MATH 154. ALGEBRAIC NUMBER THEORY

LECTURES BY BRIAN CONRAD, NOTES BY AARON LANDESMAN

CONTENTS

Thanks to Nitya Mani for note-taking on two days when Aaron Landesman was away.

1. Fermat's factorization method

Today we will discuss the proof of the following result, and highlight some of its main ideas that will be important themes in the course:

Theorem 1.1 (Fermat)**.** *For $x, y \in \mathbf{Z}$, the only solutions of $y^2 = x^3 - 2$ are $(3, \pm 5)$.*

Remark 1.2. This is an *elliptic curve* (a notion we shall not define, essentially the set of solutions to a certain type of cubic equation in two variables) with infinitely many $\mathbf{Q}$-points, a contrast with finiteness of its set of $\mathbf{Z}$-points.

Fermat's equation can be rearranged into the form $x^3 = y^2 + 2$.

Lemma 1.3. *For any $\mathbf{Z}$-solution (x, y) to $x^3 = y^2 + 2$, the value of y must be odd.*

Proof. Indeed, if y is even then x is even, so x^3 is divisible by 8. But $y^2 + 2 = 4k + 2$ is not divisible by 8. □

Fermat's first great idea is to introduce considerations in the ring

$$\mathbf{Z}[\sqrt{-2}] := \{m + n\sqrt{-2} \mid m, n \in \mathbf{Z}\}.$$

The point is that over this ring, the equation

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

can be expressed as

$$x^3 = (y + \sqrt{-2})(y - \sqrt{-2}).$$

To find solutions to this, we ask the following two questions:

Question 1.4. Do $y + \sqrt{-2}$ and $y - \sqrt{-2}$ have "gcd = 1" (meaning no nontrivial factor in common, where "nontrivial" means "not a unit")?

Question 1.5. If we know the answer to the above question is "yes," can we conclude that both $y \pm \sqrt{-2}$ are cubes in $\mathbf{Z}[\sqrt{-2}]$?

Recall the definition of unit:

Definition 1.6. For R a commutative ring, an element $u \in R$ is a *unit* if there exists $u' \in R$, so that $uu' = 1$. We let

$$R^\times := \{ \text{ units in } R\};$$

this is an abelian group.

Example 1.7. (1) $\mathbf{Z}^\times = \{\pm 1\}$.
(2) For F a field, $F^\times = F - \{0\}$ by the definition of a field.

(3) We have that $1+\sqrt{2}$ is a unit in $\mathbf{Z}[\sqrt{2}]$ because

$$(1+\sqrt{2})(-1+\sqrt{2}) = 1.$$

Thus, likewise $(1+\sqrt{2})^n \in \mathbf{Z}[\sqrt{2}]^\times$ for all $n \in \mathbf{Z}$.

Note that if $u \in R^\times$ then for any $x \in R$ we have $x = (xu^{-1})u$. Hence, for any consideration of "unique" factorization we must allow for adjusting factors by unit multiples (absorbing the inverse unit elsewhere in the factorization).

Definition 1.8. A *domain* (sometimes also called an *integral domain*) is a nonzero commutative ring R such that if $ab = 0$ with $a, b \in R$ then either $a = 0$ or $b = 0$.

For a domain R, if $a, b \in R - \{0\}$ and $a \mid b$ and $b \mid a$ then $a = bu$ for $u \in R^\times$. (Indeed, if $a = bs$ and $b = at$ then $a = a(ts)$, so $ts = 1$ since $a \neq 0$ and hence $s, t \in R^\times$.) Away from $\mathbf{Z}^+$, "$a|b, b|a \Rightarrow a = b$" is very rare.

Example 1.9. Later, we'll see

$$\mathbf{Z}[\sqrt{2}]^\times = \{\pm 1\} \times (1+\sqrt{2})^{\mathbf{Z}}.$$

This is non-obvious!

Question 1.10. What is $\mathbf{Z}[\sqrt{-2}]^\times$?

In fact, we'll answer this more generally:

Lemma 1.11. *Let d be any non-square integer > 1. We have*

$$\mathbf{Z}[\sqrt{-d}]^\times = \{\pm 1\}.$$

Proof. Consider

$$\alpha := u + v\sqrt{-d}$$

for $u, v \in \mathbf{Z}$. Let

$$\overline{\alpha} := u - v\sqrt{-d},$$

so

$$\alpha\overline{\alpha} = u^2 + dv^2 \in \mathbf{Z}_{\geq 0}.$$

Since

$$\overline{\alpha\beta} = \overline{\alpha}\overline{\beta},$$

if $\alpha\beta = 1$ then $\overline{\alpha}\overline{\beta} = \overline{1} = 1$, which implies

$$(\alpha\overline{\alpha})(\beta\overline{\beta}) = 1,$$

so

$$\alpha\overline{\alpha} = 1$$

because $\alpha\overline{\alpha} = u^2 + dv^2 \in \mathbf{Z}_{\geq 0}$. The equality $u^2 + dv^2 = 1$ with $d > 1$ forces $v = 0$ and then $u = \pm 1$. □

Remark 1.12. If we try the same argument for $\mathbf{Z}[\sqrt{d}]$, we get $u^2 - dv^2 = \pm 1$ with $u, v \in \mathbf{Z}$ (and $v \neq 0$ for units distinct from ± 1). The question of finding such non-trivial units then becomes Pell's equation, which is

$$u^2 - dv^2 = 1$$

and its variant with -1 on the right side. For the case $d = 2$, Pell's equation has infinitely many $\mathbf{Z}$-solutions with $u, v > 0$ (as we will recover later in the course from a more sophisticated point of view), beginning with:

$$(3, 2), (17, 12), \ldots$$

by considering powers $(1 + \sqrt{2})^{2n} = (3 + 2\sqrt{2})^n$ for $n > 0$.

The variant equation $u^2 - 2v^2 = -1$ also has infinitely many $\mathbf{Z}$-solutions with $u, v > 0$ by considering $(1 + \sqrt{2})^{2n+1}$ with $n \geq 0$, such as:

$$(1, 1), (7, 5), \ldots$$

We'll now resume our goal of finding the $\mathbf{Z}$-solutions to $x^3 = y^2 + 2$, first addressing if $y \pm \sqrt{-2}$ have a non-unit common factor. The answer is negative:

Lemma 1.13. *If $\delta \in \mathbf{Z}[\sqrt{-2}]$ satisfies*

$$\delta \mid (y + \sqrt{-2}) \text{ and } \delta \mid (y - \sqrt{-2})$$

then δ is a unit.

Proof. First, recall from Lemma 1.3 that y is odd. Observe that

$$\begin{aligned} \delta \mid & (y + \sqrt{-2}) - (y - \sqrt{-2}) \\ &= 2\sqrt{-2} \\ &= (\sqrt{-2})^3. \end{aligned}$$

Then, we claim the following sublemma:

Lemma 1.14. *The element $\sqrt{-2} \in \mathbf{Z}[\sqrt{-2}]$ is irreducible (i.e., it is a nonzero non-unit such that any factorization as $\alpha\beta$ must have α or β a unit).*

Proof. Assuming $\sqrt{-2} = \alpha\beta$, we have $-\sqrt{-2} = \overline{\alpha}\overline{\beta}$ by conjugating both sides. Multiplying these relations, $2 = (\alpha\overline{\alpha})(\beta\overline{\beta}) \in \mathbf{Z}_{\geq 0}$, so $\alpha\overline{\alpha} = 1$ or $\beta\overline{\beta} = 1$. This implies either α or β is a unit, as desired. □

In HW1 it will be shown that $\mathbf{Z}[\sqrt{-2}]$ is a UFD, so the irreducibility of $\sqrt{-2}$ forces $\delta = u\sqrt{-2}^e$ for some $0 \leq e \leq 3$ and some unit $u \in \mathbf{Z}[\sqrt{-2}]$. Thus, if δ is not a unit then $\sqrt{-2} \mid \delta$. Hence, to get a contradiction (and conclude δ is a unit) it is enough to show $\sqrt{-2} \nmid (y + \sqrt{-2})$ in $\mathbf{Z}[\sqrt{-2}]$.

Suppose for some $u, v \in \mathbf{Z}$ that

$$\begin{aligned} y + \sqrt{-2} &= \sqrt{-2}(u + v\sqrt{-2}) \\ &= 2v + u\sqrt{-2}. \end{aligned}$$

This forces $y = 2v$ to be even, but y is odd by Lemma 1.3. This concludes the proof that δ is a unit. □

We conclude that in the UFD $\mathbf{Z}[\sqrt{-2}]$ we have

$$\gcd(y + \sqrt{-2}, y - \sqrt{-2}) = 1.$$

This completes the preparations for:

Proof of Theorem 1.1. In any UFD, any nonzero element is a finite product of irreducibles, and by lumping together any irreducibles that agree up to unit multiple we can rewrite such a product as

$$u \cdot \prod_i \pi_i^{e_i}$$

with π_i pairwise non-associate irreducible elements (*non-associate* means one is not a unit times another; by irreducibility of the π_i's, this amounts to saying $\pi_i \nmid \pi_j$ for any $i \neq j$).

Since

$$(y + \sqrt{-2})(y - \sqrt{-2}) = x^3$$

with

$$\gcd(y + \sqrt{-2}, y - \sqrt{-2}) = 1,$$

it follows from expressing each of

$$y \pm \sqrt{-2}$$

as a unit multiple of a product of pairwise non-associate irreducibles that all irreducible factors of $y \pm \sqrt{-2}$ occur with multiplicity divisible by 3. Therefore, $y \pm \sqrt{-2}$ is the product of a unit and a cube. But now a miracle occurs: we know the units of $\mathbf{Z}[\sqrt{-2}]$ by Lemma 1.11, and from this we see that all units are themselves cubes (as $-1 = (-1)^3$)! Hence,

$$y + \sqrt{-2} = (a + b\sqrt{-2})^3$$

for some $a, b \in \mathbf{Z}$.

Therefore,

$$\begin{aligned} y+\sqrt{-2} &= (a+b\sqrt{-2})^3 \\ &= a(a^2-6b^2)+b(3a^2-2b^2)\sqrt{-2}, \end{aligned}$$

so

$$1 = b(3a^2-2b^2),$$

from which we see that $b = \pm 1$, so $b^2 = 1$. This implies $\pm 1 = 3a^2 - 2$ so

$$3a^2 = 2 \pm 1.$$

Then $3a^2$ equals either 3 or 1. The latter is impossible because we cannot have $3a^2 = 1$ with $a \in \mathbf{Z}$. Thus, $3a^2 = 3$, so $a = \pm 1$. This implies $b = 1$, so

$$\begin{aligned} y &= a(a^2-6b^2) \\ &= \pm(1-6) \\ &= \pm 5 \end{aligned}$$

and $x^3 = y^2 + 2 = 27$, so $x = 3$, concluding the proof. □

The above considerations yield the following lessons:

(1) For studying $\mathbf{Z}$-solutions to polynomial equations, it's useful to consider arithmetic in larger number systems. For example, in this case, it was useful to work in $\mathbf{Z}[\sqrt{-2}]$.

(2) We have to know about unique factorization and units in such number systems. For example, in this case, we got lucky that all elements in

$$\mathbf{Z}[\sqrt{-2}]^\times = \{\pm 1\}$$

are cubes. This is quite false for $\mathbf{Z}[\sqrt{2}]$ since (as we'll see later)

$$\mathbf{Z}[\sqrt{2}]^\times = (\pm 1) \times (1+\sqrt{2})^{\mathbf{Z}};$$

e.g., $1+\sqrt{2}$ is a non-cube unit. The need to grapple with non-cube units makes it much harder to analyze the $\mathbf{Z}$-solutions to the variant $y^2 = x^3 + 2$ of Fermat's equation because it is harder to determine the structure of the relevant unit group.

Let's now consider another question of unique factorization in an imaginary quadratic situation.

Example 1.15. Is $\mathbf{Z}[\sqrt{-3}]$ a unique factorization domain? By Lemma 1.11, the only units are $\{\pm 1\}$. We can write

$$4 = 2 \cdot 2$$

and

$$4 = (1 + \sqrt{3})(1 - \sqrt{-3}).$$

We claim:

Lemma 1.16. *The element* 2 *is irreducible in* $\mathbf{Z}[\sqrt{-3}]$.

Proof. Say $2 = \alpha\beta$, so by conjugating both sides we have $2 = \overline{\alpha}\overline{\beta}$. Multiplying both relations gives

$$4 = (\alpha\overline{\alpha})(\beta\overline{\beta}).$$

with $\alpha\overline{\alpha}$ and $\beta\overline{\beta}$ each a non-negative integer since for $\alpha = u + v\sqrt{-3}$ with $u, v \in \mathbf{Z}$ we have $\alpha\overline{\alpha} = u^2 + 3v^2$. But

$$u^2 + 3v^2 \neq 2$$

for $u, v \in \mathbf{Z}$, so either $\alpha\overline{\alpha} = 1$ or $\beta\overline{\beta} = 1$. This shows that either α or β is a unit, so 2 is irreducible in $\mathbf{Z}[\sqrt{-3}]$. □

Since neither $1 \pm \sqrt{-3}$ is divisible by 2 in $\mathbf{Z}[\sqrt{-3}]$, as clearly neither is of the form

$$2(u + v\sqrt{-3})$$

with $u, v \in \mathbf{Z}$, it follows from the two factorizations of 4 that $\mathbf{Z}[\sqrt{-3}]$ is not a UFD.

In the quadratic field $\mathbf{Q}(\sqrt{-3})$, we have

$$\mathbf{Q}(\sqrt{-3}) = \mathbf{Q}(\zeta_3)$$

with

$$\zeta_3 := \frac{-1 + \sqrt{-3}}{2}$$

a primitive cube root of unity; i.e., a root of the polynomial

$$\frac{x^3 - 1}{x - 1} = x^2 + x + 1,$$

Since $\sqrt{-3} = 2\zeta_3 + 1$ and

$$\zeta_3^2 = -1 - \zeta_3,$$

we have

$$\mathbf{Z}[\sqrt{-3}] \subset \mathbf{Z}[\zeta_3] = \mathbf{Z} \oplus \mathbf{Z}\zeta_3.$$

We'll later prove that the domain $\mathbf{Z}[\zeta_3]$ is a UFD. Thus, although $\mathbf{Z}[\sqrt{-3}]$ fails to be a UFD, perhaps it is also the "wrong" ring to consider when contemplating "arithmetic" inside $\mathbf{Q}(\sqrt{-3})$; that is, it seems that $\mathbf{Z}[\zeta_3]$ is a better ring to use. But how can one arrive at this determination in a systematic way? More broadly:

Question 1.17. For $K/\mathbf{Q}$ a finite extension, what is the "correct" notion of "ring of integers" for K?

Remark 1.18. Euler's proof of Fermat's Last Theorem for exponent 3 assumed that $\mathbf{Z}[\sqrt{-3}]$ is a UFD (which is of course false by Example 1.15!) This proof can be fixed by working in $\mathbf{Z}[\zeta_3]$.

The proof of Fermat's Last Theorem for exponent 4 is [Samuel, §1.2], after which the essential case for proving Fermat's Last Theorem is that with exponent an odd prime p. If $z^p = x^p + y^p$ for $p > 2$ prime then

$$y^p = z^p - x^p = \prod (z - \zeta_p^j x)$$

with ζ_p a primitive pth root of unity. In general, it will turn out that

$$\mathbf{Z}[\zeta_p]^\times$$

is infinite whenever $p > 3$, and rather beyond the scope of this course is the fact that $\mathbf{Z}[\zeta_p]$ is not a UFD whenever $p > 19$. Nonetheless, we will see that $\mathbf{Z}[\zeta_p]$ is the correct notion of "ring of integers" for the field $\mathbf{Q}(\zeta_p)$, and despite its failure to be a UFD in general it does have a lot of nice structural features that allow one to make serious progress on Fermat's Last Theorem for lots of p (and the final solution by Wiles uses techniques of a much more advanced nature, though building very much on the classic ideas of algebraic number theory to be developed in this course).

2. Quadratic norms

Last time, we discussed $\mathbf{Z}$-solutions to

$$y^2 = x^3 - 2$$

which we saw were $(3, \pm 5)$ via factoring in a suitable subring of the field $\mathbf{Q}(\sqrt{-2})$. It turns out that the only $\mathbf{Z}$-solutions to

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

are $(-1, \pm 1)$ via factorization in the subring $\mathbf{Z}[\sqrt{2}] \subset \mathbf{Q}(\sqrt{2})$. This is harder to prove, but follows with some work once one shows

$$\mathbf{Z}[\sqrt{2}]^\times = \pm(1+\sqrt{2})^{\mathbf{Z}}.$$

To tackle these problems, we need a way to show rings such as $\mathbf{Z}[\sqrt{\pm 2}]$ are UFD's. The most rudimentary way to do this is via the notion of Euclidean domain (since Euclidean domains are PID's, hence are UFD's).

Euclidean domains essentially are only used in first courses in algebra, and once one develops more algebraic tools, one can more easily show a wider class of rings are PID's whereas many rings we encounter in number theory that are PID's turn out not to be Euclidean domains.

The Euclidean-domain approach will work for some rings like $\mathbf{Z}[i]$ and $\mathbf{Z}[\sqrt{\pm 2}]$, but quickly runs out of steam for the study of $\mathbf{Q}(\sqrt{d})$ with squarefree $d \in \mathbf{Z}$ once $|d|$ grows beyond a small set of values.

We next review the definition of Euclidean domain:

Definition 2.1. A domain R is *Euclidean* if it admits a function

$$\nu : R \to \mathbf{Z}_{\geq 0}$$

(which is not necessarily multiplicative) so that

(1) $\nu(x) = 0 \iff x = 0$,
(2) For any $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ so that $a = bq + r$ with $\nu(r) < \nu(b)$.

Remark 2.2. There is no uniqueness assumption for q, r in the second condition of Definition 2.1.

Example 2.3. Take $R = \mathbf{Z}, \nu(x) = |x|$. Then, one could use the "greedy" division algorithm, with $-b/2 \leq r \leq b/2$. In this case r is not always unique (namely, $b/2$ can be exchanged for $-b/2$ when b is even, at the cost of changing q by 1).

Example 2.4. If $R = k[t]$ over a field k, define

$$\nu(f) = \deg f.$$

This is not quite a function satisfying the definition of Euclidean domain because ν takes value 0 on the non-zero constants too, but those are units and so it is not a real obstacle to extracting the PID conclusion; one might call such a situation "pseudo-Euclidean."

Definition 2.5. We say $\mathbf{Z}[\sqrt{d}] \subset \mathbf{Q}(\sqrt{d})$ is *norm-Euclidean* if it is Euclidean with respect to the map

$$\nu(x) = |N(x)| := x\overline{x},$$

with $x \mapsto \overline{x}$ the nontrivial element of $\mathrm{Gal}(\mathbf{Q}(\sqrt{d})/\mathbf{Q})$.

2.1. **The Gaussian integers are Euclidean.** The next result we prove is:

Theorem 2.6. *The ring* $\mathbf{Z}[i]$ *is Euclidean.*

There will be some proofs on the homework that other quadratic extensions of $\mathbf{Z}$ are Euclidean, modelled on the algebraic method below.

Before proceeding with the proof, let's first consider the general notion of quadratic norm. Let $K/\mathbf{Q}$ be a degree 2 extension. We have $K = \mathbf{Q}(\sqrt{d})$ for a unique squarefree $d \in \mathbf{Z} - \{0,1\}$. Let $x \mapsto \overline{x}$ be the nontrivial element of $\mathrm{Gal}(K/\mathbf{Q})$. If $x = u + v\sqrt{d}$ with $u, v \in \mathbf{Q}$ then

$$\overline{x} = u - v\sqrt{d}.$$

We define the *norm*

$$\begin{aligned} N := N_{K/\mathbf{Q}} \colon K &\to \mathbf{Q} \\ x &\mapsto x\overline{x} := u^2 - dv^2, \end{aligned}$$

so since $\overline{xy} = \overline{x} \cdot \overline{y}$ we have

$$N(xy) = N(x)N(y).$$

We will verify Theorem 2.6 using a ν derived from this norm operation.

Remark 2.7. The norm operation N can be defined for general finite extensions of any field, as we'll see later, but today we'll just focus on quadratic extensions of $\mathbf{Q}$.

Example 2.8. Consider

$$N : \mathbf{Z}[\sqrt{d}] \to \mathbf{Z}$$

so

$$\nu := |N| : \mathbf{Z}[\sqrt{d}] \to \mathbf{Z}_{\geq 0}$$

takes values in the non-negative integers and

$$\nu(x) = 0 \iff x = 0.$$

For example, if $d = -1$, then

$$\nu(u + vi) = u^2 + v^2$$

for $u, v \in \mathbf{Q}$. In general, note that $|N| = N$ for $d < 0$.

Proof of Theorem 2.6. We'll show the Euclidean property using the norm

$$\nu(\alpha_1 + \alpha_2 i) = \alpha_1^2 + \alpha_2^2 = \alpha\overline{\alpha}$$

for $\alpha = \alpha_1 + \alpha_2 i$ with $\alpha_j \in \mathbf{Q}$.

The idea of the proof is the following: if we are to have $a = bq + r$ with $\nu(r) < \nu(b)$ for some $q, r \in \mathbf{Z}[i]$ then

$$\mathbf{Q}(i) \ni \frac{a}{b} = q + \frac{r}{b},$$

with $q \in \mathbf{Z}[i]$ and r/b at distance < 1 from 0 since $\nu(r/b) = \nu(r)/\nu(b) < 1$. This motivates the geometric idea to *construct* q and r: we try to take q to be the point in the lattice $\mathbf{Z}[i] = \mathbf{Z} \oplus \mathbf{Z}i$ nearest to $a/b \in \mathbf{Q}(i) \subset \mathbf{C}$.

In algebraic terms, since

$$\frac{a}{b} \in \mathbf{Q}(i) = \mathbf{Q}[i] = \{u + iv \mid u, v \in \mathbf{Q}\}$$

we can write

$$\frac{a}{b} = t_1 + t_2 i$$

with $t_1, t_2 \in \mathbf{Q}$, so if $q_j \in \mathbf{Z}$ is the nearest integer to $t_j \in \mathbf{Q}$ (breaking ties arbitrarily) then

$$t_j = q_j + \varepsilon_j$$

with $q_j \in \mathbf{Z}$ and $|\varepsilon_j| \leq \frac{1}{2}$. Then,

$$\frac{a}{b} = (q_1 + q_2 i) + (\varepsilon_1 + \varepsilon_2 i) = q + \varepsilon$$

for $q \in \mathbf{Z}[i]$ and $|\varepsilon_j| \leq \frac{1}{2}$. Hence, multiplying through by b gives $a = bq + r$ with

$$r := \varepsilon b = a - bq \in \mathbf{Z}[i].$$

By multiplicativity of the norm $\mathbf{Q}(i) \to \mathbf{Q}$, we have

$$\nu(r) = \nu(\varepsilon)\nu(b) < \nu(b)$$

because

$$\nu(\varepsilon) = \varepsilon_1^2 + \varepsilon_2^2 \leq \frac{1}{4} + \frac{1}{4} = 1/2.$$

□

Warning 2.9. The very last step that

$$\nu(\varepsilon) = \varepsilon_1^2 + \varepsilon_2^2 \leq \frac{1}{4} + \frac{1}{4} = 1/2.$$

breaks when we try to adapt it to $\mathbf{Z}[\sqrt{-d}]$ for d moderately larger (such as all $d \geq 3$) because

$$\nu(\varepsilon) \leq 1/4 + d/4$$

and this upper bound is not generally below 1 anymore (but we barely scrape by successfully for $\mathbf{Z}[\sqrt{-2}]$).

Last time it was shown that for non-square integers $d > 1$, we have

$$\mathbf{Z}[\sqrt{-d}]^\times = \{\pm 1\}.$$

We can make a similar, but slightly different statement when $d = 1$ as follows. When $d = 1$ there are some additional units, namely $\pm i$, but nothing more:

Lemma 2.10. *We have*

$$\mathbf{Z}[i]^\times = \{\pm 1, \pm i\}.$$

Proof. Suppose

$$\alpha = \alpha_1 + \alpha_2 i \in \mathbf{Z}[i]^\times$$

with $\alpha_1, \alpha_2 \in \mathbf{Z}$. Then, there exists $\beta \in \mathbf{Z}[i]$ with

$$\alpha\beta = 1.$$

by definition of unit. Taking norms, we have

$$N(\alpha)N(\beta) = N(1) = 1$$

in $\mathbf{Z}_{\geq 0}$ since

$$N(u + iv) = u^2 + v^2.$$

This forces $N(\alpha) = 1$, so $\alpha_1^2 + \alpha_2^2 = 1$. One can then see that either $\alpha_1 = 0$ or $\alpha_2 = 0$, implying

$$\alpha \in \{\pm 1, \pm i\}.$$

□

Remark 2.11. As a variant, we'll see in Exercise 2(i) of Homework 2 that in $R = \mathbf{Z}[\zeta_3]$ we have

$$R^\times = \{\pm 1, \pm\zeta_3, \pm\zeta_3^2\}.$$

2.2. **Factoring in quadratic UFD's using the norm.**

Let's now use the norm to help with factoring in a quadratic UFD.

Example 2.12. Let's find the prime factorization of $7 + 4i$ in $\mathbf{Z}[i]$.

You might think "hmm... looks prime!" Who knows? Is it prime? Where do we search? What do we do?

Let's try taking the norm to get some guidance from our experience with prime factorization in $\mathbf{Z}$. Observe that this has norm $65 = 5 \cdot 13$. If we could write $7 + 4i = \alpha\beta$ with *non-units* α, β then we would have

$$65 = N(\alpha)N(\beta) \quad \text{with } N(\alpha), N(\beta) \in \mathbf{Z}_{>1}.$$

Therefore, if anything is going to work we can swap the roles of α and β if necessary to arrange that

$$\begin{aligned} N(\alpha) &= \alpha_1^2 + \alpha_2^2 = 5, \\ N(\beta) &= \beta_1^2 + \beta_2^2 = 13. \end{aligned}$$

To cut down on some of the subsequent case-checking to find any possible α, β that *might* exist, without loss of generality we can arrange $\alpha_1 > 0$ by negating both α and β if necessary. Since $5 = 1^2 + 2^2$ and by inspection this is the *unique* way (up to order of terms) to write 5 as a sum of two squares, either $\alpha_1 = 1$ or 2, so $\alpha = 1 \pm 2i$ or $\alpha = 2 \pm i$ are the only possibilities. We similarly can uniquely write $13 = 2^2 + 3^2$, so $\pm\beta = 2 \pm 3i, 3 \pm 2i$ are the only possibilities for β up to an overall sign.

Let's just try $1 + 2i$ for α and $2 + 3i$ for β and see if we get lucky:

$$\begin{aligned} (1 + 2i)(2 + 3i) &= -4 + 7i \\ &= i(7 + 4i). \end{aligned}$$

This is off from $7 + 4i$ by an overall factor of i, so we just absorb a factor of i into one of the chosen values for α and β. Hence, we can take $\alpha = 1 + 2i$, $\beta = 3 - 2i$ to arrange that

$$\alpha\beta = 7 + 4i.$$

Remark 2.13. We got a bit lucky here: if we instead tried $\alpha = 1 + 2i$ and $\beta = 2 - 3i$ then we would get

$$\alpha\beta = (1 + 2i)(2 - 3i) = 8 + i$$

and there is no way to fix this up to get $7 + 4i$ by adjusting the choices of α and β by a unit. So in a way there is still a bit of an art to this process. We will see how to make it more systematic later.

Note further that (up to unit multiplies) this is the prime factorization of $7 + 4i$. Indeed, we only need to show that α and β are themselves irreducible, and that holds because both have prime norm, permitting us to apply:

Lemma 2.14. *Suppose $\gamma \in \mathbf{Z}[i]$ satisfies that $N(\gamma) = p \in \mathbf{Z}$ is prime. Then γ is irreducible in $\mathbf{Z}[i]$.*

Proof. Certainly $\gamma \notin \mathbf{Z}[i]^\times$. If $\gamma = xy$ for some $x, y \in \mathbf{Z}[i]$ then $p = N(\gamma) = N(x)N(y)$. Since p is prime, one of the positive $N(x)$ or $N(y)$ must equal 1, implying that x or y is a unit (with inverse given by its conjugate). □

We'd next like to address the following question:

Question 2.15. How do we find all primes in $\mathbf{Z}[i]$?

As a variant whose significance will become apparent later, we also have:

Question 2.16. For $p \in \mathbf{Z}^+$ prime, how does it factor in $\mathbf{Z}[i]$, or is it still prime in $\mathbf{Z}[i]$?

Note first that taking norms does not help for the second question (in contrast with the case of $7+4i$) since $N(p) = p^2$, which is useless.

Example 2.17. Let's try $p = 5$. In this case

$$5 = (1+2i)(1-2i)$$

since $5 = 1^2 + 2^2$. Note that $1 \pm 2i$ are not associate (as the only units are $\pm 1, \pm i$).

Example 2.18. Let's next try $p = 2$. Here, we get

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

In this case, 2 has an irreducible factor that occurs with multiplicity > 1 (accounting for unit multiples) in $\mathbf{Z}[i]$.

Example 2.19. What about $3, 7, 11$?

Lemma 2.20. *If $p \equiv 3 \bmod 4$, then p is prime in $\mathbf{Z}[i]$.*

Proof. Suppose $p = \alpha\beta$ for $\alpha, \beta \in \mathbf{Z}[i]$ non-units. Then,

$$p^2 = N(p) = N(\alpha)N(\beta)$$

with $N(\alpha), N(\beta) > 1$, so both of these norms are equal to p since p is a positive prime in $\mathbf{Z}$. Therefore, $p = \alpha_1^2 + \alpha_2^2$. Consider this modulo 4. Since the squares modulo 4 are 0 and 1, a sum of two squares can only be $0, 1, 2 \bmod 4$, so it cannot be $3 \bmod 4$. Therefore, we cannot have $N(\alpha) = p$, and hence any $p \equiv 3 \bmod 4$ does not have a non-trivial factorization in $\mathbf{Z}[i]$. □

3. Quadratic factorization

Last time we saw that $\mathbf{Z}[i]$ is a UFD with unit group $\{\pm 1, \pm i\}$, and we saw a few factorization results:

- if $\alpha \in \mathbf{Z}[i]$ and $N(\alpha)$ is prime in $\mathbf{Z}$ then α is irreducible in $\mathbf{Z}[i]$,
- $2 = (1+i)(1-i) = -i(1+i)^2$ with $1+i$ irreducible (since its norm 2 is prime in $\mathbf{Z}$),

- any $p \equiv 3 \bmod 4$ remains irreducible in $\mathbf{Z}[i]$.

(Whenever we write "p" without qualification, we implicitly mean an element of $\mathbf{Z}^+$ that is prime.)

In the remaining case $p \equiv 1 \bmod 4$, the first two values (5 and 13) were seen to be reducible in $\mathbf{Z}[i]$ due to being a sum of two squares: $5 = 1^2 + 2^2 = (1+2i)(1-2i)$ and $13 = 2^2 + 3^2 = (2+3i)(2-3i)$. This phenomenon is completely general:

Theorem 3.1 (Fermat)**.** *If $p \equiv 1 \bmod 4$ then $p = x^2 + y^2$ for some $x, y \in \mathbf{Z}$.*

Before addressing the proof, we make some observations. Once it is known that x, y exist, clearly each is nonzero and $p = (x+iy)(x-iy) = N(x+iy)$, so $\pi := x + iy$ is irreducible in $\mathbf{Z}[i]$. We claim that the irreducible factors π and $\overline{\pi}$ of p are *non-associate* (in contrast with the case $p = 2$), so p really has two distinct irreducible factors in $\mathbf{Z}[i]$ (up to unit multiple).

To see this, first note that $\gcd_{\mathbf{Z}}(x, y) = 1$ since the square of this gcd divides $x^2 + y^2 = p$. The only units in $\mathbf{Z}[i]$ are ± 1 and $\pm i$, so if $x - iy$ is a unit multiple of $x + iy$ then necessarily (since $x, y \neq 0$) we must have $y = \pm x$, so $x^2 | p$ in $\mathbf{Z}[i]$ and hence $x^2 | p$ in $\mathbf{Z}$ (since $\mathbf{Z}[i] \cap \mathbf{Q} = \mathbf{Z}$). But then $x, y = \pm 1$, so $p = x^2 + y^2 = 2$, a contradiction. This affirms that such $\pi, \overline{\pi}$ are non-associate.

Since $\mathbf{Z}[i]^\times = \{\pm 1, \pm i\}$ consists of four elements, the preceding interpretation of $x + iy$ as one of the two distinct irreducible factors of p (up to unit multiple!) in the *unique factorization domain* $\mathbf{Z}[i]$ accounts for $4 \times 2 = 8$ obvious ordered pairs $(u, v) \in \mathbf{Z}^2$ satisfying $u^2 + v^2 = p$ obtained from (x, y) by introducing signs and swapping the roles of x and y. This establishes:

Corollary 3.2. *The ordered pair (x, y) in Theorem* 3.1 *is unique up to signs and swapping the roles of x and y.*

Note how this corollary makes essential use of a *ring-theoretic* property of $\mathbf{Z}[i]$; it is not something we prove by bare-hands manipulation of equations.

Let us now discuss the proof of Theorem 3.1. The details are worked out step-by-step in HW1, so here we just make some basic observations. Since $p \equiv 1 \bmod 4$, we know $-1 \equiv \square \bmod p$ (because $\mathbf{F}_p^\times$ is cyclic of order $p-1$). For $n \in \mathbf{Z}$ satisfying $n^2 \equiv -1 \bmod p$, we have $p | (n^2+1) = (n+i)(n-i)$ in $\mathbf{Z}[i]$. If p were irreducible then the UFD property of $\mathbf{Z}[i]$ would force p to divide either $n + i$ or $n - i$, either of which leads to a contradiction by inspection (due to i having $\mathbf{Z}$-coefficient ± 1 in $n \pm i$). So since p is clearly a nonzero non-unit in $\mathbf{Z}[i]$, the only remaining option is that p must be *reducible*. From this one can get Fermat's result via norm considerations.

Now that we have understood with the help of norms how ordinary primes behave for factorization in $\mathbf{Z}[i]$, let's harness that information to characterize in terms of norms when a general element of $\mathbf{Z}[i]$ is irreducible:

Proposition 3.3. *A nonzero non-unit $\pi \in \mathbf{Z}[i]$ is irreducible $\Leftrightarrow$ $N(\pi)$ falls into any of the following three mutually exclusive cases:*

(i) $N(\pi) = p^2$ *with* $p \equiv 3 \bmod 4$ (*and then* $\pi = \pm p, \pm ip$),

(ii) $N(\pi) = p \equiv 1 \bmod 4$ (*and then* $\pi, \overline{\pi}$ *are the irreducible factors of* p *in* $\mathbf{Z}[i]$ *up to units*),

(iii) $N(\pi) = 2$ (*and then* $\pi = u(1+i)$ *for some* $u \in \mathbf{Z}[i]^\times$).

Proof. Let's first handle the easier implication "$\Leftarrow$". In case (i) we have

$$p^2 = N(\pi) = \pi\overline{\pi}$$

with p irreducible in $\mathbf{Z}[i]$. Hence, the left side has two irreducible factors occurring (up to unit multiple), so the same must hold on the right side since $\mathbf{Z}[i]$ is a UFD. But each of π and $\overline{\pi}$ is a nonzero non-unit, so each contributes at least one irreducible factor to the right side through its irreducible factorization, and there is no room for more irreducible factors on the right side (since there are only two such on the left side). Hence, π must be irreducible. By the uniqueness (up to unit multiples!) of irreducible factorization, π must agree with p up to units, which is to say $\pi = \pm p, \pm ip$. This settles (i), and both (ii) and (iii) are clear since primality of $N(\alpha)$ forces α to be irreducible in $\mathbf{Z}[i]$.

Now consider "$\Rightarrow$", so π is irreducible and hence $N(\pi) \in \mathbf{Z}_{>1}$. We can pick a prime factor p of $N(\pi)$ in $\mathbf{Z}^+$, so

$$\pi\overline{\pi} = N(\pi) = p(\cdots)$$

in $\mathbf{Z}[i]$ with $\pi, \overline{\pi}$ irreducible in $\mathbf{Z}[i]$. If $p \equiv 3 \bmod 4$ then p is an irreducible factor on the right side in $\mathbf{Z}[i]$, so by "unique factorization" in $\mathbf{Z}[i]$ it follows that p coincides with π or $\overline{\pi}$ up to unit multiple. But $\overline{p} = p$, so $p = u\pi$ for some $u \in \mathbf{Z}[i]^\times$. Applying the norm to this latter relation then gives $p^2 = N(\pi)$, so we are in case (i).

If instead $p \equiv 1 \bmod 4$ then by Theorem 3.1 we have $p = \gamma\overline{\gamma}$ for some irreducible $\gamma \in \mathbf{Z}[i]$, so

$$\pi\overline{\pi} = p(\cdots) = \gamma\overline{\gamma}(\cdots)$$

in $\mathbf{Z}[i]$. Once again using the UFD property, the irreducible factor γ on the right side must be a unit multiple of one of the irreducible factors π or $\overline{\pi}$ on the left side, so $p = N(\gamma) = N(\pi)$. This puts us into case (ii).

We have addressed all cases when $N(\pi) \in \mathbf{Z}_{>1}$ has an odd prime factor, so the only remaining case to address is when $N(\pi) = 2^e$ for some $e > 0$.

But then

$$\pi\overline{\pi} = 2^e = u(1+i)^{2e}$$

with $1+i$ irreducible and $u \in \mathbf{Z}[i]^\times$. The irreducible factor π on the left side must then coincide with $1+i$ up to a unit multiple, so $N(\pi) = N(1+i) = 2$. This is case (iii). $\square$

Since the irreducibles in $\mathbf{Z}[i]$ have now been characterized via norms, let's now see how to find the irreducible factorization of a general nonzero nonunit $\alpha \in \mathbf{Z}[i]$, at least modulo our ability to carry out prime factorization in $\mathbf{Z}$. If we write $\alpha = \alpha_1 + \alpha_2 i$ with $\alpha_j \in \mathbf{Z}$ not both zero, for $n := \gcd_{\mathbf{Z}}(\alpha_1, \alpha_2) \in \mathbf{Z}_{>0}$ we can write

$$\alpha = n(\beta_1 + \beta_2 i)$$

with $\beta_j \in \mathbf{Z}$ satisfying $\gcd_{\mathbf{Z}}(\beta_1, \beta_2) = 1$. In particular, $\beta := \beta_1 + \beta_2 i$ is not divisible in $\mathbf{Z}[i]$ by any $p \equiv 3 \bmod 4$ (as it isn't even divisible in $\mathbf{Z}[i]$ by any integer $m > 1$, since $m(u+iv) = mu + imv$ yet $\gcd_{\mathbf{Z}}(\beta_1, \beta_2) = 1$).

Thus, the irreducible factorization of β in the "power" formulation (collecting associate irreducibles into a power of a single irreducible, up to a unit multiple as always) is

$$\beta = u(1+i)^e \prod_j \pi_j^{e_j}$$

with $u \in \mathbf{Z}[i]^\times$, $e \geq 0$, and non-associate irreducibles π_j satisfying $N(\pi_j) = p_j \equiv 1 \bmod 4$ (and $e_j \geq 1$). Note that for any π_j that occurs, its (non-associate!) conjugate $\overline{\pi}_j$ does not occur as a factor, since otherwise $\pi_j\overline{\pi}_j = p_j$ would be a factor of β in $\mathbf{Z}[i]$, contradicting that $\gcd_{\mathbf{Z}}(\beta_1, \beta_2) = 1$. Thus, there are *no repetitions* among the p_j's, so the formula

$$N(\beta) = 2^e \prod_j p_j^{e_j}$$

is the *prime factorization* of $N(\beta)$ in $\mathbf{Z}^+$ (and is even precisely when $e > 0$).

To summarize, after extracting the factor n (which we factor into primes in $\mathbf{Z}$, and then turn into the irreducible factorization of n in $\mathbf{Z}[i]$ using our knowledge of how all primes of $\mathbf{Z}^+$ factor into irreducibles in $\mathbf{Z}[i]$), we can read off the irreducible factorization of $\beta = \alpha/n$ (up to units) from the prime factorization of $N(\beta)$ in $\mathbf{Z}^+$. The only caveat is that for each $p \equiv 1 \bmod 4$ that occurs in $N(\beta)$, we have to figure out which among the two (conjugate, but *non-associate*) irreducible factors of p in $\mathbf{Z}[i]$ is the one that actually divides β (and its multiplicity in β is *the same* as that of p in $N(\beta)$). This latter task is achieved via the old trick of "rationalizing the denominator": we pick an irreducible factor π of p (by writing p as a sum of two squares) and compute

$$\frac{\beta}{\pi} = \frac{\beta\overline{\pi}}{\pi\overline{\pi}} = \frac{\beta\overline{\pi}}{p} \in \mathbf{Q}[i].$$

If this belongs to $\mathbf{Z}[i]$ then we chose well, and if it is not in $\mathbf{Z}[i]$ then (as in the Indiana Jones movie) we chose poorly (and $\overline{\pi}$ is what occurs in β).

It should now be clear that the UFD property of $\mathbf{Z}[i]$ is a very powerful fact. One may then wonder:

Question 3.4. For which squarefree $d \in \mathbf{Z} - \{0,1\}$ is $\mathbf{Z}[\sqrt{d}]$ a UFD?

For $d > 1$, this is known to hold for $d = 2, 3, 6, 7, 11, 13, 14, \ldots$ (though the "Euclidean domain" method for proving such a UFD property quickly runs out of steam, and this condition must be attacked in an entirely different manner via the notion of "class group" that we will study later). It is widely believed that this holds for infinitely many d, but this remains unsolved.

For $d < 0$, there are only finitely many d for which the UFD property holds. This was also conjectured in a precise form by Gauss (the *class number* 1 *problem*), and it was first solved in 1952 by a German high school math teacher named Kurt Heegner. His paper had some errors and was written in a form that was hard to read, so (since he was moreover a complete unknown) his paper was disregarded. In the late 1960's the problem was solved (again) by some professional number theorists. A bit later Heegner's paper was re-examined and it was realized that his errors were fairly minor and that in effect he really had solved the problem. Unfortunately Heegner had died by that time, but his name lives on through constructions in the arithmetic theory of elliptic curves (Heegner points, etc.).

Question 3.5. Is $\mathbf{Z}[\sqrt{d}]$ the "right" ring to focus on in $\mathbf{Q}[\sqrt{d}]$?

In general, the answer is "no". We have already seen this for $\mathbf{Q}(\sqrt{-3}) = \mathbf{Q}(\zeta)$ for $\zeta = (-1+\sqrt{-3})/2$ a primitive cube root of 1: $\mathbf{Z}[\sqrt{-3}]$ is not a UFD, but in HW2 you'll show $\mathbf{Z}[\zeta]$ is a UFD. Since $\zeta^2 = 1 + \zeta$ we have $\mathbf{Z}[\zeta] = \mathbf{Z} \oplus \mathbf{Z}\zeta$ and this contains $\mathbf{Z}[\sqrt{-3}]$ with index 2.

Example 3.6. For $K = \mathbf{Q}(\sqrt{5})$, it turns out that $\mathbf{Z}[\sqrt{5}]$ is not a UFD but for the "Golden Ratio" $\Phi := (1+\sqrt{5})/2$ that satisfies $\Phi^2 = \Phi + 1$ we have $\mathbf{Z}[\Phi] = \mathbf{Z} \oplus \mathbf{Z}\Phi$ and later we will show $\mathbf{Z}[\Phi]$ is a UFD (containing $\mathbf{Z}[\sqrt{5}]$ with index 2).

Consider the ring $\mathbf{Z}[i/2] = \{\alpha/2^n, \mid \alpha \in \mathbf{Z}[i], n \geq 0\}$. In this domain $1+i$ is a unit but the other irreducibles of $\mathbf{Z}[i]$ remain irreducible (since $1+i$ is the only irreducible factor of 2 up to $\mathbf{Z}[i]^\times$-multiple) and that accounts for all irreducibles in $\mathbf{Z}[i/2]$ which is moreover a UFD. This is very analogous to $\mathbf{Z}[1/7]$ as a UFD (with units $\pm 7^{\mathbf{Z}}$).

Although $\mathbf{Z}[i/2]$ is a UFD, there is a sense in which this is "worse" than $\mathbf{Z}[i]$: it is not *finitely generated* as a $\mathbf{Z}$-module. Indeed, the UFD-property of $\mathbf{Z}[i]$ allows us to write fractions $\alpha/2^n$ in "reduced form" (up to possibly a single factor of $1+i$ in the numerator, since 2 is a unit multiple of $(1+i)^2$) and so in this way we see that there are such fractions whose "reduced form" involves an exponent n as large as we wish. That is an obstruction to being finitely generated as a $\mathbf{Z}$-module: for any *finite* collection of such fractions $\alpha_j/2^{n_j}$, a $\mathbf{Z}$-linear combination will never yield a "reduced form" fraction $\alpha/2^n$ with $n > \max_j n_j$. (A toy analogue is that $\mathbf{Z}[1/7]$ is not finitely generated as a $\mathbf{Z}$-module: the $\mathbf{Z}$-linear combinations of any *finite* set of such fractions will never yield $1/7^n$ for arbitrarily large n.)

The failure of $\mathbf{Z}[i/2]$ to be finitely generated as a $\mathbf{Z}$-module will be seen next time to encode the failure of $i/2$ to be an "algebraic integer" in the sense of the following definition:

Definition 3.7. A *number field* is a finite-degree extension field K over $\mathbf{Q}$. An *algebraic integer* is an element α of a number field K such that $f(\alpha) = 0$ for some monic $f \in \mathbf{Z}[X]$ (i.e., the leading coefficient is 1).

We know from Galois theory that any element of a number field is a root of a monic polynomial over $\mathbf{Q}$, and we can clear denominators to make that a polynomial with coefficients in $\mathbf{Z}$ at the cost of losing monicity. The monicity condition on $f \in \mathbf{Z}[X]$ is the really crucial feature of the definition of an algebraic integer, as we will see next time.

Beware that in the definition of being an algebraic integer, it is not required that f is the minimal polynomial of α over $\mathbf{Q}$; i.e., we do not demand that the (monic) minimal polynomial has coefficients in $\mathbf{Z}$. Fortunately, we will see that this latter property *is* equivalent to a given α algebraic over $\mathbf{Q}$ being an algebraic integer. However, to set up a robust general theory it would be very bad to make that concrete condition be the initial definition.

Example 3.8. Inside the number field $\mathbf{Q}$, the algebraic integers are precisely the elements of $\mathbf{Z}$. Indeed, this is a consequence of the rational root theorem from high school. Recall that the rational root theorem says that if $f = a_d X^d + \cdots + a_1 X + a_0 \in \mathbf{Z}[X]$ with $a_d \neq 0$ has a root $q \in \mathbf{Q}$ that is written as m/n with $m \in \mathbf{Z}$ and $n \in \mathbf{Z}^+$ satisfying $\gcd(m,n) = 1$ then m divides the constant term a_0 and n divides the leading coefficient a_d. Hence, if $a_d = 1$ (i.e., f is monic) then $n = 1$ and so $q = m/1 \in \mathbf{Z}$, as desired.

4. Integrality

Definition 4.1. For a number field K, say $\alpha \in K$ is an *algebraic integer* if $f(\alpha) = 0$ for a monic $f \in \mathbf{Z}[x]$.

Example 4.2. Consider $K = \mathbf{Q}(\sqrt{d})$ for $d \in \mathbf{Z} - \{0, 1\}$ squarefree. Let

$$\alpha = u + v\sqrt{d}$$

with $u, v \in \mathbf{Z}$ and $v \neq 0$.

This has minimal polynomial over $\mathbf{Q}$ equal to $x^2 - 2ux + (u^2 - dv^2)$.

You may hear people use the phrase "rational integer" meaning the algebraic integers in $\mathbf{Q}$ (which is just $\mathbf{Z}$, by the rational root theorem).

Example 4.3. The element

$$\frac{1 + \sqrt{5}}{2}$$

is an algebraic integer as it is a root of $x^2 - x - 1$. Similarly,

$$\frac{-1 + \sqrt{-3}}{2}$$

is an algebraic integer as it is a root of $x^2 + x + 1$.

Remark 4.4. Recall that we only require α to be a root of some such f, and not necessarily that f is the minimal polynomial of α over $\mathbf{Q}$. However, we'll show later, in Homework 2, that $\alpha \in K$ is an algebraic integer if and only if its minimal polynomial over $\mathbf{Q}$ has $\mathbf{Z}$-coefficients.

Here are a couple basic questions about algebraic integers we want to answer.

Question 4.5. If $\alpha, \beta \in K$ are algebraic integers, is $\alpha + \beta$ an algebraic integer? Is $\alpha\beta$ an algebraic integer?

Question 4.6. For $K/\mathbf{Q}$ a quadratic extension, how do we find all algebraic integers in K?

We'll come to the second question next time, but for today we'll focus on the first question. The issue is that it is not easy to express the minimal polynomial of $\alpha + \beta$ or $\alpha\beta$ in terms of those of α and β. Hence, to answer Question 4.5 we need a more robust way to think about integrality.

Recall the following result from field theory:

Proposition 4.7. *Let L/k be a field extension. If $\alpha, \beta \in L$ are algebraic over k then $\alpha + \beta$ and $\alpha\beta$ are also algebraic over k.*

Proof. Consider the subring

$$k[\alpha, \beta] := \{\sum c_{ij}\alpha^i\beta^j \mid c_{ij} \in k\} \subset L$$

that lies between k and L. Note that since $k[\alpha] = k(\alpha)$ (a field!) by algebraicity of α over k, and likewise for β, we have

$$k \subset k(\alpha), k(\beta) \subset k[\alpha, \beta] \subset L.$$

The crucial point is that $k[\alpha, \beta]$ is finite-dimensional over k since in the expressions $\sum c_{ij}\alpha^i\beta^j$ we can always rewrite this using only $i < d$ and $j < d'$ where d and d' are the respective degrees of the minimal polynomials of α and β over k.

It follows that if $N = \dim_k k[\alpha, \beta]$ then for any $\gamma \in k[\alpha, \beta]$ there must be a nontrivial k-linear dependence relation among the $N + 1$ elements

$$\{1, \gamma, \gamma^2, \dots, \gamma^N\}.$$

Such a relation cannot only involve 1, so it exhibits γ as being algebraic over k. By taking $\gamma = \alpha + \beta$ and $\gamma = \alpha\beta$ (both lie in $k[\alpha, \beta]$!) we thereby conclude each of these is algebraic over k. □

Warning 4.8. The proof of Proposition 4.7 crucially uses linear algebra and in particular the notion of dimension over the field k. Therefore, it doesn't apply to create monic relations over $\mathbf{Z}$, hence doesn't apply as written to our questions of integrality. Nonetheless, we will adapt some of the ideas in that proof.

Definition 4.9. Let $\mathscr{O}_K$ denote the set of algebraic integers in K.

We now have the following goal:

Theorem 4.10. *Let K be a number field. Then:*

(1) $\mathscr{O}_K$ is a subring,
(2) $\mathscr{O}_K$ is finitely generated as a $\mathbf{Z}$-module.

Today, we'll show that $\mathscr{O}_K$ is a subring of K and we'll prove the second part later.

Warning 4.11. Beware that although we can write $K = \mathbf{Q}(\alpha)$ (by the primitive element theorem), there is no "primitive element theorem" for rings: $\mathscr{O}_K$ need not admit a description as $\mathbf{Z}[\beta]$ for some $\beta \in \mathscr{O}_K$. In fact, later we'll see that for any $n > 1$ there exist K such that $\mathscr{O}_K$ is not even generated as a ring over $\mathbf{Z}$ by n elements. The upshot is that rings of integers are rather more subtle objects than their fraction fields from an algebraic point of view, and in particular the techniques we require to analyze their structure will be more indirect than in the theory of field extensions.

Example 4.12 (Dedekind)**.** If one takes

$$K = \mathbf{Q}(\theta)$$

for θ a root of the irreducible

$$x^3 + x^2 - 2x + 8 \in \mathbf{Q}[x]$$

then $\mathscr{O}_K$ is not monogenic over $\mathbf{Z}$.

We will see why this $\mathscr{O}_K$ cannot be generated by a single element later, after we have learned some basic facts concerning Dedekind domains.

Remark 4.13. For quadratic and cyclotomic fields K a convenient miracle will happen: there will be an explicit single generator for $\mathscr{O}_K$ over $\mathbf{Z}$ as a ring. In general we cannot hope for such a miracle.

Observe that if α is an algebraic integer then

$$\mathbf{Z}[\alpha] = \{\sum_j c_j\alpha^j \mid c_j \in \mathbf{Z}\}$$

(finite sums) is finitely generated as a $\mathbf{Z}$-module. Indeed, since

$$\alpha^d + r_{d-1}\alpha^{d-1} + \cdots + r_1\alpha + r_0 = 0$$

for some $r_j \in \mathbf{Z}$, we have

$$\mathbf{Z}[\alpha] = \mathbf{Z} + \mathbf{Z}\alpha + \cdots + \mathbf{Z}\alpha^{d-1}$$

by repeatedly using the degree-d *monic* relation for α over $\mathbf{Z}$.

In contrast, the ring $\mathbf{Z}[2/3]$ is not finitely generated as a $\mathbf{Z}$-module (consider powers of 3 in the denominator, akin to what we saw last time).

The module-finiteness of $\mathbf{Z}[\alpha]$ will turn out not only to be a consequence of α being an algebraic integer, but will even *imply* it is so. The link to module-finiteness of certain rings will be the key to explaining why sums and products of algebraic integers are algebraic integers. To explain this, we shall work more generally, not just in the context of number fields:

Definition 4.14. For an injective map of rings $A \hookrightarrow B$, we say $b \in B$ is *integral over A* if $f(b) = 0$ for some monic $f \in A[x]$.

Example 4.15. Consider the case $\mathbf{Z} \hookrightarrow K$. The elements of K integral over $\mathbf{Z}$ are precisely the elements of $\mathscr{O}_K$.

Example 4.16. If A and B are fields, the elements of B integral over A are precisely the elements of B that are algebraic over A.

To see this, note that if we're working over a field we can always divide by the (non-zero!) leading coefficient of a nonzero polynomial to make it monic. Therefore, when A is a field, the elements algebraic over A are integral over A.

Definition 4.17. For $A \hookrightarrow B$ an extension (i.e., injective map) of rings, the *integral closure* of A in B is defined to be

$$\{b \in B \mid b \text{ is integral over } A\}.$$

We do not yet know that the integral closure of A in B is a subring of B, though we shall soon see this. Note that if $b \in B$ is integral over A then

$$A[b] := \{\sum_j a_j b^j\} \subset B$$

(finite sums) is a subring containing A that is finitely generated as an A-module: it is generated by $\{1, b, b^2, \dots, b^{d-1}\}$ where $f(b) = 0$ for a monic $f \in A[x]$ with degree $d > 0$.

If A is a domain then the integral closure $\widetilde{A}$ of A is by definition the integral closure of A in its own fraction field.

Example 4.18. Let $A = \mathbf{Z}[\sqrt{-3}] \subset B = \mathbf{Q}(\sqrt{-3})$. We have that $\zeta_3 \in B$ lies outside A but is even integral over $\mathbf{Z} \subset A$. Thus, A is not its own integral closure. (In this case, we say A is *not integrally closed*.) Later, we'll show $\widetilde{A} = \mathbf{Z}[\zeta_3]$.

We next want to show that $A[b]$ is finitely generated as an A-module if and only if b is integral over A. That is, we want to show this integrality property is precisely captured by module-finiteness of certain subrings. This is encoded in the following general result:

Theorem 4.19. *For any $b \in B$, we have that b is integral over A if and only if $b \in R \subset B$ for a subring $R \supset A$ that is finitely generated as an A-module.*

Using this theorem, we can easily prove Theorem 4.10(i) as follows (and then we will prove Theorem 4.19):

Proof of Theorem 4.10(i). Say we have $b, b' \in B$ which are both integral over A. Then $b + b'$ and $b \cdot b'$ both belong to the subring $R := A[b, b'] \subset B$, where

$$A[b, b'] := \{\sum a_{ij} b^i b'^j\}$$

(finite sums). The key point is that R is a finitely generated A-module since it is spanned as an A-module by $b^i b'^j$ for $i < d, j < d'$, where d and d' are the degrees of respective monic relations over A for b and b'. Hence, it follows that all elements of R are integral over A, and in particular $b + b'$ and bb' are integral over R. □

We now give:

Proof of Theorem 4.19. The implication "⇒" is easy: take R to be $A[b]$.

Now consider the converse. By hypothesis $R = \sum_{i=1}^{N} Ar_i$ with $r_1, \ldots, r_N \in R$. We want to show that any element $b \in R$ is integral over A. Informally, the idea is that although there is no linear independence condition on the r_i's over A, so A-linear maps $R \to R$ aren't "the same" as $N \times N$ matrices with entries in A, we will nonetheless study polynomial relations for b over A by instead considering polynomial relations for the corresponding A-linear multiplication operator $m_b : R \to R$ defined by $r \mapsto br$ and apply a "Cayley-Hamilton Theorem" to a matrix "computing" m_b.

