mathfrak{p}$ is an abelian extension of $\mathbb{Q}_\mathfrak{p}$ and the local Kronecker-Weber theorem implies that $K_\mathfrak{p} \subseteq \mathbb{Q}_p(\zeta_{m_p})$ for some $m_p \in \mathbb{Z}_{\geq 1}$. Let $n_p := v_p(m_p)$, put $m := \prod_p p^{n_p}$ (this is a finite product), and let $L = K(\zeta_m)$. We will show $L = \mathbb{Q}(\zeta_m)$, which implies $K \subseteq \mathbb{Q}(\zeta_m)$.

The field $L = K \cdot \mathbb{Q}(\zeta_m)$ is a compositum of Galois extensions of $\mathbb{Q}$, and is therefore Galois over $\mathbb{Q}$ with $\mathrm{Gal}(L/\mathbb{Q})$ isomorphic to a subgroup of $\mathrm{Gal}(K/\mathbb{Q}) \times \mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q})$, hence abelian (as recalled below, the Galois group of a compositum $K_1 \cdots K_r$ of Galois extensions K_i/F is isomorphic to a subgroup of the direct product of the $\mathrm{Gal}(K_i/F)$). Let $\mathfrak{q}$ be a prime of L lying above a ramified prime $\mathfrak{p}|p$; as above, the completion $L_\mathfrak{q}$ of L at $\mathfrak{q}$ is a finite abelian extension of $\mathbb{Q}_p$, since $L/\mathbb{Q}$ is finite abelian, and we have $L_\mathfrak{q} = K_\mathfrak{p} \cdot \mathbb{Q}_p(\zeta_m)$. Let $F_\mathfrak{q}$ be the maximal unramified extension of $\mathbb{Q}_p$ in $L_\mathfrak{q}$. Then $L_\mathfrak{q}/F_\mathfrak{q}$ is totally ramified

and $\mathrm{Gal}(L_\mathfrak{q}/F_\mathfrak{q})$ is isomorphic to the inertia group $I_p := I_\mathfrak{q} \subseteq \mathrm{Gal}(L/\mathbb{Q})$, by Theorem 11.23 (the $I_\mathfrak{q}$ all coincide because $L/\mathbb{Q}$ is abelian).

It follows from Corollary 10.20 that $K_\mathfrak{p} \subseteq F_\mathfrak{q}(\zeta_{p^{n_p}})$, since $K_\mathfrak{p} \subseteq \mathbb{Q}_p(\zeta_{m_p})$ and $\mathbb{Q}_p(\zeta_{m_p/p^{n_p}})$ is unramified, and that $L_\mathfrak{q} = F_\mathfrak{q}(\zeta_{p^{n_p}})$, since $\mathbb{Q}_p(\zeta_{m/p^{n_p}})$ is unramified. Moreover, we have $F_\mathfrak{q} \cap \mathbb{Q}_p(\zeta_{p^{n_p}}) = \mathbb{Q}_p$, since $\mathbb{Q}_p(\zeta_{p^{n_p}})/\mathbb{Q}_p$ is totally ramified, and it follows that

$$I_p \simeq \mathrm{Gal}(L_\mathfrak{q}/F_\mathfrak{q}) \simeq \mathrm{Gal}(\mathbb{Q}_p(\zeta_{p^{n_p}})/\mathbb{Q}_p) \simeq (\mathbb{Z}/p^{n_p}\mathbb{Z})^\times.$$

Now let I be the group generated by the union of the groups $I_p \subseteq \mathrm{Gal}(L/\mathbb{Q})$ for $p|m$. Since $\mathrm{Gal}(L/\mathbb{Q})$ is abelian, we have $\bigcup I_p \subseteq \prod I_p$, thus

$$\#I \le \prod_{p|m} \#I_p = \prod_{p|m} \#(\mathbb{Z}/p^{n_p}\mathbb{Z})^\times = \prod_{p|m} \phi(p^{n_p}) = \phi(m) = [\mathbb{Q}(\zeta_m) : \mathbb{Q}].$$

Each inertia field L^{I_p} is unramified at p (see Proposition 7.12), as is $L^I \subseteq L^{I_p}$. So $L^I/\mathbb{Q}$ is unramified, and therefore $L^I = \mathbb{Q}$, by Corollary 14.25. Thus

$$[L : \mathbb{Q}] = [L : L^I] = \#I \le [\mathbb{Q}(\zeta_m) : \mathbb{Q}],$$

and $\mathbb{Q}(\zeta_m) \subseteq L$, so $L = \mathbb{Q}(\zeta_m)$ as claimed and $K \subseteq L = \mathbb{Q}(\zeta_m)$. □

To prove the local Kronecker-Weber theorem we first reduce to the case of cyclic extensions of prime-power degree. Recall that if L_1 and L_2 are two Galois extensions of a field K then their compositum $L := L_1L_2$ is Galois over K with Galois group

$$\mathrm{Gal}(L/K) \simeq \{(\sigma_1, \sigma_2) : \sigma_1|_{L_1\cap L_2} = \sigma_2|_{L_1\cap L_2}\} \subseteq \mathrm{Gal}(L_1/K) \times \mathrm{Gal}(L_2/K).$$

The inclusion on the RHS is an equality if and only if $L_1 \cap L_2 = K$. Conversely, if $\mathrm{Gal}(L/K) \simeq H_1 \times H_2$ then by defining $L_2 := L^{H_1}$ and $L_1 := L^{H_2}$ we have $L = L_1L_2$ with $L_1 \cap L_2 = K$, and $\mathrm{Gal}(L_1/K) \simeq H_1$ and $\mathrm{Gal}(L_2/K) \simeq H_2$.

It follows from the structure theorem for finite abelian groups that we may decompose any finite abelian extension L/K into a compositum $L = L_1 \cdots L_n$ of linearly disjoint cyclic extensions L_i/K of prime-power degree. If each L_i lies in a cyclotomic field $\mathbb{Q}(\zeta_{m_i})$, then so does L. Indeed, $L \subseteq \mathbb{Q}(\zeta_{m_1}) \cdots \mathbb{Q}(\zeta_{m_n}) = \mathbb{Q}(\zeta_m)$, where $m := m_1 \cdots m_n$.

To prove the local Kronecker-Weber theorem it thus suffices to consider cyclic extensions $K/\mathbb{Q}_p$ of prime power degree ℓ^r. There two distinct cases: $\ell \ne p$ and $\ell = p$.

20.2 The local Kronecker-Weber theorem for $\ell \ne p$

Proposition 20.4. *Let $K/\mathbb{Q}_p$ be a cyclic extension of degree ℓ^r for some prime $\ell \ne p$. Then K lies in a cyclotomic field $\mathbb{Q}_p(\zeta_m)$.*

Proof. Let F be the maximal unramified extension of $\mathbb{Q}_p$ in K; then $F = \mathbb{Q}_p(\zeta_n)$ for some $n \in \mathbb{Z}_{\ge 1}$, by Corollary 10.19. The extension K/F is totally ramified, and it must be tamely ramified, since the ramification index is a power of $\ell \ne p$. By Theorem 11.10, we have $K = F(\pi^{1/e})$ for some uniformizer π, with $e = [K : F]$. We may assume that $\pi = -pu$ for some $u \in \mathcal{O}_F^\times$, since $F/\mathbb{Q}_p$ is unramified: if $\mathfrak{q}|p$ is the maximal ideal of $\mathcal{O}_F$ then the valuation $v_\mathfrak{q}$ extends v_p with index $e_\mathfrak{q} = 1$ (by Theorem 8.20), so $v_\mathfrak{q}(-pu) = v_p(-p) = 1$. The field $K = F(\pi^{1/e})$ lies in the compositum of $F((-p)^{1/e})$ and $F(u^{1/e})$, and we will show that both fields lie in a cyclotomic extension of $\mathbb{Q}_p$.

The extension $F(u^{1/e})/F$ is unramified, since $v_{\mathfrak{q}}(\mathrm{disc}(x^e-u)) = 0$ for $p \nmid e$, so $F(u^{1/e})/\mathbb{Q}_p$ is unramified and $F(u^{1/e}) = \mathbb{Q}_p(\zeta_k)$ for some $k \in \mathbb{Z}_{\geq 1}$. The field $K(u^{1/e}) = K \cdot \mathbb{Q}_p(\zeta_k)$ is a compositum of abelian extensions, so $K(u^{1/e})/\mathbb{Q}_p$ is abelian, and it contains the subextension $\mathbb{Q}_p((-p)^{1/e})/\mathbb{Q}_p$, which must be Galois (since it lies in an abelian extension) and totally ramified (by Theorem 11.5, since it is an Eisenstein extension). The field $\mathbb{Q}_p((-p)^{1/e})$ contains ζ_e (take ratios of roots of $x^e + p$) and is totally ramified, but $\mathbb{Q}_p(\zeta_e)/\mathbb{Q}_p$ is unramified (since $p \nmid e$), so we must have $\mathbb{Q}_p(\zeta_e) = \mathbb{Q}_p$. Thus $e|(p-1)$, and by Lemma 20.5 below,

$$\mathbb{Q}_p((-p)^{1/e}) \subseteq \mathbb{Q}_p((-p)^{1/(p-1)}) = \mathbb{Q}_p(\zeta_p).$$

It follows that $F((-p)^{1/e}) = F \cdot \mathbb{Q}_p((-p)^{1/e}) \subseteq \mathbb{Q}_p(\zeta_n) \cdot \mathbb{Q}_p(\zeta_p) \subseteq \mathbb{Q}_p(\zeta_{np})$. We then have $K \subseteq F(u^{1/e}) \cdot F((-p)^{1/e}) \subseteq \mathbb{Q}(\zeta_k) \cdot \mathbb{Q}(\zeta_{np}) \subseteq \mathbb{Q}(\zeta_{knp})$ and may take $m = knp$. □

Lemma 20.5. *For any prime p we have $\mathbb{Q}_p\big((-p)^{1/(p-1)}\big) = \mathbb{Q}_p(\zeta_p)$.*

Proof. Let $\alpha = (-p)^{1/(p-1)}$. Then α is a root of the Eisenstein polynomial $x^{p-1} + p$, so the extension $\mathbb{Q}_p((-p)^{1/(p-1)}) = \mathbb{Q}_p(\alpha)$ is totally ramified of degree $p-1$, and α is a uniformizer (by Lemma 11.4 and Theorem 11.5). Let $\pi = \zeta_p - 1$. The minimal polynomial of π is

$$f(x) := \frac{(x+1)^p - 1}{x} = x^{p-1} + px^{p-2} + \cdots + p,$$

which is Eisenstein, so $\mathbb{Q}_p(\pi) = \mathbb{Q}_p(\zeta_p)$ is also totally ramified of degree $p-1$, and π is a uniformizer. We have $u := -\pi^{p-1}/p \equiv 1 \bmod \pi$, so u is a unit in the ring of integers of $\mathbb{Q}_p(\zeta_p)$. If we now put $g(x) = x^{p-1} - u$ then $g(1) \equiv 0 \bmod \pi$ and $g'(1) = p - 1 \not\equiv 0 \bmod \pi$, so by Hensel's Lemma 9.15 we can lift 1 to a root β of $g(x)$ in $\mathbb{Q}_p(\zeta_p)$.

We then have $p\beta^{p-1} = pu = -\pi^{p-1}$, so $(\pi/\beta)^{p-1} + p = 0$, and therefore $\pi/\beta \in \mathbb{Q}_p(\zeta_p)$ is a root of the minimal polynomial of α. Since $\mathbb{Q}_p(\zeta_p)$ is Galois, this implies that $\alpha \in \mathbb{Q}_p(\zeta_p)$, and since $\mathbb{Q}_p(\alpha)$ and $\mathbb{Q}_p(\zeta_p)$ both have degree $p-1$, the two fields coincide. □

To complete the proof of the local Kronecker-Weber theorem, we need to address the case $\ell = p$. Before doing so, we first recall some background on Kummer extensions.

20.3 A brief introduction to Kummer theory

Let n be a positive integer and let K be a field of characteristic prime to n that contains a primitive nth root of unity ζ_n. While we are specifically interested in the case where K is a local or global field, in this section K can be any field that satisfies these conditions.

For any $a \in K$, the field $L = K(\sqrt[n]{a})$ is the splitting field of $f(x) = x^n - a$ over K; the notation $\sqrt[n]{a}$ denotes a particular nth root of a, but it does not matter which root we pick because all the nth roots of a lie in L (if $f(\alpha) = f(\beta) = 0$ then $\alpha/\beta \in \zeta_n^i \in K$ for some $0 \leq i < n$ and $K(\alpha) = K(\beta)$). The polynomial $f(x)$ is separable, since n is prime to the characteristic of K, so L is a Galois extension of K, and $\mathrm{Gal}(L/K)$ is cyclic, since we have an injective homomorphism

$$\begin{aligned} \mathrm{Gal}(L/K) &\hookrightarrow \langle \zeta_n \rangle \simeq \mathbb{Z}/n\mathbb{Z} \\ \sigma &\mapsto \frac{\sigma(\sqrt[n]{a})}{\sqrt[n]{a}}. \end{aligned}$$

This homomorphism is an isomorphism if and only if $x^n - a$ is irreducible.

Kummer's key observation is that the converse holds. In order to prove this we first recall a basic (but often omitted) lemma from Galois theory, originally due to Dedekind.

Lemma 20.6. *Let L/K be a finite extension of fields. The set $\mathrm{Aut}_K(L)$ is a linearly independent subset of the L-vector space of functions $L \to L$.*

Proof. Suppose not. Let $f := c_1\sigma_1 + \cdots + c_r\sigma_r = 0$ with $c_i \in L$, $\sigma_i \in \mathrm{Aut}_K(L)$, and r minimal; we must have $r > 1$, the c_i nonzero, and the σ_i distinct. Choose $\alpha \in L$ so $\sigma_1(\alpha) \neq \sigma_r(\alpha)$ (possible since $\sigma_1 \neq \sigma_r$). We have $f(\beta) = 0$ for all $\beta \in L$, and the same applies to $f(\alpha\beta) - \sigma_1(\alpha)f(\beta)$, which yields a shorter relation $c_2'\sigma_2 + \cdots + c_r'\sigma_r = 0$, where $c_i' = c_i\sigma_i(\alpha) - c_i\sigma_1(\alpha)$ with $c_1' = 0$, which is nontrivial because $c_r' \neq 0$, a contradiction. □

Corollary 20.7. *Let L/K be a cyclic field extension of degree n with Galois group $\langle\sigma\rangle$ and suppose L contains an nth root of unity ζ_n. Then $\sigma(\alpha) = \zeta_n\alpha$ for some $\alpha \in L$.*

Proof. The automorphism σ is a linear transformation of L with characteristic polynomial $x^n - 1$; by Lemma 20.6, this must be its minimal polynomial, since $\{1, \sigma^1, \ldots, \sigma^{n-1}\}$ is linearly independent. Therefore ζ_n is eigenvalue of σ, and the lemma follows. □

Remark 20.8. Corollary 20.7 is a special case of Hilbert's Theorem 90, which replaces ζ_n with any element u of norm $\mathrm{N}_{L/K}(u) = 1$; see [4, Thm. VI.6.1], for example.

Lemma 20.9. *Let K be a field, let $n \geq 1$ be prime to the characteristic of K, and assume $\zeta_n \in K$. If L/K is a cyclic extension of degree n then $L = K(\sqrt[n]{a})$ for some $a \in K$.*

Proof. Let L/K be a cyclic extension of degree n with $\mathrm{Gal}(L/K) = \langle\sigma\rangle$. By Corollary 20.7, there exists an element $\alpha \in L$ for which $\sigma(\alpha) = \zeta_n\alpha$. We have

$$\sigma(\alpha^n) = \sigma(\alpha)^n = (\zeta_n\alpha)^n = \alpha^n,$$

thus $a = \alpha^n$ is invariant under the action of $\langle\sigma\rangle = \mathrm{Gal}(L/K)$ and therefore lies in K. Moreover, the orbit $\{\alpha, \zeta_n\alpha, \ldots, \zeta_n^{n-1}\alpha\}$ of α under the action of $\mathrm{Gal}(L/K)$ has order n, so $L = K(\alpha) = K(\sqrt[n]{a})$ as desired. □

Definition 20.10. Let K be a field with algebraic closure $\overline{K}$, let $n \geq 1$ be prime to the characteristic of K, and assume $\zeta_n \in K$. The *Kummer pairing* is the map

$$\begin{aligned}\langle\cdot,\cdot\rangle\colon \mathrm{Gal}(\overline{K}/K) \times K^\times &\to \langle\zeta_n\rangle\\ (\sigma, a) &\mapsto \frac{\sigma(\sqrt[n]{a})}{\sqrt[n]{a}}\end{aligned}$$

where $\sqrt[n]{a}$ is any nth root of a in $\in \overline{K}^\times$. If α and β are two nth roots of a, then $(\alpha/\beta)^n = 1$, so $\alpha/\beta \in \langle\zeta_n\rangle \subseteq K$ is fixed by σ and $\sigma(\beta)/\beta = \sigma(\beta)/\beta \cdot \sigma(\alpha/\beta)/(\alpha/\beta) = \sigma(\alpha)/\alpha$, so the value of $\langle\sigma, a\rangle$ does not depend on the choice of $\sqrt[n]{a}$. If $a \in K^{\times n}$, then $\langle\sigma, a\rangle = 1$ for all $\sigma \in \mathrm{Gal}(\overline{K}/K)$, so the Kummer pairing depends only on the image of a in $K^\times/K^{\times n}$; thus we may also view it as a pairing on $\mathrm{Gal}(\overline{K}/K) \times K^\times/K^{\times n}$.

Theorem 20.11. *Let K be a field, let $n \geq 1$ be prime to the characteristic of K with $\zeta_n \in K$. The Kummer pairing induces an isomorphism*

$$\begin{aligned}\Phi\colon K^\times/K^{\times n} &\to \mathrm{Hom}\big(\mathrm{Gal}(\overline{K}/K), \langle\zeta_n\rangle\big)\\ a &\mapsto \big(\sigma \mapsto \langle\sigma, a\rangle\big).\end{aligned}$$

Proof. For each $a \in K^\times - K^{\times n}$, if we pick an nth root $\alpha \in \overline{K}$ of a then the extension $K(\alpha)/K$ will be non-trivial and some $\sigma \in \mathrm{Gal}(\overline{K}/K)$ must act nontrivially on α. For this σ we have $\langle \sigma, a\rangle \neq 1$, so $a \notin \ker \Phi$; thus Φ is injective.

Now let $f\colon \mathrm{Gal}(\overline{K}/K) \to \langle \zeta_n\rangle$ be a homomorphism, and put $d := \#\operatorname{im} f$, $H := \ker f$, and $L := \overline{K}^H$. Then $\mathrm{Gal}(L/K) \simeq \mathrm{Gal}(\overline{K}/K)/H \simeq \mathbb{Z}/d\mathbb{Z}$, so L/K is a cyclic extension of degree d, and Lemma 20.9 implies that $L = K(\sqrt[d]{a})$ for some $a \in K$. If we put $e = n/d$ and consider the homomorphisms $\Phi(a^{me})$ for $m \in (\mathbb{Z}/d\mathbb{Z})^\times$, these homomorphisms are all distinct (because the a^{me} are distinct modulo $K^{\times n}$ and Φ is injective), and they all have the same kernel and image as f (their kernels have the same fixed field L because L contains all the dth roots of a). There are $\#(\mathbb{Z}/d\mathbb{Z})^\times = \#\mathrm{Aut}(\mathbb{Z}/d\mathbb{Z})$ distinct isomorphisms $\mathrm{Gal}(\overline{K}/K)/H \simeq \mathbb{Z}/d\mathbb{Z}$, one of which corresponds to f, and each corresponds to one of the $\Phi(a^{me})$. It follows that $f = \Phi(a^{me})$ for some $m \in (\mathbb{Z}/d\mathbb{Z})^\times$, thus Φ is surjective. □

Given a finite subgroup A of $K^\times/K^{\times n}$, we can choose $a_1, \dots, a_r \in K^\times$ so that the images $\bar{a}_i$ of the a_i in $K^\times/K^{\times n}$ form a basis for the abelian group A; this means

$$A = \langle \bar{a}_1\rangle \times \cdots \times \langle \bar{a}_r\rangle \simeq \mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_r\mathbb{Z},$$

where $n_i | n$ is the order of $\bar{a}_i$ in A. For each a_i, the fixed field of the kernel of $\Phi(\bar{a}_i)$ is a cyclic extension of K isomorphic to $L_i := K(\sqrt[n_i]{a_i})$, as in the proof of Theorem 20.11. The fields L_i are linearly disjoint over K (because the a_i correspond to independent generators of A), and their compositum $L = K(\sqrt[n_1]{a_1}, \dots \sqrt[n_r]{a_r})$ has Galois group $\mathrm{Gal}(L/K) \simeq A$, an abelian group whose exponent divides n; such fields L are called *n-Kummer extensions* of K.

Conversely, given an n-Kummer extension L/K, we can iteratively apply Lemma 20.9 to put L in the form $L = K(\sqrt[n_1]{a_1}, \dots, \sqrt[n_r]{a_r})$ with each $a_i \in K^\times$ and $n_i | n$, and the images of the a_i in $K^\times/K^{\times n}$ then generate a subgroup A corresponding to L as above. We thus have a 1-to-1 correspondence between finite subgroups of $K^\times/K^{\times n}$ and (finite) n-Kummer extensions of K (this correspondence also extends to infinite subgroups provided we put a suitable topology on the groups).

So far we have been assuming that K contains all the nth roots of unity. To help handle situations where this is not necessarily the case, we rely on the following lemma.

Lemma 20.12. *Fix $n \in \mathbb{Z}_{>1}$, let F be a field of characteristic prime to n, let $K = F(\zeta_n)$, and let $L = K(\sqrt[n]{a})$ for some $a \in K^\times$. Define the homomorphism $\omega\colon \mathrm{Gal}(K/F) \to (\mathbb{Z}/n\mathbb{Z})^\times$ by $\zeta_n^{\omega(\sigma)} = \sigma(\zeta_n)$. If L/F is abelian then $\sigma(a)/a^{\omega(\sigma)} \in K^{\times n}$ for all $\sigma \in \mathrm{Gal}(K/F)$.*

Proof. Let $G = \mathrm{Gal}(L/F)$, let $H = \mathrm{Gal}(L/K) \subseteq G$, and let A be the subgroup of $K^\times/K^{\times n}$ generated by a. The Kummer pairing induces a bilinear pairing $H \times A \to \langle \zeta_n\rangle$ that is compatible with the Galois action of $\mathrm{Gal}(K/F) \simeq G/H$. In particular, we have

$$\langle h, a^{\omega(\sigma)}\rangle = \langle h, a\rangle^{\omega(\sigma)} = \sigma(\langle h, a\rangle) = \langle h^\sigma, \sigma(a)\rangle = \langle h, \sigma(a)\rangle$$

for all $\sigma \in \mathrm{Gal}(K/F)$ and $h \in H$; the Galois action on H is by conjugation (lift σ to G and conjugate there), but it is trivial because G is abelian (so $h^\sigma = h$). The isomorphism Φ induced by the Kummer pairing is injective, so $a^{\omega(\sigma)} \equiv \sigma(a) \bmod K^{\times n}$. □

20.4 The local Kronecker-Weber theorem for $\ell = p > 2$

We are now ready to prove the local Kronecker-Weber theorem in the case $\ell = p > 2$.

Theorem 20.13. *Let $K/\mathbb{Q}_p$ be a cyclic extension of odd degree p^r. Then K lies in a cyclotomic field $\mathbb{Q}_p(\zeta_m)$.*

Proof. There are two obvious candidates for K, namely, the cyclotomic field $\mathbb{Q}_p(\zeta_{p^{p^r}-1})$, which by Corollary 10.19 is an unramified extension of degree p^r, and the index $p-1$ subfield of the cyclotomic field $\mathbb{Q}_p(\zeta_{p^{r+1}})$, which by Corollary 10.20 is a totally ramified extension of degree p^r (the p^{r+1}-cyclotomic polynomial $\Phi_{p^{r+1}}(x)$ has degree $\phi(p^{r+1}) = p^r(p-1)$ and remains irreducible over $\mathbb{Q}_p$). If K is contained in the compositum of these two fields then $K \subseteq \mathbb{Q}_p(\zeta_m)$, where $m := (p^{p^r}-1)(p^{r+1})$ and the theorem holds. Otherwise, the field $K(\zeta_m)$ is a Galois extension of $\mathbb{Q}_p$ with

$$\mathrm{Gal}(K(\zeta_m)/\mathbb{Q}_p) \simeq \mathbb{Z}/p^r\mathbb{Z} \times \mathbb{Z}/p^r\mathbb{Z} \times \mathbb{Z}/(p-1)\mathbb{Z} \times \mathbb{Z}/p^s\mathbb{Z},$$

for some $s > 0$; the first factor comes from the Galois group of $\mathbb{Q}_p(\zeta_{p^{p^r}-1})$, the second two factors come from the Galois group of $\mathbb{Q}_p(\zeta_{p^{r+1}})$ (note $\mathbb{Q}_p(\zeta_{p^{r+1}}) \cap \mathbb{Q}_p(\zeta_{p^{p^r}-1}) = \mathbb{Q}_p$), and the last factor comes from the fact that we are assuming $K \not\subseteq \mathbb{Q}_p(\zeta_m)$, so $\mathrm{Gal}(K(\zeta_m)/\mathbb{Q}_p(\zeta_m))$ is nontrivial and must have order p^s for some $s \in [1, r]$.

It follows that the abelian group $\mathrm{Gal}(K(\zeta_m)/\mathbb{Q}_p)$ has a quotient isomorphic to $(\mathbb{Z}/p\mathbb{Z})^3$, and the subfield of $K(\zeta_m)$ corresponding to this quotient is an abelian extension of $\mathbb{Q}_p$ with Galois group isomorphic $(\mathbb{Z}/p\mathbb{Z})^3$. By Lemma 20.14 below, no such field exists. □

To prove that $\mathbb{Q}_p$ admits no $(\mathbb{Z}/p\mathbb{Z})^3$-extensions our strategy is to use Kummer theory to show that the corresponding subgroup of $\mathbb{Q}_p(\zeta_p)^\times/\mathbb{Q}_p(\zeta_p)^{\times p}$ given by Theorem 20.11 must have p-rank 2 and therefore cannot exist. For an alternative proof that uses higher ramification groups instead of Kummer theory, see Problem Set 10.

Lemma 20.14. *For $p > 2$ no extension of $\mathbb{Q}_p$ has Galois group isomorphic to $(\mathbb{Z}/p\mathbb{Z})^3$.*

Proof. Suppose for the sake of contradiction that K is an extension of $\mathbb{Q}_p$ with Galois group $\mathrm{Gal}(K/\mathbb{Q}_p) \simeq (\mathbb{Z}/p\mathbb{Z})^3$. Then $K/\mathbb{Q}_p$ is linearly disjoint from $\mathbb{Q}_p(\zeta_p)/\mathbb{Q}_p$, since the order of $G := \mathrm{Gal}(\mathbb{Q}_p(\zeta_p)/\mathbb{Q}_p) \simeq (\mathbb{Z}/p\mathbb{Z})^\times$ is not divisible by p, and $\mathrm{Gal}(K(\zeta_p)/\mathbb{Q}_p(\zeta_p)) \simeq (\mathbb{Z}/p\mathbb{Z})^3$ is a p-Kummer extension. There is thus a subgroup $A \subseteq \mathbb{Q}_p(\zeta_p)^\times/\mathbb{Q}_p(\zeta_p)^{\times p}$ isomorphic to $(\mathbb{Z}/p\mathbb{Z})^3$, for which $K(\zeta_p) = \mathbb{Q}_p(\zeta_p, A^{1/p})$, where $A^{1/p} := \{\sqrt[p]{a} : a \in A\}$ (here we identify elements of A by representatives in $\mathbb{Q}_p(\zeta_p)^\times$ that are determined only up to pth powers).

For any $a \in A$, the extension $\mathbb{Q}_p(\zeta_p, \sqrt[p]{a})/\mathbb{Q}_p$ is abelian, so by Lemma 20.12, we have

$$\sigma(a)/a^{\omega(\sigma)} \in \mathbb{Q}_p(\zeta_p)^{\times p} \tag{1}$$

for all $\sigma \in G$, where $\omega\colon G \xrightarrow{\sim} (\mathbb{Z}/p\mathbb{Z})^\times$ is the isomorphism defined by $\sigma(\zeta_p) = \zeta_p^{\omega(\sigma)}$.

The field $\mathbb{Q}_p(\zeta_p)$ is a totally tamely ramified extension of $\mathbb{Q}_p$ of degree $p-1$ with residue field $\mathbb{Z}/p\mathbb{Z}$; as shown in the proof of Lemma 20.5, we may take $\pi := \zeta_p - 1$ as a uniformizer. For each $a \in A$ we have

$$v_\pi(a) = v_\pi(\sigma(a)) \equiv \omega(\sigma) v_\pi(a) \bmod p,$$

thus $(1-\omega(\sigma))v_\pi(a) \equiv 0 \bmod p$, for all $\sigma \in G$, hence for all $\omega(\sigma) \in \omega(G) = (\mathbb{Z}/p\mathbb{Z})^\times$; for $p > 2$, this implies $v_\pi(a) \equiv 0 \bmod p$. Now a is determined only up to pth-powers, so after multiplying by $\pi^{-v_\pi(a)}$ we may assume $v_\pi(a) = 0$, and after multiplying by a suitable power of $\zeta_{p-1}^p = \zeta_{p-1}$, we may assume $a \equiv 1 \bmod \pi$, since the image of ζ_{p-1} generates the multiplicative group $(\mathbb{Z}/p\mathbb{Z})^\times$ of the residue field.

We may thus assume that $A \subseteq U_1/U_1^p$, where $U_1 := \{u \equiv 1 \bmod \pi\}$. Each $u \in U_1$ can be written as a power series in π with integer coefficients in $[0, p-1]$ and constant coefficient 1.

We have $\zeta_p \in U_1$, since $\zeta_p = 1 + \pi$, and $\zeta_p^b = 1 + b\pi + O(\pi^2)$ for integers $b \in [0, p-1]$.[1] For $a \in A \subseteq U_1$, we can choose b so that for some integer $c \in [0, p-1]$ and $e \in \mathbb{Z}_{\geq 2}$ we have

$$a = \zeta_p^b(1 + c\pi^e + O(\pi^{e+1})).$$

For $\sigma \in G$ we have

$$\frac{\sigma(\pi)}{\pi} = \frac{\sigma(\zeta_p - 1)}{\zeta_p - 1} = \frac{\zeta_p^{\omega(\sigma)} - 1}{\zeta_p - 1} = \zeta_p^{\omega(\sigma)-1} + \cdots + \zeta_p + 1 \equiv \omega(\sigma) \bmod \pi,$$

since each term in the sum is congruent to 1 modulo $\pi = (\zeta_p - 1)$; here we are representing $\omega(\sigma) \in (\mathbb{Z}/p\mathbb{Z})^\times$ as an integer in $[1, p-1]$. Thus $\sigma(\pi) \equiv \omega(\sigma)\pi \bmod \pi^2$ and

$$\sigma(a) = \zeta_p^{b\omega(\sigma)}(1 + c\omega(\sigma)^e\pi^e + O(\pi^{e+1})).$$

We also have

$$a^{\omega(\sigma)} = \zeta_p^{b\omega(\sigma)}(1 + c\omega(\sigma)\pi^e + O(\pi^{e+1})).$$

As we showed for a above, any $u \in U_1$ can be written as $u = \zeta_p^b u_1$ with $u_1 \equiv 1 \bmod \pi^2$. Each interior term in the binomial expansion of $u_1^p = (1 + O(\pi^2))^p$ other than leading 1 is a multiple of $p\pi^2$ with $v_\pi(p\pi^2) = p - 1 + 2 = p + 1$, and it follows that $u^p = u_1^p \equiv 1 \bmod \pi^{p+1}$. Thus every element of U_1^p is congruent to 1 modulo π^{p+1}, and as you will show on the problem set, the converse holds, that is, $U_1^p = \{u \equiv 1 \bmod \pi^{p+1}\}$.

We know from (1) that $\sigma(a)/a^{\omega(\sigma)} \in U_1^p$, so $\sigma(a) = a^{\omega(\sigma)}(1 + O(\pi^{p+1}))$ and therefore

$$\sigma(a) \equiv a^{\omega(\sigma)} \bmod \pi^{p+1}.$$

For $e \leq p$ this is possible only if $\omega(\sigma) = \omega(\sigma)^e$ for every $\sigma \in G$, equivalently, for every $\omega(\sigma) \in \sigma(G) = (\mathbb{Z}/p\mathbb{Z})^\times$, but then $e \equiv 1 \bmod (p-1)$ and we must have $e \geq p$, since $e \geq 2$.

We have shown that every $a \in A$ is represented by an element $\zeta_p^b(1 + c\pi^p + O(\pi^{p+1})) \in U_1$ with $b, c \in \mathbb{Z}$, and therefore lies in the subgroup of U_1/U_1^p generated by ζ_p and $(1 + \pi^p)$, which is an abelian group of exponent p generated by 2 elements, hence isomorphic to a subgroup of $(\mathbb{Z}/p\mathbb{Z})^2$. But this contradicts $A \simeq (\mathbb{Z}/p\mathbb{Z})^3$. □

Remark 20.15. In the proof of Lemma 20.14 above, the elements of $\mathbb{Q}_p(\zeta_p)^\times/\mathbb{Q}_p(\zeta_p)^{\times p}$ that lie in A are quite special. For most $a \in \mathbb{Q}_p(\zeta_p)^\times$ the extension $\mathbb{Q}_p(\zeta_p, \sqrt[p]{a})/\mathbb{Q}_p$ will not be abelian, even though the extensions $\mathbb{Q}_p(\sqrt[p]{a})/\mathbb{Q}_p(\zeta_p)$ and $\mathbb{Q}_p(\zeta_p)/\mathbb{Q}_p$ both are, and we typically will not have $v_\pi(a) \equiv 0 \bmod p$ (consider $a = \pi$). The key point is that we started with an abelian extension $K/\mathbb{Q}_p$, so $K(\zeta_p) = K \cdot \mathbb{Q}_p(\zeta_p)$ is an abelian extension containing $A^{1/p}$; this ensures that for $a \in A$ the fields $\mathbb{Q}_p(\zeta_p, \sqrt[p]{a})$ are abelian.

Remark 20.16. There is an alternative proof to Lemma 20.14 that is much more explicit. One can show that for $p > 2$ the field $\mathbb{Q}_p$ admits exactly $p + 1$ cyclic extensions of degree p: the unramified extension $\mathbb{Q}_p(\zeta_{p^p-1})$ and the extensions $\mathbb{Q}_p[x]/(x^p + px^{p-1} + p(1 + ap))$, for integers $a \in [0, p-1]$; see [3, Prop. 2.3.1]. This implies that $\mathbb{Q}_p$ cannot have a $(\mathbb{Z}/p\mathbb{Z})^3$ extension, since this would imply the existence of $p^2 + p + 1$ cyclic extensions of degree p, one for each index p subgroup of $(\mathbb{Z}/p\mathbb{Z})^3$.

[1] The expression $O(\pi^n)$ denotes a power series in π that is divisible by π^n.

For $p = 2$ there is an extension of $\mathbb{Q}_2$ with Galois group isomorphic to $(\mathbb{Z}/2\mathbb{Z})^3$, the cyclotomic field $\mathbb{Q}_2(\zeta_{24}) = \mathbb{Q}_2(\zeta_3) \cdot \mathbb{Q}_2(\zeta_8)$, so the proof we used for $p > 2$ will not work. However we can apply a completely analogous argument.

Theorem 20.17. *Let $K/\mathbb{Q}_2$ be a cyclic extension of degree 2^r. Then K lies in a cyclotomic field $\mathbb{Q}_2(\zeta_m)$.*

Proof. The unramified cyclotomic field $\mathbb{Q}_2(\zeta_{2^{2^r}-1})$ has Galois group $\mathbb{Z}/2^r\mathbb{Z}$, and the totally ramified cyclotomic field $\mathbb{Q}_2(\zeta_{2^{r+2}})$ has Galois group $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2^r\mathbb{Z}$ (up to isomorphism). Let $m = (2^{2^r} - 1)(2^{r+2})$. If K is not contained in $\mathbb{Q}_2(\zeta_m)$ then

$$\mathrm{Gal}(K(\zeta_m)/\mathbb{Q}_2) \simeq \begin{cases} \mathbb{Z}/2\mathbb{Z} \times (\mathbb{Z}/2^r\mathbb{Z})^2 \times \mathbb{Z}/2^s\mathbb{Z} & \text{with } 1 \le s \le r \\ \text{or} & \\ (\mathbb{Z}/2^r\mathbb{Z})^2 \times \mathbb{Z}/2^s\mathbb{Z} & \text{with } 2 \le s \le r \end{cases}$$

and thus admits a quotient isomorphic to $(\mathbb{Z}/2\mathbb{Z})^4$ or $(\mathbb{Z}/4\mathbb{Z})^3$. By Lemma 20.18 below, no extension of $\mathbb{Q}_2$ has either of these Galois groups, thus K must lie in $\mathbb{Q}_2(\zeta_m)$. □

Lemma 20.18. *No extension of $\mathbb{Q}_2$ has Galois group isomorphic to $(\mathbb{Z}/2\mathbb{Z})^4$ or $(\mathbb{Z}/4\mathbb{Z})^3$.*

Proof. As you proved on Problem Set 4, there are exactly 7 quadratic extensions of $\mathbb{Q}_2$; it follows that no extension of $\mathbb{Q}_2$ has Galois group $(\mathbb{Z}/2\mathbb{Z})^4$, since this group has 15 subgroups of index 2 whose fixed fields would yield 15 distinct quadratic extension of $\mathbb{Q}_2$.

As you proved on Problem Set 5, there are only finitely many extensions of $\mathbb{Q}_2$ of any fixed degree d, and these can be enumerated by considering Eisenstein polynomials in $\mathbb{Q}_2[x]$ of degrees dividing d up to an equivalence relation implied by Krasner's lemma. One finds that there are 59 quartic extensions of $\mathbb{Q}_2$, of which 12 are cyclic; you can find a list of them here. It follows that no extension of $\mathbb{Q}_2$ has Galois group $(\mathbb{Z}/4\mathbb{Z})^3$, since this group has 28 subgroups whose fixed fields would yield 28 distinct cyclic quartic extensions of $\mathbb{Q}_2$. □

References

[1] David Hilbert, *Ein neuer Beweis des Kroneckerschen Fundamentalsatzes über Abelsche Zahlkörper*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klass (1896), 29–39.

[2] Leopold Kronecker, *Uber die algebraisch auflösbaren Gleichungen I* (1853), in *Leopold Kronecker's Werke, Part 4* (ed. K. Hensel), AMS Chelsea Publishing, 1968.

[3] John W. Jones and David P. Roberts, *A database of local fields*, J. Symbolic Comput. **41** (2006), 80–97.

[4] Serge Lang, *Algebra*, 3rd edition, Springer, 2002.

[5] Olaf Neumann, *Two proofs of the Kronecker-Weber theorem "according to Kronecker, and Weber"*, J. Reine Angew. Math. **323** (1981),105–126.

[6] Lawrence C. Washington, *Introduction to cyclotomic fields*, 2nd edition, Springer, 1997.

[7] Heinrich M. Weber, *Theorie der Abel'schen Zahlkörper*, Acta Mathematica **8** (1886), 193–263.

21 Class field theory: ray class groups and ray class fields

In the previous lecture we proved that every abelian extension L of $\mathbb{Q}$ is contained in a cyclotomic field $\mathbb{Q}(\zeta_m)$. The isomorphism $\mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q}) \simeq (\mathbb{Z}/m\mathbb{Z})^\times$ then allows us to view $\mathrm{Gal}(L/\mathbb{Q})$ as a quotient of $(\mathbb{Z}/m\mathbb{Z})^\times$. We would like to replace the base field $\mathbb{Q}$ with an arbitrary number field K, but we need analogs of the cyclotomic fields $\mathbb{Q}(\zeta_m)$ and the abelian Galois groups $(\mathbb{Z}/m\mathbb{Z})^\times$. These analogs are *ray class fields*, and their Galois groups are isomorphic to *ray class groups*. Ray class fields are not, in general, cyclotomic extensions of K; their construction is rather more complicated. Before defining them, let us first recall some properties of the Artin map we defined in Lecture 7.

21.1 The Artin map

Let L/K be a finite Galois extension of global fields, and let $\mathfrak{p}$ be a prime of K. Recall that the Galois group $\mathrm{Gal}(L/K)$ acts on the set $\{\mathfrak{q}|\mathfrak{p}\}$ (primes $\mathfrak{q}$ of L lying above $\mathfrak{p}$) and the stabilizer of $\mathfrak{q}|\mathfrak{p}$ is the decomposition group $D_\mathfrak{q} \subseteq \mathrm{Gal}(L/K)$. By Proposition 7.9, we have a surjective homomorphism

$$\begin{aligned} \pi_\mathfrak{q} \colon D_\mathfrak{q} &\to \mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p}) \\ \sigma &\mapsto \overline{\sigma} := (\overline{\alpha} \mapsto \overline{\sigma(\alpha)}), \end{aligned}$$

where $\alpha \in \mathcal{O}_L$ is any lift of $\overline{\alpha} \in \mathbb{F}_\mathfrak{q} := \mathcal{O}_L/\mathfrak{q}$ to $\mathcal{O}_L$ and $\overline{\sigma(\alpha)}$ is the reduction of $\sigma(\alpha) \in \mathcal{O}_L$ to $\mathbb{F}_\mathfrak{q}$; kernel of $\pi_\mathfrak{q}$ is the inertia group $I_\mathfrak{q}$. If $\mathfrak{q}$ is unramified then $I_\mathfrak{q}$ is trivial and $\pi_\mathfrak{q}$ is an isomorphism. The *Artin symbol* (Definition 7.18) is defined by

$$\left(\frac{L/K}{\mathfrak{q}}\right) := \sigma_\mathfrak{q} := \pi_\mathfrak{q}^{-1}(x \mapsto x^{\#\mathbb{F}_\mathfrak{p}}),$$

where $(x \mapsto x^{\#\mathbb{F}_\mathfrak{p}}) \in \mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p})$ is the Frobenius automorphism, a canonical generator for the cyclic group $\mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p})$. Equivalently, $\sigma_\mathfrak{q}$ is the unique element of $\mathrm{Gal}(L/K)$ for which

$$\sigma_\mathfrak{q}(x) \equiv x^{\#\mathbb{F}_\mathfrak{p}} \bmod \mathfrak{q}$$

for all $x \in \mathcal{O}_L$. For $\mathfrak{q}|\mathfrak{p}$ the Frobenius elements $\sigma_\mathfrak{q}$ are all conjugate (they form the Frobenius class $\mathrm{Frob}_\mathfrak{p}$), and when L/K is abelian they coincide, in which case we may write $\sigma_\mathfrak{p}$ instead of $\sigma_\mathfrak{q}$ (or use $\mathrm{Frob}_\mathfrak{p} = \{\sigma_\mathfrak{p}\}$ to denote $\sigma_\mathfrak{p}$), and we may write the Artin symbol as

$$\left(\frac{L/K}{\mathfrak{p}}\right) := \sigma_\mathfrak{p}.$$

Now assume L/K is abelian, let $\mathfrak{m}$ be an $\mathcal{O}_K$-ideal divisible by every ramified prime of K, and let $\mathcal{I}_K^\mathfrak{m}$ denote the subgroup of fractional ideals $I \in \mathcal{I}_K$ for which $v_\mathfrak{p}(I) = 0$ for all $\mathfrak{p}|\mathfrak{m}$. The Artin map (Definition 7.21) is the homomorphism

$$\begin{aligned} \psi_{L/K}^\mathfrak{m} \colon \mathcal{I}_K^\mathfrak{m} &\to \mathrm{Gal}(L/K) \\ \prod_{\mathfrak{p}\nmid\mathfrak{m}} \mathfrak{p}^{n_\mathfrak{p}} &\mapsto \prod_{\mathfrak{p}\nmid\mathfrak{m}} \left(\frac{L/K}{\mathfrak{p}}\right)^{n_\mathfrak{p}}. \end{aligned}$$

A key ingredient of class field theory that we will prove in this lecture is surjectivity of the Artin map $\psi_{L/K}^\mathfrak{m}$. This allows us to identify $\mathrm{Gal}(L/K)$ with the quotient $\mathcal{I}_K^\mathfrak{m}/\ker\psi_{L/K}^\mathfrak{m}$.

Every $\mathfrak{p} \in \ker \psi_{L/K}^{\mathfrak{m}}$ is unramified and has the property that the Frobenius elements $\sigma_{\mathfrak{q}}$ are trivial for all $\mathfrak{q}|\mathfrak{p}$, meaning that all the residue field extensions $\mathbb{F}_{\mathfrak{q}}/\mathbb{F}_{\mathfrak{p}}$ are trivial. This implies that $\mathfrak{p}$ *splits completely* in L (it is unramified and primes above it have residue degree one). Conversely, every prime $\mathfrak{p} \in \mathcal{I}_K^{\mathfrak{m}}$ that splits completely in L lies in $\ker \psi_{L/K}^{\mathfrak{m}}$.

Proposition 21.1. *Let $K \subseteq L \subseteq M$ be a tower of finite abelian extension of global fields and let $\mathfrak{m}$ be an $\mathcal{O}_K$-ideal divisible by all primes $\mathfrak{p}$ of K that ramify in M. We have a commutative diagram*

$$\begin{array}{ccc} \mathcal{I}_K^{\mathfrak{m}} & \xrightarrow{\psi_{M/K}^{\mathfrak{m}}} & \mathrm{Gal}(M/K) \\ & \searrow^{\psi_{L/K}^{\mathfrak{m}}} & \downarrow \mathrm{res} \\ & & \mathrm{Gal}(L/K) \end{array}$$

where the vertical map is the homomorphism $\sigma \to \sigma_{|L}$ induced by restriction.

Proof. It suffices to check commutativity at primes $\mathfrak{p} \nmid \mathfrak{m}$, which are necessarily unramified. The proposition then follows from Proposition 7.20. □

21.2 Class field theory for $\mathbb{Q}$

We now specialize to $K = \mathbb{Q}$. The Kronecker-Weber theorem tells us that every abelian extension L/K lies in a cyclotomic field $\mathbb{Q}(\zeta_m)$. Each $\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q})$ is determined by its action on ζ_m, and we have an isomorphism

$$\omega \colon \mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q}) \xrightarrow{\sim} (\mathbb{Z}/m\mathbb{Z})^\times$$

defined by $\sigma(\zeta_m) = \zeta_m^{\omega(\sigma)}$. The primes p that ramify in $\mathbb{Q}(\zeta_m)$ are precisely those that divide m (by Corollary 10.20). For each prime $p \nmid m$ the Frobenius element σ_p is the unique $\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q})$ for which $\sigma(x) \equiv x^p \bmod \mathfrak{q}$ for any (equivalently, all) $\mathfrak{q}|(p)$. Thus $\omega(\sigma_p) = p \bmod m$, and it follows that the Artin map induces an inverse isomorphism $(\mathbb{Z}/m\mathbb{Z})^\times \to \mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q})$: for every integer a coprime to m we have $(a) \in \mathcal{I}_{\mathbb{Q}}^m$ and

$$\omega^{-1}(\bar{a}) = \left(\frac{\mathbb{Q}(\zeta_m)/\mathbb{Q}}{(a)}\right),$$

where $\bar{a} = a \bmod m$. As you showed on Problem Set 4, the surjectivity of the Artin map follows immediately, since a ranges over all integers coprime to m.

Now let L be a subfield of $\mathbb{Q}(\zeta_m)$. We cannot apply ω to $\mathrm{Gal}(L/\mathbb{Q})$, since $\mathrm{Gal}(L/\mathbb{Q})$ is a quotient of $\mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q})$, not a subgroup, but the Artin map $\mathcal{I}_{\mathbb{Q}}^m \to \mathrm{Gal}(L/\mathbb{Q})$ is available; notice that the modulus m works for L as well as $\mathbb{Q}(\zeta_m)$, since any primes that ramify in L also ramify in $\mathbb{Q}(\zeta_m)$ and therefore divide m. By Proposition 21.1, the Artin map factors through the surjective homomorphism $\mathrm{Gal}(\mathbb{Q}(\zeta_m)/\mathbb{Q}) \to \mathrm{Gal}(L/\mathbb{Q})$ induced by restriction and thus induces a surjective homomorphism $(\mathbb{Z}/m\mathbb{Z})^\times \to \mathrm{Gal}(L/\mathbb{Q})$.

To sum up, we can now say the following about abelian extensions of $\mathbb{Q}$:

- **Existence**: for each integer m we have a *ray class field* $\mathbb{Q}(\zeta_m)$: an abelian extension ramified only at $p|m$ with Galois group isomorphic to the *ray class group* $(\mathbb{Z}/m\mathbb{Z})^\times$.
- **Completeness**: every abelian extension of $\mathbb{Q}$ lies in a ray class field $\mathbb{Q}(\zeta_m)$.

- **Reciprocity**: if L is an abelian extension of $\mathbb{Q}$ contained in the ray class field $\mathbb{Q}(\zeta_m)$, the Artin map $\mathcal{I}_{\mathbb{Q}}^m \to \mathrm{Gal}(L/\mathbb{Q})$ induces a surjective homomorphism from the ray class group $(\mathbb{Z}/m\mathbb{Z})^\times$ to $\mathrm{Gal}(L/\mathbb{Q})$, letting us view $\mathrm{Gal}(L/\mathbb{Q})$ as a quotient of $(\mathbb{Z}/m\mathbb{Z})^\times$.

All of these statements will be made more precise; in particular, we will refine the first two statements so that ray class fields are uniquely determined by the modulus m, and we will give an explicit description of the kernel of the Artin map that allows us to identify $\mathrm{Gal}(L/\mathbb{Q})$ with a quotient of $(\mathbb{Z}/m\mathbb{Z})^\times$. But let us first consider how to generalize these statements to number fields other than $\mathbb{Q}$ and define the terms *ray class field*, and *ray class group*. In order to do so, we first need to make the role of the integer m more precise by introducing the notion of a *modulus*.

21.3 Moduli and ray class groups

Recall that for a global field K we use M_K to denote its set of places (equivalence classes of absolute values). We generically denote places by the symbol v, but for finite places, those arising from a discrete valuation associated to a prime $\mathfrak{p}$ of K (a nonzero prime ideal of $\mathcal{O}_K$), we may write $\mathfrak{p}$ in place of v. We write $v|\infty$ to indicate that v is an infinite place (one not arising from a prime of K); recall that when K is a number field all infinite places are archimedean, and they may be real ($K_v \simeq \mathbb{R}$) or complex ($K_v \simeq \mathbb{C}$).

Definition 21.2. Let K be a number field. A *modulus* (or *cycle*) $\mathfrak{m}$ for K is a function $M_K \to \mathbb{Z}_{\geq 0}$ with finite support such that for $v|\infty$ we have $\mathfrak{m}(v) \leq 1$ with $\mathfrak{m}(v) = 0$ unless v is a real place. We view $\mathfrak{m}$ as a formal product $\prod v^{\mathfrak{m}(v)}$ over M_K, which we may factor as

$$\mathfrak{m} = \mathfrak{m}_0\mathfrak{m}_\infty, \qquad \mathfrak{m}_0 := \prod_{\mathfrak{p}\nmid\infty} \mathfrak{p}^{\mathfrak{m}(\mathfrak{p})}, \qquad \mathfrak{m}_\infty := \prod_{v|\infty} v^{\mathfrak{m}(v)},$$

where $\mathfrak{m}_0$ is an $\mathcal{O}_K$-ideal and $\mathfrak{m}_\infty$ represents a subset of the real places of K; we use $\#\mathfrak{m}_\infty$ to denote the number of real places in the support of $\mathfrak{m}$. If $\mathfrak{m}$ and $\mathfrak{n}$ are moduli for K we say that $\mathfrak{m}$ *divides* $\mathfrak{n}$ and write $\mathfrak{m}|\mathfrak{n}$ if $\mathfrak{m}(v) \leq \mathfrak{n}(v)$ for all $v \in M_K$. We define the product modulus $\mathfrak{mn}$ by $\mathfrak{mn}(v) := \mathfrak{m}(v) + \mathfrak{n}(v)$ for $v \nmid \infty$ and $\mathfrak{mn}(v) := \max(\mathfrak{m}(v) + \mathfrak{n}(v), 1)$ for $v \mid \infty$; we also define $\gcd(\mathfrak{m}, \mathfrak{n})(v) := \min(\mathfrak{m}(v), \mathfrak{n}(v)$ and $\mathrm{lcm}(\mathfrak{m}, \mathfrak{n})(v) := \max(\mathfrak{m}(v), \mathfrak{n}(v))$. The zero function is the *trivial modulus*, with $\mathfrak{m}_0 = (1)$ and $\#\mathfrak{m}_\infty = 0$.
We use $\mathcal{I}_K$ to denote the ideal class group of $\mathcal{O}_K$ and define the following notation:[1]

- a fractional ideal $\mathfrak{a} \in \mathcal{I}_K$ is *coprime to* $\mathfrak{m}$ (or *prime to* $\mathfrak{m}$) if $v_\mathfrak{p}(\mathfrak{a}) = 0$ for all $\mathfrak{p}|\mathfrak{m}_0$.
- $\mathcal{I}_K^\mathfrak{m} \subseteq \mathcal{I}_K$ is the subgroup of fractional ideals coprime to $\mathfrak{m}$.
- $K^\mathfrak{m} \subseteq K^\times$ is the subgroup of elements $\alpha \in K^\times$ for which $(\alpha) \in \mathcal{I}_K^\mathfrak{m}$.
- $K^{\mathfrak{m},1} \subseteq K^\mathfrak{m}$ is the subgroup of elements $\alpha \in K^\mathfrak{m}$ with $v_\mathfrak{p}(\alpha - 1) \geq v_\mathfrak{p}(\mathfrak{m}_0)$ for all $\mathfrak{p}|\mathfrak{m}_0$ and $\alpha_v > 0$ for $v|\mathfrak{m}_\infty$ (here α_v is the image of α under $K \hookrightarrow K_v \simeq \mathbb{R}$).
- $\mathcal{R}_K^\mathfrak{m} \subseteq \mathcal{I}_K^\mathfrak{m}$ is the subgroup of principal fractional ideals $(\alpha) \in \mathcal{I}_K^\mathfrak{m}$ with $\alpha \in K^{\mathfrak{m},1}$.

The groups $\mathcal{R}_K^\mathfrak{m}$ are called *rays* or *ray groups*.

[1] This notation varies from author to author; there is no universally accepted notation for these objects (in particular, the modulus $\mathfrak{m}$ may appear as a subscript rather than a superscript). Things will improve when we come to the adelic/idelic formulation of class field theory where there is more consistency.

Definition 21.3. Let $\mathfrak{m}$ be a modulus for a number field K. The *ray class group* for the modulus $\mathfrak{m}$ is the quotient

$$\mathrm{Cl}_K^{\mathfrak{m}} := \mathcal{I}_K^{\mathfrak{m}}/\mathcal{R}_K^{\mathfrak{m}}.$$

A finite abelian extension L/K that is unramified at all places[2] not in the support of $\mathfrak{m}$ for which the kernel of the Artin map $\psi_{L/K}^{\mathfrak{m}}\colon \mathcal{I}_K^{\mathfrak{m}} \to \mathrm{Gal}(L/K)$ is equal to the ray group $\mathcal{R}_K^{\mathfrak{m}}$ is a *ray class field* for the modulus $\mathfrak{m}$.

When $\mathfrak{m}$ is the trivial modulus, the ray class group is the same as the usual class group $\mathrm{Cl}_K := \mathrm{cl}(\mathcal{O}_K)$, but in general the class group Cl_K is a quotient of the ray class group $\mathrm{Cl}_K^{\mathfrak{m}}$ (as we will prove shortly). While not immediately apparent from the definition, we will see that ray class fields are uniquely determined by $\mathfrak{m}$, so it makes sense to speak of *the* ray class field for the modulus $\mathfrak{m}$ (assuming existence).

Remark 21.4. The definitions above make sense for any global field, but in our ideal-theoretic treatment of class field theory we will mostly restrict our attention to number fields. Our adelic/idelic formulation of class field theory will address all global fields.

Remark 21.5. If $\mathfrak{m}(v) = 1$ for every real place v of K then $\mathrm{Cl}_K^{\mathfrak{m}}$ is a *narrow ray class group*. The narrow ray class group with $\mathfrak{m}_0 = (1)$ is the *narrow class group*; the usual class group $\mathrm{Cl}_K = \mathrm{cl}\,\mathcal{O}_K$ is sometimes called the *wide class group* to distinguish the two. Note that the wide class group is a quotient of the narrow class group, thus smaller in general; this terminology can be confusing, but the thing to remember is that narrow equivalence is *stronger* than ordinary equivalence, so there are *more* narrow equivalence classes, in general. Of course for number fields with no real places (imaginary quadratic fields, in particular) there is no distinction.

Example 21.6. For $K = \mathbb{Q}$ with the modulus $\mathfrak{m} = (5)$ we have $K^{\mathfrak{m}} = \{a/b : a, b \not\equiv 0 \bmod 5\}$ and $K^{\mathfrak{m},1} = \{a/b : a \equiv b \not\equiv 0 \bmod 5\}$. Thus

$$\begin{aligned}\mathcal{I}_K^{\mathfrak{m}} &= \{(1), (1/2), (2), (1/3), (2/3), (3/2), (3), (1/4), (3/4), (4/3), (4), (1/6), (6), \ldots\},\\ \mathcal{R}_K^{\mathfrak{m}} &= \{(1), (2/3), (3/2), (1/4), (4), (6), (1/6), (2/7), (7/2), \ldots\}.\end{aligned}$$

You might not have expected $(2/3) \in \mathcal{R}_K^{\mathfrak{m}}$, since $2/3 \notin K^{\mathfrak{m},1}$, but note that $-2/3 \in K^{\mathfrak{m},1}$ and $(-2/3) = (2/3)$. The ray class group is

$$\mathrm{Cl}_K^{\mathfrak{m}} = \mathcal{I}_K^{\mathfrak{m}}/\mathcal{R}_K^{\mathfrak{m}} = \{[(1)], [(2)]\} \simeq (\mathbb{Z}/5\mathbb{Z})^\times/\{\pm 1\},$$

which is isomorphic to the Galois group of the totally real subfield $\mathbb{Q}(\zeta_5)^+$ of $\mathbb{Q}(\zeta_5)$, which is the ray class field for this modulus. If we change the modulus to $\mathfrak{m} = (5)\infty$ we instead get $\mathcal{R}_K^{\mathfrak{m}} = \{(1), (6), (1/6), (2/7), (7/2), \ldots\}$, $\mathrm{Cl}_K^{\mathfrak{m}} \simeq (\mathbb{Z}/5\mathbb{Z})^\times$, and the ray class field is $\mathbb{Q}(\zeta_5)$.

Lemma 21.7. *Let A be a Dedekind domain and let $\mathfrak{a}$ be an A-ideal. Every ideal class in* $\mathrm{cl}(A)$ *contains an A-ideal coprime to $\mathfrak{a}$.*

Proof. Let I be a nonzero fractional ideal of A. For each prime $\mathfrak{p}|\mathfrak{a}$ we can pick $\pi_{\mathfrak{p}} \in \mathfrak{p}$ such that $v_{\mathfrak{q}}(\pi_{\mathfrak{p}}) = v_{\mathfrak{q}}(\mathfrak{p})$ for all $\mathfrak{q}|\mathfrak{a}$, by Corollary 3.21. If we then put $\alpha := \prod_{\mathfrak{p}|\mathfrak{a}} \pi_{\mathfrak{p}}^{-v_{\mathfrak{p}}(I)}$, then $v_{\mathfrak{p}}(\alpha I) = 0$ for all $\mathfrak{p}|\mathfrak{a}$; thus αI is coprime to $\mathfrak{a}$ and $[\alpha I] = [I]$.

[2] Archimedean places of K are unramified in L except for real places v with a complex place w of L above them. But if L is unramified at all $\mathfrak{p} \nmid \mathfrak{m}_0$ (necessary for $\psi_{L/K}^{\mathfrak{m}}$ to be defined), and $\ker \psi_{L/K}^{\mathfrak{m}} = \mathcal{R}_K^{\mathfrak{m}}$, then L will necessarily be unramified at all infinite places $v \nmid \mathfrak{m}_\infty$; so in the definition of a ray class field it is enough for L to be unramified away from $\mathfrak{m}_0$.

Now let S be the finite set of primes $\mathfrak{p}$ for which $v_\mathfrak{p}(\alpha I) < 0$ and pick $\pi_\mathfrak{p} \in \mathfrak{p}$ such that $v_\mathfrak{q}(\pi_\mathfrak{p}) = v_\mathfrak{q}(\mathfrak{p})$ for all $\mathfrak{q} \in S$ and $\mathfrak{q}|\mathfrak{a}$ (again using Corollary 3.21). If we now put $a \coloneqq \prod_{\mathfrak{p}\in S} \pi_\mathfrak{p}^{-v_\mathfrak{p}(\alpha I)} \in A$, then $v_\mathfrak{p}(a\alpha I) \geq 0$ for all $\mathfrak{p}$ and $v_\mathfrak{p}(a\alpha I) = 0$ for all $\mathfrak{p}|\mathfrak{a}$. Thus $a\alpha I$ is an A-ideal coprime to $\mathfrak{a}$ and $[a\alpha I] = [I]$. □

Theorem 21.8. *Let $\mathfrak{m}$ be a modulus for a number field K. We have an exact sequence*

$$1 \longrightarrow \mathcal{O}_K^\times \cap K^{\mathfrak{m},1} \longrightarrow \mathcal{O}_K^\times \longrightarrow K^\mathfrak{m}/K^{\mathfrak{m},1} \longrightarrow \mathrm{Cl}_K^\mathfrak{m} \longrightarrow \mathrm{Cl}_K \longrightarrow 1$$

and a canonical isomorphism

$$K^\mathfrak{m}/K^{\mathfrak{m},1} \simeq \{\pm 1\}^{\#\mathfrak{m}_\infty} \times (\mathcal{O}_K/\mathfrak{m}_0)^\times.$$

Proof. Let us consider the composition of the maps $K^{\mathfrak{m},1} \subseteq K^\mathfrak{m}$ and $\alpha \mapsto (\alpha)$:

$$K^{\mathfrak{m},1} \xrightarrow{f} K^\mathfrak{m} \xrightarrow{g} \mathcal{I}_K^\mathfrak{m}.$$

The kernel of f is trivial, the kernel of $g \circ f$ is $\mathcal{O}_K^\times \cap K^{\mathfrak{m},1}$ (since $(\alpha) = (1) \iff \alpha \in \mathcal{O}_K^\times$), the kernel of g is $\mathcal{O}_K^\times$, the cokernel of f is $K^\mathfrak{m}/K^{\mathfrak{m},1}$, the cokernel of $g \circ f$ is $\mathrm{Cl}_K^\mathfrak{m} = \mathcal{I}_K^\mathfrak{m}/\mathcal{R}_K^\mathfrak{m}$ (by definition), and the cokernel of g is Cl_K (by Lemma 21.7). Applying the snake lemma (see [2, Lemma 5.13], for example) to the following commutative diagram with exact rows

$$\begin{array}{ccccccccc} 1 & \longrightarrow & K^{\mathfrak{m},1} & \xhookrightarrow{f} & K^\mathfrak{m} & \longrightarrow & K^\mathfrak{m}/K^{\mathfrak{m},1} & \longrightarrow & 1 \\ & & \downarrow{\scriptstyle g\circ f} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle \pi} & & \\ 1 & \longrightarrow & \mathcal{I}_K^\mathfrak{m} & \xrightarrow{\sim} & \mathcal{I}_K^\mathfrak{m} & \longrightarrow & 1 & & \end{array}$$

yields the exact sequence $\ker g \circ f \to \ker g \to \ker \pi \to \operatorname{coker} g \circ f \to \operatorname{coker} g \to \operatorname{coker} \pi$:

$$1 \longrightarrow \mathcal{O}_K^\times \cap K^{\mathfrak{m},1} \longrightarrow \mathcal{O}_K^\times \longrightarrow K^\mathfrak{m}/K^{\mathfrak{m},1} \longrightarrow \mathrm{Cl}_K^\mathfrak{m} \longrightarrow \mathrm{Cl}_K \longrightarrow 1,$$

where the initial 1 follows from the fact that f is injective (and $\ker \pi = \operatorname{coker} f$).

We can write each $\alpha \in K^\mathfrak{m}$ as $\alpha = a/b$ with $a, b \in \mathcal{O}_K$ such that (a) and (b) are coprime to $\mathfrak{m}_0$ and to each other. The ideals (a) and (b) are uniquely determined by α, since $a/b = a'/b' \Rightarrow ab' = a'b \Rightarrow (a)(b') = (a')(b)$, and since (a) and (b) are coprime we must have $(a) = (a')$ and $(b) = (b')$ (by unique factorization of ideals).

We now define the homomorphism

$$\begin{aligned} \varphi\colon K^\mathfrak{m} &\to \left(\prod_{v|\mathfrak{m}_\infty} \{\pm 1\}\right) \times (\mathcal{O}_K/\mathfrak{m}_0)^\times \\ \alpha &\mapsto \left(\prod_{v|\mathfrak{m}_\infty} \mathrm{sgn}(\alpha_v)\right) \times (\bar\alpha), \end{aligned}$$

where $\bar\alpha = \bar a \bar b^{-1} \in (\mathcal{O}_K/\mathfrak{m}_0)^\times$ (here $\bar a, \bar b$ are the images of $a, b \in \mathcal{O}_K$ in $\mathcal{O}_K/\mathfrak{m}_0$, and they both lie in $(\mathcal{O}_K/\mathfrak{m}_0)^\times$ because (a) and (b) are coprime to $\mathfrak{m}_0$). The ring $(\mathcal{O}_K/\mathfrak{m}_0)^\times$ is isomorphic to $\prod_{\mathfrak{p}|\mathfrak{m}_0} (\mathcal{O}_K/\mathfrak{p}^{\mathfrak{m}(\mathfrak{p})})^\times$, by the Chinese remainder theorem, and weak approximation (Theorem 8.5) implies that φ is surjective. The kernel of φ is clearly $K^{\mathfrak{m},1}$, thus φ induces an isomorphism $K^\mathfrak{m}/K^{\mathfrak{m},1} \simeq \{\pm 1\}^{\#\mathfrak{m}_\infty} \times (\mathcal{O}_K/\mathfrak{m}_0)^\times$. This isomorphism is canonical, because $\bar\alpha$ depends only on the uniquely determined ideals (a) and (b) (if we replace a with $a' = au$ for some $u \in \mathcal{O}_K^\times$ we must replace b with $b' = bu$). □

Corollary 21.9. *Let K be a number field and let $\mathfrak{m}$ be a modulus for K. The ray class group $\mathrm{Cl}_K^{\mathfrak{m}}$ is a finite abelian group whose cardinality $h_K^{\mathfrak{m}} := \#\mathrm{Cl}_K^{\mathfrak{m}}$ is given by*

$$h_K^{\mathfrak{m}} = \frac{\phi(\mathfrak{m}) h_K}{[\mathcal{O}_K^\times : \mathcal{O}_K^\times \cap K^{\mathfrak{m},1}]},$$

where $h_K := \#\mathrm{Cl}_K$ and $\phi(\mathfrak{m}) := \#(K^{\mathfrak{m}}/K^{\mathfrak{m},1}) = \phi(\mathfrak{m}_\infty)\phi(\mathfrak{m}_0)$, with

$$\phi(\mathfrak{m}_\infty) = 2^{\#\mathfrak{m}_\infty}, \qquad \phi(\mathfrak{m}_0) = \#(\mathcal{O}_K/\mathfrak{m}_0)^\times = \mathrm{N}(\mathfrak{m}_0) \prod_{\mathfrak{p}|\mathfrak{m}_0} (1 - \mathrm{N}(\mathfrak{p})^{-1}).$$

In particular, h_K divides $h_K^{\mathfrak{m}}$ and $h_K^{\mathfrak{m}}$ divides $h_K \phi(\mathfrak{m})$.

Proof. The exact sequence implies $\phi(\mathfrak{m})/[\mathcal{O}_K^\times : \mathcal{O}_K^\times \cap K^{\mathfrak{m},1}] = h_K^{\mathfrak{m}}/h_K$, and that both sides of this equality are integers. □

Computing the ray class number $h_K^{\mathfrak{m}}$ is not a trivial problem, but there are algorithms for doing so; see [1], which considers this problem in detail.

21.4 Polar density

We now want to prove the surjectivity of the Artin map for finite abelian extensions L/K of number fields (as noted in §21.2, we already know this for $K = \mathbb{Q}$). In order to do so we first introduce a new way to measure the density of a set of primes that is defined in terms of a generalization of the Dedekind zeta function. Throughout this section and the next, all number fields are assumed to lie in some fixed algebraic closure of $\mathbb{Q}$.

Definition 21.10. Let K be a number field and let S be a set of primes of K. The *partial Dedekind zeta function* associated to S is the complex function

$$\zeta_{K,S}(s) := \prod_{\mathfrak{p} \in S} (1 - \mathrm{N}(\mathfrak{p})^{-s})^{-1},$$

which converges to a holomorphic function on $\mathrm{Re}(s) > 1$ (by the same argument we used for $\zeta_K(s)$ in Lecture 18).

If S is finite then $\zeta_{K,S}(s)$ is certainly holomorphic (and nonzero) on a neighborhood of 1. If S contains all but finitely many primes of K then it differs from $\zeta_K(s)$ by a holomorphic factor and therefore extends to a meromorphic function with a simple pole at $s = 1$, by Theorem 19.12.

Between these two extremes the function $\zeta_{K,S}(s)$ may or may not extend to a function that is meromorphic on a neighborhood of 1, but if it does, or more generally, if some power of it does, then we can use the order of the pole at 1 (or the absence of a pole) to measure the density of S.

Definition 21.11. If for some integer $n \geq 1$ the function $\zeta_{K,S}^n$ extends to a meromorphic function on a neighborhood of 1, the *polar density* of S is defined by

$$\rho(S) := \frac{m}{n}, \qquad m = -\mathrm{ord}_{s=1} \zeta_{K,S}^n(s)$$

(so m is the order of the pole at $s = 1$, if one is present). Note that if $\zeta_{K,S}^{n_1}$ and $\zeta_{K,s}^{n_2}$ both extend to a meromorphic function on a neighborhood of 1 then we necessarily have

$$n_2 \mathrm{ord}_{s=1} \zeta_{K,S}^{n_1}(s) = \mathrm{ord}_{s=1} \zeta_{K,S}^{n_1 n_2} = n_1 \mathrm{ord}_{s=1} \zeta_{K,S}^{n_2}(s),$$

which implies that $\rho(S)$ does not depend on the choice of n. We will show below that (whenever it is defined) $\rho(S)$ is a rational number in the interval $[0,1]$.

In Lecture 17 we encountered two other notions of density, the *Dirichlet density*

$$d(S) := \lim_{s\to 1^+} \frac{\sum_{\mathfrak{p}\in S} \mathrm{N}(\mathfrak{p})^{-s}}{\sum_{\mathfrak{p}} \mathrm{N}(\mathfrak{p})^{-s}} = \lim_{s\to 1^+} \frac{\sum_{\mathfrak{p}\in S} \mathrm{N}(\mathfrak{p})^{-s}}{\log \frac{1}{s-1}},$$

(the equality of the two expressions for $d(S)$ follows from the fact that $\zeta_K(s)$ has a simple pole at $s=1$, see Problem Set 9), and the *natural density*

$$\delta(S) := \lim_{x\to\infty} \frac{\#\{\mathfrak{p}\in S : \mathrm{N}(\mathfrak{p}) \le x\}}{\#\{\mathfrak{p} : \mathrm{N}(\mathfrak{p}) \le x\}}.$$

On Problem Set 9 you proved that if S has a natural density then it has a Dirichlet density and the two coincide. We now show that the same is true of the polar density.

Proposition 21.12. *Let S be a set of primes of a number field K. If S has a polar density then it has a Dirichlet density and the two are equal. In particular, $\rho(S) \in [0,1]$ whenever it is defined.*

Proof. Suppose S has polar density $\rho(S) = m/n$. By taking the Laurent series expansion of $\zeta_{K,S}^n(s)$ at $s=1$ and factoring out the leading nonzero term we can write

$$\zeta_{K,S}(s)^n = \frac{a}{(s-1)^m}\left(1 + \sum_{r\ge 1} a_r (s-1)^r\right),$$

for some $a \in \mathbb{C}^\times$. We must have $a \in \mathbb{R}_{>0}$, since $\zeta_{K,S}(s) \in \mathbb{R}_{>0}$ for $s \in \mathbb{R}_{>1}$ and therefore $\lim_{s\to 1^+}(s-1)^m \zeta_{K,S}(s)^n$ is a positive real number. Taking logs of both sides yields

$$n \sum_{\mathfrak{p}\in S} \mathrm{N}(\mathfrak{p})^{-s} \sim m \log \frac{1}{s-1} \qquad (\text{as } s \to 1^+),$$

which implies that S has Dirichlet density $d(S) = m/n$ (note that $\log(a) = O(1)$ plays no role, since $-m\log(s-1) \to \infty$ as $s \to 1^+$). □

Corollary 21.13. *Let S be a set of primes of a number field K. If S has both a polar density and a natural density then the two coincide.*

We should note that not every set of primes with a natural density has a polar density, since the later is always a rational number while the former need not be.

Recall that a degree-1 prime in a number field K is a prime with residue field degree 1 over $\mathbb{Q}$, equivalently, a prime $\mathfrak{p}$ whose absolute norm $\mathrm{N}(\mathfrak{p}) = [\mathcal{O}_K : \mathfrak{p}] = \#\mathbb{F}_\mathfrak{p}$ is prime.

Proposition 21.14. *Let S and T denote sets of primes in a number field K, let $\mathcal{P}$ be the set of all primes of K, and let $\mathcal{P}_1$ be the set of degree-1 primes of K. The following hold:*

(a) *If S is finite then $\rho(S) = 0$; if $\mathcal{P} - S$ is finite then $\rho(S) = 1$.*

(b) *If $S \subseteq T$ both have polar densities, then $\rho(S) \le \rho(T)$.*

(c) *If two sets S and T have finite intersection and any two of the sets S, T, and $S \cup T$ have polar densities then so does the third and $\rho(S \cup T) = \rho(S) + \rho(T)$.*

(d) *We have $\rho(\mathcal{P}_1) = 1$, and $\rho(S \cap \mathcal{P}_1) = \rho(S)$ whenever S has a polar density.*

Proof. We first note that for any finite set S, the function $\zeta_{K,S}(s)$ is a finite product of nonvanishing entire functions and therefore holomorphic and nonzero everywhere (including at $s = 1$). If the symmetric difference of S and T is finite, then $\zeta_{K,S}(s)f(s) = \zeta_{K,T}(s)g(s)$ for some nonvanishing functions $f(s)$ and $g(s)$ holomorphic on $\mathbb{C}$. Thus if S and T differ by a finite set, then $\rho(S) = \rho(T)$ whenever either set has a polar density

Part (a) follows, since $\rho(\emptyset) = 0$ and $\rho(\mathcal{P}) = 1$ (note that $\zeta_{K,\mathcal{P}}(s) = \zeta_K(s)$, and $\operatorname{ord}_{s=1}\zeta_K(s) = -1$, by Theorem 19.12).

Part (b) follows from the analogous statement for Dirichlet density proved on Problem Set 9.

For (c) we may assume S and T are disjoint (by the argument above), in which case $\zeta_{K,S\cup T}(s)^n = \zeta_{K,S}(s)^n\zeta_{K,T}(s)^n$ for all $n \geq 1$, and the claim follows.

For (d), let $\mathcal{P}_2 \coloneqq \mathcal{P} - \mathcal{P}_1$ so that $\mathcal{P} = \mathcal{P}_1 \sqcup \mathcal{P}_2$. For each rational prime p there are at most $n \coloneqq [K : \mathbb{Q}]$ (in fact $n/2$) primes $\mathfrak{p}|p$ in $\mathcal{P}_2$, each of which has absolute norm $\mathrm{N}(\mathfrak{p}) \geq p^2$. It follows by comparison with $\zeta(2s)^n$ that the product defining $\zeta_{K,\mathcal{P}_2}(s)$ converges absolutely to a holomorphic function on $\mathrm{Re}(s) > 1/2$ and is therefore holomorphic (and nonvanishing, since it is an Euler product) on a neighborhood of 1; thus $\rho(\mathcal{P}_2) = 0$ and $\rho(\mathcal{P}_1) = 1$. We therefore have $\rho(S \cap \mathcal{P}_2) = 0$, so $\rho(S) = \rho(S \cap \mathcal{P}_1)$ whenever $\rho(S)$ exists, by (c). □

For a Galois extension of number fields L/K, let $\mathrm{Spl}(L/K)$ denote the set of primes of K that split completely in L. When K is clear from context we may just write $\mathrm{Spl}(L)$.

Theorem 21.15. *Let L/K be a Galois extension of number fields of degree n. Then*

$$\rho(\mathrm{Spl}(L)) = 1/n.$$

Proof. Let S be the set of degree-1 primes of K that split completely in L; it suffices to show $\rho(S) = 1/n$, by Proposition 21.14. Recall that $\mathfrak{p}$ splits completely in L if and only if both the ramification index $e_\mathfrak{p}$ and residue field degree $f_\mathfrak{p}$ are equal to 1. Let T be the set of primes $\mathfrak{q}$ of L that lie above some $\mathfrak{p} \in S$. For each $\mathfrak{q} \in T$ lying above $\mathfrak{p} \in S$ we have $N_{L/K}(\mathfrak{q}) = \mathfrak{p}^{f_\mathfrak{p}} = \mathfrak{p}$, so $\mathrm{N}(\mathfrak{q}) = \mathrm{N}(N_{L/K}(\mathfrak{q})) = \mathrm{N}(\mathfrak{p})$, thus $\mathfrak{q}$ is a degree-1 prime, since $\mathfrak{p}$ is.

On the other hand, if $\mathfrak{q}$ is any unramified degree-1 prime of L and $\mathfrak{p} = \mathfrak{q} \cap \mathcal{O}_K$, then $\mathrm{N}(\mathfrak{q}) = \mathrm{N}(N_{L/K}(\mathfrak{q})) = \mathrm{N}(\mathfrak{p}^{f_\mathfrak{p}})$ is prime, so we must have $f_\mathfrak{p} = 1$, and $e_\mathfrak{p} = 1$ since $\mathfrak{q}$ is unramified, which implies that $\mathfrak{p}$ is a degree-1 prime that splits completely in L and is thus an element of S. Only finitely many primes ramify, so all but finitely many of the degree-1 primes in L lie in T, thus $\rho(T) = 1$, by Proposition 21.14. Each $\mathfrak{p} \in S$ has exactly n primes $\mathfrak{q} \in T$ lying above it (since $\mathfrak{p}$ splits completely), and we have

$$\zeta_{L,T}(s) = \prod_{\mathfrak{q}\in T}(1 - \mathrm{N}(\mathfrak{q})^{-s})^{-1} = \prod_{\mathfrak{q}\in T}(1 - \mathrm{N}(N_{L/K}(\mathfrak{q}))^{-s})^{-1} = \prod_{\mathfrak{p}\in S}(1 - \mathrm{N}(\mathfrak{p})^{-s})^{-n} = \zeta_{K,S}(s)^n.$$

It follows that $\rho(S) = \frac{1}{n}\rho(T) = \frac{1}{n}$ as desired. □

Corollary 21.16. *If L/K is a finite extension of number fields with Galois closure M/K of degree n, then $\rho(\mathrm{Spl}(L)) = \rho(\mathrm{Spl}(M)) = 1/n$.*

Proof. A prime $\mathfrak{p}$ of K splits completely in L if and only if it splits completely in all the conjugates of L in M; the Galois closure M is the compositum of the conjugates of L, so $\mathfrak{p}$ splits completely in L if and only if it splits completely in M. □

Corollary 21.17. *Let L/K be a Galois extension of number fields with Galois group $G := \mathrm{Gal}(L/K)$ and let H be a normal subgroup of G. The set S of primes for which $\mathrm{Frob}_\mathfrak{p} \subseteq H$ has polar density $\rho(S) = \#H/\#G$.*

Proof. Let $F = L^H$; then F/K is Galois (since H is normal) and $\mathrm{Gal}(F/K) \simeq G/H$. For each unramified prime $\mathfrak{p}$ of K, the Frobenius class $\mathrm{Frob}_\mathfrak{p}$ lies in H if and only if every $\sigma_\mathfrak{q} \in \mathrm{Frob}_\mathfrak{p}$ acts trivially on $L^H = F$, which occurs if and only if $\mathfrak{p}$ splits completely in F. By Theorem 21.15, the density of this set of primes is $1/[F:K] = \#H/\#G$. □

If S and T are sets of primes whose symmetric difference is finite, then either $\rho(S) = \rho(T)$ or neither set has a polar density. Let us write $S \sim T$ to indicate that two sets of primes have finite symmetric difference (this is clearly an equivalence relation), and partially order sets of primes by defining $S \precsim T \Leftrightarrow S \sim S \cap T$ (in other words, $S - T$ is finite). If S and T have polar densities, then $S \precsim T$ implies $\rho(S) \le \rho(T)$, by Proposition 21.14.

Theorem 21.18. *If L/K and M/K are two Galois extensions of number fields then*

$$L \subseteq M \iff \mathrm{Spl}(M) \precsim \mathrm{Spl}(L) \iff \mathrm{Spl}(M) \subseteq \mathrm{Spl}(L),$$
$$L = M \iff \mathrm{Spl}(M) \sim \mathrm{Spl}(L) \iff \mathrm{Spl}(M) = \mathrm{Spl}(L),$$

and the map $L \mapsto \mathrm{Spl}(L)$ is an injection from the set of finite Galois extensions of K (inside some fixed algebraic closure) to sets of primes of K that have a positive polar density.

Proof. The implications $L \subseteq M \Rightarrow \mathrm{Spl}(M) \subseteq \mathrm{Spl}(L) \Rightarrow \mathrm{Spl}(M) \precsim \mathrm{Spl}(L)$ are clear, so it suffices to show that $\mathrm{Spl}(M) \precsim \mathrm{Spl}(L) \Rightarrow L \subseteq M$.

A prime $\mathfrak{p}$ of K splits completely in the compositum LM if and only if it splits completely in both L and M: the forward implication is clear and for the reverse, note that if $\mathfrak{p}$ splits completely in both L and M then it certainly splits completely in $L \cap M$, so we may assume $K = L \cap M$; we then have $\mathrm{Gal}(LM/K) \simeq \mathrm{Gal}(L/K) \times \mathrm{Gal}(M/K)$, and if the decomposition subgroups of all primes above $\mathfrak{p}$ are trivial in both $\mathrm{Gal}(L/K)$ and $\mathrm{Gal}(M/K)$ then the same applies in $\mathrm{Gal}(LM/K)$. Thus $\mathrm{Spl}(LM) = \mathrm{Spl}(L) \cap \mathrm{Spl}(M)$.

It follows that $\mathrm{Spl}(M) \precsim \mathrm{Spl}(L) \Rightarrow \mathrm{Spl}(LM) \sim \mathrm{Spl}(M)$. By Theorem 21.15, we have $\rho(\mathrm{Spl}(M)) = 1/[M:K]$ and $\rho(\mathrm{Spl}(LM)) = 1/[LM:K]$, thus $\mathrm{Spl}(LM) \sim \mathrm{Spl}(M)$ implies

$$[LM:K] = \rho(\mathrm{Spl}(LM))^{-1} = \rho(\mathrm{Spl}(M))^{-1} = [M:K],$$

in which case $LM = M$ and $L \subseteq M$. This proves $\mathrm{Spl}(M) \precsim \mathrm{Spl}(L) \Rightarrow L \subseteq M$, so the three conditions in the first line of biconditionals are all equivalent, and this immediately implies the second line of biconditionals. The last statement of the theorem is clear, since $\mathrm{Spl}(L)$ has positive polar density, by Theorem 21.15. □

21.5 Ray class fields and Artin reciprocity

As a special case of Corollary 21.16, if F/K is a finite extension of number fields in which all but finitely many primes split completely, then $[F:K] = 1$ and therefore $F = K$. We will use this fact to prove that the Artin map is surjective.

Theorem 21.19. *Let L/K be an abelian extension of number fields and $\mathfrak{m}$ a modulus divisible by all ramified primes. Then the Artin map $\psi_{L/K}^\mathfrak{m} \colon \mathcal{I}_K^\mathfrak{m} \to \mathrm{Gal}(L/K)$ is surjective.*

Proof. Let $H \subseteq \mathrm{Gal}(L/K)$ be the image of $\psi_{L/K}^{\mathfrak{m}}$ and let $F = L^H$ be its fixed field, which we note is a Galois extension of K, since H is normal (because $\mathrm{Gal}(L/K)$ is abelian). For each prime $\mathfrak{p} \in \mathcal{I}_K^{\mathfrak{m}}$ the automorphism $\psi_{L/K}^{\mathfrak{m}}(\mathfrak{p}) \in H$ acts trivially on $F = L^H$, therefore $\mathfrak{p}$ splits completely in F. The group $\mathcal{I}_K^{\mathfrak{m}}$ contains all but finitely many primes $\mathfrak{p}$ of K, so the polar density of the set of primes of K that split completely in F is 1. Thus $[F : K] = 1$ and $H = \mathrm{Gal}(L/K)$, by Corollary 21.16. □

We now show that the kernel of the Artin map $\psi_{L/K}^{\mathfrak{m}}$ uniquely determines the field L.

Theorem 21.20. *Let $\mathfrak{m}$ be a modulus for a number field K and let L and M be finite abelian extensions of K unramified at all primes not in the support of $\mathfrak{m}$. If $\ker \psi_{L/K}^{\mathfrak{m}} = \ker \psi_{M/K}^{\mathfrak{m}}$ then $L = M$. In particular, ray class fields are unique whenever they exist.*

Proof. Let S be the set of primes of K that do not divide $\mathfrak{m}$. Each prime $\mathfrak{p}$ in S is unramified in both L and M, and $\mathfrak{p}$ splits completely in L (resp. M) if and only if it lies in the kernel of $\psi_{L/K}^{\mathfrak{m}}$ (resp. $\psi_{M/K}^{\mathfrak{m}}$). If $\ker \psi_{L/K}^{\mathfrak{m}} = \ker \psi_{M/K}^{\mathfrak{m}}$ then

$$\mathrm{Spl}(L) \sim (S \cap \ker \psi_{L/K}^{\mathfrak{m}}) = (S \cap \ker \psi_{M/K}^{\mathfrak{m}}) \sim \mathrm{Spl}(M),$$

and therefore $L = M$, by Theorem 21.18. □

Theorem 21.19 implies that we have an exact sequence

$$1 \to \ker \psi_{L/K}^{\mathfrak{m}} \to \mathcal{I}_K^{\mathfrak{m}} \to \mathrm{Gal}(L/K) \to 1.$$

One of the key results of class field theory is that for a suitable choice of the modulus $\mathfrak{m}$, we have $\mathcal{R}_K^{\mathfrak{m}} \subseteq \ker \psi_{L/K}^{\mathfrak{m}}$. This implies that the Artin map induces an isomorphism between $\mathrm{Gal}(L/K)$ and a quotient of the ray class group $\mathrm{Cl}_K^{\mathfrak{m}} = \mathcal{I}_K^{\mathfrak{m}}/\mathcal{R}_K^{\mathfrak{m}}$. When L is the ray class field for the modulus $\mathfrak{m}$, the Artin map allows us to relate subfields of L to quotients of the ray class group $\mathrm{Cl}_K^{\mathfrak{m}} \simeq \mathrm{Gal}(L/K)$ in a way that we will make more precise in the next lecture; this is known as *Artin reciprocity*.

References

[1] Henri Cohen, *Advanced topics in computational number theory*, Springer, 2000.

[2] Allen Altman and Steven Kleiman, *A term of commutative algebra*, Worldwide Center of Mathematics, 2013.

22 The main theorems of global class field theory

In this lecture we refine the correspondence between quotients of ray class groups and subfields of ray class fields given by the Artin map so that we can more precisely state the main theorems of global class field theory (for number fields) in their ideal-theoretic form. Let us first recall the notational setup.

We have a number field K and a modulus $\mathfrak{m}\colon M_K \to \mathbb{Z}_{\geq 0}$ that we view as a formal product over the places of K; we may write $\mathfrak{m} = \mathfrak{m}_0\mathfrak{m}_\infty$, where $\mathfrak{m}_0 := \prod \mathfrak{p}^{\mathfrak{m}(\mathfrak{p})}$ is a product over primes (finite places) of K and $\mathfrak{m}_\infty := \prod_{v|\infty} v^{\mathfrak{m}(v)}$ defines a subset of the real places of K (recall that for $v|\infty$ we have $\mathfrak{m}(v) \leq 1$ with $\mathfrak{m}(v) = 0$ if v is not real). The moduli for K are partially ordered by the divisibility relation $\mathfrak{m}|\mathfrak{n}$, which holds if and only if $\mathfrak{m}(v) \leq \mathfrak{n}(v)$ for all $v \in M_K$. We then define

- $\mathcal{I}_K^{\mathfrak{m}} \subseteq \mathcal{I}_K$, the subgroup of fractional ideals prime to $\mathfrak{m}$ (equivalently, $\mathfrak{m}_0$);
- $K^{\mathfrak{m}} \subseteq K^\times$, the subgroup of $\alpha \in K^\times$ for which $(\alpha) \in \mathcal{I}_K^{\mathfrak{m}}$;
- $K^{\mathfrak{m},1} \subseteq K^{\mathfrak{m}}$, the subgroup of $\alpha \in K^{\mathfrak{m}}$ for which $v_{\mathfrak{p}}(\alpha - 1) \geq v_{\mathfrak{p}}(\mathfrak{m}_0)$ for $\mathfrak{p}|\mathfrak{m}_0$ and $\alpha_v > 0$ for $v|\mathfrak{m}_\infty$ (here $\alpha_v \in \mathbb{R}$ is the image of α under the real-embedding v);
- $\mathcal{R}_K^{\mathfrak{m}} \subseteq \mathcal{I}_K^{\mathfrak{m}}$ the subgroup of ideals $(\alpha) \in \mathcal{I}_K^{\mathfrak{m}}$ with $\alpha \in K^{\mathfrak{m},1}$ (the *ray group* for $\mathfrak{m}$);
- $\mathrm{Cl}_K^{\mathfrak{m}} := \mathcal{I}_K^{\mathfrak{m}}/\mathcal{R}_K^{\mathfrak{m}}$ (the *ray class group* for $\mathfrak{m}$);
- $\mathrm{Spl}(L) := \mathrm{Spl}(L/K)$, the set of primes of K that split completely in an extension L;
- $\psi_{L/K}^{\mathfrak{m}}\colon \mathcal{I}_K^{\mathfrak{m}} \to \mathrm{Gal}(L/K)$, Artin map of an abelian extension L/K unramified at $\mathfrak{p} \nmid \mathfrak{m}$.

In the previous lecture we defined the *ray class field* of K for the modulus $\mathfrak{m}$ as a finite abelian extension L/K unramified at all $\mathfrak{p} \nmid \mathfrak{m}$ such that the kernel of the Artin map $\psi_{L/K}^{\mathfrak{m}}$ is equal to the the ray group $\mathcal{R}_K^{\mathfrak{m}}$. We did not prove that such fields exist, but we did prove that there is at most one of them; see Theorem 21.20. Let $K(\mathfrak{m})$ denote this field.

Assuming the ray class field $K(\mathfrak{m})$ exists, it follows from the surjectivity of the Artin map $\psi_{K(\mathfrak{m})/K}^{\mathfrak{m}}\colon \mathcal{I}_K^{\mathfrak{m}} \to \mathrm{Gal}(K(\mathfrak{m})/K)$ proved in Theorem 21.19 that we have a canonical isomorphism

$$\mathrm{Cl}_K^{\mathfrak{m}} = \mathcal{I}_K^{\mathfrak{m}}/\mathcal{R}_K^{\mathfrak{m}} \simeq \mathrm{Gal}(K(\mathfrak{m})/K)$$

between the ray class group and the Galois group of the ray class field. More generally, if L is any intermediate field between K and $K(\mathfrak{m})$, the kernel of the Artin map is a subgroup $\mathcal{C} \subseteq \mathcal{I}_K^{\mathfrak{m}}$ that contains the ray group

$$\mathcal{R}_K^{\mathfrak{m}} \subseteq \mathcal{C} \subseteq \mathcal{I}_K^{\mathfrak{m}},$$

and we have an isomorphism

$$\mathcal{I}_K^{\mathfrak{m}}/\mathcal{C} \simeq \mathrm{Cl}_K^{\mathfrak{m}}/\overline{\mathcal{C}} \simeq \mathrm{Gal}(L/K)$$

where $\overline{\mathcal{C}}$ denotes the image of $\mathcal{C}$ in $\mathrm{Cl}_K^{\mathfrak{m}} = \mathcal{I}_K^{\mathfrak{m}}/\mathcal{R}_K^{\mathfrak{m}}$ under the quotient map.

Thus if L is a subfield of $K(\mathfrak{m})$ then $\ker \psi_{L/K}^{\mathfrak{m}}$ is a subgroup of $\mathcal{I}_K^{\mathfrak{m}}$ containing $\mathcal{R}_K^{\mathfrak{m}}$ (a *congruence subgroup*, as defined below). To prove that a given abelian extension L/K lies in a ray class field, it is enough to show that there exists a modulus $\mathfrak{m}$ for K such that $\mathcal{R}_K^{\mathfrak{m}} \subseteq \ker \psi_{L/K}^{\mathfrak{m}}$, since we then have $\mathrm{Spl}(K(\mathfrak{m})) \precsim \mathrm{Spl}(L)$ and $L \subseteq K(\mathfrak{m})$, by Theorem 21.18.

In this lecture we want to better understand the structure of congruence subgroups, and to specify a minimal modulus $\mathfrak{m}$ for which we should expect a given finite abelian extension L/K to lie in a subfield of the ray class field $K(\mathfrak{m})$; this minimal modulus is known as the *conductor* of the extension. So far we have not addressed this question even for $K = \mathbb{Q}$ (but see Problem Set 10); our proof of the Kronecker-Weber theorem showed that every abelian extension lies in some cyclotomic field $\mathbb{Q}(\zeta_m)$, but we made no attempt to determine such an integer m (or more precisely, a modulus $\mathfrak{m}$ of the form $\mathfrak{m} = (m)\infty$ or $\mathfrak{m} = (m)$).

22.1 Congruence subgroups

Our presentation of congruence subgroups in this section follows [1, 3.3], but our notation differs slightly.

Definition 22.1. Let K be a number field and let $\mathfrak{m}$ be a modulus for K. A *congruence subgroup* for the modulus $\mathfrak{m}$ is a subgroup $\mathcal{C}$ of $\mathcal{I}_K^{\mathfrak{m}}$ that contains $\mathcal{R}_K^{\mathfrak{m}}$. We use $\overline{\mathcal{C}}$ to denote the image of $\mathcal{C}$ in $\mathcal{I}_K^{\mathfrak{m}}/\mathcal{R}_K^{\mathfrak{m}} = \mathrm{Cl}_K^{\mathfrak{m}}$ under the quotient map.

As explained above, congruence subgroups are precisely the groups we expect to arise as the kernel of an Artin map $\psi_{L/K}^{\mathfrak{m}} \colon \mathcal{I}_K^{\mathfrak{m}} \to \mathrm{Gal}(L/K)$ associated to a finite abelian extension L/K, for a suitable choice of modulus $\mathfrak{m}$. The choice of $\mathfrak{m}$ is critical; as can be seen in Example 22.2 below, $\ker \psi_{L/K}^{\mathfrak{m}}$ need not be a congruence subgroup for the modulus $\mathfrak{m}$; there are constraints on the modulus $\mathfrak{m}$ that must be satisfied beyond the basic requirement that $\mathfrak{m}$ must be divisible by all the primes of K that ramify in L (so that $\psi_{L/K}^{\mathfrak{m}}$ is defined).

Example 22.2. Let $K = \mathbb{Q}$, and consider the cyclic cubic extension $L := \mathbb{Q}[x]/(x^3 - 3x - 1)$, which is ramified only at 3. The Artin map $\psi_{L/K}^{\mathfrak{m}}$ is well-defined for any modulus $\mathfrak{m}$ divisible by (3). The ray class field for $\mathfrak{m} = (3)$ is $\mathbb{Q}(\zeta_3)^+ = \mathbb{Q}$, and the ray class field for $\mathfrak{m} = (3)\infty$ is $\mathbb{Q}(\zeta_3) = \mathbb{Q}(\sqrt{-3})$, neither of which contains L, so $\ker \psi_{L/K}^{\mathfrak{m}}$ does not contain $\mathcal{R}_K^{\mathfrak{m}}$ for either of these moduli and is not a congruence subgroup. On the other hand, L is equal to $\mathbb{Q}(\zeta_9)^+$, the ray class field for $\mathfrak{m} = (9)$, so $\ker \psi_{L/K}^{\mathfrak{m}}$ contains (and is equal to) $\mathcal{R}_K^{\mathfrak{m}}$, and is thus a congruence subgroup for the modulus $\mathfrak{m} = (9)$.

If $\ker \psi_{L/K}^{\mathfrak{m}}$ is a congruence subgroup for the modulus $\mathfrak{m}$, then $\ker \psi_{L/K}^{\mathfrak{n}}$ is a congruence subgroup for each modulus $\mathfrak{n}$ divisible by $\mathfrak{m}$. If $\mathfrak{m}$ divides $\mathfrak{n}$ then $\mathcal{R}_K^{\mathfrak{n}} \subseteq \mathcal{R}_K^{\mathfrak{m}}$ and $\psi_{L/K}^{\mathfrak{n}}$ is the restriction of $\psi_{L/K}^{\mathfrak{m}}$ to $\mathcal{I}_K^{\mathfrak{n}}$, which contains $\mathcal{R}_K^{\mathfrak{n}}$. If $\mathfrak{m}$ and $\mathfrak{n}$ are supported on the same primes, then $\mathcal{I}_K^{\mathfrak{m}} = \mathcal{I}_K^{\mathfrak{n}}$ and $\psi_{L/K}^{\mathfrak{m}} = \psi_{L/K}^{\mathfrak{n}}$, but the ray groups $\mathcal{R}_K^{\mathfrak{m}}$ and $\mathcal{R}_K^{\mathfrak{n}}$ may differ.

To deal with these complications, we define an equivalence relation on congruence subgroups and show that each equivalence class has a canonical representative whose modulus divides the modulus of every equivalent congruence subgroup.

Definition 22.3. Let K be a number field with moduli $\mathfrak{m}_1$ and $\mathfrak{m}_2$. If $\mathcal{C}_1$ is a congruence subgroup for $\mathfrak{m}_1$ and $\mathcal{C}_2$ is a congruence subgroup for $\mathfrak{m}_2$, then we say that $(\mathcal{C}_1, \mathfrak{m}_1)$ and $(\mathcal{C}_2, \mathfrak{m}_2)$ are *equivalent* and write $(\mathcal{C}_1, \mathfrak{m}_1) \sim (\mathcal{C}_2, \mathfrak{m}_2)$ whenever

$$\mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{C}_2 = \mathcal{I}_K^{\mathfrak{m}_2} \cap \mathcal{C}_1.$$

Note that when $\mathfrak{m}_1 = \mathfrak{m}_2$ this reduces to $\mathcal{C}_1 = \mathcal{C}_2$.

Proposition 22.4. *Let K be a number field. The relation $(\mathcal{C}_1, \mathfrak{m}_1) \sim (\mathcal{C}_2, \mathfrak{m}_2)$ is an equivalence relation. If $(\mathcal{C}_1, \mathfrak{m}_1) \sim (\mathcal{C}_2, \mathfrak{m}_2)$ then $\mathcal{I}_K^{\mathfrak{m}_1}/\mathcal{C}_1 \simeq \mathcal{I}_K^{\mathfrak{m}_2}/\mathcal{C}_2$ are related by a canonical isomorphism that preserves cosets of fractional ideals prime to both $\mathfrak{m}_1$ and $\mathfrak{m}_2$.*

Proof. The relation $\sim$ is clearly symmetric, and reflexive. To show that it is transitive, let $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ be congruence subgroups for moduli $\mathfrak{m}_1, \mathfrak{m}_2, \mathfrak{m}_3$ and suppose $(\mathcal{C}_1, \mathfrak{m}_1) \sim (\mathcal{C}_2, \mathfrak{m}_2)$ and $(\mathcal{C}_2, \mathfrak{m}_2) \sim (\mathcal{C}_3, \mathfrak{m}_3)$. Let $I \in \mathcal{I}_K^{\mathfrak{m}_3} \cap \mathcal{C}_1$ and pick $\alpha \in K^{\mathfrak{m}_1\mathfrak{m}_3,1}$ so that $\alpha I \in \mathcal{I}_K^{\mathfrak{m}_1\mathfrak{m}_2\mathfrak{m}_3}$ (this is possible by Lemma 21.7 and Theorem 8.5). Then $(\alpha) \in \mathcal{R}_K^{\mathfrak{m}_1\mathfrak{m}_3} \subseteq \mathcal{R}_K^{\mathfrak{m}_1} \subseteq \mathcal{C}_1$ and $I \subseteq \mathcal{C}_1$, so $\alpha I \in \mathcal{C}_1$, and we also have $\alpha I \in \mathcal{I}_K^{\mathfrak{m}_1\mathfrak{m}_2\mathfrak{m}_3} \subseteq \mathcal{I}_K^{\mathfrak{m}_2}$, so

$$\alpha I \in \mathcal{I}_K^{\mathfrak{m}_2} \cap \mathcal{C}_1 = \mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{C}_2 \subseteq \mathcal{C}_2,$$

since $\mathcal{C}_1 \sim \mathcal{C}_2$, and $\alpha I \in \mathcal{I}_K^{\mathfrak{m}_1\mathfrak{m}_2\mathfrak{m}_3} \subseteq \mathcal{I}_K^{\mathfrak{m}_3}$, so

$$\alpha I \in \mathcal{I}_K^{\mathfrak{m}_3} \cap \mathcal{C}_2 = \mathcal{I}_K^{\mathfrak{m}_2} \cap \mathcal{C}_3 \subseteq \mathcal{C}_3,$$

since $\mathcal{C}_2 \sim \mathcal{C}_3$. We have $(\alpha) \in \mathcal{R}_K^{\mathfrak{m}_1\mathfrak{m}_3} \subseteq \mathcal{R}_K^{\mathfrak{m}_3}$, so $(\alpha) \in \mathcal{C}_3$ and therefore $(\alpha)^{-1} \in \mathcal{C}_3$, since $\mathcal{C}_3$ is a group. Thus $\alpha^{-1}\alpha I = I \in \mathcal{C}_3$, and we also have $I \in \mathcal{C}_1 \subseteq \mathcal{I}_K^{\mathfrak{m}_1}$, so $I \in \mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{C}_3$. Since $I \in \mathcal{I}_K^{\mathfrak{m}_3} \cap \mathcal{C}_1$ was chosen arbitrarily, this proves that

$$I_K^{\mathfrak{m}_3} \cap \mathcal{C}_1 \subseteq \mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{C}_3.$$

The reverse inclusion follows by symmetry, so $(\mathcal{C}_1, \mathfrak{m}_1) \sim (\mathcal{C}_3, \mathfrak{m}_3)$ as desired.

For the last statement, for any fractional ideal $I \in \mathcal{I}_K^{\mathfrak{m}_1}$ we can pick $\alpha \in K^{\mathfrak{m}_1,1}$ so that $\alpha I \in \mathcal{I}_K^{\mathfrak{m}_2}$ (via Lemma 21.7 and Theorem 8.5). The image of αI in $\mathcal{I}_K^{\mathfrak{m}_2}/\mathcal{C}_2$ does not depend on the choice of α, since for any $\alpha' \in K^{\mathfrak{m}_1,1}$ with $\alpha' I \in \mathcal{I}_K^{\mathfrak{m}_2}$ we have $(\alpha I)/(\alpha' I) = (\alpha/\alpha') \in \mathcal{I}_K^{\mathfrak{m}_2}$ and $(\alpha/\alpha') \in \mathcal{R}_K^{\mathfrak{m}_1}$, so $(\alpha/\alpha') \in \mathcal{I}_K^{\mathfrak{m}_2} \cap \mathcal{R}_K^{\mathfrak{m}_1} = \mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{R}_K^{\mathfrak{m}_2} \subseteq \mathcal{R}_K^{\mathfrak{m}_2}$. This defines a group homomorphism $\varphi\colon \mathcal{I}_K^{\mathfrak{m}_1} \to \mathcal{I}_K^{\mathfrak{m}_2}/\mathcal{C}_2$. For $I \in \mathcal{C}_1$, we have $\alpha I \in \mathcal{I}_K^{\mathfrak{m}_2} \cap \mathcal{C}_1 = \mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{C}_2 \subseteq \mathcal{C}_2$, but for $I \in \mathcal{I}_K^{\mathfrak{m}_1} - \mathcal{C}_1$ we have $\alpha I \in \mathcal{I}_K^{\mathfrak{m}_2} - \mathcal{C}_1$ and therefore $\alpha I \notin \mathcal{C}_2$, so $\ker\varphi = \mathcal{C}_1$. It follows that φ induces an injective homomorphism $\mathcal{I}_K^{\mathfrak{m}_1}/\mathcal{C}_1 \to \mathcal{I}_K^{\mathfrak{m}_2}/\mathcal{C}_2$, and by symmetry we have an injective homomorphism in the opposite direction, so $\mathcal{I}_K^{\mathfrak{m}_1}/\mathcal{C}_1 \simeq \mathcal{I}_K^{\mathfrak{m}_2}/\mathcal{C}_2$ as claimed.

This isomorphism is independent of the choice of α used to define it (hence canonical), and for fractional ideals I coprime to both $\mathfrak{m}_1$ and $\mathfrak{m}_2$ we can choose $\alpha = 1$, in which case the coset of I in $\mathcal{I}_K^{\mathfrak{m}_1}/\mathcal{C}_1$ will be identified with the coset of I in $\mathcal{I}_K^{\mathfrak{m}_2}/\mathcal{C}_2$. □

We now observe that if $\mathcal{C}$ is a congruence subgroup for two moduli $\mathfrak{m}_1$ and $\mathfrak{m}_2$, then $(\mathcal{C}, \mathfrak{m}_1) \sim (\mathcal{C}, \mathfrak{m}_2)$. In particular, each subgroup of $\mathcal{I}_K$ lies in at most one equivalence class of congruence subgroups. We can thus view the equivalence relation $(\mathcal{C}_1, \mathfrak{m}_1) \sim (\mathcal{C}_2, \mathfrak{m}_2)$ as an equivalence relation on the congruence subgroups of $\mathcal{I}_K$ and write $\mathcal{C}_1 \sim \mathcal{C}_2$ without ambiguity. It follows from Proposition 22.4 that each equivalence class of congruence subgroups uniquely determines a finite abelian group that is the quotient of a ray class group.