To be precise, we can write each br_j (as for any element of R) in the form

$$br_j = \sum_i a_{ij} r_i$$

for some $a_{ij} \in A$ (that need not be unique, but we don't care). In the notation of matrices, this system of equations is expressed as the equality:

$$\begin{pmatrix} b & \cdots & 0 \\ \vdots & \ddots & \cdots \\ 0 & \cdots & b \end{pmatrix} \begin{pmatrix} r_1 \\ \vdots \\ r_N \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \cdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix} \begin{pmatrix} r_1 \\ \vdots \\ r_N \end{pmatrix}$$

Subtracting the right side from the left side, we obtain

$$\begin{pmatrix} b - a_{11} & \cdots & -a_{1N} \\ \vdots & \ddots & \cdots \\ -a_{N1} & \cdots & b - a_{NN} \end{pmatrix} \begin{pmatrix} r_1 \\ \vdots \\ r_N \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Let M be the matrix on the left, so we can form its adjugate matrix

$$M^{\mathrm{adj}} := (m'_{ij})$$

with $m'_{ij} = (-1)^{i+j} \det M^{ij}$, where M^{ij} denotes the ij-minor of M, gotten from M by removing row i and column j. Cramer's formula is the universal matrix identity:

$$M^{\mathrm{adj}} M = \begin{pmatrix} \det M & \cdots & 0 \\ \vdots & \ddots & \cdots \\ 0 & \cdots & \det M \end{pmatrix}$$

If you are uncomfortable with matrices with entries in general commutative rings, feel free to consider only the case when B and A are domains, so the preceding calculations can be viewed with entries in the field $\mathrm{Frac}(B)$, putting us in the more familiar setting of linear algebra over fields where Cramer's Formula is a known general matrix identity.

Multiplying by M^{adj} thereby gives

$$\begin{pmatrix} \det M & \cdots & 0 \\ \vdots & \ddots & \cdots \\ 0 & \cdots & \det M \end{pmatrix} \begin{pmatrix} r_1 \\ \vdots \\ r_N \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

where the matrix on the left is the diagonal matrix with entry $\det M$ along the diagonal. This says $\det M \cdot r_j = 0$ in R for all j. But every $r \in R$ is an A-linear combination of the r_j's, so $\det M \cdot r = 0$ for all $r \in R$. Taking $r = 1$, we get $\det M = 0$ in $R \subset B$.

But if we go back to how M was defined, and more specifically how b appears in M, the vanishing of $\det M$ expresses exactly a monic relation for b over A, as we now explain. Expanding the determinant as a signed sum of products of N entries at a time (one from each row and column), there is only *one* such product that involves N occurrences of b, namely the product $\prod_i (b - a_{ii})$ along the diagonal; all others involves at most $N - 1$ occurrences of b. Thus, when we expand out $\det M$ we have

$$\det M = \prod_i (b - a_{ii}) + g(b)$$

where $g \in A[X]$ (possibly not monic) has degree at most $N - 1$. But

$$\prod_i (X - a_{ii}) = X^N + h(X)$$

with $h \in A[X]$ (possibly not monic) of degree at most $N - 1$, so $\det M = f(b)$ for $f := X^N + h(X) + g(X) \in A[X]$ monic of degree N. Thus, the relation $f(b) = 0$ due to the vanishing of $\det M$ implies b is integral over A. (Explicitly, f is just the "characteristic polynomial" of (a_{ij}).) □

Remark 4.20. The method used in the argument is called the "determinant trick" and the argument applies with A an arbitrary commutative ring (once one knows that Cramer's Formula actually holds for matrices with entries in any such ring), far beyond the setting of domains or subrings of number fields. This is very useful, and is explained more fully in the handout "Generalized Cayley-Hamilton and Integrality".

Here is a related question:

Question 4.21. If A is a domain, is its integral closure in $\mathrm{Frac}(A)$ actually finitely generated as an A-module?

This is often true (non-obviously) but not always true. It is a subtle problem in commutative algebra.

5. Finiteness properties of $\mathscr{O}_K$

Before we take up the general module-finiteness for $\mathscr{O}_K$ over $\mathbf{Z}$, we compute $\mathscr{O}_K$ in a basic important case:

Theorem 5.1. *Let $K = \mathbf{Q}(\sqrt{d})$ for $d \in \mathbf{Z} - \{0, 1\}$ squarefree. Then,*

$$\mathscr{O}_K = \begin{cases} \mathbf{Z}[\sqrt{d}] & \text{if } d \equiv 2, 3 \bmod 4, \\ \mathbf{Z}\left[\frac{1+\sqrt{d}}{2}\right] & \text{if } d \equiv 1 \bmod 4. \end{cases}$$

Remark 5.2. One useful way to remember the 1 mod 4 case is separate is by remembering the case $d = -3$. In this case, for $d = -3$, we can remember the result as $\mathscr{O}_K = \mathbf{Z}[\zeta_3]$ with $\zeta_3 = (-1 + \sqrt{-3})/2$, and since $(1 + \sqrt{-3})/2 = 1 + \zeta_3$ we clearly have $\mathbf{Z}[\zeta_3] = \mathbf{Z}[(1 + \sqrt{-3})/2]$.

Remark 5.3. Note that in the setting of Theorem 5.1, $\mathscr{O}_K$ is free of rank 2 over $\mathbf{Z}$ since

$$\mathbf{Z} \oplus \mathbf{Z} \cdot \alpha = \mathbf{Z}[\alpha]$$

for α any root of a monic $f \in \mathbf{Z}[x]$ of degree 2 that is irreducible over $\mathbf{Q}$, such as f equal to either $x^2 - d$ or (for $d \equiv 1 \bmod 4$) $x^2 - x + (1-d)/4$.

Warning 5.4. Note that replacing d by $n^2 d$ for $n \in \mathbf{Z}_{>1}$ has no effect on $\mathbf{Q}(\sqrt{d})$ in the sense that

$$\mathbf{Q}(\sqrt{n^2 d}) = \mathbf{Q}(\sqrt{d}),$$

but at the level of the ring of integers there is a huge effect; e.g.,

$$\mathbf{Z}[\sqrt{28}] = \mathbf{Z} \oplus 2\mathbf{Z}\sqrt{7} \neq \mathbf{Z}[\sqrt{7}]$$

since $\sqrt{7} \notin \mathbf{Z}[\sqrt{28}]$. So for the asserted formula for $\mathscr{O}_K$ for quadratic fields K, it is really essential that we are taking d to be a squarefree integer (as we may certainly always arrange to be the case when describing the field K) and not merely a non-square integer.

Proof of Theorem 5.1. By inspection $\sqrt{d}$ and $\frac{1+\sqrt{d}}{2}$ are integral in the respective cases, so $\mathbf{Z}[\sqrt{d}] \subset \mathscr{O}_K$ when $d \equiv 2, 3 \bmod 4$ and $\mathbf{Z}\left[\frac{1+\sqrt{d}}{2}\right] \subset \mathscr{O}_K$ when $d \equiv 1 \bmod 4$ since $\mathscr{O}_K$ is a subring of K.

It remains to prove the reverse containment in all cases. Consider a general element

$$\alpha := x + y\sqrt{d} \in \mathscr{O}_K$$

with $x, y \in \mathbf{Q}$. Denote the unique nontrivial automorphism of K over $\mathbf{Q}$ as

$$z \mapsto \overline{z};$$

this satisfies $\sqrt{d} \mapsto -\sqrt{d}$. This carries $\mathscr{O}_K$ isomorphically to $\mathscr{O}_K$, since more generally for any inclusion of rings $A \to B$ any A-automorphism of B as a ring sends A-integral elements of B to A-integral elements of B because applying such an automorphism to a monic polynomial relation over A preserves the monic polynomial relation.

We conclude that

$$\overline{\alpha} = x - y\sqrt{d} \in \mathscr{O}_K,$$

so

$$\alpha + \overline{\alpha}, \alpha\overline{\alpha} \in \mathscr{O}_K \cap \mathbf{Q} = \mathbf{Z}.$$

This says

$$2x, x^2 - dy^2 \in \mathbf{Z}.$$

Hence, either $x \in \mathbf{Z}$ or $x = \frac{n}{2}$ for some odd $n \in \mathbf{Z}$. There are now two cases.

First, suppose $x \in \mathbf{Z}$. Since $x^2 - dy^2 \in \mathbf{Z}$, we have $dy^2 \in \mathbf{Z}$. But d is squarefree, so this forces $y \in \mathbf{Z}$ (by considering the possibility of d cancelling the entire denominator of y^2). Hence, $\alpha \in \mathbf{Z}[\sqrt{d}]$. Note that for $d \equiv 1 \bmod 4$ we have $\sqrt{d} \in \mathbf{Z}[(1+\sqrt{d})/2]$, so $\mathbf{Z}[\sqrt{d}] \subset \mathbf{Z}[(1+\sqrt{d})/2]$ in such cases too.

Next, suppose $x \notin \mathbf{Z}$. This is the more difficult case. Since $2x \in \mathbf{Z}$, we have $x = \frac{n}{2}$ for some odd $n \in \mathbf{Z}$, so

$$\frac{n^2}{4} - dy^2 \in \mathbf{Z};$$

this forces $y \notin \mathbf{Z}$. Clearly $n^2 - 4dy^2 \in 4\mathbf{Z}$, so $4dy^2 \in \mathbf{Z}$. This implies $y = \frac{m}{2}$ for odd m, where we are again using the fact that d is squarefree.

Thus, $n^2 - dm^2 \equiv 0 \bmod 4$. But m, n are odd, so $n^2, m^2 \equiv 1 \mod 4$. Hence, the relation $n^2 - dm^2 \equiv 0 \bmod 4$ says $1 - d \equiv 0 \bmod 4$, or equivalently $d \equiv 1 \bmod 4$. Moreover,

$$\begin{aligned}\alpha &= \frac{n}{2} + \frac{m}{2}\sqrt{d} \\ &= \frac{1+\sqrt{d}}{2} + (\frac{n-1}{2} + \frac{m-1}{2}\sqrt{d}) \in \mathbf{Z}\left[\frac{1+\sqrt{d}}{2}\right],\end{aligned}$$

using that

$$\frac{n-1}{2} + \frac{m-1}{2}\sqrt{d} \in \mathbf{Z}[\sqrt{d}],$$

since n and m are odd. □

Remark 5.5. Beyond degree 2, one has to grapple with coefficients beyond the rather concrete trace and norm in the quadratic case, and it becomes hard to establish "general descriptions". For cubefree $d \in \mathbf{Z} - \{0, 1\}$, the inclusion

$$\mathbf{Z}[d^{1/3}] \subset \mathscr{O}_{\mathbf{Q}(d^{1/3})}$$

can fail to be an equality; e.g., we will see later that this happens when $d = 10$.

There is another important case of a collection of number fields for which the ring of integers admits a clean monogenic description. Later we will prove:

Theorem 5.6 (Kummer). *For $K = \mathbf{Q}(\zeta_m) = \mathrm{split}_{\mathbf{Q}}(x^m - 1)$, we have*

$$\mathscr{O}_K = \mathbf{Z}[\zeta_m].$$

This lies much deeper than the case of quadratic fields (though for $m = 2, 3$ it recovers our descriptions of the ring of integers of $\mathbf{Q}(\sqrt{d})$ for $d = -1, -3$).

Our remaining goal for today is to prove:

Theorem 5.7. *For a number field K, the ring $\mathscr{O}_K$ is finitely generated as a $\mathbf{Z}$-module. It is free of rank $n := [K : \mathbf{Q}]$.*

Before giving the proof, we make some preliminary observations. Writing $K = \mathbf{Q}e_1 \oplus \cdots \oplus \mathbf{Q}e_n$ upon choosing a $\mathbf{Q}$-basis of K, any $\alpha \in K$ can be written as

$$\alpha = \sum_i c_i e_i$$

for $c_i \in \mathbf{Q}$. Thus, if $\mathscr{O}_K$ is finitely generated as a $\mathbf{Z}$-module then all $\mathbf{Q}$-coefficients of all elements of $\mathscr{O}_K$ with respect to the basis $\{e_i\}$ admit a single common denominator (since that holds for any finite subset of K, and then for all $\mathbf{Z}$-linear combinations of such a finite subset).

Further, the converse also holds. Indeed, if

$$\mathscr{O}_K \subset \sum \frac{1}{d}\mathbf{Z}e_i$$

for some $d \in \mathbf{Z} - \{0\}$ then $\mathscr{O}_K$ is a submodule of a finitely generated free $\mathbf{Z}$-module (generated by the elements e_i/d), and all submodules of a finite free (meaning the span of a finite set of linearly independent elements) $\mathbf{Z}$-module are again finite free due to the structure theorem for modules over a PID (as explained in the early handout "Modules over a PID"). Hence, it would follow that $\mathscr{O}_K$ is a finite free $\mathbf{Z}$-module.

Inspired by this reasoning, the method of the proof of Theorem 5.7 will involve finding a common denominator for $\mathbf{Q}$-coefficients of all elements of $\mathscr{O}_K$ with respect to a suitable $\mathbf{Q}$-basis $\{e_i\}$. The proof will also show that $\mathscr{O}_K$ has rank n as a free $\mathbf{Z}$-module (but without producing an explicit $\mathbf{Z}$-basis).

As a first step in the proof, we want to scale a choice of $\mathbf{Q}$-basis of K so that all elements of the basis belong to $\mathscr{O}_K$. This will be achieved via the useful:

Lemma 5.8. *For any $\alpha \in K$, we have $b\alpha \in \mathscr{O}_K$ for some $b \in \mathbf{Z}^+$ (perhaps depending on α). In particular,*

$$\alpha = \frac{b\alpha}{b} \in \frac{1}{b}\mathscr{O}_K.$$

Proof. We know α is a root of some

$$f = x^m + c_{m-1}x^{m-1} + \cdots + c_1 X + c_0 \in \mathbf{Q}[x].$$

Say b is a common denominator of all c_j. Multiplying through by b^m gives

$$b^m f(x) = (bx)^m + bc_{m-1}(bx)^{m-1} + \cdots + (b^{m-1}c_1)(bx) + b^m c_0.$$

Note that $b^{m-i}c_i \in \mathbf{Z}$ by construction of b, so $b^m f(x) \in \mathbf{Z}[x]$. More specifically, we have just seen that $b^m f(x) = h(bx)$ for some monic $h \in \mathbf{Z}[x]$. But $h(b\alpha) = b^m f(\alpha) = 0$, so $b\alpha \in \mathscr{O}_K$. □

Proof of Theorem 5.7. By Lemma 5.8, we can replace a $\mathbf{Q}$-basis $e_1, \ldots, e_n$ of K by $Me_1, \ldots, Me_n$ for $M \in \mathbf{Z}^+$ sufficiently divisible so that $e_i \in \mathscr{O}_K$ for all i.

Now consider any $x = \sum_i c_i e_i \in \mathscr{O}_K$ with $c_i \in \mathbf{Q}$. We seek a common denominator d of all c_i such that d is *independent of such x*, as then

$$\mathscr{O}_K \subset \sum \mathbf{Z} \cdot (e_i/d)$$

(so $\mathscr{O}_K$ is free of rank at most n, by the structure theorem for modules over a PID). Moreover, we would also have

$$\oplus_{i=1}^n \mathbf{Z}e_i \subset \mathscr{O}_K \subset \oplus_{i=1}^n \mathbf{Z} \cdot (e_i/d)$$

so comparing ranks throughout gives

$$n \leq \mathrm{rk}_{\mathbf{Z}}\, \mathscr{O}_K \leq n,$$

forcing $\mathrm{rk}_{\mathbf{Z}}\, \mathscr{O}_K = n$ as desired.

The main issue is to find such a uniform denominator d. To achieve this, we introduce a $\mathbf{Q}$-valued "dot-product" on K that is $\mathbf{Z}$-valued on $\mathscr{O}_K$: define

$$\langle x, y \rangle = \mathrm{Tr}_{K/\mathbf{Q}}(xy) \in \mathbf{Q}.$$

The crucial feature of this $\mathbf{Q}$-bilinear pairing $K \times K \to \mathbf{Q}$ is that it carries $\mathscr{O}_K \times \mathscr{O}_K$ into $\mathbf{Z}$, or more specifically that $\mathrm{Tr}_{K/\mathbf{Z}}(\mathscr{O}_K) \subset \mathbf{Z}$. To prove this latter containment, we compute the trace with the help of Galois theory: for a Galois closure $L/\mathbf{Q}$ of K (or really any finite extension L of K that is Galois over $\mathbf{Q}$) we have

$$\mathrm{Tr}_{K/\mathbf{Q}}(\xi) = \sum_{j:K \to L} j(\xi) \in \mathscr{O}_L$$

(since any $\mathbf{Q}$-embedding $K \to L$ certainly carries algebraic integers to algebraic integers). But this trace belongs to $\mathbf{Q}$, and $\mathscr{O}_L \cap \mathbf{Q} = \mathbf{Z}$ (!), so indeed $\mathrm{Tr}_{K/\mathbf{Q}}(\mathscr{O}_K) \subset \mathbf{Z}$ as claimed. In particular, as noted above, for $x, y \in \mathscr{O}_K$ we have $\langle x, y \rangle \in \mathbf{Z}$.

Recall that for an orthonormal basis relative to the usual dot product, one can extract coefficients of a vector relative to that basis via its dot product against the members of such a basis. It is essentially never going to happen that $\{e_i\}$ is orthonormal relative to the trace-pairing we have built, but nonetheless let us consider the pairing of $x \in \mathscr{O}_K$ against each e_j and relate such values to the actual coefficients $c_i = c_i(x)$ that depend on x.

By design we have $e_j \in \mathscr{O}_K$ for all j, so

$$\mathbf{Z} \ni \langle x, e_j \rangle = \sum_i c_i \langle e_i, e_j \rangle$$

for all j. Thus,

$$\mathbf{Z}^n \ni \begin{pmatrix} \langle x, e_1 \rangle \\ \vdots \\ \langle x, e_n \rangle \end{pmatrix} = \begin{pmatrix} \langle e_1, e_1 \rangle & \cdots & \langle e_1, e_n \rangle \\ \vdots & \ddots & \vdots \\ \langle e_n, e_1 \rangle & \cdots & \langle e_n, e_n \rangle \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

where $c_i = c_i(x) \in \mathbf{Q}$. Note that the matrix $M = (\langle e_i, e_j \rangle)$ is a symmetric matrix with integer entries that is *independent of* x. We can compute the rational vector of c_i's in terms of the integer vector of $\langle x, e_j \rangle$'s by inverting M provided that M is actually invertible. Fortunately, we have:

Lemma 5.9. *The $n \times n$ matrix $M = (\langle e_i, e_j \rangle)$ is invertible. In particular, the integer* $\det M \in \mathbf{Z}$ *is nonzero.*

The significance of such determinant constructions will be understood later when we take up the finer structure theory of rings of integers.

Proof. As a preliminary step, we note that the invertibility or not is independent of the choice of $\mathbf{Q}$-basis for reasons explained in the handout "Norm and Trace", where the determinant is what we called the "discriminant" with respect to the $\{e_i\}$. We also saw there that under a change of basis this determinant changes by a nonzero square multiplier.

Hence, it suffices to prove the non-vanishing of such a determinant with $\{e_i\}$ replaced with any single $\mathbf{Q}$-basis of K. But for a separable field extension k'/k of finite degree (such as $K/\mathbf{Q}$) it was seen in the handout "Norm and Trace" via a van der Monde determinant that the determinant relative to a power basis $\{1, \alpha, \dots, \alpha^{m-1}\}$ for a primitive element α of k'/k is equal to $\pm \mathrm{N}_{k'/k}(f'(\alpha))$ where $f \in k[x]$ is the minimal polynomial of α over k. This is nonzero because $f'(\alpha) \in k'^{\times}$ due to separability of k'/k. □

The upshot is that there is a matrix M with entries in $\mathbf{Z}$ and independent of x such that

$$\begin{pmatrix} c_1(x) \\ \vdots \\ c_n(x) \end{pmatrix} = M^{-1}(\mathbf{Z}^n) \in (1/d)M^{\mathrm{adj}}(\mathbf{Z}^n) \subset (1/d)\mathbf{Z}^n$$

for the adjugate matrix M^{adj} with entries in $\mathbf{Z}$ and $d := \det(M) \in \mathbf{Z} - \{0\}$. This shows that d is a universal denominator for $\mathscr{O}_K$. □

6. Irreducible elements and prime ideals

Consider $K = \mathbf{Q}(\sqrt{-3})$. We have a finite-index inclusion $\mathbf{Z}[\sqrt{-3}] \subset \mathscr{O}_K = \mathbf{Z}[\zeta_3]$ where $\mathbf{Z}[\sqrt{-3}]$ is not a UFD but $\mathbf{Z}[\zeta_3]$ is a UFD. More interestingly, for $F = \mathbf{Q}(\sqrt{-5})$ even the ring of integers $\mathscr{O}_F = \mathbf{Z}[\sqrt{-5}]$ (whose unit group is $\{\pm 1\}$) fails to be a UFD. Indeed, we claim that the two factorizations

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

of 6 have irreducible factors; since these are not related through unit multiplications, this would then show that the UFD property does not hold.

To show the indicated factors of 6 are irreducible in $\mathbf{Z}[\sqrt{-5}]$, suppose $2 = \alpha\beta$ with α, β non-units in $\mathbf{Z}[\sqrt{-5}]$. Then, $4 = N\alpha N\beta$ with each norm a *positive* integer distinct from 1, so $N\alpha = 2$; this says

$$u^2 + 5v^2 = 2$$

where $\alpha = u + v\sqrt{-5}$ with $u, v \in \mathbf{Z}$, but obviously no such integers u and v exist. A similar proof works for showing 3 is irreducible in $\mathbf{Z}[\sqrt{-5}]$.

Finally, we show irreducibility of $1 + \sqrt{-5}$ (and then $1 - \sqrt{-5}$ is handled similarly, or follows by applying the conjugation automorphism since an automorphism of a domain carries irreducibles to irreducibles). Suppose

$$1 + \sqrt{-5} = \alpha\beta$$

with non-units $\alpha, \beta \in \mathbf{Z}[\sqrt{-5}]$. Then taking norms of both sides gives

$$6 = N\alpha \cdot N\beta.$$

One of these norms must equal 2, and the other must equal 3, but we have already seen that neither 2 nor 3 are norms from $\mathbf{Z}[\sqrt{-5}]$. This completes the proof that $\mathbf{Z}[\sqrt{-5}]$ is not a UFD.

Remark 6.1. The property of $\mathbf{Z}[\sqrt{-5}]$ not being a UFD is "fake" in the sense that it will be explained by a variant of the following simple calculation in $\mathbf{Z}$: we have two distinct factorizations of 210 in $\mathbf{Z}$ given by

$$14 \cdot 15 = 210 = 6 \cdot 35,$$

but these are related to each other via two separate ways of grouping together pairs from the full factorization of 210 into prime factors, namely

$$(2 \cdot 7)(3 \cdot 5) = (2 \cdot 3)(5 \cdot 7).$$

By working with ideals rather than elements in $\mathbf{Z}[\sqrt{-5}]$, we will be able to factor certain principal ideals of $\mathbf{Z}[\sqrt{-5}]$ into a product of *non-principal* prime ideals (a notion to be defined shortly), and by grouping together such prime ideal factors in different ways we will recover the two different factorizations of 6 into pairs of irreducible elements.

We now review the notion of an ideal, and some operations on ideals.

Definition 6.2. For a commutative ring R, an *ideal* I of R is an R-submodule of R.

You may have previously encountered other ways of defining the concept of an ideal (e.g., an additive subgroup of R carried into itself under multiplication by any element of R), but if you think it through you'll see it is really the same thing as the above definition.

For two ideals $I, J \subset R$ we define their *sum* to be

$$I + J := \{x + y \mid x \in I, y \in J\}.$$

Exercise 6.3. Verify this is indeed an ideal.

We define the *product* ideal

$$IJ := \{\sum x_i y_i \mid x_i \in I, y_i \in J\}$$

(using finite collections of x_i's and y_i's). It is also left as an exercise to check this is an ideal.

Definition 6.4. An ideal is *principal* if it is of the form $(r) := \{sr \mid s \in R\}$ for some $r \in R$.

Example 6.5. Suppose $I = (r)$ and $J = (r')$ are principal. Then, $IJ = (rr')$ is also principal but

$$I + J = \{ar + br' \mid a, b \in R\}$$

is rarely principal (we will see many such examples, including some later today). In the special case that R is a PID we have $I + J = (\gcd(r, r'))$; this is very distant from $(r + r')$!

We list a few important exercises that one should check follow directly from the definitions.

Exercise 6.6. For ideals $I, J_1, \ldots, J_m$ of R, show

$$I(J_1 + \cdots + J_m) = IJ_1 + \cdots + IJ_m.$$

Exercise 6.7. For ideals I_1, I_2, I_3 of R, show

$$I_1(I_2 \cdot I_3) = (I_1 \cdot I_2)I_3.$$

Exercise 6.8. Verify $(1) = R$. (We call this the *unit ideal*; obviously $(1) \cdot J = J$ for any ideal J.)

Define the notation

$$(r_1, \ldots, r_n) := \left\{ \sum_{i=1}^{n} x_i r_i \mid x_i \in R \right\}.$$

It is easy to check (do it!) that $(r_1, \ldots, r_n)(r'_1, \ldots, r'_m)$ is the ideal generated by all products $r_i r'_j$.

Example 6.9. Note that for $a, b \in R$ with R a ring, we have

$$\begin{aligned} a \mid b &\iff b = ax \\ &\iff b \in (a) \\ &\iff (b) \subset (a). \end{aligned}$$

If you have trouble remembering this, just think about the case $a = 2, b = 10$ for $R = \mathbf{Z}$ rather than memorize anything.

Example 6.10. For $R = \mathbf{Z}$ and $c, a, b \in \mathbf{Z}^+$ we have

$$c = ab \iff (c) = (ab) = (a)(b)$$

since $\mathbf{Z}^\times = \{1, -1\}$ meets $\mathbf{Z}^+$ in $\{1\}$. Here, we are using that for a domain D and $\alpha, \beta \in D - \{0\}$, $(\alpha) = (\beta)$ if and only if $\alpha = \beta u$ for some $u \in D^\times$.

We now come back to explaining Remark 6.1.

Example 6.11. Consider $R = \mathbf{Z}[\sqrt{-5}]$ and the ideals

$$\begin{aligned} \mathfrak{p} &:= (2, 1 + \sqrt{-5}) \\ \mathfrak{q} &:= (3, 1 + \sqrt{-5}). \end{aligned}$$

We'll show soon that $\mathfrak{p}, \mathfrak{q}$ are indeed non-principal (and they will later be seen to be instances of prime ideals, hence the notation for them). We have

$$\overline{\mathfrak{q}} := (3, 1 - \sqrt{-5})$$

where the overline denotes complex conjugation (applying the nontrivial element of $\mathrm{Gal}(\mathbf{Q}(\sqrt{-5})/\mathbf{Q})$). Observe that $\mathfrak{p} + \mathfrak{q} = (1)$ just because $2 \in \mathfrak{p}$, $3 \in \mathfrak{q}$, and we have $2\mathbf{Z} + 3\mathbf{Z} = \mathbf{Z}$ as ideals of $\mathbf{Z}$.

Lemma 6.12. *The ideals $\mathfrak{p}, \mathfrak{q}$ are not principal.*

It follows by applying the conjugation *automorphism* that $\overline{\mathfrak{q}}$ is also not principal.

Proof. Suppose $\mathfrak{p} = (\alpha)$ for some $\alpha \in \mathbf{Z}[\sqrt{-5}]$, so $2 \in \mathfrak{p} = (\alpha)$ yet 2 is irreducible, so up to units (which is all that matters for any possible α) the only options are $\alpha = 2$ or $\alpha = 1$. We shall rule out both possibilities.

We have

$$(2, 1 + \sqrt{-5}) \neq (2)$$

because $1 + \sqrt{-5}$ clearly cannot be expressed in the form $2(u + v\sqrt{-5})$ with $u, v \in \mathbf{Z}$.

To show $\mathfrak{p} \neq (1)$, we first recall that for any ring R and ideal I the additive quotient group R/I has a natural ring structure using multiplication of representatives in R. We will show that the quotient ring $\mathbf{Z}[\sqrt{-5}]/\mathfrak{p}$ is nonzero: via HW2 we have $\mathbf{Z}[x]/(f) \simeq \mathbf{Z}[\beta]$ for any algebraic integer β that is a root of a monic $f \in \mathbf{Z}[x]$ that is irreducible over $\mathbf{Q}$ (the isomorphism carries x to β), so via the resulting isomorphism $\mathbf{Z}[x]/(x^2+5) \simeq \mathbf{Z}[\sqrt{-5}]$ carrying x to $\sqrt{-5}$ we have (using the definition $\mathfrak{p} = (2, 1 + \sqrt{-5})$)

$$\begin{aligned}
\mathbf{Z}[\sqrt{-5}]/\mathfrak{p} &= (\mathbf{Z}[x]/(x^2+5))/\mathfrak{p} \\
&= \mathbf{Z}[x]/(x^2+5, 2, 1+x) \\
&= \mathbf{F}_2[x]/(x^2+5, 1+x) \\
&= \mathbf{F}_2[x]/(x+1) \\
&= \mathbf{F}_2.
\end{aligned}$$

This completes the proof that $\mathfrak{p}$ is non-principal. The case of $\mathfrak{q}$ goes similarly. □

Now, observe that

$$\begin{aligned}
\mathfrak{p}\mathfrak{q} &= (6, 2(1+\sqrt{-5}), 3(1+\sqrt{-5}), (1+\sqrt{-5})^2) \\
&= (1+\sqrt{-5})(1-\sqrt{-5}, 2, 3, 1+\sqrt{-5}) \\
&= (1+\sqrt{-5})(1) \\
&= (1+\sqrt{-5}).
\end{aligned}$$

Similarly,

$$\begin{aligned}\mathfrak{p}\overline{\mathfrak{q}} &= (2, 1+\sqrt{-5})(3, 1-\sqrt{-5})\\ &= (6, 2(1-\sqrt{-5}), 3(1+\sqrt{-5}), 6)\\ &= (6, 2(1-\sqrt{-5}), (-2+\sqrt{-5})(1-\sqrt{-5}), 6)\\ &= (1-\sqrt{-5})(1+\sqrt{-5}, 2, -2+\sqrt{-5})\\ &= (1-\sqrt{-5})(1)\\ &= (1-\sqrt{-5}).\end{aligned}$$

Therefore,

$$\begin{aligned}(6) &= (1+\sqrt{-5})(1-\sqrt{-5})\\ &= (\mathfrak{p}\mathfrak{q})(\mathfrak{p}\overline{\mathfrak{q}})\\ &= \mathfrak{p}^2(\mathfrak{q}\overline{\mathfrak{q}}),\end{aligned}$$

where the final step entailed some rearrangement of terms reminiscent of what was seen in the factorization of 210 in two different ways.

Now we fulfill Remark 6.1 by showing that the preceding rearrangement of the non-principal ideal factors $\mathfrak{p}, \mathfrak{q}, \overline{\mathfrak{q}}$ can be interpreted in terms of principal ideals to arrive at the more familiar factorization of 6 as a product of the irreducibles 2 and 3 in $\mathbf{Z}[\sqrt{-5}]$:

Lemma 6.13. *We have* $\mathfrak{p}^2 = (2)$ *and* $\mathfrak{q}\overline{\mathfrak{q}} = (3)$.

Proof. These are direct calculations:

$$\begin{aligned}\mathfrak{p}^2 &= (2, 1+\sqrt{-5})(2, 1+\sqrt{-5})\\ &= (4, 2(1+\sqrt{-5}), -4+2\sqrt{-5})\\ &= (2)(2, 1+\sqrt{-5}, -2+\sqrt{-5})\\ &= (2)(1)\\ &= (2)\end{aligned}$$

and

$$\begin{aligned}\mathfrak{q}\overline{\mathfrak{q}} &= (3, 1+\sqrt{-5})(3, 1-\sqrt{-5})\\ &= (9, 3(1+\sqrt{-5}), 3(1+\sqrt{-5}), 6)\\ &= (3)(3, 1-\sqrt{-5}, 1+\sqrt{-5}, 2)\\ &= (3).\end{aligned}$$

□

Now that we have seen that the failure of $\mathbf{Z}[\sqrt{-5}]$ to be a UFD via the different irreducible factorizations of 6 is explained by rearranging a factorization of (6) as a product of 4 non-principal ideals, we want to explain the sense in which those 4 non-principal ideals deserve to be called "prime". As a warm-up, we consider a perspective on irreducibility for UFD's:

Lemma 6.14. *For any UFD R and $\pi \in R$ a nonzero non-unit, $R/(\pi)$ is a domain if and only if π is irreducible.*

Proof. By definition, a domain is not the zero ring. Since π is a non-unit, $R/(\pi)$ is not the zero ring. A nonzero ring is a domain precisely when any product of nonzero elements is nonzero. For $R/(\pi)$, this is precisely the condition in R that $\pi \mid ab \implies \pi \mid a$ or $\pi \mid b$. Since R is a UFD, it is not difficult to check (by considering the factorization of the nonzero non-unit π into irreducibles) that this is precisely the condition that π is irreducible. □

Warning 6.15. Note that Lemma 6.14 does not hold without the assumption that R is a UFD. For example, we have seen that 3 is irreducible in the non-UFD $R = \mathbf{Z}[\sqrt{-5}]$ but

$$\begin{aligned} R/(3) &= \mathbf{Z}[x]/(x^2+5,3) \\ &= \mathbf{F}_3[x]/(x^2-1) \end{aligned}$$

is not a domain because $(x+1)(x-1) = x^2-1$ but $x+1$ and $x-1$ are nonzero modulo the quadratic x^2-1.

Definition 6.16. An ideal $I \subset R$ is *prime* if the quotient ring R/I is a domain.

Since domains are non-zero (i.e., the zero ring is not a domain) by definition, prime ideals are always proper ideals (i.e., not the unit ideal) by definition. Thus, an ideal I in a ring R is a prime ideal if and only if I is a proper ideal and $ab \in I \Rightarrow a \in I$ or $b \in I$; this is sometimes presented as the initial definition of primality for ideals.

Example 6.17. Here is an example of a proper ideal that is not prime, for $R = \mathbf{Z}[\sqrt{-3}]$. Consider $I = (1+\sqrt{-3})$. One can verify $1+\sqrt{-3}$ is irreducible by the usual norm considerations, but

$$R/I = \mathbf{Z}[x]/(x^2+3, 1+x) = \mathbf{Z}/((-1)^2+3) = \mathbf{Z}/4\mathbf{Z}$$

(using $\mathbf{Z}[x]/(1+x) \simeq \mathbf{Z}$ for the second equality). Note that $\mathbf{Z}/4\mathbf{Z}$ is not a domain, since $2 \cdot 2 \equiv 0 \bmod 4$; explicitly, $2 \notin (1+\sqrt{-3})$ (as can easily be checked directly: do it!) whereas

$$2 \cdot 2 = 4 = (1+\sqrt{-3})(1-\sqrt{-3}).$$

Example 6.18. We can "fix" the failure of primality in the preceding example by passing to the slightly bigger ideal

$$\mathfrak{p} := (2, 1+\sqrt{-3}) \subset \mathbf{Z}[\sqrt{-3}].$$

Indeed,

$$\begin{aligned}\mathbf{Z}[\sqrt{-3}]/\mathfrak{p} &= \mathbf{Z}[x]/(x^2+3, 2, 1+x)\\ &= \mathbf{F}_2[x]/(x^2+3, 1+x)\\ &= \mathbf{F}_2[x]/(x+1)\\ &= \mathbf{F}_2\end{aligned}$$

is clearly a domain.

In terms of our new terminology, if a domain R is not a UFD then for a nonzero non-unit $r \in R$ the primality of (r) is *not* equivalent to r being irreducible (e.g., $r = 3$ in $R = \mathbf{Z}[\sqrt{-5}]$ as in Warning 6.15) whereas the equivalence does hold when R is a UFD.

Definition 6.19. An ideal $\mathfrak{m}$ in a commutative ring R is *maximal* if $R/\mathfrak{m}$ is a field.

Exercise 6.20. Verify $\mathfrak{m}$ is maximal if and only if $\mathfrak{m} \neq (1)$ and $\mathfrak{m}$ is not strictly contained inside another proper ideal (thereby explaining the terminology).

Note that maximal ideals are prime because fields are domains.

Example 6.21. For R a PID, the prime ideals are (r) precisely for r irreducible and for $r = 0$. (The ideal (0) is prime in a ring R precisely when R is a domain. There are very good reasons for permitting (0) as a possibility for the definition of prime ideal, but this will be better appreciated later in life after you learn some modern algebraic geometry.) The maximal ideals in a PID are precisely the ideals generated by the irreducible elements (such ideals really are not strictly contained in any other proper ideal: why not?).

In particular, the prime ideals of $\mathbf{Z}$ are just the ideals (p) for prime $p \in \mathbf{Z}^+$ and the ideal (0), whereas the maximal ideals of $\mathbf{Z}$ are precisely (p) for prime $p \in \mathbf{Z}^+$.

7. Primes in $\mathscr{O}_K$

Let K be a number field and let $\mathscr{O}_K$ be its ring of integers. Today, we'll discuss some basic properties of nonzero ideals in $\mathscr{O}_K$, generalizing the case $K = \mathbf{Q}$ for which nonzero ideals are $m\mathbf{Z} \subset \mathbf{Z}$ for unique $m \in \mathbf{Z}^+$.

Theorem 7.1. *Let K be a number field of degree $n := [K : \mathbf{Q}]$. Then, the following results hold:*

(1) For each nonzero ideal $\mathfrak{a} \subset \mathscr{O}_K$, $\mathfrak{a} \simeq \mathbf{Z}^n$ as a $\mathbf{Z}$-module and $\#\mathscr{O}_K/\mathfrak{a} < \infty$.
(2) All nonzero prime ideals $\mathfrak{p} \subset \mathscr{O}_K$ are maximal ideals.
(3) For all maximal ideals $\mathfrak{p} \subset \mathscr{O}_K$, we have $\mathfrak{p} \cap \mathbf{Z} = p\mathbf{Z}$ for some prime $p \in \mathbf{Z}^+$.

Remark 7.2. We'll see later that for every prime $p \in \mathbf{Z}^+$ there exists a maximal ideal $\mathfrak{p} \subset \mathscr{O}_K$ such that $\mathfrak{p} \cap \mathbf{Z} = p\mathbf{Z}$.

Proof. We prove the statements in order.

(1) Choose some nonzero $\alpha \in \mathfrak{a}$. We have that $\alpha\mathscr{O}_K \subset \mathfrak{a} \subset \mathscr{O}_K \simeq \mathbf{Z}^n$, using Theorem 5.7. It follows that $\mathfrak{a} \simeq \mathbf{Z}^r$ for some $r \leq n$. Since multiplication by α induces an isomorphism $\mathscr{O}_K \simeq \alpha\mathscr{O}_K$ as $\mathbf{Z}$-modules, we have $\alpha\mathscr{O}_K \simeq \mathbf{Z}^n$ as $\mathbf{Z}$-modules. Thus, $\mathbf{Z}^n \simeq \alpha\mathscr{O}_K \subset \mathfrak{a} \simeq \mathbf{Z}^r$, so $n \leq r$. It follows that $r = n$, so $\mathfrak{a} \simeq \mathbf{Z}^n$ as $\mathbf{Z}$-modules.

Now we need to show that $\mathscr{O}_K/\mathfrak{a}$ is finite. Since $\mathscr{O}_K \simeq \mathbf{Z}^n$ and $\mathfrak{a} \simeq \mathbf{Z}^n$, this reduces to the following lemma:

Lemma 7.3. *Any injective homomorphism $\mu : \mathbf{Z}^n \to \mathbf{Z}^n$ has finite cokernel.*

Proof. This follows from the structure theorem for finitely generated modules over a PID, which gives that for any inclusion $M' \subset M$ between free finitely generated modules over a PID R, there exists a basis $\{e_1, \ldots, e_m\}$ of M and $\{e'_1, \ldots, e'_{m'}\}$ of M' (with $m' \leq m$) such that $e'_j = r_j e_j$ for some $r_j \in R - \{0\}$; in effect, $M' \hookrightarrow M$ looks diagonal relative to suitable R-bases of M' and M. (Beware that this does *not* assert that we can first choose a basis of M' and then find a basis of M for which the above "diagonal" formula holds.)

Applying this to make suitable $\mathbf{Z}$-bases $\{e'_i\}$ of $\mathfrak{a}$ and $\{e_i\}$ of $\mathscr{O}_K$ such that $\mu(e'_i) = d_i e_i$ for all i with some $d_i \in \mathbf{Z} - \{0\}$, we have $\operatorname{coker} \mu \simeq \prod_{i=1}^n (\mathbf{Z}/d_i\mathbf{Z})$, which is visibly finite. □

(2) For $\mathfrak{p} \subset \mathscr{O}_K$ a nonzero prime ideal, we would like to show that $\mathscr{O}_K/\mathfrak{p}$ is a field. Since $\mathscr{O}_K/\mathfrak{p}$ is a finite domain, we can apply:

Lemma 7.4. *Suppose R is a finite domain. Then R is a field.*

Proof. For $r \in R - \{0\}$, the multiplication-by-r map $m_r : R \to R$ defined by $x \mapsto rx$ is injective because R is a domain. Since R is finite, m_r must therefore be surjective too! This implies there exists some s with $rs = 1$. This s serves as the multiplicative inverse for r, so R is a field. □

(3) Let $\mathfrak{p} \subset \mathscr{O}_K$ be a nonzero prime ideal. We can realize $\mathfrak{p} \cap \mathbf{Z}$ as the kernel of the composite ring map

$$\mathbf{Z} \longrightarrow \mathscr{O}_K \longrightarrow \mathscr{O}_K/\mathfrak{p}. \tag{7.1}$$

Call this composition ϕ. Note that $\ker \phi \neq 0$ since $\mathscr{O}_K/\mathfrak{p}$ is finite whereas $\mathbf{Z}$ is infinite, so $\ker \phi = m\mathbf{Z}$ for $m \in \mathbf{Z}_{>0}$. It remains to show m is prime. However, we have an injection of rings $\mathbf{Z}/m\mathbf{Z} \to \mathscr{O}_K/\mathfrak{p}$. Since a subring of a domain is a domain, we have that $\mathbf{Z}/m\mathbf{Z}$ is a domain, so m is prime (as otherwise $m = ab$ with $a, b > 1$, so $a, b \in \mathbf{Z}/m\mathbf{Z}$ are nonzero elements whose product is zero, contradicting the domain property).

□

Remark 7.5. Suppose $\mathfrak{p} \subset \mathscr{O}_K$ is a nonzero prime with $p\mathbf{Z} \subset \mathfrak{p} \cap \mathbf{Z}$. Then, $\mathfrak{p} \cap \mathbf{Z} = p\mathbf{Z}$ since $p\mathbf{Z}$ is a maximal ideal and $1 \notin \mathfrak{p} \cap \mathbf{Z}$.

Example 7.6. Let $K = \mathbf{Q}(\sqrt{d})$ for $d \equiv 2, 3 \bmod 4$. What are the prime ideals containing $p\mathscr{O}_K$ for $p \in \mathbf{Z}$ a prime?

To answer this question, and in particular to see that at least one such prime of $\mathscr{O}_K$ exists for each p, recall that $\mathscr{O}_K = \mathbf{Z}[\sqrt{d}] \simeq \mathbf{Z}[x]/(x^2 - d)$. Moreover, any prime $\mathfrak{p}$ of $\mathscr{O}_K$ containing $p\mathbf{Z}$ must contain $p\mathscr{O}_K$ since $\mathfrak{p}$ contains all $\mathscr{O}_K$-multiples of any of its elements (such as p), so the set of prime ideals $\mathfrak{p}$ of $\mathscr{O}_K$ containing p is in bijection with the set of prime ideals of $\mathscr{O}_K/p\mathscr{O}_K$. (Here we are using that for any ring R and ideal I there is a bijection $J \mapsto J/I$ from the set of ideals of R containing I onto the set of ideals of R/I, under which primes correspond to primes in both directions since $(R/I)/(J/I) \simeq R/J$ as rings.)

We have $\mathscr{O}_K/p\mathscr{O}_K \simeq \mathbf{F}_p[x]/(x^2 - d)$, and this ring structure can be determined by the Chinese Remainder Theorem (keep in mind that for $p = 2$ and odd d we have $x^2 - d = (x-1)^2$ in $\mathbf{F}_2[x]$):

$$\mathbf{F}_p[x]/(x^2 - d) = \begin{cases} \mathbf{F}_p[x]/(x-1)^2 & \text{if } p = 2 \text{ and } d \text{ is odd },\\ \mathbf{F}_{p^2} & \text{if } p \neq 2, p \nmid d, d \not\equiv \square \bmod p,\\ \mathbf{F}_p \times \mathbf{F}_p & \text{if } p \neq 2, p \nmid d, d \equiv \square \bmod p,\\ \mathbf{F}_p[x]/x^2 & \text{if } p \mid d. \end{cases}$$

In the first of these four cases there is only one prime ideal in the indicated ring (since any field quotient of that ring must kill the nilpotent $x - 1$), corresponding to a unique prime ideal containing $p = 2$, namely $\mathfrak{p} = (2, \sqrt{d} - 1) = (2, \sqrt{d} + 1)$. Likewise, in the second case $\mathfrak{p} = p\mathscr{O}_K$ is the unique prime ideal containing p, but in the third case with $d \equiv u^2 \bmod p$ we see that the vanishing of $(1, 0) \cdot (0, 1) = (0, 0)$ in $\mathbf{F}_p \times \mathbf{F}_p$ forces any field quotient *as a ring* to kill one of $(1, 0)$ or $(0, 1)$ and hence to be projection to either of the two evident copies of $\mathbf{F}_p$ (corresponding to killing $x + u$ or $x - u$). Thus, there are exactly two such $\mathfrak{p}$ in the third case, namely $(p, \sqrt{d} + u)$ and

$(p, \sqrt{d} - u)$. For the final case there is again exactly one prime containing p (as in the first case), namely $\mathfrak{p} = (p, \sqrt{d})$.

Observe that in the preceding example we always have that each ordinary prime $p \in \mathbf{Z}^+$ lies in either exactly one prime ideal of $\mathscr{O}_K$ or exactly two such prime ideals. Further, $\mathscr{O}_K/p\mathscr{O}_K$ has nonzero nilpotents only for finitely many p, namely for those $p \mid d$ along with $p = 2$ when d is odd.

Later we shall see that such properties generalize to any number field K: there are only finitely many primes $p \in \mathbf{Z}$ for which $\mathscr{O}_K/(p)$ has nonzero nilpotents (which will turn out to be exactly when $p\mathscr{O}_K$ has a repeated prime ideal factor, akin to the case of (2) in $\mathbf{Z}[i]$; such p will be called *ramified* in K), and the number of prime ideals of $\mathscr{O}_K$ containing p is always between 1 and $[K : \mathbf{Q}]$.

Remark 7.7. A standard convention is to say "prime of K" to mean "nonzero prime ideal of $\mathscr{O}_K$" (or equivalently "maximal ideal of $\mathscr{O}_K$") and "rational prime" to denote a prime $p \in \mathbf{Z}^+$. We will often use this convention.

As the course develops, we will see that a common scenario will be to consider an extension K'/K of number fields and the corresponding ring extension $\mathscr{O}_K \subset \mathscr{O}_{K'}$. Since $\mathscr{O}_{K'}$ is finitely generated as a $\mathbf{Z}$-module, it is certainly finitely generated as an $\mathscr{O}_K$-module. However, whereas $\mathscr{O}_{K'}$ is free as a $\mathbf{Z}$-module (since $\mathbf{Z}$ is a PID), it will typically happen that $\mathscr{O}_{K'}$ is *not* free as an $\mathscr{O}_K$-module (we will encounter examples of quadratic K with $\mathscr{O}_K$ not a PID). An important and useful task will be to understand the factorization of $\mathfrak{p}\mathscr{O}_{K'}$ for prime ideals $\mathfrak{p}$ of $\mathscr{O}_K$; for $K = \mathbf{Q}$ and $K' = \mathbf{Q}(i)$ this was the key to Fermat's 2-square theorem. Many Diophantine problems will also be related to such factorization questions.

Our next goal is to understand the theory of factorization of nonzero ideals in rings of integers of number fields. The main highlights will be:

(1) Nonzero ideals $\mathfrak{a} \subset \mathscr{O}_K$ always have at most 2 generators (a fact that is nice to know – just barely not a PID! – but in practice usually useless).
(2) Each nonzero ideal $\mathfrak{a}$ factors uniquely (up to rearrangement) as a finite product $\mathfrak{a} = \prod_{i=1}^{r} \mathfrak{p}_i$ of maximal ideals $\mathfrak{p}_i \subset \mathscr{O}_K$.
(3) There exists a positive integer h_K, called the *the class number*, such that $\mathfrak{a}^{h_k}$ is principal for all nonzero ideals $\mathfrak{a} \subset \mathscr{O}_K$.

The first step on the road toward understanding such matters is to build up more experience with rings of integers beyond the quadratic case, and in particular to develop methods to prove in some non-quadratic cases that $\mathscr{O}_K \simeq \mathbf{Z}[\alpha]$. This will rests on an "integral" theory of discriminants, refining the notion that has arisen already in the case of separable finite-degree field extensions: at the level of such fields we defined a notion of discriminant that is nonzero and well-defined up to square multiple. In the integral theory, we will define a more refined notion of discriminant that is an actual specific number, with *no scaling ambiguity* at all!

8. Discriminants of number fields

We will define the discriminant of a number field, or really the "discriminant over $\mathbf{Z}$" of $\mathscr{O}_K$, to be denoted $\mathrm{disc}_{\mathbf{Z}}(\mathscr{O}_K)$.

Goal 8.1. There are two initial applications (the first of which we will reach later today) for the integral version of discriminants that we shall define:

(1) For $\alpha \in \mathscr{O}_K$ such that $K = \mathbf{Q}(\alpha)$ (so the minimal polynomial $f \in \mathbf{Q}[x]$ of α lies in $\mathbf{Z}[x]$ and has degree $n = [K : \mathbf{Q}]$),

$$\mathbf{Z}^n \simeq \oplus_{j=0}^{n-1} \mathbf{Z}\alpha^j = \mathbf{Z}[\alpha] \subset \mathscr{O}_K \simeq \mathbf{Z}^n,$$

so $\mathbf{Z}[\alpha] \subset \mathscr{O}_K$ has finite index. How can we bound this index multiplicatively (i.e., find an explicit nonzero integer that is a multiple of this index) *without* having already found $\mathscr{O}_K$ inside K?

(2) For a prime $p \in \mathbf{Z}^+$, we have $p\mathscr{O}_K = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_g^{e_g}$ for distinct maximal ideals $\mathfrak{p}_j$ and $e_j \geq 1$. We'd like to show that there are only finitely many p for which some $e_j > 1$; how do we show this, and in practice find these finitely many "bad" p?

Example 8.2. In the special case $K = \mathbf{Q}(\sqrt{d})$ with a squarefree integer $d \neq 0, 1$ we have mentioned that the property in (2) that some $e_j > 1$ will be equivalent to $\mathscr{O}_K / p\mathscr{O}_K$ having a nonzero nilpotent element, a property that for $d \equiv 2, 3 \bmod 4$ we have seen happens if and only if $p | d$ or $p = 2$ with d odd. How does this explicit characterization of such p generalize beyond these quadratic cases?

Toy version of discriminants. In (1) above, we have

$$\mathbf{Z}[x]/(f) \simeq \mathbf{Z}[\alpha] \subset \mathscr{O}_K$$

and the evident $\mathbf{Z}$-basis $\{1, \alpha, \ldots, \alpha^{n-1}\}$ of $\mathbf{Z}[\alpha]$ has associated determinant

$$\mathrm{disc}_{\mathbf{Q}}(1, \alpha, \ldots, \alpha^{n-1}) := \det(\mathrm{Tr}_{K/\mathbf{Q}}(\alpha^i \alpha^j)) \in \mathbf{Z} - \{0\} \in \mathbf{Z}$$

since all traces in this matrix are in $\mathbf{Z}$ (and the non-vanishing was seen in Remark 3.1 of the handout "Norm and Trace", applicable to any ordered basis of a finite-degree separable extension of fields, such as $K/\mathbf{Q}$). It will be seen later today that this explicit determinant is computable and always divisible by $[\mathscr{O}_K : \mathbf{Z}[\alpha]]$ (something even better than this will be proved), providing a-priori control on what $\mathscr{O}_K$ might be as a lattice containing the explicit $\mathbf{Z}[\alpha]$.

Recall from the handout "Norm and Trace" that for a finite extension F/k with $\{e_i\}$ an ordered k-basis of F, the determinant

$$D(e_1, \ldots, e_n) = \det(\mathrm{Tr}_{F/k}(e_i e_j)) \in k$$

depends on $\{e_i\}$ up to multiplication by a nonzero square in k (via a formula we'll review shortly), and is nonzero when F/k is separable.

Example 8.3. The field $K = \mathbf{Q}(\sqrt{d})$ has $\mathbf{Q}$-basis $\{1, \sqrt{d}\}$, and

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

In the preceding example, if we change the $\mathbf{Q}$-basis of K then the determinant changes by a nonzero rational square. But if we limit ourselves to a class of $\mathbf{Q}$-bases that have a fixed $\mathbf{Z}$-span in K then the scaling ambiguity in the determinant will be completely eliminated! We illustrate this with the most important case first:

Proposition 8.4. *If $\{e_i\}$ and $\{e'_i\}$ are $\mathbf{Z}$-bases of the same $\mathbf{Z}$-lattice in K then*

$$D(e_1, \ldots, e_n) = D(e'_1, \ldots, e'_n).$$

By a "$\mathbf{Z}$-lattice" we mean the $\mathbf{Z}$-span of a $\mathbf{Q}$-basis of K; i.e., the $\mathbf{Z}$-span of $n = [K : \mathbf{Q}]$ elements linearly independent over $\mathbf{Q}$. (For example, every nonzero ideal of $\mathscr{O}_K$ is a $\mathbf{Z}$-lattice in K, as is $\mathbf{Z}[\alpha]$ for any $\alpha \in K$ for which $\mathbf{Q}(\alpha) = K$.)

Proof. We can write $e'_i = \sum c_{ij} e_j$ for a matrix $M = (c_{ij})$ that is invertible over $\mathbf{Q}$, and

$$D(e'_1, \ldots, e'_n) = \det(\mathrm{Tr}_{K/\mathbf{Q}}(e'_i e'_j)) = \det(B(e'_i, e'_j))$$

where B denotes the symmetric bilinear form

$$\begin{aligned} B \colon K \times K &\to \mathbf{Q} \\ (x, y) &\mapsto \mathrm{Tr}_{K/\mathbf{Q}}(xy). \end{aligned}$$

The matrix $(B(e'_i, e'_j))$ computes B when vectors are written relative to the basis $\{e'_i\}$. How is this related to the matrix that computes B relative to the basis $\{e_i\}$? If we write $[v]_e$ and $[v]_{e'}$ to denote the elements of $\mathbf{Q}^n$ ("column vectors") expressing the coordinates of a given $v \in K$ relative to the respective ordered $\mathbf{Q}$-bases $\{e_i\}$ and $\{e'_i\}$ then since $[v]_e = M[v]_{e'}$ (keep in mind how M was defined!) we have

$$\begin{aligned} B(x,y) &= [x]_e^T(B(e_i,e_j))[y]_e \\ &= [x]_{e'}^T M^T(B(e_i,e_j))M[y]_{e'} \\ &= [x]_{e'}^T(M^T(B(e_i,e_j))M)[y]_{e'}, \end{aligned}$$

so $(B(e'_i,e'_j)) = M^T(B(e_i,e_j))M$. (This is the general formula for how the "matrix" computing a given bilinear form $B : V \times V \to k$ on a finite-dimensional vector space V over a field k changes under a change of ordered basis.) It follows that

$$D(e'_1,\dots,e'_n) = \det(M)^2 D(e_1,\dots,e_n)$$

since $\det M^T = \det M$.

So far we have not used that $\{e_i\}, \{e'_i\}$ are $\mathbf{Z}$-bases of the same $\mathbf{Z}$-lattice L. But given this latter property we have $M \in \mathrm{Mat}_n(\mathbf{Z})$ and $M^{-1} \in \mathrm{Mat}_n(\mathbf{Z})$ (since each of $\{e_i\}$ and $\{e'_i\}$ expressed in terms of the other involves only $\mathbf{Z}$-coefficients, as each lies in the $\mathbf{Z}$-span of the other due to both having the same $\mathbf{Z}$-span by hypothesis). It follows that $\det M \in \mathbf{Z}^\times = \{\pm 1\}$, so $\det(M)^2 = 1$! □

Definition 8.5. The *discriminant* of $K/\mathbf{Q}$ is $D(e_1,\dots,e_n) \in \mathbf{Z} - \{0\}$ for any $\mathbf{Z}$-basis $\{e_i\}$ of $\mathscr{O}_K$; this is denoted by the notations $\mathrm{disc}_{\mathbf{Z}}(\mathscr{O}_K)$, $\mathrm{disc}(\mathscr{O}_K/\mathbf{Z})$, or $\mathrm{disc}(K/\mathbf{Q})$.

For any $\mathbf{Z}$-lattice $L \subset K$, define $\mathrm{disc}_{\mathbf{Z}}(L) := D(e_1,\dots,e_n) \in \mathbf{Q}^\times$ for any $\mathbf{Z}$-basis $\{e_i\}$ of L. (This belongs to $\mathbf{Z}$ when $L \subset \mathscr{O}_K$.)

These notions of discriminant are well-defined elements of $\mathbf{Q}$ (no scaling ambiguity!) due to Proposition 8.4, and they are always nonzero since such non-vanishing holds for these determinant constructions applied to *any* $\mathbf{Q}$-basis of K (as seen for separable finite-degree extensions of fields in the handout "Norm and Trace"). The notation $\mathrm{disc}(K/\mathbf{Q})$ to mean $\mathrm{disc}(\mathscr{O}_K/\mathbf{Z})$ is perhaps a bit abusive (much like "prime of K" to mean "maximal ideal of $\mathscr{O}_K$"), but it is ubiquitous in practice.