Within an equivalence class of congruence subgroups there can be at most one congruence subgroup for each modulus (since $\mathcal{C}_1 \sim \mathcal{C}_2 \Leftrightarrow \mathcal{C}_1 = \mathcal{C}_2$ whenever $\mathcal{C}_1$ and $\mathcal{C}_2$ are congruence subgroups for the same modulus). The following lemma gives a criterion for determining when there exists a congruence subgroup of a given modulus within a given equivalence class.

Lemma 22.5. *Let $\mathcal{C}_1$ be a congruence subgroup of modulus $\mathfrak{m}_1$ for a number field K. There exists a congruence subgroup $\mathcal{C}_2$ of modulus $\mathfrak{m}_2|\mathfrak{m}_1$ equivalent to $\mathcal{C}_1$ if and only if*

$$\mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{R}_K^{\mathfrak{m}_2} \subseteq \mathcal{C}_1,$$

in which case $\mathcal{C}_2 = \mathcal{C}_1\mathcal{R}_K^{\mathfrak{m}_2}$.

Proof. Note that $\mathfrak{m}_2|\mathfrak{m}_1$ implies $\mathcal{I}_K^{\mathfrak{m}_1} \subseteq \mathcal{I}_K^{\mathfrak{m}_2}$, so $\mathcal{C}_1 \subseteq \mathcal{I}_K^{\mathfrak{m}_1} \subseteq \mathcal{I}_K^{\mathfrak{m}_2}$.

Suppose $\mathcal{C}_2 \sim \mathcal{C}_1$ has modulus $\mathfrak{m}_2$. Then $\mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{C}_2 = \mathcal{I}_K^{\mathfrak{m}_2} \cap \mathcal{C}_1 = \mathcal{C}_1$, and $\mathcal{R}_K^{\mathfrak{m}_2} \subseteq \mathcal{C}_2$, so $\mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{R}_K^{\mathfrak{m}_2} \subseteq \mathcal{C}_1$ as claimed. Now suppose $\mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{R}_K^{\mathfrak{m}_2} \subseteq \mathcal{C}_1$, and let $\mathcal{C}_2 := \mathcal{C}_1\mathcal{R}_K^{\mathfrak{m}_2}$. Then $\mathcal{C}_2$ is a congruence subgroup of modulus $\mathfrak{m}_2$ and we have

$$\mathcal{I}_K^{\mathfrak{m}_2} \cap \mathcal{C}_1 = \mathcal{C}_1 = \mathcal{C}_1(\mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{R}_K^{\mathfrak{m}_2}) = \mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{C}_1\mathcal{R}_K^{\mathfrak{m}_2} = \mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{C}_2,$$

so $\mathcal{C}_1 \sim \mathcal{C}_2$. The equivalence class of $\mathcal{C}_1$ contains at most one congruence subgroup of modulus $\mathfrak{m}_2$, so if one exists it must be $\mathcal{C}_2 = \mathcal{C}_1\mathcal{R}_K^{\mathfrak{m}_2}$. □

Proposition 22.6. *Let $\mathcal{C}_1 \sim \mathcal{C}_2$ be congruence subgroups of modulus $\mathfrak{m}_1$ and $\mathfrak{m}_2$, respectively. There exists a congruence subgroup $\mathcal{C} \sim \mathcal{C}_1 \sim \mathcal{C}_2$ with modulus $\mathfrak{n} := \gcd(\mathfrak{m}_1, \mathfrak{m}_2)$.*

Proof. Put $\mathfrak{m} := \operatorname{lcm}(\mathfrak{m}_1, \mathfrak{m}_2)$ and $\mathcal{D} := \mathcal{I}_K^{\mathfrak{m}_2} \cap \mathcal{C}_1 = \mathcal{I}_K^{\mathfrak{m}_1} \cap \mathcal{C}_2$; then

$$\mathcal{R}_K^{\mathfrak{m}} = \mathcal{R}_K^{\mathfrak{m}_1} \cap \mathcal{R}_K^{\mathfrak{m}_2} \subseteq \mathcal{D} \subseteq \mathcal{I}_K^{\mathfrak{m}},$$

so $\mathcal{D}$ is a congruence subgroup of modulus $\mathfrak{m}$, and we have

$$\mathcal{I}_K^{\mathfrak{m}} \cap \mathcal{R}_K^{\mathfrak{m}_1} \subseteq \mathcal{D} \qquad \text{and} \qquad \mathcal{I}_K^{\mathfrak{m}} \cap \mathcal{R}_K^{\mathfrak{m}_2} \subseteq \mathcal{D},$$

so $\mathcal{D} \sim \mathcal{C}_1 \sim \mathcal{C}_2$, by Lemma 22.5. To prove the existence of an equivalent congruence subgroup $\mathcal{C}$ of modulus $\mathfrak{n}$ it suffices to show $\mathcal{I}_K^{\mathfrak{m}} \cap \mathcal{R}_K^{\mathfrak{n}} \subseteq \mathcal{D}$ (again by Lemma 22.5).

So let $\mathfrak{a} = (\alpha) \in \mathcal{I}_K^{\mathfrak{m}} \cap \mathcal{R}_K^{\mathfrak{n}}$, and choose $\beta \in K^{\mathfrak{m}} \cap K^{\mathfrak{m}_2,1}$ so that $\alpha\beta \in K^{\mathfrak{m}_1,1}$ (this is possible by Theorem 8.5 because $\mathfrak{m} = \operatorname{lcm}(\mathfrak{m}_1, \mathfrak{m}_2)$ and $\mathfrak{n} = \gcd(\mathfrak{m}_1, \mathfrak{m}_2)$). Then $(\beta) \in \mathcal{D}$ and $\beta\mathfrak{a} \in \mathcal{I}_K^{\mathfrak{m}} \cap \mathcal{R}_K^{\mathfrak{m}_1} \subseteq \mathcal{D}$, so $\beta^{-1}\beta\mathfrak{a} = \mathfrak{a} \in \mathcal{D}$. Thus $\mathcal{I}_K^{\mathfrak{m}} \cap \mathcal{R}_K^{\mathfrak{n}} \subseteq \mathcal{D}$ and therefore $\mathcal{C} = \mathcal{D}\mathcal{R}_K^{\mathfrak{n}}$ is a congruence subgroup of modulus $\mathfrak{n}$ equivalent to $D \sim \mathcal{C}_1 \sim \mathcal{C}_2$. □

Corollary 22.7. *Let $\mathcal{C}$ be a congruence subgroup of modulus $\mathfrak{m}$ for a number field K. There is a unique congruence subgroup in the equivalence class of $\mathcal{C}$ whose modulus $\mathfrak{c}$ divides the modulus of every congruence subgroup equivalent to $\mathcal{C}$.*

Definition 22.8. Let $\mathcal{C}$ be a congruence subgroup for a number field K. The unique modulus $\mathfrak{c} := \mathfrak{c}(\mathcal{C})$ given by Corollary 22.7 is the *conductor* of $\mathcal{C}$, and we say that $\mathcal{C}$ is *primitive* if $\mathcal{C} = \mathcal{C}\mathcal{R}_K^{\mathfrak{c}}$ (the unique congruence subgroup of modulus $\mathfrak{c}$ equivalent to $\mathcal{C}$).

Proposition 22.9. *Let $\mathcal{C}$ be a primitive congruence subgroup of modulus $\mathfrak{m}$ for a number field K. Then $\mathfrak{m}$ is the conductor of every congruence subgroup of modulus $\mathfrak{m}$ contained in $\mathcal{C}$; in particular, $\mathfrak{m}$ is the conductor of $\mathcal{R}_K^{\mathfrak{m}}$.*

Proof. Let $\mathcal{C}_0 \subseteq \mathcal{C}$ be a congruence subgroup of modulus $\mathfrak{m}$ and let $\mathfrak{c}$ be its conductor. Then $\mathfrak{c}|\mathfrak{m}$ and $\mathcal{I}_K^{\mathfrak{m}} \cap \mathcal{R}_K^{\mathfrak{c}} \subseteq \mathcal{C}_0 \subseteq \mathcal{C}$, by Lemma 22.5, and this implies that there is a congruence subgroup of modulus $\mathfrak{c}$ equivalent to $\mathcal{C}$, and therefore $\mathfrak{m}|\mathfrak{c}$, so $\mathfrak{c} = \mathfrak{m}$. □

The proposition implies that a modulus $\mathfrak{m}$ occurs as a conductor if and only if $\mathcal{R}_K^{\mathfrak{m}}$ is primitive. This does not always hold: consider $K = \mathbb{Q}$ and $\mathfrak{m} = (2)$, for example; the conductor of $\mathcal{R}_{\mathbb{Q}}^{(2)} = \mathcal{I}_{\mathbb{Q}}^{(2)}$ is (1), since $\mathcal{R}_{\mathbb{Q}}^{(2)} \cap \mathcal{I}_{\mathbb{Q}}^{(1)} = \mathcal{I}_{\mathbb{Q}}^{(1)} \cap \mathcal{I}_{\mathbb{Q}}^{(2)}$ implies $\mathcal{R}_{\mathbb{Q}}^{(2)} \sim \mathcal{I}_{\mathbb{Q}}^{(1)}$. Thus (2) is not the conductor of any congruence subgroup for $\mathbb{Q}$.

22.2 Ray class characters

We now want to prove a generalization of Dirichlet's theorem on primes in arithmetic progressions. Given a congruence subgroup $\mathcal{C}$ for a modulus $\mathfrak{m}$ we would like to compute the Dirichlet density $d(\mathcal{C}) := d(\{\mathfrak{p} \in \mathcal{C}\})$ of the set of prime ideals $\mathfrak{p} \in \mathcal{I}_K^{\mathfrak{m}}$ that lie in $\mathcal{C}$. We first need to generalize our notion of a Dirichlet character.

Definition 22.10. Let K be a number field and let $\chi\colon \mathcal{I}_K \to \mathbb{C}$ be a totally multiplicative function with finite image; so $\chi(\mathcal{O}_K) = 1$, $\chi(IJ) = \chi(I)\chi(J)$ for all $I, J \in \mathcal{I}_K$, and χ restricts to a homomorphism from a subgroup of $\mathcal{I}_K$ to a finite subgroup of $\mathrm{U}(1)$ whose kernel we denote $\ker\chi$. If $\mathfrak{m}$ is a modulus for K such that $\chi^{-1}(\mathrm{U}(1)) = \mathcal{I}_K^{\mathfrak{m}}$ and $\mathcal{R}_K^{\mathfrak{m}} \subseteq \ker\chi$, then χ is a *ray class character of modulus* $\mathfrak{m}$ and its kernel is a congruence subgroup of modulus $\mathfrak{m}$. Equivalently, χ is the *extension by zero* of a character of the finite abelian group $\mathrm{Cl}_K^{\mathfrak{m}} = \mathcal{I}_K^{\mathfrak{m}}/\mathcal{R}_K^{\mathfrak{m}}$ defined by setting $\chi(I) = 0$ for $I \notin \mathcal{I}_K^{\mathfrak{m}}$.

Example 22.11. For $K = \mathbb{Q}$ there is a one-to-one correspondence between Dirichlet characters $\chi\colon \mathbb{Z} \to \mathbb{C}$ and ray class characters $\chi'\colon \mathcal{I}_{\mathbb{Q}} \to \mathbb{C}$ with $\chi(a) = \chi'((a))$ for all $a \in \mathbb{Z}_{\geq 1}$. Each Dirichlet character χ of modulus m corresponds to a ray class character of modulus $\mathfrak{m} = (m)\infty$ whose conductor divides (m) if and only if χ is an *even* Dirichlet character, meaning that $\chi(-1) = 1$.

Definition 22.12. Let χ_1, χ_2 be ray class characters of moduli $\mathfrak{m}_1, \mathfrak{m}_2$ of a number field K, with $\mathfrak{m}_1 | \mathfrak{m}_2$. If $\chi_2(I) = \chi_1(I)$ for all $I \in \mathcal{I}_K^{\mathfrak{m}_2}$, then χ_2 is *induced* by χ_1. A ray class character is *primitive* if it is not induced by any ray class character other than itself.

Definition 22.13. The *conductor* of a ray class character χ is the conductor $\mathfrak{c}(\chi) := \mathfrak{c}(\ker\chi)$ of its kernel (as a congruence subgroup).

Theorem 22.14. *A ray class character is primitive if and only if its kernel is primitive. Every ray class character χ is induced by a unique primitive ray class character $\widetilde{\chi}$.*

Proof. Let χ be a ray class character of modulus $\mathfrak{m}$, let $\kappa\colon \mathcal{I}_K^{\mathfrak{m}}/(\ker\chi) \to \mathrm{U}(1)$ be the group character induced by χ, and let $\mathcal{C}$ be the primitive congruence subgroup equivalent to $\ker\chi$ with modulus $\mathfrak{c} = \mathfrak{c}(\chi)$ dividing $\mathfrak{m}$ given by Corollary 22.7. By Proposition 22.4, we have a canonical isomorphism $\varphi\colon \mathcal{I}_K^{\mathfrak{c}}/\mathcal{C} \xrightarrow{\sim} \mathcal{I}_K^{\mathfrak{m}}/(\ker\chi)$ that we can use to define a ray class character $\widetilde{\chi}$ of modulus $\mathfrak{c}$ as the extension by zero of the character $\kappa \circ \varphi$ of $\mathcal{I}_K^{\mathfrak{c}}/\mathcal{C}$. The isomorphism φ preserves cosets of fractional ideals in $\mathcal{I}_K^{\mathfrak{m}} \subseteq \mathcal{I}_K^{\mathfrak{c}}$, so $\widetilde{\chi}(I) = \chi(I)$ for all $I \in \mathcal{I}_K^{\mathfrak{m}}$ and χ is induced by $\widetilde{\chi}$.

If χ_2 is a ray class character of conductor $\mathfrak{m}_2$ induced by a ray class character χ_1 of conductor $\mathfrak{m}_1$, then $\ker\chi_1 \cap \mathcal{I}_K^{\mathfrak{m}_2} = \ker\chi_2 = \ker\chi_2 \cap \mathcal{I}_K^{\mathfrak{m}_1}$ and $\ker\chi_1 \sim \ker\chi_2$, and we also note that if $\chi_1 \neq \chi_2$ then $\mathcal{I}_K^{\mathfrak{m}_1} \neq \mathcal{I}_K^{\mathfrak{m}_2}$ and $\mathfrak{m}_1 \neq \mathfrak{m}_2$. It follows that $\widetilde{\chi}$ is primitive, it is the unique primitive ray class character that induces χ. Thus χ is primitive if and only if it is equal to $\widetilde{\chi}$, which holds if and only if $\ker\chi = \ker\widetilde{\chi}$ is primitive. □

Theorem 22.14 is a direct generalization of Theorem 18.13 for Dirichlet characters. For a modulus $\mathfrak{m}$ of K we use $X(\mathfrak{m})$ to denote the set of primitive ray class characters of conductor dividing $\mathfrak{m}$, which we note is in bijection with the character group of $\mathrm{Cl}_K^{\mathfrak{m}}$, and thus has a group structure given by $\widetilde{\chi_1}\widetilde{\chi_2} = \widetilde{\chi_1\chi_2}$. Indeed, for each character of $\mathrm{Cl}_K^{\mathfrak{m}}$, its extension by zero is a ray class character χ of modulus $\mathfrak{m}$ induced by a primitive ray class character $\widetilde{\chi}$ whose conductor divides $\mathfrak{m}$, and each primitive ray class character $\widetilde{\chi}$ of conductor dividing $\mathfrak{m}$ induces a ray class character χ of modulus $\mathfrak{m}$ that determines a character of $\mathrm{Cl}_K^{\mathfrak{m}}$; these two maps are inverses, hence bijections. This generalizes Corollary 18.16.

Definition 22.15. A ray class character χ is *principal* if $\ker \chi = \chi^{-1}(\mathrm{U}(1))$. We use $\mathbb{1}$ to denote the unique primitive principal ray class character.

Remark 22.16. For Dirichlet characters, $\mathbb{1}$ is the unique Dirichlet character of conductor 1, but for ray class characters this holds only when the class group Cl_K is trivial (as when $K = \mathbb{Q}$). In general, the extension by zero of any character of Cl_K is a ray class character of conductor (1) and need not be principal (but is necessarily primitive).

Like Dirichlet characters, each ray class character has an associated L-function.

Definition 22.17. The *Weber L-function* $L(s,\chi)$ of a ray class character χ for a number field K is the complex function

$$L(s,\chi) := \prod_{\mathfrak{p}} \big(1 - \chi(\mathfrak{p})\mathrm{N}(\mathfrak{p})^{-s}\big)^{-1} = \sum_{\mathfrak{a}} \chi(\mathfrak{a})\mathrm{N}(\mathfrak{a})^{-s},$$

where the the product is over prime ideals of $\mathcal{O}_K$ and the sum is over nonzero $\mathcal{O}_K$-ideals; the product and sum both converge to a non-vanishing holomorphic function on $\mathrm{Re}(s) > 1$ (this follows from comparison with the Dedekind zeta function $\zeta_K(s)$, since $|\chi(\mathfrak{a})| \le 1$).

Example 22.18. For $K = \mathbb{Q}$, Weber L-functions are Dirichlet L-functions. For any number field K, the Weber L-function for $\mathbb{1}$ is the Dedekind zeta function: $L(s,\mathbb{1}) = \zeta_K(s)$.

More generally, we have the following theorem, which is analogous to Theorem 19.15 but avoids the need to assume the existence of a ray class field.

Proposition 22.19. *Let χ be a ray class character of modulus $\mathfrak{m}$ for a number field K of degree n. Then $L(s,\chi)$ extends to a meromorphic function on $\mathrm{Re}(s) > 1 - \frac{1}{n}$ that has at most a simple pole at $s = 1$ and is holomorphic if χ is non-principal.*

Proof. Associated to each ray class $\gamma \in \mathrm{Cl}_K^{\mathfrak{m}}$ we have a Dirichlet series

$$\zeta_{K,\gamma}(s) := \sum_{\mathfrak{a}\in\gamma} \mathrm{N}(\mathfrak{a})^{-s}$$

that is holomorphic on $\mathrm{Re}(s) > 1$. For the trivial modulus $\mathfrak{m}$, our proof of analytic class number formula (Theorem 19.12) implies that $\zeta_{K,\gamma}(s)$ has a meromorphic continuation to $1 - \frac{1}{n}$ with a simple pole at $s = 1$ and residue $\rho = 2^r(2\pi)^2 R_K/(\omega_K|D_K|^{1/2})$, independent of γ. Recall that in our proof of Theorem 19.8 we treated each $\gamma \in \mathrm{Cl}_K = \mathrm{cl}(\mathcal{O}_K)$ separately and obtained the same value of ρ for each γ, leading to the residue $\rho_K = h_K\rho$ that appears in Theorem 19.12.

The same proof works for $\mathrm{Cl}_K^{\mathfrak{m}}$, *mutatis mutandi*: replace $\mathrm{covol}(\mathcal{O}_K)$ with $\mathrm{covol}(\mathfrak{m}_0)$, replace the regulator $R_K = \mathrm{covol}(\pi(\mathrm{Log}(\mathcal{O}_K^\times)))$ with $R_K^{\mathfrak{m}} := \mathrm{covol}(\pi(\mathrm{Log}(\mathcal{O}_K^\times \cap K^{\mathfrak{m},1})))$, and replace $w_K = \#\mu_K$ with $w_K^{\mathfrak{m}} := \#(\mu_K \cap K^{\mathfrak{m},1})$. The exact value of ρ is not important to us here, the key point is that $\zeta_{K,\gamma}(s)$ has a meromorphic continuation to $\mathrm{Re}(s) > 1 - \frac{1}{n}$ with a simple pole at $s = 1$ whose residue ρ depends only on K and $\mathfrak{m}$ (not γ).

We then have

$$\begin{aligned} L(s,\chi) &= \sum_{\gamma\in\mathrm{Cl}_K^{\mathfrak{m}}} \chi(\gamma)\zeta_{K,\gamma}(s) \\ &= \sum_{\gamma\in\mathrm{Cl}_K^{\mathfrak{m}}} \chi(\gamma)\big(\zeta_{K,\gamma}(s) - \rho\,\zeta(s)\big) + \sum_{\gamma\in\mathrm{Cl}_K^{\mathfrak{m}}} \chi(\gamma)\rho\,\zeta(s), \end{aligned}$$

The first sum is a finite sum of functions holomorphic on $\mathrm{Re}(s) > 1 - \frac{1}{n}$ (since $\zeta(s)$ has a simple pole at $s = 1$ with residue 1), and the second sum vanishes whenever χ is non-principal (by Corollary 18.37). The proposition follows. □

We now prove a generalization of Dirichlet's theorem on primes on arithmetic progressions for arbitrary number fields. We proved the nonvanishing of Dirichlet L-functions $L(1,\chi)$ for non-principal χ using the analytic class number formula for $\mathbb{Q}(\zeta_m)$, the ray class field $\mathbb{Q}((m)\infty)$, by writing the Dedekind zeta function for $\mathbb{Q}(\zeta_m)$ as a product of Dirichlet L-functions (see Theorem 19.15). A similar approach works for Weber L-functions, assuming the existence of ray class fields $K(\mathfrak{m})$: the Dedekind zeta function of $K(\mathfrak{m})$ is equal to the product of the Weber L-functions for $\chi \in X(\mathfrak{m})$. But we will prove the non-vanishing of $L(1,\chi)$ for non-principal χ without assuming the existence of ray class fields.

For a congruence subgroup $\mathcal{C}$, let $X(\mathcal{C})$ denote the set of primitive ray class characters whose kernels contain $\mathcal{C}$. If $\mathcal{C}$ is a congruence subgroup of modulus $\mathfrak{m}$ then $X(\mathcal{C})$ is a subgroup of $X(\mathfrak{m})$ isomorphic to the character group of $\mathcal{I}_K^{\mathfrak{m}}/\mathcal{C}$ and we may view $X(\mathcal{C})$ as the the character group of $\mathcal{I}_K^{\mathfrak{m}}/\mathcal{C}$.

Theorem 22.20. *Let $\mathcal{C}$ be a congruence subgroup of modulus $\mathfrak{m}$ for a number field K and let $n := [\mathcal{I}_K^{\mathfrak{m}} : \mathcal{C}]$. The set of primes $\{\mathfrak{p} \in \mathcal{C}\}$ has Dirichlet density*

$$d(\mathcal{C}) = \begin{cases} \frac{1}{n} & \text{if } L(1,\chi) \neq 0 \text{ for all } \chi \neq \mathbb{1} \text{ in } X(\mathcal{C}), \\ 0 & \text{otherwise.} \end{cases}$$

In fact $d(\mathcal{C}) = \frac{1}{n}$ always holds, as we will prove in Corollary 22.22 below, but it is easier to prove the theorem as stated and then use this to derive the corollary.

Proof. We proceed as in the proof of Dirichlet's theorem on primes in arithmetic progressions (see §18.4). We first construct the indicator function for the set $\{\mathfrak{p} \in \mathcal{C}\}$:

$$\frac{1}{n} \sum_{\chi \in X(\mathcal{C})} \chi(\mathfrak{p}) = \begin{cases} 1 & \text{if } \mathfrak{p} \in \mathcal{C}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that summing over $\chi \in X(\mathcal{C})$ is equivalent to summing over the character group of $\mathcal{I}_K^{\mathfrak{m}}/\mathcal{C}$, so Corollary 18.37 applies: therefore $\sum \chi(\mathfrak{p}) = 0$ unless the image of $\mathfrak{p}$ in $I_K^{\mathfrak{m}}/\mathcal{C}$ is the identity, meaning that $\mathfrak{p} \in \mathcal{C}$, in which case $\sum \chi(\mathfrak{p}) = \#X(\mathcal{C}) = n$.

As $s \to 1^+$ we have

$$\log L(s,\chi) \sim \sum_{\mathfrak{p}} \chi(\mathfrak{p}) \mathrm{N}(\mathfrak{p})^{-s},$$

and therefore

$$\begin{aligned} \sum_{\chi \in X(\mathcal{C})} \log L(s,\chi) &\sim \sum_{\chi \in X(\mathcal{C})} \sum_{\mathfrak{p}} \chi(\mathfrak{p}) \mathrm{N}(\mathfrak{p})^{-s} \\ &\sim n \sum_{\mathfrak{p} \in \mathcal{C}} \mathrm{N}(\mathfrak{p})^{-s}. \end{aligned}$$

By Proposition 22.19, we may write

$$L(s,\chi) = (s-1)^{e(\chi)} g(s)$$

for some function $g(s)$ that is holomorphic and nonvanishing on a neighborhood of 1, where $e(\chi) := \operatorname{ord}_{s=1} L(s,\chi)$ is -1 when $\chi = \mathbb{1}$, and $e(\chi) \geq 0$ otherwise. We have

$$\log \frac{1}{s-1} - \sum_{\chi \neq \mathbb{1}} e(\chi) \log \frac{1}{s-1} \sim n \sum_{\mathfrak{p} \in \mathcal{C}} \mathrm{N}(\mathfrak{p})^{-s}.$$

Dividing both sides by $n \log \frac{1}{s-1}$ yields

$$\frac{1 - \sum_{\chi \neq \mathbb{1}} e(\chi)}{n} \sim \frac{\sum_{\mathfrak{p} \in \mathcal{C}} \mathrm{N}(\mathfrak{p})^{-s}}{\log \frac{1}{s-1}} \qquad (\text{as } s \to 1^+),$$

thus

$$d(\mathcal{C}) = \lim_{s \to 1^+} \frac{\sum_{\mathfrak{p} \in \mathcal{C}} \mathrm{N}(\mathfrak{p})^{-s}}{\log \frac{1}{s-1}} = \frac{1 - \sum_{\chi \neq \mathbb{1}} e(\chi)}{n}.$$

The $e(\chi)$ are integers and the Dirichlet density is nonnegative, so either $e(\chi) = 0$ for all $\chi \neq \mathbb{1}$, in which case $L(1,\chi) \neq 0$ for all $\chi \neq \mathbb{1}$ and $d(\mathcal{C}) = \frac{1}{n}$, or $e(\chi) = 1$ for exactly one of the $\chi \neq \mathbb{1}$ and $d(\mathcal{C}) = 0$. (in fact this never happens, as noted above). □

Proposition 22.21. *Let $\mathcal{C}$ be a congruence subgroup of modulus $\mathfrak{m}$ for a number field K and let $n := [\mathcal{I}_K^{\mathfrak{m}} : \mathcal{C}]$. For every $I \in \mathcal{I}_K^{\mathfrak{m}}$ the set $\{\mathfrak{p} \in I\mathcal{C}\}$ has Dirichlet density*

$$d(I\mathcal{C}) = \begin{cases} \frac{1}{n} & \text{if } L(1,\chi) \neq 0 \text{ for all characters } \chi \neq \mathbb{1} \text{ in } X(\mathcal{C}), \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The proof is the same as in Theorem 22.20, except we now use the indicator function

$$\frac{1}{n} \sum_{\chi \in X(\mathcal{C})} \chi(I)^{-1} \chi(\mathfrak{p}) = \begin{cases} 1 & \text{if } \mathfrak{p} \in I\mathcal{C}, \\ 0 & \text{otherwise,} \end{cases}$$

and obtain

$$\sum_{\chi \in X(\mathcal{C})} \chi(I)^{-1} \log L(s,\chi) \sim \sum_{\chi \in X(\mathcal{C})} \sum_{\mathfrak{p}} \chi(I)^{-1} \chi(\mathfrak{p}) \mathrm{N}(\mathfrak{p})^{-s} \sim n \sum_{\mathfrak{p} \in I\mathcal{C}} \mathrm{N}(\mathfrak{p})^{-s}.$$

The rest of the proof is the same. □

Corollary 22.22. *Let $\mathcal{C}$ be a congruence subgroup of modulus $\mathfrak{m}$ for a number field K and let $n := [\mathcal{I}_K^{\mathfrak{m}} : \mathcal{C}]$. For every ideal $I \in \mathcal{I}_K^{\mathfrak{m}}$ the set $\{\mathfrak{p} \in I\mathcal{C}\}$ has Dirichlet density $1/n$, and for every $\chi \neq \mathbb{1}$ in $X(\mathcal{C})$ we have $L(1,\chi) \neq 0$.*

Proof. Let $I_1, \dots, I_n \in \mathcal{I}_K^{\mathfrak{m}}$ be a complete set of coset representatives for $\mathcal{C} \subseteq \mathcal{I}_K^{\mathfrak{m}}$. All but finitely many primes $\mathfrak{p}$ of K lie in $\mathcal{I}_K^{\mathfrak{m}}$, hence in one of the cosets $I_j\mathcal{C}$ partitioning $\mathcal{I}_K^{\mathfrak{m}}$, therefore

$$d(I_1\mathcal{C}) + \cdots + d(I_n\mathcal{C}) = 1.$$

By Proposition 22.21, every term in this sum is either 0 or $1/n$, and the equality implies they must all be equal to $1/n$, which then implies $L(1,\chi) \neq 0$ for all $\chi \neq \mathbb{1}$ in $X(\mathcal{C})$. □

Corollary 22.23. *Let L/K be an abelian extension of number fields and let $\mathcal{C}$ be a congruence subgroup for a modulus $\mathfrak{m}$ of K. If $\mathrm{Spl}(L) \precsim \{\mathfrak{p} \in \mathcal{C}\}$ then*

$$[\mathcal{I}_K^{\mathfrak{m}} : \mathcal{C}] \le [L : K],$$

with equality whenever $\mathrm{Spl}(L) \sim \{\mathfrak{p} \in \mathcal{C}\}$.

Proof. We know from Theorem 21.15 that $\mathrm{Spl}(L)$ has polar density $1/[L:K]$, and this is also its Dirichlet density, by Proposition 21.12. The set $\{\mathfrak{p} \in \mathcal{C}\}$ has Dirichlet density $1/[\mathcal{I}_K^{\mathfrak{m}} : \mathcal{C}]$, by Theorem 22.22, and $\mathrm{Spl}(L) \precsim \{\mathfrak{p} \in \mathcal{C}\}$ (by assumption), so

$$\frac{1}{[L:K]} = d(\mathrm{Spl}(L)) \;\le\; d(\mathcal{C}) = \frac{1}{[\mathcal{I}_K^{\mathfrak{m}} : \mathcal{C}]}.$$ □

22.3 The conductor of an abelian extension

We now introduce another notion of conductor, one attached to an abelian extension of number fields, which is defined as a product of local conductors attached to corresponding abelian extensions of the local field K_v for each place $v \in M_K$.

Definition 22.24. Let L/K be a finite abelian extension of local fields. The *conductor* $\mathfrak{c}(L/K)$ is defined as follows.[1] If K is archimedean then $\mathfrak{c}(L/K) = 1$ when $K \simeq \mathbb{R}$ and $L \simeq \mathbb{C}$ and $\mathfrak{c}(L/K) = 0$ otherwise. If K is nonarchimedean and $\mathfrak{p}$ is the maximal ideal of its valuation ring $\mathcal{O}_K$, then

$$\mathfrak{c}(L/K) := \min\{n : 1 + \mathfrak{p}^n \subseteq \mathrm{N}_{L/K}(L^\times)\}$$

(here $1 + \mathfrak{p}^n$ is a subgroup of $\mathcal{O}_K^\times$, with $1 + \mathfrak{p}^0 := \mathcal{O}_K^\times$). If L/K is a finite abelian extension of global fields then its conductor is the modulus

$$\begin{aligned} \mathfrak{c}(L/K) : M_K &\to \mathbb{Z} \\ v &\mapsto \mathfrak{c}(L_w/K_v) \end{aligned}$$

where K_v is the completion of K at v and L_w is the completion of L at a place $w|v$. (the fact that L/K is Galois ensures that $\mathfrak{c}(L_w/K_v)$ is the same for every $w|v$). As with any modulus, we may view the finite part of $\mathfrak{c}(L/K)$ as an $\mathcal{O}_K$-ideal and the infinite part as a subset of ramified infinite places.

It is not hard to show that conductor is supported on ramified places (in particular, it has finite support, as required for a modulus). More generally, we have the following.

Proposition 22.25. *Let L/K be a finite abelian extension of local or global fields. For each prime $\mathfrak{p}$ of K we have*

$$v_{\mathfrak{p}}(\mathfrak{c}(L/K)) = \begin{cases} 0 & \text{if and only if } \mathfrak{p} \text{ is unramified,} \\ 1 & \text{if and only if } \mathfrak{p} \text{ is ramified tamely,} \\ \ge 2 & \text{if and only if } \mathfrak{p} \text{ is ramified wildly.} \end{cases}$$

Proof. See Problem Set 11. □

[1] Many authors use $\mathfrak{f}(L/K)$ rather than $\mathfrak{c}(L/K)$, we use $\mathfrak{c}$ to avoid confusion with the residue field degree.

The finite part of the conductor of an abelian extension divides the discriminant ideal and is divisible by the same set of primes, but the valuation of the conductor at these primes is typically smaller than that of the discriminant. For example, the discriminant of the extension $\mathbb{Q}(\zeta_p)/\mathbb{Q}$ is $(p)^{p-2}$, but its conductor is $(p)\infty$.

Lemma 22.26. *Let L_1/K and L_2/K be two finite abelian extensions of a local or global field K. If $L_1 \subseteq L_2$ then $\mathfrak{c}(L_1/K)$ divides $\mathfrak{c}(L_2/K)$.*

Proof. If $K \simeq \mathbb{R}, \mathbb{C}$ the result is clear, and for nonarchimedean local K we may apply $N_{L_2/K}(L_2^\times) = N_{L_1/K}(N_{L_2/L_1}(L_2^\times)) \subseteq N_{L_1/K}(L_1^\times)$. The global case follows. □

22.4 Norm groups

We can now identify a candidate for the kernel of the Artin map $\psi_{L/K}^{\mathfrak{m}} \colon \mathcal{I}_K^{\mathfrak{m}} \to \mathrm{Gal}(L/K)$. Recall from Lecture 6 that the norm map $N_{L/K} \colon \mathcal{I}_L \to \mathcal{I}_K$ can be defined by

$$\prod_i \mathfrak{q}_i^{n_i} \mapsto \prod_i \mathfrak{p}_i^{n_i f_i},$$

where $\mathfrak{p}_i \coloneqq \mathfrak{q}_i \cap \mathcal{O}_K$ and $f_i := [\mathbb{F}_{\mathfrak{q}_i} : \mathbb{F}_{\mathfrak{p}_i}]$ is the residue field degree.

Definition 22.27. Let L/K be a finite abelian extension of number fields and let $\mathfrak{m}$ be a modulus for K divisible by the conductor of L/K. The *norm group* (or *Takagi group*) associated to $\mathfrak{m}$ is the congruence subgroup

$$T_{L/K}^{\mathfrak{m}} := \mathcal{R}_K^{\mathfrak{m}} N_{L/K}(\mathcal{I}_L^{\mathfrak{m}}),$$

where $\mathcal{I}_L^{\mathfrak{m}}$ denotes the subgroup of fractional ideals in $\mathcal{I}_L$ that are coprime to $\mathfrak{m}\mathcal{O}_L$.

Proposition 22.28. *Let L/K be a finite abelian extension of number fields and let $\mathfrak{m}$ be a modulus for K divisible by the conductor of L/K. Then $\ker \psi_{L/K}^{\mathfrak{m}} \subseteq T_{L/K}^{\mathfrak{m}}$.*

Proof. Let $\mathfrak{p}$ be a prime of K that lies in $\ker \psi_{L/K}^{\mathfrak{m}}$. Then $\mathfrak{p}$ is coprime to $\mathfrak{m}$ and splits completely in L, so $e_{\mathfrak{p}} = f_{\mathfrak{p}} = 1$. There is at least one prime $\mathfrak{q}$ of L above $\mathfrak{p}$, and for this prime we have $N_{L/K}(\mathfrak{q}) = \mathfrak{p}^{f_{\mathfrak{p}}} = \mathfrak{p}$ (by Theorem 6.10), so $\mathfrak{p} \in N_{L/K}(\mathcal{I}_L^{\mathfrak{m}}) \subseteq T_{L/K}^{\mathfrak{m}}$. □

To prove Artin reciprocity we need to establish the reverse inclusion, which requires a different approach (we will prove it for the trivial modulus $\mathfrak{m}$ over the next two lectures). But we can record the following theorem, historically known as the "first" fundamental inequality of class field theory (in modern terminology it is typically known as the second, even though it was proved first, by Weber).

Theorem 22.29. *Let L/K be a finite abelian extension of number fields and let $\mathfrak{m}$ be a modulus for K divisible by the conductor of L/K. Then*

$$[\mathcal{I}_K^{\mathfrak{m}} : T_{L/K}^{\mathfrak{m}}] \le [L : K].$$

Proof. Proposition 22.28 implies $[\mathcal{I}_K^{\mathfrak{m}} : T_{L/K}^{\mathfrak{m}}] \le [\mathcal{I}_K^{\mathfrak{m}} : \ker \psi_{L/K}^{\mathfrak{m}}] = [L : K]$, where the equality follows from the surjectivity of the Artin map (Theorem 21.19). □

22.5 The main theorems of class field theory (ideal-theoretic version)

We can give a more precise statement of the main theorems of class field theory. Let $\mathfrak{m}$ be a modulus for a number field K. The three main theorems of class field theory state that:

- **Existence**: The ray class field $K(\mathfrak{m})$ exists.
- **Completeness**: If L/K is finite abelian then $L \subseteq K(\mathfrak{m})$ if and only $\mathfrak{c}(L/K) \mid \mathfrak{m}$. In particular, every finite abelian L/K lies in a ray class field.
- **Artin reciprocity**: For each subextension L/K of $K(\mathfrak{m})$ we have $\ker \psi^{\mathfrak{m}}_{L/K} = T^{\mathfrak{m}}_{L/K}$ with conductor $\mathfrak{c}(L/K)|\mathfrak{m}$ and a canonical isomorphism $\mathcal{I}^{\mathfrak{m}}_K/T^{\mathfrak{m}}_{L/K} \simeq \mathrm{Gal}(L/K)$.

Artin reciprocity gives us a commutative diagram of canonical bijections:

$$\begin{array}{ccc} \{\text{abelian } L/K \text{ with } \mathfrak{c}(L/K) \mid \mathfrak{m}\} & \xrightarrow{L \mapsto T^{\mathfrak{m}}_{L/K}} & \{\text{congruence subgroups } \mathcal{C} \subseteq \mathcal{I}^{\mathfrak{m}}_K\} \\ \Big\downarrow{\scriptstyle L \mapsto \mathrm{Gal}(L/K)} & & \Big\downarrow{\scriptstyle \mathcal{C} \mapsto \mathcal{I}^{\mathfrak{m}}_K/\mathcal{C}} \\ \{\text{quotients of } \mathrm{Gal}(K(\mathfrak{m})/K)\} & \xleftarrow[\psi^{\mathfrak{m}}_{L/K}]{} & \{\text{ quotients of } \mathrm{Cl}^{\mathfrak{m}}_K\} \end{array}$$

22.6 The Hilbert class field

Definition 22.30. Let K be global field. The *Hilbert class field* of K is the maximal unramified abelian extension of K (the compositum of all finite unramified abelian extensions of K inside a fixed separable closure of K).

While it is not obvious from the definition, it follows from the completeness theorem of class field theory that the Hilbert class field must be the ray class field for the trivial modulus, and in particular, that it is a finite extension of K. This is a remarkable result (which we will prove in a later lecture), since infinite unramified extensions of number fields do exist (they are necessarily nonabelian).

Indeed, one way to construct such an extension is by considering a tower of Hilbert class fields. Starting with a number field $K_0 := K$, for each integer $n \geq 0$ define K_{n+1} to be the be the Hilbert class field of K_n. This yields an infinite tower of finite abelian extensions

$$K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots,$$

and we may then consider the field $L := \bigcup_n K_n$. There are two possibilities: either we eventually reach a field K_n with class number 1, in which case $K_m = K_n$ for all $m \geq n$ and L/K is a finite unramified extension of K, or this does not happen and L/K is an infinite unramified extension of K (which is necessarily nonabelian). It was a longstanding open question as to whether the latter could occur, but in 1964 Golod and Shafarevich proved that indeed it can; in particular, the field

$$K_0 = \mathbb{Q}(\sqrt{-30030}) = \mathbb{Q}(\sqrt{-2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13})$$

is the base of an infinite tower of Hilbert class field extensions. One might ask whether one can use an imaginary quadratic field of smaller discriminant than this. It is known that no imaginary quadratic field of discriminant $|D| \leq 420$ has an infinite Hilbert class field tower [3]; they all stabilize at either K_2 or K_3.

Extensions arising from Hilbert class field towers are necessarily solvable, since they are towers of finite abelian extensions. One might ask whether infinite nonsolvable unramified extensions exist. As shown by Maire [2], they do, and this can happen even when the base field has class number one and the Hilbert class field tower is trivial. Indeed, the biquadratic extension

$$\mathbb{Q}(\sqrt{17601097}, \sqrt{17380678572169893})$$

has class number one and its maximal unramified extension is an infinite extension.

References

[1] Henri Cohen, *Advanced topics in computational number theory*, Springer, 2000.

[2] Christian Maire, *On infinite unramfieid extensions*, Pacific J. Math. **192** (2000), 135–142.

[3] Ken Yamamura, *Maximal unramified extensions of imaginary quadratic number fields of small conductors*, Journal de Théorie des Nombres de Bordeaux **9** (1997) 405–448.

23 Tate cohomology

In this lecture we introduce a variant of group cohomology known as *Tate cohomology*, and we define the *Herbrand quotient* (a ratio of cardinalities of two Tate cohomology groups), which will play a key role in our proof of Artin reciprocity. We begin with a brief review of group cohomology, restricting our attention to the minimum we need to define the Tate cohomology groups we will use. At a number of points we will need to appeal to some standard results from homological algebra whose proofs can be found in Section 23.6. For those seeking a more thorough introduction to group cohomology, see [1]; for general background on homological algebra, we recommend [7].

23.1 Group cohomology

Definition 23.1. Let G be a group. A *G-module* is an abelian group A equipped with a G-action compatible with its group structure: $g(a+b) = ga + gb$ for all $g \in G, a, b \in A$.[1] This implies $|ga| = |a|$ (where $|a| := \#\langle a\rangle$ is the order of a); in particular $ga = 0 \Leftrightarrow a = 0$.

A *trivial* G-module is an abelian group with trivial G-action: $ga = a$ for all $g \in G, a \in A$ (so every abelian group can be viewed as a trivial G-module). A morphism of G-modules is a morphism of abelian groups $\alpha\colon A \to B$ satisfying $\alpha(ga) = g\alpha(a)$. Kernels, images, quotients, and direct sums of G-modules are also G-modules.

Definition 23.2. Let A be a G-module. The *G-invariants* of A constitute the G-module

$$A^G := \{a \in A : ga = a \text{ for all } g \in G\}$$

consisting of elements fixed by G. It is the largest trivial G-submodule of A.

Definition 23.3. Let A be a G-module and let $n \in \mathbb{Z}_{\geq 0}$. The group of *$n$-cochains* is the abelian group $C^n(G,A) := \mathrm{Map}(G^n, A)$ of maps of sets $f\colon G^n \to A$ under pointwise addition. We have $C^0(G,A) \simeq A$, since $G^0 = \{1\}$ is a singleton set. The *nth coboundary map* $d^n\colon C^n(G,A) \to C^{n+1}(G,A)$ is the homomorphism of abelian groups defined by

$$\begin{aligned} d^n(f)(g_0,\dots,g_n) := {} & g_0 f(g_1,\dots,g_n) - f(g_0g_1, g_2,\dots,g_n) + f(g_0, g_1g_2,\dots,g_n) \\ & \cdots + (-1)^n f(g_0,\dots,g_{n-2}, g_{n-1}g_n) + (-1)^{n+1} f(g_0,\dots,g_{n-1}). \end{aligned}$$

The group $C^n(G,A)$ contains subgroups of *n-cocycles* and *n-coboundaries* defined by

$$Z^n(G,A) := \ker d^n \qquad \text{and} \qquad B^n(G,A) := \operatorname{im} d^{n-1},$$

with $B^0(G,A) := \{0\}$.

The coboundary map satisfies $d^{n+1} \circ d^n = 0$ for all $n \geq 0$ (this can be verified directly, but we will prove it in the next section), thus $B^n(G,A) \subseteq Z^n(G,A)$ for $n \geq 0$ and the groups $C^n(G,A)$ with connecting maps d^n form a *cochain complex*

$$0 \longrightarrow C^0(G,A) \xrightarrow{d^0} C^1(G,A) \xrightarrow{d^1} C^2(G,A) \longrightarrow \cdots$$

that we may denote $\mathcal{C}_A$. In general, a cochain complex (of abelian groups) is simply a sequence of homomorphisms d^n that satisfy $d^{n+1} \circ d^n = 0$. Cochain complexes form a category whose morphisms are commutative diagrams with cochain complexes as rows.

[1] Here we put the G-action on the left (one can also define right G-modules), and for the sake of readability we write A additively, even though we will be primarily interested in cases where A is a multiplicative group.

Definition 23.4. Let A be a G-module. The nth *cohomology group of G with coefficients in A* is the abelian group

$$H^n(G,A) \coloneqq Z^n(G,A)/B^n(G,A).$$

Example 23.5. We can work out the first few cohomology groups explicitly by writing out the coboundary maps and computing kernels and images:

- $d^0\colon C^0(G,A) \to C^1(G,A)$ is defined by $d^0(a)(g) \coloneqq ga - a$ (note $C^0(G,A) \simeq A$).
- $H^0(G,A) \simeq \ker d^0 = A^G$ (note $B^0(G,A) = \{0\}$).
- $\operatorname{im} d^0 = \{f\colon G \to A \mid \exists a \in A : f(g) = ga - a \text{ for all } g \in G\}$ (*principal crossed homomorphisms*).
- $d^1\colon C^1(G,A) \to C^2(G,A)$ is defined by $d^1(f)(g,h) \coloneqq gf(h) - f(gh) + f(g)$.
- $\ker d^1 = \{f\colon G \to A \mid f(gh) = f(g) + gf(h) \text{ for all } g,h \in G\}$ (*crossed homomorphisms*).
- $H^1(G,A) =$ *crossed homomorphisms* modulo *principal crossed homomorphisms*.
- If A is a trivial G-module then $H^1(G,A) \simeq \operatorname{Hom}(G,A)$.

Lemma 23.6. *Let $\alpha\colon A \to B$ be a morphism of G-modules. We have induced group homomorphisms $\alpha^n\colon C^n(G,A) \to C^n(G,B)$ defined by $f \mapsto \alpha \circ f$ that commute with the coboundary maps. In particular, α^n maps cocycles to cocycles and coboundaries to coboundaries and thus induces homomorphisms $\alpha^n\colon H^n(G,A) \to H^n(G,B)$ of cohomology groups, and we have a morphism of cochain complexes $\alpha\colon \mathcal{C}_A \to \mathcal{C}_B$:*

$$\begin{array}{ccccccccc}
0 & \longrightarrow & C^0(G,A) & \xrightarrow{d^0} & C^1(G,A) & \xrightarrow{d^1} & C^2(G,A) & \xrightarrow{d^2} & \cdots \\
 & & \downarrow{\scriptstyle \alpha^0} & & \downarrow{\scriptstyle \alpha^1} & & \downarrow{\scriptstyle \alpha^2} & & \\
0 & \longrightarrow & C^0(G,B) & \xrightarrow{d^0} & C^1(G,B) & \xrightarrow{d^1} & C^2(G,B) & \xrightarrow{d^2} & \cdots
\end{array}$$

Proof. Consider any $n \geq 0$. For all $f \in C^n(G,A)$, and $g_0,\dots,g_n \in G$ we have

$$\begin{aligned}
\alpha^{n+1}(d^n(f)(g_0,\dots,g_n)) &= \alpha^{n+1}\big(g_0 f(g_1,\dots,g_n) - \cdots + (-1)^{n+1} f(g_0,\dots,g_{n-1})\big) \\
&= g_0(\alpha\circ f)(g_1,\dots,g_n) - \cdots + (-1)^{n+1}(\alpha\circ f)(g_0,\dots,g_{n-1}) \\
&= d^n(\alpha\circ f)(g_0,\dots,g_n) = d^n(\alpha^n(f))(g_0,\dots,g_n),
\end{aligned}$$

thus $\alpha^{n+1}\circ d^n = d^n \circ \alpha^n$. The lemma follows. □

Lemma 23.6 implies that we have a family of functors $H^n(G,\bullet)$ from the category of G-modules to the category of abelian groups (note that $\mathrm{id}\circ f = f$ and $(\alpha\circ\beta)\circ f = \alpha\circ(\beta\circ f)$), and also a functor from the category of G-modules to the category of cochain complexes.

Lemma 23.7. *Suppose that we have a short exact sequence of G-modules*

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0.$$

Then for every $n \geq 0$ we have a corresponding exact sequence of n-cochains

$$0 \longrightarrow C^n(G,A) \xrightarrow{\alpha^n} C^n(G,B) \xrightarrow{\beta^n} C^n(G,C) \longrightarrow 0.$$

Proof. The injectivity of α^n follows from the injectivity of α. If $f \in \ker \beta^n$, then $\beta \circ f = 0$ and $\operatorname{im} f \subseteq \ker \beta = \operatorname{im} \alpha$; via the bijection $\alpha^{-1}\colon \operatorname{im}\alpha \to A$ we can define $\alpha^{-1} \circ f \in C^n(G, A)$, and therefore $\ker \beta^n \subseteq \operatorname{im} \alpha^n$. We also have $\operatorname{im} \alpha^n \subseteq \ker \beta^n$, since $\beta \circ \alpha \circ f = 0 \circ f = 0$ for all $f \in C^n(G, A)$, and exactness at $C^n(G, B)$ follows. Every $f \in C^n(G, C)$ satisfies $\operatorname{im} f \subseteq C = \operatorname{im} \beta$, and we can define $h \in C^n(G, B)$ satisfying $\beta \circ h = f$: for each $g_0, \dots, g_n$ let $h(g_0, \dots, g_n)$ be any element of $\beta^{-1}(f(g_0, \dots, g_n))$. Thus $f \in \operatorname{im} \beta^n$ and β^n is surjective. □

Lemmas 23.6 and 23.7 together imply that we have an exact functor from the category of G-modules to the category of cochain complexes.

Theorem 23.8. *Every short exact sequence of G-modules*

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0$$

induces a long exact sequence of cohomology groups

$$0 \to H^0(G, A) \xrightarrow{\alpha^0} H^0(G, B) \xrightarrow{\beta^0} H^0(G, C) \xrightarrow{\delta^0} H^1(G, A) \longrightarrow \cdots$$

and commutative diagrams of short exact sequences of G-modules induce corresponding commutative diagrams of long exact sequences of cohomology groups.

Proof. Lemmas 23.6 and 23.7 give us the commutative diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & C^n(G, A) & \xrightarrow{\alpha^n} & C^n(G, B) & \xrightarrow{\beta^n} & C^n(G, C) & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle d^n} & & \downarrow{\scriptstyle d^n} & & \downarrow{\scriptstyle d^n} & & \\
0 & \longrightarrow & C^{n+1}(G, A) & \xrightarrow{\alpha^{n+1}} & C^{n+1}(G, B) & \xrightarrow{\beta^{n+1}} & C^{n+1}(G, C) & \longrightarrow & 0
\end{array}$$

We have $B^n(G, A) \subseteq Z^n(G, A) \subseteq C^n(G, A) \xrightarrow{d^n} B^{n+1}(G, A) \subseteq Z^{n+1}(G, A) \subseteq C^{n+1}(G, A)$, thus d^n induces a homomorphism $d^n\colon C^n(G, A)/B^n(G, A) \to Z^{n+1}(G, A)$, and similarly for the G-modules B and C. The fact that α^n maps coboundaries to coboundaries and cocycles to cocycles implies that we have induced maps $C^n(G, A)/B^n(G, A) \to C^n(G, B)/B^n(G, B)$ and $Z^{n+1}(G, A) \to Z^{n+1}(G, B)$; similar comments apply to β^n.

We thus have the following commutative diagram:

$$\begin{array}{ccccccccc}
 & & \frac{C^n(G,A)}{B^n(G,A)} & \xrightarrow{\alpha^n} & \frac{C^n(G,B)}{B^n(G,B)} & \xrightarrow{\beta^n} & \frac{C^n(G,C)}{B^n(G,C)} & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle d^n} & & \downarrow{\scriptstyle d^n} & & \downarrow{\scriptstyle d^n} & & \\
0 & \longrightarrow & Z^{n+1}(G, A) & \xrightarrow{\alpha^{n+1}} & Z^{n+1}(G, B) & \xrightarrow{\beta^{n+1}} & Z^{n+1}(G, C) & &
\end{array}$$

The kernels of the vertical maps d^n are (by definition) the cohomology groups $H^n(G, A)$, $H^n(G, B)$, $H^n(G, C)$, and the cokernels are $H^{n+1}(G, A)$, $H^{n+1}(G, B)$, $H^{n+1}(G, C)$. Applying the snake lemma yields the exact sequence

$$H^n(G, A) \xrightarrow{\alpha^n} H^n(G, B) \xrightarrow{\beta^n} H^n(G, C) \xrightarrow{\delta^n} H^{n+1}(G, A) \xrightarrow{\alpha^{n+1}} H^{n+1}(G, B) \xrightarrow{\beta^{n+1}} H^{n+1}(G, C),$$

where α^n and β^n are the homomorphisms in cohomology induced by α and β (coming from α^n and β^n in the previous diagram via Lemma 23.6), and the connecting homomorphism δ^n given by the snake lemma can be explicitly described as

$$\begin{aligned}
\delta^n\colon H^n(G, C) &\to H^{n+1}(G, A) \\
[f] &\mapsto [\alpha^{-1} \circ d^n(\hat{f})]
\end{aligned}$$

where $[f]$ denotes the cohomology class of a cocycle $f \in C^n(G,C)$ and $\hat{f} \in C^n(G,B)$ is a cochain satisfying $\beta \circ \hat{f} = f$. Here α^{-1} denotes the inverse of the isomorphism $A \to \alpha(A)$. The fact that δ^n is well defined (independent of the choice of $\hat{f}$) is part of the snake lemma. The map $H^0(G,A) \to H^0(G,B)$ is the restriction of $\alpha\colon A \to B$ to A^G, and is thus injective (recall that $H^0(G,A) \simeq A^G$). This completes the first part of the proof.

For the second part, suppose we have the following commutative diagram of short exact sequences of G-modules

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \xrightarrow{\alpha} & B & \xrightarrow{\beta} & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle\phi} & & \downarrow{\scriptstyle\psi} & & \downarrow{\scriptstyle\varphi} & & \\ 0 & \longrightarrow & A' & \xrightarrow{\alpha'} & B' & \xrightarrow{\beta'} & C' & \longrightarrow & 0 \end{array}$$

By Lemma 23.6, to verify the commutativity of the corresponding diagram of long exact sequences in cohomology we only need to check commutativity at squares of the form

$$\begin{array}{ccc} H^n(G,C) & \xrightarrow{\delta^n} & H^{n+1}(G,A) \\ \downarrow{\scriptstyle\varphi^n} & & \downarrow{\scriptstyle\phi^{n+1}} \\ H^n(G,C') & \xrightarrow{\delta'^n} & H^{n+1}(G,A') \end{array} \tag{1}$$

Let $f\colon G^n \to C$ be a cocycle and choose $\hat{f} \in C^n(G,B)$ such that $\beta \circ \hat{f} = f$. We have

$$\phi^{n+1}(\delta^n([f])) = \phi^{n+1}([\alpha^{-1} \circ d^n(\hat{f})]) = [\phi \circ \alpha^{-1} \circ d^n(\hat{f})].$$

Noting that $\varphi \circ f = \varphi \circ \beta \circ \hat{f} = \beta' \circ \psi \circ \hat{f}$ and $\phi \circ \alpha^{-1} = \alpha'^{-1} \circ \psi$ (as maps $\alpha(A) \to A'$) yields

$$\delta'^n(\varphi^n([f])) = \delta'^n([\beta' \circ \psi \circ f]) = [\alpha'^{-1} \circ d^n(\psi \circ \hat{f})] = [\alpha'^{-1} \circ \psi \circ d^n(\hat{f})] = [\phi \circ \alpha^{-1} \circ d^n(\hat{f})],$$

thus diagram (1) commutes as desired. □

Definition 23.9. A family of functors F^n from the category of G-modules to the category of abelian groups that associates to each short exact sequence of G-modules a long exact sequence of abelian groups such that commutative diagrams of short exact sequences yield commutative diagrams of long exact sequences is called a *δ-functor*. A δ-functor is said to be *cohomological* if the connecting homomorphisms in long exact sequences are of the form $\delta^n\colon F^n(G,C) \to F^{n+1}(G,A)$. If we instead have $\delta^n\colon F^{n+1}(G,C) \to F^n(G,A)$ then the δ-functor is *homological*.

Theorem 23.54 implies that the family of functors $H^n(G,\bullet)$ is a cohomological δ-functor. In fact is the universal cohomological δ-functor (it satisfies a universal property that determines it up to a unique isomorphism of δ-functors), but we will not explore this further.

23.2 Cohomology via free resolutions

Recall that the *group ring* $\mathbb{Z}[G]$ consists of formal sums $\sum_g a_g g$ indexed by $g \in G$ with coefficients $a_g \in \mathbb{Z}$, all but finitely many zero. Multiplication is given by $\mathbb{Z}$-linearly extending the group operation in G; the ring $\mathbb{Z}[G]$ is commutative if and only if G is. As an abelian group under addition, $\mathbb{Z}[G]$ is the free $\mathbb{Z}$-module with basis G, equivalently, the group of finitely supported functions $G \to \mathbb{Z}$ under pointwise addition.

The notion of a G-module defined in the previous section is equivalent to that of a (left) $\mathbb{Z}[G]$-module: to define multiplication by $\mathbb{Z}[G]$ one must define a G-action, and the G-action on a G-module extends $\mathbb{Z}$-linearly, since every G-module is also a $\mathbb{Z}$-module. The multiplicative identity 1 of the ring $\mathbb{Z}[G]$ is the identity element of G; the additive identity 0 is the empty sum, which acts on A by sending $a \in A$ to the identity element of A.[2]

For any $n \geq 0$ we view $\mathbb{Z}[G^n]$ as a G-module with G acting diagonally on the left: $g \cdot (g_1, \dots, g_n) \coloneqq (gg_1, \dots, gg_n)$. This makes $\mathbb{Z}[G^0] = \mathbb{Z}$ a trivial G-module (here we are viewing the empty tuple as the identity element of the trivial group G^0).

Definition 23.10. Let G be a group. The *standard resolution of $\mathbb{Z}$ by G-modules* is the exact sequence of G-module homomorphisms

$$\cdots \longrightarrow \mathbb{Z}[G^{n+1}] \xrightarrow{d_n} \mathbb{Z}[G^n] \longrightarrow \cdots \xrightarrow{d_1} \mathbb{Z}[G] \xrightarrow{d_0} \mathbb{Z} \longrightarrow 0,$$

where the boundary maps d_n are defined by

$$d_n(g_0, \dots, g_n) \coloneqq \sum_{i=0}^{n} (-1)^i (g_0, \dots, \hat{g}_i, \dots, g_n)$$

and extended $\mathbb{Z}$-linearly (the notation $\hat{g}_i$ means omit g_i from the tuple). The map d_0 sends each $g \in G$ to 1, and extends to the map $\sum_g a_g g \mapsto \sum_g a_g$, which is also known as the *augmentation map* and may be denoted ε.

Let us verify the exactness of the standard resolution.

Lemma 23.11. *The standard resolution of $\mathbb{Z}$ by G-modules is exact.*

Proof. The map d_0 is clearly surjective. To check $\operatorname{im} d_{n+1} \subseteq \ker d_n$ it suffices to note that for any $g_0, \dots, g_n \in G$ we have

$$d_n(d_{n+1}(g_0, \dots, g_n)) = \sum_{0 \leq i \leq n} \Bigg(\sum_{0 \leq j < i} (-1)^{i+j} (g_0, \dots, \hat{g}_j, \dots, \hat{g}_i \dots, g_n) + \sum_{i < j \leq n} (-1)^{i+j-1} (g_0, \dots, \hat{g}_i, \dots, \hat{g}_j, \dots, g_n) \Bigg) = 0.$$

Let G_1^{n+1} be the subgroup $1 \times G^n$ of G^{n+1}, and let $h \colon \mathbb{Z}[G^{n+1}] \to \mathbb{Z}[G_1^{n+2}] \subseteq \mathbb{Z}[G^{n+2}]$ be the $\mathbb{Z}$-linear map defined by $(g_0, \dots, g_{n+1}) \mapsto (1, g_0, \dots, g_{n+1})$. For $x \in \mathbb{Z}[G^{n+1}]$ we have $d_{n+1}(h(x)) \in x + \mathbb{Z}[G_1^{n+1}]$, and if $x \in \ker d_n$ then $x - d_{n+1}(h(x)) \in \ker d_n \cap \mathbb{Z}[G_1^{n+1}]$, since $\operatorname{im} d_{n+1} \subseteq \ker d_n$. To prove $\ker d_n \subseteq \operatorname{im} d_{n+1}$, it suffices to show $\ker d_n \cap \mathbb{Z}[G_1^{n+1}] \subseteq \operatorname{im} d_{n+1}$. For $n = 0$ we have $\ker d_0 \cap \mathbb{Z}[G_1^1] = \{0\}$, and we now proceed by induction on $n \geq 1$.

Let $G_{11}^{n+1} \coloneqq 1 \times 1 \times G^{n-1} \subseteq G_1^{n+1}$. We can write the free $\mathbb{Z}$-module $\mathbb{Z}[G_1^{n+1}]$ as the internal direct sum $\mathbb{Z}[G_1^{n+1}] + X$, where X is the free $\mathbb{Z}$-module generated by elements of the form $(1, g_1, \dots, g_n)$ with $g_1 \neq 1$. For $g_1 \neq 1$ the image of $(1, g_1, \dots)$ under d_n has the form $(g_1, \dots, g_n) + y$ with $y \in G_1^n$, and it follows that the restriction of d_n to X is injective and thus has trivial kernel. It therefore suffices to show $\ker d_n \cap \mathbb{Z}[G_{11}^{n+1}] \subseteq \operatorname{im} d_{n+1}$.

Let $x \in \ker d_n \cap \mathbb{Z}[G_{11}^{n+1}]$. If $n = 1$ then $x = d_2(h(x)) \in \operatorname{im} d_{n+1}$. For $n \geq 2$, let $\pi \colon \mathbb{Z}[G^{n+1}] \to \mathbb{Z}[G^{n-1}]$ be the $\mathbb{Z}$-linear map defined by $(g_0, g_1, g_2, \dots, g_n) \mapsto (g_2, \dots, g_n)$. We have $\pi(x) \in \ker d_{n-2} \subset \operatorname{im} d_{n-1}$ (by the inductive hypothesis), and for any $y \in d_{n-1}^{-1}(\pi(x))$ we have $x = d_{n+1}(h_{11}(y)) \in \operatorname{im} d_{n+1}$, where $h_{11} \colon \mathbb{Z}[G^{n-1}] \to \mathbb{Z}[G^{n+1}]$ is the $\mathbb{Z}$-linear map defined by $(g_0, \dots, g_{n-1}) \mapsto (1, 1, g_0, \dots, g_{n-1})$. Therefore $\ker d_n \cap \mathbb{Z}[G_{11}^{n+1}] \subseteq \operatorname{im} d_{n+1}$. □

[2] When A is written multiplicatively its identity is denoted 1 and one should think of 0 as acting via exponentiation (but for the moment we continue to use additive notation and view A as a left $\mathbb{Z}[G]$-module).

Definition 23.12. Let R be a (not necessarily commutative) ring. A *free resolution* P of a (left) R-module M is an exact sequence of free (left) R-modules P_n

$$\cdots \xrightarrow{d_{n+1}} P_{n+1} \xrightarrow{d_n} P_n \xrightarrow{d_{n-1}} \cdots \xrightarrow{d_1} P_1 \xrightarrow{d_0} M \longrightarrow 0.$$

Free resolutions arise naturally as presentations of an R-module. Every R-module M admits a surjection from a free module (one can always take P_1 to be the free R-module with basis M). This yields an exact sequence $P_1 \to M \to 0$, and the kernel of the homomorphism on the left is itself an R-module that admits a surjection from a free R-module P_2; continuing in this fashion yields a free resolution.

Now let A be an abelian group. If we *truncate* the free resolution P by removing the R-module M and apply the contravariant left exact functor $\operatorname{Hom}_R(\bullet, A)$ we obtain a cochain complex of R-modules[3]

$$\cdots \xleftarrow{d_{n+1}^*} P_{n+1}^* \xleftarrow{d_n^*} P_n^* \xleftarrow{d_{n-1}^*} \cdots \xleftarrow{d_1^*} P_1^* \longleftarrow 0.$$

where $d_n^*(\varphi) \coloneqq \varphi \circ d_n$. The maps d_n^* satisfy $d_{n+1}^* \circ d_n^* = 0$: for all $\varphi \in \operatorname{Hom}_R(P_n, A)$ we have

$$(d_{n+1}^* \circ d_n^*)(\varphi) = (d_n \circ d_{n+1})^*(\varphi) = \varphi \circ d_n \circ d_{n+1} = \varphi \circ 0 = 0.$$

This cochain complex need not be exact, because the functor $\operatorname{Hom}_R(\bullet, A)$ is not right-exact,[4] so we have potentially nontrivial cohomology groups $\ker d_{n+1}^* / \operatorname{im} d_n^*$, which are denoted $\operatorname{Ext}_R^n(M, A)$. A key result of homological algebra is that (up to isomorphism) these cohomology groups do not depend on the resolution P, only on A and M; see Theorem 23.71.

Recall that $\mathbb{Z}[G]$ is a free $\mathbb{Z}$-module (with basis G), and for all $n \geq 0$ we have

$$\mathbb{Z}[G^{n+1}] \simeq \bigoplus_{(g_1,\dots,g_n)\in G^n} \mathbb{Z}[G](1, g_1, \dots, g_n).$$

It follows that the standard resolution is a free resolution of $\mathbb{Z}$ by $\mathbb{Z}[G]$-modules; note that $\mathbb{Z}$, like any abelian group, can always be viewed as a trivial G-module, hence a $\mathbb{Z}[G]$-module.

With a free resolution in hand, we now want to consider the cochain complex

$$0 \to \operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G], A) \longrightarrow \cdots \longrightarrow \operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^n], A) \xrightarrow{d_n^*} \operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^{n+1}], A) \longrightarrow \cdots$$

where d_n^* is defined by $\varphi \mapsto \varphi \circ d_n$. Let $\mathcal{S}_A$ denote this cochain complex.

Proposition 23.13. *Let A be a G-module. For every $n \geq 0$ we have an isomorphism of abelian groups*

$$\Phi^n \colon \operatorname{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^{n+1}], A) \xrightarrow{\sim} C^n(G, A)$$

that sends $\varphi \colon \mathbb{Z}[G^{n+1}] \to A$ to the function $f \colon G^n \to A$ defined by

$$f(g_1, \dots, g_n) \coloneqq \varphi(1,\ g_1,\ g_1g_2,\ \dots,\ g_1g_2\cdots g_n).$$

The isomorphisms Φ^n satisfy $\Phi^{n+1} \circ d_{n+1}^* = d^n \circ \Phi^n$ for all $n \geq 0$ and thus define an isomorphism of cochain complexes $\Phi_A \colon \mathcal{S}_A \to \mathcal{C}_A$.

[3] The intuition here is that P contains a presentation of M that effectively serves as a replacement for M.

[4] Applying $\operatorname{Hom}_{\mathbb{Z}}(\bullet, \mathbb{Z})$ to $0 \to \mathbb{Z} \to \mathbb{Q} \to \mathbb{Q}/\mathbb{Z} \to 0$ yields $0 \leftarrow \mathbb{Z} \leftarrow 0 \leftarrow 0 \leftarrow 0$, for example.

Proof. We first check that Φ^n is injective. Let $\varphi \in \ker \Phi^n$. Given $g_0, \dots, g_n \in G$, let $h_i := g_{i-1}^{-1} g_i$ for $1 \le i \le n$ so that $h_1 \cdots h_i = g_0^{-1} g_i$ and observe that

$$\varphi(g_0, \dots, g_n) = g_0 \varphi(1, g_0^{-1} g_1, \dots, g_0^{-1} g_n) = g_0 \varphi(1, h_1, h_1 h_2, \dots, h_1 \cdots h_n) = 0.$$

so $\varphi = 0$ as desired. For surjectivity, let $f \in C^n(G, A)$ and define $\varphi \in \mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^{n+1}], A)$ via $\varphi(g_0, \dots, g_n) := g_0 f(g_0^{-1} g_1, g_1^{-1} g_2, \dots, g_{n-1}^{-1} g_n)$. For any $g_1, \dots, g_n \in G$ we have

$$\Phi^n(\varphi)(g_1, \dots, g_n) = \varphi(1, g_1, g_1 g_2, \dots, g_1 g_2 \cdots g_n) = f(g_1, \dots, g_n),$$

so $f \in \mathrm{im}\, \Phi^n$ and Φ^n is surjective.

It is clear from the definition that $\Phi^n(\varphi_1 + \varphi_2) = \Phi^n(\varphi_1) + \Phi^n(\varphi_2)$, so Φ^n is a bijective group homomorphism, hence an isomorphism. Finally, for any $\varphi \in \mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^{n+1}], A)$ and $g_1, \dots, g_{n+1} \in G$ we have

$$\begin{aligned}
\Phi^{n+1}(d_{n+1}^*(\varphi))(g_1, \dots, g_{n+1}) &= d_{n+1}^*(\varphi)(1, g_1, g_1 g_2, \dots, g_1 \cdots g_{n+1}) \\
&= \varphi(d_{n+1}(1, g_1, g_1 g_2, \dots, g_1 \cdots g_{n+1})) \\
&= \sum_{i=0}^{n+1} (-1)^i \varphi(1, g_1, \dots, g_1 \cdots g_{i-1}, g_1 \cdots g_{i+1}, \dots, g_1 \cdots g_{n+1}) \\
&= g_1 \Phi^n(\varphi)(g_2, \dots, g_{n+1}) \\
&\quad + \sum_{i=1}^{n} (-1)^i \Phi^n(\varphi)(g_1, \dots, g_{i-2}, g_{i-1} g_i, g_{i+1}, \dots, g_{n+1}) \\
&\quad + (-1)^{n+1} \Phi^n(\varphi)(g_1, \dots, g_n) \\
&= d^n(\Phi^n(\varphi))(g_1, \dots, g_{n+1}),
\end{aligned}$$

which shows that $\Phi^{n+1} \circ d_{n+1}^* = d_n^* \circ \Phi^n$ as claimed.

□

Corollary 23.14. *Let A be a G-module. The cochain complexes $\mathcal{S}_A$ and $\mathcal{C}_A$ have the same cohomology groups, in other words, $H^n(G, A) \simeq \mathrm{Ext}^n_{\mathbb{Z}[G]}(\mathbb{Z}, A)$ for all $n \ge 0$, and we can compute $H^n(G, A)$ using any free resolution of $\mathbb{Z}$ by G-modules.*

Proof. This follows immediately from Proposition 23.13 and Theorem 23.71. □

Corollary 23.15. *For any G-modules A and B we have*

$$H^n(G, A \oplus B) \simeq H^n(G, A) \oplus H^n(G, B)$$

for all $n \ge 0$, and the isomorphism commutes with the natural inclusion and projection maps for the direct sums on both sides.

Proof. By Lemma 23.73, the functor $\mathrm{Ext}^n_{\mathbb{Z}[G]}(\mathbb{Z}, \bullet)$ is an additive functor. □

Definition 23.16. A category containing finite coproducts (such as direct sums) in which each set of morphisms between objects has the structure of an abelian group whose addition distributes over composition (and vice versa) is called an *additive category*. A functor F between additive categories is an *additive functor* if it maps zero objects to zero objects and satisfies $F(X \oplus Y) \simeq F(X) \oplus F(Y)$, where the isomorphism commutes with the natural inclusion and projection maps for the direct sums on both sides.

Definition 23.17. Let G be a group and let A be an abelian group. The abelian group

$$\mathrm{CoInd}^G(A) := \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}[G], A)$$

with G-action defined by $(g\varphi)(z) := \varphi(zg)$ is the *coinduced* G-module associated to A.

Warning 23.18. Some texts [3, 5] use $\mathrm{Ind}^G(A)$ instead of $\mathrm{CoInd}^G(A)$ to denote the G-module $\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}[G], A)$ and refer to it is as "induced" rather than "coinduced". Here we follow [1, 4, 7] and reserve the notation $\mathrm{Ind}^G(A)$ for the induced G-module $\mathbb{Z}[G] \otimes_{\mathbb{Z}} A$ defined below (see Definition 23.25). As shown by Lemma 23.27, this clash in terminology is fairly harmless when G is finite, since we then have $\mathrm{Ind}^G(A) \simeq \mathrm{CoInd}^G(A)$.

Lemma 23.19. *Let G be a group and A an abelian group. Then $H^0(G, \mathrm{CoInd}^G(A)) \simeq A$ and $H^n(G, \mathrm{CoInd}^G(A)) = 0$ for all $n \geq 1$.*

Proof. For all $n \geq 1$ we have an isomorphisms of abelian groups

$$\begin{aligned} \alpha\colon \mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^n], \mathrm{CoInd}^G(A)) &\xrightarrow{\sim} \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}[G^n], A) \\ \varphi &\mapsto (z \mapsto \varphi(z)(1)) \\ (z \mapsto (y \mapsto \phi(yz))) &\mapsfrom \phi \end{aligned}$$

Indeed,

$$\begin{aligned} \alpha(\alpha^{-1}(\phi)) &= \alpha(z \mapsto (y \mapsto \phi(yz))) = (z \mapsto \phi(z)) = \phi, \\ \alpha^{-1}(\alpha(\varphi)) &= \alpha^{-1}(z \mapsto \varphi(z)(1)) = (z \mapsto (y \mapsto \varphi(yz)(1))) = (z \mapsto \varphi(z)) = \varphi. \end{aligned}$$

Thus computing $H^n(G, \mathrm{CoInd}^G(A))$ using the standard resolution P of $\mathbb{Z}$ by G-modules is the same as computing $H^n(\{1\}, A)$ using the resolution P viewed as a resolution of $\mathbb{Z}$ by $\{1\}$-modules (abelian groups); note that $\mathbb{Z}[G^n]$ is also a free $\mathbb{Z}[\{1\}]$-module, and the G-module morphisms d_n in the standard resolution are also $\{1\}$-module morphisms (morphisms of abelian groups). Therefore $H^n(G, \mathrm{CoInd}^G(A)) \simeq H^n(\{1\}, A)$ for all $n \geq 0$.

But we can also compute $H^n(\{1\}, A)$ using the free resolution $\cdots \to 0 \to \mathbb{Z} \to \mathbb{Z} \to 0$, which implies $H^n(\{1\}, A) = 0$ for $n \geq 1$ and $H^0(\{1\}, A) \simeq \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}, A) \simeq A$. □

23.3 Homology via free resolutions

In the previous section we applied the contravariant functor $\mathrm{Hom}_{\mathbb{Z}[G]}(\bullet, A)$ to the truncation of the standard resolution of $\mathbb{Z}$ by G-modules to get a cochain complex with cohomology groups $H^n(G, A) \simeq \mathrm{Ext}^n_{\mathbb{Z}[G]}(\mathbb{Z}, A)$. If we do the same thing using the covariant functor $\bullet \otimes_{\mathbb{Z}[G]} A$ we get a *chain complex* (of $\mathbb{Z}$-modules)

$$\cdots \longrightarrow \mathbb{Z}[G^{n+1}] \otimes_{\mathbb{Z}[G]} A \xrightarrow{d_{n*}} \mathbb{Z}[G^n] \otimes_{\mathbb{Z}[G]} A \longrightarrow \cdots \longrightarrow \mathbb{Z}[G] \otimes_{\mathbb{Z}[G]} A \longrightarrow 0,$$

where d_{n*} is defined by $(g_0, \ldots, g_n) \otimes a \mapsto d_n(g_0, \ldots, g_n) \otimes a$. One minor technical point: in order for these tensor products to make sense we need to view $\mathbb{Z}[G^n]$ as a *right* $\mathbb{Z}[G]$-module, so we define $(g_1, \ldots, g_n) \cdot g := (g_1 g, \ldots, g_n g)$; the corresponding G-module is isomorphic to the left $\mathbb{Z}[G]$-module defined above (right action by g corresponds to left action by g^{-1}).

We then have *homology groups* $\ker d_{n*} / \operatorname{im} d_{n+1*}$. As with the groups $\mathrm{Ext}^n_{\mathbb{Z}[G]}(\mathbb{Z}, A)$, we get the same homology groups using any free resolution of $\mathbb{Z}$ by right $\mathbb{Z}[G]$-modules, and they are generically denoted $\mathrm{Tor}_n^{\mathbb{Z}[G]}(\mathbb{Z}, A)$; see Theorem 23.75.

Definition 23.20. Let A be a G-module. The nth *homology group of G with coefficients in A* is the abelian group $H_n(G,A) \coloneqq \operatorname{Tor}_n^{\mathbb{Z}[G]}(\mathbb{Z},A)$. If $\alpha\colon A \to B$ is a morphism of G-modules, the natural morphism $\alpha_n\colon H_n(G,A) \to H_n(G,B)$ is given by $x \otimes a \mapsto x \otimes \varphi(a)$. Each $H_n(G,\bullet)$ is a functor from the category of G-modules to the category of abelian groups.

The family of functors $H_n(G,\bullet)$ is a homological δ-functor.

Theorem 23.21. *Every short exact sequence of G-modules*

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0$$

induces a long exact sequence of homology groups

$$\cdots \longrightarrow H_1(G,C) \xrightarrow{\delta_0} H_0(G,A) \xrightarrow{\alpha_0} H_0(G,B) \xrightarrow{\beta_0} H_0(G,C) \longrightarrow 0,$$

and commutative diagrams of short exact sequences of G-modules induce corresponding commutative diagrams of long exact sequences of homology groups.

Proof. The proof is directly analogous to that of Theorem 23.8 (or see Theorem 23.50). □

As with $H^n(G,\bullet)$, the functors $H_n(G,\bullet)$ are additive functors.

Corollary 23.22. *For any G-modules A and B we have*

$$H_n(G, A \oplus B) \simeq H_n(G,A) \oplus H_n(G,B)$$

for all $n \geq 0$, and the isomorphism commutes with the natural inclusion and projection maps for the direct sums on both sides.

Proof. By Lemma 23.77, the functor $\operatorname{Tor}_n^{\mathbb{Z}[G]}(\mathbb{Z},\bullet)$ is an additive functor. □

For $n = 0$ we have

$$H_0(G,A) \coloneqq \operatorname{Tor}_0^{\mathbb{Z}[G]}(\mathbb{Z},A) = \mathbb{Z} \otimes_{\mathbb{Z}[G]} A,$$

where we are viewing $\mathbb{Z}$ as a (right) $\mathbb{Z}[G]$-module with G acting trivially; see Lemma 23.78 for a proof of the second equality. This means that $\sum a_g g \in \mathbb{Z}[G]$ acts on $\mathbb{Z}$ via multiplication by the integer $\sum a_g$. This motivates the following definition.

Definition 23.23. Let G be a group. The *augmentation map* $\varepsilon\colon \mathbb{Z}[G] \to \mathbb{Z}$ is the ring homomorphism $\sum a_g g \mapsto \sum a_g$.[5] The *augmentation ideal* I_G is the kernel of the augmentation map; it is a free $\mathbb{Z}$-module with basis $\{g - 1 : g \in G\}$.

The augmentation ideal I_G is precisely the annihilator of the $\mathbb{Z}[G]$-module $\mathbb{Z}$; therefore

$$\mathbb{Z} \otimes_{\mathbb{Z}[G]} A \simeq A/I_G A.$$

Definition 23.24. Let A be a G-module. The group of *G-coinvariants* of A is the G-module

$$A_G \coloneqq A/I_G A;$$

it is the largest trivial G-module that is a quotient of A.

[5] The augmentation map is the boundary map d_0 in the standard resolution of $\mathbb{Z}$ by G-modules.

We thus have $H_0(G,A) \simeq A_G$ and $H^0(G,A) \simeq A^G$.

Definition 23.25. Let G be a group and let A be an abelian group. The abelian group

$$\mathrm{Ind}^G(A) \coloneqq \mathbb{Z}[G] \otimes_{\mathbb{Z}} A$$

with G-action defined by $g(z \otimes a) = (gz) \otimes a$ is the *induced* G-module associated to A.

Lemma 23.26. *Let G be a group and A an abelian group. Then $H_0(G, \mathrm{Ind}^G(A)) \simeq A$ and $H_n(G, \mathrm{Ind}^G(A)) = 0$ for all $n \geq 1$.*

Proof. Viewing $\mathbb{Z}[G^n]$ as a right $\mathbb{Z}[G]$-module and $\mathbb{Z}[G]$ as a left $\mathbb{Z}[G]$-module, for all $n \geq 1$,

$$\mathbb{Z}[G^n] \otimes_{\mathbb{Z}[G]} (\mathbb{Z}[G] \otimes_{\mathbb{Z}} A) \simeq (\mathbb{Z}[G^n] \otimes_{\mathbb{Z}[G]} \mathbb{Z}[G]) \otimes_{\mathbb{Z}} A \simeq \mathbb{Z}[G^n] \otimes_{\mathbb{Z}} A,$$

by associativity of the tensor product (and the fact that $M \otimes_R R \simeq M$ for any right R-module M). This implies that computing $H_n(G, \mathrm{Ind}^G(A))$ using the standard resolution P of $\mathbb{Z}$ by (right) G-modules is the same as computing $H_n(\{1\}, A)$ using the resolution P viewed as a resolution of $\mathbb{Z}$ by $\{1\}$-modules (abelian groups). Thus

$$H_n(G, \mathrm{Ind}^G(A)) = \mathrm{Tor}_n^{\mathbb{Z}[G]}(\mathbb{Z}, \mathrm{Ind}^G(A)) \simeq \mathrm{Tor}_n^{\mathbb{Z}}(\mathbb{Z}, A) = H_n(\{1\}, A).$$

But we can also compute $H_n(\{1\}, A)$ using the free resolution $\cdots \to 0 \to \mathbb{Z} \to \mathbb{Z} \to 0$, which implies $H_n(\{1\}, A) = 0$ for $n \geq 1$ and $H_0(\{1\}, A) \simeq \mathbb{Z} \otimes A \simeq A$. □

Lemma 23.27. *Let G be a finite group and A an abelian group. The G-modules $\mathrm{Ind}^G(A)$ and $\mathrm{CoInd}^G(A)$ are isomorphic.*

Proof. We claim that we have a canonical G-module isomorphism given by

$$\begin{aligned} \alpha\colon \mathrm{CoInd}^G(A) &\xrightarrow{\sim} \mathrm{Ind}^G(A) \\ \varphi &\mapsto \sum_{g \in G} g^{-1} \otimes \varphi(g) \\ (g^{-1} \mapsto a) &\leftarrow\!\shortmid g \otimes a \end{aligned}$$

where $(g^{-1} \mapsto a)(h) = 0$ for $h \in G - \{g^{-1}\}$. It is obvious that α and α^{-1} are inverse homomorphisms of abelian groups, we just need to check that there are morphisms of G-modules. For any $h \in G$ and $\varphi \in \mathrm{CoInd}^G(A)$ we have

$$\alpha(h\varphi) = \sum_{g \in G} g^{-1} \otimes (h\varphi)(g) = h \sum_{g \in G} (gh)^{-1} \otimes \varphi(gh) = h \sum_{g \in G} g^{-1} \otimes \varphi(g) = h\alpha(\varphi),$$

and for any $h \in G$ and $g \otimes a \in \mathrm{Ind}^G(A)$ we have

$$\alpha^{-1}(h(g \otimes a)) = \alpha^{-1}(hg \otimes a) = ((hg)^{-1} \mapsto a) = h(g^{-1} \mapsto a) = h\alpha^{-1}(g \otimes a),$$

since for $\varphi = (g^{-1} \mapsto a)$ the identity $(h\varphi)(z) = \varphi(zh)$ implies $h\varphi = ((hg)^{-1} \mapsto a)$. □

Corollary 23.28. *Let G be a finite group, A be an abelian group, and let B be $\mathrm{Ind}^G(A)$ or $\mathrm{CoInd}^G(A)$. Then $H_0(G,B) \simeq H^0(G,B) \simeq A$ and $H_n(G,B) = H^n(G,B) = 0$ for all $n \geq 1$.*

Proof. This follows immediately from Lemmas 23.19, 23.26, 23.27. □

23.4 Tate cohomology

We now assume that G is a finite group.

Definition 23.29. The *norm element* of $\mathbb{Z}[G]$ is $N_G := \sum_{g\in G} g$.

Lemma 23.30. *Let A be a G-module and let $N_G\colon A \to A$ be the G-module endomorphism $a \mapsto N_G a$. We then have $I_G A \subseteq \ker N_G$ and $\operatorname{im} N_G \subseteq A^G$, thus N_G induces a morphism $\hat{N}_G\colon A_G \to A^G$ of trivial G-modules.*

Proof. We have $gN_G = N_G$ for all $g \in G$, so $\operatorname{im} N_G \subseteq A^G$, and $N_G(g-1) = 0$ for all $g \in G$, so N_G annihilates the augmentation ideal I_G and $I_G A \subseteq \ker N_G$. The lemma follows. □

Definition 23.31. Let A be a G-module for a finite group G. For $n \geq 0$ the *Tate cohomology and homology groups* are defined by

$$\hat{H}^n(G,A) := \begin{cases} \operatorname{coker} \hat{N}_G & \text{for } n = 0 \\ H^n(G,A) & \text{for } n > 0 \end{cases} \qquad \hat{H}_n(G,A) := \begin{cases} \ker \hat{N}_G & \text{for } n = 0 \\ H_n(G,A) & \text{for } n > 0 \end{cases}$$

$$\hat{H}^{-n}(G,A) := \hat{H}_{n-1}(G,A) \qquad \hat{H}_{-n}(G,A) := \hat{H}^{n-1}(G,A).$$

Note that $\hat{H}^0(G,A)$ is a quotient of $H^0(G,A) \simeq A^G$ (the largest trivial G-module in A) and $\hat{H}_0(G,A)$ is a submodule of $H_0(G,A) \simeq A_G$ (the largest trivial G-module quotient of A).

Thus any morphism of G-modules induces natural morphisms of Tate cohomology and homology groups in degree $n = 0$ (and all other degrees, as we already know). We thus have functors $\hat{H}^n(G,\bullet)$ and $\hat{H}_n(G,\bullet)$ from the category of G-modules to the category of abelian groups.

Given that every Tate homology group is also a Tate cohomology group, in practice one usually refers only to the groups $\hat{H}^n(G,A)$, but the notation $\hat{H}_n(G,A)$ can be helpful to highlight symmetry.

Theorem 23.32. *Let G be a finite group. Every short exact sequence of G-modules*

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0$$

induces a long exact sequence of Tate cohomology groups

$$\cdots \longrightarrow \hat{H}^n(G,A) \xrightarrow{\hat{\alpha}^n} \hat{H}^n(G,B) \xrightarrow{\hat{\beta}^n} \hat{H}^n(G,C) \xrightarrow{\hat{\delta}^n} \hat{H}^{n+1}(G,A) \longrightarrow \cdots,$$

equivalently, a long exact sequence of Tate homology groups

$$\cdots \longrightarrow \hat{H}_n(G,A) \xrightarrow{\hat{\alpha}_n} \hat{H}_n(G,B) \xrightarrow{\hat{\beta}_n} \hat{H}_n(G,C) \xrightarrow{\hat{\delta}_n} \hat{H}_{n-1}(G,A) \longrightarrow \cdots.$$

Commutative diagrams of short exact sequences of G-modules induce commutative diagrams of long exact sequences of Tate cohomology groups (equivalently, Tate homology groups).

Proof. It follows from Theorems 23.8 and 23.21 that it is enough to prove exactness at the terms $\hat{H}^0(G,\bullet) = \hat{H}_{-1}(G,\bullet)$ and $\hat{H}_0(G,\bullet) = \hat{H}^{-1}(G,\bullet)$. We thus consider the diagram

$$\begin{array}{ccccccccc} H_1(C,G) & \xrightarrow{\delta_0} & A_G & \xrightarrow{\alpha_0} & B_G & \xrightarrow{\beta_0} & C_G & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \hat{N}_G} & & \downarrow{\scriptstyle \hat{N}_G} & & \downarrow{\scriptstyle \hat{N}_G} & & \\ 0 & \longrightarrow & A^G & \xrightarrow{\alpha^0} & B^G & \xrightarrow{\beta^0} & C^G & \xrightarrow{\delta^0} & H^1(A,G) \end{array}$$

whose top and bottom rows are the end and beginning of the long exact sequences in homology and cohomology given by Theorems 23.21 and 23.8, respectively; here we are using $H_0(G,\bullet) \simeq \bullet_G$ and $H^0(G,\bullet) \simeq \bullet^G$.

For any $[a] \in A_G = A/I_G A$ we have $\hat{N}_G(\alpha_0([a])) = N_G\alpha(a) = \alpha(N_G a) = \alpha^0(\hat{N}_G([a]))$, so the first square commutes, as does the second (by the same argument). Applying the snake lemma yields an exact sequence of kernels and cokernels of $\hat{N}_G$

$$\hat{H}_0(G,A) \xrightarrow{\hat{\alpha}_0} \hat{H}_0(G,B) \xrightarrow{\hat{\beta}_0} \hat{H}_0(G,C) \xrightarrow{\hat{\delta}} \hat{H}^0(G,A) \xrightarrow{\hat{\alpha}^0} \hat{H}^0(G,B) \xrightarrow{\hat{\beta}^0} \hat{H}^0(G,C),$$

where $\hat{\delta}([c]) = [a]$ for any $a \in A$, $b \in B$, $c \in C$ with $\alpha(a) = N_G b$ and $\beta(b) = c \in \ker N_G$ (that this uniquely defines the connecting homomorphism $\hat{\delta}$ is part of the snake lemma). Note that $\operatorname{im}\delta_0 = \ker\alpha_0 = \ker\hat{\alpha}_0 \subseteq \ker\hat{N}_G$, since α^0 is injective, so δ_0 gives a well-defined map $\hat{\delta}_0\colon \hat{H}_1(G,C) \to \hat{H}_0(G,A)$ that makes the sequence is exact at $\hat{H}_0(G,A)$. Similarly, $\operatorname{im}\hat{N}_G \subseteq \operatorname{im}\beta^0 = \ker\delta^0$, since β_0 is surjective, so δ^0 induces a well-defined map $\hat{\delta}^0\colon \hat{H}^0(G,C) \to H^1(A,G)$ that makes the sequence exact at $\hat{H}^0(G,C)$.

For the last statement of the theorem, suppose we have the following commutative diagram of exact sequences of G-modules

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \xrightarrow{\alpha} & B & \xrightarrow{\beta} & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle\phi} & & \downarrow{\scriptstyle\psi} & & \downarrow{\scriptstyle\varphi} & & \\ 0 & \longrightarrow & A' & \xrightarrow{\alpha'} & B' & \xrightarrow{\beta'} & C' & \longrightarrow & 0 \end{array}$$

By Theorems 23.21 and 23.8, we only need to verify the commutativity of the square

$$\begin{array}{ccc} \hat{H}_0(G,C) & \xrightarrow{\hat{\delta}} & \hat{H}^0(G,A) \\ \downarrow{\scriptstyle\varphi_0} & & \downarrow{\scriptstyle\phi^0} \\ \hat{H}_0(G,C') & \xrightarrow{\hat{\delta}'} & \hat{H}^0(G,A') \end{array}$$

Let $a \in A$, $b \in B$, $c \in C$ satisfy $\alpha(a) = N_G b$ and $\beta(b) = c \in \ker N_G$ as in the definition of $\hat{\delta}$ above, so that $\hat{\delta}([c]) = [a]$. Now let $a' = \phi(a)$, $b' = \psi(b)$, $c = \varphi(c)$. Then

$$\begin{aligned} \alpha'(a') &= \alpha'(\phi(a)) = \psi(\alpha(a)) = \psi(N_G b) = N_G\psi(b) = N_G b' \\ \beta'(b') &= \beta'(\psi(b)) = \varphi(\beta(b)) = \varphi(c) = c' \in \ker N_G, \end{aligned}$$

where we have used $N_G c' = N_G\varphi(c) = \varphi(N_G c) = \varphi(0) = 0$. Thus $\hat{\delta}'([c']) = [a']$ and

$$\phi^0(\hat{\delta}([c])) = \phi^0([a]) = [\phi(a)] = [a'] = \hat{\delta}'([c']) = \hat{\delta}'([\varphi(c)]) = \hat{\delta}'(\varphi_0([c])),$$

so $\phi^0 \circ \hat{\delta} = \hat{\delta}' \circ \varphi_0$ as desired. □

Theorem 23.32 implies that the family $\hat{H}^n(G,\bullet)$ is a cohomological δ-functor, and that the family $\hat{H}_n(G,\bullet)$ is a homological δ-functor.

Corollary 23.33. *Let G be a finite group. For any G-modules A and B we have*

$$\hat{H}^n(G, A \oplus B) \simeq \hat{H}^n(G,A) \oplus \hat{H}^n(G,B),$$

for all $n \in \mathbb{Z}$, and the isomorphisms commute with the natural inclusion and projection maps for the direct sums on both sides.

Proof. For $n \neq 0, -1$ this follows from Corollaries 23.15 and 23.22. For $n = 0, -1$ it suffices to note that N_G acts on $A \oplus B$ component-wise, and the induced morphism $\hat{N}_G$ thus acts on $(A \oplus B)_G = A_G \oplus B_G$ component-wise. □

Theorem 23.34. *Let G be a finite group and let B be an induced or co-induced G-module associated to some abelian group A. Then $\hat{H}^n(G, B) = \hat{H}_n(G, B) = 0$ for all $n \in \mathbb{Z}$.*

Proof. By Corollary 23.28, we only need to show $\hat{H}_0(G, B) = \hat{H}^0(G, B) = 0$, and by Lemma 23.27 it suffices to consider the case $B = \mathrm{Ind}^G(A) = \mathbb{Z}[G] \otimes_{\mathbb{Z}} A$. Equivalently, we need to show that $N_G \colon B \to B$ has kernel $I_G B$ and image B^G. By definition, the $\mathbb{Z}[G]$-action on $B = \mathbb{Z}[G] \otimes_{\mathbb{Z}} A$ only affects the factor $\mathbb{Z}[G]$, so this amounts to showing that, as an endomorphism of $\mathbb{Z}[G]$, we have $\ker N_G = I_G$ and $\operatorname{im} N_G = \mathbb{Z}[G]^G$. But this is clear: the action of N_G on $\mathbb{Z}[G]$ is $\sum_{g \in G} a_g g \mapsto (\sum_{g \in G} a_g) N_G$. The kernel of this action is the augmentation ideal I_G, and its image is $\mathbb{Z}[G]^G = \{\sum_{g \in G} a_g g : \text{all } a_g \in \mathbb{Z} \text{ equal}\} = N_G \mathbb{Z}$. □

Remark 23.35. Theorem 23.34 explains a major motivation for using Tate cohomology. It is the minimal modification needed to ensure that induced (and co-induced) G-modules have trivial homology and cohomology in all degrees.

Corollary 23.36. *Let G be a finite group and let A be a free $\mathbb{Z}[G]$-module. Then $\hat{H}_n(G, A) = \hat{H}^n(G, A) = 0$ for all $n \in \mathbb{Z}$.*

Proof. Let S be a $\mathbb{Z}[G]$-basis for A and let B be the free $\mathbb{Z}$-module with basis S. Then $A \simeq \mathrm{Ind}^G(B)$ and the corollary follows from Theorem 23.34. □

23.5 Tate cohomology of cyclic groups

We now assume that G is a cyclic group $\langle g \rangle$ of finite order. In this case the augmentation ideal I_G is principal, generated by $g - 1$ (as an ideal in the ring $\mathbb{Z}[G]$, not as a $\mathbb{Z}$-module). For any G-module A we have a free resolution

$$\cdots \longrightarrow \mathbb{Z}[G] \xrightarrow{N_G} \mathbb{Z}[G] \xrightarrow{g-1} \mathbb{Z}[G] \xrightarrow{N_G} \mathbb{Z}[G] \xrightarrow{g-1} \mathbb{Z}[G] \xrightarrow{\varepsilon} \mathbb{Z} \longrightarrow 0. \tag{2}$$

The fact that augmentation ideal $I_G = (g - 1)$ is principal (because G is cyclic) ensures $\operatorname{im} N_G = \ker(g - 1)$, making the sequence exact.

The group ring $\mathbb{Z}[G]$ is commutative, since G is abelian, so we need not distinguish left and right $\mathbb{Z}[G]$-modules. For any G-module A we can view $\mathbb{Z}[G] \otimes_{\mathbb{Z}[G]} A$ as a G-module via $g(h \otimes a) = gh \otimes a = h \otimes ga$ and view $\mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G], A)$ as a G-module via $(g\varphi)(h) := \varphi(gh)$.[6]

Theorem 23.37. *Let $G = \langle g \rangle$ be a finite cyclic group and let A be a G-module. For all $n \in \mathbb{Z}$ we have $\hat{H}^{2n}(G, A) \simeq \hat{H}_{2n-1}(G, A) \simeq \hat{H}^0(G, A)$ and $\hat{H}_{2n}(G, A) \simeq \hat{H}^{2n-1}(G, A) \simeq \hat{H}_0(G, A)$.*

Proof. We have canonical G-module isomorphisms $\mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G], A) \simeq A \simeq \mathbb{Z}[G] \otimes_{\mathbb{Z}[G]} A$ induced by $\varphi \mapsto \varphi(1)$ and $a \mapsto 1 \otimes a$, respectively, and these isomorphisms preserve the multiplication-by-g endomorphisms (that is, $(g\varphi)(1) = g\varphi(1)$ and $1 \otimes ga = g(1 \otimes a)$). Using the free resolution in (2), we can thus compute $H^n(G, A)$ using the cochain complex

$$0 \longrightarrow A \xrightarrow{g-1} A \xrightarrow{N_G} A \xrightarrow{g-1} A \xrightarrow{N_G} A \cdots,$$

[6]Note that we must have $g_1 g_2 \varphi(h) = g_1(g_2\varphi)(h) = (g_2\varphi)(g_1 h) = \varphi(g_2 g_1 h) = g_2 g_1 \varphi(h)$ in order for φ to be both a $\mathbb{Z}[G]$-module morphism and an element of a $\mathbb{Z}[G]$-module, so this will not work if G is not abelian.

and we can compute $H_n(G,A)$ using the chain complex

$$\cdots \longrightarrow A \xrightarrow{N_G} A \xrightarrow{g-1} A \xrightarrow{N_G} A \xrightarrow{g-1} A \longrightarrow 0.$$

We now observe that $A^G = \ker(g-1)$, so for all $n \geq 1$ we have

$$H^{2n}(G,A) = H_{2n-1}(G,A) = \ker(g-1)/\operatorname{im} N_G = \operatorname{coker} \hat{N}_G = \hat{H}^0(G,A),$$

so $\hat{H}^{2n}(G,A) = \hat{H}_{2n-1}(G,A) = \hat{H}^0(G,A)$ for all $n \in \mathbb{Z}$, since $\hat{H}^{-2n}(G,A) = \hat{H}_{2n-1}(G,A)$ and $\hat{H}_{-2n+1} = \hat{H}^{2n}$ for all $n \geq 0$.

We also note that $\operatorname{im}(g-1) = I_G A$, so for all $n \geq 1$ we have

$$H_{2n}(G,A) = H^{2n-1}(G,A) = \ker N_G/\operatorname{im}(g-1) = \ker \hat{N}_G = \hat{H}_0(G,A),$$

so $\hat{H}_{2n}(G,A) = \hat{H}^{2n-1}(G,A) = \hat{H}_0(G,A)$ for all $n \in \mathbb{Z}$, since $\hat{H}_{-2n}(G,A) = \hat{H}^{2n-1}(G,A)$ and $\hat{H}^{-2n+1} = \hat{H}_{2n}$ for all $n \geq 0$. □

It follows from Theorem 23.37 that when G is a finite cyclic group, all of the Tate homology/cohomology groups are determined by $\hat{H}_0(G,A) = \ker \hat{N}_G = \ker N_G/\operatorname{im}(g-1)$ and $\hat{H}^0(G,A) = \operatorname{coker} \hat{N}_G = \ker(g-1)/\operatorname{im} N_G$. This motivates the following definition.

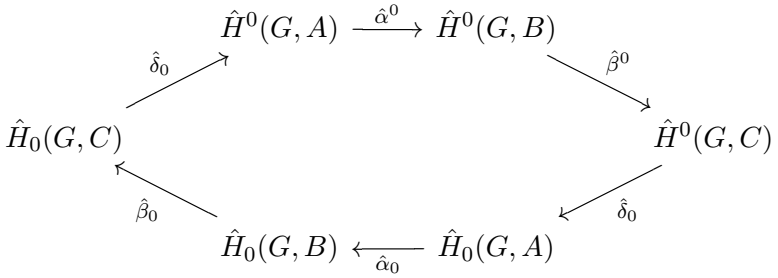
Definition 23.38. Let G be a finite cyclic group and let A be a G-module. We define $h^n(A) \coloneqq h^n(G,A) \coloneqq \#\hat{H}^n(G,A)$ and $h_n(A) \coloneqq h_n(G,A) \coloneqq \#\hat{H}_n(G,A)$. Whenever $h^0(A)$ and $h_0(A)$ are both finite, we also define the *Herbrand quotient* $h(A) \coloneqq h^0(A)/h_0(A) \in \mathbb{Q}$.

Remark 23.39. Some authors define the Herbrand quotient via $h(A) \coloneqq h^0(A)/h^1(A)$ or $h(A) \coloneqq h^0(A)/h^{-1}(A)$ or $h(A) \coloneqq h^2(A)/h^1(A)$, but Theorem 23.37 implies that these definitions are all the same as ours. The notation $q(A)$ is often used instead of $h(A)$, and one occasionally sees the Herbrand quotient defined as the reciprocal of our definition (as in [2], for example), but this is less standard.

Corollary 23.40. *Let G be a finite cyclic group. Given an exact sequence of G-modules*

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0$$

we have a corresponding exact hexagon



Proof. This follows immediately from Theorems 23.32 and 23.37. □

Corollary 23.41. *Let G be a finite cyclic group. For any exact sequence of G-modules*

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0,$$

if any two of $h(A), h(B), h(C)$ are defined then so is the third and $h(B) = h(A)h(C)$.

Proof. Using the exact hexagon given by Corollary 23.40 we can compute the cardinality

$$h^0(A) = \#\hat{H}^0(G,A) = \#\ker\hat{\alpha}^0\#\operatorname{im}\hat{\alpha}^0 = \#\ker\alpha^0\#\ker\beta^0.$$

Applying a similar calculation to $\hat{H}^0(G,C)$ and $\hat{H}^1(G,B)$ yields

$$h^0(A)h^0(C)h_0(B) = \#\ker\hat{\alpha}^0\#\ker\hat{\beta}^0\#\ker\hat{\delta}^0\#\ker\hat{\alpha}_0\#\ker\hat{\beta}_0\#\ker\hat{\delta}_0.$$

Doing the same for $\hat{H}^0(G,B)$, $\hat{H}_0(G,A)$, $\hat{H}_0(G,C)$ yields

$$h^0(B)h_0(A)h_0(C) = \#\ker\hat{\beta}^0\#\ker\hat{\delta}^0\#\ker\hat{\alpha}_0\#\ker\hat{\beta}_0\#\ker\hat{\delta}_0\#\ker\hat{\alpha}^0 = h^0(A)h^0(C)h_0(B).$$

If any two of $h(A), h(B), h(C)$ are defined then at least four of the groups in the exact hexagon are finite, and the remaining two are non-adjacent, but these two must then also be finite. In this case we can rearrange the identity above to obtain $h(B) = h(A)h(C)$. □

Corollary 23.42. *Let G be a finite cyclic group, and let A and B be G-modules. If $h(A)$ and $h(B)$ are defined then so is $h(A\oplus B) = h(A)h(B)$.*

Proof. Apply Corollary 23.41 to the split exact sequence $0 \to A \to A\oplus B \to B \to 0$. □

Lemma 23.43. *Let $G=\langle g\rangle$ be a finite cyclic group. If A is an induced or finite G-module then $h(A) = 1$.*

Proof. If A is an induced G-module then $h_0(A) = h^0(A) = h(A) = 1$, by Theorem 23.34. If A is finite, then the exact sequence

$$0 \longrightarrow A^G \longrightarrow A \xrightarrow{g-1} A \longrightarrow A_G \longrightarrow 0$$

implies $\#A^G = \#\ker(g-1) = \#\operatorname{coker}(g-1) = \#A_G$, and therefore

$$h_0(A) = \#\ker\hat{N}_G = \#\operatorname{coker}\hat{N}_G = h^0(A),$$

so $h(A) = h^0(A)/h_0(A) = 1$. □

Corollary 23.44. *Let G be a finite cyclic group and let A be a G-module that is a finitely generated abelian group. Then $h(A) = h(A/A_{\mathrm{tor}})$ whenever either is defined.*

Proof. Apply Corollary 23.41 and Lemma 23.43 to $0 \to A_{\mathrm{tor}} \to A \to A/A_{\mathrm{tor}} \to 0$. □

Remark 23.45. The hypothesis of Corollary 23.44 actually guarantees that $h(A)$ is defined, but we won't prove this here.

Corollary 23.46. *Let G be a finite cyclic group and let A be a trivial G-module that is a finitely generated abelian group. Then $h(A) = (\#G)^r$, where r is the rank of A.*

Proof. We have $A/A_{\mathrm{tor}} \simeq \mathbb{Z}^r$, where $\mathbb{Z}$ is a trivial G-module. Then $\mathbb{Z}_G = \mathbb{Z} = \mathbb{Z}^G$, and $\hat{N}_G\colon \mathbb{Z}_G \to \mathbb{Z}^G$ is multiplication by $\#G$, so $h(\mathbb{Z}) = \#\operatorname{coker}\hat{N}_G/\#\ker\hat{N}_G = \#G$. Now apply Corollaries 23.42 and 23.44. □

Lemma 23.47. *Let G be a finite cyclic group and let $\alpha\colon A \to B$ be a morphism of G-modules with finite kernel and cokernel. If either $h(A)$ or $h(B)$ is defined then $h(A) = h(B)$.*

Proof. Applying Corollary 23.41 to the exact sequences

$$0 \to \ker \alpha \to A \to \operatorname{im} \alpha \to 0$$
$$0 \to \operatorname{im} \alpha \to B \to \operatorname{coker} \alpha \to 0$$

yields $h(A) = h(\ker \alpha)h(\operatorname{im} \alpha) = h(\operatorname{im} \alpha) = h(\operatorname{im} \alpha)h(\operatorname{coker} \alpha) = h(B)$, by Lemma 23.43, since $\ker \alpha$ and $\operatorname{coker} \alpha$ are finite. The lemma follows. □

Corollary 23.48. *Let G be a finite cyclic group and let A be a G-module containing a sub-G-module B of finite index. Then $h(A) = h(B)$ whenever either is defined.*

Proof. Apply Lemma 23.47 to the inclusion $B \to A$. □

23.6 A little homological algebra

In an effort to keep these notes self-contained, in this final section we present proofs of the homological results that were used above. For the sake of concreteness we restrict our attention to modules, but everything in this section generalizes to suitable abelian categories. We use R to denote an arbitrary (not necessarily commutative) ring (in previous section R was the group ring $\mathbb{Z}[G]$). Statements that use the term R-module without qualification are understood to apply in both the category of left R-modules and the category of right R-modules.