Example 8.6. Let $L = \mathbf{Z}[\alpha]$ for $\alpha \in \mathscr{O}_K$ such that $K = \mathbf{Q}(\alpha)$. (We have seen that many such α exist: begin with any primitive element for K over $\mathbf{Q}$ and multiply it by a sufficiently divisible nonzero integer.) This lattice admits as a $\mathbf{Z}$-basis $\{1,\alpha,\dots,\alpha^{n-1}\}$, so the intrinsic $\mathrm{disc}_{\mathbf{Z}}(L)$ coincides with the determinant $D(1,\alpha,\dots,\alpha^{n-1})$. By Remark 3.1 in the handout "Norm and Trace", if f is the minimal polynomial of α over $\mathbf{Q}$ then a van der Monde calculation yields

$$D(1,\alpha,\dots,\alpha^{n-1}) = (-1)^{n(n-1)/2} N_{K/\mathbf{Q}}(f'(\alpha))$$

with $f'(\alpha) \in K^\times$ since $K/\mathbf{Q}$ is separable.

Example 8.7. If $n = 2$ and $f = x^2 + ax + b$ then

$$\begin{aligned} D(1, \alpha) &= -N_{K/\mathbf{Q}}(2\alpha + a) \\ &= -(2\alpha + a)(2\overline{\alpha} + a) \\ &= a^2 - 4b. \end{aligned}$$

Example 8.8. If $n = 3$ and $f = x^3 + ax + b$ then

$$D(1, \alpha, \alpha^2) = -4a^3 - 27b^2;$$

see [Samuel, p. 41] for the more general case with irreducible f of the form $x^n + ax + b$ for any $n > 1$.

Example 8.9. If $K = \mathbf{Q}(\sqrt{d})$ for some square-free $d \in \mathbf{Z} - \{0, 1\}$ then $\mathscr{O}_K = \mathbf{Z}[\alpha]$ where

$$\alpha = \begin{cases} \sqrt{d} & \text{if } d \equiv 2, 3 \bmod 4, \\ \frac{1+\sqrt{d}}{2} & \text{if } d \equiv 1 \bmod 4. \end{cases}$$

Thus, by a matrix calculation that we already saw for $d \equiv 2, 3 \bmod 4$ and leave as an exercise for $d \equiv 1 \bmod 4$,

$$\operatorname{disc}(K/\mathbf{Q}) = \begin{cases} 4d & \text{if } d \equiv 2, 3 \bmod 4, \\ d & \text{if } d \equiv 1 \bmod 4. \end{cases}$$

The sign of the discriminant detects if the quadratic field is real or imaginary quadratic, and we observe by inspection (using that d is square-free) that any two non-isomorphic quadratic fields have distinct discriminants. Later in the course we will encounter a remarkable theorem of Hermite (see [Samuel, §4.3]) that there are only *finitely many* number fields K with a given discriminant (no hypothesis on $[K : \mathbf{Q}]$!); the importance of this will be appreciated later as well.

Remark 8.10. The discriminant can be given a geometric meaning in terms of volume. This is going to be significant in our later work with the "geometry of numbers", and for now we illustrate it in the first non-trivial case: real quadratic fields.

Consider $K = \mathbf{Q}(\sqrt{d})$ with square-free $d > 1$, and for simplicity assume $d \equiv 2, 3 \bmod 4$. (What we are about to do also works for $d \equiv 1 \bmod 4$ in terms of the validity of the final conclusion.) We want to think about K as an abstract field, in order not to prejudice ourselves among its two embeddings into $\mathbf{R}$, so let's write K as $\mathbf{Q}(\alpha)$ with $\alpha^2 = d$. (The point is that in $\mathbf{R}$ there is a genuine distinction between the two square roots of d, with exactly one being positive for the unique order structure on $\mathbf{R}$, but K does not have a preferred order structure and so there is no way whatsoever to intrinsically distinguish among the two square roots of d in K; after all, $\mathrm{Gal}(K/\mathbf{Q})$ swaps these square roots!)

We have two field embeddings

$$\begin{aligned} \tau_{\pm}\colon K &\rightrightarrows \mathbf{R} \\ \alpha &\mapsto \pm\sqrt{d}. \end{aligned}$$

Using both at once defines an injective $\mathbf{Q}$-linear map

$$\begin{aligned} \sigma\colon K &\to \mathbf{R}\times\mathbf{R} \\ x &\mapsto (\tau_+(x), \tau_-(x)). \end{aligned}$$

Let's see where $\mathscr{O}_K$ goes. Since $\mathscr{O}_K = \mathbf{Z}\oplus\mathbf{Z}\alpha$ (recall $\alpha^2 = d$), we have

$$\sigma(\mathscr{O}_K) = \mathbf{Z}\sigma(1) + \mathbf{Z}\sigma(\alpha).$$

This is the $\mathbf{Z}$-span of the vectors $\sigma(1) = (1,1)$ and $\sigma(\alpha) = (\sqrt{d}, -\sqrt{d})$; such a $\mathbf{Z}$-span is visibly a "lattice" in the plane $\mathbf{R}^2$ for which a fundamental parallelogram (i.e., $\{t\sigma(1) + t'\sigma(\alpha) \,|\, 0 \le t, t' \le 1\}$) has area

$$\left|\det\begin{pmatrix} 1 & \sqrt{d} \\ 1 & -\sqrt{d}\end{pmatrix}\right| = 2\sqrt{d} = \sqrt{|\operatorname{disc}(K/\mathbf{Q})|}.$$

Now it is finally time to begin to reveal the real power of discriminants, starting with the following innocuous-looking property that we shall see has striking applications:

Theorem 8.11. *For $\mathbf{Z}$-lattices $L' \subset L$ inside K, $[L' : L] < \infty$ and*

$$\operatorname{disc}_{\mathbf{Z}}(L') = [L' : L]^2 \cdot \operatorname{disc}_{\mathbf{Z}}(L).$$

In particular, if $L' \subset \mathscr{O}_K$ and the integer $\operatorname{disc}_{\mathbf{Z}}(L') \in \mathbf{Z} - \{0\}$ *is squarefree then $[\mathscr{O}_K : L'] = 1$; i.e., $L' = \mathscr{O}_K$!*

The key point is the final assertion, since we can compute $\operatorname{disc}_{\mathbf{Z}}(L')$ without knowing exactly what $\mathscr{O}_K$ is (e.g., this can be applied to $L' = \mathbf{Z}[\alpha]$ for $\alpha \in \mathscr{O}_K$ that is a primitive element for $K/\mathbf{Q}$). This will be seen to impose strong constraints on the possibilities for $\mathscr{O}_K$.

Proof. Since L and L' are $\mathbf{Z}$-free of the same finite rank, by the structure theorem for modules over a PID, we can pick $\mathbf{Z}$-bases $\{e_i\}$ of L and $\{e'_i\}$ of L' so that $e'_i = d_i e_i$ for some $d_i \in \mathbf{Z} - \{0\}$. Thus, $\mathrm{Tr}_{K/\mathbf{Q}}(e'_i e'_j) = d_i d_j \,\mathrm{Tr}_{K/\mathbf{Q}}(e_i e_j)$, so we have the matrix relation

$$\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} (\mathrm{Tr}_{K/\mathbf{Q}}(e_ie_j)) \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} = (\mathrm{Tr}_{K/\mathbf{Q}}(e_i'e_j')).$$

Applying the determinant to both sides yields

$$\begin{aligned} \mathrm{disc}_{\mathbf{Z}}(L') &= (\prod d_i^2) \cdot \mathrm{disc}_{\mathbf{Z}}(L) \\ &= (\prod |d_i|)^2 \cdot \mathrm{disc}_{\mathbf{Z}}(L). \end{aligned}$$

But the compatible **Z**-bases also allow us to identify the quotient L/L' with $\oplus_i(\mathbf{Z}/d_i\mathbf{Z})$ (make sure you understand this!), so

$$[L' : L] = \#(L/L') = \prod \#(\mathbf{Z}/d_i\mathbf{Z}) = \prod |d_i|.$$

Hence, $[L' : L]^2 = (\prod |d_i|)^2$, so we are done. □

Now we can illustrate the power of this "integral" notion of discriminant: the key point is that if $\alpha \in K$ is a primitive element over **Q** and $\alpha \in \mathscr{O}_K$ with $\mathbf{Z}[\alpha]$ having *squarefree* discriminant then its square factor $[\mathscr{O}_K : \mathbf{Z}[\alpha]]^2$ is forced to be 1, which is to say $\mathscr{O}_K = \mathbf{Z}[\alpha]$!

Example 8.12. Take $K = \mathbf{Z}(\theta)$ where $\theta^3 - \theta - 1 = 0$. (It is easy via the rational root theorem to check that $X^3 - X - 1$ is irreducible in $\mathbf{Q}[X]$.) Note that $\mathbf{Z}[\theta] \subset \mathscr{O}_K$. By the general formula for discriminants given above in the cubic case with $f = X^3 + aX + b$, we have

$$D(1, \theta, \theta^2) = -4b^3 - 27a^2 = 4 - 27 = -23.$$

This is squarefree, so by Theorem 8.11 we have $\mathbf{Z}[\theta] = \mathscr{O}_K$.

We can also compute the matrix of traces explicitly (and then compute its determinant by hand):

$$(\mathrm{Tr}_{K/\mathbf{Q}}(\theta^{i+j}))_{0\leq i,j\leq 2} = \begin{pmatrix} 3 & 0 & 2 \\ 0 & 2 & 3 \\ 2 & 3 & 2 \end{pmatrix}$$

Indeed, it is obvious that $\mathrm{Tr}_{K/\mathbf{Q}}(1) = [K : \mathbf{Q}] = 3$ and $\mathrm{Tr}_{K/\mathbf{Q}}(\theta) = 0$ since the minimal polynomial of θ has vanishing quadratic coefficient. Once we figure out $\mathrm{Tr}_{K/\mathbf{Q}}(\theta^2)$ we can work out $\mathrm{Tr}_{K/\mathbf{Q}}(\theta^m)$ for any $m \geq 3$ since the cubic relation for θ allows us to write any such θ^m as a **Z**-linear combination of 1, θ, and θ^2, and $\mathrm{Tr}_{K/\mathbf{Q}}$ is **Z**-linear.

To compute $\mathrm{Tr}_{K/\mathbf{Q}}(\theta^2)$, one method is to go back to the definition of field trace and compute the matrix for multiplication m_{θ^2} explicitly relative to the basis $\{1, \theta, \theta^2\}$ and thereby check that its trace is 2: this matrix is (check!)

$$[m_{\theta^2}] = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Another method is to manipulate symmetric functions in Galois conjugates over $\mathbf{Q}$: if $\{\theta_1, \theta_2, \theta_3\}$ are the roots of $X^3 - X - 1$ in a splitting field over $\mathbf{Q}$ then

$$\mathrm{Tr}_{K/\mathbf{Q}}(\theta^2) = \sum \theta_i^2 = (\sum \theta_i)^2 - 2\sum_{i<j} \theta_i \theta_j = 0^2 - 2(-1) = 2.$$

With the traces of $1, \theta, \theta^2$ known, one can compute the traces of higher powers via repeated substitutions into the relation $\theta^3 = 1 + \theta$ as mentioned above; e.g., $\theta^4 = \theta + \theta^2$, so $\mathrm{Tr}_{K/\mathbf{Q}}(\theta^4) = \mathrm{Tr}_{K/\mathbf{Q}}(\theta) + \mathrm{Tr}_{K/\mathbf{Q}}(\theta^2) = 2$.

Example 8.13. A root θ of the irreducible $x^3 + x + 1 \in \mathbf{Q}[x]$ satisfies

$$D(1, \theta, \theta^2) = -31$$

and a root θ of the irreducible $x^3 + 10x + 1 \in \mathbf{Q}[x]$ satisfies

$$D(1, \theta, \theta^2) = -4027$$

(a prime!), so in each of these cases $\mathscr{O}_K = \mathbf{Z}[\theta]$.

Sometimes life is not so simple: $x^3 - x - 4$ is an irreducible cubic over $\mathbf{Q}$ and a root θ of this cubic satisfies

$$(\mathrm{Tr}_{K/\mathbf{Q}}(\theta^{i+j}))_{0 \le i,j \le 2} = \begin{pmatrix} 3 & 0 & 2 \\ 0 & 2 & 12 \\ 2 & 12 & 2 \end{pmatrix}$$

(it is a good exercise to check the correctness of this matrix of traces); the determinant of this matrix is $-428 = -4 \cdot 107$. The maximal square factor of this integer is 4, so for $K = \mathbf{Q}(\theta)$ the (finite) index of $\mathbf{Z}[\theta]$ inside $\mathscr{O}_K$ is equal to either 1 or 2; i.e., either $\mathscr{O}_K = \mathbf{Z}[\theta]$ or $\mathscr{O}_K$ contains $\mathbf{Z}[\theta]$ with index 2. Hence, $\mathscr{O}_K$ lies somewhere between $\mathbf{Z}[\theta]$ and $(1/2) \cdot \mathbf{Z}[\theta]$.

We have constrained $\mathscr{O}_K$ to at most 8 possibilities: $\mathbf{Z}[\theta]$ or $\mathbf{Z}[\theta] + \mathbf{Z}v$ for v representing one of the 7 nonzero cosets in $(1/2) \cdot \mathbf{Z}[\theta]/\mathbf{Z}[\theta] = ((1/2)\mathbf{Z}/\mathbf{Z})^2$ (a 3-dimensional vector space over $\mathbf{F}_2$). Have we just "chosen poorly" for θ as an integral primitive element for $K/\mathbf{Q}$, or is the index really 2? Next time we will explore this example further.

9. Some monogenic integer rings

We now resume the example we left off with at the end of last time.

Example 9.1. Let $K = \mathbf{Q}(\alpha)$ with $\alpha^3 - \alpha - 4 = 0$. We would like to determine the ring of integers $\mathscr{O}_K$. To start, we should (as always) verify that the polynomial $x^3 - x - 4$ is irreducible over $\mathbf{Q}$. To do this, by the rational root theorem it suffices to check that no factors of the constant term -4 (i.e., $\pm 4, \pm 2, \pm 1$) are roots. This is not too difficult or time-consuming (even to do in one's head) since the coefficients of the cubic are tiny (and no quadratic term appears) and 4 has few factors (including negatives!).

As we noted last time,

$$[\mathscr{O}_K : \mathbf{Z}[\alpha]]^2 \operatorname{disc}_{\mathbf{Z}}(\mathscr{O}_K) = \operatorname{disc}_{\mathbf{Z}}(\mathbf{Z}[\alpha]) = -2^2 \cdot 107.$$

Thus, $\mathbf{Z}[\alpha]$ has index in $\mathscr{O}_K$ dividing 2, and hence $2\mathscr{O}_K \subset \mathbf{Z}[\alpha]$, so $\mathbf{Z}[\alpha] \subset \mathscr{O}_K \subset \frac{1}{2}\mathbf{Z}[\alpha]$. To find $\mathscr{O}_K$, all we need to do is check among coset representatives of the nonzero elements in $\frac{1}{2}\mathbf{Z}[\alpha]/\mathbf{Z}[\alpha]$ which are algebraic integers; a choice of such representatives is given by

$$\beta := \frac{a_0}{2} + \frac{a_1}{2}\alpha + \frac{a_2}{2}\alpha^2$$

with $a_0, a_1, a_2 \in \{0, 1\}$, and not all a_i equal to 0. Since the index is 1 or 2, at most one such coset class belongs to $\mathscr{O}_K$.

We also know $\frac{1}{2} \notin \mathscr{O}_K$, so we can assume either a_1 or a_2 is nonzero (i.e., is equal to 1); hence, there are only 6 possible β for which we need to check if $\beta \in \mathscr{O}_K$. By HW2, this is the same as the minimal polynomial of such β over $\mathbf{Q}$ having coefficients in $\mathbf{Z}$. How do we find these minimal polynomials?

At this step we employ a small trick: $[K : \mathbf{Q}] = 3$ is prime, and such β have been designed to not lie in $\mathbf{Q}$, so all such β have cubic minimal polynomial over $\mathbf{Q}$ and hence it suffices to find *some* monic cubic over $\mathbf{Q}$ satisfied by β (that must then be the minimal polynomial). To find such a cubic, consider the $\mathbf{Q}$-linear endomorphism

$$\begin{aligned} m_\beta \colon K &\to K \\ x &\mapsto \beta x \end{aligned}$$

of the 3-dimensional $\mathbf{Q}$-vector space K. This operator has monic cubic characteristic polynomial $\chi_\beta(T)$. By Cayley Hamilton, as operators $K \to K$ we have $\chi_\beta(m_\beta) = 0$. Evaluating both sides on the vector $1 \in K$, we get

$$\chi_\beta(\beta) = 0$$

in K, since $m_\beta^n(1) = \beta^n$ for any $n \geq 0$.

The upshot is that in each of the 6 cases we compute the matrix for m_β relative to some $\mathbf{Q}$-basis (e.g., $\{1, \alpha, \alpha^2\}$, repeatedly using the relation $\alpha^3 = \alpha + 4$) and then from that matrix compute χ_β as its characteristic polynomial. If this ever lies in $\mathbf{Z}[T]$ then we have found an extra element β of $\mathscr{O}_K$ so that $\mathscr{O}_K = \mathbf{Z}[\alpha] + \mathbf{Z}\beta$ (and hence $\mathscr{O}_K = \mathbf{Z}[\alpha, \beta]$); if no χ_β lie in $\mathbf{Z}[T]$ then $\mathscr{O}_K = \mathbf{Z}[\alpha]$.

Let's try the case

$$\beta := \frac{\alpha + \alpha^2}{2} = \frac{1}{2}\alpha + \frac{1}{2}\alpha^2.$$

The matrix $[m_\beta]$ for multiplication by β with respect to the $\mathbf{Q}$-basis $\{1, \alpha, \alpha^2\}$ of K is determined by the computations

$$1 \mapsto \frac{1}{2}\alpha + \frac{1}{2}\alpha^2, \;\; \alpha \mapsto 2 + \frac{1}{2}\alpha + \frac{1}{2}\alpha^2, \;\; \alpha^2 \mapsto 2 + \frac{5}{2}\alpha + \frac{1}{2}\alpha^2$$

(using as always the relation $\alpha^3 = \alpha + 4$). In other words,

$$[m_\beta] = \begin{pmatrix} 0 & 2 & 2 \\ 1/2 & 1/2 & 5/2 \\ 1/2 & 1/2 & 1/2 \end{pmatrix}$$

Computing the characteristic polynomial, we get

$$\chi_\beta = T^3 - T^2 - 3T - 2 \in \mathbf{Z}[T],$$

so $\mathscr{O}_K = \mathbf{Z}[\alpha] + \mathbf{Z} \cdot \beta = \mathbf{Z}[\alpha, \beta]$. Having discovered $\beta \in \mathscr{O}_K - \mathbf{Z}[\alpha]$, so $\mathbf{Z}[\alpha]$ has index 2 in $\mathscr{O}_K$ (rather than index 1), we obtain

$$\operatorname{disc}_{\mathbf{Z}} \mathscr{O}_K = -107$$

since $\mathbf{Z}[\alpha]$ has discriminant $-4 \cdot 107$.

Exercise 9.2. Curiosity may lead us to wonder how deep inside $\mathscr{O}_K$ the subring $\mathbf{Z}[\beta]$ is. One can check

$$D(1, \beta, \beta^2) = -107;$$

make sure to do this by hand, not with a computer! We deduce that $\mathscr{O}_K = \mathbf{Z}[\beta]$, so $\mathscr{O}_K$ is monogenic in this case (even though our first guess for a ring generator, α, didn't work).

Let's now try another example:

Example 9.3 (Dedekind)**.** Let $f := x^3 - x^2 - 2x - 8$. For a root α of f we will consider $K = \mathbf{Q}(\alpha)$. To start, we check that f is irreducible over $\mathbf{Q}$. There are a couple of ways to do this:

(1) We can use the rational root test, but it's starting to get a bit annoying to check all integral factors of the constant term: there are now more factors (can't ignore the negative ones!), and the polynomial is getting more complicated to evaluate in our head.

(2) We can use Gauss' criterion:

Lemma 9.4. *If $f \in \mathbf{Z}[T]$ is monic and its reduction $\overline{f} \in \mathbf{F}_p[T]$ is irreducible for some prime p then f is irreducible over $\mathbf{Q}$.*

Proof. This is shown in HW3 in the wider context of PID's (in place of $\mathbf{Z}$). □

Using Gauss' criterion, it suffices to check that $f \bmod p \in \mathbf{F}_p[x]$ is irreducible for *one* prime p, for which it is enough (due to being a cubic) that no element of $\mathbf{F}_p$ is a root. For small p this is something we can check with much less of a hassle than the rational root test method over $\mathbf{Q}$. Since $p = 2$ doesn't work, we try $p = 3$: it is easy to check that f has no roots modulo 3 by plugging in $0, 1$, and -1. Therefore, f really is irreducible over $\mathbf{Q}$.

We would like to find the ring of integers $\mathscr{O}_K$. Might it coincide with its finite-index subring $\mathbf{Z}[\alpha]$? As we have seen, the first order of business is to compute the discriminant of $\mathbf{Z}[\alpha]$ (by systematically using the cubic relation $f(\alpha) = 0$):

Exercise 9.5. Verify that

$$D(1, \alpha, \alpha^2) = -2012 = -4 \cdot 503.$$

It follows that the inclusion

$$\mathbf{Z}[\alpha] \subset \mathscr{O}_K$$

has index 1 or 2. As in the previous example, there are 7 possible coset representatives to check (one of them is $\frac{1}{2}$, so we can rule that out and reduce to checking 6 possibilities). It turns out that one of these elements β does lie in the ring of integers, so the index is 2. But what makes this cubic example more interesting than the previous one is that one doesn't get "lucky" and find that $\mathscr{O}_K$ is monogenic with β as a generator; in fact, whichever β you try, you will find

$$\operatorname{disc}_{\mathbf{Z}} \mathbf{Z}[\beta] \neq -503,$$

so $\mathbf{Z}[\beta] \neq \mathscr{O}_K$.

It turns out that $\mathscr{O}_K \neq \mathbf{Z}[\theta]$ for all $\theta \in \mathscr{O}_K - \mathbf{Z}$! What is the obstruction? If $\theta \in \mathscr{O}_K$ is a primitive element for $K/\mathbf{Q}$ with minimal polynomial h over $\mathbf{Q}$, so necessarily $h \in \mathbf{Z}[x]$, then there are at most p ring homomorphisms

$$\mathbf{Z}[\theta] = \mathbf{Z}[x]/(h) \twoheadrightarrow \mathbf{F}_p$$

(sending x to an element of $\mathbf{F}_p$ at which h vanishes modulo p). But if for some p there are *more* than p maximal ideals $\mathfrak{p}_j$ containing p such that $\mathscr{O}_K/\mathfrak{p}_j \simeq \mathbf{F}_p$ then we would get more than p ring homomorphisms $\mathscr{O}_K \twoheadrightarrow \mathbf{F}_p$ (pairwise distinct kernel ideals $\mathfrak{p}_j$), and this would rule out the possibility that $\mathscr{O}_K$ could have the form $\mathbf{Z}[\theta]$. In effect, having "too many" maximal ideals containing a given small prime p provides an obstruction to monogenicity.

Using the structure theory of Dedekind domains that we shall develop in the coming lectures, which requires more sophisticated techniques than those we have seen so far (and is really where algebraic number theory gets off the ground), we will show as Dedekind did that for this specific cubic field K there are *three* distinct maximal ideals $\mathfrak{m} \subset \mathscr{O}_K$ with $\mathscr{O}_K/\mathfrak{m} \simeq \mathbf{F}_2$. Since $3 > 2$, this property implies that $\mathscr{O}_K$ cannot be monogenic! This will later be pushed much further, to show that for any $n > 1$ there are number fields K such that $\mathscr{O}_K$ doesn't admit n ring generators.

Before proceeding to develop the theory of Dedekind domains, we discuss one further important class of examples: cyclotomic fields. These are number fields of the form $K = \mathbf{Q}(\zeta_n)$, the splitting field over $\mathbf{Q}$ of $x^n - 1$. Recall the following theorem from field theory, also proved in [Samuel, §1.6]:

Theorem 9.6. *If k is a field, any finite subgroup of $k^\times$ is cyclic.*

For example, we can apply this to the set of solutions to $x^n - 1$ in a splitting field k'/k (as this set of solutions is certainly a finite subgroup of $k'^\times$). If follows that if $\operatorname{char} k \nmid n$, so $x^n - 1$ is separable over k, in k' the set of roots is cyclic of size n. We call a generator for this group a *primitive nth root of unity*.

Example 9.7. When $k = \mathbf{C}$, we can visualize the cyclicity via the vertices of a regular n-gon centered at 0 with one vertex at 1 (and so vertices equal to the points $e^{2\pi ij/n}$ for $j \in \mathbf{Z}/(n)$). In this case we can take a primitive nth root of unity to be $e^{2\pi i/n}$.

It is absolutely essential to remember that from an algebraic point of view, any primitive nth root of unity is just as good as any other: there is no intrinsic algebraic sense in which $e^{\pm 2\pi i/n}$ are preferred if one is working inside $\mathbf{C}$, for example, and for many later purposes it will be important to maintain the viewpoint of a number field as an abstract field (not equipped with a specific embedding into $\mathbf{C}$). Thus, *never* write "$e^{2\pi i/n}$" when you wish to refer to a primitive nth root of unity; always denote it as ζ_n. This notation promotes clearer thinking.

Remark 9.8. Recall that any cyclic group of size n has $\phi(n)$ generators. For example, if we choose a primitive nth root of unity ζ_n in a field k (with $\operatorname{char} k \nmid n$) then the primitive nth roots of unity in k are precisely ζ_n^a with $\gcd(a, n) = 1$ (i.e., $a \in (\mathbf{Z}/n\mathbf{Z})^\times$).

Since any single primitive nth root of unity generates *all* roots to $x^n - 1$, a splitting field for $x^n - 1$ over a field k with $\operatorname{char} k \nmid n$ always has the form $k(\zeta_n)$ for ζ_n a primitive nth root of 1. The Galois conjugates of such a ζ_n over k are also primitive nth roots of unity (because they must have the same finite multiplicative order n that ζ_n does), so since there are $\phi(n)$ primitive nth roots of 1 we deduce:

Lemma 9.9. *If k is a field with* $\operatorname{char} k \nmid n$ *then* $[k(\zeta_n) : k] \leq \phi(n)$.

Upon choosing some ζ_n, we define the injective (!) map of sets

$$\begin{aligned} a : \operatorname{Gal}(k(\zeta_n)/k) &\to (\mathbf{Z}/n\mathbf{Z})^\times \\ \sigma &\mapsto a(\sigma) \end{aligned}$$

where $a(\sigma)$ is defined by the formula $\sigma(\zeta_n) = \zeta_n^{a(\sigma)}$. Now observe that for any $j \in \mathbf{Z}/n\mathbf{Z}$ whatsoever, $\sigma(\zeta_n^j) = \sigma(\zeta_n)^j = (\zeta_n^{a(\sigma)})^j = (\zeta_n^j)^{a(\sigma)}$. In other words, for *all* ζ satisfying $\zeta^n = 1$ we have $\sigma(\zeta) = \zeta^{a(\sigma)}$, so the same exponent $a(\sigma) \in (\mathbf{Z}/n\mathbf{Z})^\times$ computes the effect of σ on all nth roots of unity. That is, $a(\sigma)$ does not depend on the initial choice of ζ_n. It follows that a is a homomorphism since for any ζ at all and $\tau, \sigma \in \operatorname{Gal}(k(\zeta_n)/k)$ we have

$$\tau(\sigma(\zeta)) = \tau(\zeta^{a(\sigma)}) = (\tau(\zeta))^{a(\sigma)} = (\zeta^{a(\tau)})^{a(\sigma)} = \zeta^{a(\tau)a(\sigma)}$$

when $\zeta^n = 1$; this shows that $a(\tau\sigma) = a(\tau)a(\sigma)$.

Example 9.10. If $k = \mathbf{R}$ and $n > 1$ is odd then $k(\zeta_n) = \mathbf{C}$ and the map a carries complex conjugation to -1 (since the complex conjugate of z is $1/z = z^{-1}$ for any z on the unit circle in $\mathbf{C}$).

We conclude that in general the inequality $[k(\zeta_n) : k] \leq \phi(n)$ is an equality if and only if $a : \operatorname{Gal}(k(\zeta_n)/k) \hookrightarrow (\mathbf{Z}/n\mathbf{Z})^\times$ is an isomorphism, which is precisely the case when *all* primitive nth roots of 1 are Galois conjugates to each other over k. In such cases one could truly say that all primitive nth roots of 1 are "created equal" over k, since there is no way to distinguish them from each other algebraically over k (as the Galois group permutes them transitively in such cases). For $k = \mathbf{R}$ and $n > 2$ the minimal polynomials over $\mathbf{R}$ of the primitive nth roots of 1 are all quadratic, namely

$$X^2 - 2\cos(2\pi j/n)X + 1 \in \mathbf{R}[X]$$

for $j \in (\mathbf{Z}/n\mathbf{Z})^\times$, so algebraically over $\mathbf{R}$ the various primitive nth roots of 1 over $\mathbf{C}$ are distinguishable up to the effect of complex conjugation (passing to the other root of the same minimal polynomial over $\mathbf{R}$).

But over $\mathbf{Q}$ the situation is much better: it is an important fact, originally due to Dedekind and proved in Galois theory (and whose proof we will review later), that

$$[\mathbf{Q}(\zeta_n) : \mathbf{Q}] = \phi(n).$$

We will use this as an ingredient in our proof that

$$\mathscr{O}_{\mathbf{Q}(\zeta_n)} = \mathbf{Z}[\zeta_n],$$

a property of much historical importance for Kummer's development of many features of algebraic number theory in the special case of cyclotomic fields. Kummer was lucky that even though he didn't know about the concept of algebraic integer, the naive guess $\mathbf{Z}[\zeta_n]$ for the ring to focus upon for arithmetic in $\mathbf{Q}(\zeta_n)$ turns out to be the "right one" (in the sense of having many good properties from the general theory, due to being the actual ring of integers of the number field).

Our development of properties of $\mathbf{Q}(\zeta_n)$ will proceed in stages, first with n a prime power and then the general case with aid from discriminant calculations. So now suppose $n = p^r$ with p a prime and $r \geq 1$. A primitive p^rth root of 1 has multiplicative order exactly p^r, which is to say that its p^rth-power is equal to 1 but its p^{r-1}th-power does not equal 1. In other words, this is precisely the condition of being a root of

$$\frac{X^{p^r} - 1}{X^{p^{r-1}} - 1} = \frac{Y^p - 1}{Y - 1}$$

for $Y := X^{p^{r-1}}$. The right side is $Y^{p-1} + Y^{p-2} + \cdots + Y + 1$, motivating us to define

$$\Phi_{p^r}(X) := X^{p^{r-1}(p-1)} + X^{p^{r-1}(p-2)} \cdots + X^{p^{r-1}} + 1 \in \mathbf{Z}[X];$$

we have just argued that the roots of Φ_{p^r} are precisely the primitive p^rth roots of 1. Dedekind's theorem on equality for $[\mathbf{Q}(\zeta_n) : \mathbf{Q}] \leq \phi(n)$ in the case $n = p^r$ amounts to:

Lemma 9.11. *The polynomial Φ_{p^r} is irreducible over $\mathbf{Q}$.*

This is proved in HW3 via yet another irreducibility criterion over fraction fields of PID's called *Eisenstein's criterion*.

10. Prime-power cyclotomic rings

Fix a prime p and $r \geq 1$. Let $K = \mathbf{Q}(\zeta_{p^r}) = \mathrm{split}_{\mathbf{Q}}(X^{p^r} - 1)$. We want to show:

Theorem 10.1. $\mathbf{Z}[\zeta_{p^r}] = \mathscr{O}_K$

Before proving this, we review some results from last time. Recall that

$$\begin{aligned}\Phi_{p^r}(x) &:= \frac{x^{p^r} - 1}{x^{p^{r-1}-1}} \\ &= x^{p^{r-1}(p-1)} + x^{p^{r-1}(p-2)} + \cdots + x^{p^{r-1}} + 1\end{aligned}$$

has as its roots exactly the primitive p^rth roots of unity, and by Eisenstein's irreducibility criterion it is irreducible over $\mathbf{Q}$. Thus,

$$\begin{aligned}\mathbf{Z}[x]/(\Phi_{p^r}) &= \mathbf{Z}[\zeta_{p^r}] \\ &\subset K \\ &= \mathbf{Q}[x]/(\Phi_{p^r})\end{aligned}$$

(in particular, $[K : \mathbf{Q}] = \deg \Phi_{p^r} = \phi(p^r)$) and the inclusion

$$\begin{aligned}a : \mathrm{Gal}(K/\mathbf{Q}) &\hookrightarrow (\mathbf{Z}/p^r\mathbf{Z})^\times \\ \sigma &\mapsto a(\sigma)\end{aligned}$$

(defined by $\sigma(\zeta) = \zeta^{a(\sigma)}$ for *all* p^rth roots of unity ζ) is an equality for size reasons (as $\#(\mathbf{Z}/p^r\mathbf{Z})^\times = \phi(p^r) = p^{r-1}(p-1) = \deg \Phi_{p^r} = [K : \mathbf{Q}]$).

For the rest of today, we write ζ for ζ_{p^r} (a primitive p^rth root of unity). Since the algebraic integer ζ is a primitive element for $K/\mathbf{Q}$, we know that the inclusion

$$\mathbf{Z}[\zeta] \subset \mathscr{O}_K$$

has finite index. We want it to have index 1.

We know

$$\mathrm{disc}_{\mathbf{Z}}(\mathbf{Z}[\zeta]) = [\mathscr{O}_K : \mathbf{Z}[\zeta]]^2 \cdot \mathrm{disc}_{\mathbf{Z}}\, \mathscr{O}_K.$$

In particular, $[\mathscr{O}_K : \mathbf{Z}[\zeta]] \mid \mathrm{disc}_{\mathbf{Z}}(\mathbf{Z}[\zeta])$. For abelian groups $A \subset A'$ with A'/A finite, multiplication by $[A' : A]$ kills A'/A and hence

$$[A' : A]A' \subset A.$$

Thus, for

$$\begin{aligned}d &:= D(1, \zeta, \ldots, \zeta^{p^{r-1}(p-1)-1}) \\ &= \mathrm{disc}_{\mathbf{Z}}(\mathbf{Z}[\zeta]) \\ &\neq 0\end{aligned}$$

we have

$$d \cdot \mathscr{O}_K \subset \mathbf{Z}[\zeta].$$

In other words,

$$\mathscr{O}_K \subset \frac{1}{d} \cdot \mathbf{Z}[\zeta].$$

We now have now reached the heart of the proof. We will state the following two lemmas, show why this implies the desired theorem, and then come back to prove the lemmas.

Lemma 10.2. *The discriminant d is a p-power up to a sign, with $p|d$ when $p^r > 2$.*

Lemma 10.3. *We have $((1/p)\mathbf{Z}[\zeta]) \cap \mathscr{O}_K = \mathbf{Z}[\zeta]$.*

Granting the lemmas, we argue as follows. If $\alpha \in \mathscr{O}_K$, by Lemma 10.2 we have $\alpha = \beta/p^e$ for $\beta \in \mathbf{Z}[\zeta]$ and some integer $e \geq 0$. We'll show by induction on general $e \geq 0$ that

$$(p^{-e}\mathbf{Z}[\zeta]) \cap \mathscr{O}_K = \mathbf{Z}[\zeta].$$

The case $e = 0$ is trivial and the case $e = 1$ is Lemma 10.3. Suppose $e > 1$ and assume the asserted equality holds for $e - 1$. Hence, for any $\alpha = \beta/p^e \in \mathscr{O}_K$ with $\beta \in \mathbf{Z}[\zeta]$ we have

$$p\alpha = \beta/p^{e-1} \in \mathscr{O}_K$$

with $\beta \in \mathbf{Z}[\zeta]$, so by induction $p\alpha \in \mathbf{Z}[\zeta]$. But then

$$\alpha \in ((1/p)\mathbf{Z}[\zeta]) \cap \mathscr{O}_K = \mathbf{Z}[\zeta],$$

where we used Lemma 10.3 for the final equality.

To complete the proof of Theorem 10.1, we only need prove Lemma 10.2 and Lemma 10.3.

Proof of Lemma 10.2. The minimal polynomial of ζ over $\mathbf{Q}$ is

$$f := \Phi_{p^r} = X^{p^{r-1}(p-1)} + \cdots + X^{p^{r-1}} + 1,$$

so

$$D(1, \zeta, \ldots, \zeta^r) = \pm N_{\mathbf{Q}(\zeta)/\mathbf{Q}}(f'(\zeta)).$$

Recall that

$$f = \frac{X^{p^r} - 1}{X^{p^{r-1}} - 1}.$$

Therefore,

$$f' = \frac{(X^{p^{r-1}} - 1)p^r X^{p^r - 1} - (X^{p^r} - 1)p^{r-1}X^{p^{r-1}-1}}{(X^{p^{r-1}} - 1)^2}.$$

Evaluating at ζ kills the second term in the numerator, and then cancelling a remaining common factor of $(X^{p^{r-1}}-1)|_{X=\zeta}$ in the numerator and denominator yields

$$f'(\zeta)=\frac{p^r\zeta^{-1}}{\zeta_p-1},$$

with $\zeta_p:=\zeta^{p^{r-1}}$. (We have used that $\zeta^{p^r-1}=\zeta^{-1}$ since $\zeta^{p^r}=1$.) Since $N_{\mathbf{Q}(\zeta)/\mathbf{Q}}(\zeta)=\pm 1$ (either by looking at the constant term of the minimal polynomial Φ_{p^r} of ζ over $\mathbf{Q}$, or by using that this norm is rational yet also a root of unity due to multiplicativity of the norm), we get

$$N_{\mathbf{Q}(\zeta)/\mathbf{Q}}(f'(\zeta))=\frac{(p^r)^{p^{r-1}(p-1)}(\pm 1)}{N_{\mathbf{Q}(\zeta)/\mathbf{Q}}(\zeta_p-1)},$$

where the exponent $p^{r-1}(p-1)$ is coming from the degree of K over $\mathbf{Q}$. Since $N_{\mathbf{Q}(\zeta)/\mathbf{Q}}(\zeta_p-1)\in\mathbf{Z}$, the nonzero integer $N_{\mathbf{Q}(\zeta)/\mathbf{Q}}(f'(\zeta))$ is a p-power up to sign (as desired).

We have shown $\pm d$ is a p-power, which is all we shall actually need. To see $p|d$ when $p^r>2$, we need to determine $N_{\mathbf{Q}(\zeta)/\mathbf{Q}}(\zeta_p-1)$ as a p-power up to a sign. Below we will see this is equal to $N_{\mathbf{Q}(\zeta_p)/\mathbf{Q}}(\zeta_p-1)^{p^{r-1}}$ with $N_{\mathbf{Q}(\zeta_p)/\mathbf{Q}}(\zeta_p-1)=\pm p$, so this norm in the denominator is $\pm p^{p^{r-1}}$. Comparing with the p-power in the numerator, we get $p|d$ except precisely when $r(p-1)=1$, which is to say $p^r=2$. □

Remark 10.4. The denominator in the above expression for $N_{\mathbf{Q}(\zeta)/\mathbf{Q}}(f'(\zeta))$ can be simplified a bit as follows. For any finite extension of fields k'/k and $a\in k'$ we have $N_{k'/k}(a)=N_{k(a)/k}(N_{k'/k(a)}(a))=N_{k(a)/k}(a^{[k':k(a)]})=N_{k(a)/k}(a)^{[k':k(a)]}$, so

$$N_{\mathbf{Q}(\zeta)/\mathbf{Q}}(\zeta_p-1)=N_{\mathbf{Q}(\zeta_p)/\mathbf{Q}}(\zeta_p-1)^{p^{r-1}}$$

since $[\mathbf{Q}(\zeta):\mathbf{Q}(\zeta_p)]=p^{r-1}$.

Before proving Lemma 10.3, we need the following lemma.

Lemma 10.5. *The following three statements hold:*

(1) $N_{\mathbf{Q}(\zeta)/\mathbf{Q}}(\zeta-1)=\pm p$.

(2) For $j\in(\mathbf{Z}/p^r\mathbf{Z})^\times$ we have

$$\frac{\zeta^j-1}{\zeta-1}\in\mathbf{Z}[\zeta]^\times.$$

(3) We have $p=u(\zeta-1)^{\phi(p^r)}$ for some $u\in\mathbf{Z}[\zeta]^\times$.

Proof. For (2), first note that $(\zeta^j - 1)/(\zeta - 1) \in \mathbf{Z}[\zeta]$ since $(Y^j - 1)/(Y - 1) = 1 + Y + \cdots + Y^{j-1}$. We need to show $(\zeta - 1)/(\zeta^j - 1) \in \mathbf{Z}[\zeta]$. Observe that $\zeta' := \zeta^j$ is a primitive p^rth root of unity (since $p \nmid j$), so $\zeta = \zeta'^h$ for some $h \in (\mathbf{Z}/p^r\mathbf{Z})^\times$. (Explicitly, h is multiplicative inverse to j modulo p^r.) Hence,

$$\frac{\zeta - 1}{\zeta^j - 1} = \frac{\zeta'^h - 1}{\zeta' - 1} \in \mathbf{Z}[\zeta'] \subset \mathbf{Z}[\zeta].$$

This settles (2).

Next, we reduce (3) to (1), using (2). Let $\Gamma = \mathrm{Gal}(\mathbf{Q}(\zeta)/\mathbf{Q}) = (\mathbf{Z}/p^r\mathbf{Z})^\times$. We have

$$\begin{aligned} N_{\mathbf{Q}(\zeta)/\mathbf{Q}}(\zeta - 1) &= \prod_{g \in \Gamma} (g(\zeta - 1)) \\ &= \prod_{g \in \Gamma} (g(\zeta) - 1) \\ &= \prod_{j \in (\mathbf{Z}/p^r\mathbf{Z})^\times} (\zeta^j - 1), \end{aligned}$$

and by (2) this final product is a $\mathbf{Z}[\zeta]^\times$-multiple of $(\zeta - 1)^{\phi(p^r)}$ as desired. Therefore, (3) is reduced to (1).

Finally, to prove (1) we note that

$$Y^{p^{r-1}(p-1)} + \cdots + Y^{p^{r-1}} + 1 = \Phi_{p^r}(Y) = \prod_{j \in (\mathbf{Z}/p^r\mathbf{Z})^\times} (Y - \zeta^j),$$

so evaluating at 1 gives

$$p = 1 + \cdots + 1 = \Phi_{p^r}(1) = \pm N_{\mathbf{Q}(\zeta)/\mathbf{Q}}(\zeta - 1)$$

where the sign comes from swapping the order of subtraction in the preceding product description of the norm (so the sign is actually $(-1)^{\phi(p^r)}$, which is equal to 1 except when $p^r = 2$ since $\phi(p^r)$ is easily seen to be even whenever $p^r > 2$ by considering the cases $p \neq 2$ and $p = 2$ separately). □

Proof of Lemma 10.3. We want to show

$$\mathscr{O}_K \cap (1/p)\mathbf{Z}[\zeta] = \mathbf{Z}[\zeta].$$

It is equivalent to show that

$$(p\mathscr{O}_K) \cap \mathbf{Z}[\zeta] = p\mathbf{Z}[\zeta].$$

We'll write elements of $\mathbf{Z}[\zeta]$ in terms of $\mathbf{Z}$-linear combinations of powers of $\zeta - 1$ rather than ζ, as we may certainly do (since $\zeta = (\zeta - 1) + 1$):

$$\alpha = c_0 + c_1(\zeta - 1) + \cdots + c_{n-1}(\zeta - 1)^{n-1}$$

where $n = [K : \mathbf{Q}] = \phi(p^r)$ and $c_j \in \mathbf{Z}$. We want to show that if $\alpha \in p\mathscr{O}_K$ then $c_j \in p\mathbf{Z}$ for all $0 \le j < n$.

Note that it is harmless to replace α by any $\beta \in \mathbf{Z}[\zeta]$ that differs from α by elements of $p\mathbf{Z}[\zeta] = p\mathbf{Z}[\zeta - 1]$. We start at $j = 0$ and then increase j. Since

$$c_0 = \alpha - (c_1(\zeta - 1) + \cdots + c_{n-1}(\zeta - 1)^{n-1}) \in p\mathscr{O}_K + (\zeta - 1)\mathscr{O}_K,$$

by Lemma 10.5(3) we see that c_0 even lies in $(\zeta - 1)\mathscr{O}_K$, so $c_0 \in p\mathbf{Z}$ due to:

Lemma 10.6. *We have*

$$\mathbf{Z} \cap (\zeta - 1)\mathscr{O}_K = p\mathbf{Z}.$$

Proof. The containment

$$\mathbf{Z} \cap (\zeta - 1)\mathscr{O}_K \supset p\mathbf{Z}.$$

is clear. To check the reverse containment, for $a \in \mathbf{Z} \cap (\zeta - 1)\mathscr{O}_K$ we have by Lemma 10.5(3) that

$$a^{\phi(p^r)} \in \mathbf{Z} \cap p\mathscr{O}_K = p((1/p)\mathbf{Z} \cap \mathscr{O}_K) = p\mathbf{Z},$$

which implies $a \in p\mathbf{Z}$. □

By Lemma 10.6, we can replace α with $\alpha - c_0$. Now we may suppose inductively that

$$\alpha = c_{i_0}(\zeta - 1)^{i_0} + \cdots + c_{n-1}(\zeta - 1)^{n-1} \in p\mathscr{O}_K$$

for some $i_0 < n$. We want to show $c_{i_0} \in p\mathbf{Z}$, so it is enough to show $c_{i_0} \in (\zeta - 1)\mathscr{O}_K$ (by Lemma 10.6). By Lemma 10.5(3) we have

$$(\zeta - 1)^n\mathscr{O}_K = p\mathscr{O}_K \ni \alpha = c_{i_0}(\zeta - 1)^{i_0} + \cdots .$$

All terms on the right beyond the i_0-term (if any such occur) are divisible by $(\zeta - 1)^{i_0+1}$, as is the left side since $i_0 < n$, so dividing by $(\zeta - 1)^{i_0}$ yields

$$c_{i_0} \in (\zeta - 1)\mathscr{O}_K \cap \mathbf{Z} = p\mathbf{Z}.$$

□

We have finished the proof of Theorem 10.1.

Corollary 10.7. *The ideal $(1 - \zeta) \subset \mathscr{O}_K = \mathbf{Z}[\zeta]$ containing p is prime and is the unique prime ideal containing p.*

Proof. To see primality, we compute that the quotient by this ideal is a domain:

$$\mathbf{Z}[\zeta]/(\zeta - 1) = \mathbf{Z}[x]/(x - 1, \Phi_{p^r}(x)) = \mathbf{Z}/(\Phi_{p^r}(1)) = \mathbf{Z}/p\mathbf{Z}.$$

It remains to show $\mathscr{O}_K/p\mathscr{O}_K$ has a unique prime ideal. We have

$$\begin{aligned}\mathscr{O}_K/p\mathscr{O}_K &= \mathbf{Z}[x]/(p, \Phi_{p^r}(x)) \\ &= \mathbf{F}_p[x]/(\Phi_{p^r}).\end{aligned}$$

The identity $x^{p^r} - 1 = \Phi_{p^r}(x) \cdot (x^{p^{r-1}} - 1)$ in $\mathbf{Z}[x]$ reduces to the same identity in $\mathbf{F}_p[x]$, but in $\mathbf{F}_p[x]$ we have $x^{p^j} - 1 = (x-1)^{p^j}$ for any $j \geq 0$ (since the p-power map is additive in $\mathbf{F}_p[x]$), so we can cancel $(x-1)^{p^{r-1}}$ from both sides to get $\Phi_{p^r} = (x-1)^{p^r - p^{r-1}}$ in $\mathbf{F}_p[x]$ (not in $\mathbf{Z}[x]$!). In other words,

$$\mathscr{O}_K/p\mathscr{O}_K = \mathbf{F}_p[x]/(x-1)^{\deg \Phi_{p^r}},$$

so $x-1$ is nilpotent in this ring. It follows that in any domain quotient of $\mathscr{O}_K/(p)$, $x-1$ must vanish. Thus, all such domains are quotients of $\mathbf{F}_p[x]/(x-1) = \mathbf{F}_p$, which has only itself as a domain quotient. This says there is only one prime ideal over p (corresponding to $(p, \Phi_{p^r}(x), x-1) \subset \mathbf{Z}[x]$, or equivalently $(p, \zeta - 1) = (\zeta - 1)$ in $\mathbf{Z}[\zeta] = \mathscr{O}_K$). □

11. General cyclotomic integer rings

Today, we want to show that for all $N \geq 1$, we have $\mathscr{O}_{\mathbf{Q}(\zeta_N)} = \mathbf{Z}[\zeta_N]$. In Theorem 10.1, we settled the case that N is a prime power. As in that case, we next need to understand $\mathrm{Gal}(\mathbf{Q}(\zeta_N)/\mathbf{Q})$ in general:

Theorem 11.1 (Dedekind). *The injective homomorphism*

$$\begin{aligned}\mathrm{Gal}(\mathbf{Q}(\zeta_N)/\mathbf{Q}) &\to (\mathbf{Z}/N\mathbf{Z})^\times \\ \sigma &\mapsto a(\sigma)\end{aligned}$$

is an isomorphism, where $\sigma(\zeta) = \zeta^{a(\sigma)}$ for all Nth roots of unity ζ.

Remark 11.2. Note that $\Phi_N(x) := \prod_{\text{primitive}\,\zeta}(x - \zeta) \in \mathbf{Z}[x]$ because its coefficients lie in $\mathbf{Q}$ by Galois theory and are algebraic integers by inspection. The map $\mathrm{Gal}(\mathbf{Q}(\zeta_N)/\mathbf{Q}) \hookrightarrow (\mathbf{Z}/N\mathbf{Z})^\times$ is an isomorphism if and only if $\Phi_N(x)$ is irreducible over $\mathbf{Q}$, due to degree considerations since $\deg \Phi_N = \#(\mathbf{Z}/N\mathbf{Z})^\times$. Therefore, Theorem 11.1 is equivalent to $\Phi_N(x)$ being the minimal polynomial of ζ_N over $\mathbf{Q}$.

Proof of Theorem 11.1. Choose a primitive Nth root of unity ζ and let $f \in \mathbf{Q}[x]$ be its minimal polynomial over $\mathbf{Q}$, so $f \in \mathbf{Z}[x]$ and f is monic. It suffices to show $f(\zeta^e) = 0$ for all $e \in (\mathbf{Z}/N\mathbf{Z})^\times$ (so $[\mathbf{Q}(\zeta_N) : \mathbf{Q}] = \deg f \geq \phi(N)$, forcing equality since the reverse inequality is already known due to the injection of the Galois group into $(\mathbf{Z}/N\mathbf{Z})^\times$).

Observe that every element $e \in (\mathbf{Z}/N\mathbf{Z})^\times$ is represented by some $m > 1$. So, it is enough to treat the case $e = p \bmod N$ for $p \nmid N$, just using that every residue class is a product of primes. Here it is crucial that we are treating *all* primitive Nth roots of unity on an equal footing in our basic claim with a prime exponent, so that with a general exponent $m > 1$ written as a product $\prod p_j$ of primes we can iterate the prime-exponent claim: to handle $\zeta^m = (((\zeta^{p_1})^{p_2})^{p_3})^{\cdots}$ we use the primitive Nth root of unity ζ^{p_1} as a zero of f and the prime exponent p_2 to deduce that the primitive Nth root of unity $\zeta^{p_1 p_2}$ is also a root of f, and so on until we reach that ζ^m is a root of f as desired. (Note that we are *not* using the deep Dirichlet theorem on primes in arithmetic progressions here, just the very elementary fact that any $m > 1$ is a product of primes!)

Our aim is to show $f(\zeta^p) = 0$ for any prime $p \nmid N$, with f the minimal polynomial over $\mathbf{Q}$ for a primitive Nth root of unity ζ. Suppose this does not hold. Let $g \in \mathbf{Q}[x]$ be the minimal polynomial of ζ^p over $\mathbf{Q}$, so $g \in \mathbf{Z}[x]$ is monic and $g \neq f$. The monic irreducible g and f are both factors of $x^N - 1$ over $\mathbf{Q}$ (as each is the minimal polynomial of a root of $x^N - 1$, namely ζ^p and ζ respectively), so since $g \neq f$ we have $gf \mid x^N - 1 \in \mathbf{Q}[x]$. Observe that $gf \in \mathbf{Z}[x]$ and it is monic. We claim that in fact, the divisibility relation $gf \mid (x^N - 1)$ in $\mathbf{Q}[x]$ even holds in $\mathbf{Z}[x]$. This follows from:

Lemma 11.3. *Consider $h_1, h_2 \in A[x]$ with A a domain. Suppose $h_1 \mid h_2$ in $F[x]$ for the fraction field F of A. If h_1 is monic then $h_1 \mid h_2$ in $A[x]$.*

Proof. The point is that when one carries out long division to write

$$h_2 = h_1 q + r$$

for $q, r \in F[x]$ with $\deg(r) < \deg h_1$, the explicit division algorithm shows that if h_1 is monic then $q \in A[x]$ (so also $r = h_2 - h_1 q \in A[x]$), which is what we wanted to show. □

We have shown that $x^N - 1 = gf \cdot h$, for some $h \in \mathbf{Z}[x]$. Now it makes sense to reduce modulo p to obtain

$$x^N - 1 = \overline{g} \cdot \overline{f} \cdot \overline{h} \in \mathbf{F}_p[x]$$

(where $\overline{g}$ denotes g mod p and similarly for $\overline{f}$ and $\overline{h}$). Since $x^N - 1$ is *separable* over $\mathbf{F}_p$ because $p \nmid N$ (check!), its monic factors $\overline{g}$ and $\overline{f}$ cannot have any irreducible factor in common! Thus,

$$\gcd_{\mathbf{F}_p[x]}(\overline{g}, \overline{f}) = 1.$$

We shall now contradict this gcd calculation (so the original assumption that $f(\zeta^p) \neq 0$, which is to say $f \neq g$, is false).

Since g is the minimal polynomial over $\mathbf{Q}$ for ζ^p, so $g(\zeta^p) = 0$, we can say that the polynomial $g(x^p) \in \mathbf{Q}[x]$ vanishes at ζ. Therefore, $f \mid g(x^p)$ in

$\mathbf{Q}[x]$, so by Lemma 11.3 we have $f \mid g(x^p)$ in $\mathbf{Z}[x]$. Reducing this integral divisibility property modulo p, we obtain

$$\overline{f} \mid \overline{g(x^p)} = \overline{g}(x^p)$$

in $\mathbf{F}_p[x]$. But the p-power map on $\mathbf{F}_p[x]$ is a ring homomorphism with no effect on $\mathbf{F}_p$, so

$$\overline{g}(x^p) = \overline{g}(x)^p.$$

Thus, $\overline{f}|\overline{g}^p$ in $\mathbf{F}_p[x]$. But $\gcd_{\mathbf{F}_p[x]}(\overline{f},\overline{g}) = 1$ with $\deg \overline{f}, \deg \overline{g} > 0$, so we have reached a contradiction. □

Remark 11.4. Characteristic p is awesome! We just used characteristic p to deduce non-obvious facts in characteristic 0. The two worlds can communicate through $\mathbf{Z}$.

Corollary 11.5. *We have*

$$[\mathbf{Q}(\zeta_N) : \mathbf{Q}] = \phi(N).$$

Proof. We saw in Theorem 11.1 that $\mathbf{Q}(\zeta_N)/\mathbf{Q}$ is Galois with Galois group isomorphic to $(\mathbf{Z}/N\mathbf{Z})^\times$. Since the size of the Galois group for a Galois extension is the degree of the field extension, we are done. □

Theorem 11.6. *For any $N \geq 1$, let $K = \mathbf{Q}(\zeta_N)$.*

(1) $\mathscr{O}_{\mathbf{Q}(\zeta_N)} = \mathbf{Z}[\zeta_N]$,

(2) if N is not twice an odd integer then $p \mid \mathrm{disc}(K)$ if and only if $p \mid N$.

Remark 11.7. In the second part of Theorem 11.6, the avoidance of the case $N = 2n$ for odd n is very appropriate and harmless for the following reason: if n is odd then $-\zeta_n$ is a primitive $2n$th root of unity (think about this) yet $\mathbf{Z}[\zeta_n] = \mathbf{Z}[-\zeta_n]$. Thus, the field $\mathbf{Q}(\zeta_N)$ and the ring $\mathbf{Z}[\zeta_N]$ cannot intrinsically detect whether an odd N is replaced with $2N$. Hence, for any intrinsic property such as determining prime factors of the discriminant in terms of prime factors of N we cannot expect $2n$ and n to be treated on equal footing for odd n.

The case $N = p^r$ was done in Theorem 10.1. Note that (2) was also done since p^r is never twice an odd number when $p^r > 2$ (even if $p = 2$!).

We may now assume N has at least two prime factors and we induct on the number of prime factors. We may and do write $N = p^r N'$ for an *odd* prime p with $r > 0$, and $N' > 1$ for which $p \nmid N'$. By the preceding discussion we may and do assume N is not twice an odd integer. Since we arranged (as may) that p is odd, it is clear that N' also cannot be twice an odd integer (as N is not twice an odd integer).