23.6.1 Complexes

Definition 23.49. A *chain complex* C is a sequence of R-module morphisms

$$\cdots \xrightarrow{d_2} C_2 \xrightarrow{d_1} C_1 \xrightarrow{d_0} C_0 \longrightarrow 0,$$

with $d_n \circ d_{n+1} = 0$; the d_n are *boundary maps*. The *nth homology group* of C is the R-module $H_n(C) := Z_n(C)/B_n(C)$, where $Z_n(C) := \ker d_{n-1}$ and $B_n(C) := \operatorname{im} d_n$ are the R-modules of *cycles* and *boundaries*, respectively; for $n < 0$ we define $C_n = 0$ and d_n is the zero map.

A morphism of chain complexes $f\colon C \to D$ is a sequence of R-module morphisms $f_n\colon C_n \to D_n$ that commute with boundary maps (so $f_n \circ d_n = d_n \circ f_{n+1}$).[7] Such a morphism necessarily maps cycles to cycles and boundaries to boundaries, yielding natural morphisms $H_n(f)\colon H_n(C) \to H_n(D)$ of homology groups.[8] We thus have a family of functors $H_n(\bullet)$ from the category of chain complexes to the category of abelian groups. The category of chain complexes has kernels and cokernels (and thus exact sequences). The set $\operatorname{Hom}(C, D)$ of morphisms of chain complexes $C \to D$ is an abelian group under addition: $(f+g)_n = f_n + g_n$.

The category of chain complexes of R-modules contains direct sums and direct products: if A and B are chain complexes of R-modules then $(A \oplus B)_n := A_n \oplus B_n$ and the boundary maps $d_n\colon (A \oplus B)_{n+1} \to (A \oplus B)_n$ are defined component-wise: $d_n(a \oplus b) := d_n(a) \oplus d_n(b)$. Because the boundary maps are defined component-wise, the kernel of the boundary map of

[7] We use the symbols d_n to denote boundary maps of both C and D; in general, the domain and codomain of any boundary or coboundary map should be inferred from context.

[8] In fact $H_n(f)\colon H_n(C) \to H_n(D)$ is a morphism of R-modules, but in all the cases of interest to us, either the homology groups are all trivial (as occurs for exact chain complexes, such as the standard resolution of $\mathbb{Z}$ by $\mathbb{Z}[G]$-modules), or $R = \mathbb{Z}$ (as in the chain complexes used to define the Ext and Tor groups below), so we will generally refer to homology groups rather than homology modules.

a direct sum is the direct sum of the kernels of the boundary maps on the components, and similarly for images. It follows that $H_n(A \oplus B) \simeq H_n(A) \oplus H_n(B)$, and this isomorphism commutes with the natural inclusion and projection maps in to and out of the direct sums on both sides. In other words, $H_n(\bullet)$ is an additive functor (see Definition 23.16). This extends to arbitrary (possibly infinite) direct sums, and also to arbitrary direct products, although we will only be concerned with finite direct sums/products.[9]

Theorem 23.50. *Associated to each short exact sequence of chain complexes*

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0$$

is a long exact sequence of homology groups

$$\cdots \longrightarrow H_{n+1}(A) \xrightarrow{H_{n+1}(\alpha)} H_{n+1}(B) \xrightarrow{H_{n+1}(\beta)} H_{n+1}(C) \xrightarrow{\delta_n} H_n(A) \xrightarrow{H_n(\alpha)} H_n(B) \xrightarrow{H_n(\beta)} H_n(C) \longrightarrow \cdots$$

and this association maps morphisms of short exact sequences to morphisms of long exact sequences. In other words, the family of functors $H_n(\bullet)$ is a homological δ-functor.

For $n < 0$ we have $H_n(\bullet) = 0$, by definition, so this sequence ends at $H_0(C) \to 0$.

Proof. For any chain complex C, let $Y_n(C) \coloneqq C_n/B_n(C)$. Applying the snake lemma to

$$\begin{array}{ccccccc} & & Y_{n+1}(A) & \xrightarrow{\alpha_{n+1}} & Y_{n+1}(B) & \xrightarrow{\beta_{n+1}} & Y_{n+1}(C) \longrightarrow 0 \\ & & \downarrow{\scriptstyle d_n} & & \downarrow{\scriptstyle d_n} & & \downarrow{\scriptstyle d_n} \\ 0 & \longrightarrow & Z_n(A) & \xrightarrow{\alpha_n} & Z_n(B) & \xrightarrow{\beta_n} & Z_n(C) \end{array}$$

(where α_n, β_n, d_n denote obviously induced maps) yields the exact sequence

$$H_{n+1}(A) \xrightarrow{\alpha_{n+1}} H_{n+1}(B) \xrightarrow{\beta_{n+1}} H_{n+1}(C) \xrightarrow{\delta_n} H_n(A) \xrightarrow{\alpha_n} H_n(B) \xrightarrow{\beta_n} H_n(G).$$

The verification of the commutativity of diagrams of long exact sequences of homology groups associated to commutative diagrams of short exact sequences of chain complexes is as in the proof of Theorem 23.8, *mutatis mutandi*. □

Definition 23.51. Two morphisms $f, g \colon C \to D$ of chain complexes are *homotopic* if there exist morphisms $h_n \colon C_n \to D_{n+1}$ such that $f_n - g_n = d_n \circ h_n + h_{n-1} \circ d_{n-1}$ for all $n \geq 0$ (where $h_{-1} \coloneqq 0$); this defines an equivalence relation $f \sim g$, since (a) $f \sim f$ (take $h = 0$), (b) if $f \sim g$ via h then $g \sim f$ via $-h$, and (c) if $f_1 \sim f_2$ via h_1 and $f_2 \sim f_3$ via h_2 then $f_1 \sim f_3$ via $h_1 + h_2$.

Lemma 23.52. *Homotopic morphisms of chain complexes $f, g \colon C \to D$ induce they some morphisms of homology groups $H_n(C) \to H_n(D)$; we have $H_n(f) = H_n(g)$ for all $n \geq 0$.*

Proof. Let $[z] \in H_n(C) = Z_n(C)/B_n(C)$ denote the homology class $z \in Z_n(C)$. We have

$$f_n(z) - g_n(z) = d_n(h_n(z)) + h_{n-1}(d_{n-1}(z)) = d_n(h_n(z)) + 0 \in B_n(D),$$

thus $H_n(f)([z]) - H_n(g)([z]) = 0$. It follows that $H_n(f) = H_n(g)$ for all $n \geq 0$. □

[9] This does not imply that the Ext and Tor functors defined below commute with arbitrary direct sums and direct products; see Remarks 23.62 and 23.66.

Definition 23.53. A *cochain complex* C is a sequence of R-module morphisms

$$0 \longrightarrow C^0 \xrightarrow{d^0} C^1 \xrightarrow{d^1} C^2 \xrightarrow{d^2} \cdots$$

with $d^{n+1} \circ d^n = 0$. The *nth cohomology group* of C is the R-module $H^n(C) := Z^n(C)/B^n(C)$, where $Z^n(C) := \ker d^n$ and $B^n(C) := \operatorname{im} d^{n-1}$ are the R-modules of *cocycles* and *coboundaries*; for $n < 0$ we define $C^n = 0$ and d^n is the zero map. A morphism of cochain complexes $f\colon C \to D$ consists of R-module morphisms $f^n\colon C^n \to D^n$ that commute with coboundary maps, yielding natural morphisms $H^n(f)\colon H^n(C) \to H^n(D)$ and a functors $H^n(\bullet)$ from the category of cochain complexes to the category of abelian groups. Cochain complexes form a category with kernels and cokernels, as well as direct sums and direct products (coboundary maps are defined component-wise). Like $H_n(\bullet)$, the functor $H^n(\bullet)$ is additive and commutes with arbitrary direct sums and direct products.

The set $\operatorname{Hom}(C, D)$ of morphisms of cochain complexes $C \to D$ forms an abelian group under addition: $(f+g)^n = f^n + g^n$. Morphisms of cochain complexes $f, g\colon C \to D$ are *homotopic* if there are morphisms $h^n\colon C^{n+1} \to D^n$ such that $f^n - g^n = h^n \circ d^n + d^{n-1} \circ h^{n-1}$ for all $n \geq 0$ (where $h^{-1} := 0$); this defines an equivalence relation $f \sim g$.[10]

Theorem 23.54. *Associated to every short exact sequence of cochain complexes*

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0$$

is a long exact sequence of homology groups

$$\cdots \longrightarrow H^n(A) \xrightarrow{H^n(\alpha)} H^n(B) \xrightarrow{H^n(\beta)} H^n(C) \xrightarrow{\delta^n} H^{n+1}(A) \xrightarrow{H^{n+1}(\alpha)} H^{n+1}(B) \xrightarrow{H^{n+1}(\beta)} H^{n+1}(C) \longrightarrow \cdots$$

and this association maps morphisms of short exact sequences of morphisms of long exact sequences, that is, the family of functors $H^n(\bullet)$ is a cohomological δ-functor.

For $n < 0$ we have $H_n(\bullet) = 0$, by definition, so this sequence begins with $0 \to H^0(A)$.

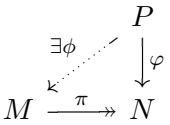
Proof. Adapt the proof of Theorem 23.50. □

Lemma 23.55. *Homotopic morphisms of cochain complexes $f, g\colon C \to D$ induce the same morphisms of cohomology groups $H^n(C) \to H^n(D)$; we have $H^n(f) = H^n(g)$ for all $n \geq 0$.*

Proof. Adapt the proof of Lemma 23.52. □

23.6.2 Projective resolutions

Recall that a *projective* R-module is an R-module P with the property that if $\pi\colon M \twoheadrightarrow N$ is a surjective morphism of R-modules, every R-module morphism $\varphi\colon P \to N$ factors through π:



Free modules are projective, since we can then fix an R-basis $\{e_i\}$ for P and define $\phi(e_i)$ by picking any element of $\pi^{-1}(\varphi(e_i))$ (note that the ϕ so constructed is in no way canonical).

[10] Note the order of composition in the homotopy relations for morphisms of chain/cochain complexes.

Definition 23.56. Let M be an R-module. A *projective resolution* of M is an exact chain complex P with $P_0 = M$ and P_n projective for all $n > 0$.

Every R-module has a projective resolution, since (as noted earlier), every R-module M has a free resolution (we can always construct $d_0 \colon P_1 \twoheadrightarrow M$ by taking P_1 to be free module with basis M, then similarly construct $d_1 \colon P_2 \twoheadrightarrow \ker d_0$, and so on).

Proposition 23.57. *Let M and N be R-modules with projective resolutions P and Q, respectively. Every R-module morphism $\alpha_0 \colon M \to N$ extends to a morphism $\alpha \colon P \to Q$ of chain complexes that is unique up to homotopy.*

Proof. We inductively construct α_n for $n \geq 1$ (the base case is given). Suppose we have constructed a commutative diagram of exact sequences

$$\begin{array}{ccccccccccccc}
\cdots & \xrightarrow{d_{n+1}} & P_{n+1} & \xrightarrow{d_n} & P_n & \xrightarrow{d_{n-1}} & P_{n-1} & \xrightarrow{d_{n-2}} & \cdots & \xrightarrow{d_1} & P_1 & \xrightarrow{d_0} & M \longrightarrow 0 \\
 & & & & \downarrow{\scriptstyle \alpha_n} & & \downarrow{\scriptstyle \alpha_{n-1}} & & \downarrow{\scriptstyle \cdots} & & \downarrow{\scriptstyle \alpha_1} & & \downarrow{\scriptstyle \alpha_0} \quad \downarrow \\
\cdots & \xrightarrow{d_{n+1}} & Q_{n+1} & \xrightarrow{d_n} & Q_n & \xrightarrow{d_{n-1}} & P_{n-1} & \xrightarrow{d_{n-2}} & \cdots & \xrightarrow{d_1} & Q_1 & \xrightarrow{d_0} & N \longrightarrow 0
\end{array}$$

Then $d_{n-1} \circ \alpha_n \circ d_n = \alpha_{n-1} \circ d_{n-1} \circ d_n = 0$, so $\operatorname{im}(\alpha_n \circ d_n) \subseteq \ker d_{n-1} = \operatorname{im} d_n$. We now define $\alpha_{n+1} \colon P_{n+1} \to Q_{n+1}$ as a pullback of the morphism $\alpha_n \circ d_n \colon P_{n+1} \to \operatorname{im} d_n$ along the surjection $d_n \colon Q_{n+1} \to \operatorname{im} d_n$ such that $d_n \circ \alpha_{n+1} = \alpha_n \circ d_n$.

Now suppose $\beta \colon P \to Q$ is another morphism of projective resolutions with $\beta_0 = \alpha_0$, and let $\gamma = \alpha - \beta$. To show that α and β are homotopic it suffices to construct maps $h_n \colon P_n \to Q_{n+1}$ such that $d_n \circ h_n = \gamma_n - h_{n-1} \circ d_{n-1}$ (where $h_{-1} = d_{-1} = 0$). We have $\gamma_0 = \alpha_0 - \beta_0 = 0$, so let $h_0 \coloneqq 0$ and inductively assume $d_n \circ h_n = \gamma_n - h_{n-1} \circ d_{n-1}$. Then

$$d_n \circ (\gamma_{n+1} - h_n \circ d_n) = d_n \circ \gamma_{n+1} - (d_n \circ h_n) \circ d_n = \gamma_n \circ d_n - (\gamma_n - h_{n-1} \circ d_{n-1}) \circ d_n = 0,$$

so $\operatorname{im}(\gamma_{n+1} - h_n \circ d_n) \subseteq B_{n+1}(Q)$. The R-module P_{n+1} is projective, so we can pullback the morphism $(\gamma_{n+1} - h_n \circ d_n) \colon P_{n+1} \to B_{n+1}(Q)$ along the surjection $d_{n+1} \colon Q_{n+1} \to B_{n+1}(Q)$ to obtain h_{n+1} satisfying $d_{n+1} \circ h_{n+1} = \gamma_{n+1} - h_n \circ d_n$ as desired. □

23.6.3 Hom and Tensor

If M and N are R-modules, the set $\operatorname{Hom}_R(M, N)$ of R-module morphisms $M \to N$ forms an abelian group under pointwise addition (so $(f + g)(m) \coloneqq f(m) + g(m)$) that we may view as a $\mathbb{Z}$-module. For each R-module A we have a contravariant functor $\operatorname{Hom}_R(\bullet, A)$ that sends each R-module M to the $\mathbb{Z}$-module

$$M^* \coloneqq \operatorname{Hom}_R(M, A)$$

and each R-module morphism $\varphi \colon M \to N$ to the $\mathbb{Z}$-module morphism

$$\begin{aligned} \varphi^* \colon \operatorname{Hom}_R(N, A) &\to \operatorname{Hom}_R(M, A) \\ f &\mapsto f \circ \varphi. \end{aligned}$$

To check this, note that

$$\varphi^*(f + g) = (f + g) \circ \varphi = f \circ \varphi + g \circ \varphi = \varphi^*(f) + \varphi^*(g),$$

so φ^* is a morphism of $\mathbb{Z}$=modules (homomorphism of abelian groups), and

$$\begin{aligned}\mathrm{id}_M^* &= (f \mapsto f \circ \mathrm{id}_M) = (f \mapsto f) = \mathrm{id}_{M^*},\\ (\phi \circ \varphi)^* &= (f \mapsto f \circ \phi \circ \varphi) = (f \mapsto f \circ \varphi) \circ (f \mapsto f \circ \phi) = \varphi^* \circ \phi^*,\end{aligned}$$

thus $\mathrm{Hom}_R(\bullet, A)$ is a contravariant functor.

Lemma 23.58. *Let $\varphi\colon M \to N$ and $\phi\colon N \to P$ be morphisms of R-modules. The sequence*

$$M \xrightarrow{\varphi} N \xrightarrow{\phi} P \longrightarrow 0$$

is exact if and only if for every R-module A the sequence

$$0 \longrightarrow \mathrm{Hom}_R(P, A) \xrightarrow{\phi^*} \mathrm{Hom}_R(N, A) \xrightarrow{\varphi^*} \mathrm{Hom}_R(M, A)$$

is exact.

Proof. ($\Rightarrow$): If $\phi^*(f) = f \circ \phi = 0$ then $f = 0$, since ϕ is surjective, so ϕ^* is injective. We have $\varphi^* \circ \phi^* = (\varphi \circ \phi)^* = 0^* = 0$, so $\mathrm{im}\,\phi^* \subseteq \ker \varphi^*$. Let $\phi^{-1}\colon P \xrightarrow{\sim} N/\ker\phi$. Each $g \in \ker\varphi^*$ vanishes on $\mathrm{im}\,\varphi = \ker\phi$ inducing $\bar{g}\colon N/\ker\phi \to A$ with $g = \bar{g} \circ \phi^{-1} \circ \phi \in \mathrm{im}\,\phi^*$.

($\Leftarrow$): For $A = P/\mathrm{im}\,\phi$ and $\pi\colon P \to P/\mathrm{im}\,\phi$ the projective map, we have $\phi^*(\pi) = 0$ and therefore $\pi = 0$, since ϕ^* is injective, so $P = \mathrm{im}\,\phi$ and ϕ is surjective. For $A = P$ we have $0 = (\varphi^* \circ \phi^*)(\mathrm{id}_P) = \mathrm{id}_P \circ \phi \circ \varphi = \phi \circ \varphi$, so $\mathrm{im}\,\varphi \subseteq \ker\phi$. For $A = N/\mathrm{im}\,\varphi$, and $\pi\colon N \to N/\mathrm{im}\,\varphi$ the projection map, we have $\pi \in \ker\varphi^* = \mathrm{im}\,\phi^*$, thus $\pi = \phi^*(\sigma) = \sigma \circ \phi$ for some $\sigma \in \mathrm{Hom}(P, A)$. Now $\pi(\ker\phi) = \sigma(\phi(\ker\phi)) = 0$ implies $\ker\phi \subseteq \ker\pi = \mathrm{im}\,\varphi$. □

Definition 23.59. A sequence of morphisms $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ is *left exact* if it is exact at A and B ($\ker f = 0$ and $\mathrm{im}\,f = \ker g$), and *right exact* if it is exact at B and C ($\mathrm{im}\,f = \ker g$ and $\mathrm{im}\,g = C$). A functor that takes exact sequences to left (resp. right) exact sequences is said to be *left exact* (resp. *right exact*).

Corollary 23.60. *For any R-module A the functor $\mathrm{Hom}_R(\bullet, A)$ is left exact.*

Proof. This follows immediately from the forward implication in Lemma 23.58. □

Corollary 23.61. *For any R-module A, the functor $\mathrm{Hom}_R(\bullet, A)$ is an additive functor.*

Proof. See [6, Lemma 12.7.2] for a proof that this follows from left exactness; it is easy to check directly in any case. □

Remark 23.62. Corollary 23.61 implies that $\mathrm{Hom}_R(\bullet, A)$ commutes with finite direct sums, but it does *not* commute with infinite direct sums (direct products are fine).

Remark 23.63. The covariant functor $\mathrm{Hom}_R(A, \bullet)$ that sends $\varphi\colon M \to N$ to $(f \mapsto \varphi \circ f)$ is also left exact.

If M is a right R-module and A is a left R-module, the tensor product $M \otimes_R A$ is an abelian group consisting of sums of pure tensors $m \otimes a$ with $m \in M$ and $a \in A$ satisfying:

- $m \otimes (a + b) = m \otimes b + m \otimes b$;
- $(m + n) \otimes a = m \otimes a + m \otimes a$;
- $mr \otimes a = m \otimes ra$.

For each left R-module A we have a covariant functor $\bullet \otimes_R A$ that sends each right R-module M to the $\mathbb{Z}$-module

$$M_* := M \otimes_R A,$$

and each right R-module morphism $\varphi\colon M \to N$ to the $\mathbb{Z}$-module morphism

$$\begin{aligned} \varphi_*\colon M \otimes_R A &\to N \otimes_R A \\ m \otimes a &\mapsto \varphi(m) \otimes a \end{aligned}$$

For each left R-module A we also have a covariant functor $\mathrm{Hom}_{\mathbb{Z}}(A, \bullet)$ that sends each $\mathbb{Z}$-module B to the right R-module $\mathrm{Hom}_{\mathbb{Z}}(A, B)$ with $\varphi(a)r := \varphi(ra)$ and each $\mathbb{Z}$-module morphism $\varphi\colon B \to C$ to the right R-module morphism $\mathrm{Hom}(A, B) \to \mathrm{Hom}(A, C)$ defined by $f \mapsto \varphi \circ f$. Note that $(\varphi rs)(a) = \varphi(rsa) = (\varphi r)(sa) = ((\varphi r)s)(a)$, so $\mathrm{Hom}_{\mathbb{Z}}(A, B)$ is indeed a *right* R-module.

For any abelian group B there is a natural isomorphism of $\mathbb{Z}$-modules

$$\begin{aligned} \mathrm{Hom}_{\mathbb{Z}}(M \otimes_R A, B) &\xrightarrow{\sim} \mathrm{Hom}_R(M, \mathrm{Hom}_{\mathbb{Z}}(A, B)) \\ \varphi &\mapsto (m \mapsto (a \mapsto \varphi(m \otimes a))) \\ (m \otimes a \mapsto \phi(m)(a)) &\leftarrow\!\shortmid \phi \end{aligned} \tag{3}$$

The functors $\bullet \otimes_R A$ and $\mathrm{Hom}_{\mathbb{Z}}(A, \bullet)$ are thus *adjoint functors*. One can view (3) as a universal property that determines $M \otimes_R A$ up to a unique isomorphism.

Lemma 23.64. *For any left R-module the functor $\bullet \otimes_R A$ is right exact.*

Proof. Let

$$0 \longrightarrow M \xrightarrow{\varphi} N \xrightarrow{\phi} P \longrightarrow 0,$$

be an exact sequence of right R-modules. For any $\sum_i p_i \otimes a_i \in P_*$ we can pick $n_i \in N$ such that $\phi(n_i) = p_i$ and then $\phi(\sum_i n_i \otimes a) = \sum_i p_i \otimes a$, thus ϕ_* is surjective. For any $\sum_i m_i \otimes a_i \in M \otimes_R A$ we have $\phi_*(\varphi_*(\sum_i m_i \otimes a_i)) = \sum_i \phi(\varphi(m_i)) \otimes a_i = \sum_i 0 \otimes a_i = 0$, so $\operatorname{im}\varphi_* \subseteq \ker \phi_*$. To prove $\operatorname{im}\varphi_* = \ker\phi_*$ it suffices to show that $N_*/\operatorname{im}\varphi_* \simeq P_*$, since the surjectivity of ϕ_* implies $N^*/\ker\varphi_* \simeq P_*$. For every abelian group B the sequence

$$0 \longrightarrow \mathrm{Hom}_R(P, \mathrm{Hom}_{\mathbb{Z}}(A, B)) \xrightarrow{\phi^*} \mathrm{Hom}_R(N, \mathrm{Hom}_{\mathbb{Z}}(A, B)) \xrightarrow{\phi^*} \mathrm{Hom}_R(M, \mathrm{Hom}_{\mathbb{Z}}(A, B))$$

is left exact (by applying Corollary 23.60 to the right R-module $\mathrm{Hom}_{\mathbb{Z}}(A, B)$; note that the corollary applies to both left and right R-modules). Equivalently, by (3),

$$0 \longrightarrow \mathrm{Hom}_{\mathbb{Z}}(P_*, B) \xrightarrow{\phi_*^*} \mathrm{Hom}_{\mathbb{Z}}(N_*, B) \xrightarrow{\varphi_*^*} \mathrm{Hom}_{\mathbb{Z}}(M_*, B),$$

Applying Lemma 23.58 and the surjectivity of ϕ_* yields the desired right exact sequence

$$M_* \xrightarrow{\varphi_*} N_* \xrightarrow{\phi} P_* \longrightarrow 0.$$

□

Corollary 23.65. *For any left R-module A, the functor $\bullet \otimes_R A$ is an additive functor.*

Proof. See [6, Lemma 12.7.2] for a proof that this follows from right exactness; it is easy to check directly in any case. □

Remark 23.66. Corollary 23.65 implies that $\bullet \otimes_R A$ commutes with finite direct sums, and in fact it commutes with arbitrary direct sums (but not direct products).

Remark 23.67. For any right R-module A the functor $A \otimes_R \bullet$ is also right exact.

If A is an R-module and C is a chain complex of R-modules, applying the functor $\mathrm{Hom}(\bullet, A)$ to the R-modules C_n and boundary maps $d_n \colon C_{n+1} \to C_n$ yields a cochain complex C^* of $\mathbb{Z}$-modules $C^n := C_n^*$ and coboundary maps $d^n := d_n^*$,[11] and morphisms $f \colon C \to D$ of chain complexes become morphisms $f^* \colon C^* \to D^*$ of cochain complexes. We thus also have a contravariant left exact functor from the category of chain complexes to the category of cochain complexes.

Proposition 23.68. *Let A be an R-module and let $\bullet^*$ denote the application of the functor $\mathrm{Hom}(\bullet, A)$. Let $f, g \colon C \to D$ be homotopic morphisms of chain complexes of R-modules. Then $f^*, g^* \colon D^* \to C^*$ are homotopic morphisms of cochain complexes of $\mathbb{Z}$-modules.*

Proof. The morphisms f and g are homotopic, so their exist morphisms $h_n \colon C_n \to D_{n+1}$ such that $f_n - g_n = d_n \circ h_n + h_{n-1} \circ d_{n-1}$ for all $n \geq 0$. Applying the contravariant functor $\mathrm{Hom}(\bullet, A)$ yields

$$f_n^* - g_n^* = h_n^* \circ d_n^* + d_{n-1}^* \circ h_{n-1}^*,$$

where $h_n^* \colon D_{n+1} \to C_n$ for all $n \geq 0$, with $h_{-1} = 0$. Thus f^* and g^* are homotopic. □

Proposition 23.69. *Let A be a left R-module and let $\bullet_*$ denote the application of the functor $\bullet \otimes_R A$. Let $f, g \colon C \to D$ be homotopic morphisms of chain complexes of right R-modules. Then $f_*, g_* \colon C_* \to D_*$ are homotopic morphisms of chain complexes of $\mathbb{Z}$-modules.*

Proof. The morphisms f and g are homotopic, so their exist morphisms $h_n \colon C_n \to D_{n+1}$ such that $f_n - g_n = d_n \circ h_n + h_{n-1} \circ d_{n-1}$ for all $n \geq 0$. Applying the covariant functor $\bullet \otimes_R A$ yields

$$f_{n*} - g_{n*} = d_{n*} \circ h_{n*} + h_{n-1*} \circ d_{n-1*},$$

where $h_{n*} \colon C_{n+1} \to D_n$ for all $n \geq 0$, with $h_{-1} = 0$. Thus f_* and g_* are homotopic. □

23.6.4 Ext and Tor functors

Definition 23.70. Let P be a projective resolution of an R-module M. The *truncation* of P is the chain complex $\overline{P}$ with $\overline{P}_0 := P_1$ and $\overline{P}_n := P_{n+1}$ for all $n > 0$ (which need not be exact at $\overline{P}_0$).[12] Any morphism of projective resolutions $f \colon P \to Q$ induces a morphism $\bar{f} \colon \overline{P} \to \overline{Q}$ of their truncations with $\bar{f}_n := f_{n+1}$.

Theorem 23.71. *Let P, Q be projective resolutions of an R-module M, let A be an R-module, and let $\bullet_A^*$ denote application of $\mathrm{Hom}_R(\bullet, A)$. Then $H^n(\overline{P}_A^*) \simeq H^n(\overline{Q}_A^*)$ for $n \geq 0$.*

Proof. We will drop the subscript A in the proof to ease the notation.

Let $f \colon P \to Q$ and $g \colon Q \to P$ be extensions of the identity morphism id_M given by Proposition 23.57. The composition $g \circ f \colon P \to P$ is an extension of id_M, as is id_P, so $g \circ f$ is homotopic to id_P, by Proposition 23.57. We have $(g \circ f)_0 = \mathrm{id}_M = (\mathrm{id}\, P)_0$, which implies that $\overline{g \circ f} = \bar{g} \circ \bar{f}$ and $\overline{\mathrm{id}_P} = \mathrm{id}_{\overline{P}}$ are also homotopic (via the same homotopy; note $h_0 = 0$ in the proof of Proposition 23.57). Similarly, $\bar{f} \circ \bar{g}$ and $\mathrm{id}_{\overline{Q}}$ are homotopic.

Applying $\mathrm{Hom}_R(\bullet, A)$ yields homotopic morphisms $\bar{f}^* \colon \overline{Q}^* \to \overline{P}^*$ and $\bar{g}^* \colon \overline{P}^* \to \overline{Q}^*$, with $\bar{f}^* \circ \bar{g}^*$ homotopic to $\mathrm{id}_{\overline{P}}^* = \mathrm{id}_{\overline{P}^*}$ and $\bar{g}^* \circ \bar{f}^*$ homotopic to $\mathrm{id}_{\overline{Q}}^* = \mathrm{id}_{\overline{Q}^*}$, by Proposition 23.68. By Lemma 23.55, $\bar{f}^*$ and $\bar{g}^*$ induce isomorphims $H^n(\overline{P}_A^*) \simeq H^n(\overline{Q}_A^*)$ for all $n \geq 0$. □

[11] This justifies our indexing the boundary maps $d_n \colon C_{n+1} \to C_n$ rather than $d_n \colon C_n \to C_{n-1}$.

[12] The intuition is that the truncation of projective resolution of M can serve as a replacement for M.

Definition 23.72. Let A and M be R-modules. $\operatorname{Ext}^n_R(M,A)$ is the abelian group $H^n(\overline{P}^*_A)$ uniquely determined by Theorem 23.71 using any projective resolution P of M. If $\alpha\colon A \to B$ is a morphism of R-modules the map $\varphi \mapsto \alpha\circ\varphi$ induces a morphism of cochain complexes $\overline{P}^*, A \to \overline{P}^*_B$ and morphisms $\operatorname{Ext}^n_R(M,\alpha)\colon \operatorname{Ext}^n_R(M,A) \to \operatorname{Ext}^n_R(M,B)$ for each $n \geq 0$.

We thus have a family of functors $\operatorname{Ext}^n_R(M,\bullet)$ from the category of R-modules to the category of abelian groups that is a cohomological δ-functor (by Theorem 23.54).

Lemma 23.73. *Let M be an R-module. The functors $\operatorname{Ext}^n_R(M,\bullet)$ are additive functors and thus commute with finite direct sums and products.*

Proof. This follows from Corollary 23.61 and the fact $H^n(\bullet)$ is an additive functor. □

Lemma 23.74. *For any two R-modules M and A we have $\operatorname{Ext}^0_R(M,A) \simeq \operatorname{Hom}_R(M,A)$.*

Proof. Let $\cdots \to P_2 \to P_1 \to M \to 0$ be a projective resolution of M. Applying $\operatorname{Hom}_R(\bullet, A)$ yields an exact sequence $0 \to M^* \to P_1^* \to P_2^* \to \cdots$, and we observe that

$$\operatorname{Ext}^0_R(M,A) = H^0(\overline{P}^*) = Z^0(\overline{P}^*)/B^0(\overline{P}^*) = \ker(P_1^* \to P_2^*)/\operatorname{im}(0 \to P_1^*) \simeq M^*. \qquad \square$$

Theorem 23.75. *Let P, Q be projective resolutions of a right R-module M. Let A be a left R-module, and let $\bullet^A_*$ denote application of $\bullet \otimes_R A$. Then $H_n(\overline{P}^A_*) \simeq H_n(\overline{Q}^A_*)$ for $n \geq 0$.*

Proof. We drop the superscript A in the proof to ease the notation.

Let $f\colon P \to Q$ and $g\colon Q \to P$ be extensions of the identity morphism id_M given by Proposition 23.57. As in the proof of Theorem 23.71, $\bar{g}\circ\bar{f}$ and $\mathrm{id}_{\overline{P}}$ are homotopic, as are $\bar{f}\circ\bar{g}$ and $\mathrm{id}_{\overline{Q}}$.

Applying $\bullet \otimes_R A$ yields homotopic morphisms $\bar{f}_*\colon \overline{P}_* \to \overline{Q}_*$ and $\bar{g}_*\colon \overline{Q}_* \to \overline{P}_*$, with $\bar{f}_*\circ\bar{g}_*$ homotopic to $\mathrm{id}_{\overline{P}_*}$ and $\bar{f}_*\circ\bar{g}_*$ homotopic to $\mathrm{id}_{\overline{Q}_*}$. By Lemma 23.52, $\bar{f}_*$ and $\bar{g}_*$ induce isomorphisms $H_n(\overline{P}_*) \simeq H_n(\overline{Q}_*)$ for all $n \geq 0$. □

Definition 23.76. Let A a left R-module and let M be a right R-module. $\operatorname{Tor}^R_n(M,A)$ is the abelian group $H_n(\overline{P}^A_*)$ uniquely determined by Theorem 23.75 using any projective resolution P of M. If $\alpha\colon A \to B$ is a morphism of left R-modules the map $x\otimes a \mapsto x\otimes\varphi(a)$ induces a morphism $\overline{P}^A_* \to \overline{P}^B_*$ and morphisms $\operatorname{Tor}^R_n(M,\alpha)\colon \operatorname{Tor}^R_n(M,A) \to \operatorname{Ext}^R_n(M,B)$ for each $n \geq 0$. This yields a family of functors $\operatorname{Tor}^R_n(M,\bullet)$ from the category of left R-modules to the category of abelian groups that is a homological δ-functor (by Theorem 23.50).

Lemma 23.77. *Let M be a right R-module. The functors $\operatorname{Tor}^R_n(M,\bullet)$ are additive functors and thus commute with finite direct sums and products.*

Proof. This follows from Corollary 23.65 and the fact $H_n(\bullet)$ is an additive functor. □

Lemma 23.78. *For any two R-modules M and A we have $\operatorname{Tor}^R_0(M,A) \simeq M \otimes_R A$.*

Proof. Let $\cdots \to P_2 \to P_1 \to M \to 0$ be a projective resolution of M. Applying $\bullet \otimes_R A$ yields the exact sequence $\cdots P_{2*} \to P_{1*} \to M_* \to 0$, and we observe that

$$\operatorname{Tor}^R_0(M,A) = H_0(\overline{P}_*) = Z_0(\overline{P}_*)/B_0(\overline{P}_*) = \ker(P_{1*} \to 0)/\operatorname{im}(P_{2*} \to P_{1*}) \simeq M_*, \qquad \square$$

Remark 23.79. One can also define $\operatorname{Ext}^n_R(M,A)$ and $\operatorname{Tor}^R_n(M,A)$ using injective resolutions; see [7, §2.7] for a proof that this yields the same results.

References

[1] K. Brown, *Cohomology of groups*, Springer, 1982.

[2] G. J. Janusz, *Algebraic number fields*, 2nd ed., AMS, 1992.

[3] J. S. Milne, *Class field theory*, version 4.02, 2013.

[4] J.-P. Serre, *Local fields*, Springer, 1979.

[5] J.-P. Serre *Galois cohomology*, Springer, 1997.

[6] Stacks Project Authors, *Stacks Project*, `http://stacks.math.columbia.edu`.

[7] C. A. Weibel, *An introduction to homological algebra*, Cambridge Univ. Press, 1994.

24 Artin reciprocity in the unramified case

Let L/K be an abelian extension of number fields. In Lecture 22 we defined the norm group $T^{\mathfrak{m}}_{L/K} \coloneqq N_{L/K}(\mathcal{I}^{\mathfrak{m}}_L)\mathcal{R}^{\mathfrak{m}}_K$ (see Definition 22.27) that we claim is equal to the kernel of the Artin map $\psi^{\mathfrak{m}}_{L/K}\colon \mathcal{I}^{\mathfrak{m}}_K \to \mathrm{Gal}(L/K)$, provided that the modulus $\mathfrak{m}$ is divisible by the conductor of L (see Definition 22.24). We showed that $T^{\mathfrak{m}}_{L/K}$ contains $\ker \psi^{\mathfrak{m}}_{L/K}$ (Proposition 22.28), and in Theorem 22.29 we proved the inequality

$$[\mathcal{I}^{\mathfrak{m}}_K\colon T^{\mathfrak{m}}_{L/K}] \;\le\; [L:K] = [\mathcal{I}^{\mathfrak{m}}_K : \ker \psi^{\mathfrak{m}}_{L/K}] \tag{1}$$

(the equality follows from the surjectivity of the Artin map proved in Theorem 21.19). It only remains to prove the reverse inequality

$$[\mathcal{I}^{\mathfrak{m}}_K\colon T^{\mathfrak{m}}_{L/K}] \ge [L:K], \tag{2}$$

which then yields an isomorphism

$$\mathcal{I}^{\mathfrak{m}}_K/T^{\mathfrak{m}}_{L/K} \xrightarrow{\sim} \mathrm{Gal}(L/K) \tag{3}$$

induced by the Artin map. This result is known as the *Artin reciprocity law*.

In this lecture we will prove (2) for cyclic extensions L/K when the modulus $\mathfrak{m}$ is trivial (which forces L/K to be unramified), and then show that this implies the Artin reciprocity law for all unramified abelian extensions.

24.1 Some cohomological calculations

If L/K is a finite Galois extension of global fields with Galois group G, then we can naturally view any of the abelian groups L, $L^\times$, $\mathcal{O}_L$, $\mathcal{O}_L^\times$, $\mathcal{I}_L$, $\mathcal{P}_L$ as G-modules.

When $G = \langle\sigma\rangle$ is cyclic we can compute the Tate cohomology groups of any of these G-modules A, and their associated Herbrand quotients $h(A)$. The Herbrand quotient is defined as the ratio of the cardinalities of

$$\hat{H}^0(A) \coloneqq \hat{H}^0(G,A) \coloneqq \operatorname{coker} \hat{N}_G = A^G/\operatorname{im}\hat{N}_G = \frac{A[\sigma-1]}{N_G(A)},$$
$$\hat{H}_0(A) \coloneqq \hat{H}_0(G,A) \coloneqq \ker \hat{N}_G = A_G[\hat{N}_G] = \frac{A[N_G]}{(\sigma-1)(A)},$$

if both are finite. We can also compute $\hat{H}_0(A) = \hat{H}^{-1}(A) \simeq \hat{H}^1(A) = H^1(A)$ as 1-cocycles modulo 1-coboundaries whenever it is convenient to do so. In the interest of simplifying the notation we omit G from our notation whenever it is clear from context.

For the multiplicative groups $\mathcal{O}_L^\times, L^\times, \mathcal{I}_L, \mathcal{P}_L$, the norm element $N_G \coloneqq \sum_{i=1}^n \sigma^i$ corresponds to the action of the field norm $\mathrm{N}_{L/K}$ and ideal norm $N_{L/K}$ that we have previously defined, provided that we identify the codomain of the norm map with a subgroup of its domain. For the groups $L^\times$ and $\mathcal{O}_L^\times$ this simply means identifying $K^\times$ and $\mathcal{O}_K^\times$ as subgroups via inclusion. For the ideal group $\mathcal{I}_K$ we have a natural extension map $\mathcal{I}_K \hookrightarrow \mathcal{I}_L$ defined by $I \mapsto I\mathcal{O}_L$ that restricts to a map $\mathcal{P}_K \hookrightarrow \mathcal{P}_L$.[1] Under this convention taking the norm of an

[1] The induced map $\mathrm{Cl}_K \to \mathrm{Cl}_L$ need not be injective; extensions of non-principal ideals may be principal. Indeed, when L is the Hilbert class field every $\mathcal{O}_K$-ideal extends to a principal $\mathcal{O}_L$-ideal; this was conjectured by Hilbert and took over 30 years to prove. You will get a chance to prove it on Problem Set 10.

element of $\mathcal{I}_L$ that is (the extension of) an element of $\mathcal{I}_K$ corresponds to the map $I \mapsto I^{\#G}$, as it should, and $\mathcal{I}_K$ is a subgroup of the G-invariants $\mathcal{I}_L^G$.[2]

When A is multiplicative, the action of $\sigma - 1$ on $a \in A$ is $(\sigma-1)(a) = \sigma(a)/a$, but we will continue to use the notation $(\sigma-1)(A)$ and $A[\sigma-1]$ to denote the image and kernel of this action. Conversely, when A is additive, the action of N_G corresponds to the trace map, not the norm map. In order to lighten the notation, in this lecture we use N to denote both the (relative) field norm $\mathrm{N}_{L/K}$ and the ideal norm $N_{L/K}$.

Theorem 24.1. *Let L/K be a finite Galois extension with Galois group $G := \mathrm{Gal}(L/K)$, and for any G-module A, let $\hat{H}^n(A)$ denote $\hat{H}^n(G, A)$ and let* N *denote the norm map* $\mathrm{N}_{L/K}$.

(i) *$\hat{H}^0(L)$ and $\hat{H}^1(L)$ are both trivial.*

(ii) *$\hat{H}^0(L^\times) \simeq K^\times/\mathrm{N}(L^\times)$ and $\hat{H}^1(L^\times)$ is trivial.*

Proof. (i) We have $L^G = K$ (by definition). The trace map $\mathrm{T}\colon L \to K$ is not identically zero (by Theorem 5.20, since L/K is separable), so it must be surjective, since it is K-linear. Thus $N_G(L) = \mathrm{T}(L) = K$ and $\hat{H}^0(L) = K/K = 0$.

Now fix $\alpha \in L$ with $\mathrm{T}(\alpha) = \sum_{\tau\in G} \tau(\alpha) = 1$, consider a 1-cocycle $f\colon G \to L$ (this means $f(\sigma\tau) = f(\sigma) + \sigma(f(\tau))$), and put $\beta := \sum_{\tau\in G} f(\tau)\tau(\alpha)$. For all $\sigma \in G$ we have

$$\sigma(\beta) = \sum_{\tau\in G} \sigma(f(\tau))\sigma(\tau(\alpha)) = \sum_{\tau\in G} (f(\sigma\tau) - f(\sigma))(\sigma\tau)(\alpha) = \sum_{\tau\in G} (f(\tau) - f(\sigma))\tau(\alpha) = \beta - f(\sigma),$$

so $f(\sigma) = \beta - \sigma(\beta)$. This implies f is a 1-coboundary, so $\hat{H}^1(L) = H^1(L)$ is trivial.

(ii) We have $(L^\times)^G = K^\times$, so $\hat{H}^0(L^\times) = K^\times/N_G L^\times = K^\times/\mathrm{N}(L^\times)$. Consider any nonzero 1-cocycle $f\colon G \to L^\times$ (now this means $f(\sigma\tau) = f(\sigma)\sigma(f(\tau))$). By Lemma 20.6, $\alpha \mapsto \sum_{\tau\in G} f(\tau)\tau(\alpha)$ is not the zero map. Let $\beta = \sum_{\tau\in G} f(\tau)\tau(\alpha) \in L^\times$ be a nonzero element in its image. For all $\sigma \in G$ we have

$$\sigma(\beta) = \sum_{\tau\in G} \sigma(f(\tau))\sigma(\tau(\alpha)) = \sum_{\tau\in G} f(\sigma\tau)f(\sigma)^{-1}(\sigma\tau)(x) = f(\sigma)^{-1}\sum_{\tau\in G} f(\tau)\tau(\alpha) = f(\sigma)^{-1}\beta,$$

so $f(\sigma) = \beta/\sigma(\beta)$. This implies f is a coboundary, so $\hat{H}^1(L^\times) = H^1(L^\times)$ is trivial. □

Corollary 24.2 (Hilbert Theorem 90)**.** *Let L/K be a finite cyclic extension with Galois group $\mathrm{Gal}(L/K) = \langle\sigma\rangle$. Then $\mathrm{N}(\alpha) = 1$ if and only if $\alpha = \beta/\sigma(\beta)$ for some $\beta \in L^\times$.*

Proof. By Theorem 23.37, $\hat{H}^1(L^\times) \simeq \hat{H}^{-1}(L^\times) = \hat{H}_0(L^\times) = L^\times[N_G]/(\sigma-1)(L^\times)$, and Theorem 24.1 implies $L^\times[N_G] = (\sigma-1)(L^\times)$. The corollary follows. □

Remark 24.3. "Hilbert Theorem 90" refers to Hilbert's text on algebraic number theory [1], although the result is due to Kummer. The result $H^1(\mathrm{Gal}(L/K), L^\times) = 0$ implied by Theorem 24.1 is also often called Hilbert Theorem 90; it is due to Noether [2].

Our next goal is to compute the Herbrand quotient of $\mathcal{O}_L^\times$ (in the case that L/K is a finite cyclic extension of number fields). For this we will apply a variant of Dirichlet's unit theorem due to Herbrand, but first we need to discuss infinite places of number fields.

If L/K is a Galois extension of global fields, the Galois group $\mathrm{Gal}(L/K)$ acts on the set of places w of L via the action $w \mapsto \sigma(w)$, where $\sigma(w)$ is the equivalence class of the absolute value defined by $\|\alpha\|_{\sigma(w)} := \|\sigma(\alpha)\|_w$. This action permutes the places w lying above a given place v of K; if v is a finite place corresponding to a prime $\mathfrak{p}$ of K, this is just the usual action of the Galois group on the set $\{\mathfrak{q}|\mathfrak{p}\}$.

[2]Note that $\mathcal{I}_L^G = \mathcal{I}_K$ only when L/K is unramified; see Lemma 24.8 below.

Definition 24.4. Let L/K be a Galois extension of global fields and let w be a place of L. The *decomposition group* of w is its stabilizer in $\mathrm{Gal}(L/K)$:

$$D_w := \{\sigma \in \mathrm{Gal}(L/K) : \sigma(w) = w\}.$$

If w corresponds to a prime $\mathfrak{q}$ of $\mathcal{O}_L$ then $D_w = D_{\mathfrak{q}}$ is also the decomposition group of $\mathfrak{q}$.

Now let L/K be a Galois extension of number fields. If we write $L \simeq \mathbb{Q}[x]/(f)$ then we have a one-to-one correspondence between embeddings of L into $\mathbb{C}$ and roots of f in $\mathbb{C}$. Each embedding of L into $\mathbb{C}$ restricts to an embedding of K into $\mathbb{C}$, and this induces a map that sends each infinite place w of L to the infinite place v of K that w extends. This map may send a complex place to a real place; this occurs when a pair of distinct complex conjugate embeddings of L restrict to the same embedding of K (which must be a real embedding). In this case we say that the place v (and w) is *ramified* in the extension L/K, and define the *ramification index* $e_v := 2$ when this holds (and put $e_v := 1$ otherwise). This notation is consistent with our notation $e_v := e_{\mathfrak{p}}$ for finite places v corresponding to primes $\mathfrak{p}$ of K. Let us also define $f_v := 1$ for $v|\infty$ and put $g_v := \#\{w|v\}$ so that the following formula generalizing Corollary 7.5 holds for all places v of K:

$$e_v f_v g_v = [L : K].$$

Definition 24.5. For a Galois extension of number fields L/K we define the integers

$$e_0(L/K) := \prod_{v\nmid\infty} e_v, \qquad e_\infty(L/K) := \prod_{v|\infty} e_v, \qquad e(L/K) := e_0(L/K)e_\infty(L/K).$$

Let us now write $L \simeq K[x]/(g)$. Each embedding of K into $\mathbb{C}$ gives rise to $[L : K]$ distinct embeddings of L into $\mathbb{C}$ that extend it, one for each root of g (use the embedding of K to view g as a polynomial in $\mathbb{C}[x]$, then pick a root of g in $\mathbb{C}$). The transitive action of $\mathrm{Gal}(L/K)$ on the roots of g induces a transitive action on these embeddings and their corresponding places. Thus for each infinite place v of K the Galois group acts transitively on $\{w|v\}$, and either every place w above v is ramified (this can occur only when v is real and $[L : K]$ is divisible by 2), or none are. It follows that each unramified place v of K has $[L : K]$ places w lying above it, each with trivial decomposition group D_w, while each ramified (real) place v of K has $[L : K]/2$ (complex) places w lying above it, each with decomposition group D_w of order 2 (its non-trivial element corresponds to complex conjugation in the corresponding embeddings), and the D_w are all conjugate.

Theorem 24.6 (Herbrand unit theorem)**.** *Let L/K be a Galois extension of number fields. Let $w_1, \dots, w_r$ be the real places of L, let $w_{r+1}, \dots, w_{r+s}$ be the complex places of L. There exist $\varepsilon_1, \dots, \varepsilon_{r+s} \in \mathcal{O}_L^\times$ such that*

(i) *$\sigma(\varepsilon_i) = \varepsilon_j$ if and only if $\sigma(w_i) = w_j$, for all $\sigma \in \mathrm{Gal}(L/K)$;*

(ii) *$\varepsilon_1, \dots, \varepsilon_{r+s}$ generate a finite index subgroup of $\mathcal{O}_L^\times$;*

(iii) *$\varepsilon_1\varepsilon_2 \cdots \varepsilon_{r+s} = 1$, and every relation among the ε_i is generated by this one.*

Proof. Pick $\epsilon_1, \dots, \epsilon_{r+s} \in \mathcal{O}_L^\times$ such that $\|\epsilon_i\|_{w_j} < 1$ for $i \neq j$; the existence of such ϵ_i follows from the strong approximation theorem that we will prove in the next lecture; the product formula then implies $\|\epsilon_i\|_{w_i} > 1$. Now let $\alpha_i := \prod_{\sigma \in D_{w_i}} \sigma(\epsilon_i) \in \mathcal{O}_L^\times$. We have

$\|\alpha_i\|_{w_i} = \prod_{\sigma\in D_{w_i}} \|\epsilon_i\|_{w_i} > 1$ and $\|\alpha_i\|_{w_j} = \prod_{\sigma\in D_{w_i}} \|\epsilon_i\|_{\sigma(w_j)} < 1$, since $\sigma \in D_{w_i}$ fixes w_i and permutes the w_j with $j \neq i$. Each α_i is fixed by D_{w_i}.

Let $G := \mathrm{Gal}(L/K)$. For $i = 1, \dots, r+s$, let $r(i) := \min\{j : \sigma(w_i) = w_j \text{ for some } \sigma \in G\}$, so that $w_{r(i)}$ is a distinguished representative of the G-orbit of w_i. For $i = 1, \dots, r+s$ let $\beta_i := \sigma(\alpha_{r(i)})$, where σ is any element of G such that $\sigma(w_{r(i)}) = w_i$. The value of $\sigma(\alpha_{r(i)})$ does not depend on the choice of σ because $\sigma_1(w_{r(i)}) = \sigma_2(w_{r(i)})$ if and only if $\sigma_2^{-1}\sigma_1 \in D_{w_{r(i)}}$ and $\alpha_{r(i)}$ is fixed by $D_{w_{r(i)}}$. The β_i then satisfy (i).

The β_i also satisfy (ii): a product $\gamma_j := \prod_{i\neq j} \beta_i^{n_i}$ cannot be trivial because $\|\gamma_j\|_{w_j} < 1$; in particular, $\beta_1, \dots, \beta_{r+s-1}$ generate a subgroup of $\mathcal{O}_L^\times$ isomorphic to $\mathbb{Z}^{r+s-1}$ which necessarily has finite index in $\mathcal{O}_L^\times \simeq \mathbb{Z}^{r+s-1} \times \mu_L$ (see Theorem 15.12). But we must have $\prod_i \beta_i^{n_i} = 1$ for some tuple $(n_1, \dots, n_{r+s}) \in \mathbb{Z}^{r+s}$ (with $n_i = n_j$ whenever w_i and w_j lie in the same G-orbit, since every $\sigma \in G$ fixes 1). The set of such tuples spans a rank-1 submodule of $\mathbb{Z}^{r+s}$ from which we choose a generator $(n_1, \dots, n_{r+s})$ (by inverting some β_i if necessary, we can make all the n_i positive if we wish). Then $\varepsilon_i := \beta_i^{n_i}$ satisfy (i), (ii), (iii) as desired. □

Theorem 24.7. *Let L/K be a cyclic extension of number fields with Galois group $G = \langle\sigma\rangle$. The Herbrand quotient of the G-module $\mathcal{O}_L^\times$ is*

$$h(\mathcal{O}_L^\times) = \frac{e_\infty(L/K)}{[L:K]}.$$

Proof. Let $\varepsilon_1, \dots, \varepsilon_{r+s} \in \mathcal{O}_L^\times$ be as in Theorem 24.6, and let A be the subgroup of $\mathcal{O}_L^\times$ they generate, viewed as a G-module. By Corollary 23.48, $h(A) = h(\mathcal{O}_L^\times)$ if either is defined, since A has finite index in $\mathcal{O}_L^\times$, so we will compute $h(A)$.