We have $\phi(N) = \phi(p^r)\phi(N')$, so in the diagram of fields

$$\begin{tikzcd}
& \mathbf{Q}(\zeta_N) & \\
\mathbf{Q}(\zeta_{p^r}) \arrow[ur] & & \mathbf{Q}(\zeta_{N'}) \arrow[ul] \\
& \mathbf{Q} \arrow[ul] \arrow[ur] &
\end{tikzcd} \tag{11.1}$$

the total degree is the product of the degrees of the middle two fields. We will exploit this after first addressing some interaction between those fields, as a special case of:

Lemma 11.8. *If* $\gcd(a, b) = 1$*, then* $\mathbf{Q}(\zeta_a)\mathbf{Q}(\zeta_b) = \mathbf{Q}(\zeta_{ab})$ *inside* $\mathbf{Q}(\zeta_{ab})$*. More specifically,*

$$\mathbf{Z}[\zeta_a]\mathbf{Z}[\zeta_b] = \mathbf{Z}[\zeta_{ab}]$$

where by definition

$$\mathbf{Z}[\alpha] \cdot \mathbf{Z}[\beta] := \{\sum f_i g_j \mid f_i \in \mathbf{Z}[\alpha], g_j \in \mathbf{Z}[\beta]\} = \mathbf{Z}[\alpha, \beta].$$

Proof. As motivation, inside $\mathbf{C}$ we have $e^{2\pi i/ab} = (e^{2\pi i/b})^x (e^{2\pi i/a})^y$ where $1 = ax + by$ with $x, y \in \mathbf{Z}$ (ensuring $1/(ab) = x/b + y/a$). The proof can then be done purely algebraically as follows: we may write $ax + by = 1$ for some $x, y \in \mathbf{Z}$, so $1/(ab) = x/b + y/a$ and one then checks directly by purely algebraic means (exercise!) that $\zeta_b^x \zeta_a^y$ is a primitive abth root of unity. □

To prove Theorem 11.6, it suffices to prove:

Theorem 11.9. *For number fields* $K, K'/\mathbf{Q}$ *with* $F = KK'$ *a compositum and* $m := [K : \mathbf{Q}]$ *and* $m' := [K' : \mathbf{Q}]$*, suppose in the field diagram*

$$\begin{tikzcd}
& F & \\
K \arrow[ur] & & K' \arrow[ul] \\
& \mathbf{Q} \arrow[ul] \arrow[ur] &
\end{tikzcd} \tag{11.2}$$

we have $[F : K'] = m$ *and* $[F : K] = m'$*; equivalently, we assume*

$$[F : \mathbf{Q}] = [K : \mathbf{Q}][K' : \mathbf{Q}].$$

If also $\gcd(\operatorname{disc} K, \operatorname{disc} K') = 1$ *then*

(1) $\mathscr{O}_K\mathscr{O}_{K'} = \mathscr{O}_F$,

(2) $\operatorname{disc} F = (\operatorname{disc} K)^{m'} \cdot (\operatorname{disc} K')^m$.

Observe that this two-part result really does enable us to push through the induction with p^r as m and N' as m' due to the nature of our inductive hypothesis (carrying along the information of the prime factors of the discriminant alongside the determination of the cyclotomic integer ring).

Warning 11.10. The condition $\gcd(\operatorname{disc} K, \operatorname{disc} K') = 1$ cannot be dropped. For example, if $K = \mathbf{Q}(\sqrt{-2n})$ and $K' = \mathbf{Q}(\sqrt{2n})$ for squarefree odd $n > 1$ then $[KK' : \mathbf{Q}] = 4$, $\mathscr{O}_K = \mathbf{Z}[\sqrt{-2n}]$, and $\mathscr{O}_{K'} = \mathbf{Z}[\sqrt{2n}]$, so

$$\mathscr{O}_K\mathscr{O}_{K'} = \{a_0 + a_1\sqrt{2n} + a_2\sqrt{-2n} + a_3 \cdot 2n\sqrt{-1} \mid a_j \in \mathbf{Z}\}.$$

Hence, $\sqrt{-1} \notin \mathscr{O}_K\mathscr{O}_{K'}$ (due to evenness of the coefficient of $\sqrt{-1}$ in the above description) yet clearly $\sqrt{-1} \in KK'$ and hence $\sqrt{-1} \in \mathscr{O}_{KK'}$.

The crucial observation to prove Theorem 11.9 is that hypothesis on the field degrees implies

$$\operatorname{Tr}_{F/K'}|_K = \operatorname{Tr}_{K/\mathbf{Q}}.$$

To see this equality, note that if $\{e_i\}$ is a $\mathbf{Q}$-basis of K then

$$F = KK' = \sum_{i=1}^{m} K'e_i$$

with $m = [F : K']$, so the K'-linear spanning set $\{e_i\}$ for F must be a K'-basis for F. This ensures that for any for $\alpha \in K$, the K'-linear multiplication map $m_\alpha : F \to F$ has the *same* matrix with respect to the K'-basis $\{e_i\}$ of F as the $\mathbf{Q}$-linear $m_\alpha : K \to K$ does with respect to the $\mathbf{Q}$-basis $\{e_i\}$ of K; the traces of these matrices therefore agree, and this is exactly the statement that $\operatorname{Tr}_{F/K'}|_K = \operatorname{Tr}_{K/\mathbf{Q}}$.

For the proof of (2) via the fact that $\operatorname{Tr}_{F/K'}|_K = \operatorname{Tr}_{K/\mathbf{Q}}$, see the handout “Discriminant of Composite Fields”; this uses (1) and amounts to some slightly tedious but ultimately rather concrete arguments in linear algebra with matrices and determinants, using that the collection of $mm' = [F : \mathbf{Q}]$ elements $\{e_ie'_j\}$ is a $\mathbf{Z}$-basis of $\mathscr{O}_K\mathscr{O}_{K'}$ for $\{e_i\}$ a $\mathbf{Z}$-basis of $\mathscr{O}_K$ and $\{e'_j\}$ a $\mathbf{Z}$-basis of $\mathscr{O}_{K'}$.

To prove (1), we have $\mathscr{O}_K\mathscr{O}_{K'} \subset \mathscr{O}_F$ and we want to show the index divides $\operatorname{disc} K$ and $\operatorname{disc} K'$. Once this is shown, then since these two discriminants have gcd equal to 1 it will follow that the index of $\mathscr{O}_K\mathscr{O}_{K'}$ inside $\mathscr{O}_F$ is 1, which is to say $\mathscr{O}_K\mathscr{O}_{K'} = \mathscr{O}_F$. Indeed, the finite index statement follows from Lemma 12.2 and the divisibility statement follows from Lemma 12.3.

The key is to study the (compatible!!) trace maps

$$\begin{tikzcd}\mathscr{O}_K \arrow[r] \arrow[d] & \mathscr{O}_F \arrow[d] \\ \mathbf{Z} \arrow[r] & \mathscr{O}_{K'}\end{tikzcd} \tag{11.3}$$

and to adapt over the base ring $\mathscr{O}_{K'}$ our earlier observations for using discriminants to control the index for subrings generated over the base ring $\mathbf{Z}$. (In effect, this will be an argument in "relative" algebraic number theory, studying F/K' via analogues of arguments seen earlier for studying $K/\mathbf{Q}$.) We will give the proof of (1) at the start of next time.

12. Noetherian rings and modules

Remark 12.1. For some interesting exercises on rings of integers, look at the Exercises for Chapter 2 of the book *Number Fields* by Daniel Marcus.

Let's now finish up a loose end from last time: proving $\mathscr{O}_K\mathscr{O}_{K'} = \mathscr{O}_F$ when a compositum $F = KK'$ of two number fields has degree $[K : \mathbf{Q}][K' : \mathbf{Q}]$ and the discriminants of K and K' are coprime. Let $m = [K : \mathbf{Q}]$, $m' = [K' : \mathbf{Q}]$. As a preliminary step, we show:

Lemma 12.2. *The subring $\mathscr{O}_K\mathscr{O}_{K'} \subset \mathscr{O}_F$ has finite index.*

Proof. Let $\{e_i\}$ be a $\mathbf{Z}$-basis of $\mathscr{O}_K$ and $\{e'_j\}$ a $\mathbf{Z}$-basis of $\mathscr{O}_{K'}$, so

$$\mathscr{O}_K\mathscr{O}_{K'} := \{\sum_r \alpha_r\alpha'_r \mid \alpha_r \in \mathscr{O}_K, \alpha'_r \in \mathscr{O}_{K'}\}$$

is thereby seen to be the $\mathbf{Z}$-span of $\{e_ie'_j\}$ (why?).

Since the mm' vectors $e_ie'_j$ $(1 \leq i \leq m, 1 \leq j \leq m')$ span $F = KK'$ over $\mathbf{Q}$ and $[F : \mathbf{Q}] = mm'$, it follows that $\mathscr{O}_K\mathscr{O}_{K'}$ is a $\mathbf{Z}$-lattice (i.e., $\mathbf{Z}$-span of a $\mathbf{Q}$-basis) in the $\mathbf{Q}$-vector space F. But we saw earlier that any inclusion of $\mathbf{Z}$-lattices in the finite-dimensional $\mathbf{Q}$-vector space has finite index, so $\mathscr{O}_K\mathscr{O}_{K'} \subset \mathscr{O}_F$ has finite index. □

It now suffices to show:

Lemma 12.3. *The finite abelian quotient group $\mathscr{O}_F/\mathscr{O}_K\mathscr{O}_{K'}$ is killed by multiplication by both* $\operatorname{disc} K$ *and* $\operatorname{disc} K'$.

Since $\gcd(\operatorname{disc} K, \operatorname{disc} K') = 1$, this implies that $\mathscr{O}_F/\mathscr{O}_K\mathscr{O}_{K'} = 0$, which is to say $\mathscr{O}_K\mathscr{O}_{K'} = \mathscr{O}_F$ as desired.

Proof. This is symmetric in K and K', so it suffices to show $\mathscr{O}_F/\mathscr{O}_K\mathscr{O}_{K'}$ is killed by $\operatorname{disc} K$. Our task amounts to showing that for any $\alpha \in \mathscr{O}_F$,

$$\alpha \in \frac{1}{\operatorname{disc} K} \cdot \mathscr{O}_K\mathscr{O}_{K'}.$$

We have

$$F = K \cdot K' = \sum_i K'e_i,$$

for any $\mathbf{Q}$-basis $\{e_i\}$ of K. But $[F : K'] = [K : \mathbf{Q}]$, so the K'-spanning set $\{e_i\}$ of F has size $[F : K']$ and hence is also a K'-basis. Thus, for $\alpha \in \mathscr{O}_F$ we may uniquely write

$$\alpha = \sum_i c_i' e_i$$

for $c_i' \in K'$. It suffices to show

$$c_i' \in \frac{1}{\operatorname{disc} K}\mathscr{O}_{K'}.$$

Inspired by the calculations we carried out to control denominators in our proof of $\mathbf{Z}$-module finiteness of rings of integers, we obtain in exactly the same way the equality

$$\begin{pmatrix} \operatorname{Tr}_{F/K'}(\alpha e_1) \\ \vdots \\ \operatorname{Tr}_{F/K'}(\alpha e_m) \end{pmatrix} = (\operatorname{Tr}_{F/K'}(e_ie_j)) \begin{pmatrix} c_1' \\ \vdots \\ c_m' \end{pmatrix}.$$

The left side belongs to ${K'}^n$, but in Exercise 0 of Homework 3 it was shown that $\operatorname{Tr}_{F/K'}$ carries $\mathscr{O}_F$ into $\mathscr{O}_{K'}$, so the left side actually belongs to $\mathscr{O}_{K'}^n$ (this is where we use the hypothesis that $\alpha \in \mathscr{O}_F$, which ensures each αe_j also belongs to $\mathscr{O}_F$!). We want to control "denominators" for the elements $c_j' \in K'$, so we wish to invert the matrix of traces in this equality.

The crucial point is that the trace $\operatorname{Tr}_{F/K'}(e_ie_j)$ coincides with $\operatorname{Tr}_{K/\mathbf{Q}}(e_ie_j)$ since $\operatorname{Tr}_{F/K'}|_K = \operatorname{Tr}_{K/\mathbf{Q}}$ (as was deduced earlier from the hypothesis on field degrees). Thus, $(\operatorname{Tr}_{F/K'}(e_ie_j)) = (\operatorname{Tr}_{K/\mathbf{Q}}(e_ie_j))$ is an integer matrix with determinant $\operatorname{disc}(K/\mathbf{Q})$. Hence, by Cramer's Formula, the inverse matrix is $1/\operatorname{disc}(K/\mathbf{Q})$ times an integer matrix, so

$$\begin{pmatrix} c_1' \\ \vdots \\ c_m' \end{pmatrix} \in \frac{1}{\operatorname{disc} K} M \cdot \begin{pmatrix} \mathscr{O}_{K'} \\ \vdots \\ \mathscr{O}_{K'} \end{pmatrix}$$

where M is a $\mathbf{Z}$-matrix. Therefore, $\alpha = \sum c_i' e_i \in \frac{1}{\operatorname{disc} K} \cdot \mathscr{O}_{K'}$, as desired. □

The determination of cyclotomic integer rings is done, and we move on to spend a few lectures setting up the basics of Dedekind domains. As motivation, we now highlight three key properties of $\mathscr{O}_K$. Here are the first two:

(1) The ring $\mathscr{O}_K$ is an integrally closed domain (see Homework 3)
(2) All nonzero prime ideals of $\mathscr{O}_K$ are maximal (proved earlier).

In fact, there are lots of maximal ideals in any ring of integers, as follows from:

Proposition 12.4. *For any prime $p \in \mathbf{Z}^+$, there exists a maximal ideal $\mathfrak{m} \subset \mathscr{O}_K$ satisfying $\mathfrak{m} \cap \mathbf{Z} = p\mathbf{Z}$.*

Remark 12.5. The property $\mathfrak{m} \cap \mathbf{Z} = p\mathbf{Z}$ is equivalent to the condition that the finite field $\mathscr{O}_K/\mathfrak{m}$ contains $\mathbf{F}_p$ (i.e., has characteristic p).

We have previously verified this proposition by hand when $[K : \mathbf{Q}] = 2$ by using the explicit description of quadratic integer rings; a more conceptual method is required for the general case. Later it will be see that for each p, the non-empty collection of maximal ideals as in this Proposition has size at most $[K : \mathbf{Q}]$, as we have also seen directly in the quadratic case.

Proof. We just need to show that $\mathscr{O}_K/p\mathscr{O}_K$ has some maximal ideal. Recall that maximal ideals are by definition maximal with respect to containment among *proper* ideals.

As a $\mathbf{Z}$-module we have $\mathscr{O}_K \simeq \mathbf{Z}^{\oplus n}$ with $n = [K : \mathbf{Q}]$, so

$$\mathscr{O}_K/p\mathscr{O}_K \simeq \mathbf{F}_p^n$$

is a nonzero finite-dimensional $\mathbf{F}_p$-vector space. Note that any ideal is an $\mathbf{F}_p$-subspace.

To prove a maximal ideal I of $\mathscr{O}_K/p\mathscr{O}_K$ exists, start with any proper ideal I_1 (e.g., (0)). If I_1 is not maximal, choose a proper ideal I_2 strictly containing I_1. If I_2 is not maximal, choose a proper ideal I_3 strictly containing I_2. This process *cannot* continue forever, since otherwise this would be a sequence of strictly increasing ideals

$$I_1 \subsetneq I_2 \subsetneq I_3 \cdots$$

in the *n-dimensional* $\mathbf{F}_p$-vector space $\mathscr{O}_K/p \simeq \mathbf{F}_p^n$. We even see that this process must stop within n steps.

The termination implies that we have found a maximal ideal in $\mathscr{O}_K/p\mathscr{O}_K$; its preimage in $\mathscr{O}_K$ is a maximal ideal of $\mathscr{O}_K$ that contains p. □

The third key property of $\mathscr{O}_K$ rests on the following concept:

Definition 12.6. A module M over a commutative ring A satisfies the *ascending chain condition* (abbreviated *acc*) if for any sequence of submodules

$$M_1 \subset M_2 \subset M_3 \subset \cdots$$

of M there is some n so that $M_n = M_{n+1} = M_{n+2} = \cdots$; i.e., the chain of submodules eventually stabilizes.

Definition 12.7. A module M over a commutative ring A is called *noetherian* if it satisfies the ascending chain condition.

Lemma 12.8. *A module M over a ring A is noetherian if and only if all A-submodules of M are finitely generated.*

Before proving the lemma, here is a crucial example:

Example 12.9. A ring A is noetherian as an A-module if and only if all ideals are finitely generated, since the A-submodules of A are precisely the ideals of A.

Proof. First we show that if M is noetherian then all submodules are finitely generated. Let $N \subset M$ be a submodule. Suppose N is not finitely generated. Pick $n_1 \in N$ (which exists as we can take $n_1 = 0$). Take $M_1 := An_1 \subset N$. Since N is not finitely generated, necessarily $M_1 \neq N$. Then we can pick $n_2 \in N - M_1$. Take

$$M_2 := M_1 + An_2 \subseteq N.$$

Since N is not finitely generated, necessarily $M_2 \neq N$, so we can pick $n_3 \in N - M_2$.

Proceeding in this way, we obtain an infinite ascending chain

$$M_1 \subsetneq M_2 \subsetneq \cdots .$$

This violates acc, so N has to be finitely generated after all!

We now show the converse implication. Suppose all submodules of M are finitely generated. Consider some ascending chain $M_1 \subset M_2 \subset \cdots$. We want show this sequence stabilizes. Let

$$N := \cup_{i \geq 0} M_i.$$

This is an A-submodule of M due to the "ascending chain" property (a union of submodules is usually not a submodule, much like a union of subspaces of a vector space is typically not a subspace; the containment relation $M_i \subset M_{i+1}$ for all i is what ensures that any A-linear combination of *finitely many* elements of N again belongs to N inside M).

By hypothesis, N is generated by finitely many elements, say $n_1, \ldots, n_r$ But then, for each j there is some $k(j)$ with $n_j \in M_{k(j)}$. Taking t to be the maximum over all j of the $k(j)$'s, we see that $n_j \in M_t$ for all j, so $N \subset M_t$.

Thus, for $t' \geq t$ we have $M_t \subset M_{t'} \subset N = M_t$, so $M_{t'} = M_t$, for all $t' \geq t$; this is the stabilization condition we wanted. □

Example 12.10 (Non-example)**.** Consider

$$\begin{aligned} A &:= \mathbf{Q}[x_1, x_2, \ldots] \\ &= \cup_{n \geq 1} \mathbf{Q}[x_1, \ldots, x_n]. \end{aligned}$$

The ideal $(x_1, x_2, \ldots)$ of A (polynomial expressions in finitely many x_i's with vanishing constant term) is not finitely generated. To see this, the sequence of ideals $M_j = (x_1, \ldots, x_j)$ in A violates acc (check!).

Example 12.11. Any principal ideal domain is a noetherian ring because every ideal is generated by a single element by definition of "PID".

The robustness of the noetherian condition is due to the second part of:

Theorem 12.12.

(1) *If M is a module over a commutative ring A and $M' \subset M$ is a submodule, then M is a noetherian A-module if and only if M' and M/M' are noetherian A-modules.*

(2) *If A is a noetherian ring, then all finitely generated A-modules are noetherian as A-modules.*

Example 12.13. To illustrate the utility of this theorem, consider $A \to B$ is a module finite ring map with A a noetherian ring, such as with A a PID; a key case for us is $\mathbf{Z} \to \mathscr{O}_K$. In such cases, we claim B is always a noetherian ring (so $\mathscr{O}_K$ is always noetherian; i.e., its ideals are finitely generated).

The key point is that ideals of B are A-submodules. Part (2) of Theorem 12.12 implies that B is a noetherian A-module, so acc holds for chains of A-submodules of B. But any chain of ideals of B is also a chain of A-submodules and hence stabilizes! This says that B satisfies acc for chains of its own ideals, which is to say B is a noetherian B-module. This means by definition that B is a noetherian ring (so all ideals of B are finitely generated!).

Proof. The proof of (1) is [Samuel, Ch. III, §3.1, Prop. 1]. Let's now see why (2) follows from (1). Let M be a finitely generated A-module, so it has some finite generating set: we can write $M = \sum_{j=1}^r Am_j$. We shall induct on r.

First, we do the inductive step. Suppose $r > 1$, and let $M' = \sum_{j=1}^{r-1} Am_j$. Then, M/M' is generated by the image of m_r. By the inductive hypothesis, we know M' is noetherian, so by (1) the noetherian property for M is reduced to that of the A-module M/M' generated by a single element.

It remains to treat the case $r = 1$. But then we have a surjection

$$\begin{aligned} A &\to M \\ a &\mapsto am, \end{aligned}$$

and we are assuming A is a noetherian A-module. It then follows from (1) that the quotient A-module M is also noetherian, so the base case that $r = 1$ is settled. □

13. Dedekind domains

Last time we saw that noetherian rings are equivalently characterized by the acc property for their ideals and by the property of all ideals being finitely generated, and moreover that any ring module-finite over a noetherian ring is itself a noetherian ring. In particular, any ring module-finite over a PID is noetherian. This is how we deduced $\mathscr{O}_K$ is noetherian for any number field K. The following definition encapsulates the key properties of $\mathscr{O}_K$ that were reviewed in the previous lecture:

Definition 13.1. A *Dedekind domain* is a domain A that is not a field such that the following three conditions hold:

(1) It is noetherian.
(2) It is integrally closed (in its fraction field).
(3) Every nonzero prime ideal $\mathfrak{p} \subset A$ is maximal.

In the presence of (3), the condition that the domain A is not a field is equivalent to the requirement that A admits a nonzero prime ideal. Indeed, in a noetherian ring every proper ideal is contained in a maximal ideal (as any counterexample would yield a strictly increasing infinite sequence of ideals, contradicting acc), so once A contains a nonzero ideal – as always occurs when A is not a field, by considering the principal ideal generated by a nonzero non-unit – it must contain a nonzero maximal ideal.

Remark 13.2. From the viewpoint of algebraic geometry, condition (3) is a "1-dimensionality" property on A. A huge supply of Dedekind domains arises as follows. Let k be a perfect field, and $f \in k[x,y]$ irreducible. (Note that $k[x,y] = (k[x])[y]$ is a UFD since $R[y]$ is a UFD whenever R is a UFD.)

The domain $k[x,y]/(f)$ is Dedekind if and only if f satisfies the "smoothness" condition that whenever x_0, y_0 are algebraic over k and $f(x_0, y_0) = 0$ then the gradient of f is nonzero at (x_0, y_0) (i.e., at least one of $\partial f/\partial x$ or $\partial f/\partial y$ is nonzero at (x_0, y_0)). This is part of the theory of algebraic curves (the smoothness condition turns out to encode exactly integral closedness; it is why k needs to be perfect). For example, if $h \in k[x]$ is squarefree nonconstant and $\mathrm{char}(k) \neq 2$ then $k[x,y]/(y^2 - h(x))$ is "smooth".

In addition to the preceding examples via smooth algebraic curves over perfect fields, the other classes of Dedekind domains that we have encountered are PID's and $\mathscr{O}_K$ for a number field K.

To give some non-examples, $\mathbf{Z}[\sqrt{-3}]$ is noetherian (due to being $\mathbf{Z}$-finite) and its nonzero prime ideals are maximal (as the proof of this property for $\mathscr{O}_K$ only used that $\mathscr{O}_K$ is $\mathbf{Z}$-finite, nothing more) but it is not integrally closed (e.g., $\zeta_3 \in \mathbf{Q}(\sqrt{-3})$ is integral over $\mathbf{Z}$ and hence over $\mathbf{Z}[\sqrt{-3}]$ but does not belong to $\mathbf{Z}[\sqrt{-3}]$). A bit more exotic is the ring

$$\overline{\mathbf{Z}} := \{a \in \mathbf{C} \mid a \text{ is integral over } \mathbf{Z}\}$$

of "all" algebraic integers; this satisfies (2) and (3) in the definition of a Dedekind domain (the proof of (3) requires some thought) but it is not noetherian (for reasons we won't get into).

Although Dedekind domains were discovered in the context of studying rings of integers of number fields, Dedekind domains in general really constitute a topic in commutative ring theory. The number theory aspect is through the special finiteness properties that are satisfied by $\mathscr{O}_K$'s but not by general Dedekind domains, such as: (i) the field $\mathscr{O}_K/\mathfrak{m}$ is finite for all maximal ideals $\mathfrak{m}$, (ii) the unit group $\mathscr{O}_K^\times$ is finitely generated. (We will discuss the proof of (ii) at the very end of the course; it is the *Dirichlet Unit Theorem*, a vast generalization of Pell's Equation for the context of real quadratic fields.) For the Dedekind domain $A = \mathbf{C}[x,y]/(y^2 - (x^3 - x))$, both (i) and (ii) fail: it turns out that $A/\mathfrak{m} = \mathbf{C}$ for all $\mathfrak{m}$ and $A^\times = \mathbf{C}^\times$ (an uncountable abelian group, so very far from finitely generated), though neither of these features of this A is evident without some knowledge about algebraic curves.

Later we will see that for Dedekind domains, the PID property is equivalent to the UFD property. The failure of the UFD property for Dedekind domains will always be governed in a precise sense by a specific associated abelian group called its *class group*. The finiteness of class groups for the Dedekind domains $\mathscr{O}_K$ will be another of the big theorems of this course, with nice applications to Diophantine questions.

Remark 13.3. Dedekind's definition of "Dedekind domain" is not the one we have introduced; he took it to mean "domain not a field for which nonzero ideals are uniquely a finite product of maximal ideals". We will show that this property of the nonzero ideals is a consequence of the definition we have given (that was introduced by Noether as part of her structural revolution for commutative ring theory in the 1930's). It turns out that conversely Dedekind's definition does imply the definition we are using, but this is now only of historical interest since the definition we are using is without any doubt the correct one upon which to base the theory.

Let us record the main result we aim to discuss for the rest of today and all of next time:

Theorem 13.4. *If A is a Dedekind domain then every nonzero proper ideal $\mathfrak{a}$ of A has the form $\prod_{j=1}^{r} \mathfrak{p}_j$ for maximal ideals $\mathfrak{p}_j$ unique up to rearrangement.*

In the special case that A is a PID, so the maximal ideals are precisely (π) for irreducible $\pi \in A$, this theorem expresses exactly the UFD property because the use of principal ideals elegantly absorbs away the intervention of units. The proof of the theorem will be non-constructive, but in the special case $A = \mathscr{O}_K = \mathbf{Z}[\alpha]$ we will see how to find the $\mathfrak{p}_j$'s in the case $\mathfrak{a} = p\mathscr{O}_K$ for prime $p \in \mathbf{Z}^+$ (a case that is most important for subsequent developments; recall that our study of general factorization in $\mathbf{Z}[i]$ had as its key ingredient the knowledge of how to factor $p\mathbf{Z}[i]$ for prime $p \in \mathbf{Z}^+$).

The first step in the proof of Theorem 13.4 is a striking application of the noetherian property (not using integral closedness):

Lemma 13.5. *For any nonzero ideal $\mathfrak{a}$ in a Dedekind domain A, there is a finite collection of maximal ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_m$ such that $\prod \mathfrak{p}_j \subset \mathfrak{a}$.*

In the context of our desired result for Dedekind A, one should regard this collection of $\mathfrak{p}_j$'s as overkill: we will need to seek out some subcollection whose product is exactly equal to $\mathfrak{a}$ (multiplying against extra maximal ideals makes the product smaller within A). So we will need a procedure to identify and remove excess $\mathfrak{p}_j$'s.

Proof. We suppose some $\mathfrak{a}$ violates the desired conclusion and we will get a contradiction. Such an $\mathfrak{a}$ cannot be maximal (otherwise it would be a singleton product of maximal ideals, so it wouldn't violate the desired conclusion). Being nonzero, it therefore cannot be prime (as all nonzero prime ideals of A are maximal, by definition of being Dedekind). Thus, there exist $x, y \in A - \mathfrak{a}$ such that $xy \in \mathfrak{a}$. Hence, the ideals

$$\mathfrak{a} + (x), \mathfrak{a} + (y)$$

each strictly contain $\mathfrak{a}$ yet their product is contained in $\mathfrak{a}$:

$$(\mathfrak{a} + (x))(\mathfrak{a} + (y)) = \mathfrak{a}^2 + x\mathfrak{a} + y\mathfrak{a} + (xy) \subset \mathfrak{a}$$

(using crucially here that $xy \in \mathfrak{a}$).

By hypothesis $\mathfrak{a}$ doesn't contain a finite product of maximal ideals, so it follows that at least one of $\mathfrak{a} + (x)$ or $\mathfrak{a} + (y)$ cannot contain any such product (if each did contain such a product then so would $(\mathfrak{a} + (x))(\mathfrak{a} + (y))$, an ideal we have seen is contained in $\mathfrak{a}$, thereby contradicting the assumption that $\mathfrak{a}$ does not contain any finite product of maximal ideals). So let $\mathfrak{a}_1$ be either of these two that doesn't contain any finite product of maximal ideals.

To summarize, we have shown that if there is a nonzero ideal $\mathfrak{a}$ not containing any finite product of maximal ideals then $\mathfrak{a}$ is strictly contained in an ideal $\mathfrak{a}_1$ (so necessarily $\mathfrak{a}_1 \neq (0)$) such that $\mathfrak{a}_1$ also doesn't contain any finite product of maximal ideals. But we can then iterate this process forever, arriving at a strictly increasing infinite chain of ideals

$$\mathfrak{a} \subsetneq \mathfrak{a}_1 \subsetneq \mathfrak{a}_2 \subsetneq \dots,$$

contradicting acc for A! Thus, no such $\mathfrak{a}$ can exist, so we win. □

Next, we introduce a construction that will be the ideal-theoretic analogue of forming the reciprocal of a nonzero element of a domain inside the fraction field: for any nonzero ideal I of a domain A with fraction field F, we define

$$\widetilde{I} := \{y \in F \,|\, yI \subset A\}.$$

Note that $\widetilde{I}$ is an A-submodule of F (why?), and clearly $A \subset \widetilde{I}$. Informally, the elements of $\widetilde{I}$ are characterized as those $y \in F$ admitting a description as a fraction a'/a where a can be any desired nonzero element of I (i.e., $I - \{0\}$ is contained in the set of "denominators" of $y \in F$); indeed, this exactly says $ya \in A$ for all $a \in I - \{0\}$ since $y = (ya)/a$ for any $a \in A - \{0\}$.

Example 13.6. In case $I = \alpha A$ is a nonzero principal ideal for a domain A with fraction field F, $\widetilde{I} = (1/\alpha)A$. Indeed, $y \in \widetilde{I}$ precisely when $y\alpha A \subset A$, which is equivalent to saying $y\alpha \in A$ (why?), and that in turn says exactly that $y \in (1/\alpha)A$.

Next time we will see that when A is noetherian, $\widetilde{I}$ is finitely generated as an A-module. These $\widetilde{I}$'s aren't generally ideals of A since they usually are not even contained in A, but in the Dedekind case they will be instances of a notion to be called "fractional ideal" (and provide a reasonable notion of "multiplicative inverse" for nonzero ideals beyond the principal case). The key lemma concerning this construction for our purposes, and which will be proved next time, is:

Lemma 13.7. *For a maximal ideal $\mathfrak{p}$ of a Dedekind domain A, the product*

$$\mathfrak{p} \cdot \widetilde{\mathfrak{p}} := \{ \sum_{\text{finite}} x_i y_i \,|\, x_i \in \mathfrak{p}, y_i \in \widetilde{\mathfrak{p}}\}$$

is equal to A.

Note that by the definition $\mathfrak{p}\widetilde{\mathfrak{p}}$ is an A-submodule of F, it is contained in A (by definition of $\widetilde{\mathfrak{p}}$), and it contains $\mathfrak{p}$ (since $1 \in \widetilde{\mathfrak{p}}$). Since an A-submodule of A is just an ideal of A by another name, it follows that $\mathfrak{p}\widetilde{\mathfrak{p}}$ is an ideal of A lying between A and $\mathfrak{p}$. Up to here, $\mathfrak{p}$ could have been any nonzero ideal of A. But $\mathfrak{p}$ is *maximal*, so the only ideals lying between $\mathfrak{p}$ and A are the obvious ones: $\mathfrak{p}$ and A.

The substance of the lemma is that the option $\mathfrak{p}\widetilde{\mathfrak{p}} = \mathfrak{p}$ cannot occur. Our proof of this will use *crucially* the integral closedness of Dedekind domains. Without integral closedness, this can really fail: Exercise 4 on HW5 will provide such an example with $A = \mathbf{Z}[\sqrt{5}]$ (not integrally closed).

Since we have some extra time today, let's make a digression to discuss a result that we will never need in this course but which illustrates most strikingly how robust the noetherian condition on rings really is. The following theorem was a tremendous breakthrough:

Theorem 13.8 (Hilbert Basis Theorem)**.** *If A is a noetherian ring then $A[x]$ is also a noetherian ring.*

We are not claiming $A[x]$ is a noetherian A-module: it never is (when $A \neq 0$)! Indeed, the A-submodules

$$M_d = \{f \in A[x] \mid \deg f \leq d\} = A + Ax + \cdots + Ax^d$$

of $A[x]$ violate acc (when $A \neq 0$).

To appreciate the importance of the Hilbert Basis Theorem, note that via the inductive recipe

$$A[X_1, \ldots, X_n] = (A[X_1, \ldots, X_{n-1}])[X_n]$$

it follows that $A[X_1, \ldots, X_n]$ is noetherian for all n when A is noetherian, and by passing to quotients modulo ideals it follows that for any ring map $A \rightarrow B$ with B finitely generated as an A-algebra (i.e., all elements of B are A-linear combinations of monomials in some fixed finite set of elements of B) necessarily B is noetherian whenever A is. This is really amazing: if $\alpha_1, \ldots, \alpha_n \in \mathbf{C}$ is any finite set of complex numbers then then subring $\mathbf{Z}[\alpha_1, \ldots, \alpha_n] \subset \mathbf{C}$ (a quotient of $\mathbf{Z}[X_1, \ldots, X_n]$) is a noetherian ring regardless of whatever algebraic relations may be satisfied among the α_j's over $\mathbf{Z}$. That is a very non-obvious fact!

Hilbert's method of proof is so non-constructive that it was very controversial at the time it was announced at the end of the 19th century (in the context of polynomial rings over $\mathbf{C}$, since the general concept of noetherian ring didn't yet exist).

Proof. Let $I \subset A[x]$ be an ideal. We want to show it is finitely generated. Hilbert's brilliant idea is to introduce an auxiliary ideal of A to control this: let $J \subset A$ be the ideal of leading coefficients of elements of I. That is, J is the set of $a \in A$ such that $a = \text{lead}(f)$ for some $f \in I$ (where it is understood that $\text{lead}(0) = 0$). To check that J really is an ideal of A the main issue is stability under addition when dealing with elements of $A[x]$ having different degrees: if $a = \text{lead}(f)$ with $n = \deg(f)$ and $b = \text{lead}(g)$ with $m = \deg(g)$

for $f, g \in I - \{0\}$ then

$$a + b = \text{lead}(X^m f + X^n g)$$

with $X^m f + X^n g \in I$, so J is indeed stable under addition.

By the noetherian property of A, the ideal J is finitely generated. We may assume $I \neq (0)$, so $J \neq (0)$, and so pick a finite generating set $a_1, \ldots, a_r$ of J with $a_i \in A - \{0\}$. Each a_i is the leading coefficient of some nonzero $f_i \in I$. Let $d_i = \deg(f_i)$ and $d = \max_i d_i$.

For any nonzero $f \in A[x]$ with $\deg(f) \geq d$, the leading coefficient $a \in A - \{0\}$ of f belongs to J, so $a = \sum b_i a_i$ for some $b_i \in A$. Hence, the difference

$$f - \sum_i b_i X^{\deg(f) - d_i} f_i$$

has its $X^{\deg(f)}$-coefficient cancelling out, due to which this difference has degree at most $\deg(f) - 1$. In other words, by subtracting from f a suitable element of $(f_1, \ldots, f_r) \subset I$ we can successively knock down the degree of f until we reach degree $< d$.

So $I = (f_1, \ldots, f_r) + (I \cap A[x]_{<d})$, where

$$A[x]_{<d} := \sum_{i=0}^{d-1} A x^i.$$

Since $A[x]_{<d}$ is a finitely generated A-module and A is noetherian, the A-submodule $I \cap A[x]_{<d}$ is finitely generated! If $\{h_1, \ldots, h_s\}$ is an A-spanning set for $I \cap A[x]_{<d}$, then we conclude that the inclusion

$$(f_1, \ldots, f_r, h_1, \ldots, h_s) \subset I$$

of ideals of $A[x]$ is an equality. □

14. Prime ideal factorization

Let A be a Dedekind domain (i.e., a noetherian integrally closed domain that is not a field and for which all nonzero prime ideals are maximal). Letting F be its fraction field, for any nonzero ideal I of A we defined last time

$$\widetilde{I} := \{y \in F \mid yI \subset A\} \subset F;$$

this is visibly an A-submodule of F and was described concretely in terms of an element of F admitting a fractional expression description with any desired element of $I - \{0\}$ as the denominator.

For any A-submodules M and N of F, we define their "product" to be

$$MN := \{\sum m_i n_i \mid m_i \in M, n_i \in N\}$$

(sums of finitely many products inside F); this is clearly an A-submodule of F and in case M and N are ideals of A (or equivalently are contained in

A) this coincides with the earlier notion of multiplication for ideals. Clearly $MN = NM$, and $AM = M$ (why?). As in the case of ideal multiplication, it is not difficult to check that this "product" operation is also associative: if $M, M', M'' \subset F$ are A-submodules then $(MM')M'' = M(M'M'')$. Thus, there is no need to worry about parentheses when we iterate this operation many times (as we shall do).

Later it will be seen that $\widetilde{I}$ plays the role of a "multiplicative inverse" to I in this sense of multiplication among A-submodules of F (in effect, $\widetilde{I}I = A$). For now, the key case we need is when I is a maximal ideal, as this will provide an ideal-theoretic replacement for the cancellation step in the proof of uniqueness of prime factorization in $\mathbf{Z}^+$. That special case was stated but not proved last time, and we now record it again (incorporating an additional finiteness property):

Lemma 14.1. *The A-module $\widetilde{I}$ is finitely generated for any nonzero ideal I of A, and for a maximal ideal $\mathfrak{p}$ of A the product*

$$\mathfrak{p} \cdot \widetilde{\mathfrak{p}} := \{ \sum_{\text{finite}} x_i y_i \,|\, x_i \in \mathfrak{p}, y_i \in \widetilde{\mathfrak{p}} \}$$

is equal to A.

Proof. By definition of $\widetilde{\mathfrak{p}}$, we have $\widetilde{\mathfrak{p}}\mathfrak{p} \subset A$ and this is an A-submodule, so it is an ideal. But $1 \in \widetilde{\mathfrak{p}}$ (why?), so $\mathfrak{p} \subset \widetilde{\mathfrak{p}}\mathfrak{p}$. Hence, $\widetilde{\mathfrak{p}}\mathfrak{p}$ is an ideal of A lying between A and the maximal ideal $\mathfrak{p}$. Thus, the only possibilities for $\widetilde{\mathfrak{p}}\mathfrak{p}$ are that it is equal to A or $\mathfrak{p}$; we will rule out the latter possibility and thereby conclude that $\widetilde{\mathfrak{p}}\mathfrak{p} = A$. In particular, we will *not* directly prove that $\widetilde{\mathfrak{p}}\mathfrak{p} = A$ (such as by exhibiting 1 in the form $\sum x_i y_i$ with $x_i \in \widetilde{\mathfrak{p}}$ and $y_i \in \mathfrak{p}$), but we will instead deduce this must be the case by ruling out the only other possibility consistent with $\mathfrak{p}$ being a maximal ideal of A.

We assume $\widetilde{\mathfrak{p}}\mathfrak{p} = \mathfrak{p}$ and seek a contradiction. It must be stressed that this step will have to use crucially that A is integrally closed: in HW5 it will be seen for a suitable maximal ideal $\mathfrak{p}$ of the non-Dedekind $\mathbf{Z}[\sqrt{5}]$ (it is not integrally closed!), the equality $\widetilde{\mathfrak{p}}\mathfrak{p} = \mathfrak{p}$ actually can happen. Since $\mathfrak{p} \subset \widetilde{\mathfrak{p}}\mathfrak{p}$, the hypothesis $\widetilde{\mathfrak{p}}\mathfrak{p} = \mathfrak{p}$ amounts to the reverse inclusion $\widetilde{\mathfrak{p}}\mathfrak{p} \subset \mathfrak{p}$, and it is from this that we shall deduce a contradiction. This inclusion is that statement that for any $\alpha \in \widetilde{\mathfrak{p}} \subset F$, the multiplication operator $m_\alpha : x \mapsto \alpha x$ on F carries $\mathfrak{p}$ into itself. Rather generally, we shall now prove:

Lemma 14.2. *For any nonzero ideal I of A, the only $\alpha \in F$ for which $m_\alpha(I) \subset I$ are the elements of A.*

Once this lemma is proved, it would follow in our situation that $\widetilde{\mathfrak{p}} \subset A$ (the reverse containment is obvious, by the way). Then to get a contradiction it would be sufficient to find an element $\beta \in \widetilde{\mathfrak{p}} - A$. We will construct such a β after we prove Lemma 14.2:

Proof. We shall use the following general construction: for any A-modules M and N, define $\mathrm{Hom}_A(M, N)$ to be the set of A-linear maps $T : M \to N$. We make $\mathrm{Hom}_A(M, N)$ into an A-module via pointwise operations on such T (for addition and A-multiplication). If M and N were free A-modules of finite rank then $\mathrm{Hom}_A(M, N)$ could be described via matrices with entries in A; in general freeness does not hold (such as with M and N non-principal ideals of A), so there is typically no matrix-theoretic way to describe such T: they simply are what they are, nothing more.

In the special case $M = N$ (which is what we will need), there is an additional operation on $\mathrm{Hom}_A(M, M)$, namely composition of such linear maps. The crucial feature, shown on Homework 5, is that for any noetherian ring R at all and any finitely generated R-modules M and N, the R-module $\mathrm{Hom}_R(M, N)$ is always *finitely generated*. (This is ultimately reduced via the magic of noetherianity to the case when M and N are free modules.)

Let's consider $\mathrm{Hom}_A(I, I)$ for a nonzero ideal I of the Dedekind domain A. As we have noted above, this is finitely generated. What are some elements of this A-module? If $\alpha \in F$ has the property that $m_\alpha(I) \subset I$ then $m_\alpha \in \mathrm{Hom}_A(I, I)$. We claim that all $T \in \mathrm{Hom}_A(I, I)$ arise in this way. More specifically, for any $a \in I - \{0\}$ we claim that the fraction $T(a)/a \in F$ is *independent* of a, in which case upon denoting it as λ_T we would have $T(a) = \lambda_T a$ for all $a \in I$ (even $a = 0$, for silly reasons); this says $T = m_{\lambda_T}$. To prove this independence of a, we simply compute that for any $a, b \in I - \{0\}$,

$$bT(a) = T(b \cdot a) = T(a \cdot b) = aT(b)$$

(using the A-linearity of $T : I \to I$ twice), so indeed $T(a)/a = T(b)/b$ as desired.

We may now define a map $\xi_I : \mathrm{Hom}_A(I, I) \to F$ by assigning to each T the scalar λ_T as made above. By inspection, if one adds two scalar multipliers then the effect is multiplication by the sum of the scalars, so ξ_I is additive. It is also easy to check that ξ_I is A-linear, and most importantly ξ_I carries composition over to multiplication in F: if we compose two scalar-multiplication operators, the net effect is multiplication by the product of those scalars.

Thus, via ξ_I we have made $\mathrm{Hom}_A(I, I)$ a *subring* of F (so $\mathrm{Hom}_A(I, I)$ has a commutative composition law); note that $\xi_I(\mathrm{id}_I) = 1$ by inspection, and ξ_I is injective since T is determined by λ_T by design ($T(x) = \lambda_T x$ for all

$x \in I$). This subring of F contains A, since certainly multiplication by any $a \in A$ carries I into itself.

But we have already noted that $\mathrm{Hom}_A(I, I)$ is module-finite over A, so (by A-linearity of ξ_I) it is a subring of F that is module-finite over A. We know that a module-finite extension of commutative rings is always integral, yet A is *integrally closed* in its fraction field by virtue of being Dedekind. Voila, so this forces $A = \mathrm{Hom}_A(I, I)$ inside F, which is to say that the *only* $\alpha \in F$ satisfying $\alpha I \subset I$ are the elements of A! This completes the proof of Lemma 14.2. □

The preceding method of argument also gives A-finiteness of $\widetilde{I}$ for any nonzero ideal I of A: for any $y \in \widetilde{I}$ we have $m_y : I \to A$ by definition of $\widetilde{I}$, so $y \mapsto m_y$ defines a map $\widetilde{I} \to \mathrm{Hom}_A(I, A)$ which is A-linear (check!) and also injective (since $m_y(1) = y$). Since $\mathrm{Hom}_A(I, A)$ is a finitely generated A-module (by Homework 5, due to A being noetherian), all of its A-submodules are also finitely generated (because A is noetherian) and hence $\widetilde{I}$ is finitely generated as an A-module!

The only remaining step to complete the proof of Lemma 14.1 is to find an element $\beta \in \widetilde{\mathfrak{p}}$ not belonging to A. For this we will use Lemma 13.5. As a warm-up, let's consider a special case:

Example 14.3. If $\mathfrak{p}$ is principal, say $\mathfrak{p} = (\alpha) = \alpha A$ for some $\alpha \in A$, necessarily $\alpha \neq 0$ and $\alpha \notin A^\times$ (why?), so we know $\widetilde{\mathfrak{p}} = (1/\alpha)A$. Thus, in such cases $\beta := 1/\alpha \in \widetilde{\mathfrak{p}} - A$.

In general of course $\mathfrak{p}$ isn't principal, but since elements of $\widetilde{\mathfrak{p}}$ (such as β that we seek to make) can be written as a fraction with *any* desired element of $\mathfrak{p} - \{0\}$ as the denominator, we pick a nonzero $\alpha \in \mathfrak{p}$ to serve as such a denominator and have by Lemma 13.5 that

$$\mathfrak{p} \supset (\alpha) \supset \mathfrak{p}_1 \cdots \mathfrak{p}_m$$

for some maximal ideals $\mathfrak{p}_1, \dots, \mathfrak{p}_m$. Let's consider such a collection of $\mathfrak{p}_j$'s with m as small as possible (relative to the fixed choice of α). If $m = 1$ then $\mathfrak{p} \supset (\alpha) \supset \mathfrak{p}_1$, but any containment among *maximal* ideals in a commutative ring is always an equality (why?), so this would force $\mathfrak{p} = \mathfrak{p}_1$ and by squeezing therefore $\mathfrak{p} = (\alpha)$ is principal, a case we have settled.

Now suppose this minimal m is larger than 1. Since the prime ideal $\mathfrak{p}$ contains the product of the ideals $\mathfrak{p}_j$, at least one of the $\mathfrak{p}_j$'s must be contained in $\mathfrak{p}$. Indeed, quite generally if a prime ideal P in a commutative ring contains a product $J_1 \cdots J_n$ of finitely many ideals $J_1, \dots, J_n$ then *some* J_i must be contained in P: if not then for each i we can pick $x_i \in J_i$ with $x_i \notin P$, but then $\prod x_i \in \prod J_i \subset P$, contradicting primality of P since $x_i \notin P$ for all i.

By rearrangement of the $\mathfrak{p}_j$'s we can suppose it is $\mathfrak{p}_1$ that is contained in $\mathfrak{p}$. But then $\mathfrak{p}_1 = \mathfrak{p}$ (again, a containment among maximal ideals must be an equality), so we have

$$\mathfrak{p} \supset (\alpha) \supset \mathfrak{p}\mathfrak{p}_2 \cdots \mathfrak{p}_m.$$

By the minimality of m, the product $\mathfrak{p}_2 \cdots \mathfrak{p}_m$ of $m-1$ maximal ideals *cannot* be contained in (α), so there exists $\alpha' \in \mathfrak{p}_2 \cdots \mathfrak{p}_m$ not lying in (α); i.e., $\alpha \nmid \alpha'$ in A, or in other words

$$\beta := \alpha'/\alpha \notin A.$$

We will show that β works, in the sense that $\beta \in \widetilde{\mathfrak{p}}$. That will finally finish the proof of Lemma 14.1. It has to be shown that $\beta\mathfrak{p} \subset A$. Using the definitions and recalling that $\mathfrak{p} = \mathfrak{p}_1$,

$$\beta\mathfrak{p} = (\alpha'/\alpha)\mathfrak{p} = (1/\alpha)(\alpha'\mathfrak{p}) \subset (1/\alpha)(\mathfrak{p}_2 \cdots \mathfrak{p}_m)\mathfrak{p} \subset (1/\alpha)(\alpha A) = A.$$

□

The hard work has been done, and now we reap the fruit of our labors, beginning with the uniqueness of prime ideal factorization:

Proposition 14.4. *If $\mathfrak{a}$ is a nonzero proper ideal of A and it is a product of maximal ideals in two ways*

$$\mathfrak{p}_1 \cdots \mathfrak{p}_m = \mathfrak{a} = \mathfrak{q}_1 \cdots \mathfrak{q}_n$$

then $m = n$ and the collection $\{\mathfrak{p}_i\}$ is a rearrangement of $\{\mathfrak{p}_j\}$.

Proof. We argue by induction on $\min(m,n) \geq 1$. First suppose this minimum is 1, so by relabeling $m = 1$:

$$\mathfrak{p}_1 = \mathfrak{a} = \mathfrak{q}_1 \cdots \mathfrak{q}_n.$$

Thus, $\mathfrak{p}_1 \subset \mathfrak{q}_j$ for all j, so $\mathfrak{p}_1 = \mathfrak{q}_j$ for all j (a containment among maximal ideals is an equality). This says $\mathfrak{p}_1 = \mathfrak{p}_1^n$, and we just need to check $n = 1$. If $n > 1$ then multiplying throughout by $\widetilde{\mathfrak{p}}_1$ yields $A = \mathfrak{p}_1^{n-1} \subset \mathfrak{p}_1$ (using $n - 1 > 0$ for the inclusion), a contradiction.

Now suppose $\min(m,n) \geq 2$. We have

$$\mathfrak{p}_1 \supset \mathfrak{a} = \mathfrak{q}_1 \cdots \mathfrak{q}_n,$$

so some $\mathfrak{q}_{j_0}$ is contained in $\mathfrak{p}_1$ (by primality of $\mathfrak{p}_1$). By rearranging, we can suppose $\mathfrak{q}_1 \subset \mathfrak{p}_1$ and so $\mathfrak{q}_1 = \mathfrak{p}_1$. Then multiplying throughout by $\widetilde{\mathfrak{p}}_1$ gives

$$\mathfrak{p}_2 \cdots \mathfrak{p}_m = \mathfrak{q}_2 \cdots \mathfrak{q}_n.$$

This equality of a product of $m-1$ and $n-1$ maximal ideals then permits us to use induction on $\min(m,n)$ to conclude. □

How about the *existence* of prime ideal factorization for nonzero proper ideals $\mathfrak{a}$ of A? We have $\mathfrak{a} \subset \mathfrak{p}$ for some maximal ideal $\mathfrak{p}$ of A (use acc to see this: if $\mathfrak{a}$ isn't maximal then $\mathfrak{a} \subsetneqq \mathfrak{a}'$ for a proper ideal $\mathfrak{a}'$, and if $\mathfrak{a}'$ isn't maximal then $\mathfrak{a}' \subsetneqq \mathfrak{a}''$ for a proper ideal $\mathfrak{a}''$, and so on; the process must reach a maximal ideal or else it violates acc). The key point is to relate such containment to divisibility in the sense of ideal multiplication:

Lemma 14.5. *If $\mathfrak{a}$ is a proper nonzero ideal of A and $\mathfrak{a} \subset \mathfrak{p}$ for maximal $\mathfrak{p}$ then $\mathfrak{a} = \mathfrak{p}\mathfrak{a}'$ for an ideal $\mathfrak{a}'$ of A that strictly contains $\mathfrak{a}$.*

The idea for the proof is that via the intuition of $\widetilde{\mathfrak{p}}$ behaving like an "inverse" to $\mathfrak{p}$, there is a natural guess for what $\mathfrak{a}'$ should be: we define $\mathfrak{a}' = \widetilde{\mathfrak{p}}\mathfrak{a}$ and hope for the best. The (visibly nonzero) A-module $\mathfrak{a}'$ inside F really is an ideal since

$$\mathfrak{a}' = \widetilde{\mathfrak{p}}\mathfrak{a} \subset \widetilde{\mathfrak{p}}\mathfrak{p} \subset A,$$

and in fact by Lemma 14.1 it "works":

$$\mathfrak{p}\mathfrak{a}' = \mathfrak{p}\widetilde{\mathfrak{p}}\mathfrak{a} = A\mathfrak{a} = \mathfrak{a}.$$

But the really important and less evident property is that $\mathfrak{a}'$ strictly contains $\mathfrak{a}$. Certainly $\mathfrak{a}' \supset \mathfrak{a}$ (since $\mathfrak{a} = \mathfrak{p}\mathfrak{a}' \subset \mathfrak{a}'$), so it amounts to showing $\mathfrak{a}' \neq \mathfrak{a}$. This requires work, explained in §3 of the handout "Unique Factorization in Dedekind Domains".

With Lemma 14.5 in hand, we get existence of prime ideal factorization via acc (in place of descending induction from the proof of existence of prime factorization in $\mathbf{Z}^+$) as follows. If $\mathfrak{a}' = A$, so $\mathfrak{a} = \mathfrak{p}$, then we are done. Suppose otherwise, so $\mathfrak{a}'$ is a nonzero proper ideal, and hence $\mathfrak{a}' \subset \mathfrak{p}'$ for some maximal ideal $\mathfrak{p}'$. Then by Lemma 14.5 we have $\mathfrak{a}' = \mathfrak{p}'\mathfrak{a}''$ for an ideal $\mathfrak{a}''$ strictly containing $\mathfrak{a}'$. Now $\mathfrak{a} = \mathfrak{p}\mathfrak{p}'\mathfrak{a}''$ with $\mathfrak{a} \subsetneqq \mathfrak{a}' \subsetneqq \mathfrak{a}''$. If $\mathfrak{a}'' = A$ we are done, and otherwise we keep repeating, and by acc the process eventually ends, giving an expression for $\mathfrak{a}$ as a finite product of maximal ideals.

In §4 of the handout "Unique Factorization in Dedekind Domains", some corollaries of prime ideal factorization are given, such as: (i) a Dedekind domain is a UFD if and only if it is a PID, (ii) a containment $\mathfrak{a} \subset \mathfrak{b}$ among nonzero ideals of A is equivalent to the divisibility $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$ for a nonzero ideal $\mathfrak{c}$ of A.

15. Norms of ideals

For a nonzero ideal $\mathfrak{a}$ in $\mathscr{O}_K$, we know that the quotient ring $\mathscr{O}_K/\mathfrak{a}$ is finite (as any inclusion between $\mathbf{Z}$-lattices in a finite-dimensional $\mathbf{Q}$-vector space has finite index), or equivalently $\mathfrak{a}$ has finite index in $\mathscr{O}_K$. We define the *norm* of $\mathfrak{a}$ to be

$$N(\mathfrak{a}) = \#(\mathscr{O}_K/\mathfrak{a}) \in \mathbf{Z}^+.$$

For example, if $\mathfrak{a} = \mathfrak{p}$ is a maximal ideal, so $\mathfrak{p} \cap \mathbf{Z} = p\mathbf{Z}$ for a unique prime $p \in \mathbf{Z}^+$ (this is the unique prime p belonging to $\mathfrak{p}$) and $p = \mathrm{char}(\mathscr{O}_K/\mathfrak{p})$, we have $N\mathfrak{p} = \#(\mathscr{O}_K/\mathfrak{p}) = p^f$ for $f := [\mathscr{O}_K/\mathfrak{p} : \mathbf{F}_p]$. Next time we will prove:

Theorem 15.1. *For nonzero ideals $\mathfrak{a}, \mathfrak{b} \subset \mathscr{O}_K$, $N(\mathfrak{a}\mathfrak{b}) = N(\mathfrak{a})N(\mathfrak{b})$.*

In view of the (unique) factorization of nonzero ideals into products of maximal ideals, it suffices to peel off one prime ideal factor at a time; i.e., to prove the general identity $N(\mathfrak{c}\mathfrak{m}) = N(\mathfrak{c})N(\mathfrak{m})$ for nonzero ideals $\mathfrak{c}$ and maximal ideals $\mathfrak{m}$ (as feeding this into itself yields $N(\mathfrak{c}) = \prod N(\mathfrak{q}_j)^{e_j}$ for the prime ideal factorization $\prod \mathfrak{q}_j^{e_j}$ of any $\mathfrak{c}$, from which the general multiplicativity of the ideal-norm operation follows). We will not use this result today.

Remark 15.2. The definition of $N(\mathfrak{a})$ makes sense for nonzero ideals in orders $\mathscr{O}$ of K; i.e., if $\mathfrak{a}$ is a nonzero ideal of such an $\mathscr{O}$ then $\mathscr{O}/\mathfrak{a}$ is finite. However, the assignment $\mathfrak{a} \mapsto \#(\mathscr{O}/\mathfrak{a})$ is generally *not* multiplicative when $\mathscr{O} \neq \mathscr{O}_K$. A counterexample will be given in Exercise 5 of Homework 6.

As justification for the "norm" terminology, we establish an important link in the case of principal ideals that will be very useful next week when working out how to factor nonzero principal ideals into a product of prime ideals in $\mathscr{O}_K$:

Proposition 15.3. *For nonzero $\alpha \in \mathscr{O}_K$, $N(\alpha\mathscr{O}_K) = |\mathrm{N}_{K/\mathbf{Q}}(\alpha)| \in \mathbf{Z}^+$.*

Proof. By definition, $\mathrm{N}_{K/\mathbf{Q}}(\alpha) = \det(m_\alpha)$ for the $\mathbf{Q}$-linear multiplication map $m_\alpha : K \to K$ defined by $x \mapsto \alpha x$; this is the determinant of the matrix for m_α relative to any desired choice of ordered $\mathbf{Q}$-basis of K.

Since the image of m_α is $\alpha\mathscr{O}_K$, $N(\alpha\mathscr{O}_K)$ is the size of $\mathrm{coker}(m_\alpha)$. Hence, it suffices to show more generally that if V is a finite-dimensional nonzero $\mathbf{Q}$-vector space, $T : V \to V$ is a $\mathbf{Q}$-linear isomorphism, and $L \subset V$ is a $\mathbf{Z}$-lattice for which $T(L) \subset L$ (e.g., $V = K$, $T = m_\alpha$, $L = \mathscr{O}_K$) then $[L : T(L)] = |\det(T)|$; note that $T(L)$ really is a $\mathbf{Z}$-lattice in V since $T(L)$ is the $\mathbf{Z}$-span of the $\mathbf{Q}$-basis $\{T(e_i)\}$ of V for any $\mathbf{Z}$-basis $\{e_i\}$ of L, so $L/T(L)$ really is finite.