For each field embedding $\phi\colon K \hookrightarrow \mathbb{C}$, let E_ϕ be the free $\mathbb{Z}$-module with basis $\{\varphi|\phi\}$ consisting of the $n := [L : K]$ embeddings $\varphi\colon L \hookrightarrow \mathbb{C}$ with $\varphi_{|K} = \phi$, equipped with the G-action given by $\sigma(\varphi) := \varphi\circ\sigma$. Let v be the infinite place of K corresponding to ϕ, and let A_v be the free $\mathbb{Z}$-module with basis $\{w|v\}$ consisting of places of L that extend v, equipped with the G-action given by the action of G on $\{w|v\}$. Let $\pi\colon E_\phi \to A_v$ be the G-module morphism sending each embedding $\varphi|\phi$ to the corresponding place $w|v$. Let $m := \#\{w|v\}$ and define $\tau := \sigma^m$; then τ is either trivial or has order 2, and in either case generates the decomposition group D_w for all $w|v$ (since G is abelian). We have an exact sequence

$$0 \to \ker\pi \longrightarrow E_\phi \xrightarrow{\pi} A_v \to 0,$$

with $\ker\pi = (\tau - 1)E_\phi$. If v is unramified then $\ker\pi = 0$ and $h(A_v) = h(E_\phi) = 1$, since $E_\phi \simeq \mathbb{Z}[G] \simeq \mathrm{Ind}^G(\mathbb{Z})$, by Lemma 23.43. Otherwise, order $\{w|v\} = \{w_0, \dots, w_{m-1}\}$ so

$$\ker\pi = (\tau-1)E_\phi = \left\{ \sum_{0\le i<m} a_i(w_i - w_{m+i}) : a_i \in \mathbb{Z} \right\},$$

and observe that $(\ker\pi)^G = 0$, since τ acts on π as negation, and $(\ker\pi)_G \simeq \mathbb{Z}/2\mathbb{Z}$, since $(\sigma-1)\ker\pi = \{\sum a_i(w_i - w_{m+i}) : a_i \in \mathbb{Z} \text{ with } \sum a_i \equiv 0 \bmod 2\}$ (which is killed by N_G). So in this case $h(\ker\pi) = 1/2$, and therefore $h(A_v) = h(E_\phi)/h(\ker\pi) = 2$, by Corollary 23.41, and in every case we have $h(A_v) = e_v$, where $e_v \in \{1, 2\}$ is the ramification index of v.

Now consider the exact sequence of G-modules

$$0 \longrightarrow \mathbb{Z} \longrightarrow \bigoplus_{v|\infty} A_v \xrightarrow{\psi} A \longrightarrow 1$$

where ψ sends each infinite place $w_1,\dots,w_{r+s}$ of L to the corresponding $\varepsilon_1,\dots,\varepsilon_{r+s} \in A$ given by Theorem 24.6 (each A_v contains either n or $n/2$ of the w_i in its $\mathbb{Z}$-basis). The kernel of ψ is the trivial G-module $(\sum_i w_i)\mathbb{Z} \simeq \mathbb{Z}$, since we have $\psi(\sum_i w_i) = \prod_i \varepsilon_i = 1$ and no other relations among the ε_i, by Theorem 24.6. We have $h(\mathbb{Z}) = \#G = [L:K]$, by Corollary 23.46, and $h(\bigoplus A_v) = \prod h(A_v) = \prod e_v$, by Corollary 23.42, so $h(A) = e_\infty(L/K)/[L:K]$. □

Lemma 24.8. *Let L/K be a cyclic extension of number fields with Galois group G. For the G-module $\mathcal{I}_L$ we have $h_0(\mathcal{I}_L) = 1$ and $h^0(\mathcal{I}_L) = e_0(L/K)[\mathcal{I}_K : N(\mathcal{I}_L)]$.*

Proof. It is clear that $I \in \mathcal{I}_L^G \Leftrightarrow v_{\sigma(\mathfrak{q})}(I) = v_{\mathfrak{q}}(I)$ for all primes $\mathfrak{q} \in \mathcal{I}_L$. If we put $\mathfrak{p} \coloneqq \mathfrak{q} \cap \mathcal{O}_K$, then for $I \in \mathcal{I}_L^G$ the value of $v_{\mathfrak{q}}(I)$ is constant on $\{\mathfrak{q}|\mathfrak{p}\}$, since G acts transitively on this set. It follows that $\mathcal{I}_L^G$ consists of all products of ideals of the form $(\mathfrak{p}\mathcal{O}_L)^{1/e_{\mathfrak{p}}}$. Therefore $[\mathcal{I}_L^G : \mathcal{I}_K] = e_0(L/K)$ and $h^0(\mathcal{I}_L) = [\mathcal{I}_L^G : N(\mathcal{I}_L)] = e_0(L/K)[\mathcal{I}_K : N(\mathcal{I}_L)]$ as claimed.

For each prime $\mathfrak{q}|\mathfrak{p}$ we have $N(\mathfrak{q}) = \mathfrak{p}^{f_{\mathfrak{p}}}$ (by Theorem 6.10). Thus if $N(I) = \mathcal{O}_K$ then $N(\prod_{\mathfrak{q}|\mathfrak{p}} \mathfrak{q}^{v_{\mathfrak{q}}(I)}) = \mathfrak{p}^{f_{\mathfrak{p}} \sum_{\mathfrak{q}|\mathfrak{p}} v_{\mathfrak{q}}(I)} = \mathcal{O}_K$, equivalently, $\sum_{\mathfrak{q}|\mathfrak{p}} v_{\mathfrak{q}}(I) = 0$, for every prime $\mathfrak{p}$ of K. Order $\{\mathfrak{q}|\mathfrak{p}\}$ as $\mathfrak{q}_1,\dots,\mathfrak{q}_r$ so that $\mathfrak{q}_{i+1} = \sigma(\mathfrak{q}_i)$ and $\mathfrak{q}_1 = \sigma(\mathfrak{q}_r)$, let $n_i \coloneqq v_{\mathfrak{q}_i}(I)$, and define

$$J_{\mathfrak{p}} \coloneqq \mathfrak{q}_1^{n_1}\mathfrak{q}_2^{n_1-n_2}\mathfrak{q}_3^{n_1-n_2-n_3}\cdots\mathfrak{q}_r^{n_1-n_2-\cdots-n_r}.$$

Then

$$\begin{aligned}\sigma(J_{\mathfrak{p}})/J_{\mathfrak{p}} &= \mathfrak{q}_2^{n_1-(n_1-n_2)}\mathfrak{q}_3^{n_1-n_2-(n_1-n_2-n_3)}\cdots\mathfrak{q}_r^{n_1-\cdots-n_{r-1}-(n_1-\cdots-n_r)}\mathfrak{q}_1^{n_1-\cdots-n_r-n_1}\\ &= \mathfrak{q}_2^{n_2}\mathfrak{q}_3^{n_3}\cdots\mathfrak{q}_r^{n_r}\mathfrak{q}_1^{-n_2-\cdots-n_r} = \mathfrak{q}_2^{n_2}\mathfrak{q}_3^{n_3}\cdots\mathfrak{q}_r^{n_r}\mathfrak{q}_1^{n_1} = \prod_{\mathfrak{q}|\mathfrak{p}}\mathfrak{q}^{v_{\mathfrak{q}}(I)},\end{aligned}$$

since $n_1+\cdots+n_r = 0$ implies $n_1 = -n_2-\cdots-n_r$. It follows that $I = \sigma(J)/J$ where $J \coloneqq \prod_{\mathfrak{p}\nmid\mathfrak{m}} J_{\mathfrak{p}}$, thus $I_L[N_G] = (\sigma-1)(I_L)$ and $h_0(\mathcal{I}_L) = 1$. □

Theorem 24.9 (Ambiguous class number formula)**.** *Let L/K be a cyclic extension of number fields with Galois group G. The G-invariant subgroup of the G-module Cl_L has cardinality*

$$\#\mathrm{Cl}_L^G = \frac{e(L/K)\#\mathrm{Cl}_K}{n(L/K)\,[L:K]},$$

where $n(L/K) \coloneqq [\mathcal{O}_K^\times : N(L^\times)\cap\mathcal{O}_K^\times] \in \mathbb{Z}_{\geq 1}$.

Proof. The ideal class group Cl_L is the quotient of $\mathcal{I}_L$ by its subgroup $\mathcal{P}_L$ of principal fractional ideals. We thus have a short exact sequence of G-modules

$$1 \longrightarrow \mathcal{P}_L \longrightarrow \mathcal{I}_L \longrightarrow \mathrm{Cl}_L \longrightarrow 1.$$

The corresponding long exact sequence in (standard) cohomology begins

$$1 \longrightarrow \mathcal{P}_L^G \longrightarrow \mathcal{I}_L^G \longrightarrow \mathrm{Cl}_L^G \longrightarrow H^1(\mathcal{P}_L) \longrightarrow 1,$$

since $H^1(\mathcal{I}_L) \simeq \hat{H}_0(\mathcal{I}_L)$ is trivial, by Lemma 24.8. Therefore

$$\#\mathrm{Cl}_L^G = [\mathcal{I}_L^G : \mathcal{P}_L^G]\; h^1(\mathcal{P}_L). \tag{4}$$

Using the inclusions $\mathcal{P}_K \subseteq \mathcal{P}_L^G \subseteq \mathcal{I}_L^G$ we can rewrite the first factor on the RHS as

$$[\mathcal{I}_L^G : \mathcal{P}_L^G] = \frac{[\mathcal{I}_L^G : \mathcal{P}_K]}{[\mathcal{P}_L^G : \mathcal{P}_K]} = \frac{[\mathcal{I}_L^G : \mathcal{I}_K][\mathcal{I}_K : \mathcal{P}_K]}{[\mathcal{P}_L^G : \mathcal{P}_K]} = \frac{e_0(L/K)\#\mathrm{Cl}_K}{[\mathcal{P}_L^G : \mathcal{P}_K]}, \tag{5}$$

where $[\mathcal{I}_L^G : \mathcal{I}_K] = e_0(L/K)$ follows from the proof of Lemma 24.8.

We now consider the short exact sequence

$$1 \longrightarrow \mathcal{O}_L^\times \longrightarrow L^\times \overset{\alpha\mapsto(\alpha)}{\longrightarrow} \mathcal{P}_L \longrightarrow 1.$$

The corresponding long exact sequence in cohomology begins

$$1 \longrightarrow \mathcal{O}_K^\times \longrightarrow K^\times \longrightarrow \mathcal{P}_L^G \longrightarrow H^1(\mathcal{O}_L^\times) \longrightarrow 1 \longrightarrow H^1(\mathcal{P}_L) \longrightarrow H^2(\mathcal{O}_L^\times) \longrightarrow H^2(L^\times), \quad (6)$$

since $H^1(L^\times)$ is trivial, by Hilbert 90 (Corollary 24.2). We have $K^\times/\mathcal{O}_K^\times \simeq \mathcal{P}_K$, thus

$$[\mathcal{P}_L^G : \mathcal{P}_K] = h^1(\mathcal{O}_L^\times) = \frac{h^0(\mathcal{O}_L^\times)}{h(\mathcal{O}_L^\times)} = \frac{h^0(\mathcal{O}_L^\times)\,[L:K]}{e_\infty(L/K)},$$

by Theorem 24.7. Combining this identity with (4) and (5) yields

$$\#\mathrm{Cl}_L^G = \frac{e(L/K)\#\mathrm{Cl}_K}{[L:K]} \cdot \frac{h^1(\mathcal{P}_L)}{h^0(\mathcal{O}_L^\times)}. \quad (7)$$

We can write the second factor on the RHS using the second part of the long exact sequence in (6). Recall that $H^2(\bullet) = \hat{H}^2(\bullet) = \hat{H}^0(\bullet)$, by Theorem 23.37, thus

$$H^1(\mathcal{P}_L) \simeq \ker\big(\hat{H}^0(\mathcal{O}_L^\times) \to \hat{H}^0(L^\times)\big) \simeq \ker(\mathcal{O}_K^\times/N(\mathcal{O}_L^\times) \to K^\times/N(L^\times)),$$

so $h^1(\mathcal{P}_L) = [\mathcal{O}_K^\times \cap N(L^\times) : N(\mathcal{O}_L^\times)]$. We have $h^0(\mathcal{O}_L^\times) = [\mathcal{O}_K^\times : N(\mathcal{O}_L^\times)]$, thus

$$\frac{h^0(\mathcal{O}_L^\times)}{h^1(\mathcal{P}_L)} = [\mathcal{O}_K^\times : N(L^\times) \cap \mathcal{O}_K^\times] = n(L/K),$$

and plugging this into (7) yields the desired formula. □

24.2 Proof of Artin reciprocity

We now have the essential ingredients in place to prove our desired inequality for unramified cyclic extensions of number fields. We first record an elementary lemma.

Lemma 24.10. *Let $f : A \to G$ be a homomorphism of abelian groups and let B be a subgroup of A containing the kernel of f. Then $A/B \simeq f(A)/f(B)$.*

Proof. Apply the snake lemma to the commutative diagram and consider the cokernels.

$$\begin{array}{ccccccccc} & & \ker f & \hookrightarrow & B & \xrightarrow{f} & f(B) & \longrightarrow & 0 \\ & & \| & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & \ker f & \hookrightarrow & A & \xrightarrow{f} & f(A) & \longrightarrow & 0. \end{array}$$

□

In the following theorem it is crucial that the extension L/K is completely unramified, including at all infinite places of K; to emphasize this, let us say that an extension of number fields L/K is *totally unramified* if $e(L/K) = 1$.

Theorem 24.11. *Let L/K be a totally unramified cyclic extension of number fields. Then*

$$[\mathcal{I}_K : N(\mathcal{I}_L)\mathcal{P}_K] \geq [L:K].$$

Proof. We have

$$[\mathcal{I}_K : N(\mathcal{I}_K)\mathcal{P}_K] = \frac{[\mathcal{I}_K : \mathcal{P}_K]}{[N(\mathcal{I}_L)\mathcal{P}_K : \mathcal{P}_K]} = \frac{\#\mathrm{Cl}_K}{[N(\mathcal{I}_L)\mathcal{P}_K : \mathcal{P}_K]}.$$

The denominator on the RHS can be rewritten as

$$\begin{aligned}
[N(\mathcal{I}_L)\mathcal{P}_K : \mathcal{P}_K] &= [N(\mathcal{I}_L) : N(\mathcal{I}_L) \cap \mathcal{P}_K] && \text{(2nd isomorphism theorem)} \\
&= [\mathcal{I}_L : N^{-1}(\mathcal{P}_K)] && \text{(Lemma 24.10)} \\
&= [\mathcal{I}_L/\mathcal{P}_L : N^{-1}(\mathcal{P}_K)/\mathcal{P}_L] && \text{(3rd isomorphism theorem)} \\
&= [\mathrm{Cl}_L : \mathrm{Cl}_L[N_G]] \\
&= \#N_G(\mathrm{Cl}_L).
\end{aligned}$$

Now $h^0(\mathrm{Cl}_L) = [\mathrm{Cl}_L^G : N_G(\mathrm{Cl}_L)]$, and applying Theorem 24.9 yields

$$[\mathcal{I}_K : N(\mathcal{I}_K)\mathcal{P}_K] = \frac{\#\mathrm{Cl}_K \cdot h^0(\mathrm{Cl}_L)}{\#\mathrm{Cl}_L^G} = \frac{h^0(\mathrm{Cl}_L)n(L/K)[L:K]}{e(L/K)} \geq [L:K],$$

since $e(L/K) = 1$, and $h^0(\mathrm{Cl}_L)$, $n(L/K) \geq 1$. □

For a totally unramified extension of number fields L/K, let $T_{L/K} \coloneqq T^{(1)}_{L/K} = N(\mathcal{I}_L)\mathcal{P}_K$.

Corollary 24.12 (Artin reciprocity law)**.** *Let L/K be a totally unramified cyclic extension of number fields. Then $[\mathcal{I}_K : T_{L/K}] = [L : K]$ and the Artin map induces an isomorphism $\mathcal{I}_K/T_{L/K} \simeq \mathrm{Gal}(L/K)$.*

Proof. Theorems 22.29 and 24.11 imply $[\mathcal{I}_K : T_{L/K}] = [L : K]$. We have $\ker \psi_{L/K} \subseteq T_{L/K}$ (Proposition 22.28), and $[\mathcal{I}_K : \ker \psi_{L/K}] = \#\mathrm{Gal}(L/K) = [L : K] = [\mathcal{I}_K : T_{L/K}]$, since $\psi_{L/K}$ is surjective (Theorem 21.19). Therefore $\ker \psi_{L/K} = T_{L/K}$, and the Corollary follows. □

Corollary 24.13. *Let L/K be a totally unramified cyclic extension of number fields. Then $\#\mathrm{Cl}_L^G = \#\mathrm{Cl}_K/[L : K]$ and the Tate cohomology groups of Cl_L are all trivial.*

Proof. By the previous corollary and the proof of Theorem 24.11: we have $n(L/K) = 1$ and $h^0(\mathrm{Cl}_L) = 1$, and $e(L/K) = 1$, so $\#\mathrm{Cl}_L^G = \#\mathrm{Cl}_L/[L : K]$ by Theorem 24.9. We also have $h(\mathrm{Cl}_K) = h^0(\mathrm{Cl}_L)/h_0(\mathrm{Cl}_L) = 1$, since Cl_L is finite, by Lemma 23.43, so $h_0(\mathrm{Cl}_L) = 1$. Thus $\hat{H}^{-1}(\mathrm{Cl}_L)$ and $\hat{H}^0(\mathrm{Cl}_L)$ are both trivial, and this implies that all the Tate cohomology groups are trivial, by Theorem 23.37. □

Corollary 24.14. *Let L/K be a totally unramified cyclic extension of number fields. Then every unit in $\mathcal{O}_K^\times$ is the norm of an element of L.*

Proof. We have $n(L/K) = [\mathcal{O}_K^\times : N(L^\times) \cap \mathcal{O}_K^\times] = 1$, so $\mathcal{O}_K^\times = N(L^\times) \cap \mathcal{O}_K^\times$. □

24.3 Generalizing to the non-cyclic case

Corollaries 24.13 and 24.14 are specific to unramified cyclic extensions, but Corollary 24.12 (Artin reciprocity) extends to all abelian extensions. Our goal in this section is to show that for any modulus $\mathfrak{m}$ for a number field K, if the Artin reciprocity law holds for all finite cyclic extensions L/K with conductor dividing $\mathfrak{m}$, then it holds for all finite abelian extensions L/K with conductor dividing $\mathfrak{m}$.

Definition 24.15. Let $\mathfrak{m}$ be a modulus for a number field K and let L/K be a finite abelian extension ramified only at primes $\mathfrak{p}|\mathfrak{m}$. We say that L is a *class field* for $\mathfrak{m}$ if $\ker \psi^{\mathfrak{m}}_{L/K} = T^{\mathfrak{m}}_{L/K}$, where $\psi^{\mathfrak{m}}_{L/K} \colon \mathcal{I}^{\mathfrak{m}}_K \to \mathrm{Gal}(L/K)$ is the Artin map.

Remark 24.16. This definition is stated more strongly than is typical, but it is convenient for our purposes; we have already proved the surjectivity of the Artin map and that $T^{\mathfrak{m}}_{L/K}$ contains $\ker \psi^{\mathfrak{m}}_{L/K}$ so there is no reason to use an (apparently) weaker definition.

Lemma 24.17. *Let $\mathfrak{m}$ be a modulus for a number field K. If L_1 and L_2 are class fields for $\mathfrak{m}$ then so is their compositum $L \coloneqq L_1L_2$.*

Proof. We first note that $L = L_1L_2$ is ramified only at primes ramified in either L_1 or L_2 (since ramification indices are multiplicative in towers), so L is ramified only at primes $\mathfrak{p}|\mathfrak{m}$. As in the proof of Theorem 21.18, a prime $\mathfrak{p} \nmid \mathfrak{m}$ splits completely in L if and only if it splits completely in L_1 and L_2, which implies $\ker \psi^{\mathfrak{m}}_{L/K} = \ker \psi^{\mathfrak{m}}_{L_1/K} \cap \ker \psi^{\mathfrak{m}}_{L_2/K}$. The norm map is transitive in towers, so if $I = N_{L/K}(J)$ then $I = N_{L_1/K}(N_{L/L_1}(J))$ and $I = N_{L_2/K}(N_{L/L_2}(J))$, thus $N(\mathcal{I}^{\mathfrak{m}}_L) \subseteq N(\mathcal{I}^{\mathfrak{m}}_{L_1}) \cap N(\mathcal{I}^{\mathfrak{m}}_{L_2})$ and therefore $T^{\mathfrak{m}}_{L/K} \subseteq T^{\mathfrak{m}}_{L_1/K} \cap T^{\mathfrak{m}}_{L_2/K}$. If L_1 and L_2 are class fields for $\mathfrak{m}$, then

$$T^{\mathfrak{m}}_{L/K} \subseteq T^{\mathfrak{m}}_{L_1/K} \cap T^{\mathfrak{m}}_{L_2/K} = \ker \psi^{\mathfrak{m}}_{L_1/K} \cap \ker \psi^{\mathfrak{m}}_{L_2} = \ker \psi^{\mathfrak{m}}_{L/K},$$

and $\ker^{\mathfrak{m}}_{L/K} \subseteq T^{\mathfrak{m}}_{L/K}$ by Proposition 22.28, so $T^{\mathfrak{m}}_{L/K} = \ker \psi^{\mathfrak{m}}_{L/K}$ and the lemma follows. □

Corollary 24.18. *Let $\mathfrak{m}$ be a modulus for a number field K. If every finite cyclic extension of K with conductor dividing $\mathfrak{m}$ is a class field for $\mathfrak{m}$ then so is every abelian extension of K with conductor dividing $\mathfrak{m}$.*

Proof. Let L/K be a finite abelian extension of conductor $\mathfrak{c}|\mathfrak{m}$. The conductor of any subextension of L divides $\mathfrak{c}$ and therefore $\mathfrak{m}$, by Lemma 22.26.

If we write $G \coloneqq \mathrm{Gal}(L/K) \simeq H_1 \times \cdots H_r$ as a product of cyclic groups and define $L_i = L^{\bar{H}_i}$ where $\bar{H}_i = \prod_{j \neq i} H_j \subseteq G$ so that $\mathrm{Gal}(L_i/K) \simeq G/\bar{H}_i \simeq H_i$ is cyclic, then $L = L_1 \cdots L_r$ is a composition of linearly disjoint cyclic extensions of K, and it follows from Lemma 24.17 that if the L_i are all class fields for $\mathfrak{m}$, so is L. □

24.4 Class field theory for unramified abelian extensions

For the trivial modulus $\mathfrak{m} = (1)$, the three main theorems of class field theory stated in Lecture 22 state that the following hold for every number field K:

- **Existence**: The ray class field $K(1)$ exists.
- **Completeness**: Every unramified abelian extension of K is a subfield of $K(1)$.
- **Artin reciprocity**: For every subextension L/K of $K(1)$ we have $\ker \psi_{L/K} = T_{L/K}$ and a canonical isomorphism $\mathcal{I}_K/T_{L/K} \simeq \mathrm{Gal}(L/K)$.

We can now prove all of this, except for the existence of $K(1)$. But if we replace $K(1)$ with the Hilbert class field H of K (the maximal unramified abelian extension of K) we can prove an analogous series of statements, including that H is a finite extension of K and that if $K(1)$ exists it must be equal to H.

Theorem 24.19. *Let K be a number field with Hilbert class field H. The following hold:*

- *H/K is a finite extension with $\mathrm{Gal}(H/K)$ isomorphic to a quotient of Cl_K.*
- *$K(1)$ exists if and only if $\mathrm{Gal}(H/K) \simeq \mathrm{Cl}_K$, in which case $K(1) = H$.*
- *Every unramified abelian extension of K is a subfield of H* (**Completeness**).
- *For every unramified abelian extension of K we have $\ker \psi_{L/K} = T_{L/K}$ and a canonical isomorphism $\mathcal{I}_K/T_{L/K} \simeq \mathrm{Gal}(L/K)$* (**Artin reciprocity**).

Proof. Corollaries 24.12 and 24.18 together imply the Artin reciprocity law for every unramified abelian extension of K. In particular, every such extension L has $\mathrm{Gal}(L/K)$ isomorphic to a quotient of Cl_K (since $T_{L/K}$ contains $\mathcal{P}_K$). Moreover, distinct unramified abelian extensions L/K correspond to distinct quotients of Cl_K, since the primes that split completely in K are precisely those that lie in the kernel of the Artin map, and this set of primes uniquely determines L, by Theorem 21.18. It follows that there is a unique quotient of Cl_K that corresponds to H, the compositum of all such fields. The theorem follows. □

References

[1] D. Hilbert, *Die Theorie der algebraischen Zahlkörper*, Jahresbericht der Deutschen Mathematiker-Vereinigung **4** (1897), 175–546.

[2] E. Noether, *Der Hauptgeschlechtssatz für relativ-galoissche Zahlkörper*, Math. Annalen **108** (1933), 411–419.

25 The ring of adeles, strong approximation

25.1 Introduction to adelic rings

Recall that we have a canonical injection

$$\mathbb{Z} \hookrightarrow \hat{\mathbb{Z}} := \varprojlim_n \mathbb{Z}/n\mathbb{Z} \simeq \prod_p \mathbb{Z}_p,$$

that embeds $\mathbb{Z}$ into the product of its nonarchimedean completions. Each of the rings $\mathbb{Z}_p$ is compact, hence $\hat{\mathbb{Z}} = \prod_p \mathbb{Z}_p$ is compact (by Tychonoff's theorem). If we consider the analogous product $\prod_p \mathbb{Q}_p$ of the completions of $\mathbb{Q}$, each of the local fields $\mathbb{Q}_p$ is locally compact (as is $\mathbb{Q}_\infty = \mathbb{R}$), but the product $\prod_p \mathbb{Q}_p$ is **not locally compact**.

To see where the problem arises, recall that for any family of topological spaces $(X_i)_{i\in I}$ (where the index set I is any set), the product topology on $X := \prod X_i$ is defined as the weakest topology that makes all the projection maps $\pi_i \colon X \to X_i$ continuous; it is thus generated by open sets of the form $\pi_i^{-1}(U_i)$ with $U_i \subseteq X_i$ open. Every open set in X is a (possibly empty or infinite) union of open sets of the form

$$\prod_{i\in S} U_i \times \prod_{i\notin S} X_i,$$

with $S \subseteq I$ finite and each $U_i \subseteq X_i$ open (these sets form a *basis* for the topology on X). In particular, every open $U \subseteq X$ satisfies $\pi_i(U) = X_i$ for all but finitely many $i \in I$. Unless all but finitely many of the X_i are compact, the space X cannot possibly be locally compact for the simple reason that no compact set C in X contains a nonempty open set (if it did then we would have $\pi_i(C) = X_i$ compact for all but finitely many $i \in I$). Recall that to be locally compact means that for every $x \in X$ there is an open U and compact C such that $x \in U \subseteq C$.

To address this issue we want to take the product of the fields $\mathbb{Q}_p$ (or more generally, the completions of any global field) in a different way, one that yields a locally compact topological ring. This is the motivation of the *restricted product*, a topological construction that was invented primarily for the purpose of solving this number-theoretic problem.

25.2 Restricted products

This section is purely about the topology of restricted products; readers already familiar with restricted products should feel free to skip to the next section.

Definition 25.1. Let (X_i) be a family of topological spaces indexed by $i \in I$, and let (U_i) be a family of open sets $U_i \subseteq X_i$. The *restricted product* $\prod(X_i, U_i)$ is the topological space

$$\prod(X_i, U_i) := \{(x_i) : x_i \in U_i \text{ for almost all } i \in I\} \subseteq \prod X_i$$

with the basis of open sets

$$\mathcal{B} := \left\{\prod V_i : V_i \subseteq X_i \text{ is open for all } i \in I \text{ and } V_i = U_i \text{ for almost all } i \in I\right\},$$

where *almost all* means all but finitely many.

For each $i \in I$ we have a projection map $\pi_i \colon \prod(X_i, U_i) \to X_i$ defined by $(x_i) \mapsto x_i$; each π_i is continuous, since if W_i is an open subset of X_i, then $\pi_i^{-1}(W_i)$ is the union of all basic open sets $\prod V_i \in \mathcal{B}$ with $V_i = W_i$, which is an open set.

As sets, we always have

$$\prod U_i \subseteq \prod(X_i, U_i) \subseteq \prod X_i,$$

but in general the restricted product topology on $\prod(X_i, U_i)$ is not the same as the subspace topology it inherits from $\prod X_i$; it has more open sets. For example, $\prod U_i$ is an open set in $\prod(X_i, U_i)$, but unless $U_i = X_i$ for almost all i (in which case $\prod(X_i, U_i) = \prod X_i$), it is not open in $\prod X_i$, and it is not open in the subspace topology on $\prod(X_i, U_i)$ because it does not contain the intersection of $\prod(X_i, U_i)$ with any basic open set in $\prod X_i$.

Thus the restricted product is a strict generalization of the direct product; the two coincide if and only if $U_i = X_i$ for almost all i. This is automatically true whenever the index set I is finite, so only infinite restricted products are of independent interest.

Remark 25.2. The restricted product does not depend on any particular U_i. Indeed,

$$\prod(X_i, U_i) = \prod(X_i, U_i')$$

whenever $U_i' = U_i$ for almost all i; note that the two restricted products are not merely isomorphic, they are identical, both as sets and as topological spaces. It is thus enough to specify the U_i for all but finitely many $i \in I$.

Each $x \in X := \prod(X_i, U_i)$ determines a (possibly empty) finite set

$$S(x) := \{i \in I : x_i \notin U_i\}.$$

Given any finite $S \subseteq I$, let us define

$$X_S := \{x \in X : S(x) \subseteq S\} = \prod_{i \in S} X_i \times \prod_{i \notin S} U_i.$$

Notice that $X_S \in \mathcal{B}$ is an open set, and we can view it as a topological space in two ways, both as a subspace of X or as a direct product of certain X_i and U_i. Restricting the basis $\mathcal{B}$ for X to a basis for the subspace X_S yields

$$\mathcal{B}_S := \left\{ \prod V_i : V_i \subseteq \pi_i(X_S) \text{ is open and } V_i = U_i = \pi_i(X_S) \text{ for almost all } i \in I \right\},$$

which is the standard basis for the product topology, so the two topologies on X_S coincide.

We have $X_S \subseteq X_T$ whenever $S \subseteq T$, thus if we partially order the finite subsets $S \subseteq I$ by inclusion, the family of topological spaces $\{X_S : S \subseteq I \text{ finite}\}$ with inclusion maps $\{i_{ST} \colon X_S \hookrightarrow X_T | S \subseteq T\}$ forms a *direct system*, and we have a corresponding *direct limit*

$$\varinjlim_S X_S := \coprod X_S / \sim,$$

which is the quotient of the coproduct space (disjoint union) $\coprod X_S$ by the equivalence relation $x \sim i_{ST}(x)$ for all $x \in S \subseteq T$.[1] This direct limit is canonically isomorphic to the restricted product X, which gives us another way to define the restricted product; before proving this let us recall the general definition of a direct limit of topological spaces.

[1]The topology on $\coprod X_S$ is the weakest topology that makes the injections $X_S \hookrightarrow \coprod X_S$ continuous; its open sets are disjoint unions of open sets in the X_S. The topology on $\coprod X_S / \sim$ is the weakest topology that makes the quotient map $\coprod X_S \to \coprod X_S / \sim$ continuous; its open sets are images of open sets in $\coprod X_S$.

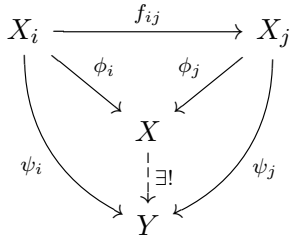
Definition 25.3. A *direct system* (or *inductive system*) in a category is a family of objects $\{X_i : i \in I\}$ indexed by a directed set I (see Definition 8.7) and a family of morphisms $\{f_{ij} \colon X_i \to X_j : i \le j\}$ such that each f_{ii} is the identity and $f_{ik} = f_{jk} \circ f_{ij}$ for all $i \le j \le k$.

Definition 25.4. Let (X_i, f_{ij}) be a direct system of topological spaces. The *direct limit* (or *inductive limit*) of (X_i, f_{ij}) is the quotient space

$$X = \varinjlim X_i := \coprod_{i \in I} X_i / \sim,$$

where $x_i \sim f_{ij}(x_i)$ for all $i \le j$. It is equipped with continuous maps $\phi_i \colon X_i \to X$ that are compositions of the inclusion maps $X_i \hookrightarrow \coprod X_i$ and quotient maps $\coprod X_i \twoheadrightarrow \coprod X_i / \sim$ and satisfy $\phi_i = \phi_j \circ f_{ij}$ for $i \le j$.

The topological space $X = \varinjlim X_i$ has the universal property that if Y is another topological space with continuous maps $\psi_i \colon X_i \to Y$ that satisfy $\psi_i = \psi_j \circ f_{ij}$ for $i \le j$, then there is a unique continuous map $X \to Y$ for which all of the diagrams



commute (this universal property defines the direct limit in any category with coproducts).

We now prove that that $\prod(X_i, U_i) \simeq \varinjlim X_S$ as claimed above.

Proposition 25.5. *Let (X_i) be a family of topological spaces indexed by $i \in I$, let (U_i) be a family of open sets $U_i \subseteq X_i$, and let $X := \prod(X_i, U_i)$ be the corresponding restricted product. For each finite $S \subseteq I$ define*

$$X_S := \prod_{i \in S} X_i \times \prod_{i \notin S} U_i \subseteq X,$$

and inclusion maps $i_{ST} \colon X_S \hookrightarrow X_T$, and let $\varinjlim X_S$ be the corresponding direct limit.

There is a canonical homeomorphism of topological spaces

$$\varphi \colon X \xrightarrow{\sim} \varinjlim X_S$$

that sends $x \in X$ to the equivalence class of $x \in X_{S(x)} \subseteq \coprod X_S$ in $\varinjlim X_S := \coprod X_S / \sim$, where $S(x) := \{i \in I : x_i \notin U_i\}$.

Proof. To prove that the map $\varphi \colon X \to \varinjlim X_S$ is a homeomorphism, we need to show that it is (1) a bijection, (2) continuous, and (3) an open map.

(1) For each equivalence class $\mathcal{C} \in \varinjlim X_S := \coprod X_S / \sim$, let $S(\mathcal{C})$ be the intersection of all the sets S for which $\mathcal{C}$ contains an element of $\coprod X_S$ in X_S. Then $S(x) = S(\mathcal{C})$ for all $x \in \mathcal{C}$, and $\mathcal{C}$ contains a unique element for which $x \in X_{S(x)} \subseteq \coprod X_S$ (distinct $x, y \in X_S$ cannot be equivalent). Thus φ is a bijection.

(2) Let U be an open set in $\varinjlim X_S = \coprod X_S / \sim$. The inverse image V of U in $\coprod X_S$ is open, as are the inverse images V_S of V under the canonical injections $\iota \colon X_S \hookrightarrow \coprod X_S$. The union of the V_S in X is equal to $\varphi^{-1}(U)$ and is an open set in X; thus φ is continuous.

(3) Let U be an open set in X. Since the X_S form an open cover of X, we can cover U with open sets $U_S := U \cap X_S$, and then $\coprod U_S$ is an open set in $\coprod X_S$. Moreover, for each $x \in \coprod U_S$, if $y \sim x$ for some $y \in \coprod X_S$ then y and x must correspond to the same element in U; in particular, $y \in \coprod U_S$, so $\coprod U_S$ is a union of equivalence classes in $\coprod X_S$. It follows that its image in $\varinjlim X_S = \coprod X_S/\sim$ is open. □

Proposition 25.5 gives us another way to construct the restricted product $\prod(X_i, U_i)$: rather than defining it as a subset of $\prod X_i$ with a modified topology, we can instead construct it as a limit of direct products that are subspaces of $\prod X_i$.

We now specialize to the case of interest, where we are forming a restricted product using a family $(X_i)_{i\in I}$ of locally compact spaces and a family of open subsets (U_i) that are almost all compact. Under these conditions the restricted product $\prod(X_i, U_i)$ is locally compact, even though the product $\prod X_i$ is not unless the index set I is finite.

Proposition 25.6. *Let $(X_i)_{i\in I}$ be a family of locally compact topological spaces and let $(U_i)_{i\in I}$ be a corresponding family of open subsets $U_i \subseteq X_i$ almost all of which are compact. Then the restricted product $X := \prod(X_i, U_i)$ is locally compact.*

Proof. We first note that for each finite set $S \subseteq I$ the topological space

$$X_S := \prod_{i\in S} X_i \times \prod_{i\notin S} U_i$$

can be viewed as a finite product of locally compact spaces, since all but finitely many U_i are compact, and the product of these is compact (by Tychonoff's theorem), hence locally compact. A finite product of locally compact spaces is locally compact, since we can construct compact neighborhoods as products of compact neighborhoods in each factor (in a finite product, products of open sets are open and products of compact sets are compact); thus the X_S are locally compact, and they cover X (since each $x \in X$ lies in $X_{S(x)}$). It follows that X is locally compact, since each $x \in X_S$ has a compact neighborhood $x \in U \subseteq C \subseteq X_S$ that is also a compact neighborhood in X (the image of C under the inclusion map $X_S \to X$ is certainly compact, and U is open in X because X_S is open in X). □

25.3 The ring of adeles

Recall that for a global field K (a finite extension of $\mathbb{Q}$ or $\mathbb{F}_q(t)$), we use M_K to denote the set of places of K (equivalence classes of absolute values), and for any $v \in M_K$ we use K_v to denote the corresponding local field (the completion of K with respect to v). When v is nonarchimedean we use $\mathcal{O}_v$ to denote the valuation ring of K_v, and for archimedean v we define $\mathcal{O}_v := K_v$.[2]

Definition 25.7. Let K be a global field. The *adele ring*[3] of K is the restricted product

$$\mathbb{A}_K := \prod(K_v, \mathcal{O}_v)_{v\in M_K},$$

which we may view as a subset (but not a subspace!) of $\prod_v K_v$; indeed

$$\mathbb{A}_K = \left\{(a_v) \in \prod K_v : a_v \in \mathcal{O}_v \text{ for almost all } v\right\}.$$

[2]Per Remark 25.2, as far as the topology goes it doesn't matter how we define $\mathcal{O}_v$ at the finite number of archimedean places, but we would like each $\mathcal{O}_v$ to be a topological ring, which motivates this choice.

[3]In French one writes *adèle*, but it is common practice to omit the accent when writing in English.

For each $a \in \mathbb{A}_K$ we use a_v to denote its projection in K_v; we make $\mathbb{A}_K$ a ring by defining addition and multiplication component-wise.

For each finite set of places S we have the subring of *S-adeles*

$$\mathbb{A}_{K,S} := \prod_{v \in S} K_v \times \prod_{v \notin S} \mathcal{O}_v,$$

which is a direct product of topological rings. By Proposition 25.5, $\mathbb{A}_K \simeq \varinjlim \mathbb{A}_{K,S}$ is the direct limit of the S-adele rings, which makes it clear that $\mathbb{A}_K$ is also a topological ring.[4]

The canonical embeddings $K \hookrightarrow K_v$ induce a canonical embedding

$$\begin{aligned} K &\hookrightarrow \mathbb{A}_K \\ x &\mapsto (x, x, x, \ldots). \end{aligned}$$

Note that for each $x \in K$ we have $x \in \mathcal{O}_v$ for all but finitely many v. The image of K in $\mathbb{A}_K$ is the subring of *principal adeles* (which of course is also a field).

We extend the normalized absolute value $\| \ \|_v$ of K_v (see Definition 13.17) to $\mathbb{A}_K$ via

$$\|a\|_v := \|a_v\|_v,$$

and define the *adelic absolute value* (or *adelic norm*)

$$\|a\| := \prod_{v \in M_K} \|a\|_v \in \mathbb{R}_{\geq 0}$$

which we note converges to zero unless $\|a\|_v = 1$ for all but finitely many v, in which case it is effectively a finite product.[5] For $\|a\| \neq 0$ this is equal to the size of the *M_K-divisor* $(\|a\|_v)$ we defined in Lecture 15 (see Definition 15.1). For any nonzero principal adele a, we have $a \in K^\times$ and $\|a\| = 1$, by the product formula (Theorem 13.21).

Example 25.8. For $K = \mathbb{Q}$ the adele ring $\mathbb{A}_\mathbb{Q}$ is the union of the rings

$$\mathbb{R} \times \prod_{p \in S} \mathbb{Q}_p \times \prod_{p \notin S} \mathbb{Z}_p$$

where S varies over finite sets of primes (but note that the topology is the restricted product topology, not the subspace topology in $\prod_{p \leq \infty} \mathbb{Q}_p$). We can also write $\mathbb{A}_\mathbb{Q}$ as

$$\mathbb{A}_\mathbb{Q} = \left\{ a \in \prod_{p \leq \infty} \mathbb{Q}_p : \|a\|_p \leq 1 \text{ for almost all } p \right\}.$$

Proposition 25.9. *The adele ring $\mathbb{A}_K$ of a global field K is locally compact and Hausdorff.*

Proof. Local compactness follows from Proposition 25.6, since the local fields K_v are all locally compact and all but finitely many $\mathcal{O}_v$ are valuation rings of a nonarchimedean local field, hence compact ($\mathcal{O}_v = \{x \in K_v : \|x\|_v \leq 1\}$ is a closed ball). The product space $\prod_v K_v$ is Hausdorff, since each K_v is Hausdorff, and the topology on $\mathbb{A}_K \subseteq \prod K_v$ is finer than the subspace topology, so $\mathbb{A}_K$ is also Hausdorff. □

[4] By definition it is a topological space that is also a ring; to be a topological ring is a stronger condition (the ring operations must be continuous), but this property is preserved by direct limits so all is well.

[5] For $v \nmid \infty$, if $\|a\|_v < 1$ then $\|a\|_v \leq 1/2$, since $\|a\|_v := q^{-v(a_v)}$ for some prime power q.

Proposition 25.9 implies that the additive group of $\mathbb{A}_K$ (which is sometimes denoted $\mathbb{A}_K^+$ to emphasize that we are viewing it as a group rather than a ring) is a locally compact group, and therefore has a Haar measure that is unique up to scaling, by Theorem 13.14. Each of the completions K_v is a local field with a Haar measure μ_v, which we normalize as follows:

- $\mu_v(\mathcal{O}_v) = 1$ for all nonarchimedean v;
- $\mu_v(S) = \mu_{\mathbb{R}}(S)$ for $K_v \simeq \mathbb{R}$, where $\mu_{\mathbb{R}}(S)$ is the Lebesgue measure on $\mathbb{R}$;
- $\mu_v(S) = 2\mu_{\mathbb{C}}(S)$ for $K_v \simeq \mathbb{C}$, where $\mu_{\mathbb{C}}(S)$ is the Lebesgue measure on $\mathbb{C} \simeq \mathbb{R} \times \mathbb{R}$.

Note that the normalization of μ_v at the archimedean places is consistent with the measure μ on $K_{\mathbb{R}} \simeq \mathbb{R}^r \times \mathbb{C}^s \simeq \mathbb{R}^n$ induced by the canonical inner product on $K_{\mathbb{R}} \subseteq K_{\mathbb{C}}$ that we defined in Lecture 14 (see §14.2).

We now define a measure μ on $\mathbb{A}_K$ as follows. We take as a basis for the σ-algebra of measurable sets all sets of the form $\prod_v B_v$, where each B_v is a measurable set in K_v with $\mu_v(B_v) < \infty$ such that $B_v = \mathcal{O}_v$ for almost all v (the σ-algebra is then generated by taking countable intersections, unions, and complements in $\mathbb{A}_K$). We then define

$$\mu\left(\prod_v B_v\right) := \prod_v \mu_v(B_v).$$

It is easy to verify that μ is a Radon measure, and it is clearly translation invariant since each of the Haar measures μ_v is translation invariant and addition is defined componentwise; note that for any $x \in \mathbb{A}_K$ and measurable set $B = \prod_v B_v$ the set $x + B = \prod_v (x_v + B_v)$ is also measurable, since $x_v + B_v = \mathcal{O}_v$ whenever $x_v \in \mathcal{O}_v$ and $B_v = \mathcal{O}_v$, and this applies to almost all v. It follows from uniqueness of the Haar measure (up to scaling) that μ is a Haar measure on $\mathbb{A}_K$ which we henceforth adopt as our normalized Haar measure on $\mathbb{A}_K$.

We now want to understand the behavior of the adele ring $\mathbb{A}_K$ under base change. Note that the canonical embedding $K \hookrightarrow \mathbb{A}_K$ makes $\mathbb{A}_K$ a K-vector space, and if L/K is any finite separable extension of K (also a K-vector space), we may consider the tensor product

$$\mathbb{A}_K \otimes_K L,$$

which is also an L-vector space. As a topological K-vector space, the topology on $\mathbb{A}_K \otimes L$ is just the product topology on $[L : K]$ copies of of $\mathbb{A}_K$ (this applies whenever we take a tensor product of topological vector spaces, one of which has finite dimension).

Proposition 25.10. *Let L be a finite separable extension of a global field K. There is a natural isomorphism of topological rings*

$$\mathbb{A}_L \simeq \mathbb{A}_K \otimes_K L$$

that makes the following diagram commute

$$\begin{array}{ccc} L & \xrightarrow{\sim} & K \otimes_K L \\ \downarrow & & \downarrow \\ \mathbb{A}_L & \xrightarrow{\sim} & \mathbb{A}_K \otimes_K L \end{array}$$

Proof. On the RHS the tensor product $\mathbb{A}_K \otimes_K L$ is isomorphic to the restricted product

$$\prod_{v \in M_K} (K_v \otimes_K L,\, \mathcal{O}_v \otimes_{\mathcal{O}_K} \mathcal{O}_L).$$

Explicitly, each element of $\mathbb{A}_K \otimes_K L$ is a finite sum of elements of the form $(a_v) \otimes x$, where $(a_v) \in \mathbb{A}_K$ and $x \in L$, and there is a natural isomorphism of topological rings

$$\begin{aligned} \mathbb{A}_K \otimes_K L &\xrightarrow{\sim} \prod_{v \in M_K} (K_v \otimes_K L,\, \mathcal{O}_v \otimes_{\mathcal{O}_K} \mathcal{O}_L) \\ (a_v) \otimes x &\mapsto (a_v \otimes x). \end{aligned}$$

Here we are using the general fact that tensor products commute with direct limits (restricted direct products can be viewed as direct limits via Proposition 25.5).[6]

On the LHS we have $\mathbb{A}_L := \prod_{w \in M_L} (L_w, \mathcal{O}_w)$. But note that $K_v \otimes_K L \simeq \prod_{w|v} L_w$, by Theorem 11.23 and $\mathcal{O}_v \otimes_{\mathcal{O}_K} \mathcal{O}_L \simeq \prod_{w|v} \mathcal{O}_w$, by Corollary 11.26. These isomorphisms preserve both the algebraic and the topological structures of both sides, and it follows that

$$\mathbb{A}_K \otimes_K L \simeq \prod_{v \in M_K} (K_v \otimes_K L,\, \mathcal{O}_v \otimes_{\mathcal{O}_K} \mathcal{O}_L) \simeq \prod_{w \in M_L} (L_w, \mathcal{O}_w) = \mathbb{A}_L$$

is an isomorphism of topological rings. The image of $x \in L$ in $\mathbb{A}_K \otimes_K L$ via the canonical embedding of L into $\mathbb{A}_K \otimes_K L$ is $1 \otimes x = (1, 1, 1, \ldots) \otimes x$, whose image $(x, x, x, \ldots) \in \mathbb{A}_L$ is equal to the image of $x \in L$ under the canonical embedding of L into its adele ring $\mathbb{A}_L$. □

Corollary 25.11. *Let L be a finite separable extension of a global field K of degree n. There is a natural isomorphism of topological K-vector spaces (and locally compact groups)*

$$\mathbb{A}_L \simeq \mathbb{A}_K \oplus \cdots \oplus \mathbb{A}_K$$

that identifies $\mathbb{A}_K$ with the direct sum of n copies of $\mathbb{A}_K$, and this isomorphism restricts to an isomorphism $L \simeq K \oplus \cdots \oplus K$ of the principal adeles of $\mathbb{A}_L$ with the n-fold direct sum of the principal adeles of $\mathbb{A}_K$.

Theorem 25.12. *For each global field L the principal adeles $L \subseteq \mathbb{A}_L$ form a discrete cocompact subgroup of the additive group of the adele ring $\mathbb{A}_L$.*

Proof. Let K be the rational subfield of L (so $K = \mathbb{Q}$ or $K = \mathbb{F}_q(t)$). It follows from Corollary 25.11 that if the theorem holds for K then it holds for L, so we will prove the theorem for K. Let us identify K with its image in $\mathbb{A}_K$ (the principal adeles).

To show that the topological group K is discrete in the locally compact group $\mathbb{A}_K$, it suffices to show that 0 is an isolated point. Consider the open set

$$U = \{a \in \mathbb{A}_K : \|a\|_\infty < 1 \text{ and } \|a\|_v \le 1 \text{ for all } v < \infty\},$$

where ∞ denotes the unique infinite place of K (either the real place of $\mathbb{Q}$ or the place corresponding to the nonarchimedean valuation $v_\infty(f/g) = \deg g - \deg f$ of $\mathbb{F}_q(t)$). The product formula (Theorem 13.21) implies $\|a\| = 1$ for all $a \in K^\times \subseteq \mathbb{A}_K$, so $U \cap K = \{0\}$.

To prove that the quotient $\mathbb{A}_K/K$ is compact, we consider the set

$$W := \{a \in \mathbb{A}_K : \|a\|_v \le 1 \text{ for all } v\}.$$

[6] In general, tensor products *do not* commute with infinite direct products; there is always a natural map $(\prod_n A_n) \otimes B \to \prod_n (A_n \otimes B)$, but it need be neither a monomorphism or an epimorphism. This is another motivation for using restricted direct products to define the adeles, so that base change works as it should.

If we let $U_\infty := \{x \in K_\infty : \|x\|_\infty \le 1\}$, then

$$W = U_\infty \times \prod_{v<\infty} \mathcal{O}_v \subseteq \mathbb{A}_{K,\{\infty\}} \subseteq \mathbb{A}_K$$

is a product of compact sets and therefore compact. We will show that W contains a complete set of coset representatives for K in $\mathbb{A}_K$. This implies that $\mathbb{A}_K/K$ is the image of the compact set W under the (continuous) quotient map $\mathbb{A}_K \to \mathbb{A}_K/K$, hence compact.

Let $a = (a_v)$ be any element of $\mathbb{A}_K$. We wish to show that $a = b + c$ for some $b \in W$ and $c \in K$, which we will do by constructing $c \in K$ so that $b = a - c \in W$.

For each $v < \infty$ define $x_v \in K$ as follows: put $x_v := 0$ if $\|a_v\|_v \le 1$ (almost all v), and otherwise choose $x_v \in K$ so that $\|a_v - x_v\|_v \le 1$ and $\|x_v\|_w \le 1$ for $w \ne v$. To show that such an x_v exists, let us first suppose $a_v = r/s \in K$ with $r, s \in \mathcal{O}_K$ coprime (note that $\mathcal{O}_K$ is a PID), and let $\mathfrak{p}$ be the maximal ideal of $\mathcal{O}_v$. The ideals $\mathfrak{p}^{v(s)}$ and $\mathfrak{p}^{-v(s)}(s)$ are coprime, so we can write $r = r_1 + r_2$ with $r_1 \in \mathfrak{p}^{v(s)}$ and $r_2 \in \mathfrak{p}^{-v(s)}(s) \subseteq \mathcal{O}_K$, so that $a_v = r_1/s + r_2/s$ with $v(r_1/s) \ge 0$ and $w(r_2/s) \ge 0$ for all $w \ne v$. If we now put $x_v := r_2/s$, then $\|a_v - x_v\|_v = \|r_1/s\|_v \le 1$ and $\|x_v\|_w = \|r_2/s\|_w \le 1$ for all $w \ne v$ as desired. We can approximate any $a'_v \in K_v$ by such an $a_v \in K$ with $\|a'_v - a_v\|_v < \epsilon$ and construct x_v as above so that $\|a_v - x_v\|_v \le 1$ and $\|a'_v - x_v\|_v \le 1 + \epsilon$; but for sufficiently small ϵ this implies $\|a'_v - x_v\|_v \le 1$, since the nonarchimedean absolute value $\| \ \|_v$ is discrete.

Finally, let $x := \sum_{v<\infty} x_v \in K$ and choose $x_\infty \in \mathcal{O}_K$ so that

$$\|a_\infty - x - x_\infty\|_\infty \le 1.$$

For $a_\infty - x \in \mathbb{Q}_\infty \simeq \mathbb{R}$, we can take $x_\infty \in \mathbb{Z}$ in the real interval $[a_\infty - x - 1, a_\infty - x + 1)$. For $a_\infty - x \in \mathbb{F}_q(t)_\infty \simeq \mathbb{F}_q((t^{-1}))$ we can take $x_\infty \in \mathbb{F}_q[t]$ to be the polynomial of least degree for which $a_\infty - x - x_\infty \in \mathbb{F}_q[[t^{-1}]]$.[7]

Now let $c := \sum_{v\le\infty} x_v \in K \subseteq \mathbb{A}_K$, and let $b := a - c$. Then $a = b + c$, with $c \in K$, and we claim that $b \in W$. For each $v < \infty$ we have $x_w \in \mathcal{O}_v$ for all $w \ne v$ and

$$\|b\|_v = \|a - c\|_v = \left\| a_v - \sum_{w\le\infty} x_w \right\|_v \le \max\left(\|a_v - x_v\|_v, \max(\{\|x_w\|_v : w \ne v\})\right) \le 1,$$

by the nonarchimedean triangle inequality. For $v = \infty$ we have $\|b\|_\infty = \|a_\infty - c\|_\infty \le 1$ by our choice of x_∞, and $\|b\|_v \le 1$ for all v, so $b \in W$ as claimed and the theorem follows. □

Corollary 25.13. *For any global field K the quotient $\mathbb{A}_K/K$ is a compact group.*

Proof. As explained in Remark 14.4, this follows immediately (in particular, the fact that K is a discrete subgroup of $\mathbb{A}_K$ implies that it is closed and therefore $\mathbb{A}_K/K$ is Hausdorff). □

25.4 Strong approximation

We are now ready to prove the strong approximation theorem, an important result that has many applications. We begin with an adelic version of the Blichfeldt-Minkowski lemma.