To compare $[L : T(L)]$ and $\det(T)$, we use the structure theorem for modules over a PID, which gives that for the finite-index inclusion $T : L \to L$ there exist ordered $\mathbf{Z}$-bases $\{e_i\}$ and $\{e_i'\}$ of L such that $T(e_i') = c_i e_i$ for $c_i \in \mathbf{Z} - \{0\}$; i.e., relative to these respective bases of L as source and target of T the resulting matrix for T is diagonal with entries $c_1, \ldots, c_n$ along the diagonal. Concretely, this exhibits $L/T(L)$ as $\prod(\mathbf{Z}/(c_i))$, so $\#(L/T(L)) = \prod |\mathbf{Z}/(c_i)\mathbf{Z}| = \prod |c_i| = |\prod c_i|$. Hence, it suffices to show that $\det(T) = \pm \prod c_i$.

By definition, we can compute the determinant of $T : V \simeq V$ by using whatever *common* $\mathbf{Q}$-basis we wish on V as source and target. Say we use the basis $\{e_i'\}$ for both. The matrix M of T relative to that choice is unlikely to be diagonal, but since T has matrix $\mathrm{diag}(c_i)$ when using $\{e_i'\}$ as basis on the source and $\{e_i\}$ as basis on the target we see that

$$M = A \cdot \mathrm{diag}(c_i)$$

where A is the change-of-basis matrix (in one direction or the other!) between $\{e_i\}$ and $\{e_i'\}$ as ordered $\mathbf{Q}$-bases of V. But these two bases have the *same* $\mathbf{Z}$-span, namely L! Thus, the change of basis matrix between them in *both* directions is an integer matrix. In other words, A and A^{-1} are both integer matrices, so $\det(A), \det(A^{-1}) = 1/\det(A) \in \mathbf{Z}$, which is to say $\det(A) \in \mathbf{Z}^\times = \{\pm 1\}$. It follows that $\det(M) = \det(A)\det(\mathrm{diag}(c_i)) = \pm\prod c_i$, as desired. □

For the rest of today, we address some basic notions in what is called the "relative viewpoint"': rather than focusing on $\mathscr{O}_K$ by itself, or as an extension of $\mathbf{Z}$ (such as to contemplate how $p\mathscr{O}_K$ factors for prime $p \in \mathbf{Z}^+$), we consider a general extension of number fields K/E and study $\mathscr{O}_E \to \mathscr{O}_K$. For example, if $\mathfrak{m}$ is a maximal ideal of $\mathscr{O}_E$ then how does $\mathfrak{m}\mathscr{O}_K$ factor? Note that unlike the case $E = \mathbf{Q}$, typically $\mathscr{O}_E$ is not a PID and the $\mathscr{O}_E$-module $\mathscr{O}_K$ (finitely generated, since $\mathscr{O}_K$ is even $\mathbf{Z}$-finite) is generally *not* free; see the Remark in Exercise 0 of Homework 3.

The importance of the relative viewpoint becomes more apparent as one studies number theory further. For now, we record an important instance to be fully developed later in this course:

Example 15.4. For an odd prime p, $\mathbf{Q}(\zeta_p)/\mathbf{Q}$ is a Galois extension with Galois group $(\mathbf{Z}/p\mathbf{Z})^\times$ that is cyclic of size $p-1$ that is even, so there is a unique quadratic subfield. This subfield turns out to be $\mathbf{Q}(\sqrt{(-1|p)p})$ (as is proved classically via an explicit formula for this square root in terms of what are called Gauss sums, but which we shall prove more conceptually via the ideal-theoretic notion of "ramification"). The study of the extension $\mathbf{Q}(\zeta_p)/\mathbf{Q}(\sqrt{(-1|p)p})$ will later yield an elegant Galois-theoretic proof of quadratic reciprocity (the only proof of quadratic reciprocity that I can actually remember).

Let us now take the first steps on the road toward replacing "$\mathbf{Z} \to \mathscr{O}_K$" with suitable extensions between general Dedekind domains.

Theorem 15.5. *Let A be a Dedekind domain with fraction field F and let F'/F be a finite separable extension field. Let $A' \subset F'$ denote the integral closure of A in F.*

1. *A' is module-finite over A (so A' is noetherian) and* $\mathrm{Frac}(A') = F'$.

(2) A' is integrally closed in its fraction field F', and $A' \neq F'$.

(3) All nonzero prime ideals $\mathfrak{p}'$ of A' are maximal, and $\mathfrak{p}' \cap A$ is a maximal ideal of A.

In particular, A' is a Dedekind domain.

This result vastly generalizes many features we have seen for $\mathbf{Z} \to \mathscr{O}_K$. Whereas earlier proofs often had to rely on the structure theorem for modules over a PID (e.g., $\mathbf{Z}$-freeness of $\mathscr{O}_K$), now we know more powerful techniques in algebra (such as noetherianity); this will enable us to push through variants of earlier arguments that seemed only to work when the base ring (such as $\mathbf{Z}$) is a PID.

There is one feature that lies a bit deeper and is not recorded above: much as every maximal ideal $p\mathbf{Z}$ of $\mathbf{Z}$ does arise as $\mathfrak{p} \cap \mathbf{Z}$ for some maximal ideal $\mathfrak{p}$ of $\mathscr{O}_K$, it is also true in the above generality that every maximal ideal $\mathfrak{m}$ of A has the form $\mathfrak{m}' \cap A$ for a maximal ideal $\mathfrak{m}'$ of A'. The crucial step in the proof of this for $A = \mathbf{Z}$ was to use the $\mathbf{Z}$-module freeness of $\mathscr{O}_K$ to deduce that $\mathscr{O}_K/p\mathscr{O}_K \neq 0$. To show $A'/\mathfrak{m}A' \neq 0$ requires further development of commutative ring theory techniques (such as the notion of "localization" that will enable us to reduce a variety of problems over general Dedekind domains to the special case of PID's where some arguments over $\mathbf{Z}$ adapt very easily!).

Proof. **Proof of (1).** We first show that $\mathrm{Frac}(A') = F'$, and even more precisely that

$$\frac{1}{A - \{0\}} A' = F'$$

(i.e., every $\alpha \in F'$ can be written as a fraction a'/a for some $a' \in A'$ and $a \in A - \{0\}$). We shall apply the exact same "clear denominators" trick that was employed long ago to show that $K = (1/(\mathbf{Z} - \{0\}))\mathscr{O}_K$ for any number field K.

Pick $\alpha \in F'$. Since F'/F is a finite extension, $f(\alpha) = 0$ for some monic $f = \sum c_i X^i \in F[X]$. If $a \in A - \{0\}$ is a common denominator of the finitely many coefficients of f in $F = \mathrm{Frac}(A)$ (i.e., $af \in A[X]$) then for the degree $d > 0$ of f we have that $a^d f(X) = \sum (a^{d-i} c_i)(aX)^i$ vanishes at α. But $g(Y) = \sum (a^{d-i} c_i) Y^i$ is *monic* (as $a^0 c_d = 1$) with coefficients in A and $g(a\alpha) = a^d f(\alpha) = 0$, so $a' := a\alpha$ is an element of F' integral over A; i.e., $a' \in A'$. Since $a'/a = \alpha$, we have expressed α in the desired form.

To prove that A' is module-finite over A, we will apply a variant of the same method with a determinant of traces used to show $\mathscr{O}_K$ is module-finite over $\mathbf{Z}$. The main changes are: appeals to the structure theorem for modules over a PID will be replaced with noetherianity considerations, and invoking the rational root theorem will be replaced with the fact that A is integrally closed in F (for $A = \mathbf{Z}$, this property was deduced from the rational root theorem as for any PID).

Pick an ordered F-basis $\{e_i\}$ of F'. By applying scaling on these e_i's by suitable elements of $A - \{0\}$ (as we may do), it can be arranged that $e_i \in A'$ for all i. The trace map $\mathrm{Tr}_{F'/F} : F' \to F$ carries A' into A for exactly the same Galois-theoretic reasons as were used to show that $\mathrm{Tr}_{K/\mathbf{Q}}(\mathscr{O}_K) \subset \mathbf{Z}$ for number fields K (namely, working inside a Galois closure of F'/F, sums of A-integral elements are A-integral and the A-integral elements of F are precisely the elements of A since A is integrally closed in F). In the expansion $\alpha = \sum c_i e_i$ with $c_i \in F$, the main aim is to uniformly control denominators for such c_i's independently of $\alpha \in A'$. For $n = [F' : F]$, consider the following matrix that has *nothing to do with* α:

$$(\mathrm{Tr}_{F'/F}(e_i e_j)) \in \mathrm{Mat}_n(F);$$

these traces belong to A since $e_i e_j \in A'$ for all i, j.

This matrix has determinant $d \in A$ that is *non-zero* due to the separability of F'/F, by exactly the same reasoning as was used to prove the nonvanishing of such a determinant for number fields $K/\mathbf{Q}$. (Namely, the nonvanishing property is seen to be independent of the choice of F-basis for F', and we compute it for a power-basis via a norm of an element of F' that is nonzero due to the separability of F'/F.) The *exact same* linear algebra computations as carried out for $K/\mathbf{Q}$ yield that $c_i \in (1/d)A$ for all i (where d has nothing to do with α), so

$$A' \subset \sum (1/d) A e_i.$$

In the earlier setting with $A = \mathbf{Z}$, we then deduced $\mathbf{Z}$-finiteness of $\mathscr{O}_K$ at this step via facts about torsion-free finitely generated modules over a PID (e.g., all submodules of such are again finitely generated). Now we know more algebra and can instead argue that since A is *noetherian*, the A-submodule A' of the finitely generated (even free) A-module $\sum(1/d)Ae_i$ is finitely generated (though typically not A-free!). This completes the proof of (1).

Proof of (2). Next, we check that $A' \neq F'$ (equivalently by (1), A' is not a field) and that A' is integrally closed in F'. Since A is not a field, so there exists a nonzero non-unit $a \in A$, to show A' is not a field it is sufficient to show any such a is not a unit in A'. Well, just as the trace map $\mathrm{Tr}_{F'/F}$ carries A' into A, the norm map $\mathrm{N}_{F'/F} : F' \to F$ carries A' into A (by the same reasoning used to show $\mathrm{N}_{K/\mathbf{Q}}$ carries $\mathscr{O}_K$ into $\mathbf{Z}$). Thus, if a admits a multiplicative inverse a' in A' then applying $\mathrm{N}_{F'/F}$ to the identity $aa' = 1$ in A' yields the identity $\mathrm{N}_{F'/F}(a)\mathrm{N}_{F'/F}(a') = 1$ in A. But since $\mathrm{N}_{F'/F}|_F : F \to F$ is the power map $t \mapsto t^n$ for $n = [F' : F]$, we get

$$a^n \mathrm{N}_{F'/F}(a') = 1.$$

This says that the element $a^{n-1}\mathrm{N}_{F'/F}(a') \in A$ is a multiplicative inverse to a, contradicting that we chose a to be a nonzero non-unit in A. This completes the proof that $A' \neq F'$.

The integral closedness of A' in its fraction field F' goes via exactly the same argument used to show (in Exercise 0 of HW3) that $\mathscr{O}_K$ is integrally closed in its fraction field K, namely using transitivity of integrality through a composition of ring extensions (e.g., if $\alpha \in F'$ is integral over A' then via integrality of A' over A we deduce α is integral over A, so in fact $\alpha \in A'$). The proof of (2) is now done.

Proof of (3). Finally, we show that every nonzero prime ideal $\mathfrak{p}'$ of A' is maximal, and that $\mathfrak{p}' \cap A$ is a maximal ideal of A. The intersection $\mathfrak{p} := A \cap \mathfrak{p}'$ is a prime ideal of A since the natural map

$$A/\mathfrak{p} \to A'/\mathfrak{p}'$$

is an *injective* ring homomorphism whose target is a domain, forcing the subring $A/\mathfrak{p}$ to also be a domain; this conclusion says that $\mathfrak{p}$ is prime. (Recall that by definition the zero ring is *not* a domain. More specifically, since $\mathfrak{p}'$ is a proper ideal, so $1 \notin \mathfrak{p}'$, likewise $1 \notin A \cap \mathfrak{p}' =: \mathfrak{p}$, so $\mathfrak{p}$ is a proper ideal of A.) The Dedekind property of A gives that all nonzero prime ideals of A are maximal, so to conclude that the prime ideal $\mathfrak{p} = \mathfrak{p}' \cap A$ of A is maximal we just need to check that it is nonzero. (It still needs to be shown that $\mathfrak{p}'$ is maximal!)

More generally, for $\mathfrak{a}'$ *any* nonzero ideal of A' we claim $A \cap \mathfrak{a}' \neq 0$. Pick a nonzero element $a' \in \mathfrak{a}'$, and let $f \in F[X]$ be its minimal polynomial over F. The constant term $f(0)$ is nonzero since $a' \neq 0$, and the coefficients of f belong to A due to $a' \in A'$ being integral over A (by an argument *identical* to that by which it was shown in Exercise 3 of Homework 2 that the minimal polynomial in $\mathbf{Q}[X]$ of any algebraic integer necessarily has its coefficients in $\mathbf{Z}$). Writing $f = \sum c_i X^i$ with all $c_i \in A$ and $c_0 = f(0) \in A - \{0\}$, the vanishing of $f(a')$ implies

$$c_0 = \sum_{i>0} -c_i a'^i = a'(-\sum_{i>0} c_i a'^{i-1}) \in \mathfrak{a}'$$

since $a' \in \mathfrak{a}'$. Hence, the nonzero c_0 belongs to $A \cap \mathfrak{a}'$, so $A \cap \mathfrak{a}' \neq 0$.

With $\mathfrak{p} = A \cap \mathfrak{p}'$ now known to be a *maximal* ideal of A, let's revisit the ring-theoretic inclusion

$$A/\mathfrak{p} \to A'/\mathfrak{p}'.$$

This expresses the domain $A'/\mathfrak{p}'$ as a ring extension of the domain $A/\mathfrak{p}$ that is a field (as $\mathfrak{p}$ is maximal), and this ring extension is module-finite because we have shown that A' is module-finite over A. Voila, the ring $A'/\mathfrak{p}'$ is a domain that is finite-dimensional over its subring $A/\mathfrak{p}$ that is a field, and in such situations we know that the ambient domain must be a field (see Exercise 4(ii) in Homework 3). With $A'/\mathfrak{p}'$ now shown to be a field, it follows that $\mathfrak{p}'$ is maximal in A'! □

16. Factoring $p\mathscr{O}_K$: the quadratic case

We'll start by finishing up a loose end from last time. Let K be a number field and let $\mathscr{O}_K$ be its ring of integers. Recall that the norm of a nonzero ideal I of $\mathscr{O}_K$, denoted $N(I)$, is the size of $\mathscr{O}_K/I$.

Theorem 16.1. *For nonzero ideals $\mathfrak{a}, \mathfrak{b} \subset \mathscr{O}_K$, we have*

$$N(\mathfrak{a} \cdot \mathfrak{b}) = N(\mathfrak{a}) \cdot N(\mathfrak{b}).$$

Proof. We can write $\mathfrak{b}$ as a product of prime ideals: $\mathfrak{p}_1 \cdots \mathfrak{p}_k$. Using this, it is enough to do the case that $\mathfrak{b}$ is a prime ideal by applying the formula recursively via induction on k since then

$$\begin{aligned} N(\mathfrak{a} \cdot \mathfrak{p}_1 \cdots \mathfrak{p}_k) &= N(\mathfrak{a}\mathfrak{p}_1 \cdots \mathfrak{p}_{k-1}) \cdot N(\mathfrak{p}_k) \\ &= N(\mathfrak{a})N(\mathfrak{p}_1)N(\mathfrak{p}_2) \cdots N(\mathfrak{p}_k) \\ &= N(\mathfrak{a}) \cdot N(\mathfrak{p}_1 \cdot \mathfrak{p}_2) \cdot N(\mathfrak{p}_3) \cdots N(\mathfrak{p}_k) \\ &= \dots \\ &= N(\mathfrak{a}) \cdot N(\mathfrak{b}). \end{aligned}$$

Thus, for the remainder of the proof, we assume $\mathfrak{b}$ is prime.

We have

$$N(\mathfrak{a} \cdot \mathfrak{b}) = \#(\mathscr{O}_K/(\mathfrak{a} \cdot \mathfrak{b})).$$

We have containments

$$\mathfrak{a}\mathfrak{b} \subset \mathfrak{a} \subset \mathscr{O}_K,$$

and one of the isomorphism theorems yields

$$\begin{aligned} \#(\mathscr{O}_K/\mathfrak{a}\mathfrak{b}) &= \#(\mathscr{O}_K/\mathfrak{a})\#(\mathfrak{a}/\mathfrak{a}\mathfrak{b}) \\ &= N(\mathfrak{a}) \cdot \#(\mathfrak{a}/\mathfrak{a}\mathfrak{b}). \end{aligned}$$

Thus, it suffices to show

$$\#(\mathfrak{a}/\mathfrak{a}\mathfrak{b}) \overset{?}{=} N(\mathfrak{b}) := \#(\mathscr{O}_K/\mathfrak{b})$$

for maximal ideals $\mathfrak{b}$.

It suffices to build an $\mathscr{O}_K$-linear isomorphism between $\mathfrak{a}/\mathfrak{a} \cdot \mathfrak{b}$ and $\mathscr{O}_K/\mathfrak{b}$. To set this up, note that it is well-defined to multiply elements of $\mathfrak{a}/\mathfrak{a}\mathfrak{b}$ against elements of $\mathscr{O}_K/\mathfrak{b}$: for $x \in \mathscr{O}_K$, $y \in \mathfrak{a}$, and $b \in \mathfrak{b}$,

$$(x + b)y = xy + by \equiv \bmod \mathfrak{a}\mathfrak{b}.$$

This implies $\mathfrak{a}/\mathfrak{a}\mathfrak{b}$ is a vector space over the finite field $\mathscr{O}_K/\mathfrak{b}$. Since this vector space finite-dimensional (as $\mathfrak{a}$ is a finitely generated $\mathscr{O}_K$-module), say with dimension d, we have

$$(\mathfrak{a}/\mathfrak{a}\mathfrak{b}) \simeq (\mathscr{O}_K/\mathfrak{b})^d$$

as $\mathscr{O}_K/\mathfrak{b}$-vector spaces for some $d \geq 0$. It suffices to show that $d = 1$. Clearly $d > 0$ since the inclusion $\mathfrak{a}\mathfrak{b} \subset \mathfrak{a}$ cannot be an equality due to the uniqueness of prime factorization. (Recall $\mathfrak{b}$ is now a maximal ideal.)

Suppose $d > 1$, so $\mathfrak{a}/\mathfrak{a}\mathfrak{b}$ has a proper nonzero subspace. Passing to preimages under the $\mathscr{O}_K$-linear surjection $\mathfrak{a} \to \mathfrak{a}/\mathfrak{a}\mathfrak{b}$, we get a strict containment $\mathscr{O}_K$-submodules

$$\mathfrak{a}\mathfrak{b} \subsetneq I \subsetneq \mathfrak{a};$$

note that I is a (nonzero) ideal of $\mathscr{O}_K$ since it is an $\mathscr{O}_K$-submodule of $\mathscr{O}_K$. To see that this leads to a contradiction, recall that for nonzero ideals J, J' of $\mathscr{O}_K$, we have $J \subset J'$ if and only if J' divides J. This implies that $\mathfrak{a} \mid I$ and $I \mid \mathfrak{a}\mathfrak{b}$. Since $\mathfrak{b}$ is a prime, by uniqueness of prime factorization either $I = \mathfrak{a}$ or $I = \mathfrak{a}\mathfrak{b}$, contrary to the strict containments in the way I was made. □

Remark 16.2. If $I = \mathfrak{p}_1 \cdots \mathfrak{p}_r$ for maximal $\mathfrak{p}_j$'s, we have

$$NI = (N\mathfrak{p}_1) \cdots (N\mathfrak{p}_r).$$

Each $N\mathfrak{p}_i$ is a prime power because $\mathscr{O}_K/\mathfrak{p}_i$ is a finite field, and all finite fields have prime power order.

Explicitly, if $N\mathfrak{p}_i = q^\alpha$ for some prime integer $q \in \mathbf{Z}^+$, then in fact $q \in \mathfrak{p}_i$. To see this, recall that q is the characteristic of the finite field $\mathscr{O}_K/\mathfrak{p}_i$, so q vanishes in the quotient $\mathscr{O}_K/\mathfrak{p}_i$. Since $q \in \mathfrak{p}_i$, we obtain a containment of ideals $q\mathscr{O}_K \subset \mathfrak{p}_i$. This implies $\mathfrak{p}_i$ divides the ideal $q\mathscr{O}_K$, and that q is the unique prime in $\mathbf{Z}^+$ such that $q \in \mathfrak{p}_i$.

In summary, to factor I, one can follow the following outline in order:

(1) factor $N(I)$ (if we have some other way to compute it),
(2) for each prime $q \in \mathbf{Z}^+$ dividing $N(I)$, factor $q\mathscr{O}_K$ into prime ideals (we will explain how to do this later today and into next time),
(3) finally, figure out which of the resulting prime ideals divide I (and with what multiplicity).

For the remainder of today, we will concentrate on examples of the second step, factoring $q\mathscr{O}_K$, for K a *quadratic* field and q a prime in $\mathbf{Z}^+$.

Let $d \in \mathbf{Z} - \{0, 1\}$ be squarefree and $K = \mathbf{Q}(\sqrt{d})$. For convenience, assume $d \equiv 2$ or $3 \bmod 4$ so that $\mathscr{O}_K = \mathbf{Z}[\sqrt{d}]$. Consider a prime $q > 2$. (See the handout "Some Quadratic Factoring" for a systematic treatment of all d, permitting the case $d \equiv 1 \bmod 4$ and allowing the case $q = 2$.) Recall that either

$$\mathscr{O}_K/q\mathscr{O}_K = \begin{cases} \mathbf{F}_q \times \mathbf{F}_q & \text{if } d \text{ is a nonzero square } \bmod q \\ \mathbf{F}_{q^2} & \text{if } d \text{ is not a square } \quad \bmod q \\ \mathbf{F}_q[x]/(x^2) & \text{if } d \equiv 0 \bmod q \end{cases}$$

Primes ideals of $\mathscr{O}_K/q\mathscr{O}_K$ correspond to prime ideals of $\mathscr{O}_K$ containing $q\mathscr{O}_K$. Correspondingly, we claim that the prime ideal factors of $q\mathscr{O}_K$ are as follows:

(1) $(q, u+\sqrt{d}) \neq (q, u-\sqrt{d})$, where u satisfies $u^2 \equiv d \bmod q$.
(2) $q\mathscr{O}_K$ is prime
(3) $(q, \sqrt{d})$

The first case is called *split* because the prime is a product of two distinct prime ideals, the second case is called *inert* because the rational prime "remains a prime" in $\mathscr{O}_K$ (in the sense of the ideal that it generates), and remaining case is called *ramified* because $q\mathscr{O}_K$ is the square of a unique prime-ideal factor. To verify this claim, first note that $N(q\mathscr{O}_K) = q^2$. Since $\mathscr{O}_K = \{a + b\sqrt{d} \,|\, a, b \in \mathbf{Z}\}$ and $q\mathscr{O}_K = \{a + b\sqrt{d} \mid q|a, q|b, a \in \mathbf{Z}, b \in \mathbf{Z}\}$, we can factor $q\mathscr{O}_K$ manually in each respective case to the above cases:

(1) We have $q\mathscr{O}_K = (q, u+\sqrt{d})(q, u-\sqrt{d})$ because we know from existence and uniqueness of prime ideal factorization (and the equivalence between the relation $I|J$ and $J \subset I$ for nonzero ideals I and J of $\mathscr{O}_K$) that

$$q\mathscr{O}_K = (q, u+\sqrt{d})^\alpha (q, u-\sqrt{d})^\beta$$

for some $\alpha, \beta \geq 1$. Taking norms we see $q^2 = q^\alpha q^\beta$, and so we must have $\alpha = \beta = 1$ (as can also be verified by bare hands).
(2) We have $q\mathscr{O}_K$ is prime, so there is nothing to do.
(3) We have $q\mathscr{O}_K = (q, \sqrt{d})^2$. To see this, observe that $N(q, \sqrt{d}) = q$ because the equality

$$(q, \sqrt{d}) = \{a + b\sqrt{d} \,|\, b \in \mathbf{Z}, q \mid a\}.$$

implies $N(q, \sqrt{d}) = q$ by inspection of the lattice index. Therefore,

$$q\mathscr{O}_K = (q, \sqrt{d})^\alpha$$

for some $\alpha \geq 1$, and we see by taking norms that $q^2 = q^\alpha$, so $\alpha = 2$ (as can also be verified by bare hands).

Example 16.3. Consider $K = \mathbf{Q}(\sqrt{-5})$. We have $\mathscr{O}_K = \mathbf{Z}[\sqrt{-5}]$.

Let's factor the ideal $\mathfrak{a} = (1+\sqrt{-5})$. Following our procedure, we first take norms. Recalling the link between norms of elements and norms of principal ideas as established last time,

$$N\mathfrak{a} = |N_{K/\mathbf{Q}}(1+\sqrt{-5})| = (1+\sqrt{-5})(1-\sqrt{-5}) = 6.$$

Since $6 = 2 \cdot 3$, we must have $a = \mathfrak{p}_2\mathfrak{p}_3$ with $N\mathfrak{p}_2 = 2$, $N\mathfrak{p}_3 = 3$. Then, $\mathfrak{p}_2$ must divide $2\mathscr{O}_K$ and $\mathfrak{p}_3$ must divide $3\mathscr{O}_K$. The only prime dividing $2\mathscr{O}_K$ is $(2, 1+\sqrt{-5})$ (as explained in the handout "Some Quadratic Factoring" for factoring $2\mathscr{O}_K$). Since $-5 \equiv 1^2 \bmod 3$, we also have the prime ideal factorization

$$3\mathscr{O}_K = (3, 1+\sqrt{-5})(3, 1\pm\sqrt{-5}).$$

for exactly one of the two possibilities for the sign. The prime ideal factor of $3\mathscr{O}_K$ that occurs here must be the unique one containing $1+\sqrt{-5}$, so

$$(1+\sqrt{-5}) = (2, 1+\sqrt{-5})(3, 1+\sqrt{-5}).$$

(Again, recall that a containment among nonzero ideals in $\mathscr{O}_K$ amounts to a divisibility in the opposite direction.)

See the handout "Some Quadratic Factoring" for further explicit examples in this spirit for $K = \mathbf{Q}(\sqrt{-5})$.

17. Factoring $p\mathscr{O}_K$: the general case

Let p be a rational prime (meaning a prime number in $\mathbf{Z}^+$) and K be a number field. In the factorization

$$p\mathscr{O}_K = \mathfrak{p}_1^{e_1}\mathfrak{p}_2^{e_2}\cdots\mathfrak{p}_g^{e_g},$$

the numbers e_i are called the *ramification indices* and the numbers f_i are called the *residual degrees*. We have the following relation between ramification indices, residual degrees, and $[K:\mathbf{Q}]$.

Theorem 17.1. *If $p\mathscr{O}_K = \mathfrak{p}_1^{e_1}\cdots\mathfrak{p}_g^{e_g}$ is the prime factorization of p in $\mathscr{O}_K$, then $\sum_i e_i f_i = [K:\mathbf{Q}]$.*

Proof. In the above setup, $\mathscr{O}_K/\mathfrak{p}_i$ is an $\mathbf{F}_p$-vector space, so we have an isomorphism of vector spaces

$$\mathscr{O}_K/\mathfrak{p}_i \simeq \mathbf{F}_p^{f_i}.$$

This tells us that

$$N\mathfrak{p}_i = \#\mathscr{O}_K/\mathfrak{p}_i = p^{f_i}.$$

Hence,

$$\begin{aligned} p^{[K:\mathbf{Q}]} = \mathrm{N}_{K/\mathbf{Q}}(p) &= N(p\mathscr{O}_K) \\ &= N(\mathfrak{p}_1)^{e_1} N(\mathfrak{p}_2)^{e_2} \cdots N(\mathfrak{p}_g)^{e_g} \\ &= (p^{f_1})^{e_1} (p^{f_2})^{e_2} \cdots (p^{f_g})^{e_g} \\ &= p^{\sum e_i f_i}. \end{aligned}$$

Comparing the exponents on the powers of p gives the result. □

Remark 17.2. Since $\sum_{i=1}^{g} e_i f_i \geq \sum_{i=1}^{g} 1 = g$, we always have $g \leq [K : \mathbf{Q}]$. Hence, for a maximal ideal $p\mathbf{Z}$ of $\mathbf{Z}$, the ideal $p\mathscr{O}_K$ that it generates in $\mathscr{O}_K$ splits into a product of at most $[K : \mathbf{Q}]$ maximal ideals in $\mathscr{O}_K$.

Example 17.3. Let us examine Theorem 17.1 in the quadratic case. In the case $[K : \mathbf{Q}] = 2$, we have $\sum_{i=1}^{g} e_i f_i = 2$. Hence, there are three cases:

(1) The first case is $g = 2$, in which case necessarily $e_i = f_i = 1$ for all i. This is called the *totally split* case. Here we have
$$p\mathscr{O}_K = \mathfrak{p}_1\mathfrak{p}_2$$
for prime ideals $\mathfrak{p}_1 \neq \mathfrak{p}_2$.

(2) The second case is $g = 1, e = 2, f = 1$. Here $p\mathscr{O}_K = \mathfrak{p}^2$. This is called the *ramified case*.

(3) The third case is $g = 1, e = 1, f = 2$, so $p\mathscr{O}_K$ is itself prime. This is called the *inert case*.

How do we determine the e_i, f_i, and g for a given prime p in general? This is the task we addressed last time for K quadratic over $\mathbf{Q}$, and today we will address the general case.

Definition 17.4. Write $p\mathscr{O}_K = \mathfrak{p}_1^{e_i}\mathfrak{p}_2^{e_2} \cdots \mathfrak{p}_g^{e_g}$ with all $e_j \geq 1$. If some $e_i > 1$ then we say p *ramifies* in K.

Next time it will be shown that p ramifies in K if and only if $p \mid \mathrm{disc}(K)$. In particular, there are only finitely many primes that ramify in a given number field, since there are only finitely many primes dividing the discriminant.

Definition 17.5. If $g = [K : \mathbf{Q}]$ (equivalently, $e_i = f_i = 1$ for all i) we say p is *totally split* in K.

To factor primes in number fields, we have the following useful sufficient criterion:

Proposition 17.6 (Dedekind's Criterion)**.** *Let K be a number field, $\alpha \in \mathscr{O}_K$ a primitive element for $K/\mathbf{Q}$ (equivalently, $[\mathscr{O}_K : \mathbf{Z}[\alpha]]$ is finite). If $p \nmid [\mathscr{O}_K : \mathbf{Z}[\alpha]]$ then we can factor $p\mathscr{O}_K$ as follows.*

Let $h \in \mathbf{Z}[x]$ be the minimal polynomial of α over $\mathbf{Q}$. Factor the reduction $\overline{h} \in \mathbf{F}_p[x]$ as

$$\overline{h} = \overline{h}_1^{e_1}\overline{h}_2^{e_2} \cdots \overline{h}_g^{e_g}$$

for distinct monic irreducible $\overline{h}_i$. Choose monic $h_i \in \mathbf{Z}[x]$ reducing to $\overline{h}_i$. Then

$$p\mathscr{O}_K = \mathfrak{p}_1^{e_1}\mathfrak{p}_2^{e_2} \cdots \mathfrak{p}_g^{e_g},$$

for $\mathfrak{p}_i := (p, h_i(\alpha))$ is the prime factorization of $p \in \mathscr{O}_K$.

This result is proved in the handout "Dedekind's Factorization Criterion". We want to explain here how to use the result, as this gives more appreciation for the usefulness of the result than is gained from a study of the proof (which is nonetheless an instructive piece of ring-theoretic reasoning).

First, recall that

$$[\mathscr{O}_K : \mathbf{Z}[\alpha]]^2 \operatorname{disc} K = \pm N_{K/\mathbf{Q}}(h'(\alpha)).$$

Thus, if $p \nmid N_{K/\mathbf{Q}}(h'(\alpha))$ then $p \nmid [\mathscr{O}_K : \mathbf{Z}[\alpha]]$. This provides a computable condition on p to ensure the non-divisibility hypothesis in Dedekind's criterion applies to p for a given α even without knowing $[\mathscr{O}_K : \mathbf{Z}[\alpha]]$.

Next, when Dedekind's criterion applies we can also use it to compute the residual degrees f_i. Namely, we have

$$\begin{aligned} f_i &= \dim_{\mathbf{F}_p} \mathscr{O}_K/\mathfrak{p}_i \\ &= \dim_{\mathbf{F}_p} \mathbf{F}_p[x]/(\overline{h}_i) \\ &= \deg \overline{h}_i. \end{aligned}$$

As a final general remark, we stress that it can happen for a given K and p that Dedekind's criterion doesn't apply for *any* choice of $\alpha \in \mathscr{O}_K$ that is a primitive element of $K/\mathbf{Q}$ (i.e., for all such α, necessarily $p|[\mathscr{O}_K : \mathbf{Z}[\alpha]]$; an explicit example of this for a cubic extension $K/\mathbf{Q}$ with $p = 2$ will be discussed later.

Let us now take up an explicit cubic extension $K/\mathbf{Q}$ and several primes p (and the handout "Dedekind Factorization Criterion" works out some more examples with another cubic number field).

Example 17.7. Let $K = \mathbf{Q}(\alpha)$ with $\alpha^3 = 10$. We can apply Dedekind's criterion to any prime p not dividing the finite index of $\mathbf{Z}[\alpha]$ in $\mathscr{O}_K$. Note that $\mathscr{O}_K \neq \mathbf{Z}[\alpha]$ since

$$\beta := (1/3)(1 + \alpha + \alpha^2) \in \mathscr{O}_K,$$

as $\beta^3 - \beta^2 - 3\beta - 3 = 0$, checked by computing the characteristic polynomial for multiplication $m_\beta : K \to K$ (say computed with respect to the $\mathbf{Q}$-basis $\{1, \alpha, \alpha^2\}$ of K). Thus, some primes divide the index $[\mathscr{O}_K : \mathbf{Z}[\alpha]] > 1$. Which ones?

The minimal polynomial of α over $\mathbf{Q}$ is $h = x^3 - 10$, so $h' = 3x^2$ and hence

$$\begin{aligned} N_{K/\mathbf{Q}}(h'(\alpha)) &= N_{K/\mathbf{Q}}(3\alpha^2) \\ &= N_{K/\mathbf{Q}}(3) N_{K/\mathbf{Q}}(\alpha)^2 \\ &= 27 \cdot 10^2 \\ &= 2^2 \cdot 5^2 \cdot 3^2. \end{aligned}$$

Thus, for$p \neq 2, 3, 5$ we know for certain that p doesn't divide $[\mathscr{O}_K : \mathbf{Z}[\alpha]]$ and hence we can apply Dedekind's criterion to determine the factorization of $p\mathscr{O}_K$. Let's do this for several such p.

The case $p = 7$**.** We need to factor $x^3 - 10$ in $\mathbf{F}_7[x]$. The element $10 \in \mathbf{F}_7$ is not a cube, since the only cubes in $\mathbf{F}_7$ are ± 1 (and 0). This implies $x^3 - 10$ is irreducible in $\mathbf{F}_7[x]$. Hence, by Dedekind's criterion we conclude that $7\mathscr{O}_K$ is prime in $\mathscr{O}_K$. In particular, the ramification index is 1 and the residual degree is 3.

The case $p = 11$**.** Since $10 \equiv -1 \bmod 11$, in $\mathbf{F}_{11}[x]$ we have $x^3 - 10 = x^3 + 1 = (x+1)(x^2 - x + 1)$ and we can check this quadratic has no roots (its discriminant -3 is a non-square modulo 11, either by computing with Legendre symbols via quadratic reciprocity or by inspection since 11 is reasonably small). Hence that quadratic factor is irreducible over $\mathbf{F}_{11}$.

It follows that the prime ideal factorization is

$$11\mathscr{O}_K = (11, \alpha + 1)(11, \alpha^2 - \alpha + 1).$$

The ramification indices of both prime ideal factors is 1, for the first prime ideal factor the residual degree is $f_1 = 1$ and for the second prime it is $f_2 = 2$. Here, we have $\sum_{i=1}^{2} e_i f_i = 1 \cdot 1 + 1 \cdot 2 = 3 = [K : \mathbf{Q}]$ as we knew must hold.

The case $p = 37$**.** How can we factor $x^3 - 10$ in $\mathbf{F}_{37}[x]$? Since

$$10 \equiv -27 = (-3)^3 \bmod 37,$$

we get a factor of $x + 3$. Is the remaining quadratic factor of $x^3 - 10$ irreducible, or alternatively does 10 have 3 distinct cube roots in $\mathbf{F}_{37}$?

Since $37 \equiv 1 \bmod 3$, the cyclic group $\mathbf{F}_{37}^{\times}$ has order 36 that is divisible by 3 and hence contains 3 distinct cube roots of 1. Thus, from one cube of 10 we can make three such (multiplying it by the three cube roots of 1).

Remark 17.8. As an alternate way to see that $\mathbf{F}_{37}$ contains a non-trivial cube root of 1, we consider such a cube root ω in an extension of $\mathbf{F}_{37}$ and can check via Galois theory if it lies in $\mathbf{F}_{37}$: it is necessary and sufficient that $\omega^{37-1} = 1$. But $\omega^3 = 1$, so obviously $\omega^{36} = 1$.

Letting ω be a nontrivial cube root of unity, we have that $10 = (-3)^3 = (-3\omega)^3 = (-3\omega^2)^3$. It follows that $x^3 - 10$ factors as a product of three distinct monic degree-1 polynomials over $\mathbf{F}_{37}$. With a bit of searching to find an explicit ω in $\mathbf{F}_{37}$ (or by seeking $\sqrt{-3}$ and using $(-1+\sqrt{-3})/2$ as ω), we can make this explicit:

$$x^3 - 10 = (x-7)(x+4)(x+3) \in \mathbf{F}_{37}[x].$$

Thus,

$$37\mathscr{O}_K = (37, \alpha - 7)(37, \alpha + 4)(37, \alpha + 3).$$

is the prime factorization of (37) in $\mathscr{O}_K$. This is totally split (so $e_i = f_i = 1$ for all i). It turns out that 37 is the smallest prime $p > 5$ that is totally split in K (and even 2, 3, 5 are not, but this involves more work; we will analyze the factorization of $3\mathscr{O}_K$ next time).

Example 17.9. Now we will give an example of a number field K such that *no* rational prime p "remains prime" in $\mathscr{O}_K$ (i.e., $p\mathscr{O}_K$ is never prime). We will use $K = \mathbf{Q}(\sqrt{2}, \sqrt{3})$, so $[K : \mathbf{Q}] = 4$ and hence $\sum_i e_i f_i = 4$.

Our task is to show it is not possible to have $g = 1, e_1 = 1, f_1 = 4$. Observe that if $p\mathscr{O}_K$ is prime then $p\mathscr{O}_L$ is prime for any subfield $L \subset K$: if $p\mathscr{O}_L$ factors non-trivially in $\mathscr{O}_L$ then

$$p\mathscr{O}_L = \mathfrak{p}_1\mathfrak{p}_2\mathfrak{a}$$

for two (possibly equal) maximal ideals $\mathfrak{p}_i$ and an ideal $\mathfrak{a}$ of $\mathscr{O}_L$. But then

$$p\mathscr{O}_K = \mathfrak{p}_1\mathscr{O}_K \cdot \mathfrak{p}_2\mathscr{O}_K \cdot \mathfrak{a}\mathscr{O}_K.$$

This contradicts the primality of $p\mathscr{O}_K$ provided that each $\mathfrak{p}_i\mathscr{O}_K$ is not the unit ideal (and so admits prime ideal factors), that in turn is a special case of the general fact (to be proved later) that for any extension F'/F of number fields and any maximal ideal $\mathfrak{p}$ of $\mathscr{O}_F$, the ideal $\mathfrak{p}\mathscr{O}_{F'}$ is never the unit ideal.

We saw last time that an odd prime p remains prime in $\mathbf{Q}(\sqrt{2})$ (i.e., $p\mathbf{Z}[\sqrt{2}]$ is prime, or equivalently p is irreducible in $\mathbf{Z}[\sqrt{2}]$ that is a PID) precisely when 2 is not a square mod p. Therefore, if p remains prime in $\mathscr{O}_K$ then necessarily 2 is not a square mod p. Similarly, applying this to the subfield $\mathbf{Q}(\sqrt{3})$, we get the further necessary condition that 3 is not a square mod p. Finally, using $\mathbf{Q}(\sqrt{6})$, we obtain that 6 is not a square mod p. But in terms of Legendre symbols we then have

$$\begin{aligned}\left(\frac{6}{p}\right) &= \left(\frac{2}{p}\right)\left(\frac{3}{p}\right)\\ &= (-1)(-1)\\ &= 1\end{aligned}$$

which contradicts that $(\frac{6}{p}) = -1$.

18. Ramification

Let $\mathscr{O}_K$ be the ring of integers in a number field K. Let $p \in \mathbf{Z}^+$ be prime, and $\alpha \in \mathscr{O}_K$ such that $[\mathscr{O}_K : \mathbf{Z}[\alpha]]$ is not divisible by p. If f is the minimal polynomial of α over $\mathbf{Q}$ then

$$\mathbf{Z}[\alpha] \simeq \mathbf{Z}[t]/(f)$$

and we saw that (even if $\mathbf{Z}[\alpha] \neq \mathscr{O}_K$) the factorization of $p\mathscr{O}_K$ is determined by the factorization of f mod $p \in \mathbf{F}_p[t]$. Explicitly, $p\mathscr{O}_K = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_k^{e_k}$ with $\mathfrak{p}_i = (p, f_i(\alpha))$ where f mod $p = \prod \overline{f}_i^{e_i}$ for pairwise distinct monic irreducible $\overline{f}_i \in \mathbf{F}_p[t]$. We now revisit a cubic field K considered last time.

Example 18.1. Let $K = \mathbf{Q}(\alpha)$ and $\alpha^3 = 10$. Recall that $\mathbf{Z}[\alpha] \neq \mathscr{O}_K$ as we saw in Example 17.7. We also saw in Example 17.7 that $\operatorname{disc} \mathbf{Z}[\alpha] = -2700 = 2^2 5^2 3^3$, which is divisible by $2, 3$, and 5.

To factor $3\mathscr{O}_K$, we shall find a more convenient order than $\mathbf{Z}[\alpha]$. Consider

$$\beta = (1/3)(1 + \alpha + \alpha^2);$$

this satisfies $\beta^3 - \beta^2 - 3\beta - 3 = 0$, from which one computes $\operatorname{disc} \mathbf{Z}[\beta] = -300 = 2^2 \cdot 3 \cdot 5^2$. Recalling that

$$[\mathscr{O}_K : \mathbf{Z}[\beta]]^2 \mid \operatorname{disc} \mathbf{Z}[\beta],$$

it follows that the index $[\mathscr{O}_K : \mathbf{Z}[\beta]]$ cannot be divisible by 3 (and must divide 10). Therefore, Dedekind's criterion may be applied for $p = 3$ by using the order $\mathbf{Z}[\beta]$: the factorization of $3\mathscr{O}_K$ corresponds to the factorization of $t^3 - t^2 - 3t - 3 \in \mathbf{F}_3[t]$; this factors as $t^2(t-1)$, so

$$3\mathscr{O}_K = (3, \beta)^2(3, \beta - 1).$$

The formula $\operatorname{disc} \mathbf{Z}[\beta] = [\mathscr{O}_K : \mathbf{Z}[\beta]]^2 \operatorname{disc} \mathscr{O}_K$ also shows that $3 | \operatorname{disc} \mathscr{O}_K$.

Note that if we were sloppy and used Dedekind's criterion at $p = 3$ with $\mathbf{Z}[\alpha]$ (which violates the requirement that $p \nmid [\mathscr{O}_K : \mathbf{Z}[\alpha]]$) we would get the wrong answer since $\mathbf{Z}[\alpha]/(3) = \mathbf{F}_3[t]/(t^3 - 10) = \mathbf{F}_3[t]/(t-1)^3$ yet $3\mathscr{O}_K$ is not the cube of $(3, \alpha - 1)$ in view of the actual prime ideal factorization of $3\mathscr{O}_K$ given above.

A rational prime p is called *ramified* in number field K if in the prime ideal factorization $\prod_i \mathfrak{p}_i^{e_i}$ of $p\mathscr{O}_K$ some e_i is larger than 1. The above shows that 3 is ramified in $\mathbf{Q}(10^{1/3})$, and last time we saw that no prime $p > 5$ is ramified in this cubic field (since if $p > 5$ then $t^3 - 10$ is clearly separable in $\mathbf{F}_p[t]$, so Dedekind's criterion applicable at such p ensures $e_i = 1$ for all i).

Our aim today is to prove a general characterization of ramified primes:

Theorem 18.2. *A prime $p \in \mathbf{Z}^+$ is ramified in K if and only if $p \mid \mathrm{disc}(\mathscr{O}_K)$. In particular, at most finitely many rational primes ramify in any specific K.*

Remark 18.3. For quadratic fields, this result is shown directly in the handout "Some Quadratic Factoring" (it works even for $p = 2$!). For the cubic field K in the preceding example we saw that $3 | \mathrm{disc}(\mathscr{O}_K)$ and that 3 is ramified in K, and that every prime $p > 5$ doesn't divide $\mathrm{disc}(\mathscr{O}_K)$ and is unramified in K. For every $K \neq \mathbf{Q}$ we will later discuss that $|\,\mathrm{disc}(\mathscr{O}_K)| > 1$, so for every $K \neq \mathbf{Q}$ there exists a rational prime ramified in K.

Remark 18.4. At the end of the handout "Dedekind's Factorization Criterion" we gave an explicit cubic field K with discriminant -4027 and checked directly that for the prime $p = 4027$, the factorization of $p\mathscr{O}_K$ has two prime factors of which one occurs with multiplicity 2 (so p is ramified).

To prove Theorem 18.2, we choose a $\mathbf{Z}$-basis $\{v_i\}$ of $\mathscr{O}_K$, so $\oplus \mathbf{Z}v_i = \mathscr{O}_K$. We'd like to find the primes dividing $\mathrm{disc}\, \mathscr{O}_K = \det(\mathrm{Tr}_{K/\mathbf{Q}}\, v_i v_j)$. Reducing $\mathscr{O}_K$ modulo p, we get the ring

$$A := \mathscr{O}_K / p\mathscr{O}_K = \oplus \mathbf{F}_p \overline{v}_i$$

that is also a finite dimensional $\mathbf{F}_p$-vector space with basis consisting of the elements $\overline{v}_i := v_i \bmod p$.

For any $a \in \mathscr{O}_K$, the reduction modulo p of the matrix for the multiplication operator $m_a : \mathscr{O}_K \to \mathscr{O}_K$ relative to the $\mathbf{Z}$-basis $\{v_i\}$ is the matrix for multiplication operator $m_{a \bmod p} : A \to A$ relative to the basis $\{\overline{v}_i\}$. Hence, the residue class $\mathrm{Tr}_{K/\mathbf{Q}}(a) \bmod p \in \mathbf{F}_p$ coincides with the mod-p trace $\mathrm{Tr}_{A/\mathbf{F}_p}(a \bmod p)$. Applying this to $a = v_i v_j$ for each i, j gives

$$\mathrm{Tr}_{A/\mathbf{F}_p}(\overline{v}_i \overline{v}_j) = \mathrm{Tr}_{K/\mathbf{Q}}(v_i v_j) \bmod p,$$

so

$$\det(\mathrm{Tr}_{A/\mathbf{F}_p} \overline{v}_i \overline{v}_j) = \mathrm{disc}\, \mathscr{O}_K \bmod p.$$

Our task has reduced to showing that

$$\det(\mathrm{Tr}_{A/\mathbf{F}_p}(\overline{v}_i \overline{v}_j)) = 0 \text{ in } \mathbf{F}_p \iff p \text{ is ramified in } K.$$

Recall that if we compute the left side for some other $\mathbf{F}_p$-basis of $A = \mathscr{O}_K/(p)$ (not necessarily arising from a $\mathbf{Z}$-basis of $\mathscr{O}_K$!) then the determinant changes by multiplication against a nonzero square factor. Hence, the vanishing or not of the left side is *unaffected* by replacing $\{\overline{v}_i\}$ with any desired ordered $\mathbf{F}_p$-basis of A.

We shall now find a more convenient such $\mathbf{F}_p$-basis by using the prime factorization $\prod_i \mathfrak{p}_i^{e_i}$ of $p\mathscr{O}_K$. By the Chinese Remainder Theorem, we have a *ring isomorphism*

$$A = \mathscr{O}_K/p\mathscr{O}_K \simeq (\mathscr{O}_K/\mathfrak{p}_1^{e_1}) \times (\mathscr{O}_K/\mathfrak{p}_2^{e_2}) \times \cdots \times (\mathscr{O}_K/\mathfrak{p}_g^{e_g}).$$

Define $A_i := \mathscr{O}_K/\mathfrak{p}_i^{e_i}$; each A_i is then an $\mathbf{F}_p$-vector space.

Consider an $\mathbf{F}_p$-basis $\{w_i\}$ for A adapted to this product decomposition: $\{w_1, \dots, w_{m_1}\}$ is a basis for A_1, $\{w_{m_1+1}, \dots, w_{m_2}\}$ is a basis for A_2, and in general $\{w_{m_{s-1}+1}, \dots, w_{m_s}\}$ is a basis for A_s. Since the product of an element of A_i and an element of A_j inside A vanishes whenever $i \neq j$, the matrix

$$(\mathrm{Tr}_{A/\mathbf{F}_p}(w_i w_j))$$

has vanishing entries except possibly when w_i and w_j arise in the same factor ring A_s. Thus, this matrix of traces is a block-diagonal matrix for which the sth block corresponds to A_s. Likewise, if $a \in A_s$ then

$$\mathrm{Tr}_{A/\mathbf{F}_p}(a) = \mathrm{Tr}_{A_s/\mathbf{F}_p}(a)$$

because $m_a : A \to A$ has matrix given by $m_a : A_s \to A_s$ as one block and 0 everywhere else (since $x \mapsto ax$ from $A = \prod A_i$ into $A = \prod A_i$ kills A_i for $i \neq s$ since $a \in A_s$ and carries A_s into itself via a-multiplication on A_s).

We conclude that if w_i, w_j are part of the basis of A_s then $\mathrm{Tr}_{A/\mathbf{F}_p}(w_i w_j) = \mathrm{Tr}_{A_s/\mathbf{F}_p}(w_i w_j)$ and otherwise $\mathrm{Tr}_{A/\mathbf{F}_p}(w_i w_j) = 0$ (since $w_i w_j = 0$ in all other cases), so

$$\det(\mathrm{Tr}_{A/\mathbf{F}_p} w_i w_j) = \prod_{s=1}^{g} \det(\mathrm{Tr}_{A_s/\mathbf{F}_p} w_{m_{s-1}+i} w_{m_{s-1}+j})$$

(with the sth determinant applied to a matrix whose indices i, j vary between 1 and $m_s - m_{s-1} = \dim A_s$). In other words, up to non-zero square multipliers in $\mathbf{F}_p^\times$ that are harmless for checking vanishing, we have

$$\mathrm{disc}(A/\mathbf{F}_p) = \prod_{i=1}^{g} \mathrm{disc}(A_i/\mathbf{F}_p).$$

So to summarize, $p | \mathrm{disc}(\mathscr{O}_K)$ (equivalently, $\mathrm{disc}(A/\mathbf{F}_p) = 0$) if and only if *some* $\mathrm{disc}(A_i/\mathbf{F}_p)$ vanishes. But p is ramified in K if and only if some prime factor $\mathfrak{p}_i$ of $p\mathscr{O}_K$ occurs with multiplicity $e_i > 1$. Hence, our task reduces to:

Lemma 18.5. *For a maximal ideal $\mathfrak{p}$ of $\mathscr{O}_K$, integer $e \geq 1$, and $R := \mathscr{O}_K/\mathfrak{p}^e$,*

$$\mathrm{disc}(R/\mathbf{F}_p) = 0$$

if and only if $e > 1$.

Proof. If $e = 1$ then R is a finite field, and it is separable over $\mathbf{F}_p$ (as for any extension between finite fields), so its discriminant over $\mathbf{F}_p$ is nonzero. Now suppose $e > 1$. We want to show

$$\det(\mathrm{Tr}_{R/\mathbf{F}_p}(r_i r_j)) = 0$$

for *some* $\mathbf{F}_p$-basis $\{r_i\}$ of R (which implies the same for *all* $\mathbf{F}_p$-bases of R). We will do this by showing the corresponding matrix $(\mathrm{Tr}_{R/\mathbf{F}_p}(r_i r_j))$ has a row of 0's upon making a good choice of the $\mathbf{F}_p$-basis (which we are free to choose at will for the purpose of analyzing the vanishing or not of the determinant).

Since $e > 1$, we have $\mathfrak{p} \neq \mathfrak{p}^e$ (by unique factorization of prime ideals, for example). Hence, we may choose a nonzero element $x \in \mathfrak{p}/\mathfrak{p}^e$ and arrange that this is the first member r_1 of an ordered $\mathbf{F}_p$-basis $\{r_i\}$ of R. We claim that in this case the entire first row of the matrix of traces vanishes, which is to say each trace $\mathrm{Tr}_{R/\mathbf{F}_p}(r_1 r_j)$ vanishes. Note that $r_1 r_j \in R = \mathscr{O}_K/\mathfrak{p}^e$ belongs to $\mathfrak{p}/\mathfrak{p}^e$ since r_1 was chosen from $\mathfrak{p}/\mathfrak{p}^e$, so $(r_1 r_j)^e = 0$ in $R = \mathscr{O}_K/\mathfrak{p}^e$.

The vanishing of the trace of each $r_1 r_j$ is now seen to be a special case of the general assertion that for any nilpotent $r \in R$, $\mathrm{Tr}_{R/\mathbf{F}_p}(r) = 0$. To prove this general vanishing, note first that if $r^N = 0$ then the multiplication operator $m_r : R \to R$ is nilpotent: its Nth iterate is multiplication against $r^N = 0$. Since $\mathrm{Tr}_{R/\mathbf{F}_p}(r)$ is the trace of m_r by definition, it is enough to prove a general fact from linear algebra: if $T : V \to V$ is a nilpotent endomorphism of a finite-dimensional (nonzero) vector space V over any field k then $\mathrm{Tr}(T) = 0$.

The trace of a matrix can be computed after any extension of the field k over which we are working. We can extend k a finite amount to split the characteristic polynomial of T, so T has upper-triangular matrix M relative to a suitable basis. But when multiplying upper-triangular matrices, the effect along the diagonal is term-wise multiplication. Thus, nilpotence of an upper-triangular matrix forces all diagonal entries to *vanish*. (Concretely, all eigenvalues are equal to 0.) Hence, the trace of T vanishes since it is the sum of the diagonal entries of M. □

19. Relative factorization and rings of fractions

Always remember that when we write $\mathrm{disc}(K)$, we really mean $\mathrm{disc}(\mathscr{O}_K)$ (or more accurately $\mathrm{disc}_{\mathbf{Z}}(\mathscr{O}_K)$). This is a common abuse of notation. (It is similar to the abuse of terminology "prime of K" to mean "maximal ideal of $\mathscr{O}_K$": as a field K has only has one prime ideal, namely 0.)

Let's begin today with an interesting application of the result from last time that $p \mid \mathrm{disc}(K) \iff p\mathbf{Z}$ is ramified in $\mathscr{O}_K$ (see Theorem 18.2):

Example 19.1 (Dedekind)**.** Let $K = \mathbf{Q}(\theta)$ with $\theta^3 + \theta^2 - \theta + 8 = 0$. We saw in Example 9.3 that $\mathrm{disc}\,\mathbf{Z}[\theta] = -4 \cdot 503$, so if $\mathscr{O}_K \neq \mathbf{Z}[\theta]$ (i.e., if we can find an algebraic integer in K not belonging to $\mathbf{Z}[\theta]$) then $[\mathscr{O}_K : \mathbf{Z}[\theta]] = 2$ and $\mathrm{disc}(K) = -503$. In fact $(1/2)(\theta + \theta^2) \in \mathscr{O}_K - \mathbf{Z}[\theta]$, so $\mathrm{disc}(K) = -503$. From this, we conclude 2 is *unramified* in K since $2 \nmid -503$.

Thus, $2\mathscr{O}_K$ is either prime, a product of two distinct primes, or a product of three distinct primes. The first two of these options can be ruled out, so the final one must hold; i.e., even without knowing how to explicitly describe $\mathscr{O}_K$, it can be inferred that $2\mathscr{O}_K$ is a product of three distinct maximal ideals. This in turn can be used to show that $\mathscr{O}_K$ is *not* monogenic.

The handout "A Non-primitive Ring of Integers" provides full details on this example (including a clean proof that $(1/2)(\theta + \theta^2) \in \mathscr{O}_K$).

In order to better understand which features of $\mathscr{O}_K$'s are really "number theory" and which parts are general features of commutative ring theory or at least general features of Dedekind domains (without any reliance on finiteness of residue fields or other special properties of rings of integers), now consider the following general setup. Let A be a Dedekind domain and $F = \mathrm{Frac}(A)$. Let F' be a finite separable field extension of F with degree n, and $A' \subset F'$ the integral closure of A in F', as in the following diagram:

(19.1)
$$\begin{tikzcd} F' \arrow[d, dash] & A' \arrow[l, hook'] \arrow[d, dash] \\ F & A \arrow[l, hook'] \end{tikzcd}$$

We have seen earlier that A' is Dedekind and A-finite, with $\mathrm{Frac}(A') = F'$.

Example 19.2.
1. If $A = \mathscr{O}_K$ for a number field $F = K$ then $A' = \mathscr{O}_{F'}$.
2. Take $A = k[t]$ and $F = k(t)$ for a field k with $\mathrm{char}(k) \neq 2$, and $F' = k(t)[y]/(y^2 - f)$ for $f \in k[t]$ squarefree. Then, $A' = k[t,y]/(y^2 - f)$. This correspond to the algebraic curve "$y^2 = f(t)$"; to make precise what is meant by "algebraic curve" and how Dedekind domains inform the study of such geometric objects is one of the main points of Math 145. This provides useful geometric intuition for general questions with Dedekind domains and in particular questions in number theory (where a geometric picture isn't readily apparent).

A key case of interest for us will be the case that F'/F is Galois. The basic topic we want to understand is the following vague question:

Question 19.3. For $\mathfrak{m}$ a maximal ideal of A, what can we say about the factorization $\mathfrak{m}A' = \prod_{i=1}^{g}(\mathfrak{m}_i')^{e_i}$.

If we write $\mathfrak{m}A' = \prod_{i=1}^{g}(\mathfrak{m}_i')^{e_i}$ for maximal ideals $\mathfrak{m}_i'$ of A' and $e_i \geq 1$ then $\mathfrak{m}_i' \cap A = \mathfrak{m}$ (as the containment "$\supset$" is clear because each prime factor $\mathfrak{m}_i'$ of $\mathfrak{m}A'$ contains $\mathfrak{m}A'$ and hence contains the *maximal* ideal $\mathfrak{m}$, and $\mathfrak{m}_i' \cap A$ is a proper ideal because $1 \notin \mathfrak{m}_i'$). Moreover, the residue field extension $A/\mathfrak{m} \to A'/\mathfrak{m}_i'$ has *finite* degree $f_i = [A'/\mathfrak{m}_i' : A/\mathfrak{m}]$ because A' is A-finite. With such invariants e_i and f_i in hand, we may wonder:

Question 19.4. As we proved in the case that $A = \mathbf{Z}$, does it still hold that $\sum_{i=1}^{g} e_i f_i = n$? Our proof of this for $\mathbf{Z}$ used in an essential way that $\mathbf{Z}$ is a PID. Might a better method of proof apply more widely?

Question 19.5. Are all but finitely many such $\mathfrak{m}$ unramified in A'? That is, are there only finitely many $\mathfrak{m}$ for which some $e_i > 1$? For $A = \mathbf{Z}$ this has been seen using $\operatorname{disc}(K)$; is there a generalization of this discriminant construction to more general Dedekind A in place of $\mathbf{Z}$?

In general, we'd like to see to what extent properties we have previously proved for $\mathbf{Z}$ and rings of integers hold more widely for Dedekind domains or perhaps in more general ring-theoretic situations.

There are two basic problems with carrying out such generalizations:

I. If A is not a PID, then A' may not be A-free. In HW3 we saw that this phenomenon already occurs in cases with A the ring of integers of a quadratic field F, so this issue is already relevant when considering rings of integers as extensions of other rings of integers.

In the A-free case, as holds when A is a PID, we can show $\mathfrak{m}A'$ is a proper ideal, or equivalently that $A'/\mathfrak{m}A' \neq 0$, as follows. If $A' = A^{\oplus r}$, then

$$\begin{aligned} A'/\mathfrak{m}A' &= A^{\oplus r}/\mathfrak{m}A^{\oplus r} \\ &= A^{\oplus r}/\mathfrak{m}^{\oplus r} \\ &= (A/\mathfrak{m})^{\oplus r} \\ &\neq 0. \end{aligned}$$

Without the crutch of A-freeness for A', how can we show that $\mathfrak{m}A' \neq A'$? That is, how can we show $A'/\mathfrak{m}A' \neq 0$? The answer will be to use the technique of localization, which we will begin to develop later today; it will allow us to reduce many problems with general Dedekind domains to the PID case!

II. If $\mathfrak{m}'$ is a maximal ideal of A', then $\mathfrak{m}' \cap A \subset A$ is prime since

$$A/(\mathfrak{m}' \cap A) \hookrightarrow A'/\mathfrak{m}'$$

as rings and any subring of a field is a domain. Further, the prime ideal $\mathfrak{m}' \cap A$ of A is nonzero, as we saw earlier by looking at the constant term $f(0) \in A - \{0\}$ for $f \in F[t]$ the minimal polynomial of any nonzero $a' \in \mathfrak{m}'$. Since $\mathfrak{m}' \cap A$ is a nonzero prime ideal of A, it is maximal since A is Dedekind.

But does the fact that $\mathfrak{m}' \cap A$ is a maximal ideal of A for $\mathfrak{m}'$ a maximal ideal of A' hold for module-finite ring extensions $A \to A'$ beyond the Dedekind case?

The question raised in (II) has an affirmative answer:

Proposition 19.6. *Suppose $A \hookrightarrow A'$ is a module-finite extension of rings. For any maximal ideal $\mathfrak{m}'$ in A', $\mathfrak{m}' \cap A$ is a maximal ideal of A.*

Example 19.7. The module-finiteness hypothesis on $A \to A'$ cannot be dropped. For instance, the ring map $\mathbf{Z} \to \mathbf{Q}$ is not module-finite and the preimage of the maximal ideal (0) of $\mathbf{Q}$ is the prime ideal (0) of $\mathbf{Z}$ that is not maximal. (Here we could replace $\mathbf{Z}$ with any domain that is not a field.)

We now prove Proposition 19.6.

Proof. Observe that $A/(\mathfrak{m} \cap A') \to A'/\mathfrak{m}'$ is a module finite ring extension, with the latter a field. In particular, $A/(\mathfrak{m}' \cap A)$ is a domain. Our goal is to show $A/(\mathfrak{m}' \cap A)$ is a field. By renaming $A/(\mathfrak{m}' \cap A)$ and $A'/\mathfrak{m}'$ as A and A' respectively, it is enough to show the following lemma. □

Lemma 19.8. *Suppose $A \hookrightarrow A'$ is a module-finite map of domains. If A' is a field then A is a field.*

The converse (i.e., a domain finite-dimensional over a field is itself a field) is a result we have proved and used earlier.

Proof. Pick $a \in A - \{0\}$, so we have an element $1/a \in A'$ since A' is a field. We want to show that actually $1/a \in A$.

Since A' is A-finite, we have that $1/a$ is integral over A. Hence, we have a relation

$$(1/a)^n + c_{n-1}(1/a)^{n-1} + \cdots + c_1(1/a) + c_0$$

for some $c_0, c_1, \ldots, c_{n-1} \in A$. Clearing denominators,

$$1 + c_{n-1}a + c_{n-2}a^2 + \cdots + c_1a^{n-1} + c_0a^n = 0.$$

Hence,

$$1 = a(-c_{n-1} - c_{n-2}a - \cdots - c_1a^{n-2} - c_0a^{n-1}).$$

Thus, the element $-c_{n-1} - c_{n-2}a - \cdots - c_1a^{n-2} - c_0a^{n-1} \in A$ is the multiplicative inverse of a. This shows $A - \{0\} \subset A^\times$, so A is a field. □

To address problem (I), the issue that A may not be a PID, we now begin to develop the theory of "rings of fractions."

Definition 19.9. For a domain A, a subset $S \subset A - \{0\}$ is a *multiplicative set* if $1 \in S$ and whenever $s, s' \in S$ then $ss' \in S$.

Here are three key examples:

(1) $S = \{1, a, a^2, a^3, \ldots\}$ for $a \in A - \{0\}$.

(2) $S = A - \mathfrak{p}$ for prime $\mathfrak{p} \subset A$ is a multiplicative set exactly by the definition of prime ideal ($a, b \notin \mathfrak{p} \implies ab \notin \mathfrak{p}$). That is, by unraveling definitions, for an ideal I of a domain A one sees that $A - I$ is a multiplicative subset if and only if I is prime. (The condition $1 \in S$ forces $I \neq A$ when $A - I$ is a multiplicative set, encoding the requirement in the definition of primality that prime ideals are proper ideals.)

(3) For a subring $B \subset A$ and a prime ideal $\mathfrak{q} \subset B$, then $S = B - \mathfrak{q} \subset A$ is multiplicative. (More generally, any multiplicative set in a subring of A is also a multiplicative set in A.) A case of much interest will be $B = \mathbf{Z}, A = \mathscr{O}_K, S = \mathbf{Z} - p\mathbf{Z}$ for a prime $p \in \mathbf{Z}^+$.

Definition 19.10. For $F = \mathrm{Frac}(A)$, we define

$$S^{-1}A := \{a/s \in F \mid a \in A, s \in S\}.$$

By inspecting how one adds and multiplies fractions, the definition of S being a multiplicative set ensures that $S^{-1}A \subset F$ is a subring of F containing A (e.g., it contains A because $1 \in S$). Further, $F = \mathrm{Frac}(A) \subset \mathrm{Frac}(S^{-1}A) \subset F$, forcing $\mathrm{Frac}(S^{-1}A) = F$.

Example 19.11. For $S = A - \{0\}$ we have $S^{-1}A = F$ by definition of F.

Example 19.12. For $A = \mathbf{Z}$ and $S = \{10^r\}_{r \geq 0}$, $S^{-1}A = \mathbf{Z}[1/10]$ consists of exactly those $q \in \mathbf{Q}$ such that q can be written with denominator only divisible by 2 and 5 (since any fraction $m/(2^a 5^b)$ can be expressed as $m'/10^{\max(a,b)}$ by multiplying the numerator and denominator by some factors or 2 or 5 to bring the exponents in the denominator to a state of equality).

Example 19.13. In general, for $I \subset A$ an ideal we see that

$$S^{-1}I = \{a/s \mid a \in I, s \in S\}$$

is an ideal of $S^{-1}A$, and in fact $S^{-1}I = I \cdot S^{-1}A$ as one checks from the definitions (do it!). For example,

$$6 \cdot \mathbf{Z}[1/10] = 3 \cdot \mathbf{Z}[1/10].$$

This illustrates that from an ideal-theoretic viewpoint the role of 2 has become multiplicatively invisible since 2 is a unit in $\mathbf{Z}[1/10]$ (explicitly, $1/2 =$

$5/10 \in \mathbf{Z}[1/10]$). In a sense we will make precise next time, when we invert an element $s \in A - \{0\}$ we "remove" all the prime ideals containing s.

Example 19.14. Let's take $S = \mathbf{Z} - 7\mathbf{Z}$. Then,

$$S^{-1}\mathbf{Z} = \{q \in \mathbf{Q} \mid \ q \text{ can be written with denominator not divisible by 7 }\}.$$

In this case, all primes of $\mathbf{Z}^+$ that are not 7 have become units in this ring. The ideal generated by 7 remains prime but $2, 3, 5, 11, 13, \ldots$ have all become units. So, $S^{-1}\mathbf{Z}$ is a PID in which (7) is the only maximal ideal. This example and variations on it will be explored at length in the coming lectures and HW7.

The key lesson is that by inverting everything in the complement of the maximal ideal $7\mathbf{Z}$, we obtain a new ring in which all other maximal ideals have disappeared. It will be shown later that for any Dedekind domain at all (not necessarily a PID), applying such a construction with the complement of any maximal ideal $\mathfrak{m}$ will always yield a PID with only one maximal ideal, corresponding in a precise (and useful!) sense to $\mathfrak{m}$.

20. Localization and prime ideals

For a domain A with fraction field F and a multiplicative set $S \subset A - \{0\}$, last time we defined the subring $S^{-1}A \subset F$; this is called a *localization* of A (for reasons related to algebraic geometry that would take much too long to explain here). For I an ideal of A we defined the ideal $S^{-1}I \subset S^{-1}A$ and noted that $S^{-1}I = I \cdot S^{-1}A$.

Remark 20.1. We have $S^{-1}I = (1)$, or equivalently $1 \in S^{-1}I$, if and only if $I \cap S \neq \emptyset$. The reason is simply that if $a/s = 1$ for $a \in I$ and $s \in S$ then $a = s \in S \cap I$.

Example 20.2. Consider $A = \mathbf{Z}$, $I = 4\mathbf{Z}$, and $S = \{10^r\}_{r \geq 0}$. We have $100 \in I \cap S$ and

$$I \cdot (S^{-1}\mathbf{Z}) = 4 \cdot \mathbf{Z}[1/10] \ni 1$$

since $1 = (4 \cdot 25)/10^2$.

We now introduce the following notation:

Definition 20.3. (1) For $S = \{1, a, a^2, \ldots\}$ with $(a \in A - \{0\})$ we denote $S^{-1}A$ as A_a or $A[1/a]$.

(2) Let $S = A - \mathfrak{p}$ for prime $\mathfrak{p}$ we write $A_{\mathfrak{p}}$ to denote $S^{-1}A$; this is called "localizing at $\mathfrak{p}$."

Example 20.4. Using the above notation, for $A = \mathbf{Z}$, we have

$$\mathbf{Z}_{(7)} = \{q \in \mathbf{Q} \,|\, 7 \nmid \mathrm{denom}(q)\}$$

while

$$\mathbf{Z}_7 = \mathbf{Z}[1/7] = \{m/7^r \,|\, m \in \mathbf{Z}, r \geq 0\}.$$

The first is the localization at the prime ideal $7\mathbf{Z}$ while the second is the localization at the multiplicative set generated by 7. Do not confuse these. (Further, do not confuse either of these with the ring of 7-adic integers, which is closely related to $\mathbf{Z}_{(7)}$ and not at all related to $\mathbf{Z}_7$ but is also denoted $\mathbf{Z}_7$! We will not use p-adic integers in this course.)

Theorem 20.5. *Let A be a domain and $S \subset A - \{0\}$ be a multiplicative set.*

1. *For any ideal $J \subset S^{-1}A$ we have $J = S^{-1}I$ for an ideal $I \subset A$. Further, we may take I to be $J \cap A$, the "ideal of numerators of J".*
2. *If A is noetherian then $S^{-1}A$ is noetherian. Likewise, if A is a PID then $S^{-1}A$ is a PID.*
3. *Consider*

$$\begin{tikzcd} F' & A' \arrow[l, hook'] \\ F \arrow[u, dash] & A \arrow[l, hook'] \arrow[u, dash] \end{tikzcd} \tag{20.1}$$

with F'/F a finite extension of fields, $F := \mathrm{Frac}(A)$, and A' the integral closure of A in F'. Then $S^{-1}A'$ is the integral closure of $S^{-1}A$ in F'. In particular, using $F' = F, A' = A$: if A is integrally closed then so is $S^{-1}A$.

Proof. We prove the parts in order:

1. Consider $x \in J$, so $x = a/s$ for some $a \in A$ and $s \in S$. Then $a = sx \in A \cap J =: I$. Therefore, $x = a/s \in S^{-1}I$, so $J \subset S^{-1}I$. Further, $I \subset J$ by construction, so $S^{-1}I = I \cdot S^{-1}A \subset J$ since J is an ideal of $S^{-1}A$. Hence, $J = S^{-1}I$.
2. Given an ascending chain of ideals $\{J_n\}$ in $S^{-1}A$, we obtain an ascending chain of corresponding ideals $I_n := J_n \cap A$ in A. Therefore, since A is noetherian, $\{I_n\}$ stabilizes, so $S^{-1}I_n = J_n$ also stabilizes in $S^{-1}A$. Therefore, $S^{-1}A$ is noetherian. (Alternatively, if J is an ideal of $S^{-1}A$ then its ideal of numerators I in A has a finite set of generators $\{a_1, \ldots, a_m\}$ and so the ideal $J = I \cdot S^{-1}A$ is generated by the a_i's in $S^{-1}A$.)
3. Since $S^{-1}I = I \cdot (S^{-1}A)$, the PID implication is clear: if $J \subset S^{-1}A$ is an ideal, then $I := J \cap A$ is generated by some $a \in A$, and so a generates $I \cdot S^{-1}A = S^{-1}I = J$.

(4) To prove this statement concerning integral closure, one must chase many denominators in integrality relations. See [Samuel, Proposition 2, §5.1]. The notation there is slightly different: they use B instead of A', and R instead of F'.

□

We now want to study how localization interacts with prime ideals. This will be a powerful technique to establish for general Dedekind domains (which may not be PID's) statements we have previously shown for $\mathbf{Z}$ or for PID's.

Theorem 20.6. *We have a bijection*

$$\begin{tikzcd} \{\mathfrak{p} \subset A \,|\, \mathfrak{p} \textit{ is a prime ideal with } \mathfrak{p} \cap S = \emptyset\} \arrow[d] \\ \{\mathfrak{P} \,|\, \mathfrak{P} \textit{ is a prime ideal of } S^{-1}A\} \end{tikzcd} \tag{20.2}$$

defined by $\mathfrak{p} \mapsto S^{-1}\mathfrak{p}$ with inverse given by $\mathfrak{P} \mapsto A \cap \mathfrak{P}$. Further, this bijection is inclusion-preserving in both directions.

Moreover, if $\mathfrak{p}$ is a prime of A not meeting S, corresponding to a prime $\mathfrak{P}$ of $S^{-1}A$, then naturally as rings

$$\overline{S}^{-1}(A/\mathfrak{p}) \simeq (S^{-1}A)/\mathfrak{P}$$

where $\overline{S} \subset (A/\mathfrak{p}) - \{0\}$ denotes the image of S in $A/\mathfrak{p}$. In particular, if $\mathfrak{p}$ is maximal in A then $\mathfrak{P}$ is maximal in $S^{-1}A$ and their residue fields naturally coincide.

Warning 20.7. The above is not saying maximal ideals of $S^{-1}A$ correspond to maximal ideals of A. It may be that a maximal ideal of $S^{-1}A$ corresponds to a prime ideal of A contained in a bigger prime ideal $\mathfrak{m}$ of A (e.g., a maximal ideal of A) satisfying $\mathfrak{m} \cap S \neq \emptyset$. For example, take $A = \mathbf{Z}$, $S = A - \{0\}$ (so $S^{-1}A = \mathbf{Q}$), and $\mathfrak{P} = (0)$. Then, $\mathfrak{p} = (0)$ is not maximal in A.

Proof. Let's first check that this recipe makes sense in both directions: each operation on prime ideals lands in the asserted target set. For a prime $\mathfrak{P}$ of $S^{-1}A$, we have an inclusion of rings

$$A/(A \cap \mathfrak{P}) \hookrightarrow (S^{-1}A)/\mathfrak{P},$$

where the latter is a domain, so $A/(A \cap \mathfrak{P})$ is a domain and hence $A \cap \mathfrak{P}$ is prime. Further, $A \cap \mathfrak{P}$ is disjoint from S because of the containment

$$A \cap \mathfrak{P} \subset \mathfrak{P}$$

with $\mathfrak{P}$ a proper ideal of $S^{-1}A$ implying $\mathfrak{P} \cap (S^{-1}A)^\times = \emptyset$ (and $S \subset (S^{-1}A)^\times$).

Now, take a prime $\mathfrak{p}$ of A with $\mathfrak{p} \cap S = \emptyset$. We have $S^{-1}\mathfrak{p} \neq (1)$ because $\mathfrak{p} \cap S = \emptyset$. Next, to show $S^{-1}\mathfrak{p}$ is prime in $S^{-1}A$ we have to show that for $x = a/s, y = b/t \in S^{-1}A$ with $s, t \in S, a, b \in A$ and $xy \in S^{-1}\mathfrak{p}$ then either x or y lies in $S^{-1}\mathfrak{p}$. Clearly

$$\frac{ab}{st} = xy = \frac{c}{s'}$$

for some $c \in \mathfrak{p}, s' \in S$. Cross-multiplying, we obtain

$$s'ab = cst \in \mathfrak{p}$$

Since $s' \notin \mathfrak{p}$ (because $\mathfrak{p} \cap S = \emptyset$), either $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$, which means either x or y lies in $S^{-1}\mathfrak{p}$.

Next, let us check the two constructions on prime ideals are inverse to each other. Given a prime $\mathfrak{p} \subset A$ with $\mathfrak{p} \cap S = \emptyset$, we know $A \cap S^{-1}\mathfrak{p} \supset \mathfrak{p}$. We want to show this containment is an equality. Suppose we have $a \in A$ satisfying $a = b/s$ for some $b \in \mathfrak{p}$ and $s \in S$. We want to show $a \in \mathfrak{p}$. Clearly $sa = b \in \mathfrak{p}$ and $s \notin \mathfrak{p}$ since $s \in S$, so $a \in \mathfrak{p}$.

For the reverse composition, give $\mathfrak{P} \subset S^{-1}A$, we have $S^{-1}(\mathfrak{P} \cap A) = \mathfrak{P}$ by the argument with the ideal of numerators construction in the proof of Theorem 20.5(1) (so this direction doesn't use that $\mathfrak{P}$ is prime). This completes the verification of the desired bijection. By construction, inclusion-preservation in both directions is clear.

Next, for corresponding $\mathfrak{p}$ and $\mathfrak{P}$, we want to identify $\overline{S}^{-1}(A/\mathfrak{p})$ with $(S^{-1}A)/\mathfrak{P}$. Recall $\mathfrak{p} = A \cap \mathfrak{P}$, so we have a subring inclusion of *domains*

$$A/\mathfrak{p} \hookrightarrow (S^{-1}A)/\mathfrak{P}.$$

This induces a containment $\mathrm{Frac}(A/\mathfrak{p}) \subset \mathrm{Frac}((S^{-1}A)/\mathfrak{P})$ between their fraction fields (as for any inclusion of domains!). Thus,

$$\overline{S}^{-1}(A/\mathfrak{p}) \subset (S^{-1}A)/\mathfrak{P}$$

inside $\mathrm{Frac}((S^{-1}A)/\mathfrak{P})$ because the image of any element of $\overline{S}$ in $S^{-1}A/\mathfrak{P}$ is visibly invertible. To show the above containment is an equality, pick any $x \in (S^{-1}A)/\mathfrak{P}$ and a fraction representative $a/s \in S^{-1}A$. Letting $\overline{s} \in \overline{S}$ be the image of s, the fraction $(a \bmod \mathfrak{p})/\overline{s} \in \overline{S}^{-1}(A/\mathfrak{p})$ is carried to x, as we leave to be checked by unraveling definitions. $\square$

Corollary 20.8. *If A is Dedekind and $A' = S^{-1}A$ with $S^{-1}A \neq F := \mathrm{Frac}(A)$ then A' is Dedekind and we have a bijection*

$$\begin{aligned}\mathrm{Max}(A') &\to \{\mathfrak{m} \in \mathrm{Max}\, A \mid \mathfrak{m} \cap S = \emptyset\}\\ \mathfrak{m}' &\mapsto \mathfrak{m}' \cap A.\end{aligned}$$

Proof. Most of the conditions for A' to be a Dedekind domain are immediate from what we have already done: it is noetherian and integrally closed, with $A' \neq F$ by hypothesis (and clearly $F = \mathrm{Frac}(A')$). Further, nonzero primes of A' correspond to nonzero primes of A that are disjoint from S. But the latter are the maximal ideals of A disjoint from S (since A is Dedekind), so there are no strict containment relations among the latter primes. Thus, back in the localization A' it follows that there are no strict containment relations among its nonzero prime ideals, so all such must be maximal ideals of the noetherian domain A'. Hence, A' is Dedekind. □

21. Applications of localization

For us, localization is an algebraic mechanism to remove certain primes from a ring, while retaining information at other primes. Let's recall our running notational setup from the end of last time: let A be a Dedekind domain, $F := \mathrm{Frac}(A)$, and $A' := S^{-1}A$ for a multiplicative set $S \subset A - \{0\}$. Assume further that $A' \neq F$. Last time we saw that A' is Dedekind and that there is a bijection between $\mathrm{Max}(A')$ and the set of $\mathfrak{m} \in \mathrm{Max}(A)$ disjoint from S (via the operations $\mathfrak{m}' \mapsto \mathfrak{m}' \cap A$ and $\mathfrak{m} \mapsto S^{-1}\mathfrak{m} = \mathfrak{m}A'$). We also saw that naturally $A/\mathfrak{m} \simeq A'/\mathfrak{m}'$ because "inverting S mod $\mathfrak{m}$" in $A/\mathfrak{m}$ does nothing due to $A/\mathfrak{m}$ being a field.

The key case of this setup for our purposes is $S := A - \mathfrak{m}_0$ for a maximal ideal $\mathfrak{m}_0$ of A, in which case $A' = A_{\mathfrak{m}_0}$ and we have a bijection

$$\mathrm{Max}(A') \to \{\mathfrak{m} \in \mathrm{Max}(A) \mid \mathfrak{m} \cap S = \emptyset\}.$$

But the condition $\mathfrak{m} \cap S = \emptyset$ says exactly $\mathfrak{m} \subset \mathfrak{m}_0$, and since $\mathfrak{m}$ is maximal this is the same as saying $\mathfrak{m} = \mathfrak{m}_0$. In other words, A' has $\mathfrak{m}_0 A' = \mathfrak{m}_0 A_{\mathfrak{m}_0}$ as its *unique* maximal ideal. It follows that $A_{\mathfrak{m}_0}$ is a PID, due to:

Lemma 21.1. *If B is a Dedekind domain with only finitely many maximal ideals then B is a PID.*

Example 21.2. For example, in $\mathbf{Z}_{(p)}$ the unique maximal ideal is $\mathfrak{m} = p\mathbf{Z}_{(p)}$ and it inherits the PID property from $\mathbf{Z}$.

Proof. Say $\mathrm{Max}(B) = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_r\}$. By weak approximation, as shown in the homework, for any $e_1, \ldots, e_r \geq 0$, there exists $b \in B - \{0\}$ so that $\mathrm{ord}_{\mathfrak{p}_i}(b) = e_i$. Hence, $bB = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}\mathfrak{a}$ for a nonzero ideal $\mathfrak{a}$ not divisible by any $\mathfrak{p}_i$'s.

But the $\mathfrak{p}_i$'s exhaust all maximal ideals of B, so the only option is $\mathfrak{a} = (1)$, so $\prod \mathfrak{p}_i^{e_i} = (b)$. By prime ideal factorization for B, we see B is a PID.

(Alternatively, weak approximation provides elements $b_i \in B - \{0\}$ such that (b_i) is divisible by $\mathfrak{p}_i$ exactly once and not divisible by any other $\mathfrak{p}_j$, so from the viewpoint of prime ideal factorization this forces $(b_i) = \mathfrak{p}_i$. Thus, each maximal ideal of B is principal, so all nonzero proper ideals of B are principal due to the unique factorization into maximal ideals.) □

Remark 21.3. Here is a strange but valid (yet totally non-quantitative) argument for the infinitude of primes in $\mathbf{Z}^+$. Suppose there were only finitely many primes. Then there would only be finitely many prime ideals in $\mathscr{O}_K$ for any number field K because there are only finitely many primes in $\mathscr{O}_K$ lying over a given prime $(p) \subset \mathbf{Z}$. By the preceding lemma this would imply that $\mathscr{O}_K$ is always a PID. However, $\mathbf{Z}[\sqrt{-5}]$ is not a PID, contradiction!

This argument is due to Larry Washington. It really is not circular (i.e., up to now nothing in this course has relied on the infinitude of the set of primes in $\mathbf{Z}^+$).

Definition 21.4. A *discrete valuation ring* (dvr) is a Dedekind domain with a unique maximal ideal. Any generator of the maximal ideal is called a *uniformizer*.

Some properties of dvr's, along with a variety of explicit numerical computations with them, are developed in Homework 7.

Example 21.5. The ring $\mathbf{Z}_{(p)}$, with maximal ideal $\mathfrak{m} = p\mathbf{Z}_{(p)}$ is a dvr with uniformizer p (and any $\mathbf{Z}_{(p)}^\times$-multiple of p is also a uniformizer, such as $p/(p+1)$, $-p$, etc.). All nonzero $x \in \mathbf{Z}_{(p)}$ can be written uniquely in the form $x = p^n \cdot u$ for a unique $n \geq 0$ and $u \in \mathbf{Z}_{(p)}^\times$ (check!).

Example 21.6. Consider the ring of "formal power series"

$$\mathbf{C}[\![t]\!] := \left\{ \sum_{n \geq 0} a_n t^n \mid a_n \in \mathbf{C} \right\}$$

(one adds term-wise, and multiplies in the usual formal manner not involving convergence issues; e.g., the degree-3 coefficient of $f(t)g(t)$ only involves the coefficients of f and g in degree ≤ 3). For $f = \sum a_n t^n$ with $a_0 = f(0) \neq 0$, we can write

$$f = a_0(1 + tg)$$

for some $g \in \mathbf{C}[\![t]\!]$, so

$$\frac{1}{f} = \frac{1}{a_0} \cdot \frac{1}{1 + tg} \in \mathbf{C}[\![t]\!]$$

since the geometric series $\sum(-1)^n t^n g^n$ involves only finitely many coefficients in any specific degree due to the t^n-multiplier with $n \to \infty$ (so again there are no convergence issues when computing this as a formal power series).

It follows that any $f \neq 0$ can be written in the form $t^m \cdot u$ for a unique $m \geq 0$ and $u \in \mathbf{C}[\![t]\!]^\times$, and that (t) is the maximal ideal (in fact, the set of all non-units). In HW7 it is shown that for any domain R such that all nonzero elements are uniquely an $R^\times$-multiple of a power of a fixed nonzero non-unit π, necessarily R is a dvr with π as a uniformizer.

We next take up a variant of dvr's. Suppose we have a Dedekind domain A with fraction field F, a finite separable extension F' of F and let A' denote the A-finite integral closure of A in F', as in the diagram

(21.1)
$$\begin{tikzcd} F' & A' \arrow[l, hook'] \\ F \arrow[u, dash] & A \arrow[l, hook'] \arrow[u, dash] \end{tikzcd}$$

For a map of rings $A \to B$ and $\mathfrak{p}$ a prime in A, by abuse of notation we define $B_\mathfrak{p} := (A - \mathfrak{p})^{-1}B$.

For the unique maximal ideal $\mathfrak{m} \subset A$, let's localize throughout by $S = A - \mathfrak{m}$. The localization $S^{-1}A' =: A'_\mathfrak{m}$ is the integral closure of $A_\mathfrak{m}$ in F'/F, fitting into the diagram

(21.2)
$$\begin{tikzcd} F' & A'_\mathfrak{m} \arrow[l, hook'] \\ F \arrow[u, dash] & A_\mathfrak{m} \arrow[l, hook'] \arrow[u, dash] \end{tikzcd}$$

and we have a bijection

$$\mathrm{Max}(A'_\mathfrak{m}) \to \{\mathfrak{m}' \in \mathrm{Max}(A') \mid \mathfrak{m}' \text{ is disjoint from } A - \mathfrak{m}\}.$$

The latter condition is equivalent to $\mathfrak{m}' \cap A \subset \mathfrak{m}$. But we know $\mathfrak{m}' \cap A$ must be maximal, so necessarily $\mathfrak{m}' \cap A = \mathfrak{m}$, or equivalently (as we have seen in our earlier work with Dedekind domains) $\mathfrak{m}'$ *is a prime factor of* $\mathfrak{m}A'$. Since $A'_\mathfrak{m}$ is Dedekind with only finitely many maximal ideals (corresponding to the prime ideal factors of $\mathfrak{m}A'$ in A'), it follows that $A'_\mathfrak{m}$ is a PID.

Example 21.7. Localizing at $\mathbf{Z} - p\mathbf{Z}$, if K is a number field then the integral closure of $\mathbf{Z}_{(p)}$ in K is $\mathscr{O}_{K,(p)}$ in which the finitely many maximal ideals correspond to the prime ideal factors of $p\mathscr{O}_K$.

Lemma 21.8. *For $\mathfrak{m}'$ lying over $\mathfrak{m}A'$ the diagram of field maps*

$$\begin{tikzcd}
A'/\mathfrak{m}' \arrow[r, "\simeq"] & A'_{\mathfrak{m}}/\mathfrak{m}'A'_{\mathfrak{m}} \\
A/\mathfrak{m} \arrow[r, "\simeq"] \arrow[u] & A_{\mathfrak{m}}/\mathfrak{m}A_{\mathfrak{m}} \arrow[u]
\end{tikzcd} \tag{21.3}$$

commutes. In particular, the residue field degree $f(\mathfrak{m}'|\mathfrak{m}) := [A'/\mathfrak{m}' : A/\mathfrak{m}]$ is unaffected by localizing throughout at $\mathfrak{m}$.

Proof. This is just a matter of chasing through the definitions of the maps; it is an instructive exercise. □

Lemma 21.9. *If $\prod_{i=1}^{g}(\mathfrak{m}'_i)^{e_i}$ is the prime factorization of $\mathfrak{m}A'$ in A', then*

$$\prod_{i=1}^{g}(\mathfrak{m}'_i A'_{\mathfrak{m}})^{e_i}$$

is the prime factorization of $\mathfrak{m}A'_{\mathfrak{m}}$ in $A'_{\mathfrak{m}}$.

That the displayed product coincides with $\mathfrak{m}A'_{\mathfrak{m}}$ is the (easily verified!) compatibility of products of ideals with localization at a multiplicative set, and we have seen above that the ideals $\mathfrak{m}'_i A'_{\mathfrak{m}}$ are precisely the maximal ideals of $A'_{\mathfrak{m}}$ (no repetition!), so since the e_i's are each at least 1 it follows that we do indeed have the prime ideal factorization of $\mathfrak{m}A'_{\mathfrak{m}}$. This lemma yields the important consequence that e_i is "unaffected" by localizing at $\mathfrak{m}$.

As an application of the preceding work with localization, now we can prove Theorem 17.1 in the more general context of extensions of Dedekind domains, going beyond our earlier work for $\mathbf{Z} \to \mathscr{O}_K$:

Theorem 21.10. *For A, F, F', A' as above and $n = [F' : F]$, let e_i be the multiplicity of $\mathfrak{m}'_i$ as a factor of $\mathfrak{m}A'$ and let $f_i := [A'/\mathfrak{m}'_i : A/\mathfrak{m}]$. Then*

$$\sum_{i=1}^{g} e_i f_i = n.$$

Proof. We can pass to the localization $A_{\mathfrak{m}} \to A'_{\mathfrak{m}}$ without affecting anything, so now we can assume A is a dvr and then A' is also a PID (since it has only finitely many maximal ideals). The PID property for A ensures that A' is now A-free, which is to say $A' = A^{\oplus r}$ as an A-module for some $r \geq 1$ (freeness typically does *not* hold in the initial setup prior to localization).

We claim that $r = n$. By localizing at $A - \{0\}$, we get

$$F' = (A - \{0\})^{-1}A' = F^{\oplus r}$$

as an F-module. Thus, $r = \dim_F F' = n$, as desired. Therefore, we have now established $A' = A^{\oplus n}$.

Define π to be a uniformizer for the dvr A, so $\mathfrak{m} = \pi A$. Reducing modulo $\mathfrak{m}$, we have

$$A'/\pi A' \simeq (A/\mathfrak{m})^{\oplus n}$$

has dimension n as an $A/\mathfrak{m}$-vector space. Let's compute $A/\mathfrak{m}$-dimension of the left side in another way to get $\sum_i e_i f_i$.

By the Chinese Remainder Theorem,

$$A'/\mathfrak{m}A' \simeq \prod_{i=1}^{g} (A'/(\mathfrak{m}_i')^{e_i})$$

as an A'-module (and as rings). Consider these as vector spaces over $k := A/\mathfrak{m}$, so computing k-dimensions of both sides yields

$$n = \sum_{i=1}^{g} \dim_k(A'/(\mathfrak{m}_i')^{e_i}).$$

It is therefore enough to show that

$$e_i f_i = \dim_k(A'/(\mathfrak{m}_i')^{e_i})$$

for all i, recalling $f_i = [k_i' : k]$.

We have the chain of k-subspaces

$$A'/\mathfrak{m}_i'^{e_i} \supset \mathfrak{m}_i'/\mathfrak{m}_i'^{e_i} \supset \mathfrak{m}_i'^2/\mathfrak{m}_i'^{e_i} \supset \cdots \supset \mathfrak{m}_i'^{e_i-1}/\mathfrak{m}_i'^{e_i} \supset (0).$$

In this chain there are e_i steps, and recall that for a subspace $W \subset V$ of a finite-dimensional vector space V over a field we have

$$\dim V = \dim W + \dim V/W.$$

Thus,

$$\dim_k A'/\mathfrak{m}_i'^{e_i} = \sum_{j=1}^{e_i} \dim_k(\mathfrak{m}_i'^{j-1}/\mathfrak{m}_i'^{j}),$$

so it suffices to show each of the e_i terms in the sum equals $f_i = [k_i' : k] = \dim_k k_i'$ (as then the sum of f_i of these is equal to $e_i f_i$ as desired).

Since A' is a PID, so $\mathfrak{m}_i'$ has a generator π_i', we have a k-linear (even k_i'-linear) map

$$k_i' = A'/\mathfrak{m}_i' \to \mathfrak{m}_i'^{j-1}/\mathfrak{m}_i'^{j} = (\pi_i')^{j-1}/(\pi_i')^{j}$$

defined by $x \mapsto \pi_i'^{j-1} x$ that is easily seen (check!) to be bijective and hence an isomorphism. Therefore, each of the e_i terms in the preceding sum is indeed equal to f_i. □

Corollary 21.11. *The number g of primes $\mathfrak{m}_i'$ over $\mathfrak{m}$ is at most n, and $g = n$ if and only if $e_i = f_i = 1$ for all i.*

Proof. We have $\sum_{i=1}^{g} e_i f_i = n$, so the result follows. □

Definition 21.12. When $g = n$, we say $\mathfrak{m}$ is *totally split* in F'.

Definition 21.13. A maximal ideal $\mathfrak{m}'$ of A' over $\mathfrak{m} \in \operatorname{Max}(A)$ is called *unramified* if both of the following hold:

- $e(\mathfrak{m}'|\mathfrak{m}) = 1$ (i.e., $\mathfrak{m}'$ occurs exactly once in the prime factorization of $\mathfrak{m}A'$),
- the residue field extension $A/\mathfrak{m} \to A'/\mathfrak{m}'$ is separable.

We say $\mathfrak{m} \in \operatorname{Max} A$ is *unramified* in F'/F if all $\mathfrak{m}_i'$ dividing $\mathfrak{m}A'$ are unramified over A (i.e., $\mathfrak{m}A'$ has no repeated prime factors and all residue field extensions are separable over $A/\mathfrak{m}$).

If $\mathfrak{m}' \in \operatorname{Max} A'$ is not unramified over A, we say $\mathfrak{m}'$ is *ramified over* A. If $\mathfrak{m} \in \operatorname{Max} A$ is not unramified in F'/F (i.e., some prime ideal factor of $\mathfrak{m}A'$ has multiplicity > 1 or involves a non-separable residue field extension), we say $\mathfrak{m}$ is *ramified* in F'/F.

Remark 21.14. The condition that $A/\mathfrak{m} \to A'/\mathfrak{m}'$ is separable is often invisible in first courses in algebraic number theory because the fields $A/\mathfrak{m}$ are usually perfect (meaning that their finite extensions are all separable), as occurs when the residue fields are finite (or characteristic 0). When $A/\mathfrak{m}$ is perfect, unramifiedness is equivalent to the condition $e(\mathfrak{m}'|\mathfrak{m}) = 1$.

Next time, we'll define a nonzero discriminant ideal $\operatorname{disc}(A'/A) \subset A$ with the property that $\mathfrak{m} \mid \operatorname{disc}(A'/A)$ if and only if $\mathfrak{m}$ is ramified in F'/F. In particular, all but finitely many maximal ideals of A are unramified in F'. We will make this ideal by bootstrapping from the PID case via localization.

22. Discriminant ideals

For a module-finite extension of Dedekind domains $A \to A'$, we want to define a nonzero ideal

$$\operatorname{disc}(A'/A) \subset A$$

recovering $\operatorname{disc}(\mathscr{O}_K/\mathbf{Z}) \cdot \mathbf{Z}$ for $A = \mathbf{Z}$ and satisfying the properties:

(1) $\operatorname{disc}(S^{-1}A'/S^{-1}A) = S^{-1}\operatorname{disc}(A'/A)$ for any multiplicative set $S \subset A - \{0\}$;

(2) for $\mathfrak{m} \in \operatorname{Max} A, \mathfrak{m} \mid \operatorname{disc}(A'/A)$ if and only if $\mathfrak{m}$ is ramified in F'/F (so in particular, only finitely many $\mathfrak{m}$ are ramified in F'/F).

Before undertaking this construction, we discuss an interesting application of the relationship already established between discriminants and ramification for number fields over $\mathbf{Q}$. Consider an odd prime p in $\mathbf{Z}^+$. The cyclotomic extension

$$\begin{tikzcd}\mathbf{Q}(\zeta_p) \arrow[d, dash] \\ \mathbf{Q}\end{tikzcd}\tag{22.1}$$

has Galois group $(\mathbf{Z}/p\mathbf{Z})^\times$ cyclic of even order $p-1$, so by the Galois correspondence there is a unique quadratic extension $K/\mathbf{Q}$ of degree 2:

$$\begin{tikzcd}\mathbf{Q}(\zeta_p) \arrow[dd, dash] \arrow[dr, dash] & \\ & K \arrow[dl, dash] \\ \mathbf{Q}. & \end{tikzcd}\tag{22.2}$$

What is K?

We know $K/\mathbf{Q}$ has ramified primes since $\mathrm{disc}(\mathscr{O}_K/\mathbf{Z}) \neq \pm 1$ for quadratic fields. Any prime ℓ of $\mathbf{Q}$ ramified in K is ramified in $\mathbf{Q}(\zeta_p)$ (since the presence of a repeated prime factor in $\ell\mathscr{O}_K$ forces the presence of a repeated prime factor in the ideal generated by $\ell\mathscr{O}_K$ in any finite extension of K). But $\mathrm{disc}(\mathbf{Z}[\zeta_p]/\mathbf{Z})$ is a (non-trivial) power of p up to sign, so the only possibility is $\ell = p$. Thus, the odd prime p is the unique prime ramified in K.

Writing $K = \mathbf{Q}(\sqrt{d})$ for some squarefree $d \in \mathbf{Z} - \{0, 1\}$, the discriminant must be $\pm p$ and hence is not divisible by 4, so necessarily $d \equiv 1 \bmod 4$ by our calculation of the discriminant for quadratic extensions of $\mathbf{Q}$. It follows that $d = \pm p$ with the sign uniquely chosen (depending on $p \bmod 4$) to make $d \equiv 1 \bmod 4$; in other words, $d = (-1|p)p$. This gives a pure-thought proof that $K = \mathbf{Q}(\sqrt{(-1|p)p})$.

An explicit proof that this quadratic field lies inside $\mathbf{Q}(\zeta_p)$, by exhibiting a specific element of $\mathbf{Z}[\zeta_p]$ ("Gauss sum") whose square is $(-1|p)p$, is given in [Samuel, (6), §5.5]. That is a very different argument from the preceding one based on ramification considerations.

Returning to the issue of discriminant ideals, last time we saw that for A a Dedekind domain, $F = \mathrm{Frac}\, A$, F'/F a separable extension of degree n, and A' the integral closure of A in F' as in the diagram

$$\begin{tikzcd} F' & A' \arrow[l, hook'] \\ F \arrow[u, dash] & A \arrow[l, hook'] \arrow[u, dash] \end{tikzcd} \tag{22.3}$$

the exponents e_i in the factorization $\mathfrak{m}A' = \prod_{i=1}^{g}(\mathfrak{m}_i')^{e_i}$ and the residue field degrees $f_i = [A'/\mathfrak{m}_i' : A/\mathfrak{m}]$ satisfy

$$\sum_{i=1}^{g} e_i f_i = n.$$

Thus, $g \leq n$ with $g = n$ if and only if all $e_i = 1, f_i = 1$. We said in this case that $\mathfrak{m}$ is *totally split* in F'/F. (When $F = \mathbf{Q}$ and $[F' : F] = 2$ then this recovers the notion of "split" for a rational prime in a quadratic field.)

Example 22.1. Which rational primes p are totally split in $\mathbf{Q}(\zeta_n)$? Using

$$\begin{tikzcd} \mathbf{Q}(\zeta_n) & \mathbf{Z}[\zeta_n] = \mathbf{Z}[x]/(\Phi_n) \arrow[l, hook'] \\ \mathbf{Q} \arrow[u, dash] & \mathbf{Z} \arrow[l, hook'] \arrow[u, dash] \end{tikzcd} \tag{22.4}$$

we know from Dedekind's criterion that p is totally split if and only if $\Phi_n \in \mathbf{F}_p[x]$ is a product of distinct monic degree-1 polynomials. Note that when this happens, necessarily $p \nmid n$ since a totally split prime is unramified and hence cannot divide the discriminant (and we know that the prime factors of the discriminant of $\mathbf{Q}(\zeta_n)$ are the prime factors of n when n is not twice an odd integer, and the prime factors of $n/2$ when n is twice an odd integer).

We claim that Φ_n splits into a product of distinct monic degree-1 factors over a field k with $\mathrm{char}(k) \nmid n$ if and only if $k^\times$ contains n distinct nth roots of unity. To see this, consider the identity

$$X^n - 1 = \prod_{d|n} \Phi_d$$

in $k[X]$, which holds because we can check it even in $\mathbf{Z}[X]$ (via consideration of roots in a big enough extension of $\mathbf{Q}$). This identity implies that when $k^\times$ has n distinct nth roots of unity, so $X^n - 1$ is a product of distinct monic degree-1 factors, then the same holds for its monic factor Φ_n. Conversely, if Φ_n has even a single degree-1 factor in $k[X]$, the corresponding root must be a *primitive* nth root of 1 (so $k^\times$ has n distinct nth roots of 1). Indeed, otherwise Φ_n would have a root in common with that of $X^m - 1$ for a proper divisor m of n, and hence with Φ_d for some $d|m$, but then the factor $\Phi_n\Phi_d$

of $X^n - 1$ in $k[X]$ would provide a repeated root, contradicting that $X^n - 1$ is separable over k (as $\mathrm{char}(k) \nmid n$).

Taking the field k to be $\mathbf{F}_p$, since $\mathbf{F}_p^\times$ is cyclic of order $p - 1$, it contains n distinct nth roots of unity if and only if this cyclic group contains a subgroup of order n, which is to say $n|(p-1)$ or equivalently $p \equiv 1 \bmod n$. Thus, p splits completely in $\mathbf{Q}(\zeta_n)$ if and only if $p \equiv 1 \bmod n$.

Remark 22.2. In general, class field theory shows that in a suitable sense splitting behaviors of primes in Galois extensions of number fields are governed by "congruence conditions" (suitably defined) when the Galois group is abelian. This is beyond the level of the course, but it puts Dirichlet's theorem on primes in arithmetic progressions into a wider Galois-theoretic framework for which it makes sense to consider *non-abelian* Galois groups too, leading to a vast generalization of Dirichlet's theorem called the Chebotarev Density Theorem (proved by a mixture of representation theory and complex analysis).

Even for the study of extensions of number fields, it is useful to have a general notion of discriminant ideal. For example, if we have $K/L/\mathbf{Q}$ an extension of number fields, it will be useful to study prime factorization relative to K/L in the study of $K/\mathbf{Q}$, and typically $\mathscr{O}_L$ will not be a PID (even when $[L : \mathbf{Q}] = 2$).

Construction of the discriminant in the free case. Let's consider the special case that A' is A-free. That is, suppose $A' = \oplus_{i=1}^n Ae_i$. In this case we will define

$$\mathrm{disc}(A'/A) = \delta \cdot A$$

with

$$\delta := \det(\mathrm{Tr}_{F'/F}(e_i e_j)).$$

Let's see that this is well-posed (i.e., independent of the A-basis of A') and nonzero. Firstly, $\{e_i\}$ is an F-basis of

$$(A - \{0\})^{-1}A' = F',$$

and we know that if we change $\{e_i\}$ to any F-basis of F', then the effect on δ is scaling by an element of $(F^\times)^2$, so $\delta \neq 0$ since we can check using a power basis (thanks to the separability of F'/F).

Note also that $\mathrm{Tr}_{F'/F}(A') \subset A$, so $\delta \in A$. If one changes $\{e_i\}$ to any other A-basis of A' then δ changes by multiplication against the square of the determinant of a change-of-basis matrix, and that matrix and its inverse both have all entries in A (since we're comparing two A-bases of A'), so the multiplier belongs to $(A^\times)^2$. In particular, the nonzero principal ideal $\delta \cdot A$ is independent of $\{e_i\}$.

Let's see that this construction for A-free A' satisfies our desired properties: it interacts well with localization and encodes the ramified primes of A relative to F'/F. For any multiplicative set $S \subset A - \{0\}$, we have $S^{-1}A' = \oplus S^{-1}Ae_i$, so by using the same basis we obtain the same δ and clearly $\delta S^{-1}A = S^{-1}(\delta A)$. This establishes the compatibility with localization. Next, we have:

Lemma 22.3. *If A' is A-free then the maximal ideals dividing δA are precisely those which ramify in F'/F.*

Proof. For a nonzero ideal $J \subset A$, we see that $J_\mathfrak{m} = JA_\mathfrak{m}$ encodes the $\mathfrak{m}$-part of J (as localizing at $\mathfrak{m}$ causes any maximal ideal of A distinct from $\mathfrak{m}$ to acquire a unit element and hence to generate the unit ideal in $A_\mathfrak{m}$). Hence, we can detect whether $\mathfrak{m} \mid \delta A$ after localizing at $\mathfrak{m}$. Further, we can also detect the data of ramification indices and residue field degrees at all $\mathfrak{m}'_i \mid \mathfrak{m}A'$ after localization at $\mathfrak{m}$. Hence, for the purposes of the proof we can first localize throughout at each such $\mathfrak{m}$ to reduce to the case that A is a dvr and hence a PID.

Now the same arguments we have given for $\mathbf{Z}$ work equally well, with one small addition given by the following fact applied to residue field extensions. This is a technical digression on inseparability that can be safely ignored for the purposes of this course (by limiting attention to Dedekind domains whose residue fields are perfect if you wish, such as finite fields or characteristic 0). The fact we need is that for a finite extension of fields k'/k, $\mathrm{disc}(k'/k) \neq 0$ if and only if k'/k is separable. To see this equivalence, if k'/k is separable then we can use the usual power basis formula. For the converse, the non-vanishing of

$$\mathrm{disc}(k'/k) = \det(\mathrm{Tr}_{k'/k}(e_ie_j))$$

implies that $\mathrm{Tr}_{k'/k} : k' \to k$ is nonzero. This implies k'/k is separable, as explained in Lemma 2.2 of the handout "Norm and Trace". □

Construction of the discriminant in the general case. We now construct the discriminant in general. The starting point is to show directly that all but finitely many $\mathfrak{m}$ are unramified in F'/F, and use this to build the global discriminant from "local" data.

Lemma 22.4. *All but finitely many $\mathfrak{m}$ are unramified in F'/F.*

Proof. Pick a primitive element α for F'/F, so $F' = F[\alpha]$. We may scale α by an element of $A - \{0\}$ so that $\alpha \in A'$. Hence, $A[\alpha] \subset A'$ and A' is A-finite. Looking at the F-coefficients of expansions of each of a finite set of

A-module generators of A' relative to expansions in the power basis $\{\alpha^i\}$ for F' over F, we have

$$A[\alpha] \subset A' \subset (1/a)A[\alpha]$$

for a sufficiently divisible "common denominator" $a \in A - \{0\}$. But localizing at a throughout (i.e., permitting denominators a^n for all $n \geq 1$) collapses the inclusion between outer terms to an equality, forcing the first inclusion to also localize to an equality.

In other words, $A'[1/a] = A[\alpha][1/a] = (A[1/a])[\alpha]$, so $A'[1/a]$ is $A[1/a]$-free using the basis of powers α^j for $0 \leq j < [F' : F]$. For $\mathfrak{m} \subset A$ not dividing (a) (which misses only finitely many prime factors), localizing at $\mathfrak{m}$ can be computed by first localizing at a (since $a \in A - \mathfrak{m}$), so such an $\mathfrak{m}$ is ramified relative to F'/F if and only if $\mathfrak{m}A[1/a]$ is ramified for $A[1/a] \to A'[1/a]$ (with $A'[1/a]$ the integral closure of $A[1/a]$ in F'). But $A'[1/a]$ is $A[1/a]$-free, so by the settled free-module case we know there are only finitely many maximal ideals of $A[1/a]$ ramified if F'/F! Thus, overall there are only finitely many $\mathfrak{m}$ that can be ramified relative to F'/F (namely, possibly those dividing aA and the ones corresponding to ramified primes for $A[1/a] \to A'[1/a]$). □

We can now finally define the discriminant ideal in general, which will make sense since we have shown that there are only finitely many primes which ramify.

Definition 22.5. Define

$$\operatorname{disc}(A'/A) = \prod_{\mathfrak{m} \text{ that ramify}} \mathfrak{m}^{\varepsilon(\mathfrak{m})}$$

where $\varepsilon(\mathfrak{m}) \geq 1$ is defined by

$$\operatorname{disc}(A'_{\mathfrak{m}}/A_{\mathfrak{m}}) = (\mathfrak{m}A_{\mathfrak{m}})^{\varepsilon(\mathfrak{m})}$$

(which makes sense since $A_{\mathfrak{m}}$ is a dvr, hence a PID, so $A'_{\mathfrak{m}}$ is $A_{\mathfrak{m}}$-free!).

This construction made from "local" data at each maximal ideal is compatible with any localization at a multiplicative set in $A - \{0\}$, as such compatibility amounts to a comparison of prime ideal factorizations in a Dedekind domain that in turn can be analyzed by checking after localizing at each individual maximal ideal. So this does satisfy the desired properties, and in the A-free case it really does coincide with the original globally principal construction for such cases because that construction was seen to be compatible with localization and hence comparing the two definitions for such cases can be checked after localizing at each maximal ideal of A (which makes the contribution from all other maximal ideals disappear).

Here is a numerical example to illustrate that the discriminant ideal in general can be non-principal, so something really does have to be done in the general definition beyond the global determinant method in the A-free case.

Example 22.6. Consider $F = \mathbf{Q}(\sqrt{-17})$ and $A = \mathscr{O}_F = \mathbf{Z}[\sqrt{-17}]$. In this extension of $\mathbf{Q}$, the primes 3 and 89 split. Take $\alpha = -23 + 4\sqrt{-17}$. We have

$$N(\alpha) = 9 \cdot 89,$$

and visibly $\alpha \notin 3\mathscr{O}_F$, so $(\alpha) = \mathfrak{p}_3^2\mathfrak{p}_{89}$ with $\mathfrak{p}_3$ and $\mathfrak{p}_{89}$ among the two primes over 3 and 89 respectively. These prime ideals must contain α, so one can check that $\mathfrak{p}_{89} = (89, -28 + \sqrt{-17})$, which is not principal since no element of $\mathscr{O}_F$ has norm 89 (i.e., $x^2 + 17y^2 = 89$ has no $\mathbf{Z}$-solutions, as can be checked by hand in various ways).

By using some finer knowledge in ramification theory than we have time to develop in this course, for $F' = F(\sqrt{\alpha})$ one can show that the primes of $\mathscr{O}_F$ over 2 are really not ramified in F', and then that $\operatorname{disc}(\mathscr{O}_{F'}/\mathscr{O}_F) = \mathfrak{p}_{89}$.

23. Decomposition and inertia groups

Example 23.1. Here is a nice application of unramified extensions and discriminant ideals.

Suppose we have finite separable extensions

$$\begin{tikzcd}
& F' \arrow[dl, dash] \arrow[dr, dash] & \\
F_1 \arrow[dr, dash] & & F_2 \arrow[dl, dash] \\
& F &
\end{tikzcd} \tag{23.1}$$

with corresponding Dedekind domains and integral closures

$$\begin{tikzcd}
& A' \arrow[dl, dash] \arrow[dr, dash] & \\
A_1 \arrow[dr, dash] & & A_2 \arrow[dl, dash] \\
& A &
\end{tikzcd} \tag{23.2}$$

Assume $F' = F_1F_2$ and that the ideals $\operatorname{disc}(A_1/A)$, and $\operatorname{disc}(A_2/A)$ in A have no common factor (equivalently, every $\mathfrak{m} \in \operatorname{Max}(A)$ is unramified in at least one of the extensions F_i/F). Note that we are *not* assuming $[F' : F] = [F_1 : F][F_2 : F]$.

Then it turns out that

$$A' = A_1A_2 := \{ \sum_{\text{finite}} a_{1i}a_{2i} \}$$

For a proof of this, see the handout "Application of Ideal Discriminant to Unramifiedness". (The proof is somewhat complicated because we don't have enough commutative algebra technology in this course. I took a short proof based on much more advanced notions and reformulated it to work based on what we have developed.)

The lesson is that unramifiedness makes the integral closure behave more like a field extension. In this sense, unramified primes are "good" (and more experience will only confirm this further).

Let's now consider the case that F'/F is Galois with Galois group G, as in the diagram

(23.3)

$$\begin{tikzcd}
F' & A' \arrow[l] \arrow[d, dash] \\
F \arrow[u, dash] \arrow[r, leftrightarrow] & A
\end{tikzcd}$$

We shall investigate how the Galois property imposes relations among the e_i's and among the f_i's for a chosen maximal ideal $\mathfrak{m} \subset A$. Consider the factorizations

$$\mathfrak{m}A' = \prod (\mathfrak{m}'_i)^{e_i}.$$

For the residue field degrees

$$f_i = [A'/\mathfrak{m}'_i : A/\mathfrak{m}]$$

we know

$$\sum_i e_i f_i = n = \#G.$$

For each $\sigma \in G$, we have $\sigma(A') \subset A'$ and $\sigma^{-1}(A') \subset A'$, and the latter equivalently says $A' \subset \sigma(A')$ (by applying σ to both sides), so $\sigma(A') = A$. In other words, the F-automorphism σ of the field F' restricts to an A-automorphism of the ring A'.