[7] Note that while $\mathbb{F}_q((t^{-1})) \simeq \mathbb{F}_q((t))$, in order to view $K = \mathbb{F}_q(t)$ as canonically embedded in its completion with respect to the absolute value $|f|_\infty = q^{\deg f}$ we need to view K_∞ as the field of Laurent series in a uniformizer, which we may take to be t^{-1} (but not t), and the valuation ring of K_∞ is $\mathbb{F}_q[[-t]]$ (not $\mathbb{F}_q[[t]]$).

Lemma 25.14 (Adelic Blichfeldt-Minkowski lemma)**.** *Let K be a global field. There is a positive constant B_K such that for any $a \in \mathbb{A}_K$ with $\|a\| > B_K$ there exists a nonzero principal adele $x \in K \subseteq \mathbb{A}_K$ for which $\|x\|_v \le \|a\|_v$ for all $v \in M_K$.*

Proof. Let $b_0 := \operatorname{covol}(K)$ be the measure of a fundamental region for K in $\mathbb{A}_K$ under our normalized Haar measure μ on $\mathbb{A}_K$ (by Theorem 25.12, K is cocompact, so b_0 is finite). Now define

$$b_1 := \mu\left(\left\{z \in \mathbb{A}_K : \|z\|_v \le 1 \text{ for all } v \text{ and } \|z\|_v \le \tfrac{1}{4} \text{ if } v \text{ is archimedean}\right\}\right).$$

Then $b_1 \neq 0$, since K has only finitely many archimedean places. Now let $B_K := b_0/b_1$.

Suppose $a \in \mathbb{A}_K$ satisfies $\|a\| > B_K$. We know that $\|a\|_v \le 1$ for all almost all v, so $\|a\| \neq 0$ implies that $\|a\|_v = 1$ for almost all v. Let us now consider the set

$$T := \left\{t \in \mathbb{A}_K : \|t\|_v \le \|a\|_v \text{ for all } v \text{ and } \|t\|_v \le \tfrac{1}{4}\|a\|_v \text{ if } v \text{ is archimedean}\right\}.$$

From the definition of b_1 we have

$$\mu(T) = b_1\|a\| > b_1 B_K = b_0;$$

this follows from the fact that the Haar measure on $\mathbb{A}_K$ is the product of the normalized Haar measures μ_v on each of the K_v. Since $\mu(T) > b_0$, the set T is not contained in any fundamental region for K, so there must be distinct $t_1, t_2 \in T$ with the same image in $\mathbb{A}_K/K$, equivalently, whose difference $x = t_1 - t_2$ is a nonzero element of $K \subseteq \mathbb{A}_K$. We have

$$\|t_1 - t_2\|_v \le \begin{cases} \max(\|t_1\|_v, \|t_2\|_v) \le \|a\|_v & \text{nonarch. } v; \\ \|t_1\|_v + \|t_2\|_v \le 2 \cdot \frac{1}{4}\|a\|_v \le \|a\|_v & \text{real } v; \\ (\|t_1 - t_2\|_v^{1/2})^2 \le (\|t_1\|_v^{1/2} + \|t_2\|_v^{1/2})^2 \le (2 \cdot \frac{1}{2}\|a\|_v^{1/2})^2 \le \|a\|_v & \text{complex } v. \end{cases}$$

Here we have used the fact that the normalized absolute value $\| \ \|_v$ satisfies the nonarchimedean triangle inequality when v is nonarchimedean, $\| \ \|_v$ satisfies the archimedean triangle inequality when v is real, and $\| \ \|_v^{1/2}$ satisfies the archimedean triangle inequality when v is complex. Thus $\|x\|_v = \|t_1 - t_2\|_v \le \|a\|_v$ for all places $v \in M_K$ as desired. □

Remark 25.15. Lemma 25.14 should be viewed as an analog of Mikowski's lattice point theorem (Thoerem 14.12) and a generalization of Proposition 15.9. In Theorem 14.12 we have a discrete cocompact subgroup Λ in a real vector space $V \simeq \mathbb{R}^n$ and a sufficiently large symmetric convex set S that must contain a nonzero element of Λ. In Lemma 25.14 the lattice Λ is replaced by K, the vector space V is replaced by $\mathbb{A}_K$, the symmetric convex set S is replaced by the set

$$L(a) := \{x \in \mathbb{A}_K : \|x\|_v \le \|a\|_v \text{ for all } v \in M_K\},$$

and sufficiently large means $\|a\| > B_K$, putting a lower bound on $\mu(L(a))$. Proposition 15.9 is actually equivalent to Lemma 25.14 in the case that K is a number field: use the M_K-divisor $c := (\|a\|_v)$ and note that $L(c) = L(a) \cap K$.

Theorem 25.16 (Strong Approximation)**.** *Let $M_K = S \sqcup T \sqcup \{w\}$ be a partition of the places of a global field K with S finite. Fix $a_v \in K$ and $\epsilon_v \in \mathbb{R}_{>0}$ for each $v \in S$. There exists an $x \in K$ for which*

$$\|x - a_v\|_v \le \epsilon_v \text{ for all } v \in S,$$
$$\|x\|_v \le 1 \text{ for all } v \in T,$$

(note that there is no constraint on $\|x\|_w$).

Proof. Let $W = \{z \in \mathbb{A}_K : \|z\|_v \le 1 \text{ for all } v \in M_K\}$ as in the proof of Theorem 25.12. Then W contains a complete set of coset representatives for $K \subseteq \mathbb{A}_K$, so $\mathbb{A}_K = K + W$. For any nonzero $u \in K \subseteq \mathbb{A}_K$ we also have $\mathbb{A}_K = K + uW$: given $c \in \mathbb{A}_K$ write $u^{-1}c \in \mathbb{A}_K$ as $u^{-1}c = a + b$ with $a \in K$ and $b \in W$ and then $c = ua + ub$ with $ua \in K$ and $ub \in uW$. Now choose $z \in \mathbb{A}_K$ such that

$$0 < \|z\|_v \le \epsilon_v \text{ for } v \in S, \quad 0 < \|z\|_v \le 1 \text{ for } v \in T, \quad \|z\|_w > B \prod_{v \ne w} \|z\|_v^{-1},$$

where B is the constant in the Blichfeldt-Minkowski Lemma 25.14 (this is clearly possible: every $z = (z_v)$ with $\|z_v\|_v \le 1$ is an element of $\mathbb{A}_K$). We have $\|z\| > B$, so there is a nonzero $u \in K \subseteq \mathbb{A}_K$ with $\|u\|_v \le \|z\|_v$ for all $v \in M_K$.

Now let $a = (a_v) \in \mathbb{A}_K$ be the adele with a_v given by the hypothesis of the theorem for $v \in S$ and $a_v = 0$ for $v \notin S$. We have $\mathbb{A}_K = K + uW$, so $a = x + y$ for some $x \in K$ and $y \in uW$. Therefore

$$\|x - a_v\|_v = \|y\|_v \le \|u\|_v \le \|z\|_v \le \begin{cases} \epsilon_v & \text{for } v \in S, \\ 1 & \text{for } v \in T, \end{cases}$$

as desired. □

Corollary 25.17. *Let K be a global field and let w be any place of K. Then K is dense in the restricted product $\prod_{v \ne w}(K_v, \mathcal{O}_v)$.*

Remark 25.18. Theorem 25.16 and Corollary 25.17 can be generalized to algebraic groups; see [1] for a survey.

References

[1] Andrei S. Rapinchuk, *Strong approximation for algebraic groups*, Thin groups and superstrong approximation, MSRI Publications **61**, 2013.

26 The idele group, profinite groups, infinite Galois theory

26.1 The idele group

Let K be a global field. Having introduced the ring of adeles $\mathbb{A}_K$ in the previous lecture, it is natural to ask about its unit group

$$\mathbb{A}_K^\times = \{(a_v) \in \mathbb{A}_K : a_v \in K_v^\times \text{ for all } v \in M_K, \text{ and } a_v \in \mathcal{O}_v^\times \text{ for almost all } v \in M_K\}.$$

Here $\mathcal{O}_v^\times := K_v^\times \cap \mathcal{O}_v$ is the unit group of the valuation ring of K_v when v is nonarchimedean and isomorphic to $\mathbb{R}^\times$ or $\mathbb{C}^\times$ when v is archimedean. As noted in Lecture 25, the definition of $\mathbb{A}_K$ does not actually depend on our choice of $\mathcal{O}_v$ at the finitely many archimedean places of K, but the choice we made ensures that every $\mathcal{O}_v^\times$ is a topological group.

However, as a subspace of $\mathbb{A}_K$, the unit group $\mathbb{A}_K^\times$ is not a topological group. Indeed, the inversion map $a \mapsto a^{-1}$ is not continuous.

Example 26.1. Consider $K = \mathbb{Q}$ and for each prime p let $a(p) = (1, \dots, 1, p, 1, \dots) \in \mathbb{A}_\mathbb{Q}$ be the adele with $a(p)_p = p$ and $a(p)_q = 1$ for $q \neq p$. Every basic open set U about 1 in $\mathbb{A}_\mathbb{Q}$ has the form

$$U = \prod_{v \in S} U_v \times \prod_{V \notin S} \mathcal{O}_v,$$

with $S \subseteq M_\mathbb{Q}$ finite and $1_v \in U_v$, and it is clear that U contains $a(p)$ for all sufficiently large p. It follows that $\lim_{p\to\infty} a(p) = 1$ in the topology of $\mathbb{A}_\mathbb{Q}$. But notice that U does not contain $a(p)^{-1}$ for any sufficiently large p, so $\lim_{p\to\infty} a(p)^{-1} \neq 1^{-1}$ in $\mathbb{A}_\mathbb{Q}$. Thus the function $a \to a^{-1}$ is not continuous in the subspace topology for $\mathbb{A}_K^\times$.

This problem is not specific to rings of adeles. For a topological ring R there is in general no reason to expect its unit group $R^\times \subseteq R$ to be a topological group in the subspace topology. One notable exception is when R is a subring of a topological field (the definition of which requires inversion to be continuous), as is the case for the unit group $\mathcal{O}_K^\times$; this explains why we have not encountered this problem before now. But the ring of adeles is not naturally contained in any topological field (note that it is not an integral domain).

There is a standard solution to this problem: give the group $R^\times$ the weakest topology that makes it a topological group. This is done by embedding $R^\times$ in $R \times R$ via the map

$$\begin{aligned} \phi\colon R^\times &\to R \times R \\ r &\mapsto (r, r^{-1}). \end{aligned}$$

We now declare ϕ to be a homeomorphism; that is, we endow $R^\times$ with the topology matching the subspace topology of $\phi(R^\times) \subset R \times R$. The inversion map $r \mapsto r^{-1}$ is continuous in this topology because it is equal to composition of ϕ with the projection map $R \times R \to R$ onto its second coordinate, both of which are continuous maps.

We now consider this construction in the case of $\mathbb{A}_K^\times$. The implied topology on $\mathbb{A}_K^\times$ has a basis of open sets of the form

$$U' = \prod_{v \in S} U_v \times \prod_{v \notin S} \mathcal{O}_v^\times$$

where $U_v \subseteq K_v^\times$ and $S \subseteq M_K$ is finite. To see this, note that in terms of the embedding $\phi\colon \mathbb{A}_K^\times \to \mathbb{A}_K \times \mathbb{A}_K$ defined above, each $\phi(a) = (a, a^{-1})$ lies in a product $U \times V$ of basic

open sets $U, V \subseteq \mathbb{A}_K$, and this forces both a and a^{-1} to lie in $\mathcal{O}_v$, hence in $\mathcal{O}_v^\times$, for almost all v. The open sets U' are precisely the open sets in the restricted product $\prod(K_v^\times, \mathcal{O}_v^\times)$. This leads to the following definition.

Definition 26.2. Let K be a global field. The *idele group* of K is the topological group

$$\mathbb{I}_K := \prod\nolimits_v (K_v^\times, \mathcal{O}_v^\times)$$

with multiplication defined component-wise, which we view as the subgroup $\mathbb{A}_K^\times$ of $\mathbb{A}_K$ endowed with the restricted product topology rather than the subspace topology. The canonical embedding $K \hookrightarrow \mathbb{A}_K$ restricts to a canonical embedding $K^\times \hookrightarrow \mathbb{I}_K$, and we define the *idele class group* $C_K := \mathbb{I}_K/K^\times$, a topological group.

Remark 26.3. In the literature one finds the notations $\mathbb{I}_K$ and $\mathbb{A}_K^\times$ used interchangeably; they both denote the idele group defined above. But in this lecture we will temporarily use the notation $\mathbb{A}_K^\times$ to denote the unit group of the ring $\mathbb{A}_K$ in the subspace topology (which is not a topological group).

Example 26.4. Let us again consider the sequence $(a(p))$ defined in Example 26.1. This sequence lies in $\mathbb{A}_\mathbb{Q}^\times$ and converges to $1 \in \mathbb{A}_\mathbb{Q}^\times$ under the subspace topology. But this sequence does not converge to 1 in the topology of $\mathbb{I}_\mathbb{Q}$. Indeed, consider the basic open set $\prod_v \mathcal{O}_v^\times = \prod_p \mathbb{Z}_p^\times \times \mathbb{R}^\times$ of $\mathbb{I}_\mathbb{Q}$. None of the $a(p) = (1, \dots, 1, p, 1, \dots)$ lie in this open neighborhood of 1, so the sequence $(a(p))$ cannot converge to 1 in $\mathbb{I}_\mathbb{Q}$ (which means it cannot converge at all: if it converged to $x \neq 1$ in $\mathbb{I}_\mathbb{Q}$ it would converge to $x \neq 1$ in $\mathbb{A}_\mathbb{Q}^\times \subseteq \mathbb{A}_\mathbb{Q}$, which we know is not the case). The counterexample to the continuity of the inversion map $x \mapsto x^{-1}$ in $\mathbb{A}_\mathbb{Q}^\times$ is removed in $\mathbb{I}_\mathbb{Q}$ by adding more open sets to the topology; this makes it easier for maps to be continuous and harder for sequences to converge.

We now define a surjective homomorphism

$$\begin{aligned} \mathbb{I}_K &\to \mathcal{I}_K \\ a &\mapsto \prod \mathfrak{p}^{v_\mathfrak{p}(a)} \end{aligned}$$

where the product ranges over primes $\mathfrak{p}$ of K and $v_\mathfrak{p}(a) := v_\mathfrak{p}(a_v)$, where v is the equivalence class of the $\mathfrak{p}$-adic absolute value $\| \ \|_\mathfrak{p}$. The composition

$$K^\times \hookrightarrow \mathbb{I}_K \twoheadrightarrow \mathcal{I}_K$$

has image $\mathcal{P}_K$, the subgroup of principal fractional ideals; we thus have a surjective homomorphism of the idele class group $C_K = \mathcal{I}_K/K^\times$ onto the ideal class group $\mathrm{Cl}_K = \mathcal{I}_K/\mathcal{P}_K$ and a commutative diagram of exact sequences:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & K^\times & \longrightarrow & \mathbb{I}_K & \longrightarrow & C_K & \longrightarrow & 1 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 1 & \longrightarrow & \mathcal{P}_K & \longrightarrow & \mathcal{I}_K & \longrightarrow & \mathrm{Cl}_K & \longrightarrow & 1 \end{array}$$

Proposition 26.5. *Let K be a global field. The idele group $\mathbb{I}_K$ is a locally compact group.*

Proof. It is clear that $\mathbb{I}_K$ is Hausdorff, since its topology is finer than the topology of $\mathbb{A}_K^\times \subseteq \mathbb{A}_K$, which is Hausdorff by Proposition 25.9. For each nonarchimedean place v, the set $\mathcal{O}_v^\times = \{x \in K_v^\times : \|x\|_v = 1\}$ is a closed subset of the compact set $\mathcal{O}_v$, hence compact. This applies to almost all $v \in M_K$, and the $K_v^\times$ are all locally compact, so the restricted product $\prod(K_v^\times, \mathcal{O}_v^\times) = \mathbb{I}_K$ is locally compact, by Proposition 25.6. □

Proposition 26.6. *Let K be a global field. Then $K^\times$ is a discrete subgroup of $\mathbb{I}_K$.*

Proof. We have $K^\times \hookrightarrow K \times K \subseteq \mathbb{A}_K \times \mathbb{A}_K$. By Theorem 25.12, K is a discrete subset of $\mathbb{A}_K$, and it follows that $K \times K$ is a discrete subset of $\mathbb{A}_K \times \mathbb{A}_K$. The image of $K^\times$ in $\mathbb{A}_K \times \mathbb{A}_K$ lies in the image of $\mathbb{A}_K^\times \hookrightarrow \mathbb{A}_K \times \mathbb{A}_K$ and in the discrete image of $K \hookrightarrow \mathbb{A}_K \times \mathbb{A}_K$, and it follows that $K^\times$ is discrete in $\mathbb{A}_K^\times$ and therefore in $\mathbb{I}_K$, since having a finer topology only makes it easier for a set to be discrete. □

We proved last time that K is a discrete cocompact subgroup of $\mathbb{A}_K$, so it is natural to ask whether $K^\times$ is a cocompact in $\mathbb{A}_K^\times$ or $\mathbb{I}_K$. The answer is no, $K^\times$ is not a cocompact subgroup of $\mathbb{I}_K$, thus the idele class group C_K, while locally compact, is not compact.

Recall that for a number field K, the unit group $\mathcal{O}_K^\times$ is not a cocompact subgroup of $K_{\mathbb{R}}^\times$ because $\mathrm{Log}(\mathcal{O}_K^\times)$ is not a (full) lattice in $\mathbb{R}^{r+s} \simeq \mathrm{Log}(K_{\mathbb{R}}^\times)$; it lies in the trace zero hyperplane $\mathbb{R}_0^{r+s}$ (see Proposition 15.11). In order to get a cocompact subgroup we need to restrict $\mathbb{I}_K$ to a subgroup that corresponds to the trace zero hyperplane.

We have a continuous homomorphism of topological groups

$$\begin{aligned} \| \| \colon \mathbb{I}_K &\to \mathbb{R}_{>0}^\times \\ a &\mapsto \|a\| \end{aligned}$$

where $\|a\| := \prod_v \|a\|_v$ is the adelic norm defined in the previous lecture. We have $\|a\| > 0$ for $a \in \mathbb{I}_K$, since $a_v \in \mathcal{O}_v^\times$ for almost all v: this implies that $\|a\|_v = 1$ for almost all v and the product $\prod_v \|a\|_v$ is effectively a finite product, and it is nonzero because $a_v \in K_v^\times$ is nonzero for all $v \in M_K$.

Definition 26.7. Let K be a global field. The group of 1-*ideles* is the topological group

$$\mathbb{I}_K^1 := \ker \| \| = \{a \in \mathbb{I}_K : \|a\| = 1\},$$

which we note contains $K^\times$, by the product formula (Theorem 13.21).

A useful feature of the group of 1-ideles is that, unlike the group of ideles, its topology is the same as the subspace topology it inherits from $\mathbb{A}_K$.

Lemma 26.8. *The group of 1-ideles $\mathbb{I}_K^1$ is a closed subset of $\mathbb{A}_K$ and $\mathbb{I}_K$, and the two subspace topologies on $\mathbb{I}_K^1$ coincide.*

Proof. We first show that $\mathbb{I}_K^1$ is closed in $\mathbb{A}_K$, and therefore also in $\mathbb{I}_K$, since it has a finer topology. Consider any $x \in \mathbb{A}_K - \mathbb{I}_K^1$. We will construct an open neighborhood U_x of x that is disjoint from $\mathbb{I}_K^1$. The union of the U_x is then the open complement of the closed set $\mathbb{I}_K^1$. For each $\epsilon > 0$, finite $S \subseteq M_K$, and $x \in \mathbb{A}_K$ we define

$$U_\epsilon(x, S) := \{u \in \mathbb{A}_K : \|u - x\|_v < \epsilon \text{ for } v \in S \text{ and } \|u\|_v \le 1 \text{ for } v \notin S\},$$

which is a basic open set of $\mathbb{A}_K$ (a product of open sets U_v for $v \in S$ and $\mathcal{O}_v$ for $v \notin S$).

The case $\|x\| < 1$. Let S be a finite set containing the archimedean places $v \in M_K$ and all v for which $\|x\|_v > 1$, such that $\prod_{v \in S} \|x\|_v < 1$: such an S exists since $\|x\| < 1$ and $\|x\|_v \le 1$ for almost all v. For all sufficiently small $\epsilon > 0$ the set $U_x := U_\epsilon(x, S)$ is an open neighborhood of x disjoint from $\mathbb{I}_K^1$ because every $y \in U_x$ must satisfy $\|y\| < 1$.

The case $\|x\| > 1$. Let B be twice the product of all the $\|x\|_v$ greater than 1. Let S be the finite set containing the archimedean places $v \in M_K$, all nonarchimedean v with

residue field cardinality less than $2B$, and all v for which $\|x\|_v > 1$. For all sufficiently small $\epsilon > 0$ the set $U_x := U_\epsilon(x, S)$ is an open neighborhood of x disjoint from $\mathbb{I}_K^1$ because for every $y \in U_x$, either $\|y\|_v = 1$ for all $v \notin S$, in which case $\|y\| > 1$, or $\|y\|_v < 1$ for some $v \notin S$, in which case $\|y\|_v < 1/(2B)$ and $\|y\| < 1$.

This proves that $\mathbb{I}_K^1$ is closed in $\mathbb{A}_K$, and therefore also in $\mathbb{I}_K$. To prove that the subspace topologies coincide, it suffices to show that for every $x \in \mathbb{I}_K^1$ and open $U \subseteq \mathbb{I}_K$ containing x there exists open sets $V \subseteq \mathbb{I}_K$ and $W \subseteq \mathbb{A}_K$ such that $x \in V \subseteq U$ and $V \cap \mathbb{I}_K^1 = W \cap \mathbb{I}_K^1$; this implies that every neighborhood basis in the subspace topology of $\mathbb{I}_K^1 \subseteq \mathbb{I}_K$ is a neighborhood basis in the subspace topology of $\mathbb{I}_K^1 \subseteq \mathbb{A}_K$ (the latter is *a priori* coarser than the former).

So consider any $x \in \mathbb{I}_K^1$ and open neighborhood $U \subseteq \mathbb{I}_K$ of x. Then U contains a basic open set

$$V = \{u \in \mathbb{A}_K : \|u - x\|_v < \epsilon \text{ for } v \in S \text{ and } \|u\|_v = 1 \text{ for } v \notin S\},$$

for some $\epsilon > 0$ and finite $S \subseteq M_K$ (take $S = \{v \in M_K : \|x\| \neq 1\}$ and $\epsilon > 0$ small enough). If we now put $W := U_\epsilon(x, S)$ then $x \in V \subseteq U$ and $V \cap \mathbb{I}_K^1 = W \cap \mathbb{I}_K^1$ as desired. □

Theorem 26.9. *For any global field K, the group $K^\times$ is a discrete cocompact subgroup of the group of 1-ideles $\mathbb{I}_K^1$.*

Proof. By Proposition 26.6, $K^\times$ is discrete in $\mathbb{I}_K$, and therefore discrete in the subspace $\mathbb{I}_K^1$.

As in the proof of Theorem 25.12, to prove that $K^\times$ is cocompact in $\mathbb{I}_K^1$ it suffices to exhibit a compact set $W \subseteq \mathbb{A}_K$ for which $W \cap \mathbb{I}_K^1$ surjects onto $\mathbb{I}_K^1/K^\times$ under the quotient map (here we are using Lemma 26.8: $\mathbb{I}_K^1$ is closed so $W \cap \mathbb{I}_K^1$ is compact).

To construct W we first choose $a \in \mathbb{A}_K$ such that $\|a\| > B_K$, where B_K is the Blichfeldt-Minkowski constant in Lemma 25.14, and let

$$W := L(a) = \{x \in \mathbb{A}_K : \|x\|_v \leq \|a\|_v \text{ for all } v \in M_K\}.$$

Now consider any $u \in \mathbb{I}_K^1$. We have $\|u\| = 1$, so $\|\frac{a}{u}\| = \|a\| > B_K$, and by Lemma 25.14 there is a $z \in K^\times$ for which $\|z\|_v \leq \left\|\frac{a}{u}\right\|_v$ for all $v \in M_K$. Therefore $zu \in W$. Thus every $u \in \mathbb{I}_K^1$ can be written as $u = z^{-1} \cdot zu$ with $z^{-1} \in K^\times$ and $zu \in W \cap \mathbb{I}_K^1$. Thus $W \cap \mathbb{I}_K^1$ surjects onto $\mathbb{I}_K^1/K^\times$ under the quotient map $\mathbb{I}_K^1 \to \mathbb{I}_K^1/K^\times$, which is continuous, and it follows that $\mathbb{I}_K^1/K^\times$ is compact. □

Definition 26.10. For a global field K the compact group $C_K^1 := \mathbb{I}_K^1/K^\times$ is the *norm-1 idele class group*.

Remark 26.11. When K is a function field the norm-1 idele class group C_K^1 is totally disconnected, in addition to being a compact group, and thus a *profinite group*.

26.2 Profinite groups

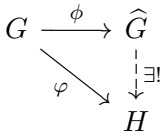
In order to state the main theorems of class field theory in our adelic/idelic setup, rather than considering each finite abelian extension L of a global field K individually, we prefer to work in K^{ab}, the compositum of all finite abelian extensions of K. This requires us to understand the infinite Galois group $\mathrm{Gal}(K^{\mathrm{ab}}/K)$, which is an example of a *profinite group*.

Definition 26.12. A *profinite group* is a topological group that is an inverse limit of finite groups with the discrete topology. Given any topological group G, we can construct a profinite group by taking the *profinite completion*

$$\widehat{G} := \varprojlim_N G/N \subseteq \prod_N G/N$$

where N ranges over finite index open normal subgroups, ordered by containment.[1] If we are given a group G without a specified topology, we can make it a topological group by giving it the *profinite topology*. This is the weakest topology that makes every finite quotient discrete and is obtained by taking all cosets of finite-index normal subgroups as a basis.

The profinite completion of G is (by construction) a profinite group, and it comes equipped with a natural homomorphism $\phi\colon G \to \widehat{G}$ that sends each $g \in G$ to the sequence of its images $(\overline{g}_N)$ in the discrete finite quotients G/N, which we may view as an element of $\prod_N G/N$. The homomorphism ϕ is not necessarily injective; this occurs if and only if the intersection of all finite-index open normal subgroups of G is the trivial group (such a G is said to be *residually finite*), but we always have the following universal property for inverse limits. For every continuous homomorphism $\varphi\colon G \to H$ with H a profinite group, there is a unique continuous homomorphism that makes the following diagram commute



There is much one can say about profinite groups but we shall limit ourselves to a few remarks and statements of the main results we need, deferring most of the proofs to Problem Set 11. See [4] for a comprehensive treatment of profinite groups.

Remark 26.13. Taking inverse limits in the category of topological groups is the same thing as taking the inverse limits in the categories of topological spaces and groups independently: the topology is the subspace topology in the product, and the group operation is the group operation in the product (defined component-wise). This might seem obvious, but the same statement does not apply to direct limits, where one must compute the limit in the category of topological groups, otherwise the group operation in the direct limit of the groups is not necessarily continuous under the direct limit topology; see [5].[2]

Remark 26.14. The profinite completion of G as a topological group is not necessarily the same thing as the profinite completion of G as a group if we forget its topology; this depends on whether the original topology on G contains the profinite topology or not. In particular, a profinite group need not equal to its profinite completion as a group; the group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ endowed with the Krull topology is an example (see below). Profinite groups that are isomorphic to their profinite completions as groups are said to be *strongly complete*; this is equivalent to requiring every finite index subgroup to be open (see Corollary 26.19 below). It is known that if G is finitely generated as a topological group (meaning it contains a finitely generated dense subgroup), then G is strongly complete [3]. This applies, for example, to $\mathrm{Gal}(\bar{\mathbb{F}}_q/\mathbb{F}_q)$ for any finite field $\mathbb{F}_q$, since the q-power Frobenius automorphism generates a dense subgroup (it is thus a *topological generator*).

Remark 26.15. For suitable restricted types of finite groups $\mathcal{C}$ (for example, all finite cyclic groups, or all finite p-groups for some fixed prime p), one can similarly define the notion of a pro-$\mathcal{C}$ group and the pro-$\mathcal{C}$ completion of a group by constraining the finite groups in the inverse system to lie in $\mathcal{C}$. One can also define profinite rings or pro-$\mathcal{C}$ rings.

[1] Recall that an inverse system has objects X_i and morphisms $X_i \leftarrow X_j$ for $i \le j$. Here we have objects G/N_i and morphisms $G/N_i \leftarrow G/N_j$ for $i \le j$; we want the indices ordered so that $i \le j$ whenever N_i contains N_j; containment induces a canonical morphism $g + N_i \leftarrow\!\shortmid g + N_j$ on the quotients.

[2] For countable direct systems of locally compact groups this issue does not arise [5, Thm. 2.7].

Example 26.16. Here are a few examples of profinite completions:

1. The profinite completion of any finite group G is isomorphic to G with the discrete topology; the natural map $G \to \widehat{G}$ is an isomorphism.
2. The profinite completion of $\mathbb{Z}$ is $\widehat{\mathbb{Z}} := \varprojlim_n \mathbb{Z}/n\mathbb{Z} = \prod \mathbb{Z}_p$, where the indices n are ordered by divisibility; the natural map $\mathbb{Z} \to \widehat{\mathbb{Z}}$ is injective but not surjective.
3. The profinite completion of $\mathbb{Q}$ is trivial because $\mathbb{Q}$ has no finite index subgroups other than itself. The natural map $\mathbb{Q} \to \widehat{\mathbb{Q}} = \{1\}$ is surjective but not injective.

Lemma 26.17. *Let G be a topological group with profinite completion $\widehat{G}$. The image of G under the natural map $\phi\colon G \to \widehat{G}$ is dense in $\widehat{G}$.*

Proof. See Problem Set 11. □

We now give a topological characterization of profinite groups that can serve as an alternative definition.

Theorem 26.18. *A topological group is profinite if and only if it is a totally disconnected compact group.*

Proof. See Problem Set 11. □

Corollary 26.19. *Let G be a profinite group. Then G is naturally isomorphic to its profinite completion. In fact,*

$$G \simeq \varprojlim G/U,$$

where U ranges over open normal subgroups (ordered by containment).

However, G is isomorphic to its profinite completion as a group (in other words, strongly complete) if and only if every finite index subgroup of G is open.

Proof. See Problem Set 11 for the first statement. For the second statement, if every finite index subgroup of G is open then every finite-index normal subgroup is open, meaning that the topology on G is finer than the profinite topology, and we get the same profinite completion under both topologies.

Conversely, if G has a finite index subgroup H that is not open, then no subgroup of H is open (since H is the union of the cosets of any of its subgroups); in particular, the intersection of all the conjugates of H, which is a normal subgroup N, is not open in G, nor are any of its subgroups. If the topological group G is isomorphic to its profinite completion $\widehat{G}$ as a group, then by the universal property of the profinite completion the natural map $\phi\colon G \to \widehat{G}$ is an isomorphism, but the image of N under ϕ is an open subgroup of $\widehat{G}$ by construction, which is a contradiction. □

26.3 Infinite Galois theory

The key issue that arises when studying Galois groups of infinite algebraic extensions (as opposed to finite ones) is that the Galois correspondence (the inclusion reversing bijection between subgroups and subextensions) fails spectacularly. As you proved on Problem Set 5 in the case $\mathrm{Gal}(\overline{\mathbb{F}}_q/\mathbb{F}_q) \simeq \widehat{\mathbb{Z}} \simeq \prod_p \mathbb{Z}_p$, this happens for a simple reason: there are too many subgroups. For a more extreme example, the absolute Galois group of $\mathbb{Q}$ has uncountably many subgroups of index 2 (all of which are necessarily normal) but $\mathbb{Q}$ has only countably many quadratic extensions, see [2, Aside 7.27].

Thus not all subgroups of an infinite Galois group $\mathrm{Gal}(L/K)$ correspond to subextensions of L/K. We are going to put a topology on $\mathrm{Gal}(L/K)$ that distinguishes those that do.

Lemma 26.20. *Let L/K be a Galois extension with Galois group $G = \mathrm{Gal}(L/K)$, If F/K is a normal subextension of L/K, then $H = \mathrm{Gal}(L/F)$ is a normal subgroup of G with fixed field F, and we have an exact sequence*

$$1 \to \mathrm{Gal}(L/F) \to \mathrm{Gal}(L/K) \to \mathrm{Gal}(F/K) \to 1,$$

where the first map is inclusion, the second map is induced by restriction, and we have

$$G/H \simeq \mathrm{Gal}(F/K).$$

This lemma is a list of things we already know to be true for finite Galois extensions, the point is simply to verify that they also hold for infinite Galois extensions; this seems prudent given the aforementioned failure of the Galois correspondence.

Proof. If F/K is a normal subextension of L/K then the restriction map $\sigma \mapsto \sigma_{|_F}$ defines a homomorphism $\mathrm{Gal}(L/K) \to \mathrm{Gal}(F/K)$ whose kernel is a normal subgroup $H = \mathrm{Gal}(L/F)$. The fixed field of H contains F by definition, and it must be equal to F: if we had $\alpha \in L^H - F$ we could construct an element of H that sends α to a distinct root $\alpha' \neq \alpha$ of its minimal polynomial f over F (this defines an element of $\mathrm{Gal}(E/F)$, where E is the splitting field of f, which can be extended to $\mathrm{Gal}(L/F) = H$ by embedding L in an algebraic closure and applying Theorem 4.9). The restriction map is surjective because any $\sigma \in \mathrm{Gal}(F/K)$ can be extended to $\mathrm{Gal}(L/K)$, by Theorem 4.9, thus the sequence in the lemma is exact, and $G/H \simeq \mathrm{Gal}(F/K)$ follows. □

Unlike the situation for finite Galois extensions, it can happen that a normal subgroup H of $\mathrm{Gal}(L/K)$ with fixed field F is **not** equal to $\mathrm{Gal}(L/F)$; it must be contained in $\mathrm{Gal}(L/F)$, but it could be a proper subgroup. This is exactly what happens for all but a countable number of the uncountably many index 2 subgroups H of $G = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$; the fixed field of H is $\mathbb{Q}$ but $H \subsetneq G$ is not the Galois group of $\overline{\mathbb{Q}}/\mathbb{Q}$, nor is the the Galois group of $\overline{\mathbb{Q}}/K$ for any subextension $K/\mathbb{Q}$. It is thus necessary to distinguish the subgroups of $\mathrm{Gal}(L/K)$ that are actually Galois groups of a subextension. This is achieved by putting an appropriate topology on the Galois group.

Definition 26.21. Let L/K be a Galois extension with Galois group $G := \mathrm{Gal}(L/K)$. The *Krull topology* on G has the basis consisting of all cosets of subgroups $H_F := \mathrm{Gal}(L/F)$, where F ranges over finite normal extensions of K in L.

Under the Krull topology every open normal subgroup necessarily has finite index, but it is typically **not** the case that every normal subgroup of finite index is open. Thus the Krull topology on $\mathrm{Gal}(L/K)$ is strictly coarser than the profinite topology, in general (this holds for $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, for example). However, the topological group we obtain by putting the Krull topology on $\mathrm{Gal}(L/K)$ is a profinite group.

Theorem 26.22. *Let L/K be a Galois extension. Under the Krull topology, the restriction maps induce a natural isomorphism of topological groups*

$$\phi\colon \mathrm{Gal}(L/K) \to \varprojlim \mathrm{Gal}(F/K),$$

where F ranges over finite Galois extensions of K in L. In particular, $\mathrm{Gal}(L/K)$ is a profinite group whose open normal subgroups are precisely those of the form $\mathrm{Gal}(L/F)$ for some finite normal extension F/K.

Proof. Every $\alpha \in L$ is algebraic over K, hence lies in some finite normal subextension F/K (take the normal closure of $K(\alpha)$). Every automorphism in $\mathrm{Gal}(L/K)$ is thus uniquely determined by its restrictions to finite normal F/K, which implies that ϕ is injective. Given an element $(\sigma_F) \in \varprojlim \mathrm{Gal}(F/K)$, we can define an automorphism $\sigma \in \mathrm{Gal}(L/K)$ by simply putting $\sigma(\alpha) = \sigma_F(\alpha)$, where F is the normal closure of $K(\alpha)$ (the fact that this actually gives an automorphism is guaranteed by the inverse system of restriction maps used to define $\varprojlim \mathrm{Gal}(F/K)$). Thus ϕ is surjective.

By Lemma 26.20, if we put $G := \mathrm{Gal}(L/K)$ and $H_F := \mathrm{Gal}(L/F)$, then we can view ϕ as the natural map

$$\phi\colon G \to \varprojlim G/H_F,$$

which is continuous, and we have shown it is a bijection. To prove that ϕ is an isomorphism of topological groups it remains only to show that it is an open map. For this it suffices to show that ϕ maps open subgroups $H \subseteq G$ to open sets in $\varprojlim G/H_F$, since every open set in G is a union of cosets of open subgroups. If $H = \mathrm{Gal}(L/F)$ then

$$\phi(H) = \{(\sigma_E) : \sigma_{E|_{E\cap F}} = \mathrm{id}_{|_{E\cap F}}\} = \pi_F^{-1}(\mathrm{id}_{|_F}),$$

where E/K ranges over finite normal subextensions of L/K and π_F is the projection map from the inverse limit to $\mathrm{Gal}(F/K)$. The singleton set $\{\mathrm{id}_{|_F}\}$ is open in the discrete group $\mathrm{Gal}(E/F)$, so its inverse image under the continuous projection π_F is open in G.

The last statement follows from Corollary 26.19 and Lemma 26.20. □

Theorem 26.23 (Fundamental theorem of Galois theory)**.** *Let L/K be a Galois extension and let $G := \mathrm{Gal}(L/K)$ be endowed with the Krull topology. The maps $F \mapsto \mathrm{Gal}(L/F)$ and $H \mapsto L^H$ define an inclusion reversing bijection between subextensions F/K of L/K and closed subgroups H of G. Under this correspondence, subextensions of finite degree n correspond to subgroups of finite index n, and normal subextensions F/K correspond to normal subgroups $H \subseteq G$ such that $\mathrm{Gal}(F/K) \simeq G/H$ as topological groups.*

Proof. We first note that every open subgroup of G is closed, since it is the complement of the union of its non-trivial cosets, all of which are open, and closed subgroups of finite index are open by the same argument.

The correspondence between finite Galois subextensions F/K and finite index closed normal subgroups H then follows the previous theorem, and we have $[F : K] = [G : H]$ because $G/H \simeq \mathrm{Gal}(F/K)$, by Lemma 26.20.

If F/K is any finite subextension with normal closure E, then $H = \mathrm{Gal}(L/F)$ contains the normal subgroup $N = \mathrm{Gal}(L/E)$ with finite index. The subgroup N is open and therefore closed, thus H is closed since it is a finite union of cosets of N. The fixed field of H is F (by the same argument as in the proof of Lemma 26.20), thus finite subextensions correspond to closed subgroups of finite index. Conversely, every closed subgroup H of finite index has a fixed field F of finite degree, since the intersection of its conjugates is a normal closed subgroup $N = \mathrm{Gal}(L/E)$ of finite index whose fixed field E contains F and has finite degree. The degrees and indices match because $[G : N] = [G : H][H : N]$ and $[E : K] = [F : K][E : F]$; by the previous argument for finite normal subextensions, $[E : K] = [G : N]$ and $[E : F] = [H : N]$ (for the second equality, replace L/K with L/F and G with H).

Any subextension F/K is the union of its finite subextensions E/K. The intersection of the corresponding closed finite index subgroups $\mathrm{Gal}(L/E)$ is equal to $\mathrm{Gal}(L/F)$, which is therefore closed. Conversely, every closed subgroup H of G is an intersection of basic

closed subgroups, all of which have the form $\mathrm{Gal}(L/E)$ for some finite subextension E/K, thus $H = \mathrm{Gal}(L/F)$, where F is the union of the E.

The isomorphism $\mathrm{Gal}(F/K) \simeq G/H$ for normal subextensions/subgroups follows directly from Lemma 26.20. □

Corollary 26.24. *Let L/K be a Galois extension and let H be a subgroup of $\mathrm{Gal}(L/K)$ with fixed field F. The closure $\overline{H}$ of H in the Krull topology is $\mathrm{Gal}(L/F)$.*

Proof. The Galois group $\mathrm{Gal}(L/F)$ contains H, since it contains every $\sigma \in \mathrm{Gal}(L/K)$ that fixes F (by definition), and $\mathrm{Gal}(L/F)$ is a closed subgroup of $\mathrm{Gal}(L/K)$ with $L^{\mathrm{Gal}(L/F)} = F$, by Theorem 26.23. We thus have $H \subseteq \overline{H} \subseteq \mathrm{Gal}(L/F)$ with the same fixed field F. The last two groups are closed and therefore equal under the bijection given by Theorem 26.23. □

We conclude this section with the following theorem due to Waterhouse [6].

Theorem 26.25 (Waterhouse 1973)**.** *Every profinite group G is isomorphic to the Galois group of some Galois extension L/K.*

Proof sketch. Let X be the disjoint union of the finite discrete quotients of G equipped with the G-action induced by multiplication. Now let k be any field and define $L = k(X)$ as a purely transcendental extension of k with indeterminates for each element of X. We can view each $\sigma \in G$ as an automorphism of L that fixes k and sends each $x \in X$ to $\sigma(x)$, and since G acts faithfully on X, we can view G as a subgroup of $\mathrm{Aut}_k(L)$. Now let $K = L^G$. Then L/K is a Galois extension with $G \simeq \mathrm{Gal}(L/K)$, by [6, Thm. 1]. □

Remark 26.26. Although this proof lets us choose any field k we like, we have no way to control K. In particular, it is not known whether every profinite group G is isomorphic to a Galois group over $K = \mathbb{Q}$; indeed, this is not even known for all finite groups G.

References

[1] Nicolas Bourbaki, *General Topology: Chapters 1-4*, Springer, 1995.

[2] J.S. Milne, *Fields and Galois theory*, version 4.51, 2015.

[3] Nikolay Nikolov and Dan Segal, *On finitely generated profinite groups I: strong completeness and uniform bounds*, Annals of Mathematics **165** (2007), 171–238.

[4] Luis Ribes and Pavel Zalesskii, *Profinite groups*, second edition, Springer, 2010.

[5] N. Tatsuuma, H. Shimomura, and T. Hirai, *On group topologies and unitary representations of inductive limits of topological groups and the case of the group of diffeomorphisms*, J. Math. Kyoto Univ. **38** (1998), 551–578.

[6] William C. Waterhouse, *Profinite groups are Galois groups*, Proceedings of the American Mathematical Society **42** (1974).

27 Local class field theory

In this lecture we give a brief overview of local class field theory. Recall that a local field is a locally compact field whose topology is induced by a nontrivial absolute value (Definition 9.1). As we proved in Theorem 9.9, every local field is isomorphic to one of the following:

- $\mathbb{R}$ or $\mathbb{C}$ (archimedean, characteristic 0);
- finite extension of $\mathbb{Q}_p$ (nonarchimedean, characteristic 0);
- finite extension of $\mathbb{F}_q((t))$ (nonarchimedean, characteristic $p > 0$).

In the nonarchimedean cases, the ring of integers of a local field is a complete DVR with finite residue field.

The goal of local class field theory is to classify all finite abelian extensions of a given local field K. Rather than considering each finite abelian extension L/K individually, we will treat them all at once, by working in the maximal abelian extension of K inside a fixed separable closure K^{sep}.

Definition 27.1. Let K be field with separable closure K^{sep}. The field

$$K^{\mathrm{ab}} := \bigcup_{\substack{L \subseteq K^{\mathrm{sep}} \\ L/K \text{ finite abelian}}} L$$

is the *maximal abelian extension of* K (in K^{sep}). We also define

$$K^{\mathrm{unr}} := \bigcup_{\substack{L \subseteq K^{\mathrm{sep}} \\ L/K \text{ finite unramified}}} L,$$

the *maximal unramified extension of* K (in K^{sep}).

The field K^{ab} contains the field K^{unr}; this is obvious in the archimedean case, where we have $K = K^{\mathrm{unr}}$ is $\mathbb{R}$ or $\mathbb{C}$ and $K^{\mathrm{ab}} = K^{\mathrm{sep}} = \mathbb{C}$ (note that the extension $\mathbb{C}/\mathbb{R}$ is ramified). In the nonarchimedean case the inclusion $K^{\mathrm{unr}} \subseteq K^{\mathrm{ab}}$ follows from Theorem 10.15, which implies that K^{unr} is isomorphic to the algebraic closure of the residue field of K, which is an abelian extension because it is pro-cyclic (every finite extension of the residue field is cyclic because the residue field is finite). We thus have a tower of field extensions

$$K \subseteq K^{\mathrm{unr}} \subseteq K^{\mathrm{ab}} \subseteq K^{\mathrm{sep}}.$$

By Theorem 26.22, the Galois group $\mathrm{Gal}(K^{\mathrm{ab}}/K)$ is the profinite group

$$\mathrm{Gal}(K^{\mathrm{ab}}/K) \simeq \varprojlim_{L} \mathrm{Gal}(L/K),$$

where L ranges over the finite extensions of K in K^{ab}, ordered by inclusion (note that every finite extension of K in K^{ab} is normal because every open subgroup of the abelian group $\mathrm{Gal}(K^{\mathrm{ab}}/K)$ is a normal subgroup).

Like all Galois groups, the profinite group $\mathrm{Gal}(K^{\mathrm{ab}}/K)$ is a totally disconnected compact group; see Problem Set 11. By Theorem 26.23, we have the Galois correspondence

$$\begin{aligned} \{\,\text{extensions of } K \text{ in } K^{\mathrm{ab}}\,\} &\longleftrightarrow \{\,\text{closed subgroups of } \mathrm{Gal}(K^{\mathrm{ab}}/K)\,\} \\ L &\longmapsto \mathrm{Gal}(K^{\mathrm{ab}}/L) \\ (K^{\mathrm{ab}})^H &\longleftarrow\!\shortmid H. \end{aligned}$$

Finite abelian extensions L/K correspond to open subgroups of $\mathrm{Gal}(K^{\mathrm{ab}}/K)$ (which must have finite index since $\mathrm{Gal}(K^{\mathrm{ab}}/K)$ is compact).

When K is an archimedean local field its abelian extensions are easy to understand; either $K = \mathbb{R}$, in which case $\mathbb{C}$ is the unique nontrivial abelian extension, or $K = \mathbb{C}$ and there are no nontrivial abelian extensions.

Now suppose K is a nonarchimedean local field with ring of integers $\mathcal{O}_K$, maximal ideal $\mathfrak{p}$, and residue field $\mathbb{F}_\mathfrak{p} := \mathcal{O}_K/\mathfrak{p}$. If L/K is a finite unramified extension with residue field $\mathbb{F}_\mathfrak{q} := \mathcal{O}_L/\mathfrak{q}$, Theorem 10.15 gives us a canonical isomorphism

$$\mathrm{Gal}(L/K) \simeq \mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p}) = \langle x \mapsto x^{\#\mathbb{F}_\mathfrak{p}} \rangle,$$

between the Galois group of L/K and the Galois group of the residue field extension $\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p}$. The group $\mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p})$ is generated by the Frobenius automorphism $x \to x^{\#\mathbb{F}_\mathfrak{p}}$, and we use $\mathrm{Frob}_{L/K} \in \mathrm{Gal}(L/K)$ to denote the corresponding element of $\mathrm{Gal}(L/K)$; note that $\mathrm{Frob}_{L/K}$ is an element, not just a conjugacy class, because $\mathrm{Gal}(L/K)$ is abelian. Every finite unramified extension of local fields L/K thus comes equipped with a canonical generator $\mathrm{Frob}_{L/K}$ for its Galois group (which is necessarily cyclic).

In this local unramified setting, the Artin map is very easy to understand. The ideal group $\mathcal{I}_K$ is the infinite cyclic group generated by the prime ideal $\mathfrak{p}$, and the Artin map

$$\begin{aligned} \psi_{L/K} \colon \mathcal{I}_K &\to \mathrm{Gal}(L/K) \\ \mathfrak{p} &\mapsto \mathrm{Frob}_{L/K}, \end{aligned}$$

corresponds to the quotient map $\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$, where $n := [L : K]$. We can extend the Artin map to $K^\times$ by defining $\psi_{L/K}(x) := \psi_{L/K}((x))$; this map sends every uniformizer π to the Frobenius element $\mathrm{Frob}_{L/K}$; note that since $\mathcal{O}_K$ is a DVR, hence a PID, every ideal in $\mathcal{I}$ is of the form (x) for some $x \in K^\times$, so defining the Artin map on $K^\times$ rather than $\mathcal{I}_K$ does not lose any information when K is a local field.

27.1 Local Artin reciprocity

Local class field theory is based on the existence of a continuous homomorphism

$$\theta_K \colon K^\times \to \mathrm{Gal}(K^{\mathrm{ab}}/K)$$

known as the *local Artin homomorphism* (or *local reciprocity map*), which is described by the following theorem.

Theorem 27.2 (Local Artin Reciprocity)**.** *Let K be a local field. There is a unique continuous homomorphism*

$$\theta_K \colon K^\times \to \mathrm{Gal}(K^{\mathrm{ab}}/K)$$

with the property that for each finite extension L/K in K^{ab}, the homomorphism

$$\theta_{L/K} \colon K^\times \to \mathrm{Gal}(L/K)$$

given by composing θ_K with the natural map $\mathrm{res}_{L/K} \colon \mathrm{Gal}(K^{\mathrm{ab}}/K) \twoheadrightarrow \mathrm{Gal}(L/K)$ satisfies:

- *if K is nonarchimedean and L/K is unramified then $\theta_{L/K}(\pi) = \mathrm{Frob}_{L/K}$ for every uniformizer π of $\mathcal{O}_K$;*
- *$\theta_{L/K}$ is surjective with kernel $\mathrm{N}_{L/K}(L^\times)$, inducing $K^\times/\mathrm{N}_{L/K}(L^\times) \simeq \mathrm{Gal}(L/K)$.*

The natural map $\mathrm{res}_{L/K}\colon \mathrm{Gal}(K^{\mathrm{ab}}/K) \twoheadrightarrow \mathrm{Gal}(L/K)$ can be viewed as any of

- the map induced by restriction $\sigma \mapsto \sigma_{|_L}$ (note that $\sigma(L) = L$ because L/K is Galois);
- the quotient map $\mathrm{Gal}(K^{\mathrm{ab}}/K) \twoheadrightarrow \mathrm{Gal}(K^{\mathrm{ab}}/K)/\mathrm{Gal}(K^{\mathrm{ab}}/L)$;
- the projection coming from $\mathrm{Gal}(K^{\mathrm{ab}}/K) = \varprojlim_L \mathrm{Gal}(L/K) \subseteq \prod_L \mathrm{Gal}(L/K)$ (where L ranges over finite extensions of K in K^{ab}).

These are equivalent descriptions of the same surjective homomorphism of topological groups (where the finite group $\mathrm{Gal}(L/K)$ has the discrete topology).

We will not have time to prove this theorem, but we would like to understand exactly what it says. The homomorphisms $\theta_{L/K}$ form a compatible system, in the sense that if $L_1 \subseteq L_2$ then $\theta_{L_1/K} = \mathrm{res}_{L_2/L_1} \circ \theta_{L_2/K}$, where res_{L_2/L_1} is the natural map from $\mathrm{Gal}(L_2/K)$ to $\mathrm{Gal}(L_1/K) = \mathrm{Gal}(L_2/K)/\mathrm{Gal}(L_2/L_1)$. Indeed, the maps res_{L_2/L_1} are precisely the maps that appear in the inverse system defining $\varprojlim_L \mathrm{Gal}(L/K) \simeq \mathrm{Gal}(K^{\mathrm{ab}}/K)$.

It is first worth contrasting local Artin reciprocity with the more complicated global version of Artin reciprocity that we saw in Lecture 21:

- There is no modulus $\mathfrak{m}$; working in K^{ab} addresses all abelian extensions of K at once.
- The ray class groups $\mathrm{Cl}_K^{\mathfrak{m}}$ are replaced by quotients of $K^\times$.
- The Takagi group $\mathrm{N}_{L/K}(\mathcal{I}_L^{\mathfrak{m}})\mathcal{R}_K^{\mathfrak{m}} \subseteq \mathcal{I}_K^{\mathfrak{m}}$ is replaced by $\mathrm{N}_{L/K}(L^\times) \subseteq K^\times$.

27.2 Norm groups

Definition 27.3. A *norm group* of a local field K is a subgroup of the form

$$\mathrm{N}(L^\times) := \mathrm{N}_{L/K}(L^\times) \subseteq K^\times,$$

for some finite abelian extension L/K.