By the definitions we see that $\sigma(\mathfrak{m}A') = \sigma(\mathfrak{m})\sigma(A')$, and this is equal to $\mathfrak{m}A'$ since σ has no effect on $\mathfrak{m} \subset A$ and we just showed $\sigma(A') = A'$. Hence, we have the diagram

$$\begin{tikzcd} A' \arrow[r, "\sigma"] & A' \\ \mathfrak{m}A' \arrow[u] \arrow[r, "\sigma"] & \mathfrak{m}A' \arrow[u] \end{tikzcd} \tag{23.4}$$

with bijective horizontal maps But $\sigma : A' \to A'$ is a ring isomorphism, so it is compatible with the formation of products of ideals and hence

$$\sigma(\prod (\mathfrak{m}_i')^{e_i}) = \prod \sigma(\mathfrak{m}_i')^{e_i}.$$

Since ring isomorphisms carry maximal ideals to maximal ideals, it follows that the $\sigma(\mathfrak{m}_i')$'s are pairwise distinct maximal ideals, so

$$\prod \sigma(\mathfrak{m}_i')^{e_i}$$

is also an expression of $\mathfrak{m}A'$ as a product of powers of a collection of pairwise distinct maximal ideals.

Thus, by *uniqueness* of prime factorization, it follows that $\{\sigma(\mathfrak{m}_i')\}_i$ is a permutation of the collection of prime ideal factors of $\mathfrak{m}A'$. In other words, G permutes the finite set of prime ideals of A' over $\mathfrak{m}$. In fact, we can say more:

Theorem 23.2. *In the above setup, the G-action on $\{\mathfrak{m}_i'\}$ is transitive; i.e., for all $i \neq j$ there exists $\sigma \in G$ such that $\mathfrak{m}_j' = \sigma(\mathfrak{m}_i')$.*

We refer to [Samuel, §6.2, Proposition 1] for the proof, which amounts to some cleverness with commutative algebra. Informally, this result says that the primes of A' over $\mathfrak{m}$ are all "created equal": the symmetry arising from G makes it impossible to intrinsically distinguish any one from any other. For our purposes, what is really important is the following consequence:

Corollary 23.3. *The integers e_i and f_i associated to $\mathfrak{m}_i'$ over $\mathfrak{m}$ are independent of $1 \leq i \leq g$ and furthermore, letting e and f denote these respective common values, we have $efg = n$.*

Proof. Since $\mathfrak{m}A' = \prod_i \sigma(\mathfrak{m}_i')^{e_i}$, the exponent $e(\sigma(\mathfrak{m}_i')|\mathfrak{m})$ for $\sigma(\mathfrak{m}_i')$ as a prime factor of $\mathfrak{m}A'$ coincides with the exponent $e_i = e(\mathfrak{m}_i'|\mathfrak{m})$ for $\mathfrak{m}_i'$ as a prime factor of $\mathfrak{m}A'$. In other words, $e(\sigma(\mathfrak{m}')|\mathfrak{m}) = e(\mathfrak{m}'|\mathfrak{m})$ for all $\mathfrak{m}'|\mathfrak{m}A'$ and all $\sigma \in G$. By transitivity, it follows that $e(\mathfrak{m}'|\mathfrak{m})$ is the same value for all $\mathfrak{m}'$, which is to say that all e_i are equal to a common value $e \in \mathbf{Z}^+$ (depending only on $\mathfrak{m}$).

Also, we have a commutative diagram of maps of fields

$$\begin{tikzcd}
A'/\mathfrak{m}' \arrow[r, "\sigma"] & A'/\sigma(\mathfrak{m}') \\
A/\mathfrak{m} \arrow[r] \arrow[u] & A/\mathfrak{m} \arrow[u]
\end{tikzcd} \tag{23.5}$$

for $\mathfrak{m}' \mid \mathfrak{m}A'$, where the bottom map is the identity and the top map is the isomorphism induced by $\sigma : A' \simeq A'$. Thus, the vertical maps have the same field degree, which is to say $f(\sigma(\mathfrak{m}')|\mathfrak{m}) = f(\mathfrak{m}'|\mathfrak{m})$ for all $\mathfrak{m}'$ and all $\sigma \in G$. By transitivity, these residual degrees are all equal to each other, which is to say all f_i are equal to a common value $f \in \mathbf{Z}^+$ (depending only on $\mathfrak{m}$).

From the equality $\sum_{i=1}^{g} e_i f_i = n$ we conclude that $efg = n$. □

Definition 23.4. The *decomposition group* at $\mathfrak{m}'$ over $\mathfrak{m}$ is

$$D(\mathfrak{m}'|\mathfrak{m}) = \{\sigma \in G \,|\, \sigma(\mathfrak{m}') = \mathfrak{m}'\}$$

viewed as a subgroup of $G = \mathrm{Gal}(F'/F)$.

Observe that we have an action of $D(\mathfrak{m}'|\mathfrak{m})$ on the field $A'/\mathfrak{m}'$ over $A/\mathfrak{m}$, since if $\sigma \in G$ preserves $\mathfrak{m}'$ then the automorphism that σ defines on A' induces an automorphism of the residue field $A'/\mathfrak{m}'$ visibly having no effect on the subfield $A/\mathfrak{m}$.

Thus, we have a group homomorphism

$$D(\mathfrak{m}'|\mathfrak{m}) \to \mathrm{Aut}_{A/\mathfrak{m}}(A'/\mathfrak{m}')$$

carrying σ to the automorphism $\overline{\sigma} : a' \bmod \mathfrak{m}' \mapsto \sigma(a') \bmod \mathfrak{m}'$ of $A'/\mathfrak{m}'$ over $A/\mathfrak{m}$.

By definition, the *inertia group* $I(\mathfrak{m}'|\mathfrak{m})$ is the kernel of this homomorphism, or equivalently

$$I(\mathfrak{m}'|\mathfrak{m}) = \{\sigma \in G \,|\, \sigma(a') \equiv a' \bmod \mathfrak{m}' \text{ for all } a' \in A'\}.$$

Remark 23.5. On Homework 8, there is a plethora of explicit examples of computing inertia groups and decomposition groups.

Lemma 23.6. *For each $\sigma \in G$, we have*

$$D(\sigma\mathfrak{m}' \mid \mathfrak{m}) = \sigma D(\mathfrak{m}'|\mathfrak{m})\sigma^{-1}, \;\; I(\sigma\mathfrak{m}' \mid \mathfrak{m}) = \sigma I(\mathfrak{m}'|\mathfrak{m})\sigma^{-1}.$$

In particular, if G is abelian then $D(\mathfrak{m}'|\mathfrak{m})$ and $I(\mathfrak{m}'|\mathfrak{m})$ are independent of $\mathfrak{m}'$ over $\mathfrak{m}$.

Proof. The proof is definition-chasing, first for D and then for $I \subset D$. (In the case of D, this just expresses the general behavior of stabilizers at points in the same orbit under a group action on a set, such as the G-action on the set of primes of A' over $\mathfrak{m}$.) Work through the details yourself! □

The key fact is:

Theorem 23.7 (Frobenius)**.** *The natural map $D(\mathfrak{m}'|\mathfrak{m}) \to \mathrm{Aut}_{A/\mathfrak{m}}(A'/\mathfrak{m}')$ is surjective. In particular, this is an isomorphism if $I(\mathfrak{m}'|\mathfrak{m}) = 1$.*

We'll see soon that $I(\mathfrak{m}'|\mathfrak{m}) = 1$ in unramified settings.

Proof. All modern texts either prove this using decomposition fields (an obsolete notion) or completions (in effect, the field of p-adic numbers, and variants thereof); the proof in our course text is of the first type (though it doesn't mention the terminology "decomposition field" for the field constructed in the middle of its proof).

However, it turns out that Frobenius' original proof, largely forgotten, is much simpler than these! His proof involves studying the polynomial

$$\prod_{\sigma \in G} (Y - \sum \sigma(a_i')X_i) \in A'[Y, X_1, X_2, \dots]$$

for $\{a_i'\}$ a finite spanning set of A' over A (i.e., $\sum Aa_i' = A'$). The G-action on A' permutes the factors of this polynomial, so the polynomial itself is an element of $(A')^G[Y, X_1, X_2, \dots] = A[Y, X_1, X_2, \dots]$. By studying its mod-$\mathfrak{m}$ reduction, Frobenius showed that any $A/\mathfrak{m}$-automorphism of $A'/\mathfrak{m}'$ actually comes from an A-automorphism of A' preserving $\mathfrak{m}'$ (and hence an F-automorphism of F' belonging to $D(\mathfrak{m}'|\mathfrak{m})$). See the handout "Frobenius' Surjectivity Theorem" for full details. □

Consequences of Theorem 23.7. Let's now see some useful consequences. Recall in the setup of Theorem 23.7, we have an action of G on the set

$$\{\mathfrak{m}_1', \dots, \mathfrak{m}_g'\}$$

of size g, and this action is transitive. Thus, the stabilizer $D(\mathfrak{m}'|\mathfrak{m})$ at a point $\mathfrak{m}'$ in this set has size $\#G/g = n/g = (efg)/g = ef$.

We also have an isomorphism

$$D/I \simeq \mathrm{Aut}_{A/\mathfrak{m}}(A'/\mathfrak{m}'),$$

and this automorphism group has a known size if $A/\mathfrak{m} \to A'/\mathfrak{m}'$ is *Galois*: it would be equal to the residual degree f, so then we could also conclude that $\#I = \#D/f = (ef)/f = e$ and hence $I = 1$ in the unramified case.

But is the residue field extension Galois? It certainly is for finite residue fields, as in the case of number fields. This actually holds much more generally, for purely algebraic reasons, as we now explain. First we recall:

Definition 23.8. A finite-degree extension of fields E/L is *normal* if for any element $a \in E$ the minimal polynomial of a over L splits completely in E.

(It is equivalent to say E/L is the splitting field of a polynomial in $L[x]$, as we learned in our study of fields prior to this course.)

Proposition 23.9. *The extension of residue fields*

$$A/\mathfrak{m} \to A'/\mathfrak{m}'$$

is always normal. In particular, it is Galois if it is separable.

Proof. We have to check that for $a' \in A'$, the minimal polynomial h of $\overline{a}' := a' \bmod \mathfrak{m}' \in k'$ over k splits completely over k'. It is enough to find some $h_1 \in k[x]$ with $\overline{a}'$ as a root such that h_1 splits completely over k', as then the same would hold for h since $h|h_1$ over k.

Consider $H_1 \in F[x]$ the minimal polynomial of a' over F, so $H_1 \in A[x]$ has reduction h_1 that is vanishing at $\overline{a}'$. However, F'/F being Galois implies that H_1 splits completely over F' with all roots in A'. Therefore, H_1 splits in $A'[x]$, so its reduction h_1 completely over $k' = A'/\mathfrak{m}'$. □

Now, assume $A/\mathfrak{m}$ is perfect. For example, this holds for all finite fields, so it will apply whenever A is the ring of integers of a number field (or localization thereof!). In such cases we have seen that $\#I = e$, so if $\mathfrak{m}$ is *unramified* in F'/F then

$$D(\mathfrak{m}'|\mathfrak{m}) \simeq \mathrm{Gal}((A'/\mathfrak{m}')/(A/\mathfrak{m})).$$

Returning to the setting of number fields suppose we have an extension K'/K with Galois group G

(23.6)
$$\begin{tikzcd} K' \arrow[d, dash] & \mathfrak{p}' \arrow[d, dash] \\ K & \mathfrak{p} \end{tikzcd}$$

Assume $\mathfrak{p}$ is *unramified* in K'/K, and pick $\mathfrak{p}'$ over $\mathfrak{p}$. Recall $N\mathfrak{p} = \#\mathscr{O}_K/\mathfrak{p}$, so the extension of finite residue fields $\mathscr{O}_K/\mathfrak{p} \to \mathscr{O}_{K'}/\mathfrak{p}'$ has Galois group with generator

$$x' \mapsto x'^{\mathrm{N}\mathfrak{p}}$$

called the *Frobenius* generator.

Thus, there exists a unique element $\mathrm{Fr}(\mathfrak{p}'|\mathfrak{p}) \in D(\mathfrak{p}'|\mathfrak{p})$ inducing the Frobenius generator of the Galois group for the extension of finite residue fields. This is remarkable: a field automorphism in characteristic p has a canonical lift to an automorphism in characteristic 0. Concretely, $\mathrm{Fr}(\mathfrak{p}'|\mathfrak{p})$ is the unique element $s \in D(\mathfrak{p}'|\mathfrak{p})$ such that $s(a') \equiv a'^{\mathrm{N}\mathfrak{p}} \bmod \mathfrak{p}'$ for all $a' \in \mathscr{O}_{K'}$. (Note that this congruence even characterizes $\mathrm{Fr}(\mathfrak{p}'|\mathfrak{p})$ as an element of G since if $s(a') \equiv a'^{\mathrm{N}\mathfrak{p}} \bmod \mathfrak{p}'$ for all $a' \in \mathscr{O}_{K'}$ then $a' \equiv 0 \bmod \mathfrak{p}' \Rightarrow s(a') \equiv 0 \bmod \mathfrak{p}'$, so s carries $\mathfrak{p}'$ into itself, which is to say $s \in D(\mathfrak{p}'|\mathfrak{p})$.)

For $\sigma \in G$, one checks as an instructive exercise that

$$\sigma\mathrm{Fr}(\mathfrak{p}'|\mathfrak{p})\sigma^{-1} = \mathrm{Fr}(\sigma(\mathfrak{p}')|\mathfrak{p}).$$

Hence, given $\mathfrak{p}$, we obtain a *conjugacy class* of "Frobenius elements" in G (by varying the choice of $\mathfrak{p}'$ over $\mathfrak{p}$). For G abelian, we thereby get a well-defined *element* $\mathrm{Frob}_{\mathfrak{p}} \in G$. Next time, we'll calculate what the Frobenius element is in the case of the abelian extension $\mathbf{Q}(\zeta_n)/\mathbf{Q}$ at unramified primes p.

These Frobenius conjugacy classes are the key to modern algebraic number theory.

24. Class groups and units

Before moving on to class groups and units, we wrap up our discussion of Frobenius elements (which we started last time) by discussing how to use Frobenius elements to give a conceptual proof of quadratic reciprocity that one can actually remember (unlike various "elementary" proofs which are too long or complicated). This builds on the entire infrastructure we have set up so far, and is quite satisfying.

Consider a Galois extension of number fields K'/K with Galois group G. Let $\mathfrak{p}$ be a prime of K unramified in K', and choose a prime $\mathfrak{p}'$ of K' lying over $\mathfrak{p}$. Let k' be the residue field of $\mathfrak{p}'$ and k be the residue field of $\mathfrak{p}$:

(24.1)
$$\begin{tikzcd} K' \arrow[d, dash] & \mathfrak{p}' \arrow[d, dash] & k' \arrow[d, dash] \\ K & \mathfrak{p} & k \end{tikzcd}$$

Recall that the generator

$$\mathrm{Fr}(\mathfrak{p}'|\mathfrak{p}) \in D(\mathfrak{p}'|\mathfrak{p}) \subset G$$

arises as follows: by our unramifiedness assumption the natural surjective map

$$D(\mathfrak{p}'|\mathfrak{p}) \to \mathrm{Gal}(k'/k)$$

is an isomorphism, and $\mathrm{Fr}(\mathfrak{p}'|\mathfrak{p})$ corresponds to $(x \mapsto x^{\#k}) \in \mathrm{Gal}(k'/k)$. This element $\mathrm{Fr}(\mathfrak{p}'|\mathfrak{p})$ is the unique element $\sigma \in G$ satisfying

$$\sigma(a') \equiv (a')^{N\mathfrak{p}} \bmod \mathfrak{p}'$$

for all $a' \in \mathscr{O}_{K'}$. (Recall that this congruence condition forces preservation of $\mathfrak{p}'$, so it incorporates the requirement of membership in $D(\mathfrak{p}'|\mathfrak{p})$.)

Example 24.1. Consider the special case $K = \mathbf{Q}$ and $K' = \mathbf{Q}(\zeta_n)$, so $G = \mathrm{Gal}(\mathbf{Q}(\zeta_n)/\mathbf{Q}) \simeq (\mathbf{Z}/n\mathbf{Z})^\times$, with $a \in (\mathbf{Z}/n\mathbf{Z})^\times$ corresponding to the element $\sigma_a \in G$ uniquely determined by the condition $\sigma_a(\zeta) = \zeta^a$ for all nth roots of unity $\zeta \in K'$ (it is the same to require this only for ζ_n, as then passing to its powers gives the same for all ζ).

Choose a positive prime $p \nmid n$, so p is unramified in $\mathbf{Q}(\zeta_n)$. There exists a unique $\mathrm{Frob}_p \in G$ preserving each $\mathfrak{p}'$ over $p\mathbf{Z}$ and inducing the p-power map on each residue field $\mathbf{Z}[\zeta_n]/\mathfrak{p}'$. What is Frob_p viewed in $(\mathbf{Z}/n\mathbf{Z})^\times$? (If $n = 2m$ for odd m, so $\mathbf{Q}(\zeta_n) = \mathbf{Q}(\zeta_m)$, note that $(\mathbf{Z}/n\mathbf{Z})^\times = (\mathbf{Z}/m\mathbf{Z})^\times$.)

We claim that $\mathrm{Frob}_p = p \bmod n$; i.e., $\mathrm{Frob}_p = \sigma_p$. To prove this, consider the ring-theoretic decomposition from the Chinese Remainder Theorem:

$$\mathbf{Z}[\zeta_n]/(p) = \prod_{\mathfrak{p}'|(p)} \mathbf{Z}[\zeta_n]/\mathfrak{p}'.$$

The automorphism σ_p preserves (p) and so induces an automorphism $\overline{\sigma}_p$ of the quotient ring $\mathbf{Z}[\zeta_n]/(p)$. Also by design $\sigma_p : \zeta_n \mapsto \zeta_n^p$. We claim that $\overline{\sigma}_p$ agrees with the p-power endomorphism of the $\mathbf{F}_p$-algebra $\mathbf{Z}[\zeta_n]/(p)$ (recall that for any $\mathbf{F}_p$-algebra A, $a \mapsto a^p$ is a ring homomorphism from A to itself). Since $\mathbf{Z}[\zeta_n]/(p)$ is generated as a ring over $\mathbf{F}_p$ by the image of ζ_n, to compare two endomorphisms of this ring it suffices to compare them on the single element $\zeta_n \bmod p$, on which we've noted that $\overline{\sigma}_p$ and the p-power endomorphism coincide.

We have shown that the effect of $\overline{\sigma}_p$ on $\mathbf{Z}[\zeta_n]/(p)$ coincides with the p-power endomorphism that in turn is clearly given by the p-power endomorphism of each factor field $\mathbf{Z}[\zeta_n]/\mathfrak{p}'$ of $\mathbf{Z}[\zeta_n]/(p)$! In particular, $\overline{\sigma}_p$ preserves the kernel of the projection to each factor field, so σ_p preserves each $\mathfrak{p}'$ (i.e., it lies in each $D(\mathfrak{p}'|p\mathbf{Z})$) and on the quotient modulo each $\mathfrak{p}'$ induces the p-power map. That exactly says that σ_p satisfies the properties uniquely characterizing Frob_p, so indeed $\sigma_p = \mathrm{Frob}_p$.

Remark 24.2. The above yields a connection between group-theoretic information in a Galois group and primes in arithmetic progressions: the statement that every arithmetic progression $a + n\mathbf{Z}$ with $\gcd(a, n) = 1$ contains a prime is the statement that each element of $\mathrm{Gal}(\mathbf{Q}(\zeta_n)/\mathbf{Q})$ is a Frobenius element at an unramified prime (ignoring the case $n = 2m$ for odd m at the unramified prime $2|n$). Needless to say, it is by no means obvious that an element in $\mathrm{Gal}(K'/K)$ should be a Frobenius element at an unramified prime, let alone at infinitely many unramified primes.

This perspective is the starting point for a vast "non-abelian" generalization of Dirichlet's theorem on primes in arithmetic progressions, lying beyond the level of this course.

Example 24.3. Consider a quadratic extension $K/\mathbf{Q}$, say with discriminant d. Consider an *odd* prime p which is unramified in K, so p doesn't divide d and p is split if and only if d is a square modulo p (as p is *odd*). We have a unique isomorphism of groups of order 2:

$$\mathrm{Gal}(K/\mathbf{Q}) \simeq \langle -1 \rangle$$

(there is only one group of order 2, containing the identity element and the other element). Under this identification, Frob_p maps to 1 or -1 corresponding to Frob_p being trivial or non-trivial respectively.

For the invariants e, f, g attached to p we have $efg = 2$, and because p is unramified we know $e = 1$, so $fg = 2$. But recall that at unramified primes the order of the Frobenius element is precisely the residue field extension degree f, so $\mathrm{Frob}_p \mapsto 1$ precisely when $f = 1$ or equivalently $g = 2$, which is to say that the unramified prime p is split in K. Hence, $\mathrm{Frob}_p \mapsto -1$ precisely when p is inert (i.e., the only other unramified situation for quadratic fields). Summarizing, we have

$$\begin{aligned} \mathrm{Frob}_p = 1 &\iff p \text{ is split} \iff \left(\frac{d}{p}\right) = 1 \\ \mathrm{Frob}_p = -1 &\iff p \text{ is inert} \iff \left(\frac{d}{p}\right) = -1. \end{aligned}$$

We conclude that via the unique isomorphism $\mathrm{Gal}(K/\mathbf{Q}) \simeq \langle -1 \rangle$,

$$\mathrm{Frob}_p = \left(\frac{d}{p}\right).$$

To apply the two preceding examples, let q be an odd positive prime not equal to the odd prime p, and consider the tower

$$\begin{tikzcd} \mathbf{Q}(\zeta_q) \arrow[d, dash] \\ K = \mathbf{Q}\left(\sqrt{(\tfrac{-1}{q})q}\right) \arrow[d, dash] \\ \mathbf{Q} \end{tikzcd} \tag{24.2}$$

Comparing Frob_p for $\mathbf{Q}(\zeta_q)/\mathbf{Q}$ and for $K/\mathbf{Q}$ (extensions in which p is unramified!), general "naturality" properties of Frobenius elements in Exercise 1 of HW8 give that $\mathrm{Gal}(\mathbf{Q}(\zeta_q)/\mathbf{Q}) \twoheadrightarrow \mathrm{Gal}(K/\mathbf{Q})$ carries $\mathrm{Frob}_p^{\mathbf{Q}(\zeta_q)/\mathbf{Q}}$ to

$\mathrm{Frob}_p^{K/\mathbf{Q}}$, so via the identification of the two maps

$$\begin{tikzcd} \mathrm{Gal}(\mathbf{Q}(\zeta_q)/\mathbf{Q}) \arrow[r] & \mathrm{Gal}(K/\mathbf{Q}) \\ (\mathbf{Z}/q\mathbf{Z})^\times \arrow[r] & \langle -1 \rangle \end{tikzcd} \tag{24.3}$$

we have that $p \bmod q \mapsto ((-1|q)q|p)$. But the cyclic group $(\mathbf{Z}/q\mathbf{Z})^\times$ has only *one* order-2 quotient, namely by the index-2 subgroup of squares modulo q, so it follows that $((-1|q)q|p) = (p|q)$. This is exactly quadratic reciprocity! (See [Samuel, Ch. VI, §5] for full details, including a related argument using $\mathbf{Q}(\zeta_8)$ to deduce the supplementary law for $(2|p)$.)

Remark 24.4. For $\mathfrak{p}$ unramified in a Galois extension K'/K, $\mathrm{Fr}(\mathfrak{p}'|\mathfrak{p})$ has order f, so since $\mathfrak{p}$ is totally split if and only if ($e = 1$ and) $f = 1$, it is totally split if and only if $\mathrm{Fr}(\mathfrak{p}'|\mathfrak{p}) = 1$ for some (equivalently, all) $\mathfrak{p}'$ over $\mathfrak{p}$.

For the rest of the course, we'll discuss the proofs and applications of two fundamental finiteness theorems. To formulate the first one, we need some notation. For a number field K, the *class group* $\mathrm{Cl}(K)$ denotes the group $\mathrm{Cl}(\mathscr{O}_K)$ studied in Homeworks 6 and 7: this is the group of fractional ideals modulo principal ideals. We call $h_K := \#\mathrm{Cl}(K)$ the *class number* of K when it is finite. Note that $h_K = 1$ if and only if $\mathscr{O}_K$ is a PID. Since a Dedekind domain is a PID if and only if it is a UFD, this is also equivalent to $\mathscr{O}_K$ being a UFD. The first big finiteness theorem is:

Theorem 24.5. *Any number field has finite class number.*

To state the second big finiteness theorem, the Dirichlet Unit Theorem, we need some more notation. For K a number field, we let r_1 denote the number of embeddings $K \to \mathbf{R}$ and let $2r_2$ denote the number of embeddings $K \to \mathbf{C}$ which do not factor through $\mathbf{R}$ (these come in conjugate pairs, since composing with complex conjugation changes any *non-real* embedding). We emphasize that these are counts of *maps*, and not of the image subfields under such maps. For example, if $K/\mathbf{Q}$ is Galois then all such embeddings have the *same* image but there are $[K : \mathbf{Q}]$ distinct embeddings (since $\mathbf{C}$ is algebraically closed).

In general, since $\mathbf{C}$ is algebraically closed (so the minimal polynomial over $\mathbf{Q}$ of a primitive element of K splits completely over $\mathbf{C}$), we have

$$r_1 + 2r_2 = \#\{K \to \mathbf{C}\} = [K : \mathbf{Q}] =: n.$$

Theorem 24.6 (Dirichlet Unit Theorem)**.** *For K a number field, $\mathscr{O}_K^\times$ is a finitely generated group of rank $r_1 + r_2 - 1$. Equivalently,*

$$\mathscr{O}_K^\times \simeq \mu \times \mathbf{Z}^{r_1+r_2-1}$$

for μ the (finite!) group of roots of unity in K.

Remark 24.7. In Exercise 4 in Homework 7, for a finite set

$$S = \{\mathfrak{p}_1, \dots, \mathfrak{p}_m\}$$

of primes of K (notation not to be confused with a multiplicative set for localization) we defined the ring $\mathscr{O}_{K,S}$ of S-integers of K (whose set of maximal ideals is that of $\mathscr{O}_K$ with S removed) and saw that

$$\mathrm{Cl}(\mathscr{O}_{K,S}) = \mathrm{Cl}(\mathscr{O}_K)/\langle[\mathfrak{p}_i]\rangle_i.$$

Thus, $\mathrm{Cl}(\mathscr{O}_{K,S})$ is trivial if we choose the $\mathfrak{p}_i$ to generate $\mathrm{Cl}(\mathscr{O}_K)$, as we can always do since each of the finitely many elements of $\mathrm{Cl}(\mathscr{O}_K)$ is represented by a nonzero ideal of $\mathscr{O}_K$ that in turn is a product of finitely many primes. (It is true that every element of $\mathrm{Cl}(\mathscr{O}_K)$ is the class of a prime ideal on the nose, but that is beyond the level of the course.) In this case, $\mathscr{O}_{K,S}$ is a UFD.

It was seen in Homework 7 that for any such S there exists $a \in \mathscr{O}_K - \{0\}$ so that $\mathscr{O}_{K,S} = \mathscr{O}_K[1/a]$, so $\mathscr{O}_K[1/a]$ is a UFD for some a. This latter property is very specific to the Dedekind domains $\mathscr{O}_K$ that arise from number fields. For instance, in contrast the Dedekind domain

$$A := \mathbf{C}[x,y]/(y^2 - (x^3 - x))$$

has uncountable class group and $A[1/a]$ is not a UFD for any $a \in A - \{0\}$; the proof involves the theory of elliptic curves. Similarly, for the Dedekind domain $B = \mathbf{Q}[x,y]/(y^2 - (x^3 - 36x))$, deeper input from the theory of elliptic curves shows $\mathrm{Cl}(B) = (\mathbf{Z}/2\mathbf{Z})^2 \oplus \mathbf{Z}$ (countably infinite!) and $B[1/b]$ is not a UFD for any $b \in B - \{0\}$.

The proof of finiteness of the class group will be effective, in the sense that it gives explicit bounds on where to search to find generators. However, the proof of Dirichlet's Unit Theorem uses a compactness argument and is not effective (in a suitable sense). To illustrate the more subtle nature of finding units, let's first see what the Unit Theorem is saying for some small values of the rank $r_1 + r_2 - 1$. Since $r_1 + 2r_2 = n$, it follows that

$$\begin{aligned} r_1 + r_2 - 1 = 0 \iff & K \text{ is either imaginary quadratic } (r_2 = 1, r_1 = 0) \\ & \text{or } K = \mathbf{Q} \ (r_2 = 0, r_1 = 1). \end{aligned}$$

Likewise we have $r_1 + r_2 - 1 = 1$ if and only if either

$r_1 = 2, r_2 = 0$ (i.e., real quadratic),
$r_1 = 0, r_2 = 1$ (i.e., quartic extensions with no real embedding),
$r_1 = 1, r_2 = 1$ (i.e., cubic extensions with one real embedding).

Let's consider the real quadratic case. For real quadratic K the only roots of unity are ± 1 (as $\mathbf{R}$ doesn't contain higher-order roots of unity), so

$$\mathscr{O}_K^\times = \langle -1 \rangle \times \varepsilon_K^{\mathbf{Z}},$$

where ε_K is a *fundamental unit*, meaning a generator of the unit group modulo the torsion subgroup $\{\pm 1\}$. There are 4 fundamental units: $\pm\varepsilon_K, \pm 1/\varepsilon_K$, so if we *fix* an embedding $j : K \hookrightarrow \mathbf{R}$ then there is exactly one fundamental unit in each of the 4 connected components of $\mathbf{R} - \{0, \pm 1\}$ (transitively permuted via iterating the operations $x \mapsto 1/x$ and $x \mapsto -x$). The unique one in $(1, \infty)$ is often called "the" fundamental unit of K (relative to j).

Example 24.8. The variation of the fundamental unit is quite erratic as we vary the real quadratic field K, as illustrated with the following 3 cases:

K	$\varepsilon_K > 1$
$\mathbf{Q}(\sqrt{381})$	$1015 + 52\sqrt{381}$
$\mathbf{Q}(\sqrt{382})$	$164998439999 + 8442054600\sqrt{382}$
$\mathbf{Q}(\sqrt{383})$	$18768 + 959\sqrt{383}$

TABLE 1. Variation of ε_K for some real quadratic fields

Example 24.9. For a higher-rank illustration of the Unit Theorem, consider $K = \mathbf{Q}(\zeta_p)$ for $p > 3$. We have $r_1 = 0$, so $2r_2 = p - 1$ and hence $r_1 + r_2 - 1 = (p-3)/2 > 0$. Thus,

$$\mathbf{Z}[\zeta_p]^\times = \mu_{2p} \times \langle \varepsilon_1, \ldots, \varepsilon_{(p-3)/2} \rangle$$

for some multiplicatively independent $\varepsilon_1, \ldots, \varepsilon_{(p-3)/2}$. In general, it is quite difficult to find such ε_j's, but one can give an explicit multiplicatively independent subgroup of finite index (called the *cyclotomic units*) as follows.

For $2 \leq j \leq p-2$ (so $p - 3$ values of j) we define

$$u_j = (\zeta_p^{1/2})^{1-j} \cdot \frac{1 - \zeta_p^j}{1 - \zeta_p}$$

(note that $\zeta_p^{1/2}$ makes sense unambiguously since one can uniquely divide by 2 in a finite abelian group of odd order). The root of unity multiplier in the definition of u_j is rigged to ensure that $u_{p-j} = -u_j$ (as one may check without difficulty). These multiplicative dependence relations (i.e., u_{p-j}/u_j is 2-torsion for all j) imply that the group of u_j's has rank at most $(p-3)/2$ (generated by u_j's for $2 \leq j \leq (p-1)/2$).

For explicit small p one can use techniques from the proof of the Unit Theorem to show that the rank of the group of cyclotomic units is exactly $(p-3)/2$ (so this group has finite index in $\mathbf{Z}[\zeta_p]^\times$), but to prove the same for all p requires a theoretical argument ultimately relying on analytic input via L-functions. It turns out that the finite index of the group of cyclotomic units inside $\mathbf{Z}[\zeta_p]^\times/\mu_{2p}$ is exactly the class number $h_{\mathbf{Q}(\zeta_p)^+}$ of the maximal real subfield $\mathbf{Q}(\zeta_p)^+ = \mathbf{Q}(\zeta_p + \zeta_p^{-1})$ of $\mathbf{Q}(\zeta_p)$, a fact that was first discovered in the course of Kummer's work on Fermat's Last Theorem.

In a handout "Diophantine Applications of Class Groups" we give some applications of class group computations to two types of arithmetic questions. Let us survey the highlights of one of those here:

Example 24.10. Consider $y^2 = x^3 - 51$. It can be shown that this has solutions in $\mathbf{Z}/m\mathbf{Z}$ for all $m > 1$ and that it has infinitely many $\mathbf{Q}$-solutions (one of which is $(1375/9, 50986/27)$). However, there are no solutions over $\mathbf{Z}$! This is interesting because it cannot be proven by congruential methods (due to the existence of solutions modulo m for all $m > 1$).

As in the first lecture of the course, one begins by rewriting it as

$$x^3 = y^2 + 51 = (y + \sqrt{-51})(y - \sqrt{-51}).$$

Now a new difficulty emerges: one has $h_{\mathbf{Q}(\sqrt{-51})} > 1$, so the ring of integers $\mathbf{Z}[\sqrt{-51}]$ is not a UFD. Thus, the analogue of the step showing that certain quantities are cubes seems doomed.

But a miracle happens: it turns out that $3 \nmid h_{\mathbf{Q}(\sqrt{-51})}$, so there is no non-trivial 3-torsion in the class group and hence any nonzero (potentially non-principal) ideal of $\mathbf{Z}[\sqrt{-51}]$ whose cube is principal must itself be principal. This allows one to push through a version of the method from the first day of class "as if" $\mathbf{Z}[\sqrt{-51}]$ were a UFD! The key in the end is that a *non-trivial* class number is nonetheless coprime to 3, a property that cannot be seen via congruential considerations in the ring of integers.

(In Kummer's work on Fermat's Last Theorem for exponent p, he made decisive progress in the cases for which $p \nmid h_{\mathbf{Q}(\zeta_p)}$, called "regular" primes. In these cases, to battle with unit issues it was crucial that Kummer had the explicit group of cyclotomic units whose index is $h_{\mathbf{Q}(\zeta_p)^+}$. The subtle interplay of units and class groups is a pervasive issue in number theory.)

25. Computing some class groups

We wish to state a more precise version of the finiteness of class numbers (Theorem 24.5). This requires some notation:

Definition 25.1. Let K be a number field and $n := [K : \mathbf{Q}]$. Let r_1 be the number of real embeddings, and $2r_2$ be the number of non-real complex embeddings, so $r_1 + 2r_2 = n$. Define the *Minkowski constant* for K to be

$$\lambda_K := \frac{n!}{n^n}(4/\pi)^{r_2}\sqrt{|\operatorname{disc} K|}$$

Here is the precise version of finiteness:

Theorem 25.2. *Let K be a number field. Each element of* $\mathrm{Cl}(K)$ *is the ideal class* $[\mathfrak{a}]$ *of a nonzero ideal* $\mathfrak{a} \subset \mathscr{O}_K$ *such that*

$$N\mathfrak{a} \leq \lambda_K.$$

Note that there are only finitely many $\mathfrak{a} = \prod_i \mathfrak{p}_i^{r_i}$ with $(N\mathfrak{p}_i)^{r_i} \leq \lambda_K$ for all i since this latter inequality bounds the possibilities for the rational prime $p_i \mid N\mathfrak{p}_i$ over which $\mathfrak{p}_i$ lies (so there are only finitely many possibilities for each $\mathfrak{p}_i$), and also bounds the possibilities for r_i since $p_i^{r_i} \leq \lambda_K$. This yields:

Corollary 25.3. *For K a number field,* $\mathrm{Cl}(K)$ *is generated by the classes* $[\mathfrak{p}]$ *for the finitely many primes* $\mathfrak{p}$ *over the finitely many primes* $p \in \mathbf{Z}^+$ *such that* $p \leq \lambda_K$.

Let's now see some consequences of the preceding more precise versions of the finiteness result.

Example 25.4. If $\lambda_K < 2$ then $h_K = 1$ since there are no $\mathfrak{p}$'s satisfying $N\mathfrak{p} \leq \lambda_K$ and hence the only possibility for $\mathfrak{a}$ in the preceding discussion is the unit ideal (so $\mathrm{Cl}(K) = \{1\}$, or equivalently $\mathscr{O}_K$ is a UFD).

Let's see what this yields when $K = \mathbf{Q}(\sqrt{d})$ for squarefree $d \in \mathbf{Z} - \{0, 1\}$.

(1) If $d > 0$, then $n = 2, r_1 = 2, r_2 = 0$, so

$$\lambda_K = \begin{cases} \frac{1}{2}\sqrt{d} & \text{if } d \equiv 1 \bmod 4, \\ \sqrt{d} & \text{if } d \equiv 2, 3 \bmod 4. \end{cases}$$

(2) If $d < 0$, then $n = 2, r_1 = 0, r_2 = 1$. We then obtain

$$\lambda_K = \begin{cases} \frac{2}{\pi}\sqrt{|d|} & \text{if } d \equiv 1 \bmod 4, \\ \frac{4}{\pi}\sqrt{|d|} & \text{if } d \equiv 2, 3 \bmod 4. \end{cases}$$

Thus, we have $\lambda_K < 2$ (and hence $h_K = 1$) precisely for

$$d = 2, 3, 5, 13, -1, -2, -3, -7.$$

In fact, these are all norm-Euclidean, so such cases are not so impressive for triviality of the class number (though admittedly the preceding approach treats all of them by a single slick method, without needing to battle with different norm formulas in many cases).

But $h_K = 1$ in many more cases by using more refined techniques to compute class groups as we shall begin to illustrate later today. For example,

$$h_K = 1 \text{ if } d = 17, 19, 21, 29, 33, -11$$

but $\lambda_K > 2$ in all of these cases. The case $K = \mathbf{Q}(\sqrt{-19})$ also satisfies $h_K = 1$ and it can be proved that $\mathscr{O}_K$ is not Euclidean (i.e., it has no possible norm function!).

For the rest of today, we answer the following question via concrete examples:

Question 25.5. When $\lambda_K \geq 2$, how do we find relations among the ideal classes $[\mathfrak{p}]$ for primes $\mathfrak{p}|p$ with $p \leq \lambda_K$?

We shall focus on some imaginary quadratic cases, taking up the harder real quadratic case next time (the extra difficulty due to units not of finite order, as we shall see).

Example 25.6. Consider $K = \mathbf{Q}(\sqrt{-14})$. We will show

$$\mathrm{Cl}(K) \simeq \mathbf{Z}/4\mathbf{Z}.$$

Let's first work out the Minkowski constant. We have $n = 2, r_1 = 0, r_2 = 1, \operatorname{disc} K = -56$, so approximately

$$\lambda_K \sim 4.764\ldots$$

Therefore, $\mathrm{Cl}(K)$ is generated by $[\mathfrak{p}]$ for $\mathfrak{p}$ over 2 and 3, so we should factor the ideals (2) and (3) into primes and then try to find as many relations as we can among the ideal classes of those prime factors. Recall that

$$\mathscr{O}_K = \mathbf{Z}[\sqrt{-14}] = \mathbf{Z}[x]/(x^2 + 14).$$

Thus, by Dedekind's method, to find the primes above 2 and 3, we factor $x^2 + 14 \bmod p$ for $p = 2$ and 3. We get

$$(2) = \mathfrak{p}_2^2,$$
$$(3) = \mathfrak{p}_3\mathfrak{p}_3',$$

for $\mathfrak{p}_2$ the unique prime above 2 and $\mathfrak{p}_3, \mathfrak{p}_3'$ the two *distinct* primes above 3. We conclude that the class group $\mathrm{Cl}(K)$ is generated by $[\mathfrak{p}_2], [\mathfrak{p}_3]$ because $[\mathfrak{p}_3'] = [\mathfrak{p}_3]^{-1}$ (as the product $\mathfrak{p}_3\mathfrak{p}_3'$ is principal). Also, $[\mathfrak{p}_2]^2 = 1$.

We'd like to determine if $[\mathfrak{p}_2]$ has order 2 or order 1; we just saw that its order divides 2. Since there is only one prime $\mathfrak{p}_2$ over 2 and its norm is 2, the principality of $\mathfrak{p}_2$ is equivalent to finding $\alpha \in \mathscr{O}_K$ with $|N\alpha| = 2$. Also, $N\alpha \geq 0$ for all $\alpha \in \mathscr{O}_K$ since K is imaginary quadratic. Explicitly, if $\alpha = u + v\sqrt{-14}$ with $u, v \in \mathbf{Z}$ then $2 = N\alpha = u^2 + 14v^2$, but this is impossible, so $[\mathfrak{p}_2]$ has order exactly 2.

What can we do about $[\mathfrak{p}_3]$? The key idea comes now: if we can find an element of norm 6 then its principal ideal would yield a relation in the class group between the classes of $\mathfrak{p}_2$ and either $\mathfrak{p}_3$ or $\mathfrak{p}_3'$ (with $[\mathfrak{p}_3'] = [\mathfrak{p}_3]^{-1}$), and in general if we can find an element α for which the only prime factors of its norm are 2 and 3 then factoring (α) will give us a relation between $\mathfrak{p}_2$ and $\mathfrak{p}_3$ in the class group. We now state this more precisely.

Norm technique. For $K = \mathbf{Q}(\theta)$, with $\theta \in \mathscr{O}_K$, and elements $\alpha := \sum_{i=0}^{n-1} c_i\theta^i$ with $c_i \in \mathbf{Z}$ we compute

$$N\alpha = N(\sum_{i=0}^{n-1} c_i\theta^i) \in \mathbf{Z}.$$

The principal ideal (α) is a product of primes $\mathfrak{p}$ over the primes $p \mid N(\alpha)$, yielding relations among such $[\mathfrak{p}]$'s in $\mathrm{Cl}(K)$. Hence, if we can find some α for which the prime factors of $N(\alpha)$ all satisfy $p \leq \lambda_K$ then we get relations in the class group among some primes $\mathfrak{p}$ over such p.

To see this technique in our example $K = \mathbf{Q}(\sqrt{-14})$, consider

$$N(m + \sqrt{-14}) = m^2 + 14$$

for various $m \in \mathbf{Z}$. We have $N(1 + \sqrt{-14}) = 15$ and $N(2 + \sqrt{-14}) = 18 = 2 \cdot 3^2$. Therefore, for some prime $\mathfrak{p}_5$ above 5 we have

$$(1 + \sqrt{-14}) = \mathfrak{p}_5 \cdot \mathfrak{p}_3 \text{ or } \mathfrak{p}_5 \cdot \mathfrak{p}_3'$$

We cannot yet distinguish if $\mathfrak{p}_3$ or $\mathfrak{p}_3'$ is a factor because we have not yet stated which prime above 3 is $\mathfrak{p}_3$. Similarly, we have that $(2 + \sqrt{-14})$ factors as one among the possibilities

$$\mathfrak{p}_2\mathfrak{p}_3^2,\ \ \mathfrak{p}_2\mathfrak{p}_3\mathfrak{p}_3',\ \ \mathfrak{p}_2(\mathfrak{p}_3')^2.$$

We can rule out the second case since $\mathfrak{p}_3\mathfrak{p}_3' = (3)$ and $3 \nmid (2 + \sqrt{-14})$ in $\mathbf{Z}[\sqrt{-14}]$. Thus, swapping $\mathfrak{p}_3$ and $\mathfrak{p}_3'$ if necessary, we may arrange that

$$(2 + \sqrt{-14}) = \mathfrak{p}_2\mathfrak{p}_3^2,$$

so

$$[\mathfrak{p}_3]^2 = [\mathfrak{p}_2]^{-1}$$

where $[\mathfrak{p}_2]$ has order 2.

Therefore, $[\mathfrak{p}_3]$ has order 4: its order divides 4, but its order does not divide 2 because its square is inverse to the element $[\mathfrak{p}_2]$ that is nontrivial (even with order exactly 2). This implies

$$\mathrm{Cl}(K) = \langle [\mathfrak{p}_3] \rangle \simeq \mathbf{Z}/4\mathbf{Z}.$$

Example 25.7. For $K = \mathbf{Q}(\sqrt{-30})$ we have $n = 2, r_1 = 0, r_2 = 1, \operatorname{disc} K = -120$, so

$$\lambda_K \sim 6.97\ldots$$

We therefore need to study the primes $\mathfrak{p}$ over $(2), (3)$, and (5). We claim $\mathrm{Cl}(K) = \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$. Using that $2, 3$, and 5 are all ramified, one then has three generators which are all 2-torsion. By norm considerations as above, these 2-torsion ideal classes are all non-trivial. In Homework 9, one finds a further relation among them and shows it is "the only relation".

Example 25.8. Consider $K = \mathbf{Q}(\sqrt{-65})$. Since $-65 \equiv 3 \bmod 4$, we have

$$\mathscr{O}_K = \mathbf{Z}[\sqrt{-65}].$$

We claim

$$\mathrm{Cl}(K) = (\mathbf{Z}/4\mathbf{Z}) \times (\mathbf{Z}/2\mathbf{Z}).$$

Computing the Minkowski constant, we get

$$\lambda_K = \frac{4}{\pi}\sqrt{65} \sim 10.26,$$

so we need to look at the factors of $(2), (3), (5)$ and (7) in $\mathbf{Z}[\sqrt{-65}]$.

By factoring $x^2 + 65 \bmod p$ for $p = 2, 3, 5, 7$, we get

$$\begin{aligned}(2) &= \mathfrak{p}_2^2\\ (3) &= \mathfrak{p}_3\mathfrak{p}_3'\\ (5) &= \mathfrak{p}_5^2\\ (7) &= \mathfrak{p}_7.\end{aligned}$$

We therefore see the class group is generated by the classes of $\mathfrak{p}_2, \mathfrak{p}_3, \mathfrak{p}_5$, since the class of $\mathfrak{p}_7$ is trivial and $[\mathfrak{p}_3'] = [\mathfrak{p}_3]^{-1}$. We further see $\mathfrak{p}_2$ and $\mathfrak{p}_5$ are 2-torsion in the class group. In fact, all of $\mathfrak{p}_2, \mathfrak{p}_3, \mathfrak{p}_5$ are not principal because their norms are $2, 3$, and 5 and one cannot write any of $2, 3$, or 5 in the form $u^2 + 65v^2$ for $u, v \in \mathbf{Z}$.

This tells us that $[\mathfrak{p}_2]$ and $[\mathfrak{p}_5]$ both have order exactly 2 in the class group. We now want to link $[\mathfrak{p}_3]$ to $[\mathfrak{p}_2]$ and $[\mathfrak{p}_5]$ using the norm technique from the previous example. We seek an element whose norm only involves $2, 3$, and 5 as prime factors. To do this, we compute

$$N(a + \sqrt{-65})$$

for various small $a \in \mathbf{Z}$. (There is no particular reason that we have a coefficient 1 in front of $\sqrt{-65}$; we are just playing around and trying different things.) That is, we hope to find some $a^2 + 65$ with all prime factors in $\{2, 3, 5\}$.

Try $a = 4$ and $a = 5$: $N(4 + \sqrt{-65}) = 81 = 3^4$ and $N(5 + \sqrt{-65}) = 90 = 2 \cdot 3^2 \cdot 5$. Therefore,

$$(4 + \sqrt{-65}) = \mathfrak{p}_3^4 \text{ or } \mathfrak{p}_3'^4$$

using that $3 \nmid (4 + \sqrt{-65})$ (to rule out any intervention by $\mathfrak{p}_3\mathfrak{p}_3'$). Let's relabel $\mathfrak{p}_3$ and $\mathfrak{p}_3'$ if necessary so that $(4 + \sqrt{-65}) = \mathfrak{p}_3^4$. Then

$$(5 + \sqrt{-65}) = \mathfrak{p}_2\mathfrak{p}_5\mathfrak{p}_3'^2$$

since we know it is a product of a prime above 2, a prime above 5, and two primes above 3 with the two primes above 3 equal to each other (i.e., $\mathfrak{p}_3$ and $\mathfrak{p}_3'$ can't both appear) because $3 \nmid (5 + \sqrt{-65})$ and not both equal to $\mathfrak{p}_3$ because $\mathfrak{p}_3 \mid (4 + \sqrt{-65})$ yet $4 + \sqrt{-65}$ and $5 + \sqrt{-65}$ generate 1 (so their associated principal ideals cannot have a common prime factor, such as $\mathfrak{p}_3$).

Therefore,

$$\begin{aligned} 1 &= [\mathfrak{p}_2][\mathfrak{p}_3']^2[\mathfrak{p}_5] \\ &= [\mathfrak{p}_2][\mathfrak{p}_3]^{-2}[\mathfrak{p}_5]. \end{aligned}$$

This relation among the ideal classes of $\mathfrak{p}_2, \mathfrak{p}_3, \mathfrak{p}_5$ allows us to drop $[\mathfrak{p}_5]$ from the list of generators: $\mathrm{Cl}(K)$ is generated by the element $[\mathfrak{p}_2]$ of order 2 and the element $[\mathfrak{p}_3]$ of order dividing 4.

We claim $[\mathfrak{p}_3]$ has order exactly 4. We know its order is either 2 or 4 since $\mathfrak{p}_3$ is not principal. To rule out the possibility $[\mathfrak{p}_3]^2 = 1$, suppose otherwise, so $\mathfrak{p}_3^2 = (\alpha)$ for some α that must satisfy $N\alpha = 9$. If we write $N\alpha = u^2 + 65v^2$, we see that the only way this can happen is if $u = \pm 3, v = 0$, which would say $\mathfrak{p}_3^2 = (\alpha) = (3) = \mathfrak{p}_3\mathfrak{p}_3'$, an absurdity (since $\mathfrak{p}_3' \neq \mathfrak{p}_3$). Therefore, we have a surjection

$$\mathbf{Z}/4\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z} \to \mathrm{Cl}(K)$$

with $\mathbf{Z}/4\mathbf{Z}$ corresponding to the subgroup generated by $[\mathfrak{p}_3]$ and $\mathbf{Z}/2\mathbf{Z}$ corresponding to the subgroup generated $[\mathfrak{p}_2]$. In particular, the class group has order dividing 8.

To show this surjective map is an isomorphism, we only need rule out the possibility that the class group doesn't have order 8, which is to say that it coincides with its subgroup of order 4 generated by $[\mathfrak{p}_3]$. In this latter case we would have $[\mathfrak{p}_3]^2 = [\mathfrak{p}_2]$ for order reasons, so we just need to rule out this latter equality.

Suppose $[\mathfrak{p}_3]^2 = [\mathfrak{p}_2]$, so

$$[\mathfrak{p}_2][\mathfrak{p}_3]^2 = [\mathfrak{p}_3]^4 = 1.$$

Therefore, we would obtain

$$\mathfrak{p}_2\mathfrak{p}_3^2 = (\alpha)$$

for some $\alpha = u + v\sqrt{-65}$ with $u, v \in \mathbf{Z}$. But then

$$u^2 + 65v^2 = N\alpha = 18,$$

which is clearly impossible. Thus,

$$\mathrm{Cl}(K) \simeq \mathbf{Z}/4\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$$

Next time, we'll investigate a real quadratic example, in which case we'll have to grapple with new difficulties involving units.

26. Norms and volumes

An example of the class group of a real quadratic extension. Last time, we worked out a number of examples of class groups of imaginary quadratic fields. We'll begin by working through an example of a real quadratic case. It will become apparent why this is harder when we encounter the plethora of units in the associated ring of integers.

Example 26.1. Let's compute $\mathrm{Cl}(K)$ for $K = \mathbf{Q}(\sqrt{82})$. We'll show $\mathrm{Cl}(K) = \mathbf{Z}/4\mathbf{Z}$. We have $n = 2, r_1 = 2, r_2 = 0, \mathrm{disc}(K) = 4 \cdot 82$, so

$$\lambda_K \sim 9.055\ldots$$

The upshot is that the class group is generated by the classes $[\mathfrak{p}]$ for primes $\mathfrak{p}$ over $2, 3, 5$, and 7.

Since $\mathscr{O}_K = \mathbf{Z}[\sqrt{82}] = \mathbf{Z}[x]/(x^2 - 82)$, to factor $p\mathscr{O}_K$ is the same as to factor $x^2 - 82 \in \mathbf{F}_p[x]$. Thus, $(2) = \mathfrak{p}_2^2$ and $(3) = \mathfrak{p}_3\mathfrak{p}_3'$ for primes $\mathfrak{p}_3 \neq \mathfrak{p}_3'$. Via quadratic reciprocity or by hand one can check that 82 is not a square modulo 5 and 7, so (5) and (7) are actually prime in $\mathscr{O}_K$. Therefore, $\mathrm{Cl}(K)$ is generated by $[\mathfrak{p}_2]$ (which is 2-torsion) and $[\mathfrak{p}_3]$ (the class of either of the primes over 3).

Let's now cross our fingers and hope we can find some nonzero element in $\mathscr{O}_K$ whose norm has as only 2 and/or 3 as its prime factors: we'll compute

$$N(a + \sqrt{82}) = a^2 - 82$$

for some small integers a. Doing this, we eventually discover

$$N(10 + \sqrt{82}) = 18 = 2 \cdot 3^2.$$

This tells us

$$(10 + \sqrt{82}) = \mathfrak{p}_2 \cdot \mathfrak{p}_3^2$$

after possibly switching $\mathfrak{p}_3$ and $\mathfrak{p}_3'$. Here we are using that 3 visibly does not divide $10 + \sqrt{82}$ in $\mathscr{O}_K = \mathbf{Z}[\sqrt{82}]$, so $(10 + \sqrt{82})$ cannot have both $\mathfrak{p}_3$ and $\mathfrak{p}_3'$ as prime factors (as their product is (3)).

Therefore,

$$[\mathfrak{p}_3]^2 = [\mathfrak{p}_2]^{-1} = [\mathfrak{p}_2].$$

We conclude that $[\mathfrak{p}_3]$ generates the class group and is 4-torsion. To show $[\mathfrak{p}_3]$ has order exactly 4, it suffices to show that the ideal class $[\mathfrak{p}_3]^2 = [\mathfrak{p}_2]$ is nontrivial; i.e., we are reduced to showing $\mathfrak{p}_2$ is not principal.

So far, we have not encountered any difficulties in dealing with a real quadratic field instead of an imaginary quadratic field. Suppose $\mathfrak{p}_2 = (\alpha)$ for some $\alpha \in \mathscr{O}_K = \mathbf{Z}[\sqrt{82}]$; we seek a contradiction. We can write $\alpha = u + v\sqrt{82}$ for some $u, v \in \mathbf{Z}$. The equality $\mathfrak{p}_2 = (\alpha)$, implies $|N_{K/\mathbf{Q}}(\alpha)| = N\mathfrak{p}_2 = 2$ so $u^2 - 82v^2 = N_{K/\mathbf{Q}}(\alpha) = \pm 2$. Conversely, if $N_{K/\mathbf{Q}}(\alpha) = \pm 2$, then $(\alpha) = \mathfrak{p}_2$ since $\mathfrak{p}_2$ is the unique prime over (2).

Hence, our task of showing $[\mathfrak{p}_2] \neq 1$ is equivalent to showing

$$x^2 - 82y^2 = \pm 2 \tag{26.1}$$

has no solutions in $\mathbf{Z}$. The huge difference in this real quadratic case is that to rule out integral solutions to such a "norm equation" we can no longer simply do a quick finite check as in the imaginary quadratic case.

In $\mathscr{O}_K^\times$, we have a fundamental unit given by

$$\varepsilon = 9 + \sqrt{82}. \tag{26.2}$$

What really matters initially is that $N(\varepsilon) = -1$, so after potentially multiplying by ε we see that if we had a solution to (26.1) for one of the signs on the right side then we get a solution for the other sign too. Hence, it doesn't actually matter whether we consider 2 or -2 on the right side.

To appreciate the subtlety involved in proving the absence of $\mathbf{Z}$-solutions, we observe two features of the conic (26.2) with $+2$ on the right. Firstly, it has infinitely many $\mathbf{Q}$-points, one of which is $(10/3, 1/3)$. Moreover, this conic even has points $\bmod m$ for all $m \geq 1$ (this requires some work, but is ultimately some effort with congruential algebra and doesn't involve anything very deep, though handling m divisible by high powers of 2 or 41 does entail some additional thought); e.g., for $m = 41$ we have $17^2 \equiv 2 \bmod 41$ and -1 is a square modulo 41, so -2 is also a square modulo 41.

The upshot is that we cannot rule out $\mathbf{Z}$-solutions just by congruence reasons, so the absence of such solutions is not an entirely "elementary" assertion. On Exercise 5 of Homework 9 you will show (26.1) has no $\mathbf{Z}$-solutions. A key input in the argument, going beyond congruential information, is that ε is a fundamental unit satisfying $N(\varepsilon) = -1$, due to which the units of norm 1 are up to a sign precisely the squares of units. One builds on that to show that if there exists a $\mathbf{Z}$-solution then one of ± 2 is a square in $\mathbf{Q}(\sqrt{82})$, which is absurd.

The geometry of numbers. Finiteness of class groups and finite generation of unit groups were understood by the 1880's, but in 1896 Minkowski came up with a way to establish such results via consideration of lattices in Euclidean spaces attached to number fields. His method, called "geometry of numbers", has applications to the arithmetic of quadratic forms and many other problems in number theory.

We shall turn the questions of finiteness of the class group and finite generation of units into one of finding nonzero vectors of lattices contained in specific bounded regions in a Euclidean space. To achieve this, we shall show the following reformulation of the refined finiteness of class numbers:

Theorem 26.2. *For any nonzero ideal $\mathfrak{b} \subset \mathscr{O}_K$, there is some $\beta \in \mathfrak{b} - \{0\}$ such that*

$$|N_{K/\mathbf{Q}}(\beta)| \leq \lambda_K \cdot N\mathfrak{b}.$$

Let's see why this implies Theorem 25.2.