Remark 27.4. Removing the word abelian does not change the definition above. If L/K is any finite extension (not necessarily Galois), then $\mathrm{N}(L^\times) = \mathrm{N}(F^\times)$, where F is the maximal abelian extension of K in L; this result is known as the Norm Limitation Theorem (see [1, Theorem III.3.5]). So we could have defined norm groups more generally. This is not relevant to classifying the abelian extension of K, but it demonstrates a key limitation of local class field theory (which extends to global class field theory): norm groups tell us nothing about nonabelian extensions of K.

Theorem 27.2 implies that the Galois group of any finite abelian extension L/K of a local fields is canonically isomorphic to the quotient $K^\times/\mathrm{N}_{L/K}(L^\times)$. In order to understand the finite abelian extensions of a local field K, we just need to understand its norm groups.

Corollary 27.5. *The map $L \mapsto \mathrm{N}(L^\times)$ defines an inclusion reversing bijection between the finite abelian extensions L/K in K^{ab} and the norm groups in $K^\times$ which satisfies*

$$(a)\ \mathrm{N}((L_1L_2)^\times) = \mathrm{N}(L_1^\times) \cap \mathrm{N}(L_2^\times) \qquad \text{and} \qquad (b)\ \mathrm{N}((L_1 \cap L_2)^\times) = \mathrm{N}(L_1^\times)\mathrm{N}(L_2^\times).$$

In particular, every norm group of K has finite index in $K^\times$, and every subgroup of $K^\times$ that contains a norm group is a norm group.

Here we write L_1L_2 for the compositum of L_1 and L_2 inside K^{ab} (the intersection of all subfields of K^{ab} that contain both L_1 and L_2).

Proof. We first note that if $L_1 \subseteq L_2$ are two extensions of K then transitivity of the field norm (Corollary 4.52) implies

$$\mathrm{N}_{L_2/K} = \mathrm{N}_{L_1/K} \circ \mathrm{N}_{L_2/L_1},$$

and therefore $\mathrm{N}(L_2^\times) \subseteq \mathrm{N}(L_1^\times)$; the map $L \mapsto \mathrm{N}(L^\times)$ thus reverses inclusions.

This immediately implies $\mathrm{N}((L_1L_2)^\times) \subseteq \mathrm{N}(L_1^\times) \cap \mathrm{N}(L_2^\times)$, since $L_1, L_2 \subseteq L_1L_2$. For the reverse inclusion, let us consider the commutative diagram

$$\begin{array}{ccc} K^\times & \xrightarrow{\theta_{L_1L_2/K}} & \mathrm{Gal}(L_1L_2/K) \\ & \searrow^{\theta_{L_1/K}\times\theta_{L_2/K}} & \downarrow{\scriptstyle \mathrm{res}\times\mathrm{res}} \\ & & \mathrm{Gal}(L_1/K) \times \mathrm{Gal}(L_2/K) \end{array}$$

By Theorem 27.2, each $x \in \mathrm{N}(L_1^\times) \cap \mathrm{N}(L_2^\times) \subseteq K^\times$ lies in the kernel of $\theta_{L_1/K}$ and $\theta_{L_2/K}$, hence in the kernel of $\theta_{L_1L_2/K}$ (by the diagram), and therefore in $\mathrm{N}((L_1L_2)^\times)$, again by Theorem 27.2. This proves (a).

We now show that $L \mapsto \mathrm{N}(L^\times)$ is a bijection; it is surjective by definition, so we just need to show it is injective. If $\mathrm{N}(L_2^\times) = \mathrm{N}(L_1^\times)$ then by (a) we have

$$\mathrm{N}((L_1L_2)^\times) = \mathrm{N}(L_1^\times) \cap \mathrm{N}(L_2^\times) = \mathrm{N}(L_1^\times) = \mathrm{N}(L_2^\times),$$

and Theorem 27.2 implies $\mathrm{Gal}(L_1L_2/K) \simeq \mathrm{Gal}(L_1/K) \simeq \mathrm{Gal}(L_2/K)$, which forces $L_1 = L_2$; thus $L \mapsto \mathrm{N}(L^\times)$ is injective.

We now prove (b). The field $L_1 \cap L_2$ is the largest extension of K that lies in both L_1 and L_2, while $\mathrm{N}(L_1^\times)\mathrm{N}(L_2^\times)$ is the smallest subgroup of $K^\times$ containing both $\mathrm{N}(L_1^\times)$ and $\mathrm{N}(L_2^\times)$; they therefore correspond under the inclusion reversing bijection $L \mapsto \mathrm{N}(L^\times)$ and we have $\mathrm{N}((L_1 \cap L_2)^\times) = \mathrm{N}(L_1^\times)\mathrm{N}(L_2^\times)$ as desired.

The fact that every norm group has finite index in $K^\times$ follows immediately from the isomorphism $\mathrm{Gal}(L/K) \simeq K^\times/N_{L/K}(L^\times)$ given by Theorem 27.2, since $\mathrm{Gal}(L/K)$ is finite.

Finally, let us prove that every subgroup of $K^\times$ that contains a norm group is a norm group. Suppose $\mathrm{N}(L^\times) \subseteq H \subseteq K^\times$, for some finite abelian L/K, and subgroup H of $K^\times$, and put $F \coloneqq L^{\theta_{L/K}(H)}$. We have a commutative diagram

$$\begin{array}{ccc} K^\times & \xrightarrow{\theta_{L/K}} & \mathrm{Gal}(L/K) \\ & \searrow^{\theta_{F/K}} & \downarrow{\scriptstyle \mathrm{res}} \\ & & \mathrm{Gal}(F/K) \end{array}$$

in which $\mathrm{Gal}(L/F) = \theta_{L/K}(H)$ is precisely the kernel of the map $\mathrm{Gal}(L/K) \to \mathrm{Gal}(F/K)$ induced by restriction. It follows from Theorem 27.2 that

$$H = \ker \theta_{F/K} = \mathrm{N}(F^\times)$$

is a norm group as claimed. □

Lemma 27.6. *Let L/K be any extension of local fields. If $\mathrm{N}(L^\times)$ has finite index in $K^\times$ then it is open.*

Proof. The lemma is clear if K is archimedean (either $L = K$ and $\mathrm{N}(L^\times) = K^\times$, or $L \simeq \mathbb{C}$, $K \simeq \mathbb{R}$, and $[K^\times : \mathrm{N}(L^\times)] = [\mathbb{R}^\times : \mathbb{R}_{>0}] = 2$), so assume K is nonarchimedean. Suppose $[K^\times : \mathrm{N}(L^\times)] < \infty$. The unit group $\mathcal{O}_L^\times$ is compact, so $\mathrm{N}(\mathcal{O}_L^\times)$ is compact (since $\mathrm{N}\colon L^\times \to K^\times$ is continuous), thus closed in the Hausdorff space $K^\times$. For any $\alpha \in L$,

$$\alpha \in \mathcal{O}_L^\times \iff |\alpha| = 1 \iff |\mathrm{N}_{L/K}(\alpha)| = 1 \iff \mathrm{N}_{L/K}(\alpha) \in \mathcal{O}_K^\times,$$

and therefore

$$\mathrm{N}(\mathcal{O}_L^\times) = \mathrm{N}(L^\times) \cap \mathcal{O}_K^\times.$$

It follows that $\mathrm{N}(\mathcal{O}_L^\times)$ is the kernel of the homomorphism $\mathcal{O}_K^\times \hookrightarrow K^\times \twoheadrightarrow K^\times/\mathrm{N}(L^\times)$ and therefore $[\mathcal{O}_K^\times : \mathrm{N}(\mathcal{O}_L^\times)] \le [K^\times : \mathrm{N}(L^\times)] < \infty$. Thus $\mathrm{N}(\mathcal{O}_L^\times)$ is a closed subgroup of finite index in $\mathcal{O}_K^\times$, hence open (its complement is a finite union of closed cosets, hence closed), and $\mathcal{O}_K^\times$ is open[1] in $K^\times$, so $\mathrm{N}(\mathcal{O}_L^\times)$ is open in $K^\times$, and therefore $\mathrm{N}(L^\times)$ is open in $K^\times$, since $\mathrm{N}(L^\times)$ is a union of cosets of the open subgroup $\mathrm{N}(\mathcal{O}_L^\times)$. □

Remark 27.7. If K is a local field of characteristic zero then one can show that in fact every finite index subgroup of $K^\times$ is open (whether it is a norm group or not), but this is not true in positive characteristic.

27.3 The main theorems of local class field theory

Corollary 27.5 implies that all norm groups of K have finite index in $K^\times$, and Lemma 27.6 then implies that all norm groups are finite index open subgroups of $K^\times$. The existence theorem of local class field theory states that the converse also holds.

Theorem 27.8 (Local Existence Theorem)**.** *Let K be a local field and let H be a finite index open subgroup of $K^\times$. There is a unique extension L/K in K^{ab} with $\mathrm{N}_{L/K}(L^\times) = H$.*

The local Artin homomorphism $\theta_K\colon K^\times \to \mathrm{Gal}(K^{\mathrm{ab}}/K)$ is not an isomorphism; indeed, it cannot be, because $\mathrm{Gal}(K^{\mathrm{ab}}/K)$ is compact and $K^\times$ is not. However, the local existence theorem implies that after taking profinite completions the local Artin homomorphism becomes an isomorphism.

Theorem 27.9 (Main Theorem of Local Class Field Theory)**.** *Let K be a local field. The local Artin homomorphism induces a canonical isomorphism*

$$\widehat{\theta}_K\colon \widehat{K^\times} \xrightarrow{\sim} \mathrm{Gal}(K^{\mathrm{ab}}/K)$$

of profinite groups.

Proof. The Galois group $\mathrm{Gal}(K^{\mathrm{ab}}/K)$ is a profinite group, isomorphic to the inverse limit

$$\mathrm{Gal}(K^{\mathrm{ab}}/K) \simeq \varprojlim_L \mathrm{Gal}(L/K), \tag{1}$$

where L ranges over the finite extensions of K in K^{ab} ordered by inclusion; see Theorem 26.22. It follows from Lemma 27.6, Theorem 27.8, and the definition of the profinite completion, that

$$\widehat{K^\times} \simeq \varprojlim_L K^\times/\mathrm{N}(L^\times), \tag{2}$$

[1]Recall that in a nonarchimedean local field, $|K^\times|$ is discrete in $\mathbb{R}_{>0}$ and we can always pick $\epsilon > 0$ so that $\mathcal{O}_K^\times = \{x \in K^\times : 1 - \epsilon < |x| < 1 + \epsilon\}$, which is clearly open in the metric topology induced by $|\ |$.

where L ranges over finite abelian extensions of K (in K^{sep}). By local Artin reciprocity (Theorem 27.2), for each finite abelian extension L/K we have an isomorphism

$$\theta_{L/K}\colon K^\times/\mathrm{N}(L^\times) \xrightarrow{\sim} \mathrm{Gal}(L/K),$$

and these isomorphisms commute with inclusion maps between finite abelian extensions of K. We thus have an isomorphism of the inverse systems appearing in (1) and (2). The isomorphism is canonical because the Artin homomorphism θ_K is unique and the isomorphisms in (1) and (2) are both canonical. □

In view of Theorem 27.9, we would like to better understand the profinite group $\widehat{K^\times}$. If K is archimedean then $\widehat{K^\times}$ is either trivial or the cyclic group of order 2, so let us assume that K is nonarchimedean. If we pick a uniformizer π for the maximal ideal $\mathfrak{p}$ of $\mathcal{O}_K$, then we can uniquely write each $x \in K^\times$ in the form $u\pi^{v(x)}$, with $u \in \mathcal{O}_K^\times$ and $v(x) \in \mathbb{Z}$. This defines an isomorphism

$$\begin{aligned} K^\times &\xrightarrow{\sim} \mathcal{O}_K^\times \times \mathbb{Z} \\ x &\longmapsto (x/\pi^{v(x)}, v(x)). \end{aligned}$$

Taking profinite completions (which commutes with products), we obtain an isomorphism

$$\widehat{K^\times} \simeq \mathcal{O}_K^\times \times \widehat{\mathbb{Z}},$$

since the unit group

$$\mathcal{O}_K^\times \simeq \mathbb{F}_\mathfrak{p}^\times \times (1+\mathfrak{p}) \simeq \mathbb{F}_\mathfrak{p}^\times \times \varprojlim_n \mathcal{O}_K/(1+\mathfrak{p}^n)$$

is already profinite (hence isomorphic to its profinite completion, by Corollary 26.19). Note that the isomorphism $\widehat{K^\times} \simeq \mathcal{O}_K^\times \times \widehat{\mathbb{Z}}$ is far from canonical; it depends on our choice of π, and there are uncountably many π to choose from.

We have a commutative diagram of exact sequences of topological groups

$$\begin{array}{ccccccccc} 1 & \longrightarrow & \mathcal{O}_K^\times & \longrightarrow & K^\times & \xrightarrow{v} & \mathbb{Z} & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \wr} & & \downarrow{\scriptstyle \theta_K} & & \downarrow{\scriptstyle \phi} & & \\ 1 & \longrightarrow & \mathrm{Gal}(K^{\mathrm{ab}}/K^{\mathrm{unr}}) & \longrightarrow & \mathrm{Gal}(K^{\mathrm{ab}}/K) & \xrightarrow{\mathrm{res}} & \mathrm{Gal}(K^{\mathrm{unr}}/K) & \longrightarrow & 1 \end{array}$$

in which the bottom row is the profinite completion of the top row. The map ϕ on the right is given by

$$\mathbb{Z} \hookrightarrow \widehat{\mathbb{Z}} \simeq \mathrm{Gal}(\overline{\mathbb{F}}_\mathfrak{p}/\mathbb{F}_\mathfrak{p}) \simeq \mathrm{Gal}(K^{\mathrm{unr}}/K),$$

and sends 1 to the sequence of Frobenius elements $(\mathrm{Frob}_{L/K})$ in the profinite group

$$\mathrm{Gal}(K^{\mathrm{unr}}/K) \simeq \varprojlim_L \mathrm{Gal}(L/K) \subseteq \prod_L \mathrm{Gal}(L/K),$$

where L ranges over finite unramified extensions of K; here we are using the canonical isomorphisms $\mathrm{Gal}(L/K) \simeq \mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p})$ given by Theorem 10.15. The Frobenius element $\phi(1)$ is a *topological generator* for $\mathrm{Gal}(K^{\mathrm{unr}}/K)$, meaning that it generates a dense subset.

Remark 27.10. The Frobenius element $\phi(1) \in \mathrm{Gal}(K^{\mathrm{unr}}/K)$ corresponds to the Frobenius automorphism $x \mapsto x^{\#\mathbb{F}_\mathfrak{p}}$ of $\mathrm{Gal}(\overline{\mathbb{F}}_\mathfrak{p}/\mathbb{F}_\mathfrak{p})$; both are canonical topological generators of the Galois groups in which they reside, and both are sometimes referred to as the *arithmetic Frobenius*. There is another obvious generator for $\mathrm{Gal}(K^{\mathrm{unr}}/K) \simeq \mathrm{Gal}(\overline{\mathbb{F}}_\mathfrak{p}/\mathbb{F}_\mathfrak{p})$, namely $\phi(-1)$, which is called the *geometric Frobenius* (for reasons we won't explain here).

The group $\mathrm{Gal}(K^{\mathrm{ab}}/K^{\mathrm{unr}}) \simeq \mathcal{O}_K^\times$ corresponds to the inertia subgroup of $\mathrm{Gal}(K^{\mathrm{ab}}/K)$. The top sequence splits (but not canonically), hence so does the bottom, and we have

$$\mathrm{Gal}(K^{\mathrm{ab}}/K) \simeq \mathrm{Gal}(K^{\mathrm{ab}}/K^{\mathrm{unr}}) \times \mathrm{Gal}(K^{\mathrm{unr}}/K) \simeq \mathcal{O}_K^\times \times \widehat{\mathbb{Z}}.$$

For each choice of a uniformizer $\pi \in \mathcal{O}_K$ we get a decomposition $K^{\mathrm{ab}} = K_\pi K^{\mathrm{unr}}$ corresponding to $K^\times = \mathcal{O}_K^\times \pi^{\mathbb{Z}}$. The field K_π is the subfield of K^{ab} fixed by $\theta_K(\pi) \in \mathrm{Gal}(K^{\mathrm{ab}}/K)$. Equivalently, K_π is the compositum of all the totally ramified finite extensions L/K in K^{ab} for which $\pi \in \mathrm{N}(L^\times)$.

Example 27.11. Let $K = \mathbb{Q}_p$ and pick $\pi = p$. The decomposition $K^{\mathrm{ab}} = K_\pi K^{\mathrm{unr}}$ is

$$\mathbb{Q}_p^{\mathrm{ab}} = \bigcup_n \mathbb{Q}_p(\zeta_{p^n}) \cdot \bigcup_{m \perp p} \mathbb{Q}_p(\zeta_m),$$

where the first union on the RHS is fixed by $\theta_K(p)$ and the second is fixed by $\theta_K(\mathcal{O}_K^\times)$.

Constructing the local Artin homomorphism is the difficult part of local class field theory. However, assuming the local existence theorem, it is easy to show that the local Artin homomorphism is unique if it exists.

Proposition 27.12. *Let K be a local field and assume every finite index open subgroup of $K^\times$ is a norm group. There is at most one homomorphism $\theta\colon K^\times \to \mathrm{Gal}(K^{\mathrm{ab}}/K)$ of topological groups that has the properties given in Theorem 27.2.*

Proof. The proposition is clear when K is archimedean, so assume it is nonarchimedean. Let $\mathfrak{p} = (\pi)$ be the maximal ideal of $\mathcal{O}_K$, and for each integer $n \geq 0$ let $K_{\pi,n}/K$ be the finite abelian extension given by Theorem 27.8 corresponding to the finite index subgroup $(1+\mathfrak{p}^n)\langle\pi\rangle$ of $K^\times$; here $1+\mathfrak{p}^n$ and $\langle\pi\rangle$ denote subgroups of $K^\times$, with $1+\mathfrak{p}^0 := \mathcal{O}_K^\times$, and we note that $K^\times \simeq \mathcal{O}_K^\times\langle\pi\rangle$.

Suppose $\theta\colon K^\times \to \mathrm{Gal}(K^{\mathrm{ab}}/K)$ is a continuous homomorphism as in Theorem 27.2. Then $\theta(\pi)$ fixes $K_\pi := \bigcup_n K_{\pi,n}$, since $\pi \in \mathrm{N}(K_{\pi,n}) = \ker \theta_{K_{\pi,n}/K}$. We also know that $\theta_{L/K}(\pi) = \mathrm{Frob}_{L/K}$ for all finite unramified extensions L/K, which uniquely determines the action of $\theta(\pi)$ on K^{unr}, and hence on $K^{\mathrm{ab}} = K_\pi K^{\mathrm{unr}}$.

Now suppose $\theta'\colon K^\times \to \mathrm{Gal}(K^{\mathrm{ab}}/K)$ is another continuous homomorphism as in Theorem 27.2. By the argument above we must have $\theta'(\pi) = \theta(\pi)$ for every uniformizer π of $\mathcal{O}_K$, and $K^\times$ is generated by its subset of uniformizers: if we fix one uniformizer π, every $x \in K^\times$ can be written as $u\pi^n = (u\pi)\pi^{n-1}$ for some $u \in \mathcal{O}_K^\times$ and $n \in \mathbb{Z}$, and $u\pi$ is another uniformizer). It follows that $\theta(x) = \theta'(x)$ for all $x \in K^\times$ and therefore $\theta = \theta'$ is unique. □

Remark 27.13. One approach to proving local class field theory uses the theory of formal groups due to Lubin and Tate to explicitly construct the fields $K_\pi = \bigcup_n K_{\pi,n}$ used in the proof of Proposition 27.12, along with a continuous homomorphism $\theta_\pi\colon \mathcal{O}_K^\times \to \mathrm{Gal}(K_\pi/K)$ that extends uniquely to a continuous homomorphism $\theta\colon K^\times \to \mathrm{Gal}(K_\pi K^{\mathrm{unr}}/K)$. One then shows that $K^{\mathrm{ab}} = K_\pi K^{\mathrm{unr}}$ (using the Hasse-Arf Theorem), and that θ does not depend on the choice of π; see [1, §I.2-4] for details.

27.4 Finite abelian extensions

Local class field theory gives us canonical bijections between the following sets:

(1) finite-index open subgroups of $K^\times$ (which are necessarily normal);

(2) open subgroups of $\mathrm{Gal}(K^{\mathrm{ab}}/K)$ (which are necessarily normal and of finite index);

(3) finite extensions of K in K^{ab} (which are necessarily normal).

The bijection from (1) to (2) is induced by the isomorphism $\widehat{K^\times} \simeq \mathrm{Gal}(K^{\mathrm{ab}}/K)$ given by Theorem 27.9 and is inclusion preserving. The bijection from (2) to (3) follows from Galois theory (for infinite extensions), and is inclusion reversing, while the bijection from (3) to (1) is via the map $L \mapsto \mathrm{N}(L^\times)$, which is also inclusion reversing.

References

[1] J.S. Milne, *Class field theory*, version 4.02, 2013.

[2] Jean-Pierre Serre, *Local fields*, Springer, 1979.

28 Global class field theory, the Chebotarev density theorem

Recall that a global field is a field with a product formula whose completions at nontrivial absolute values are local fields. By the Artin-Whaples theorem (see Problem Set 7), every such field is either

- a *number field*: finite extension of $\mathbb{Q}$ (characteristic zero);
- a *global function field*: finite extension of $\mathbb{F}_q(t)$ (positive characteristic).

In Lecture 25 we defined the *adele ring* $\mathbb{A}_K$ of a global field K as the restricted product

$$\mathbb{A}_K := \prod\nolimits_v (K_v, \mathcal{O}_v) = \big\{(a_v) \in \prod K_v : a_v \in \mathcal{O}_v \text{ for almost all } v\big\},$$

where v ranges over the places of K (equivalence classes of absolute values), K_v denotes the completion of K at v, and $\mathcal{O}_v$ is the valuation ring of K_v if v is nonarchimedean, and equal to K_v otherwise. As a topological ring, $\mathbb{A}_K$ is locally compact and Hausdorff. The field K is canonically embedded in $\mathbb{A}_K$ via the diagonal map $x \mapsto (x, x, x, \dots)$ whose image is discrete, closed, and cocompact; see Theorem 25.12.

In Lecture 26 we defined the *idele group*

$$\mathbb{I}_K := \prod (K_v^\times, \mathcal{O}_v^\times) = \big\{(a_v) \in \prod K_v^\times : a_v \in \mathcal{O}_v^\times \text{ for almost all } v\big\},$$

which coincides with the unit group of $\mathbb{A}_K$ but has a finer topology (using the restricted product topology ensures that $a \mapsto a^{-1}$ is continuous, which is not true of the subspace topology). As a topological group, $\mathbb{I}_K$ is locally compact and Hausdorff. The multiplicative group $K^\times$ is canonically embedded as a discrete subgroup of $\mathbb{I}_K$ via the diagonal map $x \mapsto (x, x, x, \dots)$, and the *idele class group* is the quotient $C_K := \mathbb{I}_K / K^\times$, which is locally compact but not compact.

28.1 The idele norm

The idele group $\mathbb{I}_K$ surjects onto the ideal group $\mathcal{I}_K$ of invertible fractional ideals of $\mathcal{O}_K$ via the surjective homomorphism

$$\begin{aligned} \varphi \colon \mathbb{I}_K &\to \mathcal{I}_K \\ a &\mapsto \prod \mathfrak{p}^{v_\mathfrak{p}(a)}, \end{aligned}$$

where $v_\mathfrak{p}(a)$ is the $\mathfrak{p}$-adic valuation of the component $a_v \in K_v^\times$ of $a = (a_v) \in \mathbb{I}_K$ at the finite place v corresponding to the absolute value $\| \ \|_\mathfrak{p}$. We have the following commutative diagram of exact sequences:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & K^\times & \longrightarrow & \mathbb{I}_K & \longrightarrow & C_K & \longrightarrow & 1 \\ & & \big\downarrow{\scriptstyle x \mapsto (x)} & & \big\downarrow{\scriptstyle \varphi} & & \big\downarrow & & \\ 1 & \longrightarrow & \mathcal{P}_K & \longrightarrow & \mathcal{I}_K & \longrightarrow & \mathrm{Cl}_K & \longrightarrow & 1 \end{array}$$

where $\mathcal{P}_K$ is the subgroup of principal ideals and $\mathrm{Cl}_K := \mathcal{I}_K / \mathcal{P}_K$ is the ideal class group.

Definition 28.1. Let L/K is a finite separable extension of global fields. The *idele norm* $\mathrm{N}_{L/K}\colon \mathbb{I}_L \to \mathbb{I}_K$ is defined by $\mathrm{N}_{L/K}(b_w) = (a_v)$, where each

$$a_v := \prod_{w|v} \mathrm{N}_{L_w/K_v}(b_w)$$

is a product over places w of L that extend the place v of K and $\mathrm{N}_{L_w/K_v}\colon L_w \to K_v$ is the field norm of the corresponding finite separable extension of local fields L_w/K_v.

It follows from Corollary 11.24 and Remark 11.25 that the idele norm $\mathrm{N}_{L/K}\colon \mathbb{I}_L \to \mathbb{I}_K$ agrees with the field norm $\mathrm{N}_{L/K}\colon L^\times \to K^\times$ on the subgroup of principal ideles $L^\times \subseteq \mathbb{I}_L$. The field norm is also compatible with the ideal norm $\mathrm{N}_{L/K}\colon \mathcal{I}_L \to \mathcal{I}_K$ (see Proposition 6.6), and we have the following commutative diagram:

$$\begin{array}{ccccc} L^\times & \longrightarrow & \mathbb{I}_L & \longrightarrow & \mathcal{I}_L \\ \downarrow{\scriptstyle \mathrm{N}_{L/K}} & & \downarrow{\scriptstyle \mathrm{N}_{L/K}} & & \downarrow{\scriptstyle \mathrm{N}_{L/K}} \\ K^\times & \longrightarrow & \mathbb{I}_K & \longrightarrow & \mathcal{I}_K \end{array}$$

The image of $L^\times$ in $\mathbb{I}_L$ under the composition of the maps on the top row is precisely the group $\mathcal{P}_L$ of principal ideals, and the image of $K^\times$ in $\mathbb{I}_K$ is similarly $\mathcal{P}_K$. Taking quotients yields induced norm maps on the idele and ideal class groups, both of which we also denote $\mathrm{N}_{L/K}$, and we have a commutative square

$$\begin{array}{ccc} C_L & \twoheadrightarrow & \mathrm{Cl}_L \\ \downarrow{\scriptstyle \mathrm{N}_{L/K}} & & \downarrow{\scriptstyle \mathrm{N}_{L/K}} \\ C_K & \twoheadrightarrow & \mathrm{Cl}_K \end{array}$$

28.2 The Artin homomorphism

We now construct the global Artin homomorphism using the local Artin homomorphisms we defined in the previous lecture. Let us first fix once and for all a separable closure K^{sep} of our global field K, and for each place v of K, a separable closure K_v^{sep} of the local field K_v. Let K^{ab} and K_v^{ab} denote maximal abelian extensions within these separable closures; henceforth all abelian extensions of K and the K_v are assumed to lie in these maximal abelian extensions.

By Theorem 27.2, each local field K_v is equipped with a local Artin homomorphism

$$\theta_{K_v}\colon K_v^\times \to \mathrm{Gal}(K_v^{\mathrm{ab}}/K_v).$$

For each finite abelian extension L/K and each place $w|v$ of L, composing θ_{K_v} with the natural map $\mathrm{Gal}(K_v^{\mathrm{ab}}/K_v) \to \mathrm{Gal}(L_w/K_v)$ yields a surjective homomorphism

$$\theta_{L_w/K_v}\colon K_v^\times \to \mathrm{Gal}(L_w/K_v)$$

with kernel $\mathrm{N}_{L_w/K_v}(L_w^\times)$. When K_v is nonarchimedean and L_w/K_v is unramified we have $\theta_{L_w/K_v}(\pi_v) = \mathrm{Frob}_{L_w/K_v}$ for all uniformizers π_v of K_v. Note that by Theorem 11.20, every finite separable extension of K_v is of the form L_w for some place $w|v$.

We now define an embedding of Galois groups

$$\begin{aligned}\varphi_w\colon \mathrm{Gal}(L_w/K_v) &\hookrightarrow \mathrm{Gal}(L/K)\\ \sigma &\mapsto \sigma_{|_L}\end{aligned}$$

The map φ_w is well defined and injective because every element of L_w can be written as ℓx for some $\ell \in L$ and $x \in K_v$ (any K-basis for L spans L_w as a K_v vector space), so each $\sigma \in \mathrm{Gal}(L_w/K_v)$ is uniquely determined by its action on L, which fixes $K \subseteq K_v$. If v is archimedean then $\varphi_w(\mathrm{Gal}(L_w/K_v))$ is either trivial or generated by the involution corresponding to complex conjugation in $L_w \simeq \mathbb{C}$. If v is a finite place and $\mathfrak{q}$ is the prime of L corresponding to $w|v$, then $\varphi_w(\mathrm{Gal}(L_w/K_v))$ is the decomposition group $D_{\mathfrak{q}} \subseteq \mathrm{Gal}(L/K)$; this follows from parts (5) and (6) of Theorem 11.23.

More generally, for any place v of K, the Galois group $\mathrm{Gal}(L/K)$ acts on the set $\{w|v\}$, via $|\alpha|_{\sigma(w)} \coloneqq |\sigma(\alpha)|_w$, and $\varphi_w(\mathrm{Gal}(L_w/K_v))$ is the stabilizer of w under this action. It thus makes sense to call $\varphi_w(\mathrm{Gal}(L_w/K_v))$ the *decomposition group* of the place w. For $w|v$ the groups $\varphi_w(\mathrm{Gal}(L_w/K_v))$ are necessarily conjugate, and in our abelian setting, equal.

Moreover, the composition $\varphi_w \circ \theta_{L_w/K_v}$ defines a map $K_v^\times \to \mathrm{Gal}(L/K)$ that is independent of the choice of $w|v$: this is easy to see when v is an unramified nonarchimedean place, since then $\varphi_w(\theta_{L_w/K_v}(\pi_v)) = \mathrm{Frob}_v$ for every uniformizer π_v of K_v, and this determines $\varphi_w \circ \theta_{L_w/K_v}$ since the π_v generate $K_v^\times$.

For each place v of K we now embed $K_v^\times$ into the idele group $\mathbb{I}_K$ via the map

$$\begin{aligned}\iota_v\colon K_v^\times &\hookrightarrow \mathbb{I}_K\\ \alpha &\mapsto (1,1,\dots,1,\alpha,1,1,\dots),\end{aligned}$$

whose image intersects $K^\times \subseteq \mathbb{I}_K$ trivially. This embedding is compatible with the idele norm in the following sense: if L/K is any finite separable extension and w is a place of L that extends the place v of K then the diagram

$$\begin{array}{ccc} L_w^\times & \xrightarrow{\mathrm{N}_{L_w/K_v}} & K_v^\times \\ \downarrow{\scriptstyle \iota_w} & & \downarrow{\scriptstyle \iota_v} \\ \mathbb{I}_L & \xrightarrow{\mathrm{N}_{L/K}} & \mathbb{I}_K \end{array}$$

commutes.

Now let L/K be a finite abelian extension. For each place v of K, let us pick a place w of L extending v and define

$$\begin{aligned}\theta_{L/K}\colon \mathbb{I}_K &\to \mathrm{Gal}(L/K)\\ (a_v) &\mapsto \prod_v \varphi_w(\theta_{L_w/K_v}(a_v)),\end{aligned}$$

where the product takes place in $\mathrm{Gal}(L/K)$. The value of $\varphi_w(\theta_{L_w/K_v}(a_v))$ is independent of our choice of $w|v$, as noted above. The product is well defined because $a_v \in \mathcal{O}_v^\times$ and v is unramified in L for almost all v, in which case

$$\varphi_w(\theta_{L_w/K_v}(a_v)) = \mathrm{Frob}_v^{v(a_v)} = 1,$$

It is clear that $\theta_{L/K}$ is a homomorphism, since each $\varphi_w \circ \theta_{L_w/K_v}$ is, and $\theta_{L/K}$ is continuous because its kernel is a union of open sets: each $a \coloneqq (a_v) \in \ker \theta_{L/K}$ lies in an open set

$U_a := U_S \times \prod_{v \notin S} \mathcal{O}_v^\times \subseteq \ker \theta_{L/K}$, where S contains all ramified v and all v for which $a_v \notin \mathcal{O}_v^\times$, and U_S is the kernel of $(a_v)_{v\in S} \mapsto \prod_{v\in S} \varphi_w(\theta_{L_w/K_v}))$, an open subsset of $\prod_{v\in S} K_v^\times$).

If $L_1 \subseteq L_2$ are two finite abelian extensions of K, then $\theta_{L_1/K}(a) = \theta_{L_2/K}(a)_{|_{L_1}}$ for all $a \in \mathbb{I}_K$. The $\theta_{L/K}$ form a compatible system of homomorphisms from $\mathbb{I}_K$ to the inverse limit $\varprojlim_L \mathrm{Gal}(L/K) \simeq \mathrm{Gal}(K^{\mathrm{ab}}/K)$, where L ranges over finite abelian extensions of K in K^{ab} ordered by inclusion. By the universal property of the profinite completion, they uniquely determine a continuous homomorphism.

Definition 28.2. Let K be a global field. The *global Artin homomorphism* is the continuous homomorphism

$$\theta_K \colon \mathbb{I}_K \to \varprojlim_L \mathrm{Gal}(L/K) \simeq \mathrm{Gal}(K^{\mathrm{ab}}/K)$$

defined by the compatible system of homomorphisms $\theta_{L/K} \colon \mathbb{I}_K \to \mathrm{Gal}(L/K)$, where L ranges over finite abelian extensions of K in K^{ab}.

The isomorphism $\mathrm{Gal}(K^{\mathrm{ab}}/K) \simeq \varprojlim \mathrm{Gal}(L/K)$ is the natural isomorphism between a Galois group and its profinite completion with respect to the Krull topology (Theorem 26.22) and is thus canonical, as is the global Artin homomorphism $\theta_K \colon \mathbb{I}_K \to \mathrm{Gal}(K^{\mathrm{ab}}/K)$.

Proposition 28.3. *Let K be global field. The global Artin homomorphism θ_K is the unique continuous homomorphism $\mathbb{I}_K \to \mathrm{Gal}(K^{\mathrm{ab}}/K)$ with the property that for every finite abelian extension L/K in K^{ab} and every place w of L lying over a place v of K the diagram*

$$\begin{array}{ccc} K_v^\times & \xrightarrow{\theta_{L_w/K_v}} & \mathrm{Gal}(L_w/K_v) \\ \downarrow{\scriptstyle \iota_v} & & \downarrow{\scriptstyle \varphi_w} \\ \mathbb{I}_K & \xrightarrow{\theta_{L/K}} & \mathrm{Gal}(L/K) \end{array}$$

commutes, where the homomorphism $\theta_{L/K}$ is defined by $\theta_{L/K}(a) := \theta_K(a)_{|_L}$.

Proof. That θ_K has this property follows from its construction. Now suppose that there is another continuous homomorphism $\theta'_K \colon \mathbb{I}_K \to \mathrm{Gal}(K^{\mathrm{ab}}/K)$ with the same property. We may view elements of $\mathrm{Gal}(K^{\mathrm{ab}}/K) \simeq \varprojlim \mathrm{Gal}(L/K)$ as elements of $\prod_{L/K} \mathrm{Gal}(L/K)$, where L varies over finite abelian extensions of K in K^{ab}. If θ_K and θ'_K are not identical, then there must be an $a \in \mathbb{I}_K$ and a finite abelian extension L/K for which $\theta_{L/K}(a) \neq \theta'_{L/K}(a)$.

Let S be a finite set of places of K that includes all places v for which $a_v \notin \mathcal{O}_v^\times$ and all ramified places of L/K. Define $b \in \mathbb{I}_K$ by $b_v := 1$ for $v \in S$ and $b_v := a_v$ for $v \notin S$, so that $a = b \prod_{v\in S} \iota_v(a_v)$. Then $\theta_{L_w/K_v}(b_v) = 1$ for all places v, so we must have $\theta_{L/K}(b) = 1 = \theta'_{L/K}(b)$, and for $v \in S$ we have

$$\theta_{L/K}(\iota_v(a_v)) = \varphi_w(\theta_{L_w/K_v}(a_v)) = \theta'_{L/K}(\iota_v(a_v)),$$

by the commutativity of the diagram in the proposition. But then

$$\theta_{L/K}(a) = \theta_{L/K}(b) \prod_{v\in S} \theta_{L/K}(\iota_v(a_v)) = \theta'_{L/K}(b) \prod_{v\in S} \theta'_{L/K}(\iota_v(a_v)) = \theta'_{L/K}(a),$$

which is a contradiction. So $\theta'_K = \theta_K$ as claimed. □

28.3 The main theorems of global class field theory

In the global version of Artin reciprocity, the idele class group $C_K := \mathbb{I}_K/K^\times$ plays the role that the multiplicative group $K_v^\times$ plays in local Artin reciprocity (Theorem 27.2).

Theorem 28.4 (Global Artin Reciprocity)**.** *Let K be a global field. The kernel of the global Artin homomorphism θ_K contains $K^\times$, and we thus have a continuous homomorphism*

$$\theta_K \colon C_K \to \mathrm{Gal}(K^{\mathrm{ab}}/K),$$

with the property that for every finite abelian extension L/K in K^{ab} the homomorphism

$$\theta_{L/K} \colon C_K \to \mathrm{Gal}(L/K)$$

obtained by composing θ_K with the natural map $\mathrm{Gal}(K^{\mathrm{ab}}/K) \twoheadrightarrow \mathrm{Gal}(L/K)$ is surjective with kernel $\mathrm{N}_{L/K}(C_L)$, inducing an isomorphism $C_K/\mathrm{N}_{L/K}(C_L) \simeq \mathrm{Gal}(L/K)$.

Remark 28.5. When K is a number field, θ_K is surjective but not injective; its kernel is the connected component of the identity in C_K. When K is a global function field, θ_K is injective but not surjective; its image consists of automorphisms $\sigma \in \mathrm{Gal}(K^{\mathrm{ab}}/K)$ corresponding to integer powers of the Frobenius automorphism of $\mathrm{Gal}(k^{\mathrm{sep}}/k)$, where k is the constant field of K (this is precisely the dense image of $\mathbb{Z}$ in $\widehat{\mathbb{Z}} \simeq \mathrm{Gal}(k^{\mathrm{sep}}/k)$).

We also have a global existence theorem.

Theorem 28.6 (Global Existence Theorem)**.** *Let K be a global field. For every finite index open subgroup H of C_K there is a unique finite abelian extension L/K in K^{ab} for which $\mathrm{N}_{L/K}(C_L) = H$.*

As with the local Artin homomorphism, taking profinite completions yields an isomorphism that allows us to summarize global class field theory in one statement.

Theorem 28.7 (Main theorem of global class field theory)**.** *Let K be a global field. The global Artin homomorphism θ_K induces a canonical isomorphism*

$$\widehat{\theta}_K \colon \widehat{C_K} \xrightarrow{\sim} \mathrm{Gal}(K^{\mathrm{ab}}/K)$$

of profinite groups.

We then have an inclusion reversing bijection

$$\begin{aligned} \{\,\text{finite index open subgroups } H \text{ of } C_K\,\} &\longleftrightarrow \{\,\text{finite abelian extensions } L/K \text{ in } K^{\mathrm{ab}}\,\} \\ H &\mapsto (K^{\mathrm{ab}})^{\theta_K(H)} \\ \mathrm{N}_{L/K}(C_L) &\mapsfrom L \end{aligned}$$

and corresponding isomorphisms $C_K/H \simeq \mathrm{Gal}(L/K)$, where $H = \mathrm{N}_{L/K}(C_L)$. We also note that the global Artin homomorphism is *functorial* in the following sense.

Theorem 28.8 (Functoriality)**.** *Let K be a global field and let L/K be any finite separable extension (not necessarily abelian). Then the following diagram commutes*

$$\begin{array}{ccc} C_L & \xrightarrow{\theta_L} & \mathrm{Gal}(L^{\mathrm{ab}}/L) \\ \downarrow{\scriptstyle N_{L/K}} & & \downarrow{\scriptstyle \mathrm{res}} \\ C_K & \xrightarrow{\theta_K} & \mathrm{Gal}(K^{\mathrm{ab}}/K). \end{array}$$

28.4 Relation to ideal-theoretic version of global class field theory

Let K be a number field and let $\mathfrak{m}\colon M_K \to \mathbb{Z}_{\geq 0}$ be a modulus for K, which we view as a formal product $\mathfrak{m} = \prod_v v^{e_v}$ over the places v of K with $e_v \leq 1$ when v is archimedean and $e_v = 0$ when v is complex (see Definition 21.2). For each place v we define the open subgroup

$$U_K^{\mathfrak{m}}(v) := \begin{cases} \mathcal{O}_v^\times & \text{if } v \nmid \mathfrak{m}, \text{ where } \mathcal{O}_v^\times := K_v^\times \text{ when } v \text{ is infinite),} \\ \mathbb{R}_{>0} & \text{if } v|\mathfrak{m} \text{ is real, where } \mathbb{R}_{>0} \subseteq \mathbb{R}^\times \simeq \mathcal{O}_v^\times := K_v^\times, \\ 1 + \mathfrak{p}^{e_v} & \text{if } v|\mathfrak{m} \text{ is finite, where } \mathfrak{p} = \{x \in \mathcal{O}_v : |x|_v < 1\}, \end{cases}$$

and let $U_K^{\mathfrak{m}} := \prod_v U_K^{\mathfrak{m}}(v) \subseteq \mathbb{I}_K$ denote the corresponding open subgroup of $\mathbb{I}_K$. The image $\overline{U}_K^{\mathfrak{m}}$ of $U_K^{\mathfrak{m}}$ in the idele class group $C_K = \mathbb{I}_K/K^\times$ is a finite index open subgroup. The idelic version of a ray class group is the quotient

$$C_K^{\mathfrak{m}} := \mathbb{I}_K/(U_K^{\mathfrak{m}} K^\times) = C_K/\overline{U}_K^{\mathfrak{m}},$$

and we have isomorphisms

$$C_K^{\mathfrak{m}} \simeq \mathrm{Cl}_K^{\mathfrak{m}} \simeq \mathrm{Gal}(K(\mathfrak{m})/K),$$

where $\mathrm{Cl}_K^{\mathfrak{m}}$ is the ray class group for the modulus $\mathfrak{m}$ (see Definition 21.3), and $K(\mathfrak{m})$ is the corresponding *ray class field*, which we can now define as the finite abelian extension L/K for which $\mathrm{N}_{L/K}(C_L) = \overline{U}_K^{\mathfrak{m}}$, whose existence is guaranteed by Theorem 28.6.

If L/K is any finite abelian extension, then $\mathrm{N}_{L/K}(C_L)$ contains $\overline{U}_K^{\mathfrak{m}}$ for some modulus $\mathfrak{m}$; this follows from the fact that the groups $\overline{U}_K^{\mathfrak{m}}$ form a fundamental system of open neighborhoods of the identity. Indeed, the conductor of the extension L/K (see Definition 22.24) is precisely the minimal modulus $\mathfrak{m}$ for which this is true. It follows that every finite abelian extension L/K lies in a ray class field $K(\mathfrak{m})$, with $\mathrm{Gal}(L/K)$ isomorphic to a quotient of a ray class group $C_K^{\mathfrak{m}}$.

28.5 The Chebotarev density theorem

We conclude this lecture with a proof of the Chebotarev density theorem, a generalization of the Frobenius density theorem you proved on Problem Set 10. Recall from Lecture 18 and Problem Set 9 that if S is a set of primes of a number field K, the *Dirichlet density* of S is defined by

$$d(S) := \lim_{s \to 1^+} \frac{\sum_{\mathfrak{p} \in S} \mathrm{N}(\mathfrak{p})^{-s}}{\sum_{\mathfrak{p}} \mathrm{N}(\mathfrak{p})^{-s}} = \lim_{s \to 1^+} \frac{\sum_{\mathfrak{p} \in S} \mathrm{N}(\mathfrak{p})^{-s}}{\log \frac{1}{s-1}},$$

whenever this limit exists. As you proved on Problem Set 9, if S has a natural density then it has a Dirichlet density and the two coincide (and similarly for polar density).

In order to state Chebotarev's density theorem we need one more definition: a subset C of a group G is said to be *stable under conjugation* if $\sigma\tau\sigma^{-1} \in C$ for all $\sigma \in G$ and $\tau \in C$. Equivalently, C is a union of conjugacy classes of G.

Theorem 28.9 (Chebotarev density theorem)**.** *Let L/K be a finite Galois extension of number fields with Galois group $G := \mathrm{Gal}(L/K)$. Let $C \subseteq G$ be stable under conjugation, and let S be the set of primes $\mathfrak{p}$ of K unramified in L with $\mathrm{Frob}_{\mathfrak{p}} \subseteq C$. Then $d(S) = \#C/\#G$.*

Note that G is not assumed to be abelian, so $\mathrm{Frob}_{\mathfrak{p}}$ is a conjugacy class, not an element. However, the main difficulty in proving the Chebotarev density theorem (and the only place where class field theory is used) occurs when G is abelian, in which case $\mathrm{Frob}_{\mathfrak{p}}$ contains a single element. The main result we need is a corollary of the generalization of Dirichlet's theorem on primes in arithmetic progressions to number fields that we proved in Lecture 22, a special case of which we record below.

Proposition 28.10. *Let $\mathfrak{m}$ be a modulus for a number field K and let $\mathrm{Cl}_K^{\mathfrak{m}}$ be the corresponding ray class group. For every ray class $c \in \mathrm{Cl}_K^{\mathfrak{m}}$ the Dirichlet density of the set of primes $\mathfrak{p}$ of K that lie in c is $1/\#\mathrm{Cl}_K^{\mathfrak{m}}$.*

Proof. Apply Corollary 22.22 to the congruence subgroup $\mathcal{C} = \mathcal{R}_K^{\mathfrak{m}}$. □

The Chebotarev density theorem for abelian extensions follows from Proposition 28.10 and the existence of ray class fields, which we now assume.[1]

Corollary 28.11. *Let L/K be a finite abelian extension of number fields with Galois group G. For every $\sigma \in G$ the Dirichlet density of the set S of primes $\mathfrak{p}$ of K unramified in L for which $\mathrm{Frob}_{\mathfrak{p}} = \{\sigma\}$ is $1/\#G$.*

Proof. Let $\mathfrak{m} = \mathrm{cond}(L/K)$ be the conductor of the extension L/K; then L is a subfield of the ray class field $K(\mathfrak{m})$ and $\mathrm{Gal}(L/K) \simeq \mathrm{Cl}_K^{\mathfrak{m}}/H$ for some subgroup H of the ray class group. For each unramified prime $\mathfrak{p}$ of K we have $\mathrm{Frob}_{\mathfrak{p}} = \{\sigma\}$ if and only if $\mathfrak{p}$ lies in one of the ray classes contained in the coset of H in $\mathrm{Cl}_K^{\mathfrak{m}}/H$ corresponding to σ. The Dirichlet density of the set of primes in each ray class is $1/\#\mathrm{Cl}_K^{\mathfrak{m}}$, by Proposition 28.10, and there are $\#H$ ray classes in each coset of H; thus $d(S) = \#H/\#\mathrm{Cl}_K^{\mathfrak{m}} = 1/\#G$. □

We now derive the general case from the abelian case.

Proof of the Chebotarev density theorem. It suffices to consider the case where C is a single conjugacy class, which we now assume; we can reduce to this case by partitioning C into conjugacy classes and summing Dirichlet densities (as proved on Problem Set 9). Let S be the set of primes $\mathfrak{p}$ of K unramified in L for which $\mathrm{Frob}_{\mathfrak{p}}$ is the conjugacy class C.

Let $\sigma \in G$ be a representative of the conjugacy class C, let $H_\sigma := \langle\sigma\rangle \subseteq G$ be the subgroup it generates, and let $F_\sigma := L^{H_\sigma}$ be the corresponding fixed field. Let T_σ be the set of primes $\mathfrak{q}$ of F_σ unramified in L for which $\mathrm{Frob}_{\mathfrak{q}} = \{\sigma\} \subseteq \mathrm{Gal}(L/F_\sigma) \subseteq \mathrm{Gal}(L/K)$ (note that the Frobenius class $\mathrm{Frob}_{\mathfrak{q}}$ is a singleton because $\mathrm{Gal}(L/F_\sigma) = H_\sigma$ is abelian). We have $d(T_\sigma) = 1/\#H_\sigma$, since L/F_σ is abelian, by Corollary 28.11.[2]

As you proved on Problem Set 9, restricting to degree-1 primes (primes whose residue field has prime order) does not change Dirichlet densities, so let us replace S and T_σ by their subsets of degree-1 primes, and define $T_\sigma(\mathfrak{p}) := \{\mathfrak{q} \in T_\sigma : \mathfrak{q}|\mathfrak{p}\}$ for each $\mathfrak{p} \in S$.

Claim: For each prime $\mathfrak{p} \in S$ we have $\#T_\sigma(\mathfrak{p}) = [G : H_\sigma]$.

Proof of claim: Let $\mathfrak{r}$ be a prime of L lying above $\mathfrak{q} \in T_\sigma(\mathfrak{p})$. Such an $\mathfrak{r}$ is unramified, since $\mathfrak{p}$ is, and we have $\mathrm{Frob}_{\mathfrak{r}} = \sigma$, since $\mathrm{Frob}_{\mathfrak{q}} = \{\sigma\}$. It follows that $\mathrm{Gal}(\mathbb{F}_{\mathfrak{r}}/\mathbb{F}_{\mathfrak{q}}) = \langle\bar{\sigma}\rangle \simeq H_\sigma$.

[1] This assumption is not necessary; indeed Chebotarev proved his density theorem in 1923 without it. With slightly more work one can derive the general case from the cyclotomic case $L = K(\zeta)$, where ζ is a primitive root of unity, which removes the need to assume the existence of ray class fields; see [4] for details.

[2] Note that the integers $\#H_\sigma$ and $[G : H_\sigma]$ do not depend on the choice of σ (the H_σ are all conjugate).

Therefore $f_{\mathfrak{r}/\mathfrak{q}} = \#H_\sigma$ and $\#\{\mathfrak{r}|\mathfrak{q}\} = 1$, since $\#H_\sigma = [L : F_\sigma] = \sum_{\mathfrak{r}|\mathfrak{q}} e_{\mathfrak{r}/\mathfrak{q}} f_{\mathfrak{r}/\mathfrak{q}}$. We have $f_{\mathfrak{r}/\mathfrak{p}} = f_{\mathfrak{r}/\mathfrak{q}} f_{\mathfrak{q}/\mathfrak{p}} = f_{\mathfrak{r}/\mathfrak{q}} = \#H_\sigma$, since $f_{\mathfrak{q}/\mathfrak{p}} = 1$ for degree-1 primes $\mathfrak{q}|\mathfrak{p}$, and $e_{\mathfrak{r}/\mathfrak{p}} = 1$, thus

$$\#G = [L : K] = \sum_{\mathfrak{r}|\mathfrak{p}} e_{\mathfrak{r}/\mathfrak{p}} f_{\mathfrak{r}/\mathfrak{p}} = \#\{\mathfrak{r}|\mathfrak{p}\}\#H_\sigma = \#T_\sigma(\mathfrak{p})\#H_\sigma,$$

so $\#T_\sigma(\mathfrak{p}) = \#G/\#H_\sigma = [G : H_\sigma]$ as claimed.

We now observe that

$$\sum_{\mathfrak{p}\in S} \mathrm{N}(\mathfrak{p})^{-s} = \sum_{\sigma\in C} \sum_{\mathfrak{p}\in S} \frac{1}{[G : H_\sigma]} \sum_{\mathfrak{q}\in T_\sigma(\mathfrak{p})} \mathrm{N}(\mathfrak{q})^{-s} = \frac{\#C}{[G : H_\sigma]} \sum_{\mathfrak{q}\in T_\sigma} \mathrm{N}(\mathfrak{q})^{-s}$$

since $\mathrm{N}(\mathfrak{q}) = \mathrm{N}(\mathfrak{p})$ for each degree-1 prime $\mathfrak{q}$ lying above a degree-1 prime $\mathfrak{p}$, and therefore

$$d(S) = \frac{\#C}{[G : H_\sigma]} d(T_\sigma) = \frac{\#C\#H_\sigma}{[G : H_\sigma]} = \frac{\#C}{\#G}.$$ □

Remark 28.12. The Chebotarev density theorem holds for any global field; the generalization to function fields was originally proved by Reichardt [3]; see [2] for a modern proof (and in fact a stronger result). In the case of number fields (but not function fields!) Chebotarev's theorem also holds for natural density. This follows from results of Hecke [1] that actually predate Chebotarev's work; Hecke showed that the primes lying in any particular ray class have a natural density.

References

[1] Erich Hecke, *Über die L-Funktionen und den Dirichletschen Primzahlsatz für einen beliebigen Zahlkörper*, Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch–Physikalische Klasse (1917) 299–318.

[2] Michiel Kosters, *A short proof of a Chebotarev density theorem for function fields*, arXiv:1404.6345.

[3] Hans Reichardt, *Der Primdivisorsatz für algebraische Funktionenkörper über einem endlichen Konstantenkörper*, Mathematische Zeitschrift **40** (1936) 713–719.

[4] Peter Stevenhagen and H.W. Lenstra Jr., *Chebotarev and his density theorem*, Math. Intelligencer **18** (1996), 26–37.

18.785 Number Theory I
Fall 2019