Proof. The fractional ideal $\mathfrak{a} := \beta\mathfrak{b}^{-1}$ lies inside $\mathscr{O}_K$ since $\beta \in \mathfrak{b}$ (think back to how $\mathfrak{b}^{-1}$ is defined), so $\mathfrak{a}$ is an ordinary nonzero ideal of $\mathscr{O}_K$. We have

$$[\mathfrak{a}] = [\mathfrak{b}]^{-1}$$

and $N\mathfrak{a} = |N_{K/\mathbf{Q}}\beta| N\mathfrak{b}^{-1} \leq \lambda_K$. The upshot is that $[\mathfrak{b}]^{-1}$ has an integral ideal representative with norm at most λ_K. However, every element of $\mathrm{Cl}(K)$ has this form, since $\mathfrak{b}$ is just an integral representative of the inverse class. □

We next aim to give a geometric interpretation of norms in terms of volumes of parallelotopes in some $\mathbf{R}^N$. Which $\mathbf{R}^N$ do we take?

The time has come to consider *all* embeddings of a number field into $\mathbf{C}$ at once (e.g., for a real quadratic field, it will be essential to treat both real embeddings on an equal footing, and not prefer one over the other). The key geometric tool is the following map:

$$\begin{aligned} \theta_K \colon K &\to \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} \\ x &\mapsto (\sigma_1(x), \ldots, \sigma_{r_1}(x), \sigma_{r_1+1}, \ldots, \sigma_{r_1+r_2}(x)) \end{aligned}$$

where we identify $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ with $\mathbf{R}^n$ for $n = r_1 + 2r_2$ via identifying $\mathbf{C}$ with $\mathbf{R} \oplus \mathbf{R} \cdot i$; here $\sigma_1, \ldots, \sigma_{r_1} : K \to \mathbf{R}$ are the real embeddings and

$$\sigma_j,\ \overline{\sigma}_j : K \rightrightarrows \mathbf{C}$$

for $r_1 + 1 \leq j \leq r_1 + r_2$ is the set of conjugate pairs of non-real embeddings.

Lemma 26.3. *For any nonzero ideal $\mathfrak{b} \subset \mathscr{O}_K$ the image $\theta_K(\mathfrak{b}) \subset \mathbf{R}^n$ is a lattice (i.e., the $\mathbf{Z}$-span of a $\mathbf{R}$-basis).*

Proof. Any $\mathbf{Z}$-basis of $\mathfrak{b}$ is a $\mathbf{Q}$-basis of K, so it is enough to show that if $\{e_1, \dots, e_n\}$ is *any* $\mathbf{Q}$-basis of K then the n vectors $\{\theta_K(e_j)\}_{1 \leq j \leq n}$ is an $\mathbf{R}$-linearly independent set in $\mathbf{R}^n$.

To show this, it is the same to show that the $n \times n$ real matrix

$$M = \begin{pmatrix} \theta_K(e_1) \\ \vdots \\ \theta_K(e_n) \end{pmatrix}$$

has nonzero determinant. We will establish this using column operations, actually working with $|\det(M)|$ and regarding M as a matrix with its n^2 entries in $\mathbf{C}$ rather than just in $\mathbf{R}$ (so we may applying column operations involving scaling against elements of $\mathbf{C}^\times$ such as $\pm i$).

More explicitly, writing $\Re(z)$ and $\Im(z)$ to denote the real and imaginary parts of $z \in \mathbf{C}$, M is given as:

$$\begin{pmatrix} \sigma_1(e_1) & \dots & \sigma_{r_1}(e_1) & \Re\sigma_{r_1+1}(e_1) & \Im\sigma_{r_1+1}(e_1) & \dots & \Re\sigma_{r_1+r_2}(e_1) & \Im\sigma_{r_1+r_2}(e_1) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_1(e_n) & \dots & \sigma_{r_1}(e_n) & \Re\sigma_{r_1+1}(e_n) & \Im\sigma_{r_1+1}(e_n) & \dots & \Re\sigma_{r_1+r_2}(e_n) & \Im\sigma_{r_1+r_2}(e_n) \end{pmatrix}$$

The rightmost $2r_2$ columns come in pairs $\Re\sigma_j$ and $\Im\sigma_j$, so for such $r_1 + 1 \leq j \leq r_1 + r_2$ if we multiply the column of $\Im\sigma_j$'s by $-i$ (which has no effect on $|\det(M)|$) and subtract this from the column of $\Re\sigma_j$'s then we turn the $\Re\sigma_j$-column into σ_j. Next, doubling each $-i\Im\sigma_j$-column is counteracted by multiplying the determinant by 2^{-r_2}. Adding the σ_j-column to the $-2i\Im(\sigma_j)$-column yields $\overline{\sigma}_j$ in place of $-2i\Im\sigma_j$. Therefore, we have arrived at the matrix in $\mathrm{Mat}_n(\mathbf{C})$ given by

$$\begin{pmatrix} \sigma_1(e_1) & \dots & \sigma_{r_1}(e_1) & \sigma_{r_1+1}(e_1) & \overline{\sigma}_{r_1+1}(e_1) & \dots & \sigma_{r_1+r_2}(e_1) & \overline{\sigma}_{r_1+r_2}(e_1) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_1(e_n) & \dots & \sigma_{r_1}(e_n) & \sigma_{r_1+1}(e_n) & \overline{\sigma}_{r_1+1}(e_n) & \dots & \sigma_{r_1+r_2}(e_n) & \overline{\sigma}_{r_1+r_2}(e_n) \end{pmatrix}$$

In other words, this is the matrix $(\sigma(e_m))$ whose entries are indexed by pairs (σ, m) for *all* embeddings $\sigma : K \to \mathbf{C}$ and $1 \leq m \leq n$. Our task is reduced to showing that this latter matrix has nonzero determinant.

This final non-vanishing property is *independent* of the choice of $\mathbf{Q}$-basis $\{e_j\}$ of K since it is easy to check that if we applying an invertible $\mathbf{Q}$-linear change of basis to $\{e_j\}$ then this final matrix is changed via multiplication against an invertible $n \times n$ matrix with rational entries. Thus, it is now enough to consider the matrix $(\sigma(e_m))_{(\sigma,m)}$ for *one* $\mathbf{Q}$-basis $\{e_j\}$ of K. We consider a power basis! That is, for $\alpha \in K$ a primitive element over $\mathbf{Q}$, take

$\{e_j\}$ to be $\{1, \alpha, \ldots, \alpha^{n-1}\}$. Then our matrix is $(\sigma(\alpha^m))$, which is precisely the matrix that came up in the proof of the non-vanishing of the discriminant of K: this is a van der Monde matrix whose determinant is nonzero since $\sigma(\alpha) \neq \tau(\alpha)$ for any $\sigma \neq \tau$ (as $\mathbf{Q}(\alpha) = K$). □

27. Volume calculations

Application of $K \to \mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ to video games. There is a video game called *Lucy and Lily* by Rich Schwartz, a math professor at Brown; it is on his webpage at Brown, and he wrote an article about it for the *American Mathematical Monthly* (also on his webpage).

Take $K = \mathbf{Q}(\zeta)$ for a primitive 5th root of unity ζ. We consider two pentagons in their own copies of $\mathbf{C}$, with center at the origin and vertices at the 5th roots of unity in $\mathbf{C}$; call these pentagons Lucy and Lily (the names of Rich Schwartz' children). Each pentagon is dissected into 5 equal sectors as usual (sector edges joining the vertices to the center), colored with a common set of 5 colors. The coloring order in the counterclockwise direction of Lily is related to that of Lucy by doubling the angles (note that 2 generates $(\mathbf{Z}/5\mathbf{Z})^\times$). Combining this information amounts to considering

$$\theta_K : \mathscr{O}_K = \mathbf{Z}[\zeta] \to \mathbf{C} \times \mathbf{C}$$

defined by $\zeta \mapsto (\zeta_5, \zeta_5^2)$ for $\zeta_5 := e^{2\pi i/5}$.

The pentagons move by flipping across an exterior edge, and whatever is done in one plane happens in the other via flipping through the exterior edge of the same color (so geometric effects in the two planes are different, due to the angle-doubling on colored sectors). The game begins by closing your eyes while the computer does many flips, and the challenge upon opening your eyes is to move Lucy back to 0 through exterior edge flips. This is impossible to win using just Lucy because the discrete $\theta_K(\mathbf{Z}[\zeta]) \subset \mathbf{C}^2$ has *dense* image under projection to either factor $\mathbf{C}$ (so the set of centers of legal positions of the pentagon is dense!); a later handout will prove density.

The "winning strategy" (explained at length in the AMM article) exploits the discreteness of $\mathscr{O}_K$ in $\mathbf{C}^2 = \mathbf{R}^4$ to overcome the density of the image of $\mathscr{O}_K$ in each copy of $\mathbf{C}$: one first moves Lucy as efficiently as possible towards the origin via exterior edge flips while Lily makes the "same" flips (based on colors) but moves quite differently. The distance loss by Lily winds up being less than the gain made by Lucy, and so if we then switch to moving Lily as efficiently as possible towards the origin we entail some loss for Lucy by it is not as bad as the gain by Lily. Then go back to efficient moves for Lucy towards the origin, etc. The linked motion corresponds to making progress towards $(0, 0) \in \mathbf{C}^2$ in the lattice $\theta_K(\mathscr{O}_K)$ as measured by distance in $\mathbf{R}^4$, so by the *discreteness* of $\theta_K(\mathscr{O}_K)$ we eventually win (i.e., reach $0 \in \mathbf{C}$)!

Remark 27.1. On the same website is a "$\sqrt{2}$-game" with boats hopping distances of 1 or $\sqrt{2}$ in $\mathbf{R}$ to illustrate the phenomenon with $K = \mathbf{Q}(\sqrt{2}) \to \mathbf{R}^2$: one sees 2-dimensional discreteness and 1-dimensional density, and why the strategy of moving optimally in alternating "parallel worlds" works.

Geometric meaning of determinant.

Lemma 27.2. *Let $\Lambda \subset \mathbf{R}^n$ be a lattice, and let $\mathbf{v} := \{v_i\}$ be an ordered $\mathbf{Z}$-basis of Λ. For the parallelotope $P_{\mathbf{v}} := \{\sum_i t_i v_i \,|\, 0 \le t_i \le 1\}$, $\mathrm{Vol}(P_{\mathbf{v}})$ is independent of $\mathbf{v}$ and is given by the $n \times n$ absolute determinants:*

$$\mathrm{Vol}\, P_{\mathbf{v}} = \left| \det \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \right| = \left| \det \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \right| \tag{27.1}$$

(*where both matrices inside the determinants are $n \times n$*).

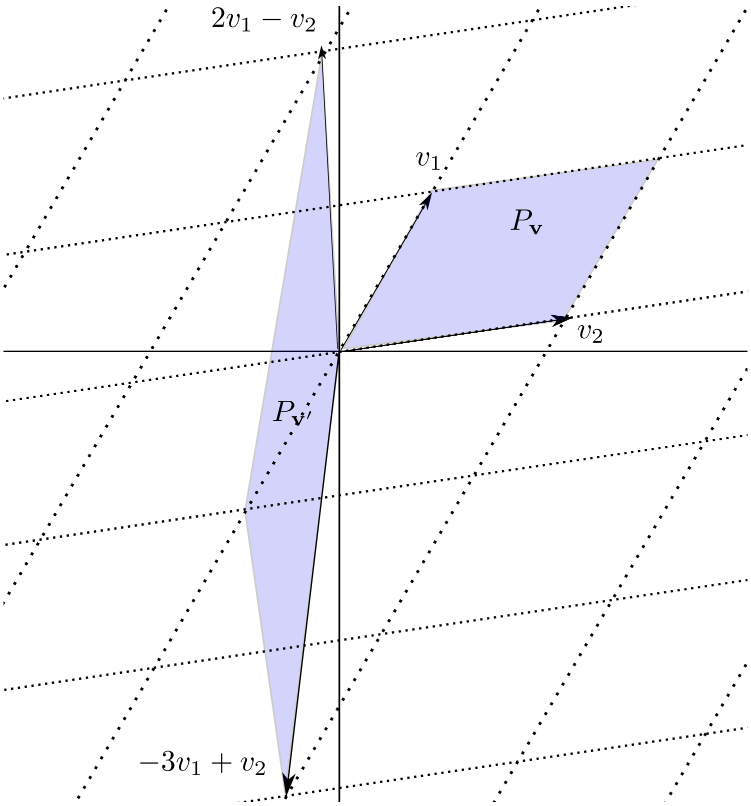

FIGURE 1. Both $\mathbf{v} = \{v_1, v_2\}$ and $\mathbf{v}' = \{2v_1 - v_2, -3v_1 + v_2\}$ have the same $\mathbf{Z}$-span, and $P_{\mathbf{v}}$ and $P_{\mathbf{v}'}$ have the same area.

Before proving the lemma, we make some observations. From a geometric viewpoint the independence of $\mathbf{v}$ may be a bit surprising. For example with $n = 2$, $\mathbf{v}' = \{2v_1 - v_2, -3v_1 + v_2\}$ is also a $\mathbf{Z}$-basis of $\mathbf{Z}v_1 \oplus \mathbf{Z}v_2$ because

$$\det \begin{pmatrix} 2 & -3 \\ -1 & 1 \end{pmatrix} = -1 \in \mathbf{Z}^\times$$

(so the inverse is also an integer matrix), yet the geometry of the parallelograms $P_{\mathbf{v}'}$ and $P_{\mathbf{v}}$ are very different (but areas agree), as shown in Figure 1.

Also, the real explanation for $P_{\mathbf{v}}$ having volume independent of $\mathbf{v}$ is that the compact coset space $\mathbf{R}^n/\Lambda$ that is intrinsic to Λ (i.e., it doesn't involve a choice of $\mathbf{v}$) inherits a natural measure from the one on $\mathbf{R}^n$ relative to which its volume is that of every $P_{\mathbf{v}}$ (informally, $P_{\mathbf{v}} \to \mathbf{R}^n/\Lambda$ is a measure-preserving isomorphism "away from the measure-zero boundary"). However, we don't want to digress into a discussion of measures on coset spaces, so we opt for the perspective of $P_{\mathbf{v}} \subset \mathbf{R}^n$ that is sufficient for our needs.

Now we take up the proof of the lemma:

Proof. By definition $P_{\mathbf{v}}$ is the image of $[0,1]^n$ under the invertible linear map

$$T : \mathbf{R}^n \to \mathbf{R}^n$$

corresponding to the $n \times n$ matrix

$$\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$$

Therefore,

$$\mathrm{Vol}(P_{\mathbf{v}}) = |\det T| \cdot \mathrm{Vol}([0,1]^n),$$

which is the right side of (27.1).

For independence of $\mathbf{v}$, if $\mathbf{v}' = \{v'_1, \ldots, v'_n\}$ with $v'_j = \sum_i a_{ij} v_i$ another ordered $\mathbf{Z}$-basis of Λ then the integer matrix (a_{ij}) has inverse matrix that is also an integer matrix (why?) and hence its determinant belongs to $\mathbf{Z}^\times = \{\pm 1\}$. But it is elementary to check (work it out for $n = 2$ first) that

$$\begin{pmatrix} v'_1 & \cdots & v'_n \end{pmatrix} = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \cdot (a_{ij})$$

as $n \times n$ matrices. Applying $|\det(\cdot)|$ to both sides gives the result. □

Definition 27.3. We write Vol_Λ to denote $\mathrm{Vol}(P_{\mathbf{v}})$ for any ordered $\mathbf{Z}$-basis $\mathbf{v}$ of Λ.

To apply the preceding lemma in the proof of finiteness of the class number, let's recall that we have previously reduced the finiteness to the following:

Goal 27.4. For a nonzero ideal $\mathfrak{b} \subset \mathscr{O}_K$, we seek $\beta \in \mathfrak{b} - \{0\}$ so that

$$|N_{K/\mathbf{Q}}\beta| \leq \lambda_K \cdot N\mathfrak{b},$$

where

$$\lambda_K = \frac{n!}{n^n}(4/\pi)^{r_2}\sqrt{|\operatorname{disc} K|}.$$

We need the following result to reinterpret our goal in terms of volumes attached to lattices:

Theorem 27.5. *For a nonzero integral ideal* $\mathfrak{b} \subset K \xrightarrow{\theta_K} \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} = \mathbf{R}^n$, *we have*

$$\mathrm{Vol}_{\mathfrak{b}} := \mathrm{Vol}_{\theta_K(\mathfrak{b})} = 2^{-r_2}\sqrt{|\operatorname{disc} K|} \cdot N\mathfrak{b}.$$

For example, in the above notation, we have

$$\mathrm{Vol}_{\theta_K(\mathscr{O}_K)} = 2^{-r_2}\sqrt{|\operatorname{disc} K|}.$$

Here is an explicit example:

Example 27.6. Take $K = \mathbf{Q}(\alpha)$ with $\alpha^2 = 3$. Let's define the embeddings $\sigma_i : K \to \mathbf{R}$ by $\sigma_1(\alpha) = \sqrt{3}$ and $\sigma_2(\alpha) = -\sqrt{3}$. Then $\mathscr{O}_K = \mathbf{Z}[\alpha]$ and

$$\theta_K(\mathbf{Z}[\alpha]) = \{\theta_K(a + b\alpha) = (a + b\sqrt{3}, a - b\sqrt{3}) \mid a, b \in \mathbf{Z}\}$$

is the $\mathbf{Z}$-span of $\theta_K(1) = (1,1)$ and $\theta_K(\alpha) = (\sqrt{3}, -\sqrt{3})$. In other words,

$$(a + b\sqrt{3}, a - b\sqrt{3}) = a \cdot (1,1) + b \cdot (\sqrt{3}, -\sqrt{3})$$

is the lattice spanned over $\mathbf{Z}$ by $\theta_K(1), \theta_K(\alpha)$. Its associated volume ("area") is

$$\left|\det\begin{pmatrix} 1 & \sqrt{3} \\ 1 & -\sqrt{3} \end{pmatrix}\right| = 2\sqrt{3} = \sqrt{12}.$$

as desired. Similarly, for the ideal $\mathfrak{b} = (1 + \alpha)$ with norm $|\mathrm{N}_{K/\mathbf{Q}}(1 + \alpha)| = |-2| = 2$ we have

$$\begin{aligned}\theta_K(\mathfrak{b}) &= \{\theta_K((1 + \alpha) \cdot (a + b\alpha)) \mid a, b \in \mathbf{Z}\} \\ &= \{((1 + \sqrt{3})(a + b\sqrt{3}), (1 - \sqrt{3})(a - b\sqrt{3})) \mid a, b \in \mathbf{Z}\} \\ &= \{a(1 + \sqrt{3}, 1 - \sqrt{3}) + b(3 + \sqrt{3}, 3 - \sqrt{3}) \mid a, b \in \mathbf{Z}\},\end{aligned}$$

so

$$\mathrm{Vol}_{\mathfrak{b}} = \left|\det\begin{pmatrix} 1 + \sqrt{3} & 3 + \sqrt{3} \\ 1 - \sqrt{3} & d - \sqrt{3} \end{pmatrix}\right| = 2\sqrt{12},$$

consistent with the fact that we saw $N\mathfrak{b} = 2$.

The proof of the general formula for $\mathrm{Vol}_{\mathfrak{b}}$ is given in the handout "Volume Attached to an Ideal". Here, we'll explain why

$$\mathrm{Vol}_{\mathfrak{b}} = N\mathfrak{b} \cdot \mathrm{Vol}_{\mathscr{O}_K};$$

the computation $\mathrm{Vol}_{\mathscr{O}_K} = 2^{-r_2}\sqrt{|\operatorname{disc} K|}$ as given in that handout is essentially a more refined version of what we did last time where we performed a linear transformation to turn the matrix for our embedding $\theta_K : K \to \mathbf{R}^n$ into a van der Monde matrix related to a discriminant.

Since $N\mathfrak{b} = [\mathscr{O}_K : \mathfrak{b}]$, the desired relation between $\mathrm{Vol}_{\mathfrak{b}}$ and $\mathrm{Vol}_{\mathscr{O}_K}$ is a special case of the more general claim for *any* pair of lattices $\Lambda' \subset \Lambda$ in $\mathbf{R}^n$ that

$$\mathrm{Vol}_{\Lambda'} = [\Lambda : \Lambda'] \cdot \mathrm{Vol}_{\Lambda}.$$

By the structure theorem for modules over a PID, we can always choose $\mathbf{Z}$-bases of the two lattices so that the $\mathbf{Z}$-basis of Λ' is given by multiples of a $\mathbf{Z}$-basis of Λ:

$$\Lambda = \oplus \mathbf{Z} v_i, \Lambda' = \oplus \mathbf{Z}(c_i v_i)$$

for some $c_i \in \mathbf{Z} - \{0\}$. Thus, $[\Lambda : \Lambda'] = \prod_i |c_i|$, and for $\mathbf{v}' = \{c_i v_i\}$ the parallelogram $P_{\mathbf{v}'}$ is covered by $\prod_i |c_i|$ translates of $P_{\mathbf{v}}$ that are "essentially disjoint" (from the viewpoint of volumes) in the sense that their pairwise overlaps are contained in hyperplanes (hence not contributing to volume):

$$P_{\mathbf{v}'} = \bigcup_{\vec{r}} (P_{\mathbf{v}} + \sum r_j v_j)$$

for varying $\vec{r} = (r_1, \ldots, r_n)$ with $0 \le r_j < |c_j|$. This implies $\mathrm{Vol}(P_{\mathbf{v}'}) = (\prod_i |c_i|) \cdot \mathrm{Vol}(P_{\mathbf{v}})$, or equivalently $\mathrm{Vol}_{\Lambda'} = [\Lambda' : \Lambda]\,\mathrm{Vol}_{\Lambda}$ as desired.

Now consider the diagram

$$\begin{tikzcd} K \arrow[r, "\theta_K"] \arrow[d, "|N_{K/\mathbf{Q}}|"] & \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} \arrow[d, "\mathcal{N}"] \\ \mathbf{Q} \arrow[r] & \mathbf{R} \end{tikzcd} \tag{27.2}$$

with $\mathcal{N} \colon \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} \to \mathbf{R}$ defined by

$$(\vec{x}, \vec{z}) \mapsto \prod |x_i| \cdot \prod |z_j|^2.$$

It is left as an exercise to show this diagram commutes. (The main point is that $|z|^2 = |z||\bar{z}|$ implies $\mathcal{N} \circ \theta_K : v \mapsto \prod_{\sigma:K\to\mathbf{C}} |\sigma(v)| = |\prod_{\sigma:K\to\mathbf{C}} \sigma(v)|$ with σ varying through *all* field embeddings $K \to \mathbf{C}$, including *both* members of each pair of conjugate non-real embeddings, so the commuting of the diagram reduces to the purely algebraic formula $N_{K/\mathbf{Q}}(v) = \prod_{\sigma:K\to\mathbf{C}} \sigma(v)$ for each $v \in K$, which in turn is proved via transitivity of field norms applied to the tower $K/\mathbf{Q}(v)/\mathbf{Q}$.)

Using this, it suffices to prove the following result in which K has essentially disappeared (except for the values of r_1 and r_2 that acquire the status of independent parameters in the formulation):

Theorem 27.7. *For any $r_1, r_2 \geq 0$ at least one of which is positive, $n := r_1 + 2r_2$, and any lattice $L \subset \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} = \mathbf{R}^n$, we have*

$$L \cap \{v \in \mathbf{R}^n \,|\, \mathcal{N}(v) \leq \frac{n!}{n^n}(8/\pi)^{r_2}\,\mathrm{Vol}_L\} \neq 0.$$

Theorem 26.2 follows from this general result for lattices applied to $L = \theta_K(\mathfrak{b})$ due to the commutative diagram (27.2) and the general formula for $\mathrm{Vol}_{\mathfrak{b}}$! We now prove Theorem 27.7:

Proof. Since $r_1 + 2r_2 = n$, for $v = (\vec{x}, \vec{z}) \in \mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ the arithmetic-mean geometric-mean inequality gives

$$\mathcal{N}(v)^{1/n} \leq (\sum_i |x_i| + 2\sum_j |z_j|)/n.$$

Thus, it is enough to show

$$L \cap \{v \in \mathbf{R}^n \,|\, ((\sum_i |x_i| + 2\sum_j |z_j|)/n)^n \leq \frac{n!}{n^n}(8/\pi)^{r_2}\,\mathrm{Vol}_L\} \neq 0$$

(note that the denominator n^n on both sides of the inequality cancels out). To prove this latter assertion, we shall use the following volume computation that is proved in [Samuel, Ch. IV, Appendix, pp. 66-67] via a clever application of Fubini's theorem and an induction on r_1 and r_2:

Lemma 27.8. *For $t > 0$, the compact region*

$$X_t := \{(x, z) \in \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} \,|\, \sum_i |x_i| + 2\sum_j |z_j| \leq t\}$$

satisfies

$$\mathrm{Vol}(X_t) = 2^{r_1}(\pi/2)^{r_2}\frac{t^n}{n!}.$$

Now, consider

$$D = \{(x, z) \in \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} \,|\, ((\sum_i |x_i| + 2\sum_j |z_j|)/n)^n \leq \frac{n!}{n^n}(8/\pi)^{r_2}\,\mathrm{Vol}_L\}.$$

This is X_{t_0} of Lemma 27.8 for t_0 satisfying

$$t_0^n = n!(8/\pi)^{r_2}\,\mathrm{Vol}_L,$$

so

$$\begin{aligned}\mathrm{Vol}(X_{t_0}) &= 2^{r_1}(\pi/2)^{r_2}(8/\pi)^{r_2}\,\mathrm{Vol}_L\\ &= 2^{r_1+2r_2}\,\mathrm{Vol}_L\\ &= 2^n\,\mathrm{Vol}_L\,.\end{aligned}$$

Note that each X_t is compact, convex (i.e., it contains the line segment joining any two of its points), and symmetric around 0 (i.e., stable under negation). Hence, we are reduced to the general result below that is truly just about lattices in $\mathbf{R}^n$ for any $n > 0$ and has nothing to do with K. □

Theorem 27.9 (Minkowski convex body theorem). *If $L \subset \mathbf{R}^n$ is a lattice and $D \subset \mathbf{R}^n$ is compact, convex, and symmetric around 0 with*

$$\mathrm{Vol}(D) \geq 2^n\,\mathrm{Vol}_L$$

then $L \cap D$ contains a nonzero element.

We will prove this next time. Note that D contains 0 (an obvious property for $D = X_t$): D is non-empty due to positivity of $\mathrm{Vol}(D)$ and for $d_0 \in D$ the symmetry forces $-d_0 \in D$, so the convexity implies that the midpoint 0 of the segment joining d_0 and $-d_0$ belongs to D.

28. Minkowski's theorem and applications

We have seen that finiteness of class numbers is reduced to Minkowski's convex body theorem stated at the end of last lecture (Theorem 27.9).

Remark 28.1. Kronecker's original proof of finiteness of class numbers in 1882 came before Minkowski's invention of the geometry of numbers in 1896 (and even before Minkowski earned his PhD in 1885). The proof (written in Latin) used a pigeonhole argument instead of volumes and got a constant λ'_K depending on a choice of $\mathbf{Z}$-basis of $\mathscr{O}_K$. Typically λ'_K is much larger than λ_K, and Minkowski's geometric method has much wider significance (as we shall see).

Before proceeding with the general proof, let's see the idea illustrated when $n = 2$. Pick a $\mathbf{Z}$-basis $\{v, v'\}$ for L. Define the "half-open" parallelogram

$$P := \{tv + tv' \mid -1 \leq t, t' < 1\},$$

so $\mathrm{Vol}\,P = 4\,\mathrm{Vol}_L$. We'll treat the case $\mathrm{Vol}(D) > 4\,\mathrm{Vol}_L$ (and when treating general n we'll discuss a trick to handle the case of equality $\mathrm{Vol}(D) = 2^n\,\mathrm{Vol}_L$). The plane $\mathbf{R}^2$ is covered by disjoint translates P_α of the parallelogram P. Consider the overlaps $D \cap P_\alpha$. Only finitely many such are nonempty because D is bounded. Letting $D_\alpha \subset P$ be the translate of $D \cap P_\alpha$

back into P, at least two such subsets $D_\alpha \subset P$ must have a non-empty overlap because

$$\sum_\alpha \mathrm{Vol}(D_\alpha) = \mathrm{Vol}(D) > 4\,\mathrm{Vol}_L = \mathrm{Vol}(P).$$

Thus, for some $\alpha_0 \neq \beta_0$ there exists $\xi \in D \cap P_{\alpha_0}, \xi' \in D \cap P_{\beta_0}$ that translate to the same point in P. These points $\xi, \xi' \in D$ are *distinct* (since P_α's are pairwise distinct) yet $\xi - \xi' \in 2L$, so $0 \neq \frac{\xi - \xi'}{2} \in L$. However, $\xi \in D$ and $-\xi' \in D$ (since D is symmetric), so by convexity the entire line segment of points $t\xi + (1-t)(-\xi')$ for $0 \leq t \leq 1$ belongs to D. Setting $t = 1/2$, the midpoint $(1/2)(\xi - \xi')$ of the segment joining ξ and $-\xi'$ belongs to D, but we also saw that this belongs to $L - \{0\}$.

Remark 28.2. It may seem that the full force of convexity is not needed in this proof, since we only used that D is stable under passing to midpoints. However, via repeated bisections of a segment, stability under midpoints *in general* implies that a dense subset of the segment joining any two points of D is inside D. Thus, when D is closed it follows that the entire segment belongs to D, so stability under taking midpoints is equivalent to convexity. One might then ask where we use that D is closed. This is relevant in handling the case $\mathrm{Vol}(D) = 4\,\mathrm{Vol}_L$ which we didn't discuss above but will discuss in the treatment of general n below (and which also uses more crucially that D is bounded).

Now we give the proof of Minkowski's convex body theorem for all n. First, we explain how to reduce the case $\mathrm{Vol}\,D = 2^n\,\mathrm{Vol}_L$ to the case $\mathrm{Vol}\,D > 2^n\,\mathrm{Vol}_L$, using closedness of D. That is, *assuming Theorem 27.9 holds for all compact convex symmetric D' with* $\mathrm{Vol}\,D' > 2^n\,\mathrm{Vol}_L$ we claim it also holds for compact convex symmetric D satisfying $\mathrm{Vol}\,D = 2^n\,\mathrm{Vol}_L$.

Pick some $\varepsilon > 0$ and consider the fattened region

$$D_\varepsilon := (1+\varepsilon)\cdot D$$

that is also compact, convex, and symmetric with volume

$$\mathrm{Vol}\,D_\varepsilon = (1+\varepsilon)^n\,\mathrm{Vol}\,D > 2^n\,\mathrm{Vol}_L.$$

Thus, $D_\varepsilon \cap L \neq \{0\}$ for all $\varepsilon > 0$. A priori, $D_\varepsilon \cap L$ is *finite* since L is a lattice $\mathbf{R}^n$ and D_ε is bounded (this is easiest to see using coordinates relative to a $\mathbf{Z}$-basis of L, keeping in mind that boundedness in $\mathbf{R}^n$ is coordinate-independent: we just observe after a coordinate change to turn L into $\mathbf{Z}^n$ that a bounded subset of $\mathbf{R}^n$ has finite intersection with $\mathbf{Z}^n$).

For $\varepsilon' < \varepsilon$ clearly $D_{\varepsilon'} \cap L \subset D_\varepsilon \cap L$ as finite sets, so for sufficiently small ε all $D_\varepsilon \cap L$ coincide. Thus, since these all contain a nonzero point of L, there exists $\ell \in L - \{0\}$ belonging to $\cap_{\varepsilon>0} D_\epsilon$. But this intersection that visibly contains D is contained in the closure $\overline{D}$ of D (writing $\ell = (1+\varepsilon)d_\varepsilon$ for $d_\varepsilon \in D$ amounts to saying $(1+\varepsilon)^{-1}\ell \in D$, and these points for $\varepsilon \to 0$ converge to ℓ, forcing $\ell \in \overline{D}$), yet D is closed by hypothesis, so this intersection is equal to D and hence D meets $L - \{0\}$ as desired. (Note that we have used that D is bounded and closed, or equivalently compact.)

We conclude that it suffices to treat the case $\mathrm{Vol}(D) > 2^n \mathrm{Vol}_L$, which will proceed basically as in the case $n = 2$ that we settled already, up to some more notation. Pick a $\mathbf{Z}$-basis $\mathbf{e} = \{e_i\}$ of L so $\mathbf{R}^n$ also has $\mathbf{e}$ as an $\mathbf{R}$-basis: $\mathbf{R}^n = \oplus_i \mathbf{R}e_i$. Define the "half-open" parallelotope

$$P_{\mathbf{e}} := \{\sum_i t_i e_i \mid -1 \leq t_i < 1\},$$

so $\mathrm{Vol}\, P_{\mathbf{e}} = 2^n \mathrm{Vol}_L$. Since

$$\mathbf{R} = \coprod_{n \in \mathbf{Z}} [2n-1, 2n+1) = \coprod_{n \in \mathbf{Z}} ([-1,1) + 2n)$$

is the disjoint union of even translates of $[-1,1)$, by applying this along each coordinate direction for $\oplus \mathbf{R}e_i = \mathbf{R}^n$ we have

$$\mathbf{R}^n = \coprod_{\ell \in 2L} (P_{\mathbf{e}} + \ell).$$

By boundedness of D, the overlap

$$D_\ell := D \cap (P_{\mathbf{e}} + \ell)$$

of D with each $2L$-translate of $P_{\mathbf{e}}$ is nonempty for only finitely many $\ell \in 2L$. (This is most easily seen by working in the $\mathbf{e}$-coordinate system of $\mathbf{R}^n$, in terms of which $2L$ corresponds to $2\mathbf{Z}^n$ and $P_{\mathbf{e}}$ becomes $[-1,1)^n$.) Now, consider the translates

$$D'_\ell := -\ell + D_\ell \subset P_{\mathbf{e}}.$$

Since

$$\begin{aligned}\sum_{\ell \in 2L} \mathrm{Vol}\, D'_\ell &= \sum_{\ell \in 2L} \mathrm{Vol}(D_\ell) \\ &= \mathrm{Vol}(D) \\ &> 2^n \mathrm{Vol}_L \\ &= \mathrm{Vol}\, P_{\mathbf{e}},\end{aligned}$$

by volume reasons there exist distinct $\ell_1, \ell_2 \in 2L$ so that

$$D'_{\ell_1} \cap D'_{\ell_2} \neq \emptyset.$$

Therefore, there are points $\xi_1 \in D \cap (P_{\mathbf{e}} + \ell_1)$ and $\xi_2 \in D \cap (P_{\mathbf{e}} + \ell_2)$ such that $-\ell_1 + \xi_1 = -\ell_2 + \xi_2$. Note that $\xi_1 \neq \xi_2$ since $P_{\mathbf{e}} + \ell_1$ and $P_{\mathbf{e}} + \ell_2$ are disjoint (as $\ell_1 \neq \ell_2$). But

$$0 \neq \frac{\xi_1 - \xi_2}{2} \in D$$

by convexity and symmetry (as in the case $n = 2$), and also

$$\frac{\xi_1 - \xi_2}{2} = \frac{\ell_1 - \ell_2}{2} \in \frac{1}{2}(2L) = L.$$

Therefore, we have constructed a *nonzero* point $(\xi_1 - \xi_2)/2 \in D \cap L$, as desired.

Let's now record some further applications of Minkowski's convex body theorem, to illustrate its power. These results, unlike finiteness of class numbers, were not known prior to Minkowski's geometry of numbers.

Theorem 28.3 (Minkowski). *If $K \neq \mathbf{Q}$ then $|\operatorname{disc} K| > 1$. In particular, for $K \neq \mathbf{Q}$ there exist rational primes ramified in K.*

Kronecker was led to initially conjecture this result via analogies between rings of integers and compact Riemann surfaces. See [Samuel, Theorem 1, p. 58] for a proof.

The next finiteness result is quite striking due to not imposing any constraint on the degree over $\mathbf{Q}$:

Theorem 28.4 (Hermite). *For any $B > 0$ there are only finitely many number fields K up to isomorphism with $|\operatorname{disc} K| \leq B$.*

This result is proved in [Samuel, Theorem 3, p. 59]. As a complement, if we don't constrain the discriminant but only its prime factors *and* we bound the field degree then we get a finiteness result over any number field:

Theorem 28.5. *For F a number field, $S := \{\mathfrak{p}_1, \ldots, \mathfrak{p}_m\}$ a finite set of primes of F, and $n > 1$, the set of extensions K/F (up to isomorphism) of degree at most n ramified at most over S is finite.*

The proof of this result requires deeper input from ramification theory (such as higher ramification groups, a refinement of inertia groups), and is given in [Neukirch, Theorem 2.13, Ch III]. Ultimately, the idea of the proof is to use n and S to deduce a uniform bound on the discriminant of K over $\mathbf{Q}$ so that one can apply Hermite's theorem.

Remark 28.6. Loosely speaking, the preceding theorem says that $\mathscr{O}_{F,S}$ has "fundamental group" with strong finiteness properties (made precise using Grothendieck's theory of étale fundamental groups). This fits well into the analogy between rings of integers and compact Riemann surfaces, an analogy that involves the theory of algebraic curves and often doesn't help to prove things but can help in deciding what to try to prove!

Next time, we'll prove the Dirichlet Unit Theorem which says that the group $\mathscr{O}_K^\times$ is finitely generated with rank $r_1 + r_2 - 1$. Recall from HW5 Exercise 1(ii) that the torsion subgroup

$$(\mathscr{O}_K^\times)_{\mathrm{tor}} = \{\text{ roots of unity in } K\}$$

of $\mathscr{O}_K^\times$ is finite. Thus, it is equivalent to show that the torsion-free group $\mathscr{O}_K^\times/(\mathscr{O}_K^\times)_{\mathrm{tor}}$ is finitely generated of rank $r_1 + r_2 - 1$. The key difficulty is to show the rank is as big as $r_1 + r_2 - 1$; we'll see that finite generatedness and the rank being at most $r_1 + r_2 - 1$ are not so deep. The main task is to construct units *not* of finite order when $r_1 + r_2 - 1 > 0$ (i.e., when K is neither $\mathbf{Q}$ nor imaginary quadratic), which we will do using pigeonhole-type arguments via the geometry of numbers.

Example 28.7. For $K = \mathbf{Q}(2^{1/4})$ we have $r_1 = 2$ and $r_2 = 1$ (since there are exactly two 4th roots of 2 in $\mathbf{R}$ and one conjugate pair of non-real 4th roots of 2 in $\mathbf{C}$), so Dirichlet's Unit Theorem says $\mathrm{rk}_{\mathbf{Z}}\, \mathscr{O}_K^\times = 2$. This is exhibited by the explicit description $\mathscr{O}_K^\times = \langle \pm 1\rangle (1 + 2^{1/4})^{\mathbf{Z}} \cdot (1 + \sqrt{2})^{\mathbf{Z}}$. (The only roots of unity in K are ± 1 since K admits a real embedding.) The proof we'll give of Dirichlet's theorem is theoretical; to compute the unit group in a specific case, one needs refinements of the theoretical methods (explained in books on computational algebraic number theory).

29. The Unit Theorem

Today we finally come to the proof of Dirichlet's Unit Theorem, after we first discuss a few more illustrative examples. The proof will rest crucially on Minkowski's geometry of numbers; Dirichlet's original proof in 1846 (18 years before Minkowski was born!) involved pigeonhole arguments instead of the volume-theoretic method we will use. The modern algorithms for actually computing generators of the unit group of an explicitly given number field rest on using "lattice reduction" techniques (for finding short vectors in lattices) to make the method of proof below more explicit.

The idea for the original proof came to Dirichlet during a concert in the Sistine Chapel in Rome. He had already proved special cases (such as for cubic fields), but the general case stumped him for some years. Minkowski invented the geometry of numbers 50 years later, in 1896.

Before proving the Unit Theorem, let's see some more examples.

Example 29.1. Suppose $[K:\mathbf{Q}]=3$. Since $r_1>0$ (why?) and $r_1+2r_2=3$ we have two possibilities for (r_1,r_2), so two possibilities for $\operatorname{rk}\mathscr{O}_K^\times$:

$$\operatorname{rk}\mathscr{O}_K^\times=\begin{cases}2 & \text{if } r_1=3, r_2=0;\\ 1 & \text{if } r_1=1, r_2=1.\end{cases}$$

An instance of the first case is $\mathbf{Q}(\alpha)$ with $\alpha^3-3\alpha+1=0$. For this, one has $\mathscr{O}_K^\times=\langle\pm1\rangle\alpha^{\mathbf{Z}}(\alpha+1)^{\mathbf{Z}}$. An instance of the second case is $\mathbf{Q}(2^{1/3})$, for which $\mathscr{O}_K^\times=\langle\pm1\rangle(2^{1/3}-1)^{\mathbf{Z}}$.

Example 29.2. Consider the biquadratic extension $K=\mathbf{Q}(\sqrt{2},\sqrt{3})$. We have $r_1=4$ and $r_2=0$, so $\operatorname{rk}\mathscr{O}_K^\times=3$. There are three quadratic subfields: $\mathbf{Q}(\sqrt{2})$, $\mathbf{Q}(\sqrt{3})$, $\mathbf{Q}(\sqrt{6})$. These give units $1+\sqrt{2}, 2+\sqrt{3}, 5-2\sqrt{6}$ which can be shown to be multiplicatively independent by using the "log map" to be defined later. Thus, the group generated by these units is of finite index in $\mathscr{O}_K^\times$ (and hence anything in $\mathscr{O}_K^\times$ has a nontrivial power contained in this subgroup). This subgroup really misses some infinite-order units. For example, it misses $\sqrt{2}-\sqrt{3}$ (with square $5-2\sqrt{6}$) and $(\sqrt{2}+\sqrt{6})/2$ (with square $2+\sqrt{3}$).

Example 29.3. Let p be an odd prime and $K=\mathbf{Q}(\zeta_p)$. Inside K the subfield $K^+:=\mathbf{Q}(\zeta_p+\zeta_p^{-1})$ equals the fixed field E of complex conjugation (as K^+ is certainly inside that "index-2" subfield E yet $[K:K^+]\le 2$ since ζ_p is easily seen to satisfy a quadratic relation over K^+, so $[K:K^+]=2$ and $K^+=E$).

Thus, $r_1^K=0, r_2^K=\frac{p-1}{2}$ whereas $r_1^{K^+}=\frac{p-1}{2}, r_2^{K^+}=0$, so $\mathscr{O}_{K^+}^\times\subset\mathscr{O}_K^\times$ has finite index. This generalizes with $\mathbf{Q}(\zeta_p)$ replaced by any CM field (where "CM" stands for *Complex Multiplication*, an important notion in the theory of abelian varieties).

To prove the Unit Theorem, we will first turn our multiplicative problem into an additive problem via the so-called *log map*

$$\mathscr{L}:\mathscr{O}_K^\times\subset K^\times\xrightarrow{\theta_K}(\mathbf{R}^\times)^{r_1}\times(\mathbf{C}^\times)^{r_2}\xrightarrow{\ell}\mathbf{R}^{r_1+r_2}$$

where $\ell\colon(\mathbf{R}^\times)^{r_1}\times(\mathbf{C}^\times)^{r_2}\to\mathbf{R}^{r_1+r_2}$ is defined by

$$(x_1,\dots,x_{r_1},z_1,\dots,z_{r_2})=:(x,z)\mapsto(\log|x_1|,\dots,\log|z_1|^2,\dots)$$

(output to be denoted below by the shorthand $(\log|x|,\log|z|^2)$). The composite map $\mathscr{L}$ is therefore given by

$$\begin{aligned}\mathscr{L}\colon\mathscr{O}_K^\times&\to\mathbf{R}^{r_1+r_2}\\ u&\mapsto(\log|\sigma(u)|,\log|\tau(u)|^2)\end{aligned}$$

where σ varies through the r_1 real embeddings and τ varies through the r_2 non-real embeddings taken up to complex conjugation. Further, since $|\tau(u)|^2 = |\tau(u)|\,|\overline{\tau}(u)|$, the sum of the coordinates of $\mathscr{L}(u)$ is precisely $\log|N_{K/\mathbf{Q}}(u)| = \log 1 = 0$ (using that $\mathrm{N}_{K/\mathbf{Q}}(\mathscr{O}_K^\times) \subset \{\pm 1\}$).

Hence, we obtain

$$\mathscr{L}(\mathscr{O}_K^\times) \subset H := \{t \in \mathbf{R}^{r_1+r_2} \mid \sum_{j=1}^{r_1+r_2} t_j = 0\} \subset \mathbf{R}^{r_1+r_2}$$

with $\dim H = r_1 + r_2 - 1$. No other linear dependence relation on the image of $\mathscr{L}$ leaps to mind (beyond $\sum t_j = 0$), so we aim show it is a lattice in H.

Example 29.4. Let $K = \mathbf{Q}(\alpha)$ with $\alpha^2 = 2$, so $\mathscr{O}_K^\times = \langle \pm 1 \rangle (1+\alpha)^{\mathbf{Z}}$. We have $\theta \colon K \to \mathbf{R}^2$ via $\alpha \mapsto (\sqrt{2}, -\sqrt{2})$, so $\theta(a + b\alpha) = (a + b\sqrt{2}, a - b\sqrt{2})$ for $a, b \in \mathbf{Q}$. Here, the log map $\ell \colon \mathbf{R}^\times \times \mathbf{R}^\times \to \mathbf{R}^2$ is

$$(x, y) \mapsto (\log|x|, \log|y|).$$

The image $\theta_K(\mathscr{O}_K^\times) \subset \mathbf{R}^\times \times \mathbf{R}^\times$ lies inside $\{|x||y| = 1\} = \{xy = \pm 1\} \subset \mathbf{R}^2$ as a discretely-placed set of points, and $\mathscr{L}(\mathscr{O}_K^\times)$ is a lattice in the line $\{t_1 + t_2 = 0\} \subset \mathbf{R}^2$, as shown in Figure 2.

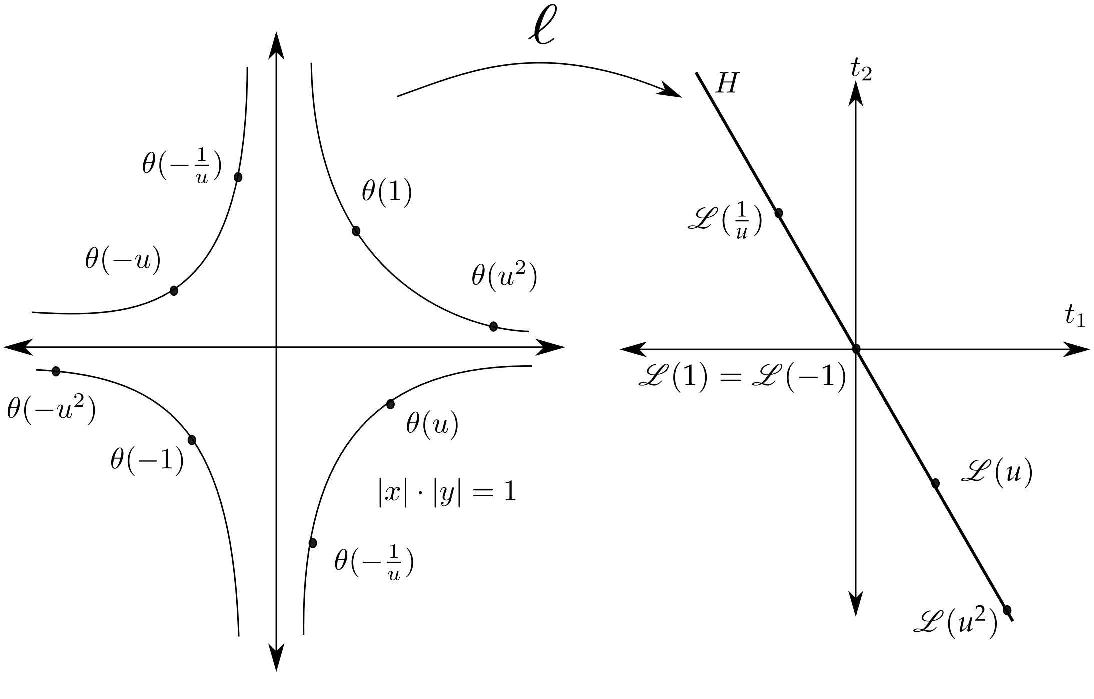

FIGURE 2. Picture showing $\theta_K(\mathscr{O}_K^\times) \subset \{|x||y| = 1\}$ and ℓ carrying it into the line $H = \{t_1 + t_2 = 0\}$ in $\mathbf{R}^2$.

For later use, we define the *norm-1 hypersurface*

$$\Sigma := \ell^{-1}(H) = \{(x,z) \in \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} \mid \prod_i |x_i| \cdot \prod_j |z_j|^2 = 1\}$$

(equivalently, this is the set of points $v \in \mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$ such that $|\mathcal{N}(v)| = 1$). Note that ℓ carries Σ onto H. In the handout "Generalized Pell Equation" we use the log map to bound any possible $\mathbf{Z}$-solution to $x^2 - dy^2 = n$ for any given $n \in \mathbf{Z} - \{0\}$ when such solutions are taken up to $\mathbf{Z}[\sqrt{d}]^\times$-scaling. The aim is to slide along the norm-n hypersurface via $\theta_K(\mathbf{Z}[\sqrt{d}]^\times)$-scaling to get oneself into some specific bounded region in the plane, and then do a search through the finitely many $\mathbf{Z}$-points in that bounded region; the handout illustrates how it goes.

Let's first see that the log map does not lose a lot of information:

Lemma 29.5. *The containment $(\mathscr{O}_K^\times)_{\rm tor} \subset \ker \mathscr{L}$ is an equality. In particular, $\ker \mathscr{L}$ is finite.*

Proof. If $\zeta \in K$ satisfies $\zeta^m = 1$ for some $m > 0$ then $\mathscr{L}(\zeta) = 0$ since all the roots of unity in $\mathbf{C}$ clearly lie on the unit circle. To show the reverse inclusion, suppose $\mathscr{L}(u) = 0$, so $|\sigma(u)| = 1$ for all $\sigma : K \to \mathbf{C}$. This is equivalent to all $\mathbf{Q}$-conjugates of u in $\mathbf{C}$ lying in the unit circle. Hence, since $[\mathbf{Q}(u) : \mathbf{Q}] \leq [K : \mathbf{Q}]$, there is a uniform bound on the $\mathbf{Z}$-coefficients of the minimal polynomial f over $\mathbf{Q}$ of u, so there are only *finitely many* possibilities for f and hence for such u. But all u^n satisfy the *same* bounds (they're all killed by $\mathscr{L}$ too!), so $u^{\mathbf{Z}}$ is finite and hence u is a root of unity. The finiteness of $\ker \mathscr{L}$ follows because there are only finitely many roots of unity in K by HW 5, Exercise 1. □

By Lemma 29.5 we know $\ker \mathscr{L}$ is finite, so for the Unit Theorem we are reduced to studying $\mathscr{L}(\mathscr{O}_K^\times) \subset H$. We can get an initial handle on this image using the following soft topological input proved in [Samuel, §4.1]:

Lemma 29.6. *If $\Gamma \subset \mathbf{R}^N$ is a subgroup that is discrete in the sense that Γ has finite intersection with any bounded region in $\mathbf{R}^N$ then Γ is finitely generated with any $\mathbf{Z}$-basis linearly independent over $\mathbf{R}$. In particular $\mathrm{rk}_{\mathbf{Z}}\, \Gamma \leq N$.*

Beware that $\mathbf{R}^N$ can have finitely generated subgroups with rank $> N$; they are not discrete. For example, $\mathbf{Z} \cdot 1 + \mathbf{Z} \cdot \sqrt{2} \subset \mathbf{R}$ is *dense* (see the handout "Density for the Ring of Integers" for a version with any number field). Working out the proof of Lemma 29.6 for $N = 1$ by yourself is instructive.

Bounding the rank from above. If Lemma 29.6 applies to $\Gamma := \mathscr{L}(\mathscr{O}_K^\times) \subset H$ then Γ (and hence $\mathscr{O}_K^\times$) is finitely generated with rank at most $\dim H =$

$r_1 + r_2 - 1$. It will be much deeper that the rank is equal to $r_1 + r_2 - 1$, for which we shall use the geometry of numbers.

To establish the discreteness of Γ in H, it suffices to show the overlap

$$\Gamma \cap ([-B, B]^{r_1} \times [-2B, 2B]^{r_2})$$

with big boxes is finite for any $B > 0$. Indeed, this would imply Γ is discrete in $\mathbf{R}^{r_1} \times \mathbf{C}^{r_2}$, and hence also in the subspace H (so Lemma 29.6 applies).

For any $\gamma := \mathscr{L}(u) \in \Gamma$ in this box, $-B \leq \log|\sigma(u)| \leq B$ for $\sigma : K \to \mathbf{R}$ and $-2B \leq \log|\tau(u)|^2 \leq 2B$ for non-real $\tau : K \to \mathbf{C}$, so $|\sigma(u)| \leq e^B$ for *all* field embeddings $\sigma : K \to \mathbf{C}$. This uniformly bounds the $\mathbf{Z}$-coefficients of the minimal polynomial over $\mathbf{Q}$ for u (as $[\mathbf{Q}(u) : \mathbf{Q}] \leq [K : \mathbf{Q}]$ for all such u), and hence permits only finitely many such u.

The idea for establishing maximal rank. To complete the proof of the Unit Theorem, we need to show the discrete subgroup $\Gamma \subset H$ has full rank; equivalently, the $\mathbf{R}$-span $\mathbf{R}\Gamma$ exhausts H (recall that we know a $\mathbf{Z}$-basis of Γ is linearly independent over $\mathbf{R}$ by Lemma 29.6, so $\dim_{\mathbf{R}} \mathbf{R}\Gamma = \mathrm{rk}_{\mathbf{Z}}\,\Gamma$).

The idea is to find some compact $\Delta \subset H$ with

$$H = \cup_{\gamma \in \Gamma}(\gamma + \Delta).$$

If we could do this, the compact region Δ gives us a set of representatives for H/Γ, implying we have a continuous surjective map $\Delta \twoheadrightarrow H/\Gamma \twoheadrightarrow H/\mathbf{R}\Gamma$. But then $H/\mathbf{R}\Gamma$ is a compact $\mathbf{R}$-vector space, forcing it to vanish, which is to say $H = \mathbf{R}\Gamma$ as desired. This argument highlights the important fact that one should regard the Unit Theorem as a compactness result (for H/Γ).

Proving the lower bound for the rank of the units. Now let's now discuss how to find Δ. Recall that the norm-1 hypersurface

$$\Sigma = \{(x, z) \in \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} \mid \prod_i |x_i| \cdot \prod_j |z_j|^2 = 1\}$$

admits a surjection $\ell : \Sigma \twoheadrightarrow H$ with $H \supset \Gamma = \mathscr{L}(\mathscr{O}_K^\times) = \ell(\theta_K(\mathscr{O}_K^\times))$. It is enough to find some compact $\Delta' \subset \Sigma$ so that

$$\Sigma = \cup_{\varepsilon \in \mathscr{O}_K^\times}(\Delta' \cdot \theta_K(\varepsilon)),$$

as then applying ℓ to this equality does the job using $\Delta = \ell(\Delta')$.

For $\xi \in (\mathbf{R}^\times)^{r_1} \times (\mathbf{C}^\times)^{r_2}$ consider the associated region

$$D := \{(x, z) \in \mathbf{R}^{r_1} \times \mathbf{C}^{r_2} \mid |x_i| \leq |\xi_i|, |z_j| \leq |\xi_j|\}.$$

Since the compact convex symmetric D is a product of r_1 closed intervals and r_2 closed discs centered at 0, its volume is $2^{r_1}\pi^{r_2}C$ for $C := |\mathscr{N}(\xi)| = \prod|\xi_i|\prod|\xi_j|^2$. Hence, choosing ξ to make C big enough ensures we can use

Minkowski's theorem to find a nonzero element of the lattice $\theta_K(\mathscr{O}_K)$ belonging to D, which is to say D contains $\theta_K(a)$ for some $a \in \mathscr{O}_K - \{0\}$.

For any $v \in \Sigma$, the domain $D \cdot v$ (using component-wise multiplication) can be described exactly like D but using the component-wise product $\xi \cdot v$ in place of ξ, so it has the same volume since $|\mathcal{N}(\xi v)| = |\mathcal{N}(\xi)||\mathcal{N}(v)| = C \cdot 1 = C$. Hence, $D \cdot v$ contains $\theta_K(a_v)$ for some $a_v \in \mathscr{O}_K - \{0\}$. The membership $\theta_K(a_v) \in D \cdot v$ implies

$$\left|\mathrm{N}_{K/\mathbf{Q}}(a_v)\right| \leq \prod |\xi_i| \prod |\xi_j|^2 = C,$$

so the nonzero principal ideals $(a_v) \subset \mathscr{O}_K$ have bounded norm as we vary through all uncountably many $v \in \Sigma$. There are only *finitely many* nonzero ideals of $\mathscr{O}_K$ with norm below a given bound (such as C), so the collection of principal ideals (a_v) must involve *a lot* of repetition as we vary v (pigeonhole principle!). Every time $(a_v) = (a_{v'})$ for $v, v' \in \Sigma$, we have $a_{v'}/a_v \in \mathscr{O}_K^\times$; this is a way to produce units (and is Dirichlet's key idea).

If one is careful with this repetition process, the units obtained in this way turn out to be sufficiently abundant to multiplicatively slide each point $v \in \Sigma$ into a specific compact subset $\Delta' \subset \Sigma$. The precise definition of Δ' and the proof that it works in this way are addressed in the handout "Minkowski Step with Units".

References

[Neukirch] J. Neukirch, *Algebraic Number Theory*, Springer-Verlag, 1999.
[Samuel] Pierre Samuel, *Algebraic Theory of Numbers*, Dover, 2013 reprint of 1970 original.

Department of Mathematics, Stanford University, Stanford, CA 94305
E-mail address, Brian Conrad: `conrad@math.stanford.edu`

Department of Mathematics, Stanford University, Stanford, CA 94305
E-mail address, Aaron Landesman: `aaronlandesman@stanford.edu`